Integrating Histology and Microarchitecture Modelling with Deep Learning for Diffusion-Weighted Magnetic Resonance Imaging

By

Vishwesh Nath

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Computer Science

May 8th, 2020

Nashville, Tennessee

Approved:

Bennett A. Landman, Ph.D.

Adam W. Anderson, Ph.D.

Catie Chang, Ph.D.

Ipek Oguz, Ph.D.

Larry T. Davis, M.D

Holger Roth, Ph.D.

Acknowledgements

This is a snapshot of my experience and an expression of gratitude towards those who have helped to shape it to be so. I am indebted to Dr. Bennett Landman for imbibing me with a life-long curiosity to think and develop ideas that hold a vision to help mankind. The door was always open for enlightening discussions and late-night email chains were always responded to. These cannot be forgotten, and I always hope you keep the same energy forever Bennett. I often reminisce back to the day of our first meeting where I didn't really understand where I was headed with just a thought of 'a little research'. Thank you for being a friend, mentor and above all a *true* guru. I have always appreciated the fact that you've pushed me beyond what I thought possible of my abilities and not so when I didn't need it. You've always been there to support me academically and in troubled times of my life. I could not have asked for a better advisor.

A word of gratitude towards Dr. Adam Anderson, just your attention and interest in my research is an honor for me. I absolutely enjoyed our Friday afternoon time every week, which we spent on the whiteboard, culminating ideas at the Vanderbilt Imaging Institute basement conference room. I deeply appreciate the mentoring that I've received from Dr. Holger Roth during my time at Nvidia, you've led me to be critical and thorough in formulation for my own ideas. I'm thankful for having the kindest committee members Dr. Catie Chang, Dr. Ipek Oguz and Dr. Larry T. Davis, you've kept me inspired and motivated throughout this amazing journey.

Academic support aside, I would not have never made if I had not been blessed with three sets of parents: Vikram & Sangita, Jared & Nina, Deepak & Shailja. For they were there with me on those days where I could not see the rays of hope coming in from any direction. To all the worry that I've caused because I was out late at night discovering things that were beyond the present times and I hope this document is proof enough to clear up misconceptions. I do not think this would have been possible without the unwavering support and concern that my family has shown me during my time as a student. Last but not the least thanks for being there Varad, you've always been by my side.

It’s been an absolute pleasure working in tandem with all the folks at the MASI Lab. I owe if not all but some of my coding practices to Justin who taught me enough to last, through my tenure as a student. Kurt you have been an amazing source of motivation and diffusion theory, you kept me going. I would specifically like to mention all diffusion folks (Prasanna, Colin, Allison, Justin, and Kurt) of our lab thank you for listening to me and simultaneously enlightening me with your own experiences. It indeed seems trivial to mention coffee, but it was the enabler and great company from fellow coffee connoisseurs: Cailey (Coffee Fairy) and Camilo. My ride with all of you adjoined with casual banter and camaraderie has been amazing and I hope that we can still stay connected (Shikha, Sandra, Gaile, Roza, Yuankai, Andrew, Rob, Ilwoo, Samuel, Shunxing, Riqiang, Yucheng, Leon, Qi, Kaiwen and Peter). Last but not the least thank you dear engineers (Justin, Josh, Karthik, PK, Stephen) for you’ve kept the systems up and alive through the nights and dealt with all the messy compute clusters so that I could play my part.

No story is complete without friends and mine is no different I’ve always had the support from Chandan, Amogh, Ayan & Sahil. You all are the first set of friends that I made at Vanderbilt. Thank you for putting up with my idiosyncrasies and for enriching me with warm company that has made this experience even better. To Karthik, Roslin, and Niranjana who I met along the way, I want to say thank you for letting me discover myself in your company. Brief bouts of musical outbreaks, hikes, road trips and comfortable silences were my sources of relaxation and I could write more on this, but I’ll save it for another day.

My being at Vanderbilt is due to multiple reasons: primary was my will to be here, but the ability was given by the spirit that I imbibed at Manipal. I would like to extend my gratitude and thanks to my prior mentors there (Dr. Savitha G. Kini, Dr. Sandhyalaxhmi G Navada and Dr. Shailesh Kumar) I would not have made it to Vanderbilt in the first place had it not been there for their support. The time that I spent back there was massively shaped by my fellow peers and comrades who are still very dear to me. A salute to them who truly made my time worthwhile at Manipal: Tamaghna, Surabhi, Siddesh, Nirosha, Aarushi, Rajarshi, Aanchal, Ishwant and Ankita.

At the end to brothers for life: Sheru, Kshitu, Jain Bhai & Vibhor Da because you guys were there

from the days that I can no longer clearly remember.

Though I've only been able to barely scratch the surface of the core of science. I feel incredibly fortunate to have been able to contribute even if only to an extremely small subset, for I only strongly believe in what I read: "*Opportunities multiply as they are seized*" - The Art of War by Sun Tzu

Table of Contents

List of Tables

Page

List of Figures

Page

List of Abbreviations

ACC: Angular Correlation Coefficient

ADC: Apparent Diffusion Coefficient

AI: Artificial Intelligence

ANTs: Advanced Normalization Tools

BLSA: Baltimore Longitudinal Study of Aging

CE: Containment Energy

CI: Containment Index

CNG: Cingulum

CNN: Convolutional Neural Network

CSD: Constrained Spherical Deconvolution

CSDLR: Lucy-Richardson Constrained Spherical Deconvolution

CSF: Cerebrospinal Fluid

CST: Corticospinal Tract

CT: Computed Tomography

DOT: Diffusion Orientation Transform

DSI: Diffusion Spectrum Imaging

DSC: Dice Sorensen Coefficient

DTI: Diffusion Tensor Imaging

DW-MRI: Diffusion-Weighted Magnetic Resonance Imaging

EPI: Echo-Planar Imaging

FA: Fractional Anisotropy

FACT: Fiber Assignment through Continuous Tracking

FNX: Fornix

FOD or fODF: Fiber Orientation Distribution Function

FSL: Free Surfer Library Toolkit

GM: Grey Matter

GFA: General Fractional Anisotropy

GQI: Generalized Q-Sampling

HARDI: High Angular Resolution Diffusion Imaging

HCP: Human Connectome Project

HDFT: High Definition Fiber Tractography

HFOD: Histology Fiber Orientation Distribution

HOMOR: Higher Order Model Outlier Rejection

ICC: Intra Class Correlation

ICVF: Intra-Cellular Volume Fraction

IFO: Inferior Fronto-Occipital

ILF: Inferior Longitudinal Fasciculus

IRB: Institutional Review Board

IRLLS: Iteratively Reweighted Linear Least Squares

MD: Mean Diffusivity

MESD: Maximum Entropy Spherical Deconvolution

MNI: Montreal Neurological Institute

MSE: Mean Squared Error

MT-CSD: Multi-Tissue Constrained Spherical Deconvolution

NODDI: Neurite Orientation Dispersion and Density Imaging

NSDN: Null Space Deep Network

ODI: Orientation Dispersion Index

PAS-MRI: Persistent Angular Structure Magnetic Resonance Imaging

PET: Positron Emission Tomography

PF: Peak Fraction

PGSE: Pulsed Gradient Spin Echo

PICo: Probabilistic Index of Connectivity

QBI: Q-ball Imaging

RESTORE: Robust Estimation of Tensors by Outlier Rejection

ResDNN: Residual Deep Neural Network

ResCNN: Residual Convolutional Neural Network

RISH: Rotationally Invariant Spherical Harmonics

RMSE: Root Mean Squared Error

ROI: Region of Interest

SAE: Symmetric Angular Error

sCSD: Super-Resolved Constrained Spherical Deconvolution

SF: Success Fraction

SH: Spherical Harmonics

SHORE: Simple Harmonic Oscillator Reconstruction

SHARD: Spherical Harmonic based Robust Outlier Detection

SLF: Superior Longitudinal Fasciculus

SMT: Spherical Mean Technique

SNR: Signal to Noise Ratio

SoH: Solid Harmonics

SyN: Symmetric Image Normalization Algorithm

UNC: Uncinate

WM: White Matter

1. Introduction

1.1. Overview

Most voluntary and involuntary actions that a human individual performs everyday are governed by the human brain. The brain and the spinal cord formulate the core nervous system in the human body which formulate/execute the actions. Currently, diffusion-weighted magnetic resonance imaging (DW-MRI) is the only imaging modality that holds the potential to capture the microstructural tissue properties in-vivo (within living) at an intra-voxel millimetric scale for the human brain. DW-MRI holds the potential to transform our understanding of neurological disorders and the capacity to help in healthcare. Throughout the life span the brain alters in various ways [1] and evolution has led for humans to have a longer life span [2]. Increased life span has been directly correlated with an expectation of higher cases of dementia related disorders for e.g. Alzheimer's disease [3]. Number of cases reported for Alzheimer's in 2015 was 5.3 million in USA alone [4]. Multiple Sclerosis in US itself accounts for over 400,000 cases with approximately 10,000 new cases being received every year [5, 6]. Current global review studies suggest that for every 10,000 children 62 are diagnosed with autism spectrum disorder [7].

Prevalence of neurological disorders has led scientists/researchers to conduct longitudinal studies, connectome-based studies to characterize governing factors between control and impaired subjects. The Human Connectome Project (HCP) [8, 9] is one such popular study with over 2400 acquired scans with a systematic protocol focusing on high quality and same subject repeat data. The Baltimore Longitudinal Study of Aging (BLSA) [10, 11] is more focused towards characterization of long term effects of neurological disorders, the data acquisition started in 1958 with over 1000 participants aged between 17 to 96 years. The imaging acquisitions began in the year 1977 [10]. The protocols that govern imaging modalities (e.g. MRI) are being acquired consistently with more variants of modalities being added as technological advancements progress. Due to the successive changes of sites and scanner hardware upgrades, it has become a critical processing step to harmonize the DW-MRI acquisitions to support large

scale statistical analysis. DW-MRI holds great potential for developing our understanding of neurological disorders large scale studies from multiple cohorts of population need to be harmonized for meaningful interpretations.

The term ‘harmonization’ can be defined as a method for ‘allowing data to be meaningfully compared when there are multiple sources of data’. Harmonization is necessitated so that models that are fit to DW-MRI do not get affected by the bias and variance of the scanner sites and hardware upgrades. Harmonization will allow for an unbiased or reduced bias estimates of the tissue properties, for which a plethora of advanced methods/models exist currently. Due to the complex acquisition schemes of DW-MRI the advanced methods are applicable for estimation of microstructure tissue properties.

The microstructure/microarchitecture models can be classified into two major categories: orientation models and compartment models. A few popular orientation models are persistent angular structure (PAS-MRI) [12], Q-ball imaging (QBI) [13], diffusion orientation transform (DOT) [14], constrained spherical deconvolution (CSD) [15]. The well-known compartment models in use are neurite orientation dispersion and density imaging (NODDI) [16] and spherical mean technique (SMT) [17] etc. The orientation models provide information about the geometrical microstructure at an intra-voxel level and the compartmental models provide us with fractional estimates of the tissue that constitutes the voxel. In general, the orientation methods are often collectively referred as high angular resolution diffusion imaging (HARDI) methods [18]. The HARDI methods are described in a more elaborate fashion in section 3.2.

The diffusion microstructure/microarchitecture methods have showed varied reproducibility when compared intra-method, inter-method, intra-session, inter-session, inter-scanner; where intra-method and intra-session exhibit high reproducibility and the others exhibit low reproducibility [19, 20]. In the context of precision/accuracy for HARDI methods and compartment methods, there are three possible validations 1.) phantom based studies, 2.) histology-based validation (animal studies) and 3.) scan-rescan studies of human in-vivo acquisitions. Although the methods are primarily intended for applicability on humans it is exceedingly difficult to perform human histology validation due to lack of such datasets. Phantom studies

have indicated sound validation of most of the methods [21, 22]. However, the same is not true for histology validation using datasets from animal studies [23]. This indicates there is scope for development of new microstructure methods that are focused towards reproducibility specifically towards inter-session, inter-scanner and show a high precision for histology-based validation[23, 24]. Earlier proposed models have not shown high correlation with the same metrics being derived from histology [25]. The most widely used methods are classical mathematical transforms, and machine learning to drive these models has only been mildly explored [25, 26] to improve reproducibility and show a higher correlation. These gaps lay out the foundation of this dissertation. Briefly, the gaps for microstructure modelling can be segregated into 1.) highly reproducible and harmonized estimates of microstructure tissue properties and 2.) microstructure estimates that show high correlation with histology-based tissue estimates.

This chapter is segregated into different sections. Section 2 covers the fundamental information regarding DW-MRI, hyper-parameters of DW-MRI, and its applications. Section 3 delves into the understanding of microstructural measures from imaging modalities and their associative synergy with histological measures. In section 4, we cover the majority of the harmonization and reproducibility methodologies involved for the neuroimaging data in section 4. The section covers how the acquisitions were normalized and registered collectively for analysis. Next in section 5 we cover the basics of deep learning and how it can be applied to DW-MRI; covered in section 5. Section 6 outline's the clinical applicability of these approaches and outlines this dissertation into contributions.

As a part of this dissertation, we lay out the characterization of empirical reproducibility in the context of HARDI methods and tractography to understand the strengths and the fallacies of prior established work from the literature. Further, we propose novel methods to reconstruct microarchitecture from external validation data (animal studies). The external validation data showed promising results which motivated us to utilize them for proposal of multiple harmonizing approaches for microstructure

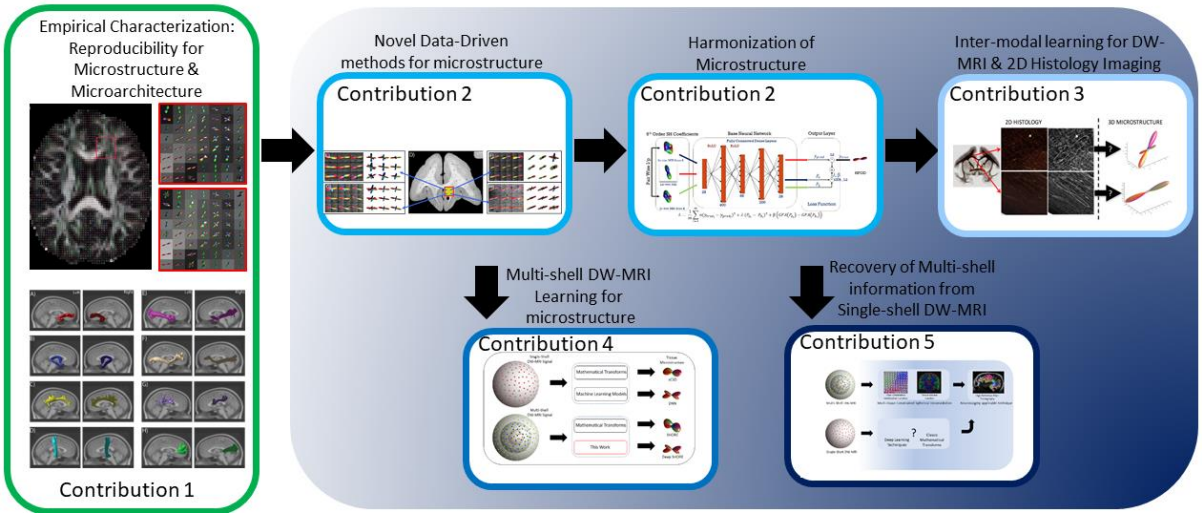

Figure I.1. Overview of the contributions proposed in this dissertation. Contribution 1: was to evaluate empirical reproducibility for microstructure in terms of HARDI methods and tractography. Contribution 2: Proposition of novel data-driven methods using rare animal histology datasets. Contribution 3: Inter-modal learning to recover 3D microstructure from 2D microscopic imaging. Contribution 4: Scanning acquisition parameter invariant learning manifolds for generalizability of data-driven learning techniques. Contribution 5: Recovery of multi-shell DW-MRI microstructural measures from single-shell DW-MRI

measurement. In parallel, we also show that recovery of 3D microstructural measures from 2D histology imaging is possible.

We briefly outline all the contributions that have been performed as a part of this dissertation. A brief overview for each contribution can also be observed in Figure I.1.

1.1.1. Contribution 1: Empirical characterization of reproducibility of microstructure

Due to the existence of a plethora of microstructure modelling techniques and multiple tractography reconstructions, a reproducibility evaluation was necessitated. Hence, we empirically characterized microstructure modelling methods using single subject large scale in-vivo acquired data. The acquired data

had multiple repeats per session with repetitions of sessions and the same was acquired on different scanners for evaluation.

Two well-known HARDI models were chosen (Q-ball and PAS-MRI) because they both belong to the same category of orientation tissue models; however, they are at the extreme ends of a spectrum in terms of anisotropic nature with Q-ball showing smooth angular structure while PAS-MRI recovering sharper angular structure. The two methods follow a fundamentally different classical transformation for modelling. This study was performed on one of the largest single subject datasets with a total number of more than 5400 brain volumes. The study design focused on intra-method and inter-method comparison. The metrics that were used for characterization were symmetric angular error, peak fraction and general fractional anisotropy. Both the methods showed a high level of disagreement when they were inter-compared for the same subject.

The global tractography (geometrically connected tract micro-structure) reproducibility challenge was conducted on a single subject acquisition acquired across two different scanners, with over 1800 brain volumes per scanner. The data was acquired in two different sessions per scanner and hence the tractographic algorithms were assessed at three levels of intra-session, inter-session and inter-scanner reproducibility. As a final outcome given all algorithms the variance was quite high for all the tracts. However, considering the top five submissions measured by intraclass correlation coefficient and dice similarity measure 8 of the tracts showed high reproducibility, 4 showed moderate and 4 showed low reproducibility.

To understand the underlying challenges associated with studies of microstructure and microarchitecture, empirical characterization was the first step that we took. There are multiple HARDI methods. For an intricate evaluation, we choose two methods PAS-MRI and Q-ball to study for intra and inter method disagreement. The contribution is covered in greater detail in chapter II and III. As a pre-processing technique, we also explored if the reproducibility could be improved by proposal of an outlier detection and imputation technique covered in chapter IV.

The moniker of microarchitecture begins with microstructure methods however, it was as critical

to assess the reproducibility of tractography. To address this, we conducted a global challenge of for the assessment of 16 different tracts that depict the neural pathway structure of the human brain. Nine Teams submitted a total of 46 different algorithm pipelines that could reconstruct the streamlines. The outcome is discussed in greater detail in chapter V.

1.1.2. Contribution 2: Novel methods to improve precision and reproducibility of microstructure using single shell DW-MRI acquisitions

Validation studies are often performed using corresponding histology with DW-MRI measures. We explore the potential of an external validation dataset to drive data-driven machine learning approaches.

The venture led us to explore a harmonized reconstruction of the microstructure by utilizing paired in-vivo data from different scanners of the same subject. The contribution is further defined by inclusion of scalar metrics as loss so that a joint harmonization framework is formed.

We inferred that the current approaches were limited when histological validations were assessed. Hence, we proposed a novel deep learning method that could recover the microstructure using external validation data. This approach is covered in more detail in chapter VI. The approach motivated us to harmonize the scanner site effects and propose a framework technique using deep learning that could predict harmonized microstructure. The framework is discussed more elaborately in chapter VII and VIII.

1.1.3. Contribution 3: Combining DW-MRI with 2D imaging to predict 3D measures from 2D imaging.

Integration of DW-MRI and histology for microstructure reconstruction has encouraged us to think about applications where there is a lack of DW-MRI acquisition measurement and only 2D imaging exists. There are no known ways to extract 3D microstructure information from 2D imaging. Hence this contribution is focused towards obtaining 3D microstructural measures from 2D imaging acquisitions.

While contribution 2 led us to proposal and implementation of harmonizing techniques, it simultaneously depicted a pathway to a new challenge where the 3D microstructural measures could not be derived from 2D imaging. Using rare datasets, we propose techniques that can recover 3D measures using 2D imaging. This work is covered in more detail in chapter IX and X. We hope that this work will find

applicability where researchers are faced with data where only 2D imaging exists, for example postmortem brain studies based on histology.

1.1.4. Contribution 4: Learning across the manifolds of single and multi-shell DW-MRI for advanced 3D microstructure reconstruction

Typical clinical DW-MRI scan acquisitions are based on single shell DW-MRI data with a limited number of gradient directions. The promising results from the prior work were limited in application to advanced multi-shell DW-MRI acquisitions where the scanning parameters could vary in number of diffusivity shells, varying diffusivity values and varying number of gradient directions per diffusivity shell. Therefore, we proposed and developed a learning manifold where the data could have varying scanning parameters.

Such methodologies will be critical for the new wave of research studies that are being conducted for a richer understanding of the DW-MRI based microstructure. Research institutes often collect advanced DW-MRI data where there are two or more diffusivity shells. A typical example is also of the Human Connectome Project where three shells of data per scan acquisition were collected.

Contribution 2 opened the possibilities for the applicability/utility for reconstruction of fiber orientation based microstructural measures, yet it was limited only towards operating on single shell DW-MRI. Although, the typical clinical acquisitions are single-shell hardware advancement in the near future will allow for multi-shell DW-MRI at clinical level. This led us to propose the learning space manifold for multi-shell DW-MRI for reconstruction of fiber orientational measures. The proposed methods were trained and validated extensively on a rare histology with corresponding DW-MRI dataset. The proposed method was tested based on reproducibility measures with in-vivo imaging of the normal human brain.

1.1.5. Contribution 5: Recovery of microstructural measures based on multi-shell DW-MRI from single shell DW-MRI

The second and fourth contribution paved the path for effective learning and reconstruction of microstructural measures across multiple varying parameters. However, an interesting

opportunity/hypothesis left to explore was the recovery of multi-shell DW-MRI from single shell DW-MRI. The existing data from clinical studies primarily was acquired with single shell DW-MRI and in present times is still a clinical standard. Hence, the hypothesis of recovery of advanced multi-shell DW-MRI microstructural measures from single shell DW-MRI.

We show that such a recovery is possible and has more information content than prior existing single shell techniques. This contribution is specifically useful for existing longitudinal studies of aging where a vast amount of clinical single shell DW-MRI data has been collected to study effects of aging and longitudinal effects of neurological disorders.

Contribution 4, while applicable for advanced multi-shell DW-MRI acquisitions, is currently limited because of the advanced acquisition scheme and this led us to the idea to explore the possibility of recovery of multi-shell DW-MRI information from single shell DW-MRI. The majority of the current existing data via clinical studies [10] being conducted are of clinical nature and typically have a single shell DW-MRI acquisition with less data.

Therefore, we developed advanced data-driven methods to recover multi-shell information from single shell. The contribution is also intended towards lowering the acquisition time further as it quite often it is not possible for a subject suffering from a neurological condition to be scanned for a longer period of time.

The scope this dissertation is outlined with the following objectives:

1.) Empirical characterization of reproducibility of microstructure
2.) Novel methods to increase precision and reproducibility of microstructure using single shell DW-MRI acquisitions
3.) Combining DW-MRI with 2D imaging to predict 3D measures from 2D imaging.
4.) Reconstruct reproducible and precise microstructure using multi-shell DW-MRI acquisitions Multi-shell micro-architecture reconstruction using single shell DW-MRI imaging.

1.2. Diffusion-Weighted MRI for Neuroimaging

1.2.1. Diffusion-Weighted MRI Acquisition Parameters

Achieving the right target contrast is the target for all medical imaging procedures. The well-known contrasts of T1, T2 and FLAIR provide an elaborate comprehension of the understanding of the connected

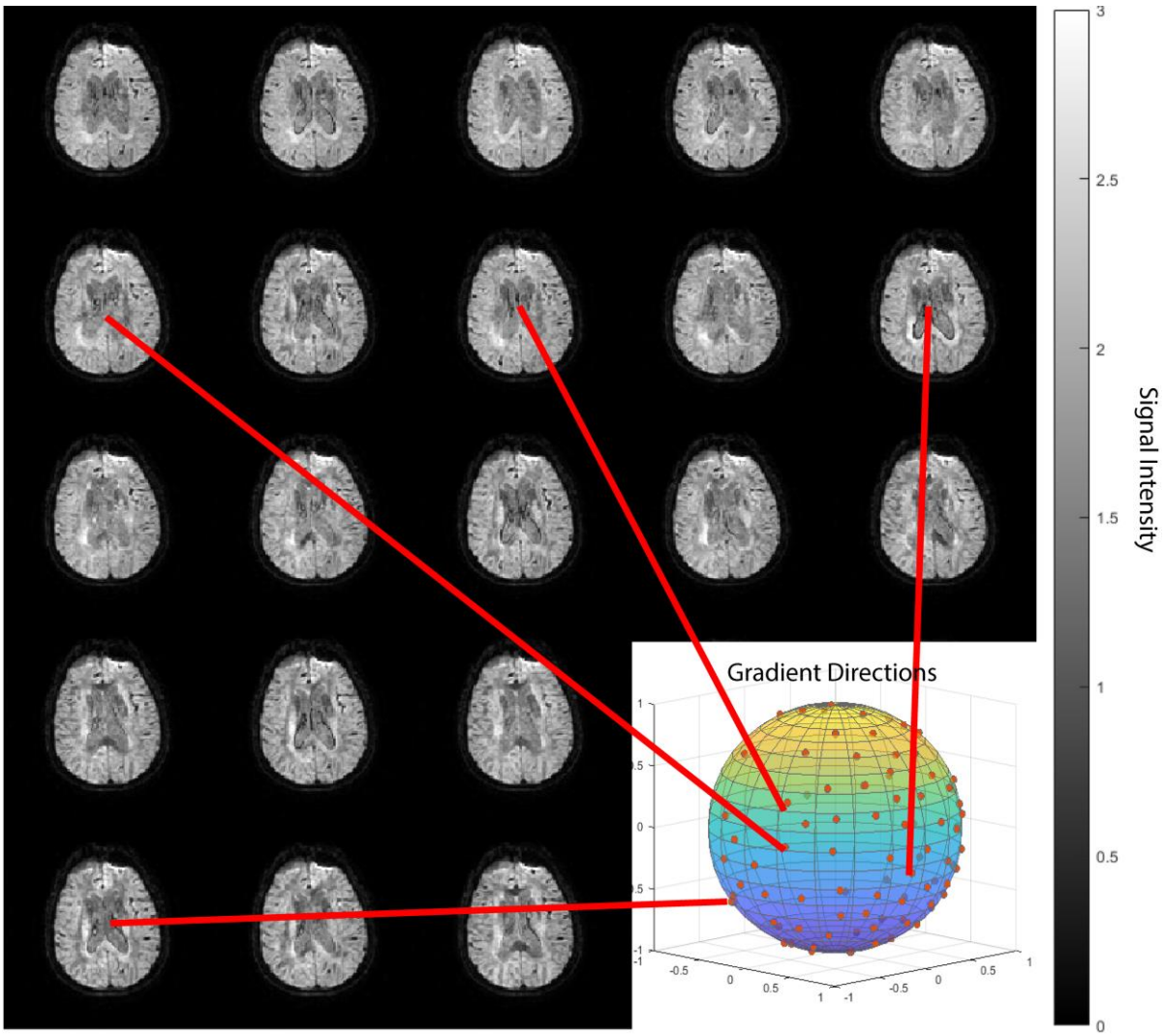

Figure I.2. The gradient directions are sampled over a sphere and for every gradient direction a 3D DW-MRI volume is acquired. Observable are 25 gradient volumes with the arrows depicting corresponding gradient volumes in those directions. Specifically, the middle axial slice of a human brain is show for all gradient directions

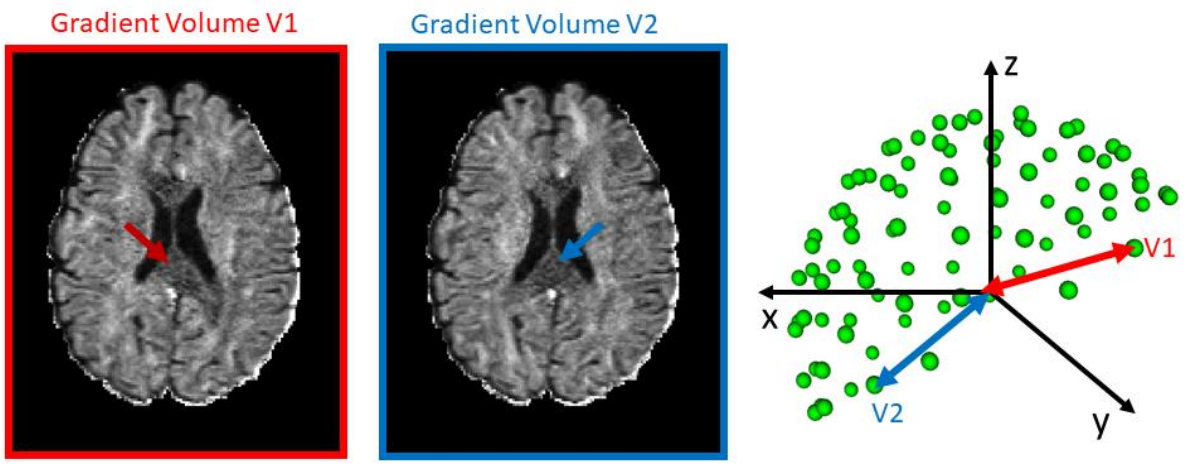

Figure I.3. Middle axial slice of a subject for two different gradient volumes is visualized. It can be observed that the arrows highlight a region of interest for the corpus callosum where the signal attenuation is different for both the gradient volumes.

tissue structure of the brain for both healthy and diseased subjects. However, these modalities are not limited and cannot acquire information regarding to intra-voxel microstructural tissue properties. Diffusion-weighted MRI allows non-invasive imaging at millimetric resolution scale per voxel of the gray and white matter tissue [27]. DW-MRI allows for a reconstruct of the micro-architecture due to the signal attenuation being the representation of the random or Brownian motion of water molecules in the human brain. The human brain is 73% water [28]. Since the inception of DW-MRI in mid-1980's [29] it has become a critical tool for imaging in-vivo neuro microarchitecture/microstructure clinically and for advancement through research.

The most commonly used diffusion pulse sequence is pulsed gradient spin echo (PGSE), the sequence was proposed by Stejskal and Tanner in 1965 [30]. The diffusion-weighting characterized by gradient pulses, for a PGSE sequence are defined as:

$$\boldsymbol{b} = \boldsymbol{\gamma^2 \delta^2 G^2 (\Delta - \frac{\delta}{3})} \qquad \textbf{Eq. 1}$$

Where γ is the gyromagnetic ratio, G is gradient amplitude, δ duration and Δ separation where the b is the diffusion weighting taking place.

The most common way that a diffusion signal can be characterized by under a Gaussian

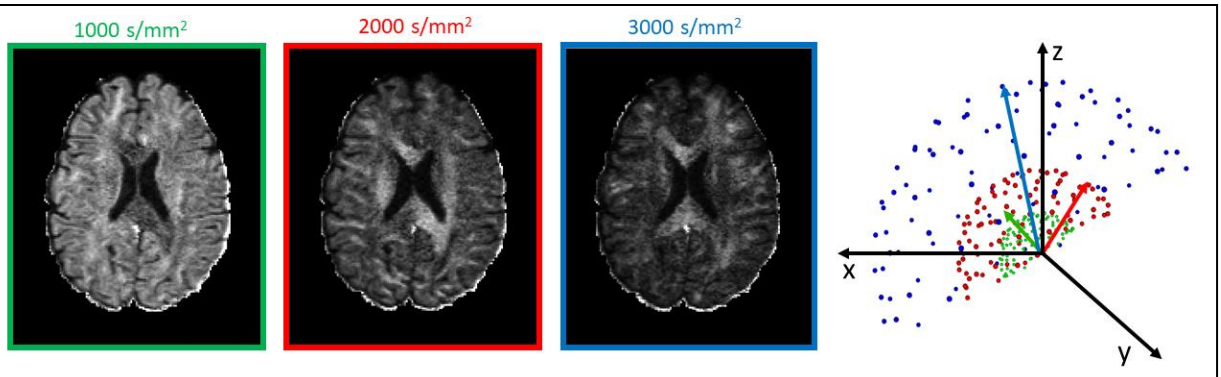

Figure I.4. Middle axial slice of a human subject is visualized for three different gradient directions. Each gradient direction corresponds to a unique diffusivity value. We can observe the change in contrast due to the change in diffusivity value.

assumption:

$$\boldsymbol{S = S_0 e^{-bD}} \qquad \textbf{Eq. 2}$$

Here, S_0 is the signal present in the absence of diffusion gradient pulses. The afore-mentioned PGSE sequence and the diffusion signal has been probed widely by the scientific community and will be used in the forthcoming chapters of this thesis.

DW-MRI acquisitions vary in terms of number of gradient volumes and diffusivity values (b-values) (Figure I.2, Figure I.3, Figure I.4). For ease of discussion, we can categorize them into clinical and advanced acquisition schemes. A major difference between the two is the amount of time required to acquire the data. For e.g. clinical acquisitions can typically be acquired in less than 5 minutes. The clinical acquisitions are usually defined by b-values of less than 1000 s/mm^2 with less than 30 gradient volumes. Anytime the number of gradient volumes is greater than 45 it is termed as high angular resolution DW-MRI (HARDI) [18]. A HARDI sequence can be considered as an advanced acquisition scheme. More advanced acquisition schemes would be where multiple b-values exist with varying number of gradient directions. A higher number of gradient directions and multiple diffusivity values allow for a higher angular and radial resolution which allows for measurement of more specific tissue properties. A typical voxel resolution for DW-MRI is 2.5mm isotropic.

1.2.2. High Angular Resolution DW-MRI

These acquisitions have become more common in recent times for research purposes. These acquisitions require longer periods of time where multiple diffusivity values are acquired with a varying number of gradient volumes [18]. Advanced acquisitions are capable of allowing the microstructure reconstruction methods to be more precise, reproducible and infer crossing, fiber fanning microstructures [31, 32]. The HCP uses an advanced acquisition protocol with diffusivity values of 1000, 2000 and 3000 s/mm^2 are acquired with a total of 90 gradient volumes per diffusivity value. The HCP data is being acquired at a voxel resolution of 1.25mm iso which is much higher than the clinical 2.5mm iso. The total volume per voxel when compared is 1.95mm^3 vs 15.6mm^3 roughly equivalent to 8 times of clinical resolution.

1.2.3. Spherical Harmonic & Simple Harmonic Oscillator Modelling of DW-MRI

There are varying scanning acquisition hyper-parameters in terms of gradient directions and diffusivity values. As also described prior in section 2.1 the DW-MRI acquisitions can be broadly classified into two major categories of single shell DW-MRI and multi-shell DW-MRI. Single shell DW-MRI can typically consist of non-diffusion weighted image and varying number of gradient volume for a fixed or specific diffusivity value (e.g b = 1000s/mm^2 and 30 gradient directions, commonly used for clinical acquisitions). Whereas, the multi-shell DW-MRI would typically contain two or more than two diffusivity values with varying gradient directions per diffusivity shell. Briefly, it can be stated that multi-shell DW-MRI consists of varying number of diffusivity shells, varying diffusivity values and varying gradient directions. As the scanning acquisition parameters vary a lot, it is generally useful have a consistent signal representation technique for further applicability to tissue models for detecting microstructure orientation and sometimes intra-voxel tissue compartmental measures as well. Spherical harmonics (SH) have been well established for the representation of single shell DW-MRI [33] however they are limited to a single shell and theoretically cannot capture the variations for a multi-shell DW-MRI acquisition. Simple harmonic oscillator reconstruction (SHORE) [34] is a modelling representation technique that was proposed to model multi-shell DW-MRI ubiquitously without compromising on the single shell DW-MRI

representation.

1.3. Microstructural Measures

1.3.1. Diffusion Tensor Imaging

Diffusion tensor imaging (DTI)is one of the first tissue models that was discovered and primarily shows the orientation of the white matter (WM) microstructure on a per voxel basis. Modelling the tissue microstructure is an inverse problem as the signal attenuation characterizes the underlying tissue microstructure present. Hence a tensor was hypothesized to fit the tissue structure and can be observed as an ellipsoid per voxel [35]. The DTI is estimated using a rank-2 symmetric positive definite matrix D:

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{D_{xx}} & \boldsymbol{D_{xy}} & \boldsymbol{D_{xz}} \\ \boldsymbol{D_{xy}} & \boldsymbol{D_{yy}} & \boldsymbol{D_{yz}} \\ \boldsymbol{D_{xz}} & \boldsymbol{D_{yz}} & \boldsymbol{D_{zz}} \end{bmatrix} \quad \textbf{Eq. 3}$$

The tensor matrix can replace the diffusion coefficient in the diffusion signal equation in the form of:

$$\boldsymbol{S} = \boldsymbol{S_0 e^{-bg^T Dg}} \quad \textbf{Eq. 4}$$

This equation can be further simplified to the form of:

$$\boldsymbol{\ln\left(\frac{S}{S_0}\right) = -\sum_{i=1}^{3}\sum_{j=1}^{3} b_{ij} D_{ij}} \quad \textbf{Eq. 5}$$

The DTI allows for estimation of other useful metrics such as fractional anisotropy (FA) and mean diffusivity (MD) which have become the ubiquitous in clinical DW-MRI and research of DW-MRI. FA describes the anisotropic nature of the orientation which is directly correlated with the amount of WM present in a voxel. MD describes the amount of diffusivity in the voxel which is used for characterizing neuropathology in the brain.

DTI has become prevalent in clinical studies however it is limited to studying only the primary orientation on a per voxel basis. It cannot resolve multiple fiber orientations [36] on a per voxel basis and hence cannot be used to study crossing, fanning fiber regions which call out for more advanced models that can infer more information using advanced DW-MRI imaging protocols (Figure I.5) .

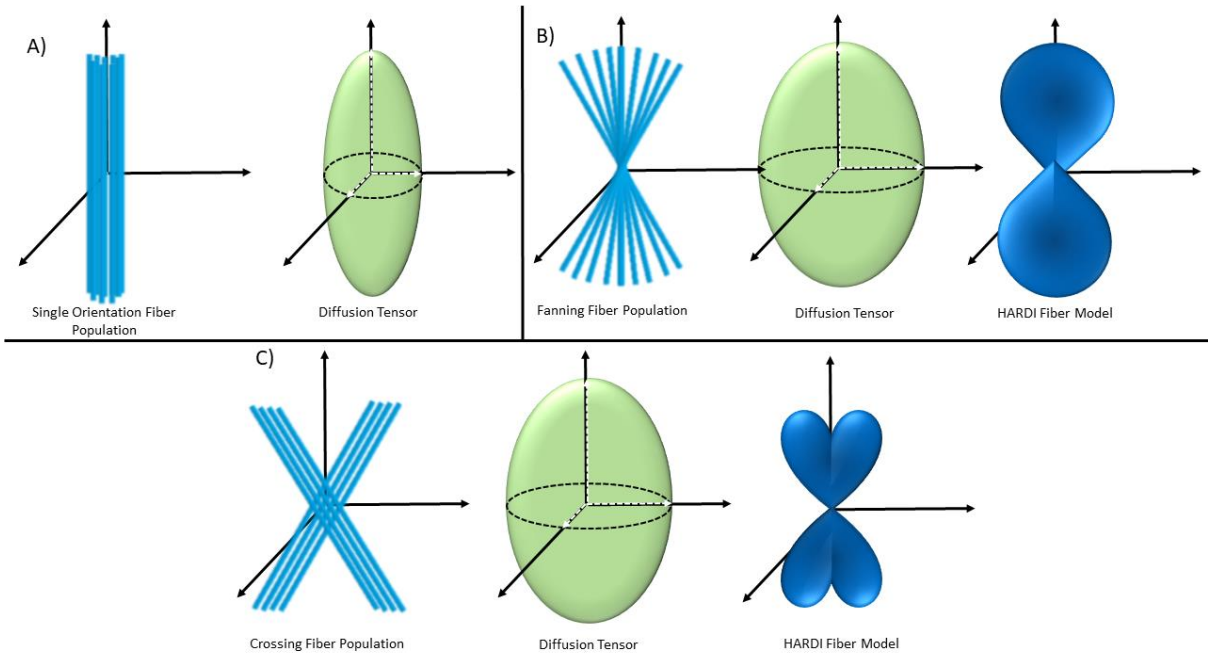

Figure I.5. A) Depicts a single fiber population and its corresponding DTI model as an ellipsoid. B) Shows a fanning fiber population and its corresponding DTI model vs the HARDI based fiber model. C) We can observe a crossing fiber population and its corresponding DTI model vs HARDI fiber model

1.3.2. HARDI Methods

With the limitation of DTI being failing to recover only a primary orientation in the WM. Multiple approaches were proposed which utilized additional information from advanced HARDI acquisitions that could recover multiple orientations on a per voxel basis. The motivation behind the proposed approaches was to incorporate multiple diffusivity values and a high number of gradient volumes in the hope of inferring a more elaborate microstructure. A few popular approaches that have been proposed are constrained spherical deconvolution (CSD), persistent angular structure (PAS-MRI), Q-ball imaging, diffusion orientation transform (DOT), diffusion spectrum imaging (DSI), simple harmonic oscillator reconstruction (SHORE), solid harmonics (SoH) and there are multiple others. Global challenge studies were performed to assimilate a collective evaluation of these approaches [22]. A few of the HARDI model

reconstructions on in-vivo data can be observed (Figure I.6). Comparisons of the proposed approaches have suggested inter-method disagreement for voxel wise orientation reconstruction. A substantial number of approaches have utilized the modelling capacities of spherical harmonics (SH) to model the DW-MRI signal and the reconstruct the underlying tissue microstructure as well. The underlying tissue orientational

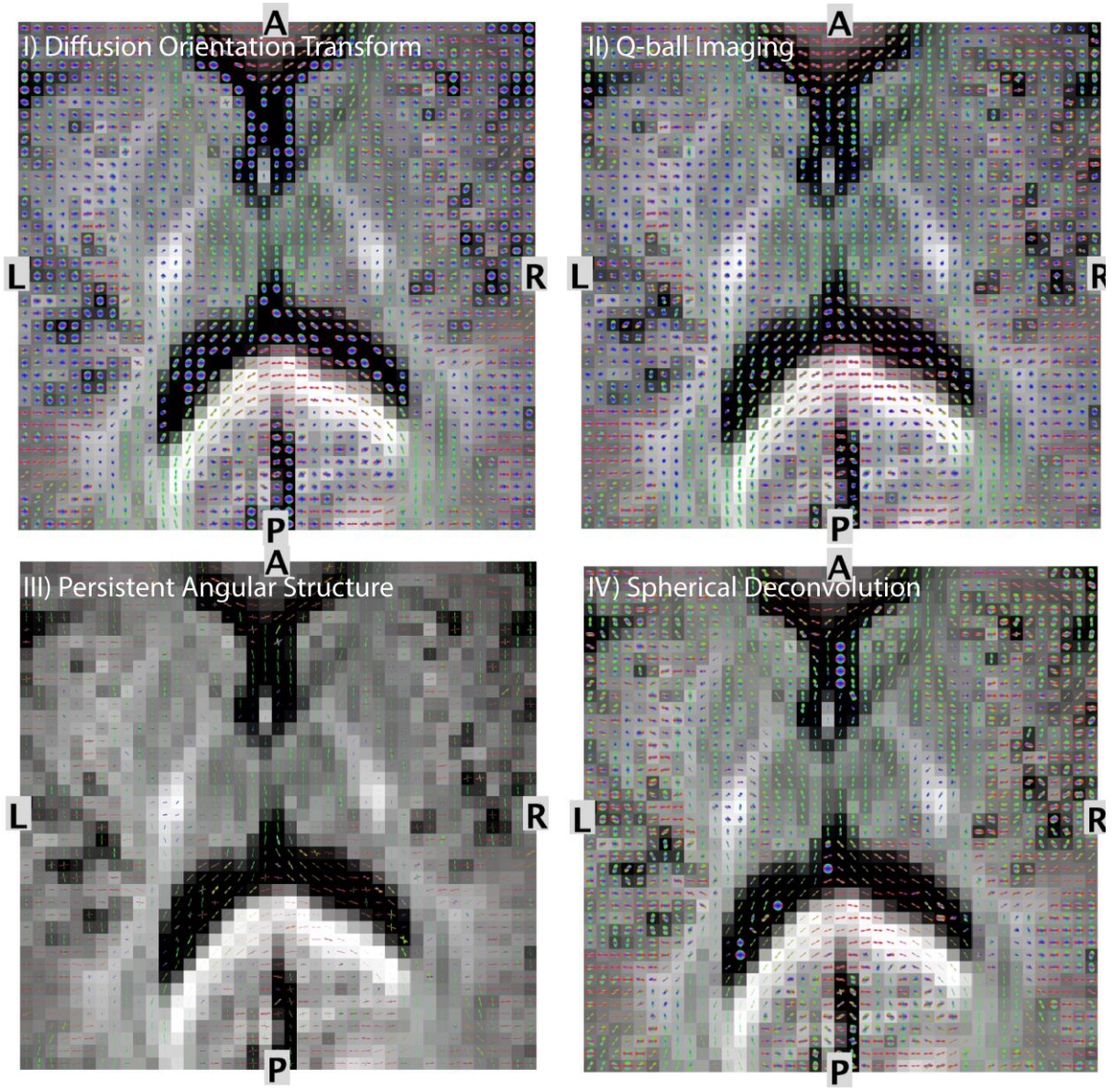

Figure I.6. HARDI methods visualization of asubject's middle axial slice. I) Reconstruction from Diffusion orientation transform. II) Q-ball Imaging III) Persistent Angular Structure IV) Spherical Deconvolution. It can be observed that each method shows a different tissue microstructure if compared for the same voxel across the different methods.

structure has most often been referred to as the fiber orientation distribution function (fODF or FOD) both abbreviations are commonly used in the literature. Development of these approaches led to formulation of metrics that could adapt to the new generation of models such as general fractional anisotropy (GFA), peak fraction (PF), angles between crossing fibers etc.

The advanced reconstruction methods have so far utilized classical mathematical transforms and mild explorations of data-driven techniques have been performed. There is a lack of methods that could be driven by external validation data (e.g animal studies).

Single Shell Spherical Harmonic Model: The assumption that only a single diffusivity shell of DW-MRI is being modeled allows for the modeling of DW-MRI on a sphere. Spherical harmonics (SH) are special functions that are defined on the surface of a sphere and allow for the formation of an orthonormal basis. The necessity for such a modeling arises due to the variability in scanning acquisition parameters in terms of the number of gradient directions for a DW-MRI scan. SH modeling allows for the representation of the signal via coefficients that are orthogonal and independent of each other. The Gaussian model of DW-MRI signal can be rewritten as below for SH DW-MRI modelling:

$$\frac{S}{S_o} = \sum_{l=0}^{L} \sum_{m=-l}^{l} \boldsymbol{C_{lm} Y_l^m(u)} \qquad \textbf{Eq. 6}$$

Here the Y_l^m is the SH component with 'l' and 'm' representing the order and the phase of SH. 'C' are the SH coefficients that will capture a representation of the signal. 'S' is the DW-MRI signal measurement and 'S_o' is the non-diffusion-weighted signal measurement.

The SH model for the DW-MRI signal utilizes a portion of Laplace's equation in spherical coordinates and are defined as:

$$\boldsymbol{Y_l^m(\theta, \phi)} = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} \boldsymbol{P_l^m(\cos\theta) e^{im\phi}} \qquad \textbf{Eq. 7}$$

Here Y_l^m represents the corresponding SH at polar coordinates of (θ, ϕ) where 'l' and 'm' correspond to the order and phase of the SH, P_l^m is the associated Legendre polynomial. The SH is used to form a basis set which can be used for estimation of SH coefficients. The basis set is a matrix of dimensions 'N x R' where 'N' represents the number of gradient directions and 'R' represents the number of

coefficients. The basis set is defined as:

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{Y_1(\theta_1\phi_1)} & \cdots & \boldsymbol{Y_R(\theta_1\phi_1)} \\ \vdots & \ddots & \vdots \\ \boldsymbol{Y_1(\theta_N\phi_N)} & \cdots & \boldsymbol{Y_R(\theta_N\phi_N)} \end{bmatrix} \quad \textbf{Eq. 8}$$

Here the elements of the basis matrix can be computed using equation __.

With utilization of the basis matrix the DW-MRI signal equation can be formulated as an over-determined system of equations:

$$\boldsymbol{X = BC} \quad \textbf{Eq. 9}$$

Where 'B' is the basis matrix and 'C' are the SH coefficients that will be estimated. The system of equations is based on the assumption that the number of gradient directions 'N' are always greater than 'R' as otherwise the system of equations becomes unsolvable. This system can be solved using linear least squares represented as:

$$\boldsymbol{C = (B^T B)^{-1} B^T X} \quad \textbf{Eq. 10}$$

However, it has been shown by prior work that regularized linear least squares offer significant representational advantages relatively [33]. It should also be noted that due to symmetric acquisition nature in terms of gradient directions only even ordered real SH are utilized. The number of coefficients for a given SH order 'R' can be estimated by:

$$\boldsymbol{R = \frac{(l+1)(l+2)}{2}} \quad \textbf{Eq. 11}$$

Here 'l' is representative of the maximum order that the SH series has to be truncated at. For greater 45 gradient directions 8th order is preferred as 45 coefficients can be estimated. It is advantageous to have more than 45 gradient volumes as some gradient volumes could be noisy due to scanner-based artifacts or as such. The primary advantage that SH modeling offers is that the signal measurements for a varying number of gradient directions can be recovered with a consistent number of coefficients. Also, the signal can be resampled to a different set of gradient directions using the coefficients due to their useful property of orthonormal basis sets.

Numerous advantages are offered by SH based modeling of DW-MRI signal for an effective feature representation. However, SH based modeling is limited for DW-MRI acquisitions where multiple diffusivity shells are present. At best the SH basis can be termed as an angular basis it does not consist of a

radial basis.

Multi-Shell Simple Harmonic Oscillator Reconstruction: With the added acquisitions of multiple diffusivity shells for the signal to be captured in a single set of coefficients a radial scaling function is applied to SH basis. The simple harmonic oscillator reconstruction (SHORE) model was proposed for this purpose and the DW-MRI signal equation can be re-written as below:

$$\frac{S}{S_o} = \sum_{k=0}^{K}\sum_{l=0}^{L}\sum_{m=-l}^{l} \boldsymbol{C_{klm}} \boldsymbol{G_{kl}}(\boldsymbol{q}, \boldsymbol{\zeta}) \boldsymbol{Y_l^m}(\boldsymbol{u}) \qquad \textbf{Eq. 12}$$

Here 'C' are the SHORE coefficients that will be estimated, 'G' depicts the radial scaling function which is combined with the SH basis Y_l^m. The radial scaling function 'G' can be defined as:

$$\boldsymbol{G_{nl}}(\boldsymbol{q}, \boldsymbol{\zeta}) = \boldsymbol{\kappa_{nl}}(\boldsymbol{\zeta}) \left(\frac{q^2}{\zeta}\right)^{\frac{l}{2}} \exp\left(-\frac{q^2}{2\zeta}\right) \boldsymbol{L}_{n-\frac{l}{2}}^{l+\frac{1}{2}}\left(\frac{q^2}{\zeta}\right) \qquad \textbf{Eq. 13}$$

Here the $\boldsymbol{\zeta}$ is a scaling parameter, q is the radius of the diffusivity value and 'L' depicts the associated Laguerre polynomial. The SHORE modeling allows for estimation of a basis matrix similar to SH based modeling which in turn allow for the estimation of coefficients using the branch of least square methods. SHORE is usually compatible up to 6th order after which it is considered unstable as per prior work [37].Tractography

The different regions of the brain are connected through the white matter nerve structure also often referred to as the microarchitecture organization of the brain. These connections serve as the passageway for transmission of information between the different regions of the brain. Therefore, an understanding of the tracts is fundamental for studying normal and diseased subjects alike. The recovery of intra-voxel fiber geometry estimates is a pre-requisite for estimating the tracts of the white matter. Delineation of these tracts is termed as tractography [38]. For over two decades tractography has been utilized for over two decades to study the different connections between regions of interest in the human brain. Also, tractography has been used for pre-surgical guidance by neurosurgeons [39, 40]. The tract delineation can be performed by multiple methods [41-43]. A visualization of tracts on the human brain is depicted in Figure I.7.

Tractography reconstruction can be segregated into two parts where the first part is the intra-voxel fiber geometric reconstruction which can provide the orientational information about the fiber

microstructure. The second part is the algorithmic that reconstructs streamlines or tracts by a continuous connection of the fiber orientations. DTI is one of the most commonly used techniques for recovery of intra-voxel fiber structure for performing tractography. HARDI methods can provide more elaborate information for tractographic algorithms and it is an active area of research. Existence of these tracts have been known through post-mortem or animal studies [44]. Estimation of tracts through DW-MRI was made possible first in 1999 [45].

The first technique invented to delineate tracts was termed as fiber assignment through continuous tracking (FACT). The technique relied on the primary fiber orientations being detected by DTI. The delineated tracts were also shown to have a correspondence with known anatomical locations of the fibers in a rat brain [46]. FACT algorithm relies on the primary orientation vector estimated per voxel using an orientational microstructure estimate. The simplest example is from DTI as it provides only a single primary orientation. Fiber tracking via FACT was formulated based on the primary orientation as follows:

$$\boldsymbol{R} = \sum_{i}^{s}\sum_{j}^{s} \boldsymbol{abs}(\boldsymbol{\lambda_{1_i}} \cdot \boldsymbol{\lambda_{1_j}})/\boldsymbol{s}(\boldsymbol{s}-\mathbf{1}) \qquad \textbf{Eq. 14}$$

Here $\lambda_1$represents the corresponding longest primary direction for a voxel and 's' is the total number of neighboring data points that are being utilized. Intuitively, if the value of R is high then it implies that neighboring voxels have fibers that are well-aligned the next voxel can be selected. While if the value of R is low then the fiber tracking is terminated as it implies discontinuity or that fiber orientations of neighboring voxels do not align.

The FACT algorithm was well received and led to a discovery for an entire field of tractography where multiple methods have been proposed some utilizing Bayesian statistics and some utilizing the information yielded by uncertainty measures. It became viable to state that there were too many tractography methods that had been proposed which led to benchmarking and challenges of tractography algorithms [42, 47].

Challenges have become a present norm in various capacities where some challenges tested reproducibility, a few exhibited phantom based validations [21] and a few tried to validate using histology [44]. The challenge studies focused towards tractography have specifically shown the current set of

problems that exist (how it differs on different scans from scanners, how some algorithms are not consistent.) Due to the varying scanning acquisition parameters of DW-MRI based acquisitions it is become important to establish that what type of scanning parameters lead to which kind of tractography based algorithms.

DW-MRI is currently the only way to delineate these tracts non-invasively and no imaging modalities can directly acquire the underlying neural pathway connections of the human in-vivo brain. Tractography can be classified into two major types of probabilistic and deterministic. It directly relies on the orientation estimates from the geometry-based microstructure reconstruction methods, which is one of the primary reasons as to why extensive research is being performed for advanced reconstruction methods

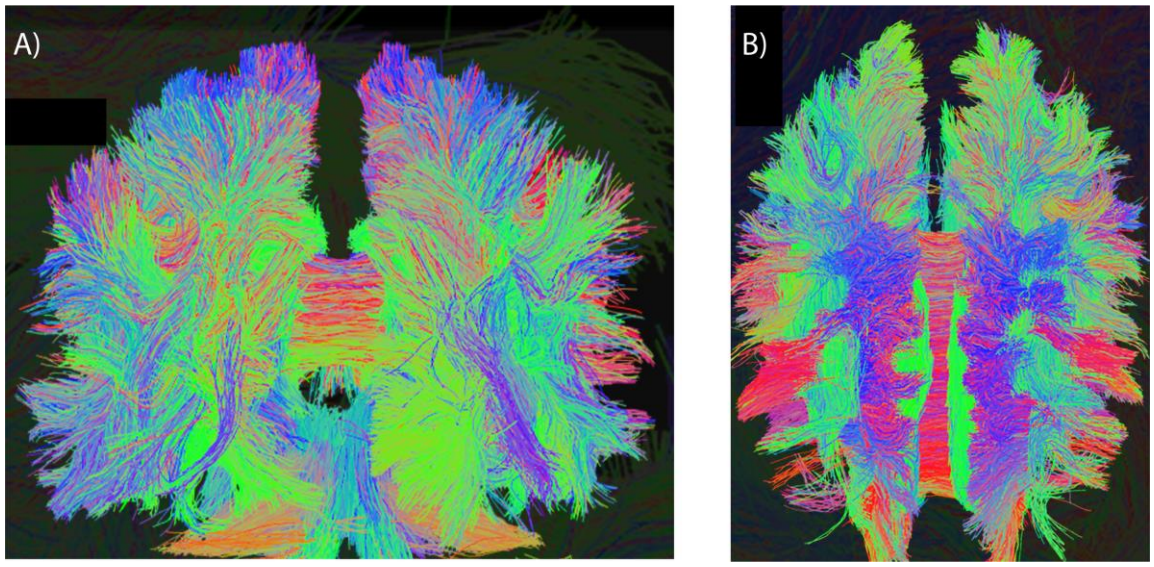

Figure I.7.Probabilistic tractography reconstruction which shows the neural pathways of an in-vivo brain on the same subject presented in two different views A) Coronal view B) Axial view. It can be observed that it is a challenge to detect which of the fiber streamlines correspond with the actual anatomy of the connectivity structure of the brain.

such as HARDI methods. To improve precision and reproducibility of tractography, it is often constrained using microstructural measures of FA or GFA [41].

1.4. Reproducibility & Harmonization of DW-MRI

"The extent to which consistent results are obtained when an experiment is repeated" is the definition of reproducibility [48]. Generally, reproducibility is a critical aspect of research due to the volatile nature of research where it is often observed that certain experiments show high variance. If the experiment does not exhibit the characteristic of reproducibility it is not possible for the specific proposition to be applicable in a real-world setting. It cannot be stressed enough on the aspect of reproducibility; recently a study published in Nature showed that over 70% of scientists were unable to reproduce the result of another scientist and more than half failed to reproduce their own results [48]. Advent of reproducibility is necessitated for the cumulative growth of scientific knowledge that is being produced every day. At the simplest level of understanding one could empirically think that if a pair of measurements are being made by repeating a specific experiment twice then the values should be reproduced with minimal error. While harmonization is a relative term, it is closely related to reproducibility but in a different context.

Harmonization can be defined as a way for "Allowing data to be meaningfully compared when there are multiple sources of data" [49]. A simple use case scenario would be harmonization of 'n' different populations from different regions/countries given the context of dietary patterns [49]. Similarly, in context of imaging modalities harmonization becomes a critical context when studies are being conducted across multiple different scanner sites [10, 50-54].

Both harmonization and reproducibility are equally important caveats for DW-MRI microstructure modelling that still need to be addressed. Given a single subject reproducibility can be measured across DW-MRI signal, different HARDI methods, tractography. In order to measure reproducibility, there are multiple pre-processing steps that must be performed for uniform analysis even for a single subject. This section is segregated further into reproducibility metrics and harmonization methodologies.

1.4.1. Reproducibility Metrics

Reproducibility of DW-MRI can be measured at different steps of the hierarchy such as directly at the signal measurement, microstructure metrics computed from different models such as DTI, or at the

tractography based reconstructed tract. The collection of metrics used for measuring the reproducibility of DW-MRI spans a wide variety ranging from mean squared error (MSE) to correlation-based similarity measurements over a sphere based on angular correlation coefficient (ACC) to Dice- Sorensen coefficient (DSC).

The mean squared error for a pair of measurements is defined as:

$$\boldsymbol{MSE} = \frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}}\left(\boldsymbol{Y_i} - \widehat{\boldsymbol{Y}}_{\boldsymbol{\iota}}\right)^{\mathbf{2}} \qquad \textbf{Eq. 15}$$

The angular correlation coefficient for a pair of SH measurements (u, v) is defined as:

$$\boldsymbol{ACC} = \frac{\sum_{j=1}^{\infty}\sum_{m=-j}^{j} u_{jm}v_{jm}^{*}}{\left[\sum_{j=1}^{\infty}\sum_{m=-j}^{j}|u_{jm}|^{2}\right]^{0.5}\cdot\left[\sum_{j=1}^{\infty}\sum_{m=-j}^{j}|v_{jm}|^{2}\right]^{0.5}} \qquad \textbf{Eq. 16}$$

Expected measure from ACC is on a scale of -1 to 1 where 1 indicates high correlation.

The Dice-Sorensen coefficient for pair of binary image measurements is defined as:

$$\boldsymbol{DSC} = \frac{\mathbf{2}|\boldsymbol{X}\cap\boldsymbol{Y}|}{|\boldsymbol{X}|+|\boldsymbol{Y}|} \qquad \textbf{Eq. 17}$$

Expected measure from DSC is on a scale of 0 to 1 where 1 indicates high similarity.

If the reproducibility is being measured for the DW-MRI signal directly where the scanning acquisition parameters are the same then standard statistics such as mean, standard deviation, MSE, root mean squared error (RMSE) are applicable on the computed apparent diffusion coefficient (ADC) [55-57]. However, these metrics are not sufficient or are not always directly applicable for use of measuring reproducibility for microstructural metrics.

There are many techniques for modelling the microstructure of DW-MRI which also often characterize a different kind of tissue property for the underlying tissue organization of the voxel. The most common orientational measure is provided by DTI [35] where FA reflects the anisotropic nature of the fiber population. FA as a scalar value allows for direct computation of standard statistics. However, the advanced HARDI methods often characterize multiple fiber populations in a single voxel which cannot be effectively captured by FA. The microstructure orientation characterized by HARDI methods is typically a set of either SH coefficients or it can be represented as a set of probabilistic values over a sphere. If reproducibility is being measured for SH coefficients for a pair of voxels then ACC [15] is used. ACC was first proposed by

Anderson et. Al in [15] and has been used widely in [58-60]. Quite often the major peaks of the fiber distributions are computed where they become 3D vectors per peak. It is usually uncommon to find more than 3-4 peaks per voxel. However, the metrics change for as the nature of data which is being measured is different. Symmetric angular error (SAE) as in [19, 61-63] is used and success fraction (SF) as in [19, 64]. For all compartmental measures such as partial volume fraction of tissues and orientation dispersion index since they are scalar values it is straightforward to apply standard statistics.

When dealing with tracts reconstructed from tractography based algorithms, the metrics utilized are different because the context changes to analysis of bundle-based streamlines. Commonly used metrics for tractography reproducibility are bundle overlap, bundle overreach and DSC as can be observed in [21, 41, 43]

We consider reproducibility metrics at three different levels:

- **Assessment of direct signal reproducibility** where it would be root mean squared error (RMSE) or mean absolute error (MAE). This could be across the brain or at voxel level.
- **Assessment of the reproducibility of tissue microstructure.** Here the metrics can change to ACC, MSE, PF and SF. These are voxel-based metrics.
- **Assessment of reproducibility for tractography.** Here the metrics change to volume-based ones such as DSC, ICC, bundle overlap and bundle overreach.

1.4.2. Harmonization Methodologies

Similar to the assessment of reproducibility at different levels harmonization can be performed at different levels as well (Refer Section 4.2). Prior harmonization methodologies have suggested techniques to harmonize the microstructure metrics directly such as FA [50]. However, the approach is limited by the analysis being focused on the singular metric. There are other approaches that have tried to harmonize the DW-MRI signal directly by calculation of rotational invariant features and registration between inter scanner sites. These approaches become limited to a specific study design. These gaps and limitations have given the motivation to propose methods that are robust to any type of study design data acquisition and

are not limited to a specific metric to perform a large-scale reproducible joint analysis.

Harmonization for analysis of multiple populations of scans coming from different scanner sites as a problem has been approached from various fronts such as harmonization at scanner level or harmonization at computed metrics for effective analysis. Earlier works in the literature focused towards direct harmonization of the computed metrics from DTI such as FA [50, 51]. The technique proposed in [51] is known as ComBat and has been commonly used as a batch-effect correction tool for genomics. Harmonization for DW-MRI as a problem can still be considered as a recent one and has not been approached as effectively till date. However, given existence of longitudinal studies such as BLSA [10] where over 20 years of longitudinal DW-MRI exists for studying the process of aging and other neurological disorders; harmonization is necessitated. Recently, signal based harmonization techniques were proposed where rotationally invariant features using spherical harmonics were utilized in [52-54]. While the above defined are a myriad of techniques for harmonization of DW-MRI data; harmonization from the perspective of microstructure modelling has not been addressed yet. Challenge studies have also been conducted for inter-site harmonization methodologies and a detailed review can be referred to in [65-67]

1.5. Machine Learning in DW-MRI

1.5.1. Advent of Deep Learning

Neural networks that could 'learn' were conceived half a century ago [68, 69] the idea was inspired from the neural circuit wiring of a nervous system. A seminal review for the field of neural networks, indicates that the original idea can be dated back to early 1800's [70]. When the initial artificial neuron model was proposed the proposed model was not intended to learn rather present a biological understanding for the neuronal activity in the brain [71]. In fact as stated in [70] "In a sense neural networks have been around even longer, since early supervised neural networks were essentially variants of linear regression methods going back at least to the early 1800s (e.g., Legendre, 1805; Gauss, 1809, 1821); Gauss also refers to his work of 1795".

Neural networks greatly evolved in 1980's; however, the ideas were limited by the hardware in that era. Only in the current (soon to end) decade has the progress been maximized with deep neural networks [72]. The early phase of research (early 2000's) on deep learning networks comprised of exploration of different combinations of the fundamental blocks of deep learning network architectures (for e.g. number of layers, width and height of convolutions, skip connections, different activation functions etc.) [73-76]. Recent aggressive research utilizing deep learning for computer vision now holds the potential for an advantageous expansion for multiple fields. A brief overview of advent of deep learning is shown in Figure

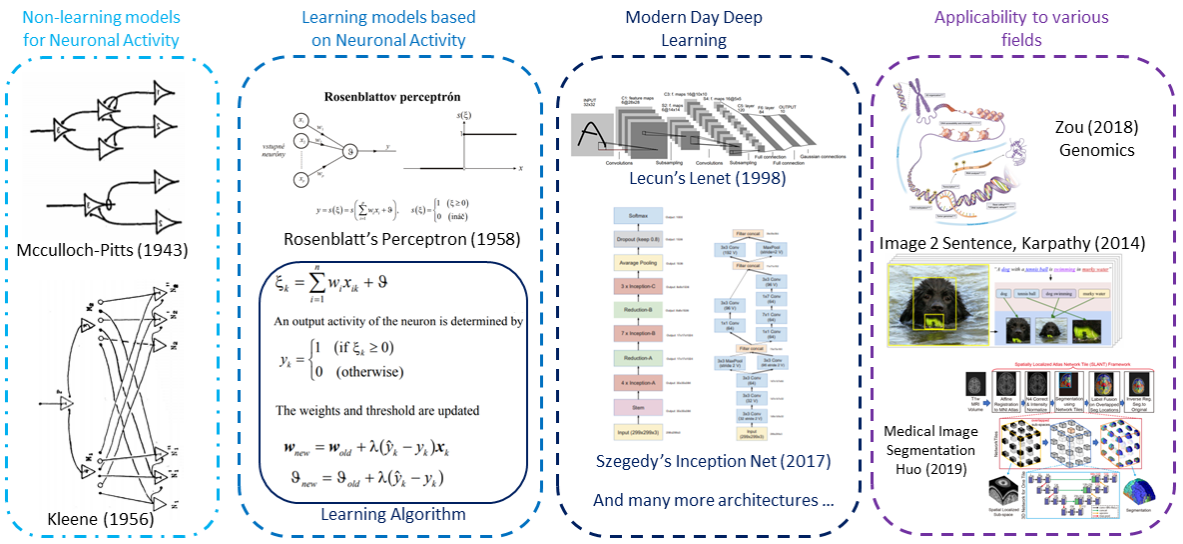

Figure I.8.Advent of deep learning and its progression from non-learning neuronal activity-based models to learning models on neuronal activity. The learning-based models formed the basis for the proposition of multiple architectures in 2000's and its current applicability and generalizability to multiple fields such as natural language processing, genomics and medical imaging.

I.8.

The deep learning revolution started snowballing once its advantageous applicability was showed for image recognition tasks such as image classification. Simple network architectures consisting of a few layers of neurons were shown to be able to classify hand written digits at extremely high accuracy as compared to classical/traditional machine learning models [73].

Machine learning is often defined as the ability of computers to perform a specific task based on a consistent set of patterns. Using a set of data or sample space where the data belongs to a class a machine

can be trained to evaluate and learn from the data. The data is broken into sets of training and testing where the training data is utilized by a machine learning model to learn a set of parameters that can be used for performing a specific task. The testing data is withheld, and the machine learning model is evaluated on the testing set of data. A simple example is a set of images where an image is labelled to a class of either "cats" or "dogs" and a machine learning model can be trained on this dataset to evaluate on a testing set to predict whether the given image belongs to the class of "dogs" or "cats" (Figure I.9). There are many categories of machine learning models such as Bayesian model, dimensionality reduction algorithms, clustering algorithms and deep learning models. Deep learning has exponentially become popular in the current decade for image processing and similarly for medical imaging as well.

Gradually the applicability of deep learning was extended to a collection of broad tasks such as image and video recognition [77], object detection as landmarks in images and videos [78], classification of multiple objects [79], generation of similar data via adversarial models [80], segmentation of objects and registration. Apart from computer vision tasks deep learning is also being utilized for completely different

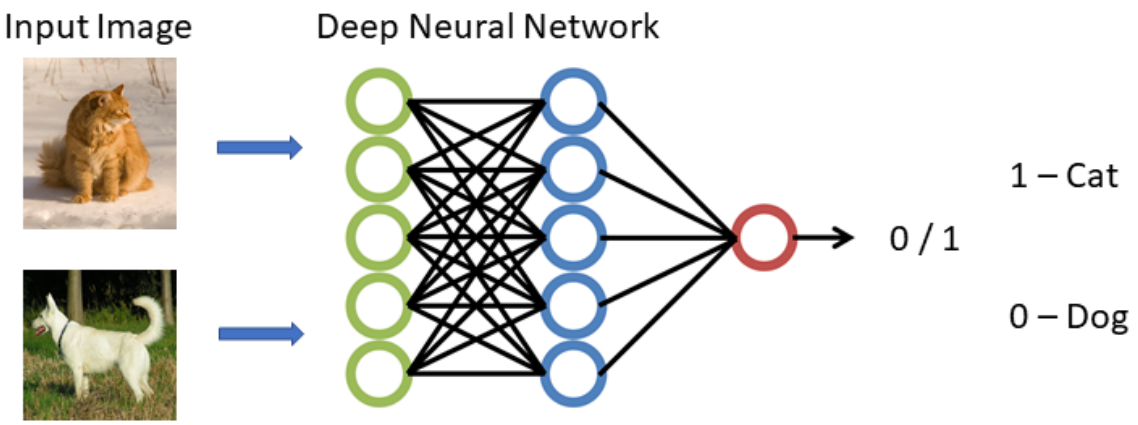

Figure I.9.Shows the workflow of a neural network where it can discern between a cat or a dog. The deep neural network is represented by lots of neurons that translate to learning parameters.

fields of natural language processing [81], genetics [82], astronomy [83], and many more. Specifically, in computer vision deep learning is also heavily being utilized for medical imaging computer vision tasks as

well [84]. Medical imaging as compared to generic computer vision is much more complex due to advanced nature of acquisitions for various imaging modalities of CT, functional MRI, DW-MRI, PET etc. The core of this thesis is focused towards opening up the possibilities of how and what deep learning can bring to the table for DW-MRI.

1.5.2. Deep Learning for DW-MRI

Unlike various other fields where deep learning has heavily evolved and become a norm for generic 2D image,video tasks and NLP.. The same does not hold true for the modality of DW-MRI due to the complex nature of the acquisition scheme. The radial and angular dependency of the 3D volumes of DW-MRI do not allow for a straightforward application of data-driven approaches which is unlike for generic 2D vision tasks where deep learning is applicable in a straight-forward fashion. MRI acquisitions are 3D volumes and for DW-MRI the dimensionality becomes 4D due to the gradient volumes. The increase in dimensionality creates a challenge for application of data-driven approaches due to limitations of hardware. Complexity is further increased with inclusion of multiple diffusivity shells as the data dimensionality becomes 5D. The primary challenge is the dimensionality of the DW-MRI acquisitions and the secondary challenge is the inconsistency of the acquisition style where the number of gradient volumes and diffusivity values vary on independent scanning acquisition parameters. Data-driven modeling for DW-MRI are at a nascent stage [25].

Generative adversarial networks have been used for acceleration of reconstruction of MRI [85]. The authors couple the adversarial loss with an innovative content loss. While [86] uses the introduced novelty above, additionally they propose to use a two-stage architecture for better adaption of loss functions and ensure learning on other features as compared to joint training. The results were interpreted on a cardiac MRI dataset. Deep learning has also been applied to diffusion tensor cardiac magnetic resonance for resolving microscopic structural organization of the myocardium [87]. A cascaded convolutional neural network architecture was used for assessment as compared to other classical methods on a simulated dataset. Thereafter, there are also a few data-driven approaches to reconstruct microstructural measures

[25] and address harmonization for DW-MRI for inter-site scanning acquisition [66, 88, 89].

2. Empirical Comparison of Multi-Fiber Reproducibility of PAS-MRI and Q-ball with Empirical Multiple b-values HARDI

This chapter has been adapted from the published work in [19].

2.1. Introduction

Diffusion tensor imaging (DTI) [35] is widely recognized as a robust method and heavily relied upon for interpreting brain structural connectivity in vivo, but it can only resolve a dominant single fiber orientation. Advanced diffusion weighted magnetic resonance imaging (MRI), such as Q-ball [13], Persistent Angular Structure (PAS-MRI) [12], Spherical Deconvolution (SD) [15, 90], Diffusion Orientation Transform (DOT)[91], seek to resolve this shortcoming by modeling complex intra-voxel structure. With DTI, empirical reproducibility studies [92, 93] provided concrete information linking observed behavior with theoretical noise sensitivities. Understanding the empirical tradeoffs between these sequences is critical to design of optimal imaging protocol and analysis specification. Yet, to date, comparative analysis of the advanced sequences (commonly known as high angular resolution diffusion imaging - HARDI) has not been performance on large standardized dataset. The overall objective of this work is to address this shortcoming.

Herein, we focus on two methods, Q-ball and PAS-MRI. Q-ball models the voxel wise orientation diffusion function (ODF) by low-order spherical harmonics [33], and is thus able to detect multiple fibers as peaks in diffusion signal in a spherical coordinate system. PAS-MRI more explicitly models the intra-voxel diffusion by a discrete number of fiber compartments [12]. Prior work with Q-ball has shown that it resolves multiple fibers in regions of fractional anisotropy (FA) < 0.4 and detects many voxels with the second direction of diffusion for FA > 0.15 at b-values greater than 1500 s/mm2 [94]. In [95], Q-ball resolved crossing angles to approximately 45 degrees with lower accuracy at b-values of higher than 4000 s/mm2, and it was estimated that lower b-values lead to under estimation in crossing fibers.

Similarly, PAS-MRI has been shown to recover multiple fiber orientations at “low” b-value of 1200

s/mm2 and has been shown more consistent than q-ball on synthetic MRI data [64]. PAS-MRI was further established as a special case of SD described as maximum entropy spherical deconvolution (MESD) which was shown to be more accurate than traditional SD at the cost of increased computational time [96]. The empirical literature has not yet shown how consistently PAS reproduces the crossing fiber angles.

Briefly, in this study (see Figure II.10), we pre-process the data with eddy and topup (FMRIB Software Library (FSL), [97, 98]) for each scan session individually and, then all sessions were registered to the Montreal Neurological Institute (MNI) template in FSL to establish a coordinate system for inter-session analysis. Thereafter, the gold sets were created by concatenating all the single scans per b-value in the acquisition order and HARDI methods were implemented.

2.2. Methods

2.2.1. DW-MRI Data Measurements

A healthy volunteer was scanned for three different sessions each two hours long at 3T (Achieva, Philips Medical Systems, Best, The Netherlands) with a 32-channel head coil on consecutive days. The first two scan sessions consist of four repetitions of 96 gradient directions per the b-values of 1000, 1500, 2000, 2500 and 3000 s/mm2. For each shell, a scanner average of 10 minimally weighted reference images ("b0") was acquired. The last scan session had three repetitions of 96 gradient directions per the same b-values above equivalent to 15 single scans. This data was acquired at the voxel resolution of 2.5mm x 2.5mm x 2.5mm with a matrix of 96 x 96 and 38 slices. Briefly, scan parameters were: Multi-Band=2; SENSE=2.2; TR= 2650 ms;TE=94 ms;partial Fourier=0.7. Fold over direction was A-P with a P fat shift. For each set of the 5 diffusion weighted shells, an additional "b0" was acquired with reverse phase encoded volumes (i.e., fold over direction A-P with A fat shift), 3 diffusion weighted directions with a b-value of 1000 s/mm2 along the imaging frame cardinal directions, and all other parameters were kept constant.

2.2.2. Data Pre-processing

The data for each session were corrected for patient movement, eddy current distortion, and susceptibility distortion through topup and eddy [97, 98]. Every diffusion weighted scan in each session had a single b0 (scanner average of ten scans). In addition, each session also had four reverse phase encoded

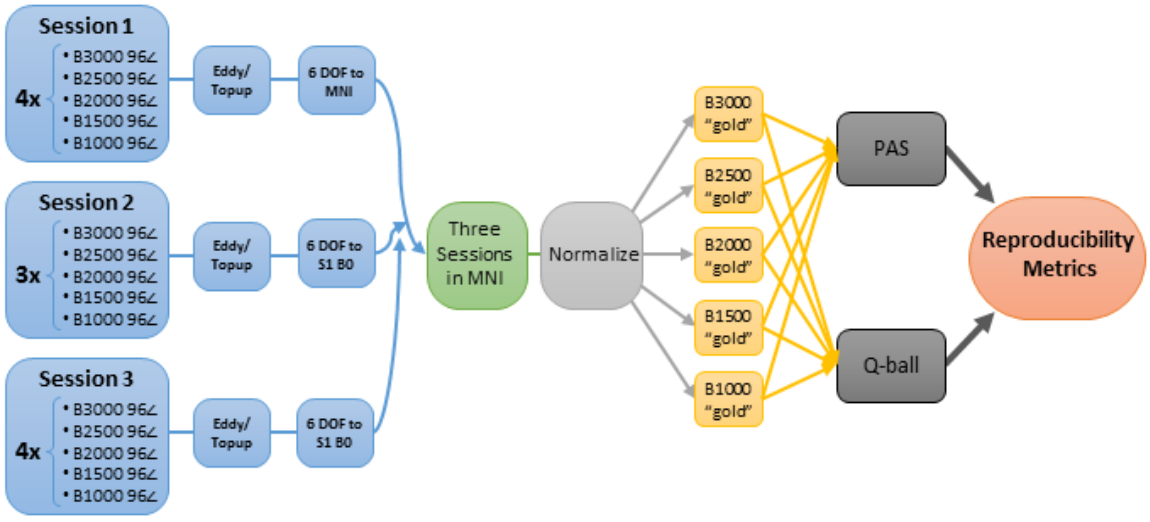

Figure II.10. 5830 volumes were acquired in 3 sessions that comprised 11 (4/4/3) repetitions of 5 shells, each with 96 diffusion weighted directions and a scanner average of ten b0 images. All data from each session was distortion corrected and the reregistered in MNI space. "Gold standard" data for each shell was created by concatenating all data into a single acquisition, while each empirical acquisition shell was also considered separately. PAS and Q-ball were run on all data and reproducibility measures were computed.

b0 volumes (scanner average of ten scans). All b0s (including reverse phase encoded b0s) in a session were concatenated together and fed as inputs to topup. For eddy, every scan in a session was concatenated and then corrected using the results from topup (note that the acquisition did not have complete reverse phase encoding acquisitions for the diffusion weighted directions). Afterwards, the distortion-corrected b0s from the first session were averaged and registered to a 2.5mm isotropic MNI T2 resample from 0.5mm weighted template using 6 degree of freedom registration with normalized mutual information metric (flirt, [99]). The volumes for the next session were brought to the same space by averaging its b0s together and registering it to the averaged b0s in the first session. The same procedure was subsequently done to the last session to bring all three corrected sessions to a common subject specific MNI space.

All sessions were normalized by the b0 corresponding to the scan to account for amplitude drift. A weighted mean b0 was created from all b0 scans in MNI space by a weighted average of all b0's. Weight has been taken as inverse of the median of all b0 scans in MNI space averaged with the mean of b0's from all sessions. The resulting normalization scan was computed by multiplying the weighted mean b0 to the ratio of diffusion weighted scan (in MNI space) to the b0 (in MNI space).

2.2.3. Segmentation:

The data was divided into 11 single scans per b-value each consisting of 96 gradient directions with a single b0. To create the reference standard ("gold model"), we concatenated the 11 single scans in the order of acquisition to create a volume of 1067 gradient directions per b-value. For the analysis a white matter mask was created a structural image using FSL's automated segmentation tool (fast, [100]) and subsequently registered to the diffusion data..

2.2.4. Q-ball Reconstruction

The Q-ball was reconstructed following [101] and using spherical harmonic basis functions [33] using the Camino Diffusion Toolkit [102]. Q-ball estimates the diffusion orientation distribution function (ODF) by approximating it through a Funk-Radon Transform of the diffusion weighted signal. Herein, we use the spherical harmonic order of 8 for Q-ball, as per the guidance in [101].

2.2.5. RE-PASMRI Reconstruction

PAS was reconstructed from maximum entropy spherical deconvolution [96]. We used the reduced encoding model for the MESD [103]. Similar to Q-ball, PAS estimates the "persistent angular structure" of the diffusion propagator which resembles the DODF from Q-ball. We used a reduced encoding of 16 as per [103] which gives a good trade-off between computation speed and accuracy for the PAS. The gold sets (amalgamated 11 single scans) for the PAS were processed across multiple CPU cores with the data being split at imaging slices.

2.2.6. Peak Reconstruction

The peak reconstruction of the ODF and the PAS was performed using sfpeaks from Camino which is an implementation for searching the local maxima of the reconstructed PAS and ODF's using Powell's method [12]. Both PAS and Qball were restricted to detecting peaks of (principal direction threshold) greater than one unit of peak strength divided by the mean of all the detected reconstruction functions (PAS/ODF). We used the recommended default parameters of the program sfpeaks. These parameters follow the precedence of [12].

2.3. Metrics

2.3.1. Reproducibility of Peaks

The consensus number of peaks was defined per voxel as the mode number of peaks across the 11 repetitions. The mean and standard deviation of agreement with consensus (1 versus 0 for disagreement) was determined within the white matter mask for each shell. To assess bias, absolute agreement was computed with the mean and standard deviation of agreement with gold standard number of peaks was determined within the white matter mask for each shell.

2.3.2. Peak Fraction

The consensus peak fraction was defined per voxel as the mean peak fraction for the first 3 peaks across the 11 repetitions. The root mean squared error (RMSE) and bias (mean difference) between each repetition and the gold standard for each shell were computed and averaged over the white matter mask.

2.3.3. Reproducibility of Crossing Fiber Angle

For any case of voxel where two or more peaks were detected, we determine the minimum angle being resolved, $\measuredangle = \underset{i,j,k}{\operatorname{argmin}}(\arccos(i,j)\ arcos(j,k)\ arcos(j,k))$ where peaks are represented as vectors i, j, k. The absolute error per voxel for the crossing fiber angle is defined as $|\measuredangle gold - \measuredangle test|$. We define a crossing fiber *success fraction* as the proportion of voxels which have error less than 20 degrees.

2.4. Results

Table II.1 Reproducibility for the peak fraction

	PAS				Q-ball			
b-Values ***s/mm²***	***RMSE Consensus***	***Std. Dev. in WM***	***RMSE Gold***	***Std. Dev. In WM (Gold)***	***RMSE Consensus***	***Std. Dev. in WM***	***RMSE Gold***	***Std. Dev. in WM (Gold)***
1000	0.075	0.032	0.103	0.049	0.128	0.060	0.211	0.120
1500	0.070	0.031	0.094	0.046	0.124	0.060	0.202	0.118
2000	0.068	0.030	0.090	0.045	0.126	0.056	0.205	0.114
2500	0.066	0.030	0.086	0.044	0.125	0.052	0.208	0.112
3000	0.068	0.030	0.087	0.043	0.126	0.049	0.215	0.110

2.4.1. Consistency in Number of Peaks and Peak Strength

The consensus for the number of peaks per voxel is consistent as a function of b-value for both PAS and Q-ball (Table II.1). PAS is more consistent than Q-ball for all experiments in Table II.1, yet the within white matter mask variance is similar. For both methods, substantial bias is evident between the single repetition and gold standard data in terms of lower consistency with the gold standard versus with

Table II.2 Reproducibility for the number of peaks

	PAS				Q-ball			
b-Values ***s/mm²***	***Mean Consensus***	***Std. Dev. in WM***	***Abs. Agreement (Gold)***	***Std. Dev. Of Abs. Agreement in WM (Gold)***	***Mean Consensus***	***Std. Dev. in WM***	***Abs. Agreement***	***Std. Dev. Of Abs. Agreement in WM (Gold)***
1000	0.792	0.170	0.678	0.302	0.6834	0.183	0.531	0.315
1500	0.799	0.168	0.713	0.277	0.691	0.184	0.546	0.312
2000	0.799	0.169	0.729	0.261	0.681	0.178	0.531	0.307
2500	0.804	0.166	0.743	0.250	0.669	0.173	0.507	0.305
3000	0.793	0.167	0.737	0.244	0.652	0.168	0.475	0.302

the consensus. PAS exhibits a trend toward lower bias an increasing b-value while q-ball appears to increase. The RMSE of the peak fraction for 11 single scans shows a decreasing trend with increasing b-value for PAS while for Q-ball remains consistent (Table II.2). Peak fraction bias decreases with increasing

b-values for PAS but remains consistent for Q-ball.

2.4.2. Crossing Fiber Angles indexed with FA

There is an increase in the number of crossing fibers voxels detecting in "gold standard" maps increasing b-value (refer Figure II.11). For all b-values, PAS resolves crossing fibers in the range of 65 to 90 degrees. At lower b-values, crossing fiber are detected up to an FA of ~0.6, but at b-value of 3000 s/mm2, the concentration of voxels with crossing fibers has a FA of <0.4. The spatial maps in Figure II.11 show qualitative agreement between the gold standard and single repetition, but there are clear spatial patterns of differences. These differences are reflected quantitatively in the bias shown in Table II.1 and Table II.2.

Figure II.12 presents a similar analysis for Q-ball. The voxels with crossing fibers largely occur at 70 to 90-degree angles; however, a clear second population crosses at ~55 degrees. While crossing fibers are often detected in voxels with FA up to ~0.4, the majority of detected crossing fibers lie in voxels with FA < 0.2. Notably, number of crossing fibers increases with b-value. Similarly, with PAS, the spatial maps for Q-ball show systematic difference in the number of fibers estimated with the gold standard relative to a single repetition, which is also reflected in Table II.1 and Table II.2.

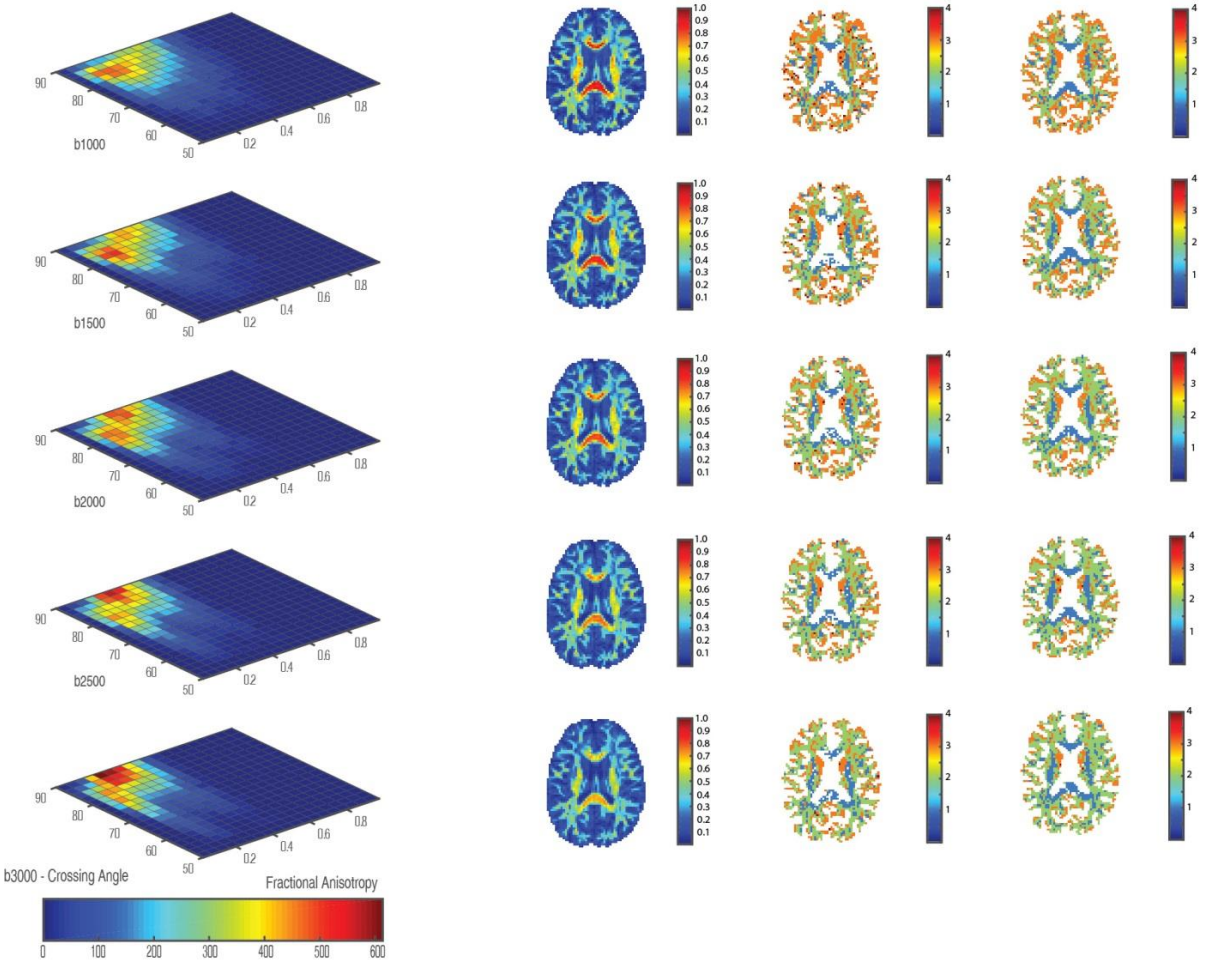

Figure II.11. Results for PAS: (A) Two-dimensional histograms of crossing fiber angles detected with the gold standard at each b-value. (B) FA maps at each b-value computed with DTI for the gold standard. (C) Number of peaks detected for the gold standard. (D) Number of peaks detected for a single repetition.

2.4.3. Success Fraction and Error

PAS resolves crossing fiber most successfully when the angle is greater than 75 degrees (Figure II.13 A), which occurs when the difference between gold is minimized (Figure II.13 B). There is a trend

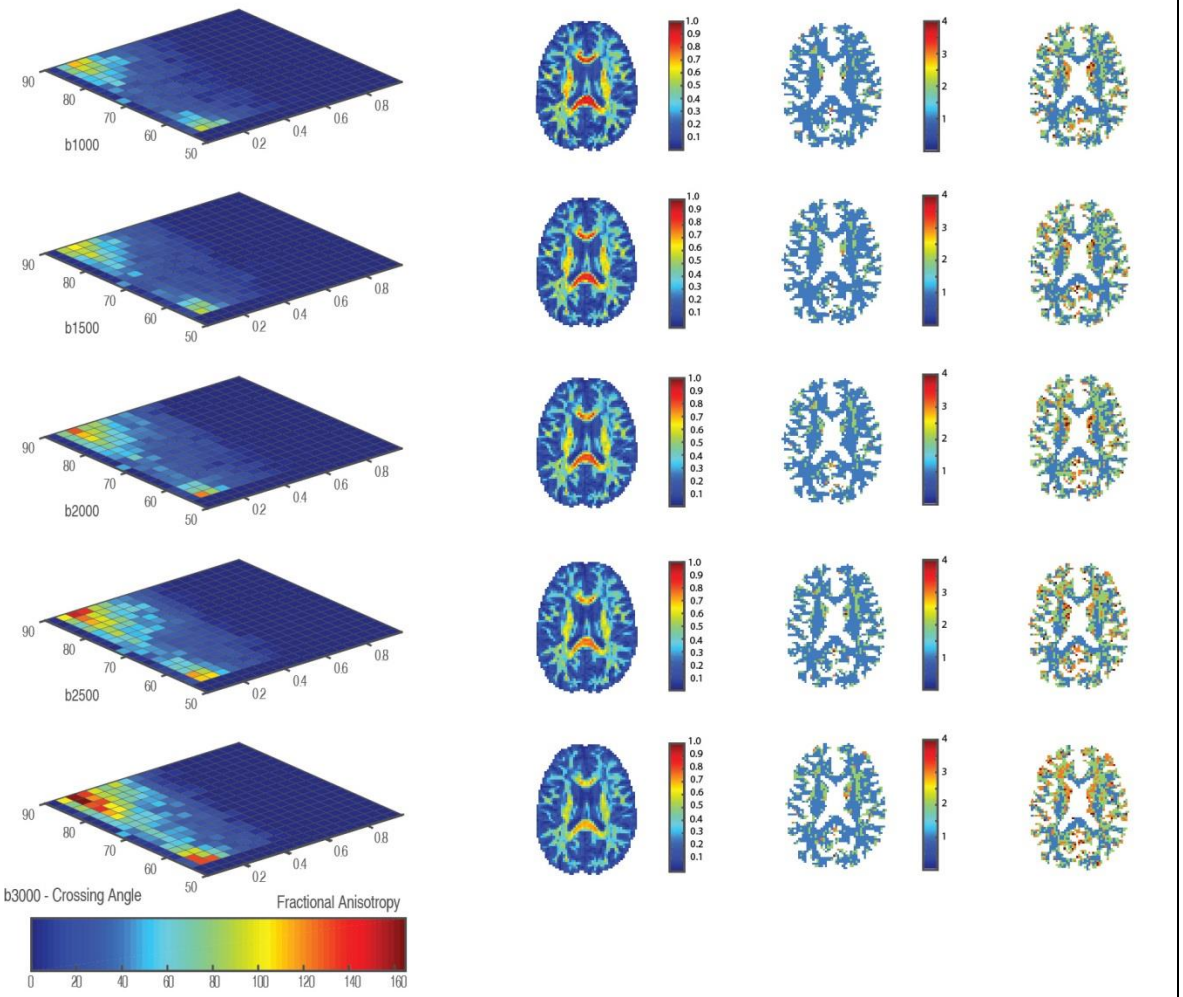

Figure II.12. Results for Q-ball: (A) Two-dimensional histograms of crossing fiber angles detected with the gold standard at each b-value. (B) FA maps at each b-value computed with DTI for the gold standard. (C) Number of peaks detected for the gold standard. (D) Number of peaks detected for a single repetition.

for more accurate crossing fiber detection at high angles for lower b-values, and more accurate low angle crossing for higher b-values. Q-ball exhibits overall lower success fraction than PAS and has a peak sensitivity for 65-75-degree crossings (Figure II.13 C). Lower angles crossing is more accurate with Q-ball, but Q-ball shows degrading performance near 90-degree crossing. The highest b-values resulted in the lowest errors for Q-ball.

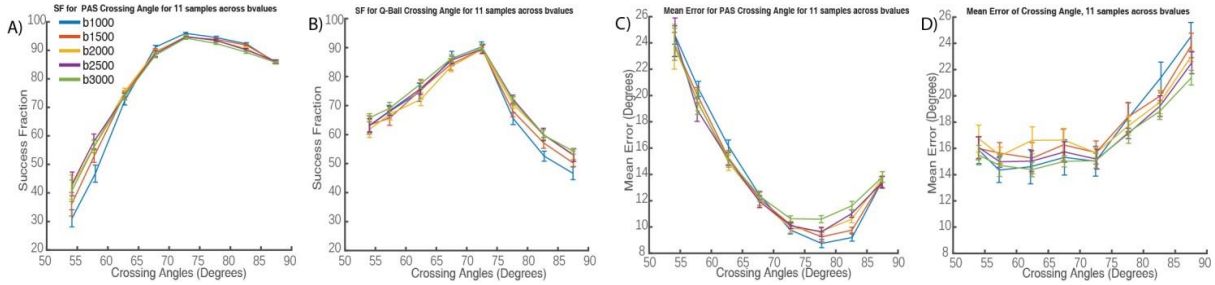

Figure II.13. (A) Crossing fiber success fraction for PAS and (B) mean error between each repetition and the gold standard for PAS. (C) Crossing fiber success fraction for Q-ball and (D) mean error between each repetition and the gold standard for Q-ball.

2.5. Conclusion

This is the first large-scale empirical reproducibility analysis to compare Q-ball at PAS at 3T in vivo. For the acquisition studied, PAS is more reproducible both in terms of number fibers per voxel (Table II.1) and the peak fraction (Table II.2). However, there is still substantial variability (15% ~ 20%) in terms of consensus for both Q-ball and PAS. It is important to note that Q-ball and PAS detect different aspects of intra-voxel structure within identical data. Note the two-dimensional histograms in Figure II.11 and Figure II.12 where PAS identifies more fibers in moderate FA voxels, while Q-ball resolves fibers within voxels with low FA. Similarly, PAS MRI is more consistent for crossing fibers at near-right angles while Q-ball appears more consistent for lower angles. A deeper consideration of the validity and consistency of cross-fiber populations across analysis method is clearly warranted and will be pursued.

Moreover, the study reveals that both Q-ball and PAS exhibit strong biases in estimated fiber fractions with the number of signal averages considered. Specifically, compare the consensus reproducibility versus the gold standard reproducibility in Table II.1 and Table II.2. Qualitatively, the differences can be appreciated in terms of the peak maps in Figure II.11 and Figure II.12 and seen in the rendering of Figure II.14. Visually, the single scan estimates for Q-ball appear consistent with prior expectations of fiber crossing locations; while the face validity differences for PAS are less clear. Deeper consideration of Q-ball and PAS estimation with signal-to-noise-ratio (SNR) and/or number of repetitions

is clearly warranted (as has been done for DTI [92, 93].

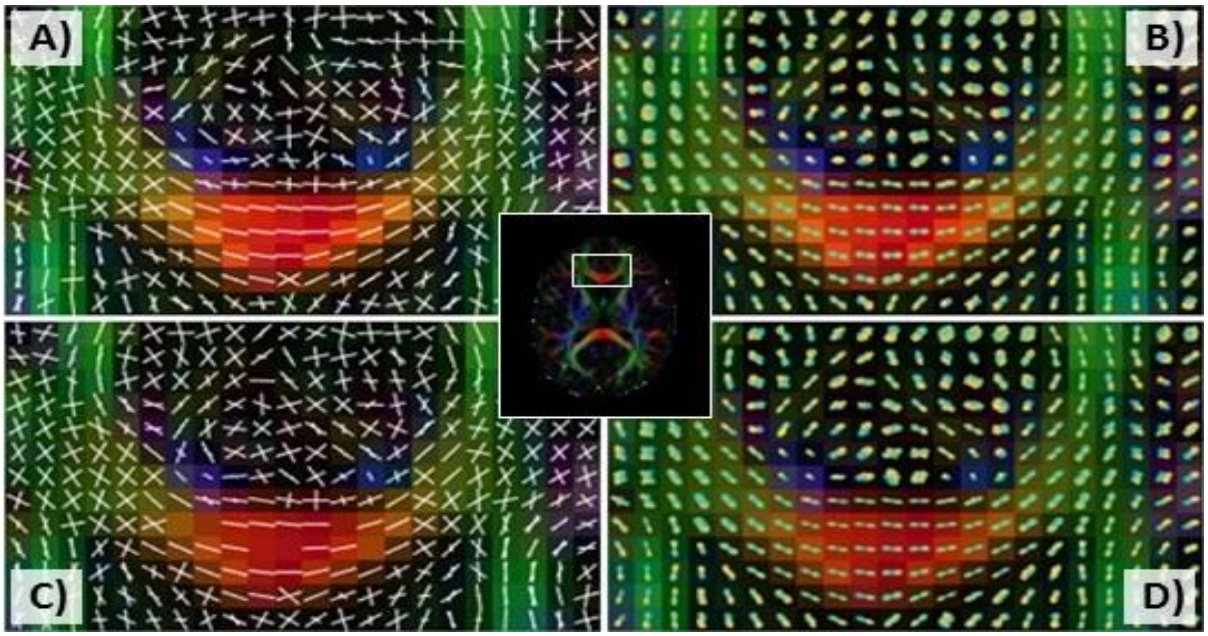

Figure II.14. Illustration of the differences in model fit for the gold standard datasets with: (A) PASMRI – 1000 s/mm2, (B) Q-ball – 1000 s/mm2, (C) PASMRI – 3000 s/mm2, and (D) Q-ball – 3000 s/mm2.

In summary, this study provides direct empirical evidence of systematic differences between Q-ball and PAS. Although b-value clearly impacts estimation of fiber structure, the choices in analysis approach (Q-ball versus PAS) and number of scans (single repetition versus gold standard) have a more substantial impact on data interpretation. Additionally, the types of crossing fibers detected by the methods were difference in terms FA, crossing angle, and number of peaks. There is clear opportunity for innovation in data interpretation to improve consistency, minimizing the impact of SNR, and unifying models of intra-voxel structure. In the meantime, PAS would appear to be a pragmatic alternative to Q-ball for detecting intra-voxel structures in voxels with moderate FA.

3. Empirical Estimation of Intra-Voxel Structure with Persistent Angular Structure and Q-ball Models of Diffusion Weighted MRI

This chapter has been adapted from the published work in [104].

3.1. Introduction

Diffusion weighted magnetic resonance imaging (MRI) enables non-invasive milli-metric mapping of local tissue orientation via sensitivity to directional diffusion on the micrometer scale. The tensor model in diffusion tensor imaging (DTI) has been wildly successful for interpreting this signal in cerebral white matter tracts and reconstructing major fiber pathways in the brain as it is sensitive to fiber orientations [35]. To address DTI's shortcomings in regions of complex intra-voxel structure, High Angular Resolution Diffusion Imaging (HARDI) methods acquire and analyze additional data to estimate multiple fiber population orientations per voxel. These methods reveal more elaborate information about the intra-voxel structure, e.g., Diffusion Orientation Transform (DOT) [91], Spherical Deconvolution (SD) [15, 105], Persistent Angular Structure (PAS-MRI) [12] or Q-ball [13].

Previous large-scale reproducibility studies with DTI have been essential for understanding the empirical behavior of the tensor estimator (e.g., [62, 63]). We perform an in-depth comparison of PAS-MRI and Q-ball as representative variants of HARDI methods. PAS-MRI models the intra-voxel diffusion using a discrete number of fiber compartments [106], while Q-ball estimates an orientation distribution function (ODF), which is assumed to reflect the underlying fiber orientation distribution through its impact on diffusivity [101]. PAS has been shown to be more consistent than Q-ball on synthetic data at lower or clinical b-values of about 1200 s/mm2 [64]. Meanwhile, Q-ball is effective at higher b-values (~3000 s/mm2 or greater) [19]. Previous work on Q-ball has also shown that it resolves multiple fiber orientations for fractional anisotropy (FA) regions < 0.4 [94]. While a minimum angle of fiber orientations has not been established for PAS yet, Q-ball has been known to resolve angles down to 45 degrees with diminishing accuracy depending on adequate signal to noise ratio (SNR) [95]. Herein, we compare the empirical

reproducibility of PAS-MRI and Q-ball focusing on b-values of 1000 and 3000 s/mm2. We have also compared the models in terms of agreement between PAS & Q-ball.

In [106], the authors show that PAS can more consistently resolve crossing fiber voxels compared to Q-ball. PAS has also been shown capable of detecting up to two crossing fibers with robustness in the method in presence of noisy data [107]. In [108], improvement in probabilistic index of connectivity (PICo) tractography was shown using a Bingham distribution with compelling PAS results compared with Q-ball. A drawback noted was that PAS-MRI showed a spurious perpendicular peak for fanning structures based on bootstrap validation. In [109], the authors state that the fiber orientation distribution methods have a high agreement with probabilistic tractography and PAS-MRI specifically having a sparse connectivity matrix. It has been strongly suggested for further validation using more complex biological phantoms techniques [110].

Previously, the bias and variance of PAS and Q-ball methods have been assessed using relatively small datasets [12, 13, 101, 111]. Region of interest (ROI) analysis has shown that Q-ball retains angular information when b-value is brought down to the range of 2000 – 2500 s/mm2 [112]. Average angular resolution of Q-ball has been estimated to be between 15º and 30º based on analysis presented in [113]. While analyses of bias have been performed using repeated phantom/synthetic data [95], verification of these results are important with in vivo data. Q-ball can identify crossing fibers, yet it fails in scenarios where there are fanning or splitting of fibers. In [114], a comparative reproducibility study of tractography between PAS-MRI and Q-ball was performed for specific ROI's. Both the methods were shown to be highly reproducible with PAS-MRI being slightly superior. Phantom studies [110] on Q-ball using tractography have suggested the need for more reliable validation methods. Here, we offer an extensive empirical scan-rescan validation set on a single subject acquired at multiple b-values so that the diffusion model fitting methods can be studied on a range of acquisition sequences and quantities of data following large-scale validation work done with DTI [62, 63].

3.2. Methods

Briefly, the data were pre-processed for each scan session individually and successively registered to the Montreal Neurological Institute (MNI) space template (Figure III.15). Gold standard models were created for each method and b-value by concatenating the 11 scans per b-value and fitting with each method. Comparisons were performed across repetitions to assess variance/reproducibility, while comparisons between the gold standard scans were performed to assess bias.

3.2.1. DW-MRI Data Measurements

A healthy volunteer was scanned in three different sessions on successive days on a 3T Phillips Scanner with a 32-channel head coil. The first and last scan sessions consisted of four repetitions of 96 gradient directions per b-values at each of 1000 and 3000 s/mm2. The second session had three repetitions of the protocol. For each b-value shell, ten minimally weighted reference images (b0's) were also acquired. Voxel resolution for the data is 2.5mm x 2.5mm x 2.5mm with a matrix of 96 x 96 and 38 slices. The scan parameters were: Multi-Band=2; SENSE=2.2; TR= 2650 ms; TE=94 ms; partial Fourier=0.7. Fold over direction was A-P with a P fat shift. For each set of 5 shells, an additional diffusion scan was acquired with reverse phase encoded volumes (i.e., fold over direction A-P with A fat shift) with a minimally weighted volume and 3 diffusion weighting directions with a b-value of 1000 s/mm2 along the imaging frame cardinal directions, and all other parameters were kept constant. All data were acquired in accordance with the Vanderbilt University Institutional Review Board (IRB) guidelines and with the signed consent of the volunteer.

3.2.2. Preprocessing

Each session was corrected for eddy current motion, patient head movement and susceptibility

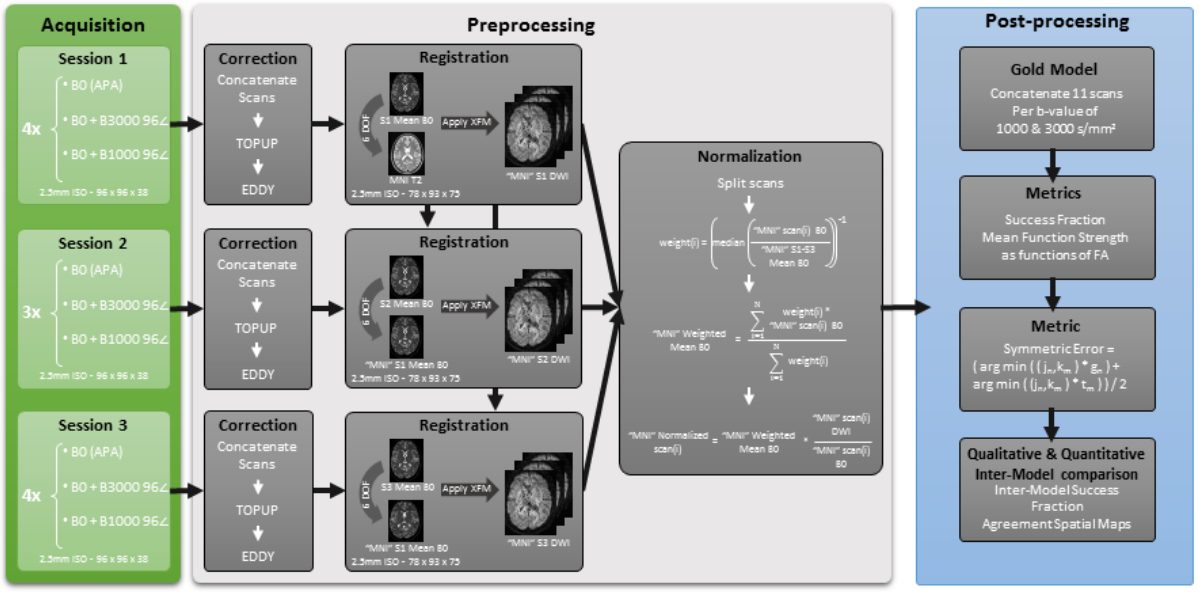

Figure III.15. Flowchart depicting the pre-processing pipeline

distortion with FSL's topup and eddy [115, 116]. Every diffusion weighted shell was preceded by a non-diffusion weighted image (b0) averaged ten times on the scanner, and each session had four reverse phase encoded b0 volumes. All b0's including reverse phase encoded b0's in a session were concatenated and fed as inputs to topup. For eddy all scans in a session were concatenated and then corrected together using the results from topup). Once topup and eddy had been performed the corrected b0's from the first session were registered to a 2.5mm isotropic MNI T2 which was resampled from a 0.5mm weighted template using six degree of freedom registration [117]. The volumes for the next session were brought to the same space by averaging the b0's together and registering the combined image to the averaged b0's in the first session. The same procedure was subsequently done to the last session to bring all three sessions into a common subject-specific pose aligned with MNI space.

Thereafter, all the sessions were normalized by the b0 corresponding to the scan to account for amplitude drift (Figure III.15). A weighted mean b0 was created from all b0 scans in MNI space by a weighted average of all b0's. The weight was taken as the inverse of the median of all b0 scans in MNI

space averaged with the mean of b0's from all sessions. The final normalization scan was computed by multiplying the weighted mean b0 to the ratio of diffusion weighted scan (in MNI space) to the b0 (in MNI space).

3.2.3. Data Segmentation:

Individual target images (T1 weighted scans) were affine registered to the MNI305 atlas [118] and bias corrected with N4 [119] using Advanced Normalization Tools (ANTs) [120] on the atlas and the target images. Non-rigid registration was performed from atlas images to the target image using Advanced Normalization Tools (ANTs) and symmetric image normalization algorithm (SyN) [121]. Image and label volumes for the atlas were then deformed to the target space with bi-cubic and nearest-neighbor interpolation and fused with non-local spatial STAPLE [122, 123] and Adaboost correction [124]. Each individual voxel in the brain was labelled to one of the 133 labels obtained from the multi-atlas labelling using the BrainCOLOR protocol [125]. T1 image labels were brought back to original target space with the ANTs inverse transformation. All WM labels were concatenated together to create a WM mask. The mask has been used for all reported results with the only exception of the visual glyphs where we have shown all data for the particular slice.

In order to avoid circularity of defining white matter by FA and then examining FA in white matter, we defined the white matter mask based on multi-atlas segmentation of the T1-weighted MRI (Figure III.17 & Figure III.16). As the histograms show, the structural definition spans the full breadth of white matter configurations from homogenous major tracks, to crossing fibers, and even to some regions of partial volume (e.g., very low FA).

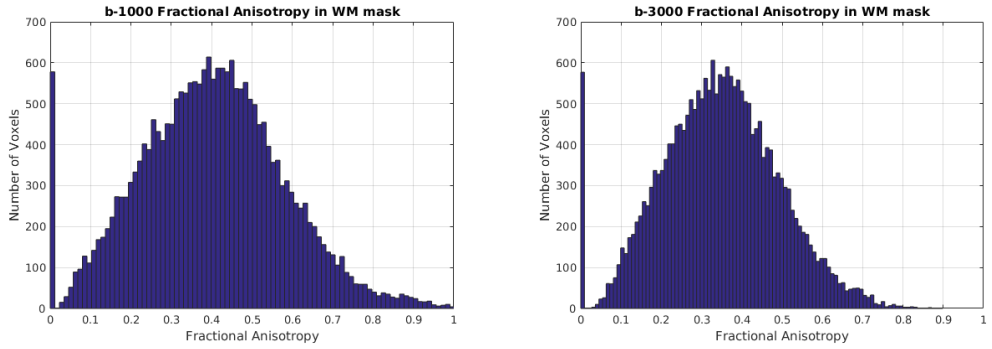

Figure III.16. Histogram of FA values in the white mask for both b-values of 1000 and 3000 s/mm^2.

3.2.4. Q-ball Reconstruction

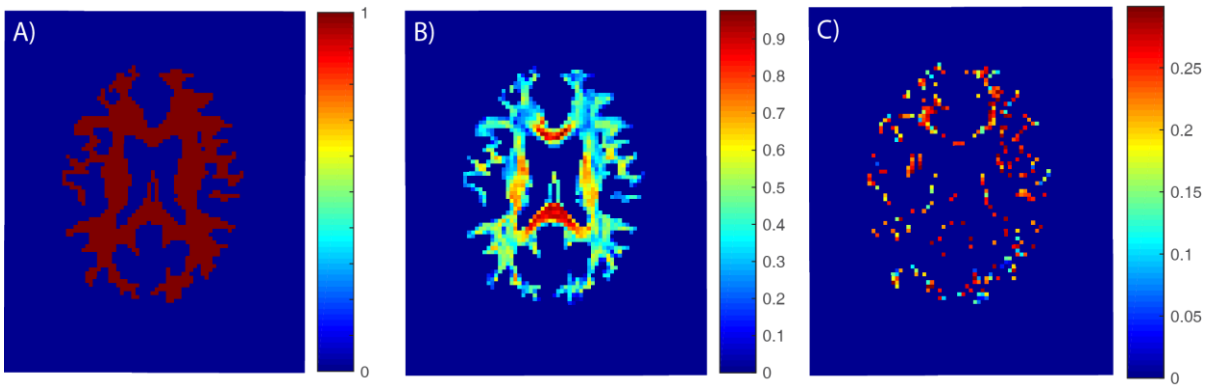

Figure III.17. Spatial maps of middle axial slice A) white matter mask B) fractional anisotropy in white matter mask C) where fractional anisotropy < 0.3 in white matter mask.

Q-ball's orientation distribution function (ODF) was originally calculated using a funk radon transform [13] using a sphere in q-space on raw diffusion data. In [101], the authors showed that a faster and robust q-ball model can be formed and regularized using spherical harmonics.

The spherical harmonic q-ball has exhibited different behavior with spherical harmonic orders, which keep even orders to retain symmetry. Previous results have revealed spherical harmonic order shows dependence on the number of gradient directions and b-value [20]. Additionally, lower order harmonics lead to less angular resolution. High spherical harmonic orders and high b-values yield decrease in angular error for high SNR sequences for b-values up till 6000 s/mm2 [20]. Herein, the parameters chosen were order 8 for the spherical harmonics and the default regularization parameter which is 0.006. Camino was used for the implementation [126].

3.2.5. RE-PASMRI Reconstruction

PAS is a special case of spherical deconvolution. The radius of sphere defined for PAS is a constant parameter and set to 1.4 as was determined in [12]. Reduced encoding PAS was shown to have a good tradeoff between speed and accuracy [111], and reduced encoding was used herein to improve computational efficiency. PAS was reconstructed from maximum entropy spherical deconvolution [96] with the reduced encoding model [111] with a reduction factor of 16. The gold standard sets of the PAS were computed in an "embarrassingly parallel" computation model (split across axis stacks) across multiple CPU's so as not to affect the algorithm but ensure computation at reasonable speeds. Camino was used for the implementation [126].

3.2.6. Peak Reconstruction

Peak search was performed using sfpeaks from camino [126]. Local maxima are determined within a fixed search radius using Powell's method. The radius was specifically set to 1.4 for PAS as per [12] from the default 0.4 which was used for Q-ball. The number of peaks being determined for both the methods was set to 3. The pdthresh defines a ratio of peak strength to mean of the basis function values was set to 1 for both the methods. In brief, the parameters chosen for the two methods have been the suggested optimized parameters for these methods given the acquisition parameters. Lowering of the pdthresh could lead to detection of spurious/false peaks.

Note that the search radius of sfpeaks is note the same parameters as the radius constant in the PAS

model. Peak finding algorithms have a separate search radius parameter however it is set as per the model. Maximum entropy models require a higher radius because of the definition of the model as per [15, 91]. While the default parameters of the sfpeaks have been used for Q-ball. Usage of higher radius such as 1.4 for Q-ball will lead to unstable or spurious results.

3.3. Metrics

3.3.1. Success Fraction

Success fraction (SF) has been effectively applied to characterize synthesized data on a small scale in vivo dataset [64], originally termed as consistency fraction in the article. Success fraction declares two intra-voxel measurements to be in agreement if (1) the number of fiber populations is equivalent and (2) the peaks are within an angular tolerance. In [64], 18 degrees was used as the tolerance; herein we use 20 degrees.

All metrics have been binned at intervals of 0.1 by FA. The binned values have been represented by the mean with standard deviations across the bin.

3.3.2. Peak Fraction

The peak fraction is a representation of the function value detected from the PAS or the ODF. We have restricted the analysis to scenarios of <= 3 fiber populations per voxel and that had been defined during the peak search algorithms as well. The ODF/PAS values were normalized in cases where fiber populations detected were > 1. It has been defined as a function of FA and is also mapped to the number of voxels. We have chosen to represent this for all the eleven single test models per b-value for both the methods, where f_i^n are the normalized peak fraction (PAS/ODF) values of the peaks detected and fi is the function value (PAS/ODF) of the peak detected:

$$f_i^n = f_i / (f_1 + f_2 + f_3) \qquad \textbf{Eq. 18}$$

3.3.3. Symmetric Angular Error

Symmetric angular error presents an insight to bias of the reproducibility of the fiber populations

being detected from the PAS and Q-ball functions and a different perspective from the SF as it combines it with quantitative peak fractions. It informs us about quantitative error presence even in cases where the gold standard model detects two populations and the test model detects three. Consider two vectors, a gold standard vector, jn, and test model vectors, km, along with gold standard weights, gn, and test weights, tm. The symmetric angular error (SAE) is a representation of the orientation error between all the peaks of the gold model and the test model:

$$\left(\arg\min\left(\measuredangle\left(\boldsymbol{j_n}, \boldsymbol{k_m}\right)\right) \cdot \boldsymbol{g_n}\right) + \arg\min\left(\measuredangle\left(\boldsymbol{j_n}, \boldsymbol{k_m}\right) \cdot \boldsymbol{t_m}\right)\Big) / 2 \qquad \textbf{Eq. 19}$$

3.4. Results

With the one-fiber case, SF for PAS and Q-ball shows increasing consistency as a function of FA (Figure III.18 A). It should be noted that SF for PAS at a b-value of 1000 s/mm2 is significantly lower. At

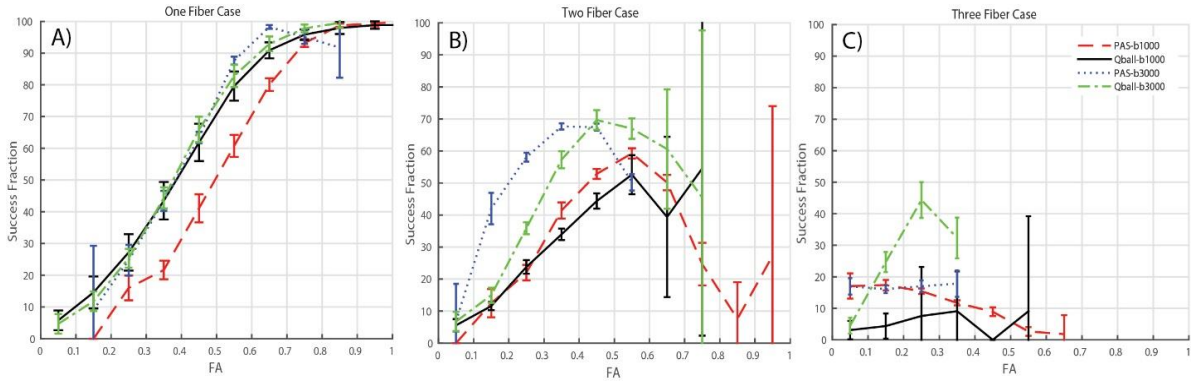

Figure III.18. SF separated in fiber population cases from PAS and Q-ball at b-values of 1000 s/mm2 and 3000 s/mm2. A) Single fiber population detected by gold standard methods. B) Two fiber population detected by gold standard methods. C) Three fiber population detected by gold standard methods. Error bars represent standard deviation across each bin.

the extremes of FA (0.8-0.9 or 0.1-0.2), limited sample sizes lead to higher variance in the estimates (e.g., PAS at a b-value of 3000 s/mm2). With the two-fiber case (Figure III.18 B), SF is maximal at intermediate FA (0.35-0.55 PAS and ~0.55 for Q-ball). With the three-fiber case, SF is generally low for both methods

while Q-ball at a b-value of 3000 s/mm2 shows the highest SF.

Figure III.19 explores both the number of voxels identified along with the FA, and peak fraction. Q-ball shows more voxels with single fiber populations than PAS at both b-values (Figure III.19A). As the b-values increase, PAS exhibits a slight increase in the single fiber voxels while Q-ball finds a substantial

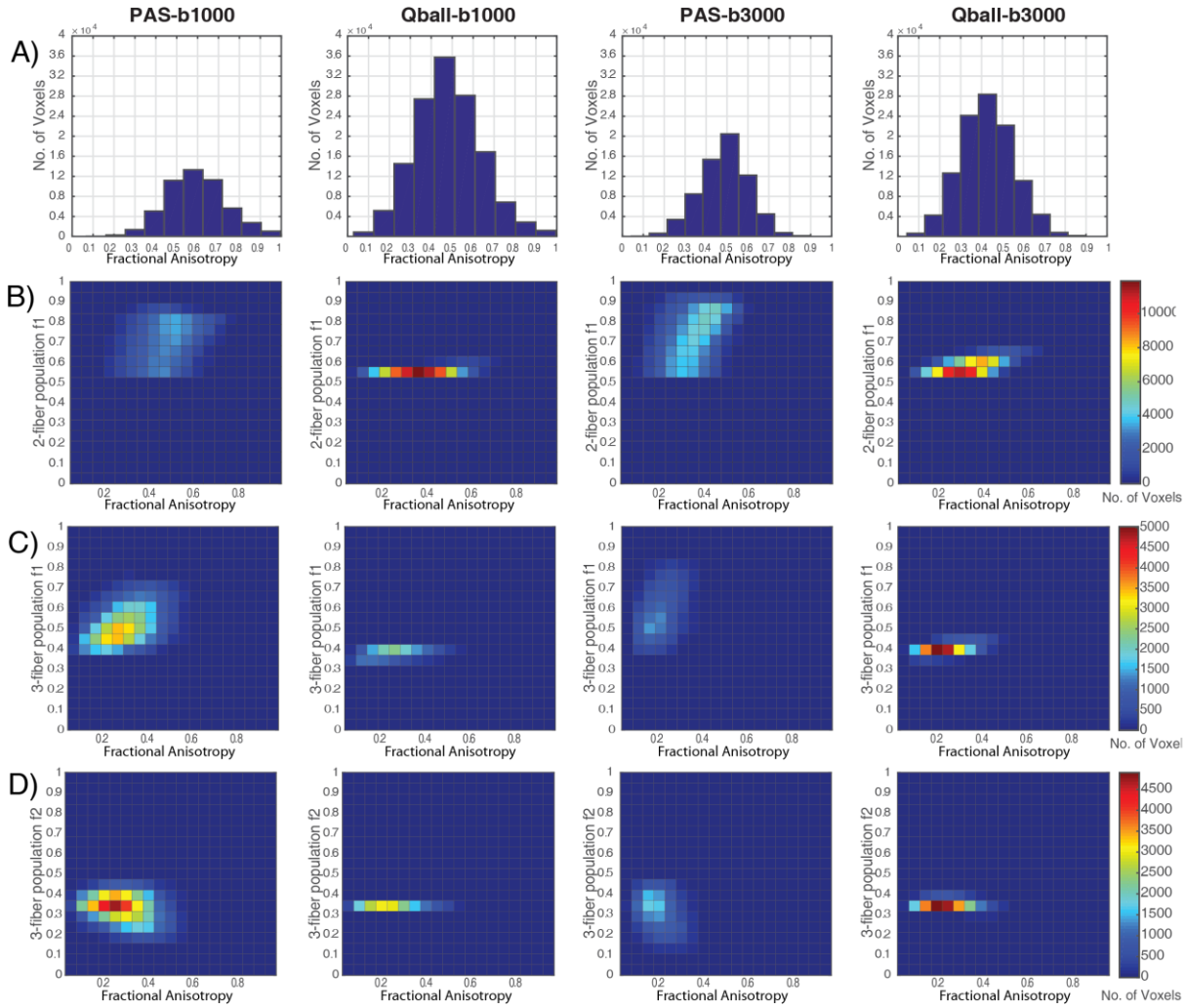

Figure III.19. Peak Fractions of the methods across eleven single scans per b-values of 1000 s/mm2 and 3000 s/mm2. A) Number of voxels mapped as a function of FA. B) Two fiber population f1 as function of FA with number of voxels as the third dimension. C) Three fiber population f1 as function of FA with number of voxels as the third dimension. D) Three fiber population f2 as function of FA with number of voxels as the third dimension.

decrease. With the two-fiber case, PAS estimates a wide spread of peak fraction (0.5-0.8) for a range of FA

(0.2-0.8), but Q-ball finds a very narrow, peak fraction of (0.5-0.6) for a more limited range of FA (0.3-0.6) (Figure III.19 B). The number of two-fiber voxels moderately increases with b-value for PAS and Q-ball. For the three-fiber case, there is substantial decrease in detected fibers with PAS, but an increase with Q-ball. For a b-value of 1000 s/mm2, PAS finds 3 fibers for an FA of 0-0.6, but for a b-value of 3000 s/mm2,

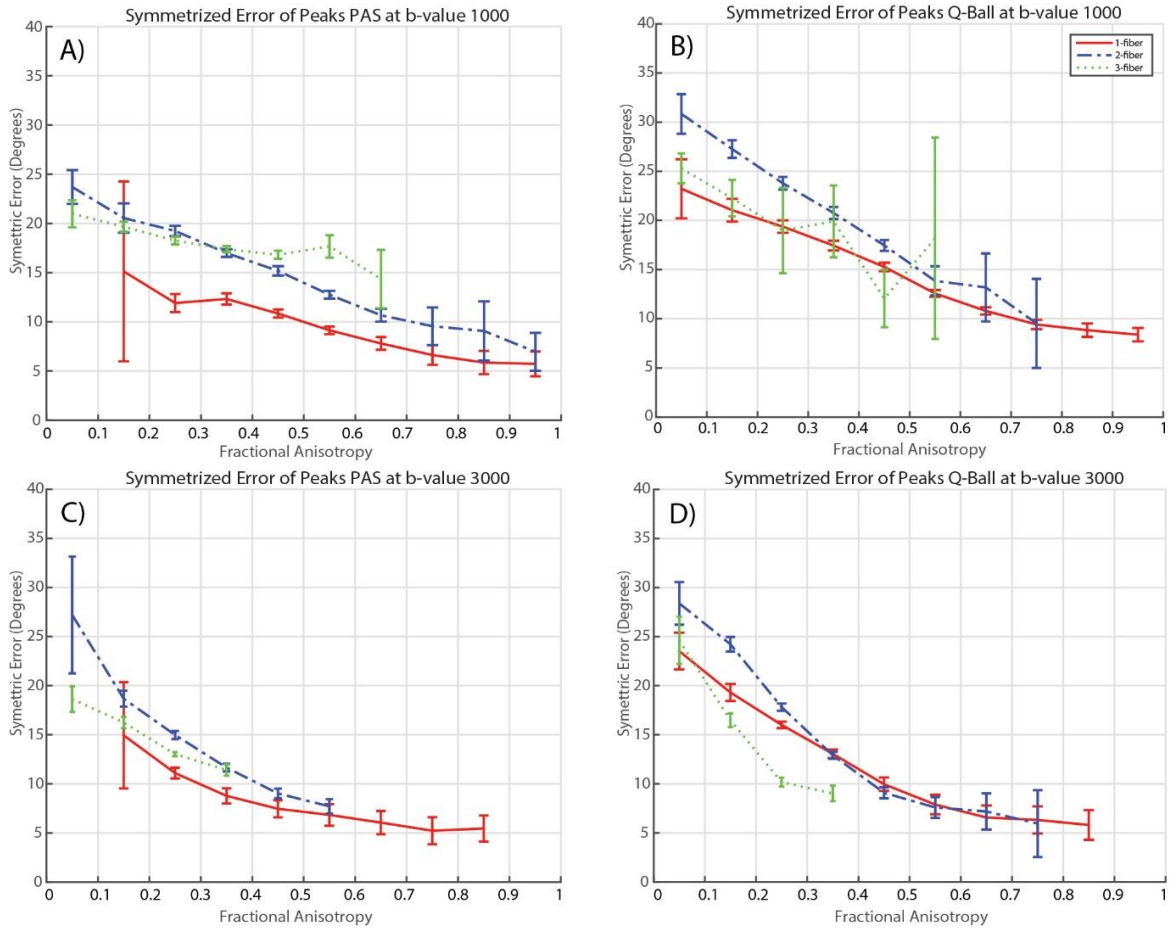

Figure III.20. Symmetric angular error is shown for A) PAS at b-value of 1000 s/mm2, B) Q-ball at b-value of 1000 s/mm2, C) PAS at b-value of 3000 s/mm2, and D) Q-ball at b-value of 3000 s/mm2. Error bars represent standard deviation across the bin.

PAS detects 3 fibers only FA < 0.4. For Q-ball, the FA range mains the same (FA in 0-0.5) with a peak fraction of ~one third for both b-values (Figure III.19 C & D).

The single fiber cases exhibit symmetric angular error less than multi-fiber cases for both PAS and Q-ball (Figure III.20). Symmetric angular error decreases with increased FA for all scenarios. PAS shows

lower symmetric angular error than Q-ball (Figure III.20A vs. B and Figure III.20 C vs. D). Interestingly, symmetric angular errors are lower for the three crossing fibers relative to the two crossing fibers. Yet, note that 7836 and 661 voxels were detected with three fibers for PAS and Q-ball at b-value of 1000 s/mm2

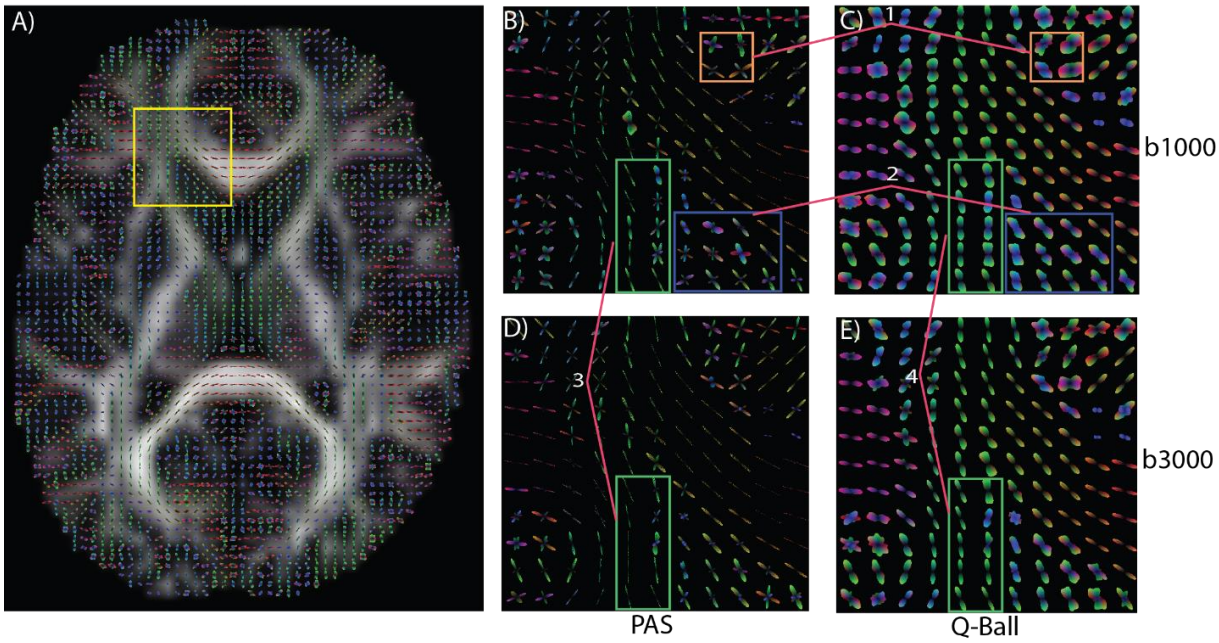

Figure III.21. Enlarged ROI contains genu of corpus callosum and left pre-frontal area: A) Q-ball is shown for the middle axial slice at b-value of 3000 s/mm2. Enlargements are presented for: B) PAS at b-value 1000 s/mm2, C) Q-ball at b-value 1000 s/mm2, D) PAS at b-value 3000 s/mm2, and E) Q-ball at b-value 3000 s/mm2. (1) Structural differences between PAS and Q-ball. (2) Fanning fibers in Q-ball while crossings detected by PAS. (3) Loss of structure for PAS. (4) Reduction in fanning for Q-ball.

while they were 2941 and 816 at b-value of 3000 s/mm2, respectively. Figure III.21 presents a qualitative comparison.

Figure III.22 compares the gold standard estimates from PAS to the individual estimates from Q-

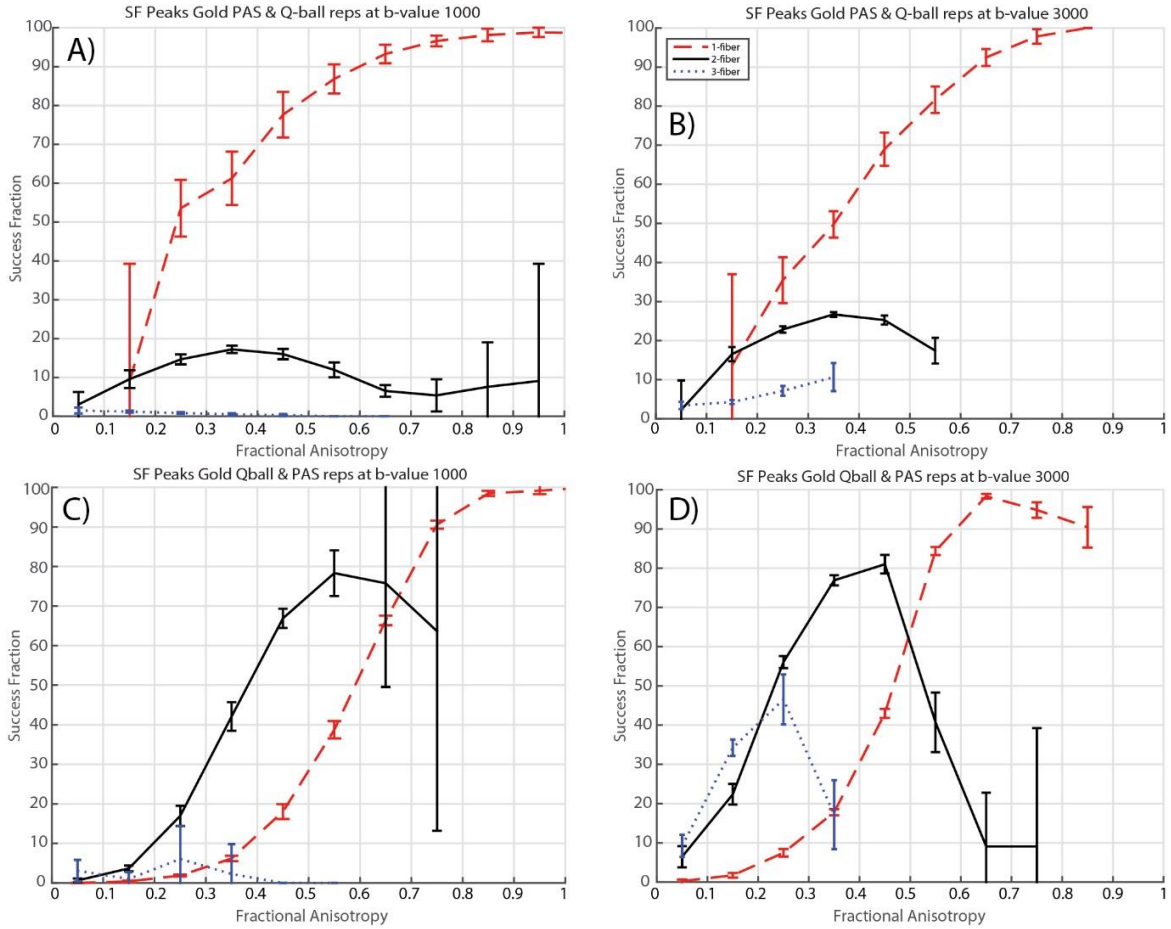

Figure III.22. A) SF gold standard model PAS and test models Q-ball at b-value of 1000 s/mm2. B) SF gold standard model PAS and test models Q-ball at b-value of 3000 s/mm2. C) SF gold standard model Q-ball and test models PAS at b-value of 1000 s/mm2. D) SF gold standard model Q-ball and test models PAS at b-value of 3000 s/mm2. Error bars represent standard deviation across the bin.

ball and vice versa. SF follows the same trend for all four scenarios for the single fiber population. For the single fiber case, SF is higher when the PAS treated as the baseline (red curves: Figure III.22A vs C and Figure III.22 B vs. D). Yet, for the two-fiber model, Q-ball ground truth generally agrees with PAS at high

b-value (>0.5), but a substantial fraction (~80%) of PAS two-fiber cases are not seen in Q-ball (black curves: Figure III.22 A vs C and Figure III.22 B vs. D). The three fiber cases are not consistent for either

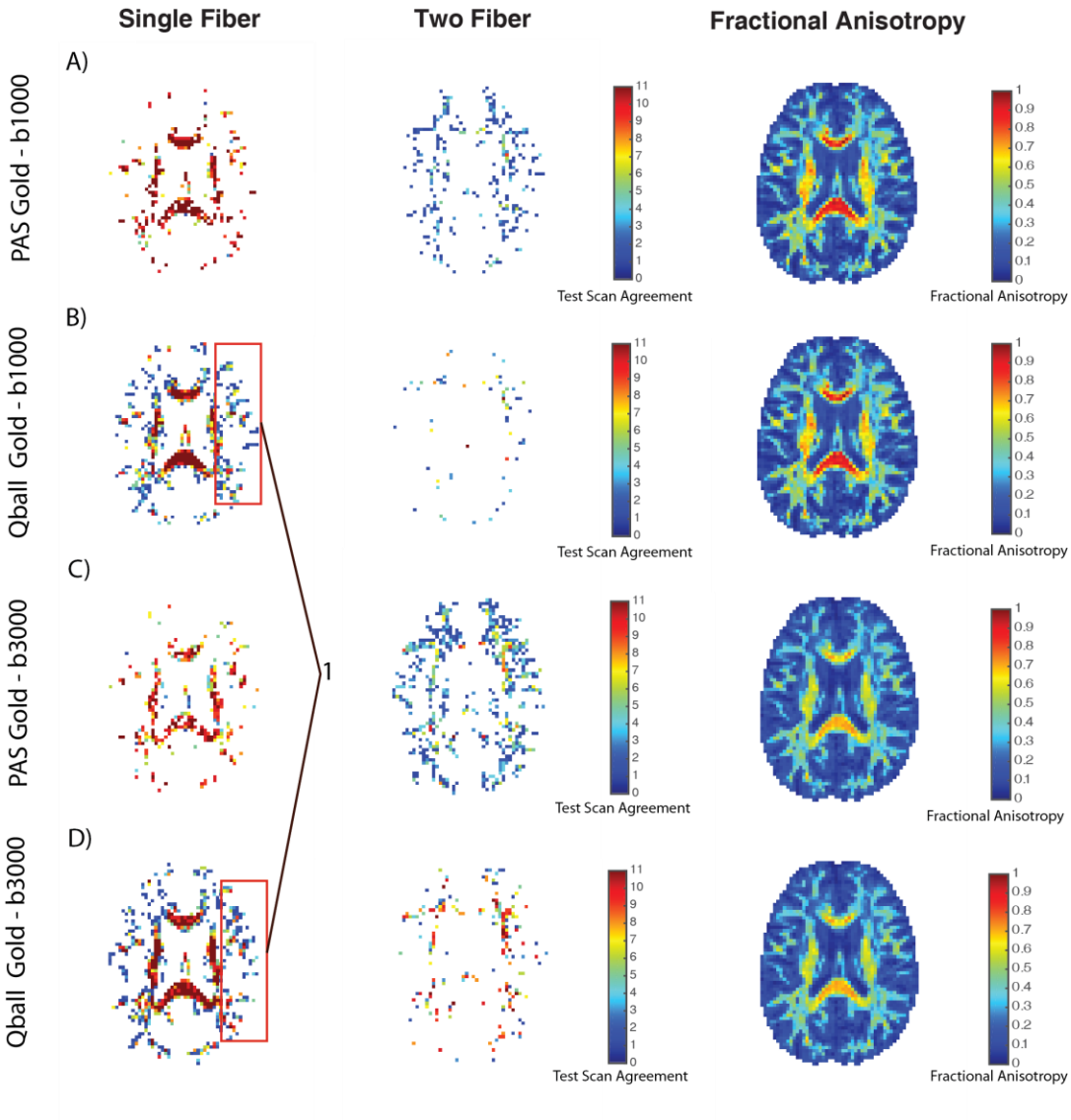

Figure III.23. Middle-axial slice spatial map for intra-model comparison. A) PAS gold standard model and Q-ball test models at b-value 1000 s/mm2. B) Q-ball gold standard model and PAS test models at b-value 1000 s/mm2. C) PAS gold standard model and Q-ball test models at b-value 3000 s/mm2. D) Q-ball gold standard model and PAS test models at b-value 3000 s/mm2. (1) False positives being detected by PAS.

baseline or b-value scenario (blue curves: Figure III.22). We can appreciate these levels of agreement qualitatively in Figure III.23. The major white matter tracts have single fiber detected and have high SF

with both approaches used as gold standards (Figure III.23 first column). The impact of the number of voxels shown with two fibers on the overall SF is shown in the second column of Figure III.23.

Table III.3. Mean SNR and the error observed per session for the acquired dataset.

Session	Mean SNR	Std SNR	Single b0 equivalent
Session 1	32.81	0.1091	10.3768
Session 2	27.84	0.0830	8.8065
Session 3	32.81	0.1091	10.3768

3.5. Discussion

This study focuses on capturing the differences between PAS and Q-ball interpretations of a single acquisition type using the established best practices. We interpret the differences observed in the results to be fundamentally attributed to algorithms, and hence discuss the relative sensitivity/reproducibility of each approach. Note that the observed differences between Q-ball and PAS may arise not only from the data model, but also from their dependencies on experimental factors (e.g., noise level, angular sampling), on the selection of reconstruction/model parameters (e.g., order/radius), peak determination algorithm parameters. We have made the data freely available on nitrc.org to allow deeper exploration of these considerations.

In the literature, Q-ball has been shown to detect crossing angles in multi-fiber regions down to 45 degrees at b-values of 8000 s/mm2. Given sufficient SNR using high quality phantom data, and Q-ball improves on accuracy with b-values > 4000 s/mm2 [95]. Typically, these b-values are not clinically feasible. Crossing fiber angles have been detected with Q-ball for FA less than 0.4 [94]. Meanwhile, PAS resolves greater crossing fiber populations on sparse diffusion data and is able to resolve crossing fibers at lower b-values which are clinically feasible [64].

Herein, we find that both the methods perform well when self-comparing each single scan with concatenated multiple single scans (gold standard model). Specifically, both methods work well on single

fiber population scenarios (however, this is not the application for which the methods were designed to address) (Figure III.18) [12]. However, Q-ball performs better than PAS for single fiber scenarios, which is interesting because it is probable that PAS is detecting false positives. Comparing (Figure III.18 & Figure III.20) we see that specifically for lower FA there is high standard deviation which is indicative of the false positives for PAS. Q-ball performs consistently for this scenario.

With the evaluated imaging sequences, Q-ball and PAS reliably detect multiple fibers only for crossing regions with no more than two fiber populations (Figure III.18). Note that PAS and Q-ball both show low reproducibility for extreme FA values which is a likely artifact. The inference is that the detection of fiber population might not be accurate. At high values of FA, it is more probable that single fiber populations would be expected rather than two fiber populations. While the lower FA regions show disagreement for two fiber populations which is not as extreme as for higher FA values. Suggestively it could be noise that is making the methods underperform in those regions. It is also possible that the methods are not able to resolve more complex architecture, which might be present in the lower anisotropic regions as there is evidence of very low reproducibility in three fiber populations regions or more. Comparison with histological validation may indicate which methods are more accurate in regions of disagreement [23, 127]. It is likely that usage of these voxels might lead to spurious tracts or false continuations.

Comparing (Figure III.18) and (Figure III.20), we find an interesting difference. The symmetric angular error (Figure III.20) shows that most of the errors detected for three fiber populations are less than the two-fiber population scenario. Yet this is contradicted by the SF (Figure III.18) which shows us that there is hardly any agreement for three fiber scenarios. Though our analysis suggests that they are not reproducible there might be a possibility to improve them and increase their reproducibility. Robust fitting of HARDI methods with removal of outlier volumes could lead to improvements.

The peak fractions of Q-ball show a self-consistent, but distinct, value in most scenarios (Figure III.19). While PAS shows them consistently as well but across a wider range. It can be expected from Q-ball that they will be consistent values because a smooth function (ODF) has been normalized. While PAS is a spikier function and hence the wider range of fiber fraction. Characterization of the reconstruction of

PAS and ODF may lead to a better understanding but it is beyond the scope of this article which deals with fiber population and their angular error. A deeper understanding of the differences between ODF and FOD methods are needed, in particular peak fractions lead to quite different interpretations for the different approaches. This is not surprising as the ODF function is quite smooth and not intended to directly model the fiber fraction. For typical clinical b-values (~1000 s/mm2) for PAS offers advantages for sensitivity and reproducibility. Q-ball can achieve similar reproducibility performance as PAS given moderate increases in SNR or directions [95]. For higher b-values, Q-ball detects more detailed micro-architecture in the brain (Figure III.16) at the higher b-value which implies more voxels detected with crossing fibers. This is supported by the fact that symmetric angular error is lower for single and two fiber populations for higher b-values (Figure III.20). PAS appears to be a more reasonable choice given higher reproducibility of crossing fiber majorly of the two fiber populations (Figure III.18) and (Figure III.20). However, care should be taken in interpretation of both methods as to which regions they are applied as both methods are unstable at the extreme of FA (Figure III.17).

Prior conference analysis of the data [19], showed that PAS resolves crossing fibers more consistently than Q-ball at moderate and higher crossing angles. Accuracy and consistency for both Q-ball and PAS have been shown to increase with increasing SNR (our empirical data is low SNR) [13], but the observed methodological effects are not mitigated by large quantities of data as indicated as at the comparison between gold standard models computing using all available data (Figure III.17) and (Figure III.18). In [111], PAS was consistent in all cases with SNR > 16, but for Q-ball to reach this level of consistent/accuracy SNR >24 (which ours is not, and is not often seen in clinical scans). Note that the SNR of the presented data were 11.828 in the centrum semiovale (WM), in the peripheral white matter, and 5.838 in the cortical gray matter in the b0 images. Also shown are values of single b0 equivalent as our scanner provides with an average of 10 b0's for a b0 (refer Table III.3 and Table III.4). Hence, single b0 equivalent is (SNR calculated)/$\sqrt{10}$. The SNR has been calculated using Reeder's difference method [128]. PAS accuracy and consistency is better than Q-ball for crossing fibers at our low/clinically feasible SNR regime,

Q-ball is more consistent than PAS in single fiber regions at a lower SNR. This is interesting to note, probably due to the false positives in PAS.

Table III.4. Mean SNR and the error observed across all sessions for the acquired dataset.

Session	Mean SNR	Std SNR	Single b0 equivalent
All sessions	33.12	0.0945	10.47

Q-ball shows fanning fiber voxels while PAS detects sharp and narrow peaks as visually evident by the glyphs (see (1), Figure III.19). Higher number of crossing fiber voxels are seen for PAS when compared to Q-ball glyphs which also reinforces the fact that there are more 2 or more fiber voxels for PAS (see (2)). Highlighting architectural differences between the two methods. The effect of b-value can be seen on both the methods as it increases from 1000 s/mm2 to 3000 s/mm2. It is evident that fanning of the fiber reduces, and peaks become narrower for Q-ball and PAS both (see (3, 4)). However, for Q-ball it is beneficial for PAS it is detrimental. It is noticeable qualitatively that the fanning structure being captured by PAS-MRI at b-value of 1000 s/mm2 is lost once the b-value is increased (see (3). This loss of structure is also being quantified by the symmetric angular error metric (Figure III.18). The crossing fiber voxels being detected till FA of 0.8-0.9 at b-value of 1000 s/mm2 has dropped to FA < 0.6.

At high values of FA (FA > 0.6), there is noticeable disagreement between Q-ball and PAS which is evident even for the two-fiber case (Figure III.20). PAS detects multiple fibers even when the tensor FA is high, which would appear to indicate false positives (Figure III.21, see (1)). Meanwhile, Q-ball consistently finds a single fiber in cases where the tensor FA is high.

In summary, the analyses for single-fiber populations are in good self-agreement with PAS, however there is the possibility of false positives. Q-ball performs qualitatively better with fair inter-model agreement. For two-fiber populations, the methods are in fair self-agreement except for extreme high or low FA regions. Overall PAS shows more reproducibility and the inter-model agreement is reasonable for mid-ranges of white matter FA (0.4 < FA < 0.6). The three-fiber population case shows low self-agreement indicating model instability for both PAS and Q-ball. There is little agreement between both the methods

in terms of crossing fibers (Figure III.20, Figure III.21). Visually also the agreement is less than expected given crossing fiber regions or two and more fiber population voxels.

4. SHARD: Spherical Harmonic Based Robust Outlier Detection for HARDI Methods

This chapter has been adapted from the published work in [129]

4.1. Introduction

Diffusion weighted magnetic resonance imaging (DW-MRI) measurements provide contrasts sensitive to the micro-architectural environment at a millimeter scale. The data have been widely used to model brain structural connectivity using diffusion tensor imaging (DTI) [35]. However, DTI is limited in that it only provides a single fiber/peak orientation per voxel. More recent methods which can be labelled collectively as high angular resolution diffusion imaging (HARDI) provide models capturing information on the existence of multiple fiber orientations per voxel. Yet, HARDI is not a singular technique; rather, numerous methods exist to model DW-MRI data with more complex intra-voxel models such as Q-ball, spherical deconvolution, PAS-MRI, diffusion orientation transform [12-14, 90, 101].

Generally, acquisitions consisting of 45 or more diffusion weighted volumes can be termed as HARDI acquisitions [130]. With increased number of gradient volumes per acquisition the probability of artifacts that might invade the acquisition also increases. Artifacts occur not only due to noise or system instabilities of a scanner but also from physiological motion of the subject [131], e.g., signal dropout, signal intensity spike, ghosting and striping (see Figure IV.24). A single artifact in a gradient volume can be detrimental for voxel-wise fitting methods and can lead to increased bias and variance for the methods. It has been shown that a single corrupted gradient volume for a HARDI acquisition can introduce an error of approximately 10% in Fractional Anisotropy (FA) and General Fractional Anisotropy (GFA) [132] using robust spherical harmonics Q-ball imaging [101]. Generally, the practical quality control advice for DW-MRI is to either remove gradient volumes or perform correction using interpolation techniques [132]. Hence, it is critically important for the artifacts to be detected and either removed or imputed for reduction of bias in HARDI methods.

Previous work on robust estimation of DTI has been characterized with weighted least squares fit

and robust M-estimators [133, 134]. Briefly, outlier detection models have been designed to exclude and then compute DTI with the remaining measurements. Subsequently, iteratively reweighted linear least squares (IRLLS) were developed to increase the speed with DTI [135]. To date, a consensus has not been reached for HARDI outlier detection. One approach uses compressed sensing on q-space resampling as an alternative to excluding gradient volumes [136], which is effective when there is signal dropout because of bulk motion. An adaption of the robust estimation of tensors by outlier rejection (RESTORE) [133] uses spherical harmonics for outlier detection and exclusion, higher order model outlier rejection (HOMOR) [137]. HOMOR is effective at removing outliers at higher b-values (up to 3000 s/mm2 where RESTORE suggests invalid outliers), while HOMOR maintains similar performance to RESTORE at lower/clinical b-values. Outlier detection and replacement has also been incorporated in the non-parametric framework of eddy [138]. Eddy's non-parametric model achieves superior performance over RESTORE [139]. Here we focus on outlier detection independent of registration and distortion correction. In light of the advanced work with tensors, one could construct statistically robust variants of each traditional HARDI model. However, such efforts would require specific coding for each HARDI variant and potentially dramatically increased computational times given the need for re-computation of fitting procedures on a voxel-wise basis. Here, we follow in HOMOR's approach for outlier detection using an iterative linear least squares model, but our approach uses residual-based re-weighting with a ridge regression with L2 penalty. HOMOR does not impute the data and is a detection technique while our approach is both a detection and replacement model on a voxel by voxel basis.

Here, we propose a robust outlier imputation model that uses an iterative regularized weighted linear least squares (IRWLLS) fit which functions per voxel. The IRWLLS model uses even order weighted spherical harmonics which can be used to calculate residuals with the original signal intensities. The weights are based on squared residuals and squared standard deviation (SD) of residual between signal intensities and regular spherical harmonics across the entire acquired volume of the brain. Based on a certain threshold multiplied with the SD of the entire brain residuals, the outliers can be detected and imputed with weighted spherical harmonic intensities. This work reveals that an imputation model can be useful for

detecting artifacts and it minimizes error when compared with an in vivo acquired ground truth (good scan).

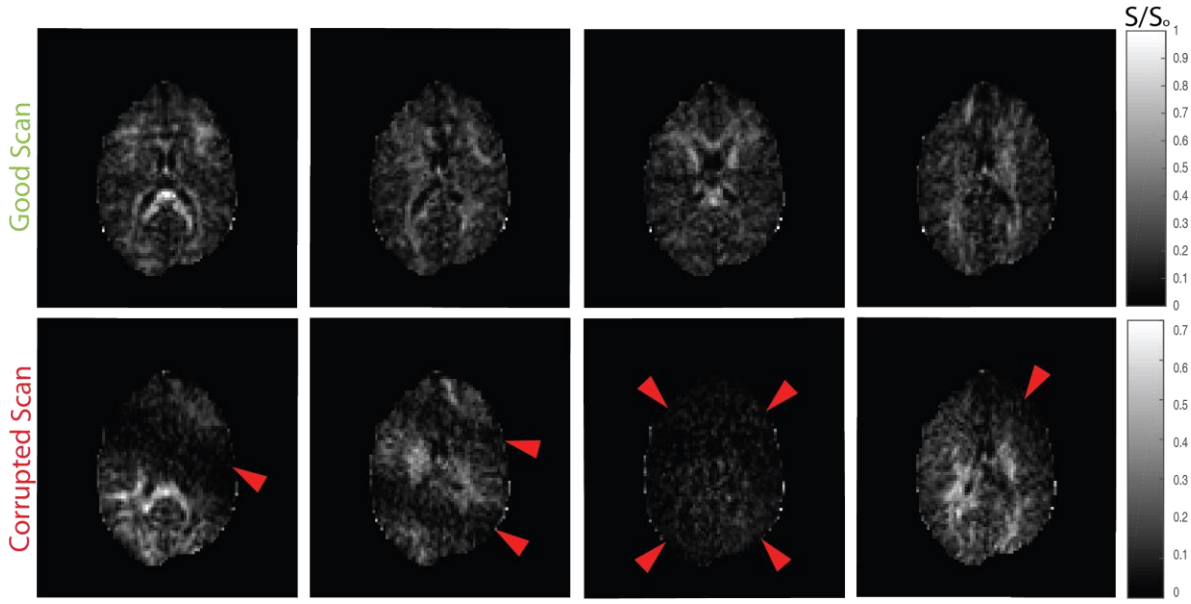

Figure IV.24. Comparison of good and corrupted scan data from an empirical in vivo acquisition. The red arrows in the second-row highlight areas of artifact.

In an empirical demonstration, the improvement is noticeable for Q-ball in terms of angular correlation coefficients.

4.2. Methods

4.2.1. DW-MRI Data Measurements and Pre-processing

A single healthy volunteer was scanned for a single session at 3T (Achieva, Philips Medical System Systems, Best, The Netherlands) with a 32-channel head coil. Two scans were acquired each with 64 gradient direction for a b-value of 3000 s/mm2 each with a minimally weighted reference (b0). Acquisition parameters voxel resolution=2.5x2.5x2.5 mm3 Multi-Band=2; SENSE=2.2; TR= 2650 ms; TE=94 ms; partial Fourier=0.7. In one of the two scans, the volunteer was specifically asked to move their head at random intervals. This was done to introduce artifacts in the acquired scan. One third of the gradient volumes show major artifacts in the middle axial slice and can be noticed easily visually (refer Figure

IV.24).

The acquired data wer preprocessed for patient movement, eddy current distortions and susceptibility distortion using eddy and topup [139, 140]. The b0's were concatenated together and fed as input to topup. Both the scans were concatenated together and then used as input to topup which was pipelined to eddy. The corrected data from eddy were successively registered to the b0's using flirt [99,

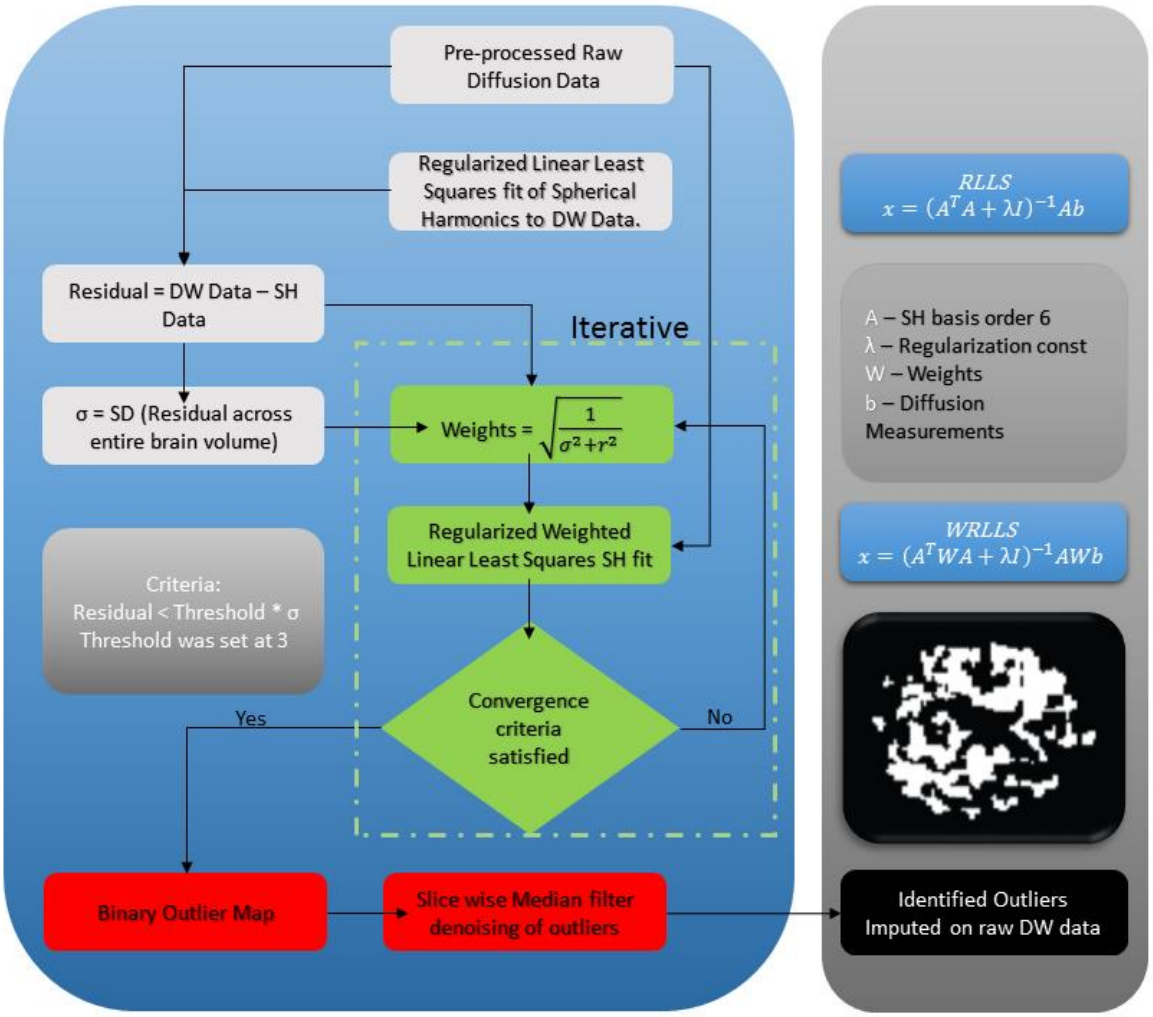

Figure IV.25. SHARD Algorithm's iterative pipeline. The first phase of SHARD (white boxes), involves a regularized spherical harmonic fit to the preprocessed diffusion data, which enables estimation of the standard deviation of the model fit across the brain (σ). The second phase (green boxes) uses iterative reweighted spherical harmonic fitting to identify and de-weight outliers on a voxel wise basis. Finally (red boxes), outliers are identified from the reweighted spherical harmonic fit, median filtered, and imputed with their model fit values. The resulting data can be used in subsequent HARDI model processing (black box).

117]. The two scans were normalized by the corresponding b0 to account for the amplitude drift.

4.2.2. SHARD Algorithm

The pre-processed data was fitted to spherical harmonics of order 6 to generate 28 basis functions. The SD was calculated of the residuals between the DW-MRI data signal intensities (S/So) and the spherical harmonic fitted signal intensities for both the scans (σ) (Figure IV.25). The spherical harmonic basis functions were fitted to DW-MRI using a regularized linear least squared fit (RLLS) and is also known as a ridge regression with L2 penalty. Regularization constant was fixed at 0.005 as the known used value from Q-ball. This algorithm functions on a voxel by voxel basis iteratively. Weights have been calculated as square root of inverse of sum of squares of σ and residual per voxel leading to an equivalent number of weights as the gradient directions. The original DW-MRI data are fitted using weighted spherical harmonics basis functions of order 6. A weighted regularized linear least squared fit (WRLLS) has been used iteratively till the convergence criteria are satisfied. The criteria were set to three deviations of σ for the residual. Failure of satisfaction of the criteria leads to recalculation of weights using any signals that were imputed. At every iteration, the outlier binary map is updated. Once the procedure has been completed for the entire brain volume a final outlier map is generated. This outlier map is denoised/interpolated using a two-dimensional median filter to get rid of spurious outliers that were detected during the iterations of the SHARD algorithm. Denoised outlier map is used for imputation of the original DW-MRI measurements which are replaced by the sampled weighted spherical harmonic signals for the true outliers that were detected. This algorithm can be used for either exclusion or imputation functionality or for both. However, in regard to this work it has been used an exclusion and imputation model.

4.2.3. Validation of SHARD in Simulation

The ground truth (good scan) was taken and artificial noise was added per voxel at gradient directions that were chosen randomly. ‘S’ defines the signal intensity for a diffusion weighted volume while ‘So’ defines a non-weighted diffusion volume. For a randomly chosen gradient direction a probability (p) was randomly generated. If $p >= 0.5$ then the corruption factor was used to divide S/So and if it was $p < 0.5$

then corruption factor was used to multiply S/So. This experiment was repeated 64 times per corrupted direction while increasing number of corrupted directions up to the maximum number of gradient directions that were acquired with this data set. This synthetic corruption model has been labelled as 'Slash Model' for future reference and in the figures as well. We used a two-step validation process the first one being the root mean squared error measured against the ground truth across all the signal intensities (S/So) over the entire brain volume comprising of all gradient directions (1). 'G' is the total number of gradient directions and 'N' is the total number of slices that were acquired.

$$\boldsymbol{RMSE}\left(\frac{\boldsymbol{S}}{\boldsymbol{S_o}}\right) = \sqrt{\frac{\Sigma_1^G \Sigma_1^N (\frac{\boldsymbol{S}}{\boldsymbol{S_o}}(\boldsymbol{truth}) - \frac{\boldsymbol{S}}{\boldsymbol{S_o}}(\boldsymbol{imputed}))^2}{\boldsymbol{G \cdot N}}} \qquad \textbf{Eq. 20}$$

The second step validation is using angular correlation coefficient (ACC) which is based on the coefficients of the HARDI method Q-ball (2). ACC is a convenient way for calculating correlations of functions of all directions over a spherical harmonic expansion. It provides an estimate of how closely two orientation distribution functions (ODF) relate to each other. It was first estimated on a voxel by voxel basis. 'u' is the test function of Q-ball coefficients while 'v' is the true set of Q-ball coefficients from the ground truth (good scan) data. It can be inferred that the ACC if taken between two separately calculated Q-ball coefficients from ground truth data its value will be equivalent to 1. This was empirically determined as well. This metric has been represented across the entire brain volume using RMSE and the mean.

$$\boldsymbol{ACC} = \frac{\Sigma_{j=1}^{\propto} \Sigma_{m=-j}^{j} \boldsymbol{u_{jm} v_{jm}^*}}{\left[\Sigma_{j=1}^{\propto} \Sigma_{m=-j}^{j} |\boldsymbol{u_{jm}}|^2\right]^{0.5} \cdot \left[\Sigma_{j=1}^{\propto} \Sigma_{m=-j}^{j} |\boldsymbol{v_{jm}}|^2\right]^{0.5}} \qquad \textbf{Eq. 21}$$

When performing this robust experiment on the ground truth (good scan) and the motion scan data the same methodology as above has been used. However instead of multiplying or dividing signal intensities by a corruption factor, random permutations of the gradient volumes were chosen from the motion scan

data. The permutations ranging from 1 to 64. At each level of corrupted directions, the experiment was repeated 64 times for robustness.

4.3. Results

4.3.1. True and False Positive

True positives have been defined as outliers that are correctly identified by the model (true outliers are known because they are simulated) while false are the ones that are detected but they are not in actual. The number of false positives being detected for both the slash model and when tested with motion scan data are close to 1000 voxels combined across all the uncorrupted gradient volumes (Figure IV.27A and C). Specifically, when motion corrupted scan volumes are swapped, it is noticeable that there are visual inconsistencies. The true positives in the corrupted gradient volumes are consistently greater and at a much higher ratio than the number of false positives being detected (Figure IV.27B and D). The number of true positives drop in cases of certain corrupted volumes and are accounted to gradient volumes that are not corrupt in the motion scan data. The ratio is ~1:60 when using motion data scan to introduce corruption. The ratio is even higher in the scenario of slash model true positives and false positives. Qualitatively, the above-mentioned result is re-enforced as it is apparent to notice that there are fewer outliers being detected in the uncorrupted gradient volumes (Figure IV.26B and D). This result is regardless of slash model corruption of volumes or swapping corrupted volumes from motion scan data.

4.3.2. RMSE of Signal Intensity

The RMSE of 64 repeated measurements with random permutations that the overall error after using this algorithm for pre-processing is always less than the original error (Figure IV.28A and B). This can be observed for both slash model corruption and motion scan corruption. A distribution plot of the 64 repeated measurements for the first 20 corrupted directions shows that either the RMSE is reduced or it is the same when compared with the original error (Figure IV.28C and D). The RMSE increases for the original corruption and after imputation with the increase of corrupted gradient directions. However, the

RMSE after imputation is generally less by approximately 0.06 for the S/So. The base error that is introduced by SHARD in the ground truth with no corrupted volumes is approximately 0.015 for S/So (Figure IV.28D). The distribution plots show that there is a higher likelihood of having a lower RMSE as compared to original RMSE after imputation.

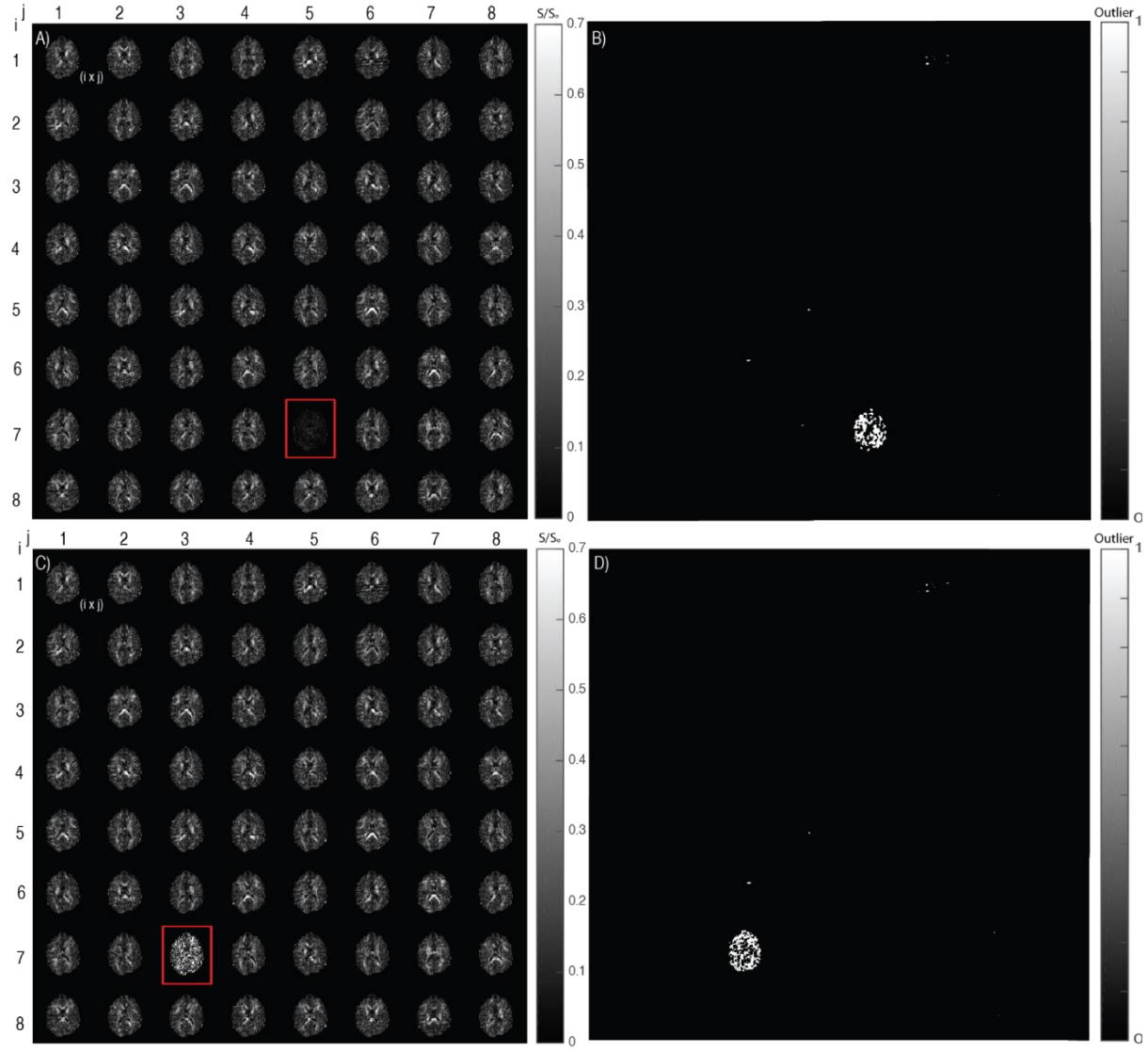

Figure IV.26. Middle axial slice shown for the diffusion data and the binary outlier map. i*j denotes the weighted gradient volume corresponding to gradient directions. A) Ground truth data when corrupted with an artifact containing gradient volume from the motion scan. B) Outlier detected on the corrupted gradient volume in (A). C) Simulated good scan data when corrupted artificially by random multiplication and division of signal intensity using a corruption factor (Slash Model). D) Outliers detected with the artificial corrupted gradient volume in (C).

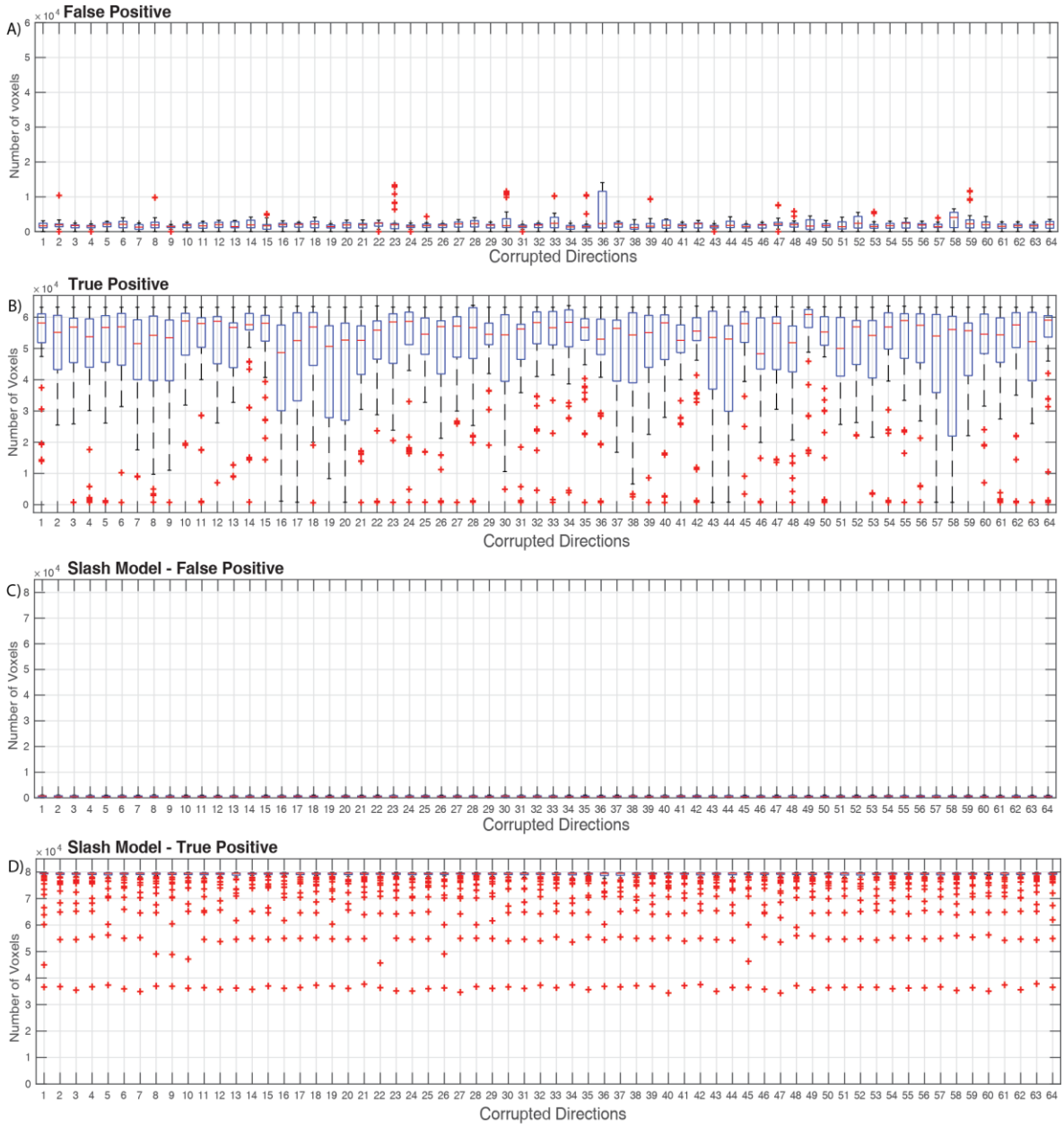

Figure IV.27. Count of outliers detected in corrupted gradient volumes (True Positive) and uncorrupted gradient volumes (False Positive). A) False positive, outliers detected in the uncorrupted gradient volume when using corrupted gradient volumes from motion scan. B) True Positive, outliers detected in the uncorrupted gradient volumes which has random volumes swapped in from the motion scan. C) False positive, outliers detected in the uncorrupted gradient volume when using Slash model corruption factor for random gradient volumes. D) True Positive, outliers detected in gradient volumes when using Slash model corruption factor for random gradient volumes.

4.3.3. RMSE of ACC

RMSE of ACC across the entire brain shows us that the error is consistently reduced when compared to the ground truth ACC (Figure IV.29A and B). This result holds true for both slash model and corruption using motion scan data. It is noticeable, however, that the error reduction in the scenario of motion scan data is comparatively lesser as compared to error reduction in slash model. The RMSE of ACC increases with increasing number of corrupted gradient directions and so does the RMSE of ACC after imputation. However, the constant error reduction can be noticed at about 0.03 for the slash model and 0.02 for the motion scan data.

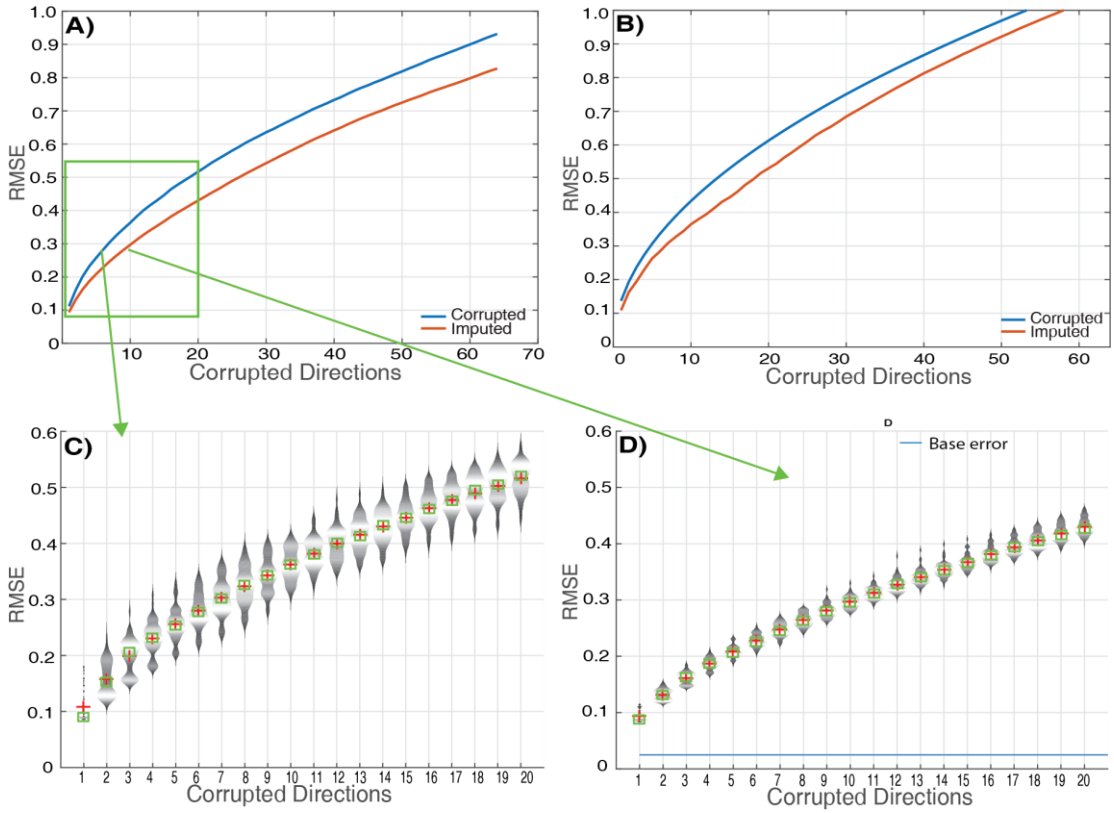

Figure IV.28. A) RMSE of corrupted and imputed signal intensities across the brain when compared with the ground truth. Corrupted gradient volumes were swapped from the motion scan. Each point represents RMSE across 64 measurements. B) RMSE of corrupted and imputed signal intensities when compared with ground truth. Gradient volumes were corrupted with a corruption factor artificially (slash model). Each point represents RMSE across 64 measurements. C) Violin plot of 64 repeated measures of RMSE per bin with random permutations before imputation for the first 20 corrupted gradient volumes. The corrupted volumes were swapped from motion scan data. D) Violin plot of 64 repeated measures of RMSE per bin with random permutations after imputation for the first 20 corrupted gradient volumes. The corrupted volumes were swapped in from motion scan data.

4.4. Discussion

The SHARD algorithm is a promising pre-processing technique for mitigating outliers in HARDI data. The base error that has been observed in the ground truth (Figure IV.29D) with no corrupted volumes can be visually attributed to artifacts that are present in the ground truth "good" data (Figure IV.30). These artifacts can sometimes be difficult to capture with visual observation (given 44 total slices per dataset) but were detected by SHARD. SHARD's utility appears to degrade once approximately 10 corrupted gradient volumes are introduction.

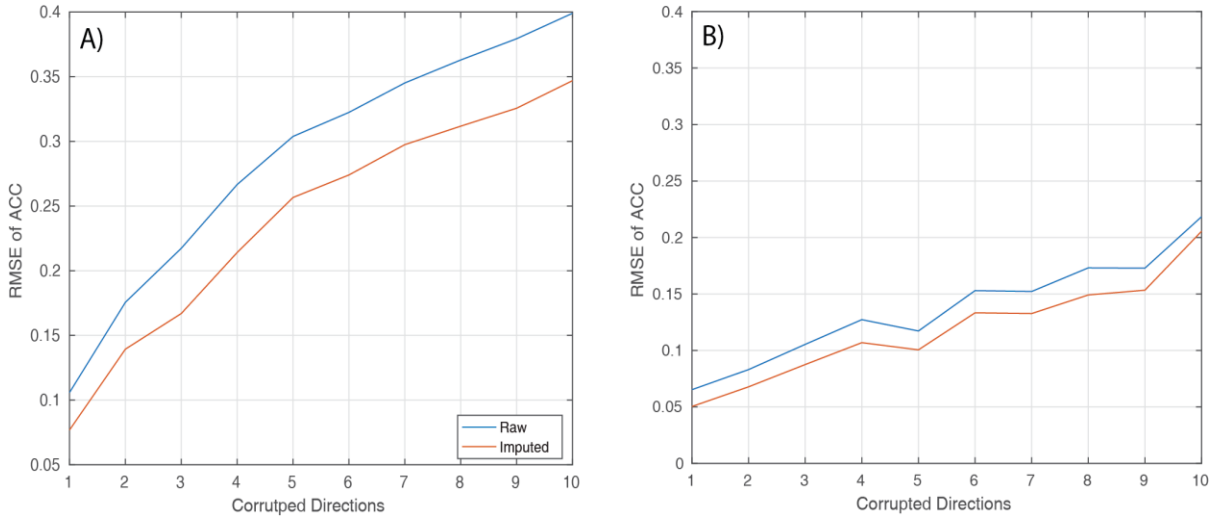

Figure IV.29. RMSE of ACC between spherical harmonics coefficients of Q-ball: A) between corrupted data (Slash model) and ground truth data. B) between corrupted data (motion scan data) and ground truth data.

The usage of spherical interpolation of data has been quite popular in the recent times for [132, 137]. In this scenario while using a weighted version of spherical harmonics only a reduction 0.06 in terms of RMSE could achieved when imputing original signal intensities. There is potential room for improvement with usage of different interpolation techniques. Though there have been outlier detection models which have used spherical harmonics of order 4 [137]. A parameter optimization analysis could

prove to be beneficial purely focused towards spherical harmonic orders. Another possible room for improvement is whether to introduce a weighted mean of the signal intensity along with the spherical harmonic fitted signal intensity which could further reduce the error between the original and the imputed signal.

A median filter approach slice wise has been effective, but it is a two-dimensional solution. A positive improvement could be made using a three-dimensional spatial regularization which could tighten the exact outliers being detected to the specific artifact. This could lead to a crucial increment in accuracy of temporal artifacts that could possibly invade the data. At the same time doing so should also remove single spurious outliers that are detected. Labelling a single voxel as an outlier is a difficult question to answer and simply labelling them as an outlier might only lead to increment in errors between original and imputed signal. Clustering of outliers holds potential in increasing this algorithm's efficiency.

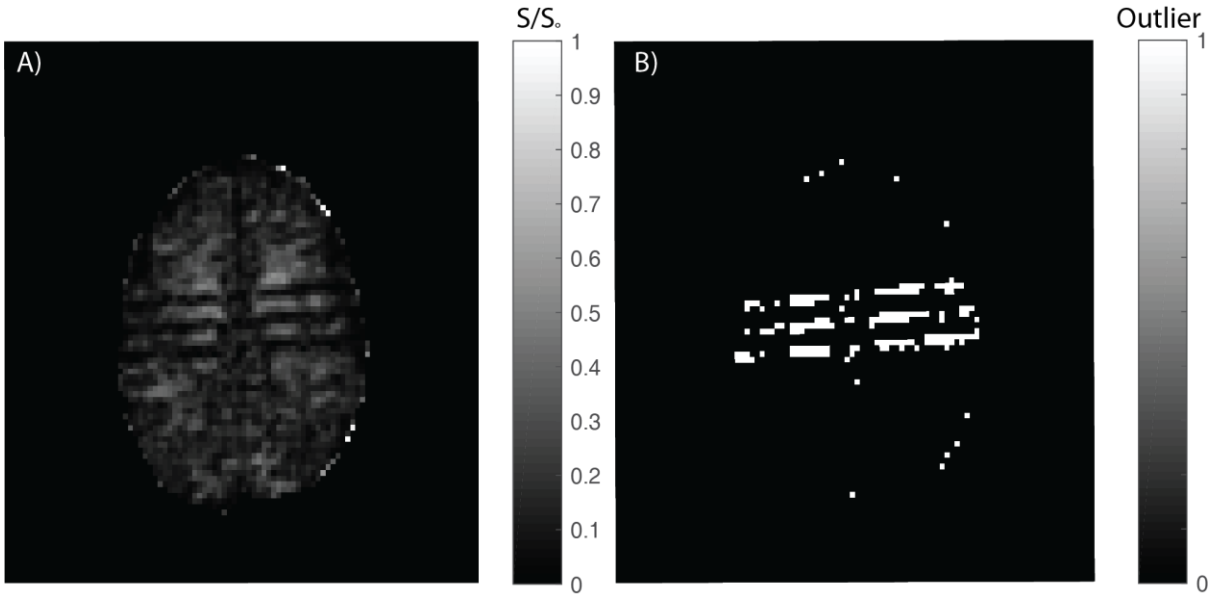

Figure IV.30. A) Middle axial slice containing a striping artifact in the ground truth. B) Outlier mask detected by the SHARD model.

4.5. Conclusion

The SHARD algorithm can detect sufficient outliers in corrupted gradient volumes which contain artifacts or contain signal dropout regions while not impacting the uncorrupted gradient volumes. Instead of exclusion of the detected outliers they can be safely imputed with the weighted spherical harmonic modelled signal intensities without introducing additional errors. Further validation and testing of this algorithm is necessary on a larger data set and using multiple HARDI models that are present. A comparison with the present outlier imputation techniques is clearly warranted. It would be interesting to estimate which techniques can detect and impute outliers without introducing additional errors.

5. Tractography Reproducibility Challenge with Empirical Data (TraCED): The 2017 ISMRM Diffusion Study Group Challenge

This chapter has been adapted from the published work in [41].

5.1. Introduction

Diffusion weighted magnetic resonance imaging (DW-MRI) is a technique which allows for non-invasive mapping of the human brain's micro-architecture at milli-metric resolution. Using voxel-wise fiber orientation reconstruction methods, tractography can provide quantitative and qualitative information for studying structural brain connectivity and continuity of neural pathways of the nervous system in vivo. There have been many algorithms, global, iterative, deterministic and probabilistic, that reconstruct streamlines using fiber reconstruction methods. Tractography was conceived [141] using one of the first fiber reconstruction method, diffusion tensor imaging (DTI) [35]. However, DTI has a well-known limitation: it cannot resolve complex fiber configurations [142]. With the advancement in acquisitions protocols allowing for better resolution and higher number of gradient values new methods for reconstruction of local fiber have been developed. These methods are commonly referred to as high angular resolution diffusion imaging (HARDI), e.g., q-ball, constrained spherical deconvolution (CSD), persistent angular structure (PAS) [12, 13, 90]. HARDI methods enable characterization of more than a single fiber direction per voxel, but have been often shown to be limited when more than two fiber populations exist per voxel [19, 143]. While there is definite gain in sensitivity when using HARDI methods, there remain critical questions of their reproducibility [144].

There have been many validation efforts that aim to assess the anatomical accuracy of tractography. Early studies investigated how well tractography followed large white matter trajectories through qualitative comparisons with dissected human samples [145], or previous primate histological tracings [146]. Later works on the macaque [147] or porcine [148] brains highlighted limitations and common errors in tractography. Recently, the sensitivity and specificity of tractography in detecting connections has been

systematically explored against tracers in the monkey [149-151], porcine [152], or mouse [153] brains. The main conclusions drawn from these are (1) that algorithms always show a tradeoff in sensitivity and specificity (i.e. those that find the most true connections have the most false connections) (2) short-range connections are more reliably detected than long-range, (3) connectivity predictions do better than chance and thus have useful predictive power, and (4) tractography performs better when assessing connectivity between relatively large-scale regions rather than identifying fine details or connectivity.

Despite the wide range of validation studies, there have been few reproducibility studies of tractography [21, 154, 155]. Rather than ask how right (or wrong) tractography is, we ask how stable are the outputs of these techniques? Because tractography is an essential part of track segmentation, network analysis, and microstructural imaging, it is important that reproducibility is high, otherwise power is lost in group analyses or in longitudinal comparisons. In this study, given a standard, clinically realistic, diffusion protocol, we aim to assess how reproducible tractography results are between repeats, between scanners, and between algorithms.

Publicly organized challenges provide unique opportunities for research communities to fairly compare algorithms in an unbiased format, resulting in quantitative measures of the reliability and limitation of competing approaches, as well as potential strategies for improving consistency. In the diffusion MRI community, challenges have focused on recovering intra-voxel fiber geometries using synthetic data [22] and physical phantoms [21, 156]. Similarly, diffusion tractography challenges [154] have provided insights into the effects of different acquisition settings, voxel-wise reconstruction techniques, and tracking parameters on tract validity by comparing results to ground truth physical phantom fiber configurations [21, 155]. Recently, more clinically relevant evaluations have been put forth. For example, a recent MICCAI challenge benchmarked DTI tractography of the pyramidal tract in neurosurgical cases presenting with tumors in the motor cortex [157]. Towards this direction, the current challenge utilized a large-scale single subject reproducibility dataset, acquired in clinically feasible scan times. This challenge was intended to study reproducibility to describe the limitations for capturing physiological and imaging considerations prevalent in human data and evaluate the newest generation of tractography algorithms.

This paper is organized as follows. First, we present the analysis structure of this challenge to characterize which tracts are the most reproducible. Second, we characterize the variance across the tractography methods by design features and compare the potential containment of tracts on a per algorithm basis.

5.2. Methods

5.2.1. DW-MRI Data Acquisition

Neuroimaging The data were acquired with a multi-shell HARDI sequence on single healthy human subject. The two scanners were both Phillips, Achieva, 3T, Best, Netherlands. These are referred to as scanner 'A' and 'B'. The three shells that were acquired: b=1000 s/mm2, 2000s/mm2 and 3000s/mm2 with 20, 48 and 64 gradient directions respectively (uniformly distributed over a hemi-sphere and independently per shell, this was done in consideration of scanner hardware.). The other parameters were kept consistent for all shells. They are as follows: Delta=~48ms, delta=~37ms, partial fourier=0.77, TE = 99 ms, TR ~= 2920 ms and voxel resolution=2.5mm isotropic. A total of 15 non-weighted diffusion volumes 'b0' images interspersed as 5 per shell were acquired. Additionally, for scanner A & B, 5 reverse phase-encoded b0 images and 3 diffusion weighted directions were acquired to aid in distortion correction. The additional 3 diffusion-weighted direction volumes were acquired for ease of acquisition from the scanner. They do not contribute to the pre-processing of the data in any way.

Additionally, a T1-weighted reference image (MPRAGE) was acquired for each session per scanner (4 volumes total). A single volume of T1 was used which was registered to the first session of scanner A where the session had already been registered to the MNI template. This was done using a 6 degree of freedom rigid body registration.

For the initial data release, a technical issue resulted in 5 non-reverse phase-encoded b0 images for scanner A. Note that at the end of the challenge, the scanner 'A' data were completely re-acquired for both sessions with 5 reverse phase-encoded b0 images and 3 diffusion weighted directions. These data were released as supplementary material, but not included in the presented challenge data. Following the protocol

for tractography in [158], we delineated six tracts cingulum (CNG) Left/Right (L/R), inferior longitudinal fasciculus (ILF) (L/R), inferior fronto-occipital (IFO) (L/R). The mean intra-class correlation (ICC) inter-scanner values for the original challenge data and the updated challenge data were 0.86 and 0.89, respectively. The mean difference between methods was 0.15 in terms of ICC. As expected, the inclusion of full reverse phase encoding for Scanner 'A' introduced a small increased in consistency relative to much larger differences between methods.

DW-MRI Data Pre-processing as illustrated in Figure V.31, the 5 repeated acquisitions from each of the four sessions (two repeated on scanner A and B) were concatenated and corrected with FSL's eddy

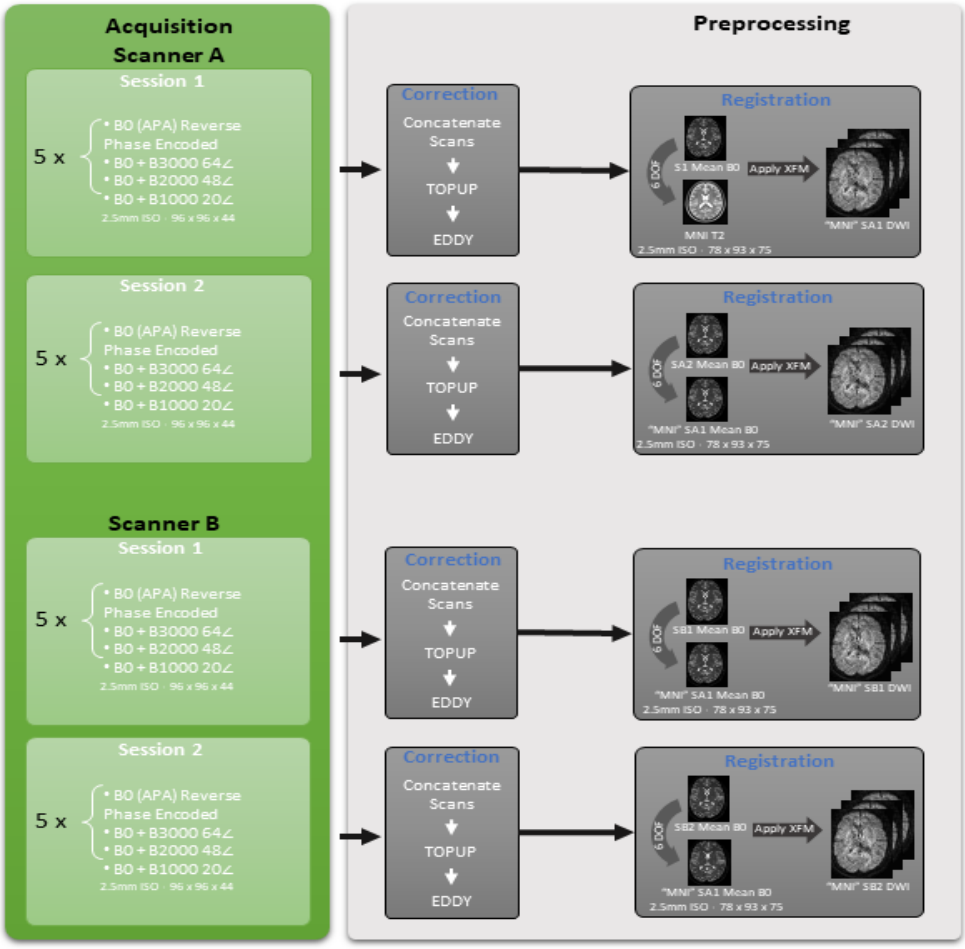

Figure V.31.Non-rigid image registration of GM probability maps of three subjects. Each color box highlights the corresponding region of interest. Right column shows detailed differences in cortical folding patterns across the subjects.

and topup [140, 159, 160]. Intensity normalization was performed by dividing each diffusion weighted scan by the mean of all non-weighted diffusion volume (B0) per session. The average B0 from scanner A of the first sessions was rigidly (six degrees of freedom) registered [117] to a 2.5 mm T2 MNI template (this was done to ensure resampling from registration was done on both datasets). Next, the average B0 from the scanner A second session was rigidly registered to the average B0 of the registered scanner A first session B0 which had already been registered to the MNI space. Successively, the sessions from scanner B were registered to the sessions of scanner A. The b-vectors were rotated to account for the registration of the DW-MRI data [161].

The T1 weighted MPRAGE was rigidly registered to the average registered b0 from the first session of scanner A. This transformation was applied to the T1 maintaining 1 mm isotropic resolution, thus providing a high-resolution segmentation that may be converted into diffusion space by performing a simple down-sampling. Multi-atlas segmentation with non-local spatial STAPLE fusion was used for the segmentation of the T1 volume to 133 different ROI's [123, 162]. Finally, Multi-atlas CRUISE (MaCRUISE) was used to identify cortical surfaces [163]. These were provided for ease of algorithm implementations.

An informed consent under the Institutional Review Board (IRB) was obtained to conduct this study.

5.2.2. Challenge Rules and Metrics

For each of the 20 HARDI datasets (5 repetitions x 2 sessions x 2 scanners), participants were asked to submit a tractogram (i.e., "fiber probability membership function") for each well-modeled fiber structures (uncinate (UNC) L/R, fornix (FNX) L/R, genu of the corpus callosum, cingulum (CNG) (L/R), corticospinal tract (CST) (L/R), splenium of the corpus callosum, inferior longitudinal fasciculus (ILF) (L/R), superior longitudinal fasciculus (SLF) (L/R), and inferior fronto-occipital (IFO) (L/R)(1)). Each tractogram is a NIFTI volume at the field of view and resolution of the T1-weighted reference space where the floating-point value (32-bit single precision) of each voxel is in [57] and indicates the probability of the

voxel belonging to the specified fiber tract. Thus, participants submitted a total of 320 (5 x 2 x 2 x 16) NIFTI volumes using the acquisition of both the scanners. Assessment of fiber fractions was supported (i.e., the sum across all tracts is <=1 with the remainder as background). However, strict probabilities where each voxel may have a high probability of 2 or more fibers with a sum greater than 1 were permitted as well.

Tractograms within a submission were compared based on reproducibility of the tracts (intra-class correlation coefficient (ICC) statistics for continuous values and Dice similarity scores based on maximum probability assignment at 0.5). Intra-session, inter-session, same scanner, and inter-scanner scanner metrics have been reported for quantitative interpretation. The ICC and dice value of unique number of combinations of pairs of repeats were used as data points for violin plots depicting results of intra-session, inter-session and inter-scanner. The unique combinations of repetitions were 40, 50 and 100 respectively for the three levels of reproducibility.

5.2.3. Containment Analysis

A key question is whether the differences in tractography are driven by different considerations of the volume of the track, i.e., the larger the volume is, the more likely the track may include the underlying true track. For example, it is plausible that a set of tractography methods could see the same underlying probabilistic connection pattern and choose to threshold it based on different preferences for the volume of tracks. If the preference was driving the tractography differences, then tractograms would essentially be able to be nested from smallest to largest. To examine this hypothesis, we define the property containment index (CI) for two tracts where

$$\boldsymbol{CI(A,B)} = \begin{cases} \boldsymbol{|A| = 0 : 1} \\ \boldsymbol{|A| \neq 0 \; and \; |B| = 0 : 0} \\ \boldsymbol{otherwise : \; |A \cap B|/|B|} \end{cases} \qquad \textbf{Eq. 22}$$

For the purposes of this discussion, we define the tractogram set to be the binary volume resulting at a 0.05 threshold of the mean of all results submitted for each algorithm. A visual understanding of containment index can be observed in Figure V.32.

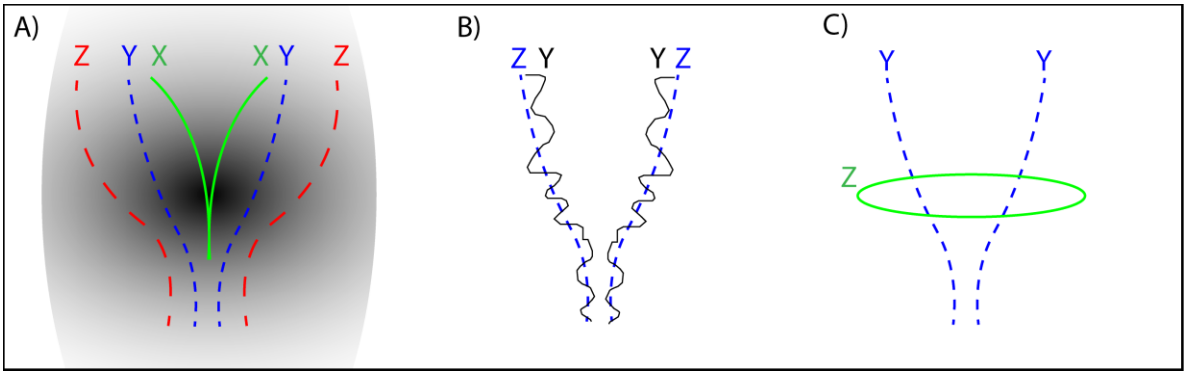

Figure V.32. A) Where the shape X is impeccably contained in Y and Y is contained in Z. The resulting containment CI(Y, X) = 1, CI(Z, X) = 1 and CI(Z, Y) = 1. B) Shape Y is a noisy representation of shape Z where CI(Y, Z) = 0.84. C) Shape Z is different from shape Y in a different orientation and the CI(Z,Y) = 0.17

Then, an optimal ordering ("nesting") of tractogram entries can be computed by maximizing the containment energy (CE, i.e., sum of CI for all tracts versus the tracts earlier than the one under consideration):

$$\underset{o\in perm(1\ldots|Entry|)}{\mathbf{argmax}}\ CE = \underset{o\in perm(1\ldots|Entry|)}{\mathbf{argmax}} \sum_{i}^{N}\sum_{j\le i}^{N} CI\left(Entry\{o_i\}, Entry\{i_j\}\right) \quad \textbf{Eq. 23}$$

Where perm denotes the permutation operator and Entry is a list of all entered tractograms. Conceptually, this procedure finds the ideal order to stack the tractograms inside each other where the first tract is "most inside" the subsequent ones and the last tract is "most outside" all others. We define <CI> as the average containment index of all nesting for the ordered entries that are smaller than or equal to an entry provides a quantitative way to examine "nesting" (note, this approach includes the self-containment index so that the first entry has a CI of 1). Then, we can see how the nesting holds up from the inner (#'1') to the outer (#'46') entry.

5.3. Results

Table V.5 presents a more detailed technical contribution of each of the works:

- Team 1, Team 5, Team 6, Team 8 and Team 9 used all three shells of b-values provided in the dataset. Team 2 used all shells with data from an additional 30 subjects from the Human Connectome

Project. Team 3 used shells of b-values 1000 and 2000 s/mm2. Team 7 and Team 4 only used the shell of b-value 3000 s/mm¬2.

- Additional pre-processing has been used by four teams. Team 4: Data was upsampled to 1mm isotropic resolution. Team 6 used image de-noising techniques and upsampled the data to 1.25mm. Team 5 and Team 9 used different styles of segmentation of the data presented for analysis.

- In terms of the fiber detection model, Team 6 and Team 3 used variants of tensor models while the others have used different variants of constrained spherical deconvolution. Notably Team 8 used a compartment analysis model using spherical harmonics.

- Considering the tractography parameters - the range of step sizes that have been used lie between 0.2-1.25mm. Threshold angle lies in the range of 20-40 degrees.

- Single fiber assumptions were considered with the condition of FA > 0.7 by teams Team 1 and Team 4. A notable observation here is that a general assumption was made by Team 6 to reject voxels which were less than 0.15 FA.

- Team 2, Team 6 and Team 8 post-processed the tractography results for removal of spurious fibers by defining different and specific constraints.

- Of note, Team 2 treated the tractography problem as a segmentation problem and developed a U-net which was trained on the HCP data. While Team 9 used a multi-atlas approach to tractography. The other teams used the general approach of probabilistic or deterministic tractography.

An overlay of all 46 submissions, for all estimated fiber pathways can be observed (Figure V.33 Column 1 & 3). Only the left side has been shown as the right side is a similar observation. There are vast differences that can be noticed in the estimated pathways. The volume of the brain occupied by each tract from different submissions varied dramatically. When all 46 submissions are overlaid, tracts occupy 14-53% of the brain volumetrically (average – 34%). Specifically, the union of all entries for FNX (L/R), CNG (L/R), IFO (L/R) and SLF (L/R) cover (30.7, 25.8), (40.9, 37.2), (42.4, 46.1), (50.6, 53.3) respectively, while CST (L/R), ILF (L/R), UNC (L/R) and Fminor and Fmajor cover (23.6, 25.4), (33.4, 33.6), (14.3, 17.4), 44.3 and 34.1. Note that individual submissions appear qualitatively reasonable (Figure V.33 Column

2 & 4).

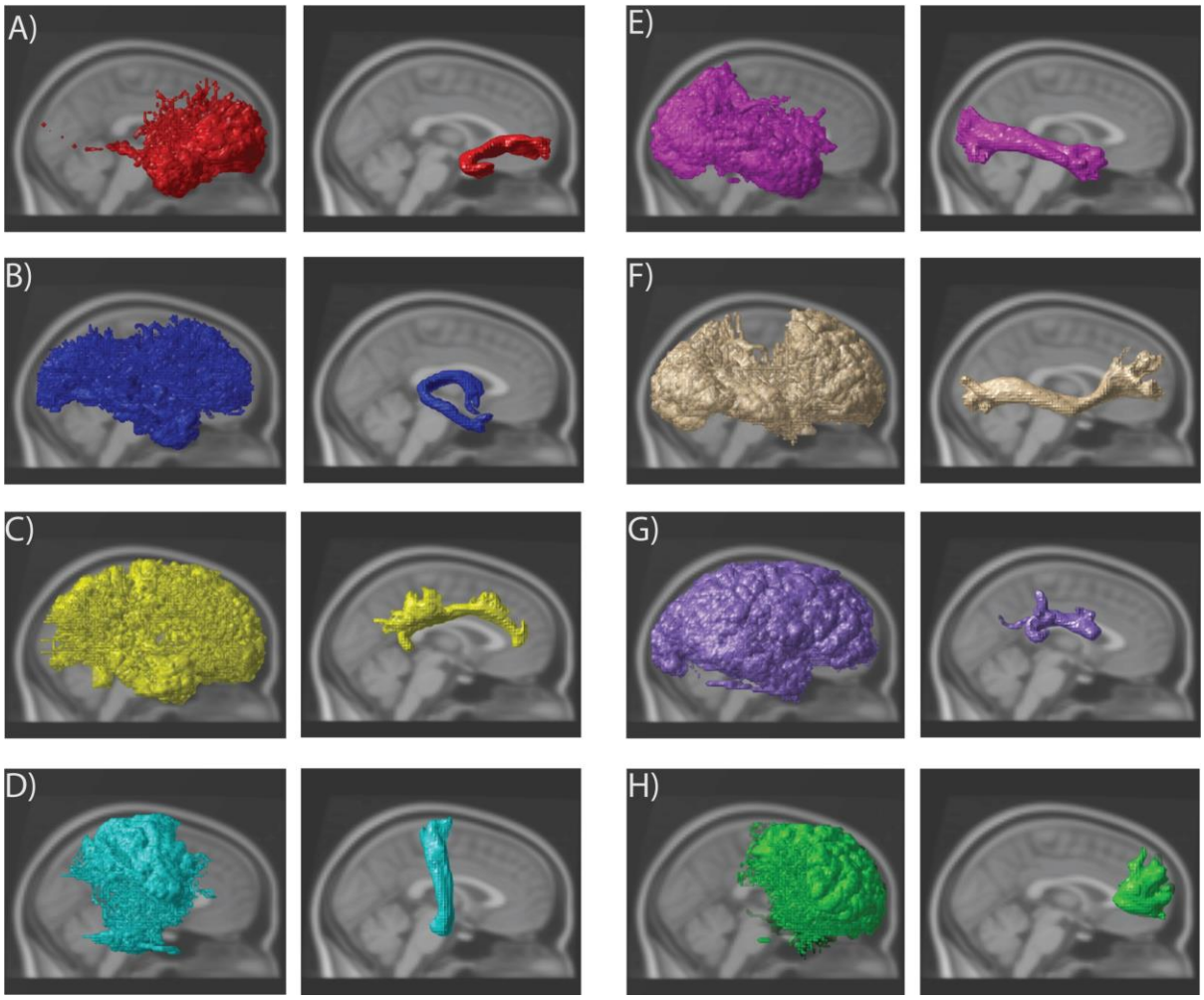

Figure V.33. Left: An overlay of all the 46 submissions from all sessions that were acquired using both scanners per tract Right: An overlay of a single submission using all sessions that were acquired using both scanners per tract A) Uncinate left B) Fornix left C) Cingulum eft D) Corticospinal tract left E) Inferior Longitudinal Fasciculus left F) Inferior Fronto-Occipital Fasciculus left G) Superior Longitudinal Fasciculus left H) Fminor.

The number of algorithmic submission's team wise are Team 1: 14, Team 2: 1, Team 3: 2, Team 4: 12, Team 5: 1, Team 6: 6, Team 7: 1, Team 8: 6 and Team 9: 3. It can be observed that the ICC range for the set of algorithms on a per team basis does not show a lot of variance. The ICC range of algorithms per team are Team 1 (0.61 – 0.77), Team 4 (0.52 – 0.58), Team 6 (0.77 – 0.85), Team 8 (0.81 – 0.89), Team

9 (0.27 – 0.69), Team 3 (0.64, 0.73), Team 2 (0.85), Team 7 (0.88) and Team 5 (0.97). The teams that submitted more than 3 algorithms show an average difference of 0.1 in terms of ICC.

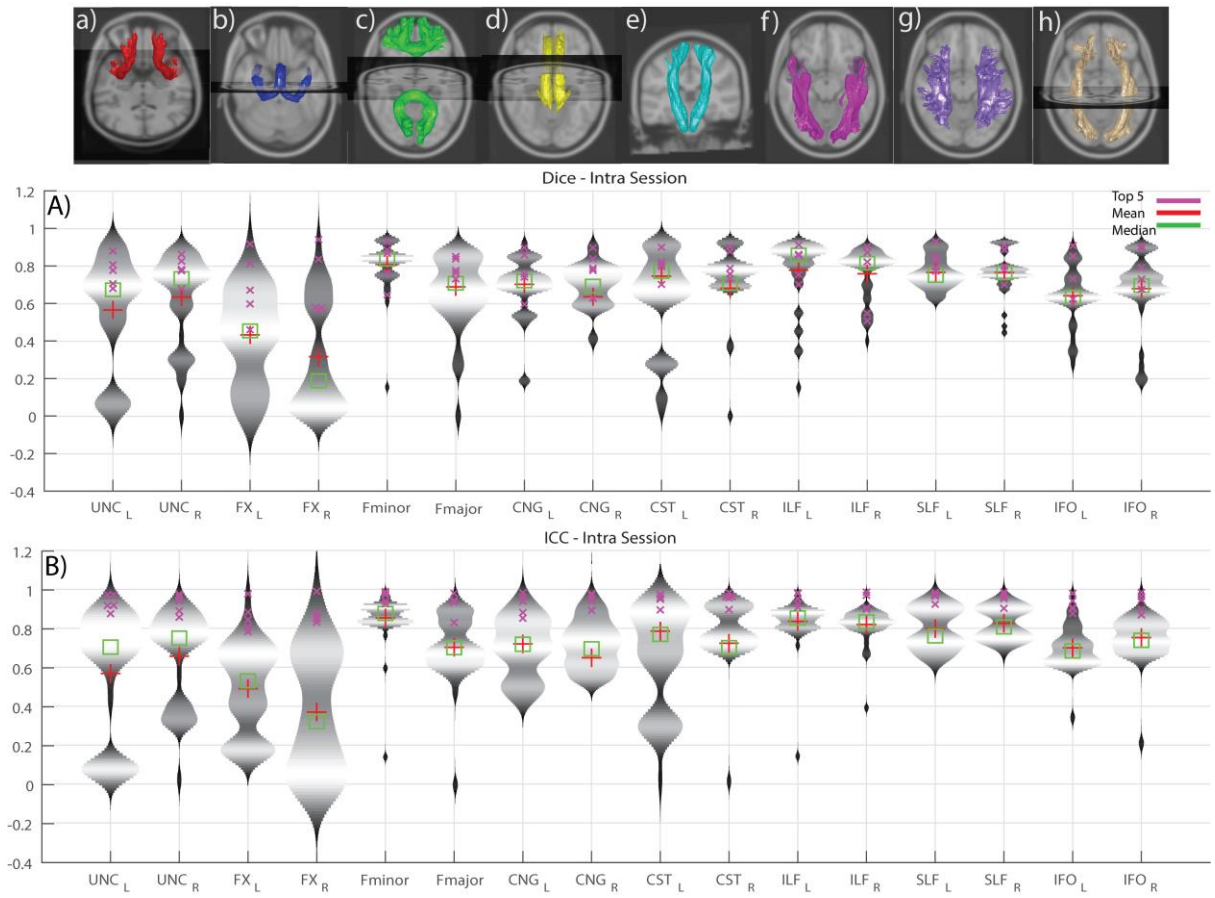

Figure V.34. Violin plots of intra-session submissions across both the scanners per tract. A) Dice similarity coefficients B) Intra-class correlation coefficients. The top row depicts the median of the top five intra session submissions. The tracts are in the following order (L/R): a) Uncinate b) Fornix c) Fminor & Fmajor d) Cingulum e) Corticospinal tract f) Inferior longitudinal fasciculus g) Superior longitudinal fasciculus h) Inferior fronto-occipital tract

Violin plots (depict the probability density of the data) of ICC and Dice for intra-session reproducibility, inter-session, and inter-scanner measures of reproducibility are presented in Figure V.34, Figure V.35 and Figure V.36 respectively. Since the observations are highly similar in the afore-mentioned

figures we only present a detailed comment on Figure V.34 which holds true for Figure V.35 and Figure V.36 as well. This figure helps in identifying the low, moderate and high reproducibility tracts. The intra-session distributions (Figure 4B) across entries for UNC (L/R) and FNX (L/R) are bi-modal with a median

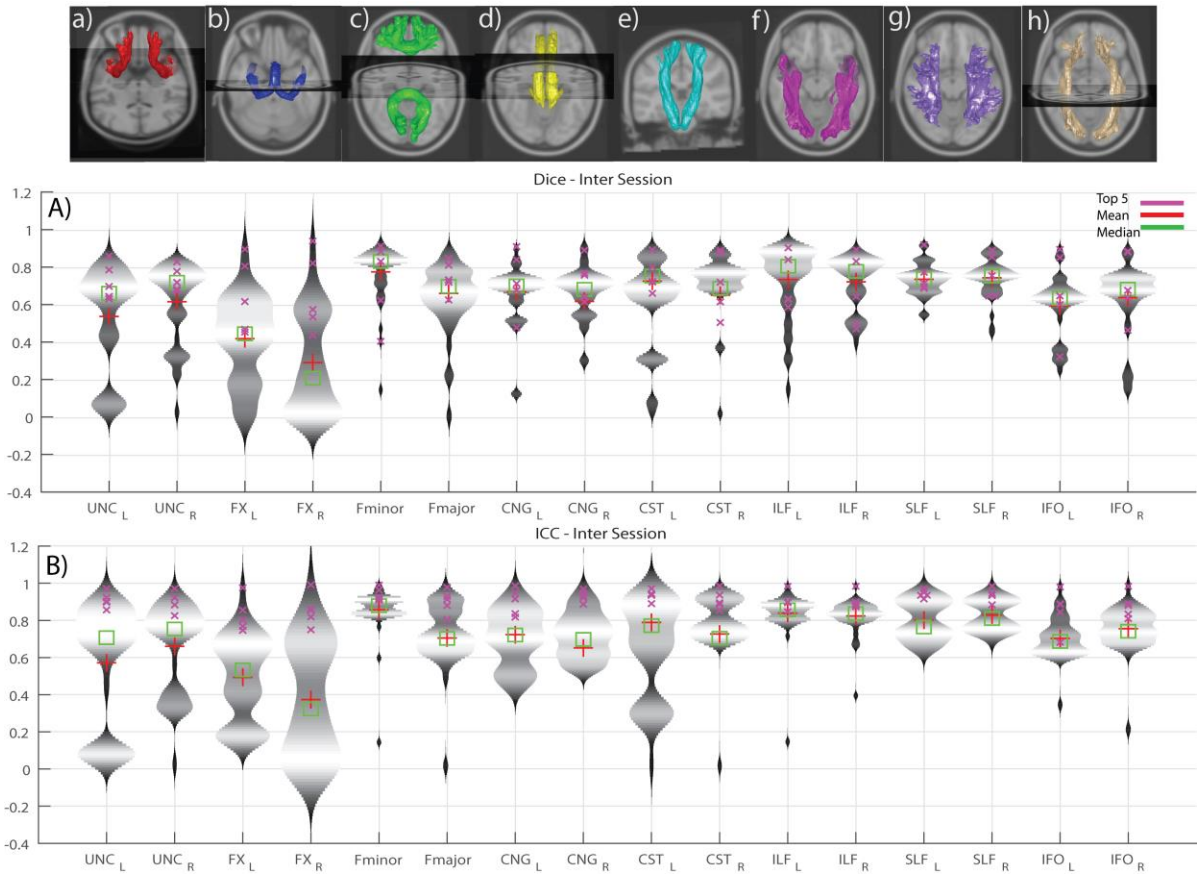

Figure V.35.Violin plots of inter-session submissions across both the scanners per tract. A) Dice similarity coefficients B) Intra-class correlation coefficients. The top row depicts the median of the top five inter session submissions. The tracts are in the following order (L/R): a) Uncinate b) Fornix c) Fminor & Fmajor d) Cingulum e) Corticospinal tract f) Inferior longitudinal fasciculus g) Superior longitudinal fasciculus h) Inferior fronto-occipital tract

of the lower mode less than 0.4 ICC. The CST (L) has a smaller fraction of the entries with ICC less than 0.4, while the remainder of the entries have only a few outlier entries less than 0.4 The inter-session (Figure V.35) and inter-scanner (Figure V.36) distributions were similar, with a slight increase in outlier entries for IFO (L/R). The patterns in the dice were similar when using a quality threshold of less than 0.4 dice.

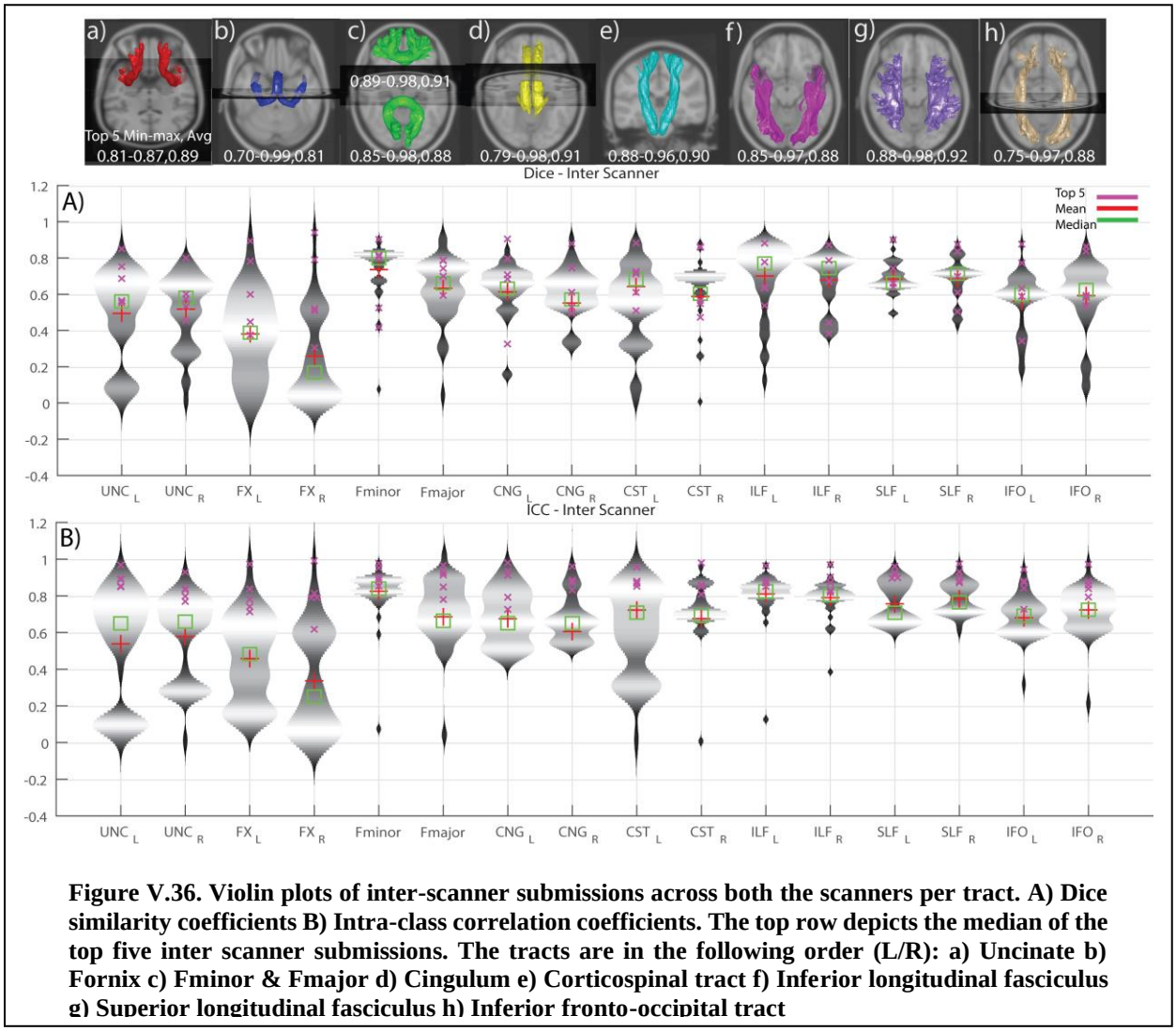

Figure V.36. Violin plots of inter-scanner submissions across both the scanners per tract. A) Dice similarity coefficients B) Intra-class correlation coefficients. The top row depicts the median of the top five inter scanner submissions. The tracts are in the following order (L/R): a) Uncinate b) Fornix c) Fminor & Fmajor d) Cingulum e) Corticospinal tract f) Inferior longitudinal fasciculus g) Superior longitudinal fasciculus h) Inferior fronto-occipital tract

We define cutoffs for high, moderate, and low reproducibility on the inter-scanner reproducibility. High reproducibility was defined as a median ICC greater than 0.6 and less than 5% of entries less than 0.4 ICC. Moderate reproducibility was defined as median ICC greater than 0.4 and less than 25% of entries less than 0.4 ICC. Low reproducibility was defined as a median ICC less than 0.4 or more than 25% of entries less than 0.4 ICC. Hence, the high reproducibility tracts were Fminor, CST (/R), ILF (L/R), SLF (L/R) and IFO (L/R). The moderate reproducibility tracts were CST (L), Fmajor, CNG (L/R). The low reproducibility tracts were UNC (L/R) and FNX (L/R). This above is observed when looking at all submissions however when observing the top 5 submissions we see higher reproducibility.

When the analysis is restricted to only the top five submissions, we see a different picture that suggests substantively reproducible methods. The inter-scanner reproducibility among the top 5 entries in ICC (min-max, average) are shown in Figure V.36.

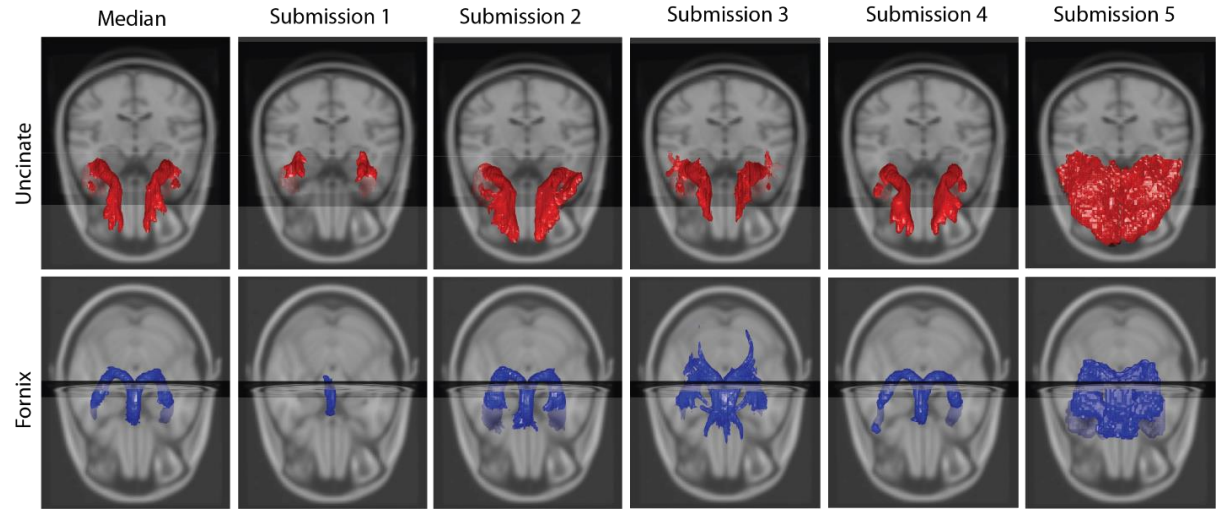

Figure V.37. First row shows the median of Uncinate (L/R) and the top five submissions. The second row shows the median and submissions of Fornix (L/R).

Figure V.37 illustrates the top five entries for the tracts with the lowest inter-scanner reproducibility alongside the volumetric median (median per voxel from five submissions) of the top five entries. Qualitatively, the volumetric profiles of the UNC (L/R) and FNX (L/R) are very different across the top five entries. The first submission has small "core" tracts labeled, while the second, third and fifth found much larger spatial extents and the fourth was mid-way between.

5.4. Discussion

The most reproducible tracts were Fminor, CST (\R), ILF (L\R), SLF (L\R), IFO (L\R), while the moderately reproducible tracts were Fmajor, CNG (L\R) and CST (\L). Lowest reproducibility tracts are UNC (L\R), FNX (L\R). These tracts have a well-spread/broad probability distribution. Note that the reproducibility of these tracts was maintained across imaging sessions and change of scanner. It is evident that all the algorithms entered are not consistently identifying the same fiber structures given the extreme

variance observed in Figure V.33. While most of the individual submissions show a reasonable detection of the tracts if observed from a ROI point of view (Figure V.33), the difference between tract volumes between methods is quite high.

The reproducibility (ICC) of the entered algorithms varied from 0.27 to 0.97 (Figure V.38 A), but most of the algorithms performed with a reproducibility of 0.6 or higher. Similar levels of reproducibility were observed for methods that used selective shells or additional data from the Human Connectome Project. Note it would be inappropriate to assume independence and there are a few methods per categorical assignment, so statistical analysis across method types was not performed. Qualitatively, CSD was the most popular approach as the pre-processing fiber reconstruction method (Figure V.38 B). Tensor and compartment models perform well but trailed slightly behind CSD when comparing maximum values that have been achieved using these methods. The modified version of CSD with the addition of Deep Learning U-net also performed well.

The choices of analysis parameters appear to have affected method performance. A comparison of different step sizes that have been used shows that the most heavily used category was 0.2mm (Figure V.38C). However, methods using all other step size choices (e.g., 0.005, 1 and 1.25mm) performed better in terms of ICC. A variety of threshold angles have been used lying in the range of 20 – 60 degrees (Figure V.38 D). The variation is hard to comment upon as this suggests that a threshold angle is specific to the type of tractography algorithm. High reproducibility has been achieved at lower threshold angles such as 20 degrees and at higher angles as well such as 45 or 60 degrees. Additional pre-processing before implementing fiber reconstruction methods shows improvement for ICC only when additional segmentation was performed (Figure V.38 E). A comparison of de-noising coupled with up-sampling and no additional pre-processing shows higher reproducibility when no additional steps are performed. While most of the algorithms did not use additional post-processing steps (Figure V.38 F), the few algorithms that used the methods of outlier rejection, spurious fiber removal and SIFT2 show improvement in reproducibility. In brief, it might be inferred that additional pre-processing and post-processing techniques are helpful in increasing the reproducibility of tractography algorithms, though a systematic test of this

would be necessary to draw accurate conclusions.

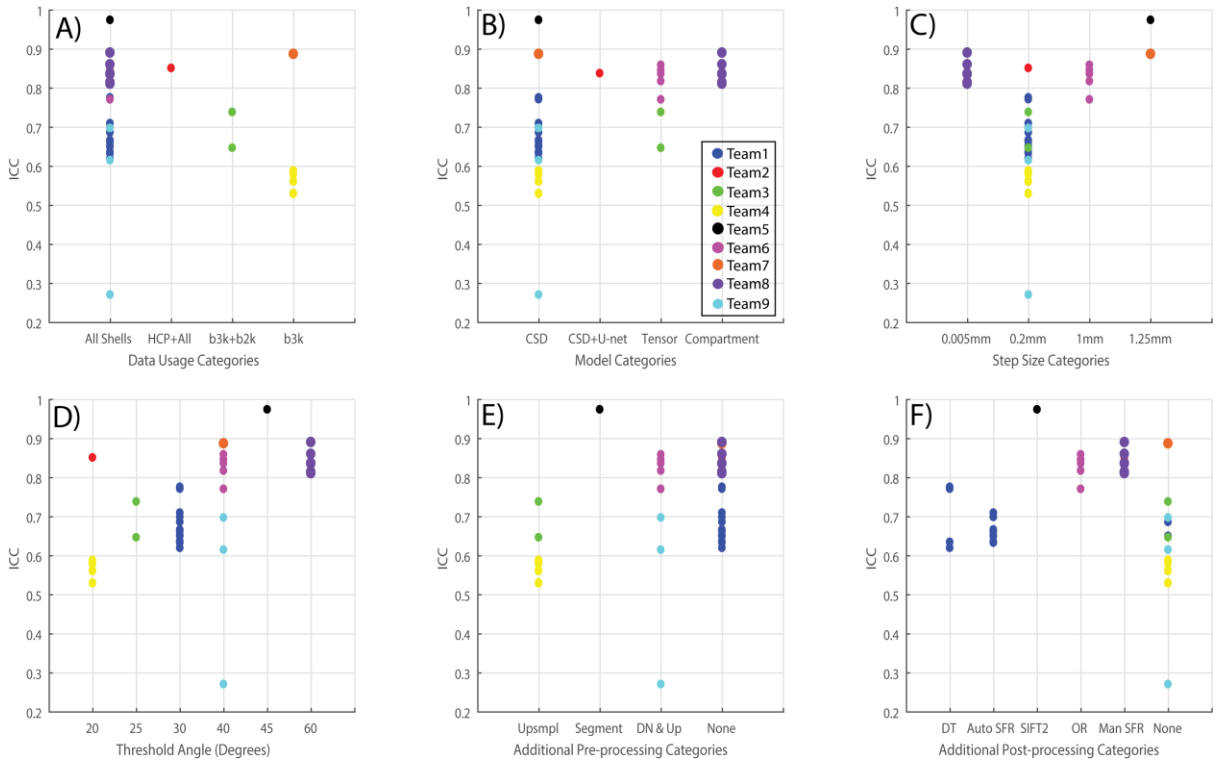

Figure V.38. A) Quantifies the number of algorithms that used a specific part of the dataset or added more from other sources. B) Quantifies the usage of HARDI/Tensor methods by different tractography algorithms as a pre-step. C & D) Quantifies the step size and threshold angle parameter for tractography algorithms. E & F) Quantify the number of additional pre-processing and post-processing techniques applied for the tractography algorithms.

While it would be expected that an algorithm with empty or inaccurate bundles could achieve an extremely high ICC which would be representative of 'null' learning. Hence, we conducted consistency analysis using the containment index as to which bundles are contained inside which ones. The inaccurate ones will lie on the outside or show up as outliers which can be observed in Figure V.39.

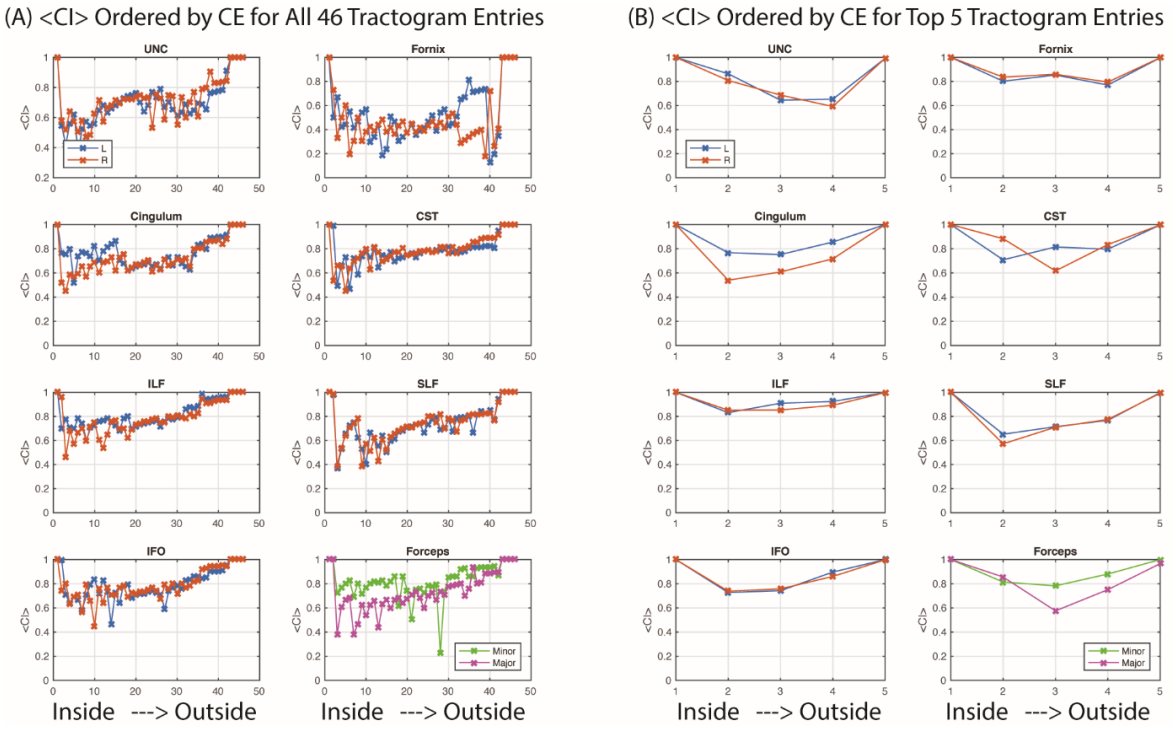

Figure V.39. Ordering entries to minimize containment energy (CE) shows that containment index is generally lower for the volumetrically smaller tractograms (toward "inside" on each subplot) and increases for the larger tractograms (toward "outside" on each subplot). Variations in containment explained the least amount of entry variability for the UNC and Fornix, while the other tracts were more consistent. The containment between all methods (A) were more variable and lower than the containment for the top five methods (B).

As seen in Figure V.39, <CI> is moderate and variable (~0.4-0.6) for the first approximately 20 entries (after ordering) and then steadily increases for the CST, Cingulum, Forceps, ILF, IFO, and SLF. Hence, for smaller tractograms, approximately 50% of the variance is explained by nesting, but there are substantial contributions from other factors. For the larger tractograms (~20-46 ordered entries), the differences appear largely driven by increasing volume of the tracts. UNC and Fornix are a bit more variable between ordered methods, which indicates associations within methods and suggests disagreements across major categories of entries. Finally, the Fornix is highly variables across methods (~<0.4 <CI>), which point towards inconsistency of tract definition between approaches. When looking across all pairs of tracts, the overall rank correlation of the method ordering was low (mean=0.25) with a high variance (standard deviation=0.27, range=-0.28 to 1.0). Therefore, the relative volumetric differences between tracts were not

consistent for methods across white matter tracts. Examining nestings of the top five tracts showed that Submission 5 (not shown) was always the largest, while Submission 1 and Submission 4 were determined to be the most inner methods half of the time. Submission 2 was the second largest for 12/16 tracts, while Submission 3 was the second largest for the others. This is consistent with a visualization interpretation of Figure V.37. The <CI> was ~1 for the fifth method, so a highly reproducible tract was feasible that encompassed the choices of the other top entries. The top 5 entries had high <CI> (>0.7) for the Fornix, IFO, ILF, but the remaining tracts were showed low CI for at least one method. Therefore, while at least one of the top methods differed from the others in a substantial manner, this could not be explained by volumetric differences of the tracts.

5.5. Conclusion

The most reproducible tracts considering all submitted algorithm outcomes are Fminor, CST (\R), ILF (L\R), SLF (L\R), IFO (L\R). The moderately reproducible ones are Fmajor, CNG (L\R) and CST (\L). Tracts with low reproducibility are UNC (L\R) and FNX (L\R). The most reproducible algorithms are 5A, 8D, 7A, 6E and 6F (Table V.5) as per criteria of ICC. The mentioned algorithms are not an example of a consistent null learning as they all lie with in a nested containment with the largest covered volume.

In conclusion, the 2017 ISMRM TraCED Challenge created a publicly available multi-scanner, multi-scan in vivo reproducibility dataset and engaged nine groups with 46 algorithm entries. The TraCED Challenge dataset is freely available at www.synapse.org. Consistent with previous studies, reproducibility of tractograms was found to vary by anatomical tract. When viewed across all entries, reproducibility was concerning (ICC <0.5); however, the cluster of top performing methods resulting in reassuringly high results (ICC > 0.85). Variation in performance were seen across processing parameters, but the challenge design did not provide sufficient number of samples to identify uniformly preferred design choices. The key novel finding of this challenge is that variations in tractography methods can be largely attributed to larger/smaller volumetric difference tradeoffs for the larger tracts, especially among methods that are tuned towards volumetrically larger tractograms. Yet, the different methods clearly result in fundamentally

different tract structures at the more conservative specificity choices (i.e., volumetrically smaller tractograms). The containment index, containment energy, and containment index framework provides a consistent approach to evaluate the nesting structure tractograms, and the freely available data and results from this challenge can be used to quantify new tractography approaches.

Table V.5.The table presents all the hyper-parameters of the different algorithms that were submitted and an overall evaluation of the algorithm in terms of ICC and Dice.

Synapse Submission id	Algorithm ID	ICC	DICE	b-value shells	HARDI/Tensor Model	Step size	Threshold angle	Additional Pre-Processing	Post-Processing
syn8533598	1A	0.7753	0.6364	All shells	CSD	0.2mm	30 degrees	NA	Distance transform of bundle volumes
syn8643780	1B	0.6857	0.6596	All shells	CSD	0.2mm	30 degrees	NA	NA
syn8643793	1C	0.6343	0.6346	All shells	CSD	0.2mm	30 degrees	NA	Distance transform of bundle volumes
syn8648608	1D	0.7707	0.5402	All shells	CSD	0.2mm	30 degrees	NA	Distance transform of bundle volumes
syn8649314	1E	0.6498	0.6508	All shells	CSD	0.2mm	30 degrees	NA	NA
syn8649322	1F	0.6192	0.6197	All shells	CSD	0.2mm	30 degrees	NA	Distance transform of bundle volumes
syn8649611	1G	0.6324	0.6332	All shells	CSD	0.2mm	30 degrees	NA	Automatic spurious fiber removal
syn8649618	1H	0.6494	0.6503	All shells	CSD	0.2mm	30 degrees	NA	Automatic spurious fiber removal
syn8649622	1I	0.6517	0.6526	All shells	CSD	0.2mm	30 degrees	NA	Automatic spurious fiber removal
syn8649650	1J	0.6662	0.6671	All shells	CSD	0.2mm	30 degrees	NA	Automatic spurious fiber removal
syn8649652	1K	0.6616	0.6624	All shells	CSD	0.2mm	30 degrees	NA	Automatic spurious fiber removal
syn8649654	1L	0.6362	0.637	All shells	CSD	0.2mm	30 degrees	NA	Automatic spurious fiber removal
syn8649656	1M	0.7093	0.7103	All shells	CSD	0.2mm	30 degrees	NA	Automatic spurious fiber removal
syn8649658	1N	0.6984	0.6994	All shells	CSD	0.2mm	30 degrees	NA	Automatic spurious fiber removal

syn8555229	2A	0.8506	0.7918	All shells + 30 HCP subjects	CSD + U-net	0.2mm	20 degrees	NA	Spurious Fiber Removal
syn8656474	3A	0.7379	0.7253	b1000 and b2000	Tensor Variant	0.2mm	25 degrees	Data Upsampling	NA
syn8656475	3B	0.6463	0.6341	b1000 and b2000	Tensor Variant	0.2mm	25 degrees	Data Upsampling	NA
syn8662707	4A	0.5285	0.5317	b3000	CSD	0.2mm	20 degrees	Data Upsampling	NA
syn8662708	4B	0.5822	0.3207	b3000	CSD	0.2mm	20 degrees	Data Upsampling	NA
syn8662709	4C	0.5881	NaN	b3000	CSD	0.2mm	20 degrees	Data Upsampling	NA
syn8662710	4D	0.5285	0.5317	b3000	CSD	0.2mm	20 degrees	Data Upsampling	NA
syn8662711	4E	0.5781	0.3182	b3000	CSD	0.2mm	20 degrees	Data Upsampling	NA
syn8662712	4F	0.5835	NaN	b3000	CSD	0.2mm	20 degrees	Data Upsampling	NA
syn8662713	4G	0.5285	0.5317	b3000	CSD	0.2mm	20 degrees	Data Upsampling	NA
syn8662714	4H	0.5291	0.4932	b3000	CSD	0.2mm	20 degrees	Data Upsampling	NA
syn8662715	4I	0.5302	NaN	b3000	CSD	0.2mm	20 degrees	Data Upsampling	NA
syn8662716	4J	0.5285	0.5317	b3000	CSD	0.2mm	20 degrees	Data Upsampling	NA
syn8662717	4K	0.5596	0.5323	b3000	CSD	0.2mm	20 degrees	Data Upsampling	NA
syn8662718	4L	0.5616	NaN	b3000	CSD	0.2mm	20 degrees	Data Upsampling	NA
syn8664905	5A	0.9738	0.8231	All shells	CSD	1.25mm	45 degrees	Additional Segmentation	SIFT2
syn8666133	6A	0.7702	0.7708	All shells	Tensor Variant	1mm	40 degrees	Denoising, Upsampling	Outlier Rejection
syn8666134	6B	0.8358	0.5742	All shells	Tensor Variant	1mm	40 degrees	Denoising, Upsampling	Outlier Rejection
syn8666135	6C	0.8171	0.7595	All shells	Tensor Variant	1mm	40 degrees	Denoising, Upsampling	Outlier Rejection
syn8666136	6D	0.817	0.7704	All shells	Tensor Variant	1mm	40 degrees	Denoising, Upsampling	Outlier Rejection
syn8666137	6E	0.8586	0.571	All shells	Tensor Variant	1mm	40 degrees	Denoising, Upsampling	Outlier Rejection

syn8666138	6F	0.8458	0.7646	All shells	Tensor Variant	1mm	40 degrees	Denoising, Upsampling	Outlier Rejection
syn8667007	7A	0.8868	0.6187	b3000	CSD	1.25mm	40 degrees	NA	NA
syn8666587	8A	0.86	0.8672	All shells	Compartment Model	0.005mm	60 degrees	NA	Spurious Fiber Removal
syn8666598	8B	0.8367	0.5166	All shells	Compartment Model	0.005mm	60 degrees	NA	Spurious Fiber Removal
syn8666602	8C	0.8349	0.5287	All shells	Compartment Model	0.005mm	60 degrees	NA	Spurious Fiber Removal
syn8666936	8D	0.8901	0.6409	All shells	Compartment Model	0.005mm	60 degrees	NA	Spurious Fiber Removal
syn8667021	8E	0.8145	0.4983	All shells	Compartment Model	0.005mm	60 degrees	NA	Spurious Fiber Removal
syn8667022	8F	0.8103	0.4773	All shells	Compartment Model	0.005mm	60 degrees	NA	Spurious Fiber Removal
syn8698866	9A	0.6145	0.6015	All shells	CSD	0.2mm	40 degrees	Addtional Segmentation	NA
syn8698867	9B	0.6968	0.6804	All shells	CSD	0.2mm	40 degrees	Addtional Segmentation	NA
syn8698868	9C	0.2703	0.2572	All shells	CSD	0.2mm	40 degrees	Addtional Segmentation	NA

6. Deep Learning Captures More Accurate Diffusion Fiber Orientation Distributions than Constrained Spherical Deconvolution

This chapter has been adapted from the published work in [164].

6.1. Introduction

Understanding the relationship between observed diffusion weighted MRI signals and true tissue microarchitecture is of fundamental concern for biophysical modeling, detecting microstructural differences, and brain tractography. Substantial efforts have been invested in interpreting the diffusion signal from both model-based (e.g., constrained spherical deconvolution - CSD [15, 165], Q-ball [13], persistent angular structure - PAS [12]) and data-driven [89] perspectives. Recently, multi-layer neural networks (or informally, deep learning or deep neural networks - DNN) have emerged as a leading class of machine learning approaches. Moreover, advances combining MRI and whole brain histology have enabled volumetric registration between MRI and histological processes, while co-registered confocal microscopy allows direct 3D observation of intra-voxel tissue orientation. Here, we apply deep learning to investigate the potential information content in single shell diffusion weighted MRI to explain histologically observed fiber orientation distribution (FOD) functions.

6.2. Data Acquisitions & Methods

6.2.1. Data Acquisitions

Three ex-vivo squirrel monkey brains were imaged on a Varian 9.4T scanner. Briefly, data were acquired with a 3D diffusion-weighted EPI sequence (b-value=6,000 s/mm2, 100 directions) at 300um isotropic resolution. After scanning, the tissue was sectioned, stained with the fluorescent DiI, and imaged on an LSM710 Confocal microscope following the procedures outlined in [23]. The histological FOD was extracted using structure tensor analysis. Finally, a multi-step registration procedure [23] was used to determine the corresponding diffusion MRI signal. A total of 567 histological voxels were processed, and a hundred random rotations were applied to each one of them for both the MR signal and the histology FOD to augment the data bringing the total to 57,267 voxels [166].

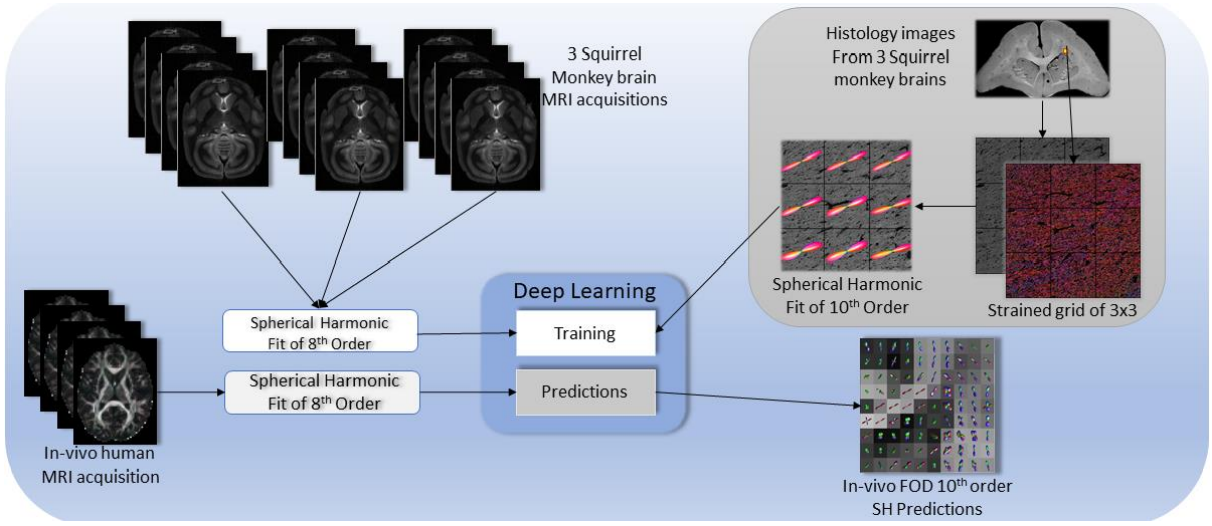

Figure VI.40. Confocal histological data provides a ground truth basis for fiber orientation distributions. The truth data was split into a training set and a testing set. Once trained, the deep learning approach was applied to both the testing set and a separate human dataset.

For qualitative validation, a single healthy human volunteer was scanned for a single session using a 3T (Achieva, Philips Medical Systems, Best, The Netherlands) with a 32-channel head coil. Four scans acquired were at a b-value of 2000 s/mm2 (which approximates the diffusion contrast of a fixed ex vivo scan at a b-value of 6000 s/mm2) with 96 gradient directions and an additional b0 per scan (2.5mm isotropic

resolution, matrix of 96x96, 38 slices, Multi-Band=2; SENSE=2.2;TR= 2650 ms; TE=94 ms; partial Fourier=0.7). Standard pre-processing with FSL (topup, eddy correction, registration, averaging across scans) was performed before analysis.

6.2.2. Methods

Both ex-vivo and in-vivo HARDI acquisitions were fit with 8th order real spherical harmonics. Outliers were manually reviewed for imaging artifacts, and 54 voxels were removed. FOD's from the histology were fitted with a 10th order real spherical harmonics. Histology data was divided into training/validation (44,541 voxels) and testing sets (7,272 voxels) without mixing augmented data (rotations). For training/validation, a 20% percent split was used with 5 folds. Mean squared error was used to assess model accuracy [167].

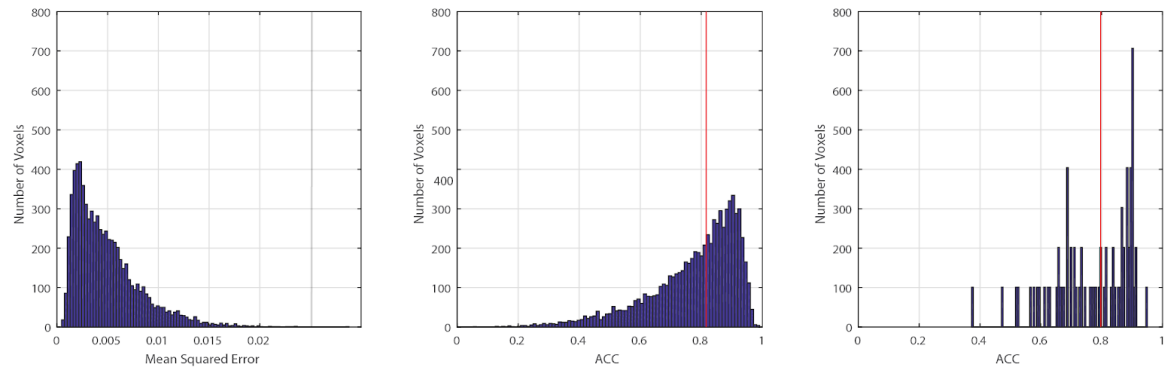

Figure VI.41. A) Histogram of MSE across all voxels between histology and DNN predicted FOD's. B) Histogram of ACC across all voxels from the test set of histology and DNN predicted FOD's. Media ACC is 0.817 C) Histogram of ACC across all voxels from the test set of histology and CSD predicted FOD's. Median ACC is 0.797

6.3. Results

The median angular correlation coefficient (ACC) for CSD (0.7965) was significantly ($p<0.05$, non-parametric signed rank test) lower than for the deep approach (0.8165) (Figure VI.41), which corresponded to a lower root mean squared error for the deep approach (0.539 versus 0.561). Qualitatively, the predicted FOD's on the human in vivo data demonstrate anatomical consistency (Figure VI.42), indicating that the deep learning approach is predicting structure in line with prior observations.

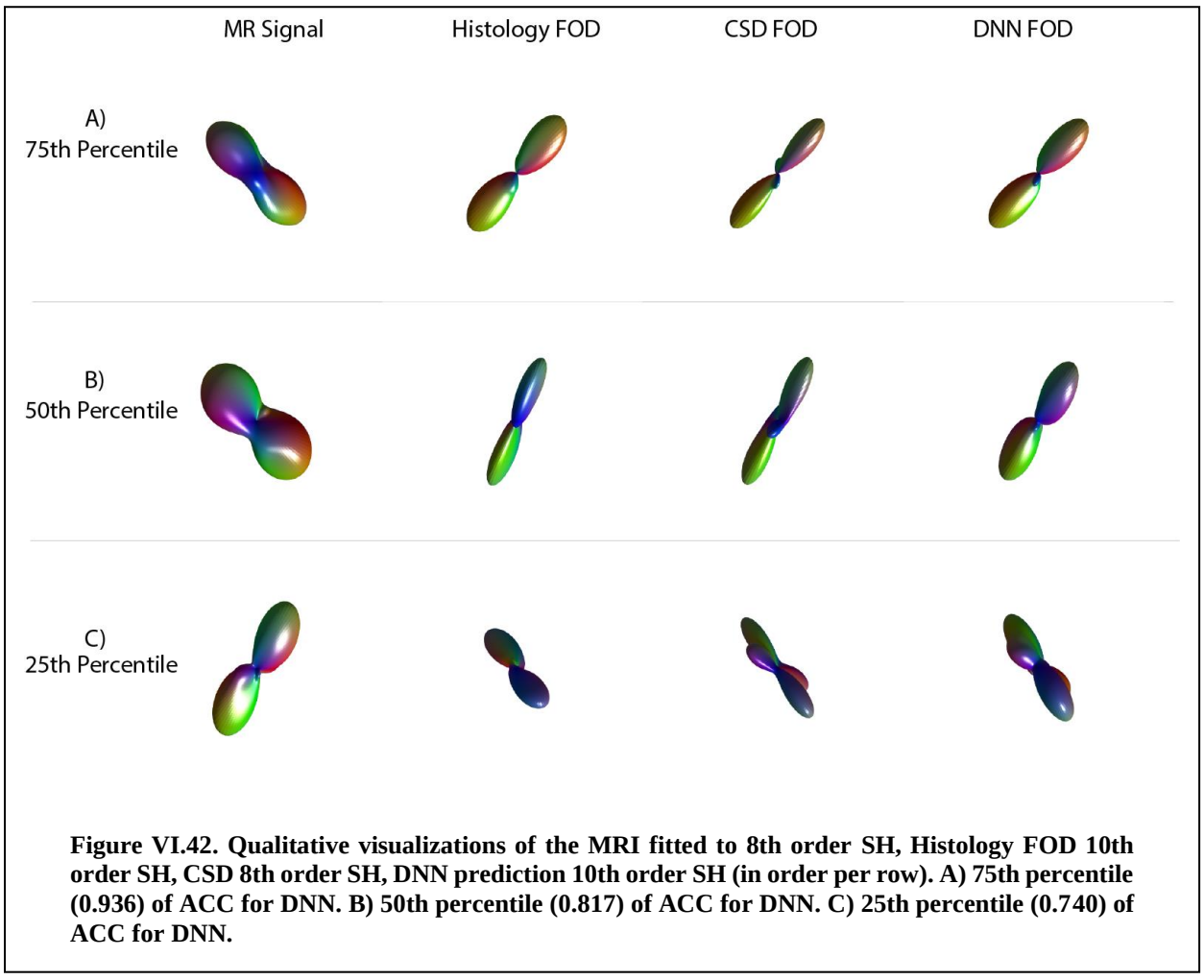

Figure VI.42. Qualitative visualizations of the MRI fitted to 8th order SH, Histology FOD 10th order SH, CSD 8th order SH, DNN prediction 10th order SH (in order per row). A) 75th percentile (0.936) of ACC for DNN. B) 50th percentile (0.817) of ACC for DNN. C) 25th percentile (0.740) of ACC for DNN.

6.4. Discussion

By demonstrating superiority of a deep learning approach over a leading model-based approach, CSD, we show that (1) there exists additional information in the diffusion signal that is not currently exploited by CSD, and (2) provide an illustrative data-driven model that makes use of this information. In a preliminary analysis, we applied the same network to ex vivo imaging at a b-value of 9000 s/mm2 and found a significantly higher ACC (0.850, $p<0.05$, non-parametric signed rank test) for deep learning which is 6.7% higher than CSD. Hence, generalizing the deep learning to use multiple shells and adapt to high b-values is a promising area of exploration. To enable others to investigate our results, the derived TensorFlow models that describe the identified MRI:histology relationships are available on the NITRC project "masimatlab".

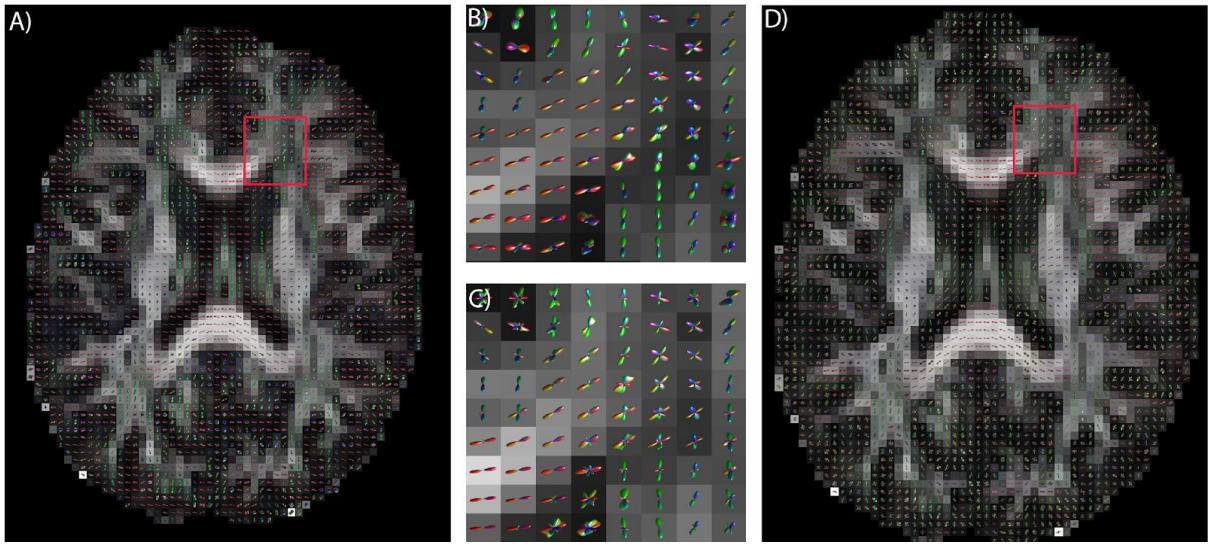

Figure VI.43. A.) Prediction of deep learning model on human in vivo data at a b-value of 2000 s/mm shown on a middle axial slice. B.) deep learning models predictions zoomed region of interest in the pons of corpus callosum. C) CSD predictions zoomed region of interest in the pons of corpus callosum. D) Predictions of CSD on human in vivo data at a b-value of 2000 s/mm shown on a middle axial slice.

Perhaps most importantly, this deep learning analysis demonstrates that current models for identifying fiber orientation distributions do not make all possible use of existing information, and additional innovation is possible. The deep learning models presented herein are preliminary and have not guaranteed optimality properties,

and further exploration of the space of multi-layer neural networks is warranted. Additionally, continued refinement of deep learning approaches could make use of both traditional data augmentation of ground truth (e.g., rotations as used herein), but also physics/diffusion simulations of modeled geometry along with image acquisition models.

7. Deep Learning Reveals Untapped Information for Local White-Matter Fiber Reconstruction in Diffusion-Weighted MRI

This chapter has been adapted from the published work in [25].

7.1. Introduction

Precise reconstruction of white-matter (WM) structural connectivity is of critical importance for the advancement of neuroscience. Though multiple approaches have been proposed for the reconstruction of WM micro-structure a consensus for precision and reproducibility has not been attained [27]. A critical challenge when reconstructing WM connectivity for human MRI in-vivo acquisitions is due to the lack of a ground truth [168, 169]. Multiple approaches have been proposed to reconstruct WM micro-structure and resolve multiple fiber orientations per voxel in the brain [22, 170]. Here, we address two key questions: 1) 'Do state-of-the-art methods precisely capture independently observed structure?' (Figure VII.44) and 2) 'Are data-driven approaches able to better capture this structure while generalizing for in-vivo data?'

Diffusion tensor imaging (DTI) [35] is a diffusion-weighted magnetic resonance imaging (DW-MRI) technique that has been widely applicable for clinical and research use. DTI can only resolve a single direction of fiber structure [92, 93], but it is well known that more complex structures are pervasive, such as crossing and fanning fibers [61, 143]. To resolve complex structures multiple classical approaches have been proposed such as CSD, Q-ball, PAS-MRI [12, 13, 15]. The proposed approaches are often collectively termed as high-angular resolution diffusion imaging (HARDI) methods [18], but HARDI methods are plagued by limited reproducibility and are difficult to assess with respect to precision due to the challenge of lack of a ground truth. Validation studies have highlighted biases, inaccuracies, and limitations of

HARDI methods in characterizing tissue microstructure [171]. This gap in performance suggests that DW-MRI signal may not be utilized completely/sufficiently.

Data-driven approaches can be useful in validation of the hypothesis of the existence of untapped

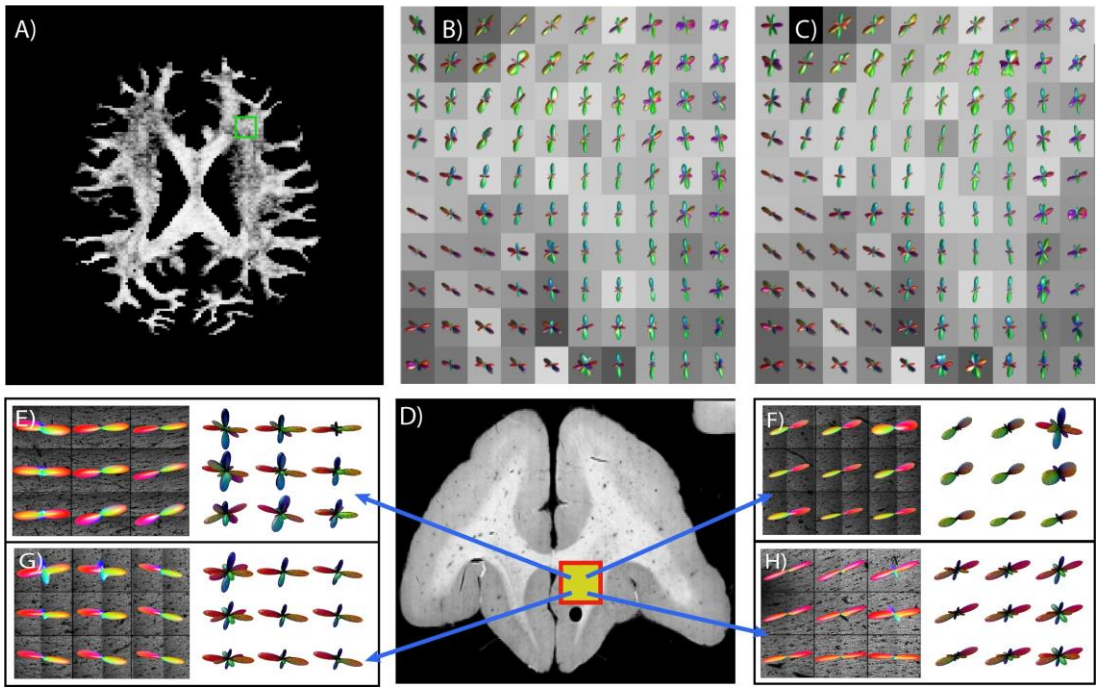

Figure VII.44.A) Middle axial slice of a human brain acquired using DW-MRI. B, C) CSD FOD's of the same subject on two different scanners showing inconsistency which cannot be validated without ground truth. D) Coronal slice of a Squirrel Monkey. E, F, G, H) Right: CSD FOD's at b-value of 3000 s/mm2 Left: Ground truth reconstruction using aggregated histological structural tensors depicting a loss in precision.

information because they generalize towards the ground truth. Deep-learning, a subset of machine learning, has been particularly useful in generalizing non-linear mappings [72]. Herein, we show the evidence of untapped information using a residual deep learning neural network (ResDNN) based on prior work. Our network was trained on a 3-D confocal histology dataset consisting of white matter (WM) voxels from two squirrel monkey brains and validated on a third, independent brain [171]. Further we also show an improvement in terms of precision and reproducibility when the newly created method is applied on scan-rescan data from in-vivo human brain DW-MRI acquisitions from the Human Connectome Project (HCP). The DNN applicable for reconstruction of local white matter reconstruction was first proposed in.

This study presents the first data-driven diffusion-weighted magnetic resonance imaging (DW-

MRI) analysis method that links in-vivo feasible imaging protocols with 3-D histological ground truth. A novelty here is an introduction of combined usage of histology and deep-learning. Under full-specimen validation, the proposed model yields higher predictive accuracy than the leading model-based approach. Hence, this study demonstrates that a significant level of information in the DW-MRI is untapped with current analysis methods. Moreover, we demonstrate that the proposed method is more reproducible in a sample of scan-rescan pairs from the Human Connectome Project. The proposed method generalizes from an ex-vivo pre-clinical context to an in-vivo human context without re-training.

7.2. Methods

7.2.1. Squirrel Monkey Data Acquisitions

All animal procedures were approved by the Vanderbilt University Animal Care and Use Committee. This data comes a prior study [171]. Three ex-vivo squirrel monkey brains were imaged on a Varian 9.4T scanner. A total of 100 gradient volumes were acquired using a diffusion-weighted echo planar imaging (EPI) sequence at a diffusivity value of 9000 s/mm2 with the isotropic resolution being 0.3mm. Once acquired the tissue was sectioned and stained with fluorescent dil and imaged on a LSM710 Confocal microscope followed [171]. The histological fiber orientation distribution was extracted using 3D structure tensor analysis [23]. A multi-step registration procedure was used to determine the corresponding diffusion MRI signal. A total of 567 histological voxels were processed. Respective number of voxels by Monkey A, B and C were 270, 162 and 135. 54 of the total number of voxels were labelled as outliers qualitatively and rejected for analysis. The outlier voxels were detected using ACC when fitted with CSD. A hundred random rotations were applied to the remaining voxels after the removing the outliers. This brought the total number of voxels to 51,813 voxels. All voxels from Monkey C were kept completely hidden for evaluation purpose of the machine learning model. Total number of voxels for Monkey C after outlier removal were 9090. The voxels from monkey A and B were used for the training and they accumulated to a total of 42,723.

7.2.2. Human Connectome Project Data

Human connectome project (HCP) data of 12 subjects with the retest acquisition was used (ID's: 103818, 105923, 111312, 114823, 115320, 122317, 125525, 130518, 139839, 143325, 144226 and 146129). The acquisitions at b-value of 3000 s/mm2 with 90 gradient directions were extracted for the study. A T1 volume of the same subject was used for WM segmentation using FAST [100]. Pre-processed HCP diffusion data was used where topup and eddy have been applied for distortion corrections [140, 159].

7.2.3. Residual Deep Neural Network Regression

The data driven models learn a set of parameters when provided with a set of corresponding inputs and outputs. The set of parameters here depict a non-linear mapping. The input was defined by spherical harmonic (SH) coefficients that were fitted to the ex-vivo DW-MRI acquisition of squirrel monkeys per voxel. This was done using regularized linear least squares fit on the DW-MRI signal. The SH coefficients have been known to characterize the diffusion signal sufficiently [33]. Ortho-normal representation allows for flexibility of different acquisitions. The DW-MRI signal was fitted to 8th order SH [33, 101]. Acquisitions with greater than 45 gradient volumes are known to be characterized well by 8th order SH coefficients. The output for the training network were SH coefficients of 8th order per voxel depicting the fiber orientation distribution (FOD) structure from histology. These were derived from histology [23].

The neural network architecture briefly is five layers deep with the number of neurons per layer being: 45 (Input), 400, 45, 200, 45, 200 and 45 (Output) respectively (Figure VII.45). The middle three layers are used as a combined residual block. Rectified linear units were used for activation only to introduce non-linearity. The output layer was not activated to introduce negativity in the ResDNN as SH coefficients can be negative. Default hyper-parameters of 'RMSProp' optimizer were used while training the network for convergence with the exception of learning rate being set to 10-4 [167]. Cross-validation

set size was set to 0.2. Number of iterations for optimal convergence were determined to be at 400.

7.2.4. HARDI Methods and Angular Correlation Coefficient

Multiple HARDI methods were implemented for comparison to the proposed ResDNN approach. A list of methods that were compared for the acquisition of the test monkey 'C' are Q-ball imaging (QBI),

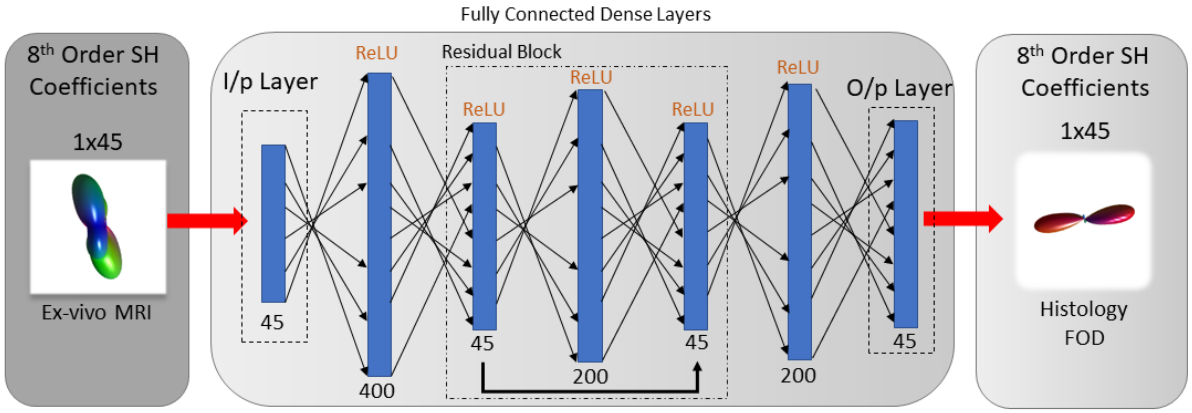

Figure VII.45. Left to Right: Input of SH coefficients of DW-MRI signal at 8th order to the DNN. The middle box depicts the architecture of the ResDNN with number of neurons and activation functions respectively. Output is denoted by SH coefficients of the FOD of a structural tensor at 8th order.

with constant solid angle (QBICS), super resolved constrained spherical deconvolution (sCSD) at 6th and 8th order, Lucy-Richardson constrained spherical deconvolution (CSDLR), diffusion orientation transform (DOT), diffusion orientation transform revisited (DOTr1). All the mentioned HARDI methods were compared using angular correlation coefficient (ACC). ACC is defined as below:

$$ACC = \frac{\sum_{j=1}^{\propto}\sum_{m=-j}^{j} u_{jm} v_{jm}^{*}}{\left[\sum_{j=1}^{\propto}\sum_{m=-j}^{j}|u_{jm}|^{2}\right]^{0.5}\cdot\left[\sum_{j=1}^{\propto}\sum_{m=-j}^{j}|v_{jm}|^{2}\right]^{0.5}} \quad \textbf{Eq. 24}$$

ACC is calculated using 'u' and 'v' where they are two sets of SH coefficients of the same order. It is a measure of similarity on a scale of -1 to 1. A measurement of 1 denotes a perfect similarity.

7.3. Results

Results are presented in three sections: 1) we compare the ground truth histology with HARDI methods and ResDNN with accuracy assessed by ACC [15]. All comparisons were made on a third squirrel monkey which was not used for the training of the network. 2) Quantitative assessment of scan-rescan pairs from HCP dataset using 12 subjects using ACC with sCSD at 6th and 8th order. 3) Qualitative assessment of the fiber orientation distribution (FOD) glyphs and spatial maps of slices from in-vivo HCP brain slices.

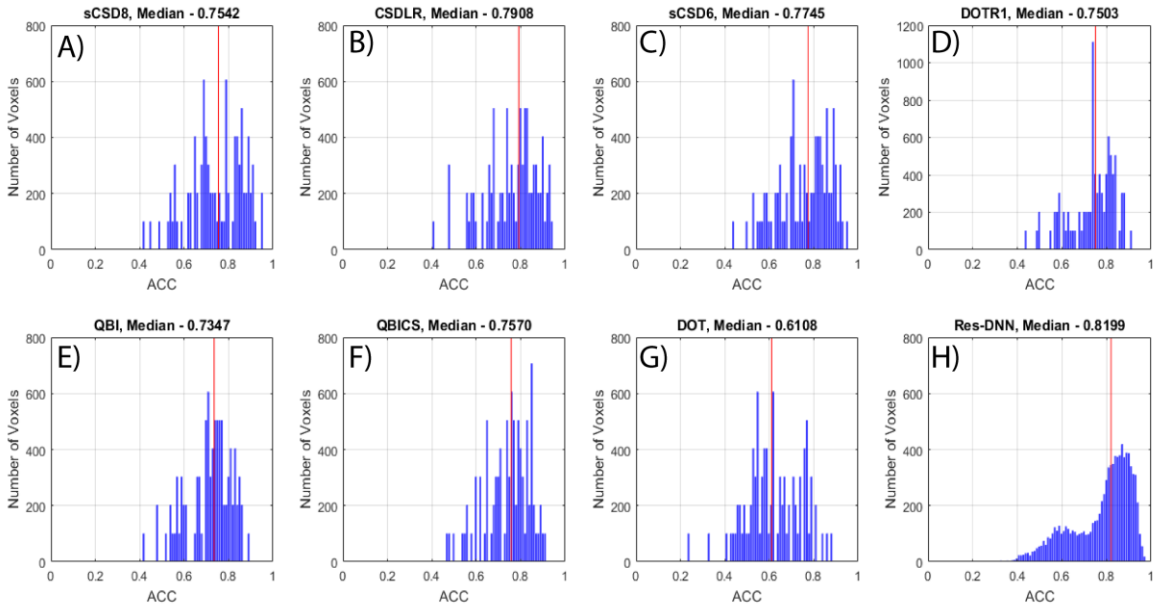

Figure VII.46. Distribution of ACC between the ground truth and the CSD FOD's of monkey C. The red line indicates the median at for the distribution. A, C) super resolved constrained spherical deconvolution at 6th and 8th order respectively with median at 0.75 and 0.77. B) Lucy-richardson constrained spherical deconvolution with median at 0.79. D, G) Diffusion orientation transform and its revisited approach with median at 0.61 and 0.75 respectively. E, F) Q-ball imaging and with constant solid angle with median at 0.73 and 0.75. H) Proposed ResDNN approach with median at 0.82

A trained ResDNN using ground truth based on histology with corresponding diffusion signal reconstructs more precise fiber structure as compared to super resolved spherical deconvolution (CSD). The median ACC of the ResDNN is 0.82 as compared to any other HARDI method approach where all median is <0.79 (Figure VII.46). Non-parametric signed rank tests for the ACC distributions shows that ResDNN is higher than other HARDI methods ($p<<0.01$).

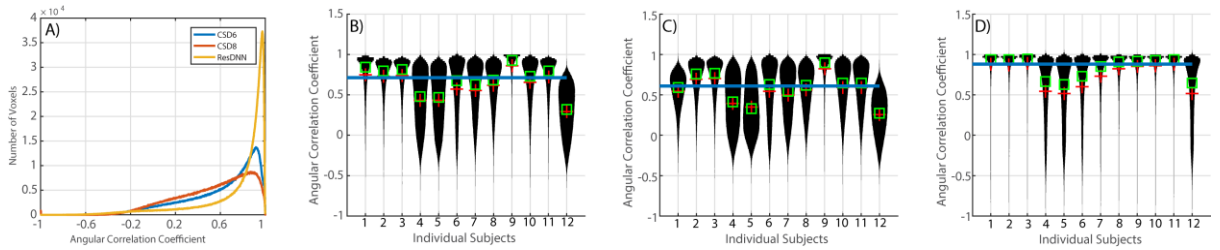

Figure VII.47. A) ACC of all paired WM voxels across the 12 subjects laid out in distribution for the ResDNN, sCSD at 6th and 8th order. ResDNN distribution is skewed towards higher correlation. B, C, D) Violin plots per subject of ACC for paired WM voxels for sCSD at 6th order, 8th order and ResDNN. Blue line denotes the median.

For quantitative assessment on in-vivo DW-MRI acquisitions the ResDNN was tested on paired human in-vivo acquisitions. Twelve test-retest acquisition subjects were acquired from the human connectome project (HCP). The distribution of ACC between the pair of scans for sCSD at 6th order SH,

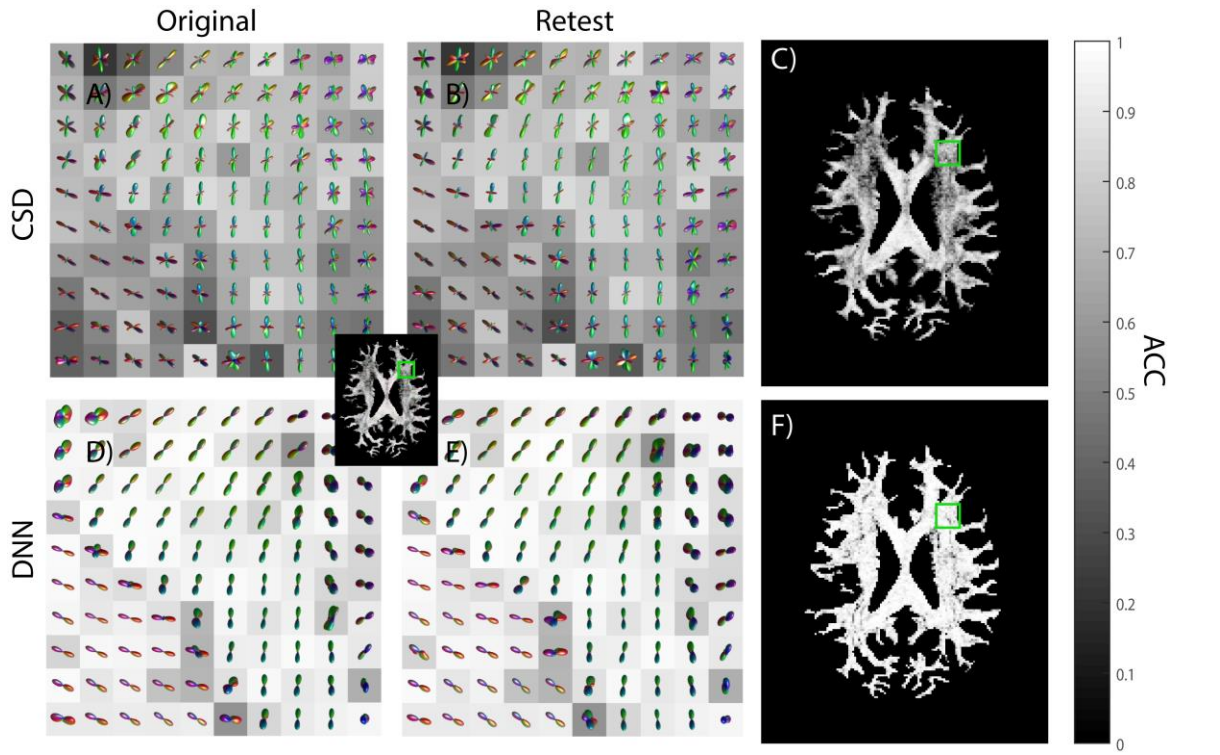

Figure VII.48. The selected ROI shows frontal lobe of WM on right side. The image underlay in A, B, D & E is ACC which shows agreement between pairs of A & B and D & E. C & F show the ACC of the middle axial slice of the brain. FOD of sCSD are shown in A & B where they show lesser agreement as compared to FOD of ResDNN in D & E.

8th order and SH ResDNN are shown in (Figure VII.47). ResDNN shows the most skewed distribution towards higher correlation (Figure VII.47 A). Observing distribution plots per subject the ResDNN shows a more skewed distribution towards higher correlation for all subjects as compared to sCSD at 6th and 8th order. The mean, median and standard deviation of the ACC across the 12 subjects for CSD at 6th order: (0.61, 0.71, 0.31), 8th order: (0.54, 0.60 and 0.31). The same for the ResDNN are (0.74, 0.88, 0.31). The reproducibility gain calculated by the difference of the ACC median of sCSD at 6th order and ResDNN is 24% and the same for 8th order is 46%. Non-parametric signed ranked test for all pairs of ACC distributions show that the ACC of ResDNN is significantly higher than sCSD ($p<<0.01$).

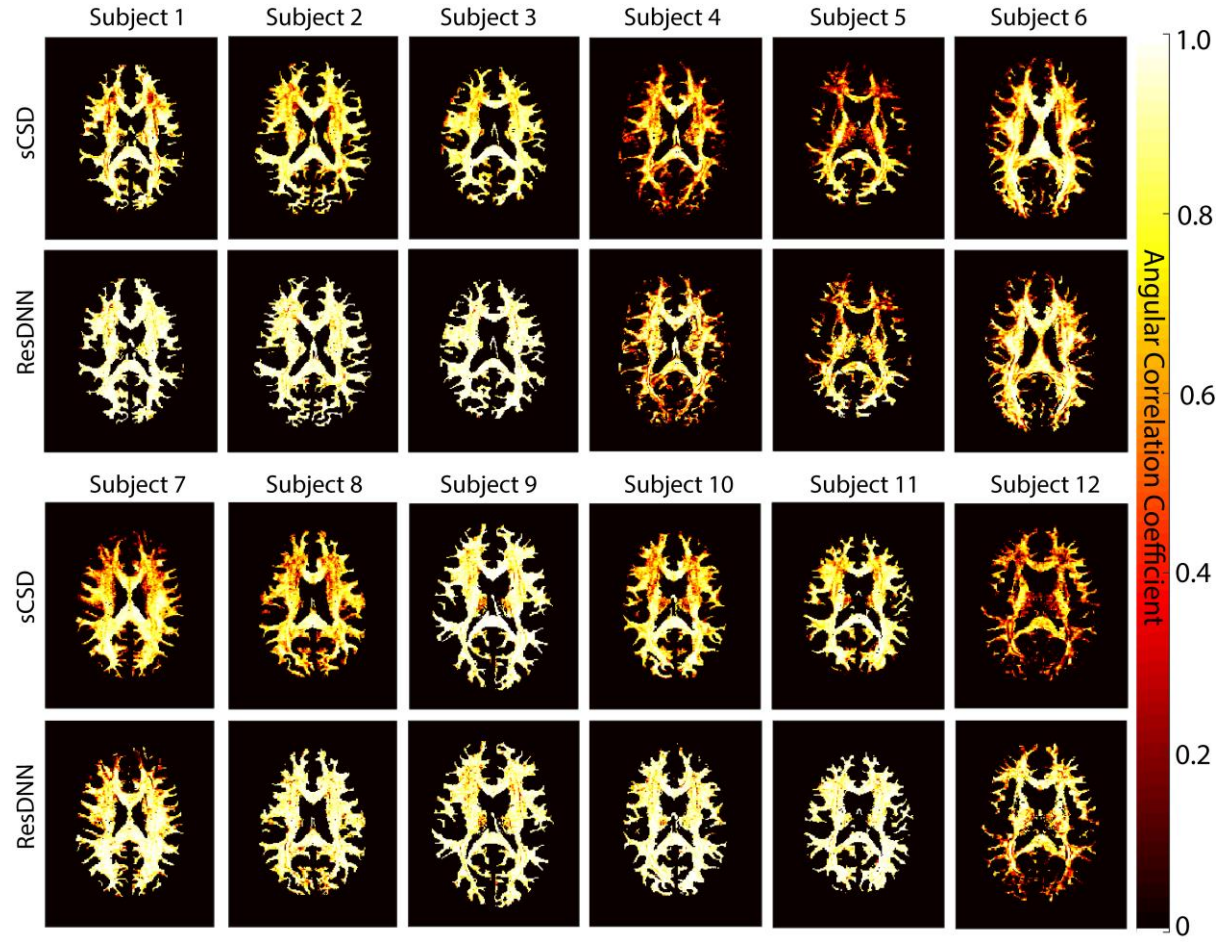

Figure VII.49. Spatial maps of ACC of middle axial slice for scan-rescan pairs are shown for both ResDNN and sCSD at 6th order. First row shows sCSD ACC spatial maps for subjects 1-6. Second row shows ResDNN ACC spatial maps for subjects 1-6. Third row shows sCSD ACC spatial maps for subjects 7-12. Fourth row shows ResDNN ACC spatial maps for subjects 7-12. It can be observed that the ResDNN shows higher ACC for all observed pairs.

We chose a region of interest (ROI) in the right frontal lobe of the white matter (WM) of the brain (Figure VII.48). Comparing sCSD (A & B) we can observe that the FOD glyphs are similar qualitatively for single fiber populations; however, notable differences can be seen for multiple fiber populations. Comparing ResDNN (D & E) the FOD glyphs visually agree more as compared to sCSD specifically glyphs with multiple fiber populations as well. Observing the ACC across the middle axial slice of the brain for sCSD (C) and ResDNN (F), it can be noticed that the ACC is higher across the brain for DNN.

Spatial maps of ACC (Figure VII.49) show that the ACC for ResDNN predicted SH coefficients are higher for all pairs of subjects when compared with ACC of sCSD at 6th order. Subject 4, 5, 6 and 12

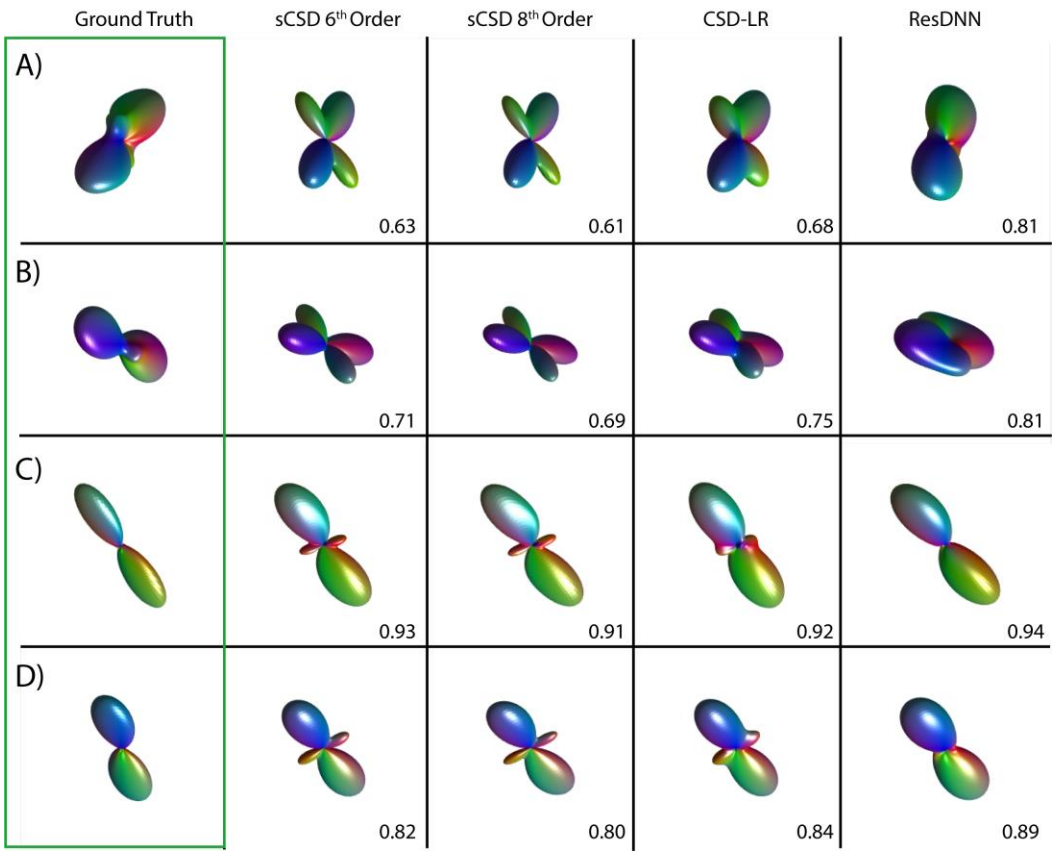

Figure VII.50. A, B, C, D) 1st Column: depicts ground truth FOD with representation as 8th order SH Coefficients. 2nd & 3rd Column: depicts sCSD predicted FOD with representation of 6th and 8th order SH. 4th Column: depicts CSDLR predicted FOD with representation as 8th order SH. 5th Column: depicts ResDNN predicted FOD as 8th order SH Coefficients. ACC is depicted at bottom right corner when compared with ground truth.

show lower ACC relatively for both ResDNN and sCSD as compared to other subjects.

FOD glyphs (Figure VII.50) of sCSD and CSDLR show a spurious detection of fiber population which is not present in the ground truth for single fiber configurations. ResDNN and sCSD both depict similar structure which resembles the ground truth for the principal orientation.

7.4. Discussion

Accurate characterization of the structural properties of the brain can lead to fundamental insights into cognition, development, and diseases. Currently, the only noninvasive method to map the structural connectivity is diffusion MRI. In this study, we use data-driven machine learning approaches to learn a relationship between the diffusion MRI signal and the underlying tissue microstructure. Together, we find that there is information in the diffusion signal that is currently unutilized or underutilized by current methods, and that this information is stable. Specifically, our study has three major takeaways: 1) ResDNN can reconstruct the FOD more precisely/accurately than current diffusion reconstruction methods. 2) The information gain is significant which is indicative of a more precise reconstruction of the micro-structure architecture of the brain, while also providing evidence that a better non-linear mapping exists between the signal and the FOD. 3) In-vivo validation on scan and rescan data used from the HCP confirms reproducibility and stability of the model, demonstrating clinical applicability. Qualitative analysis re-enforces the improved scan-rescan consistency.

While advocating the use of a data-driven approach is nascent in the domain of DW-MRI. A novelty here is the introduction of the combined usage of histology and deep learning. The numerous classical approaches to reconstruct the FOD have been reproducible and successful however the evidence suggests that a more precise one could be discovered. The result of the information gain shows us that there is untapped information present in the DW-MRI signal that is not being accounted for. This study in no way suggests as to what could be another non-linear mapping. The advancements in machine-learning specifically deep learning have made this discovery possible.

Deep learning has been widely applicable in image processing and has become a powerful tool.

While it is still treated as a black box because of the large number of parameters that a ResDNN learns, deep learning is becoming clinically accepted. Although a simple sequential ResDNN has been used for the study, more complex networks can be designed for more specific DW-MRI problems. Spatial information as features can be used for improving the performance of the ResDNN. A precise reconstruction of FOD's directly leads to improved performance for tractography algorithms.

On a side note there are multiple HARDI methods that enforce non-negativity in different ways (cite non-negative papers here). However, they cannot be directly enforced upon deep learning, hence, to enforce non-negativity we follow a proposed procedure (Figure VII.51A). Starting out with normalized DW-MRI signal it is truncated by a regularizing hyper-parameter (theta=0.005). This is needed to eliminate negative and extremely low values (<0.005) as they would translate to noisy values in log transform. The log transform of DW-MRI signal was fitted to SH leading to them being in log space. This procedure was performed on the FOD SH coefficients as well. A consistent set of 100 gradient directions was used which uniformly sampled over a sphere. While testing the withheld set of voxels were not transformed in any way to ensure integrity of blind testing. In comparison the non-negative results show a median ACC of 0.81 as compared to ResDNN ACC of 0.82. While there is a little compensation in correlation, we are able to enforce non-negativity in the network. The non-negativity requires a deeper validation to show further applicability.

Apart from precision the criteria of reproducibility have been validated as well using scan-rescan pairs from the HCP. Reproducibility while not correlated with precision has its own independent importance due to the fact of clinical applicability. Precision without reproducibility is an example of high variance in the information being reconstructed. For example, if the reproducibility is not high for a healthy control subject then the interpretation of the reconstruction on a diseased subject would be an unreliable one. Here we have shown that the ResDNN achieves high reproducibility across two different datasets: HCP and in-

house acquired data. This model marks the advent of artificial intelligence (AI) in DW-MRI.

AI has gained quite the popularity and applicability in multiple domains. The ResDNN presented

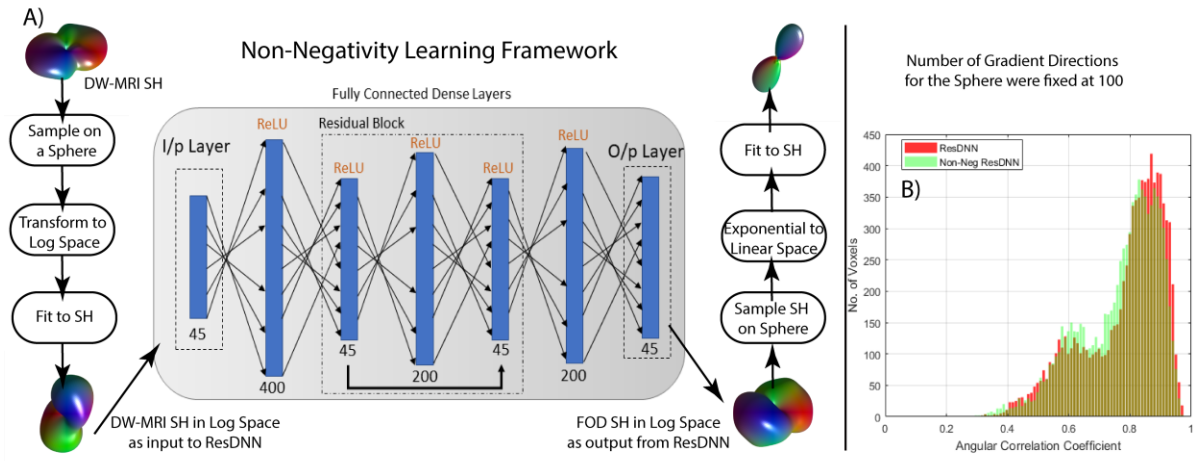

Figure VII.51. A) Describes the step procedure followed to incorporate non-negativity which consists of a log transformation to eliminate negative values and using log space SH coefficients for input and output for ResDNN network. B) Shows the comparison between the ground truth and predictions from ResDNN in linear space and log space.

here can be foreseen as a small part of an overall data-driven approach that could eventually be constructed for DW-MRI data. There are multiple clinical issues regarding DW-MRI that can be tackled such handling of multiple acquisition hyper-parameters and acquisition details.

The ResDNN presented can only learn as well as the presented ground truth to it which could be improved upon (improved registration, SNR, artifact suppression, etc.). Also, another limitation is that the DW-MRI acquisition was ex-vivo for the training of the ResDNN. However, it was validated upon human in-vivo data. Lastly, the dataset that was created was sparse. Standard data augmentation techniques could possibly be useful in making the dataset usable for training this model for the study. We hope that the work presented in this study will be helpful in guiding researchers and neuroscientists better for forward looking research in AI with DW-MRI.

7.5. Conclusion

We have shown three key findings with the presented study. 1.) A machine learning framework can be used to directly learn the non-linear mapping between DW-MRI and 3D confocal histological derived ground truth. 2.) The trained ResDNN is able to show that it is able to utilize untapped information which is present in the DW-MRI signal. 3.) The trained ResDNN model is able to perform more consistently when relatively compared to state of the art sCSD.

8. Inter-scanner Harmonization of High Angular Resolution DW-MRI using Null Space Deep Learning

This chapter has been adapted from the published work in [88].

8.1. Introduction

The Diffusion-weighted MRI (DW-MRI) provides orientation and acquisition-dependent imaging contrasts that are uniquely sensitive to the tissue microarchitecture at a millimeter scale [27]. Substantial effort has gone into modeling the relationship between observed signals and underlying biology, with a tensor model of Gaussian processes being the most commonly used model [35]. Voxel-wise models that characterize higher order spatial dependence than tensors fall under the moniker of higher angular resolution diffusion imaging (HARDI) [130]. Recently, a myriad of techniques has emerged to estimate local structure from these diffusion measures [12, 13, 15, 165]. However, broad adoption and clinical translation of specific methods has been hindered by a lack of re-producibility [61, 94], inter-scanner stability [172, 173] , and anatomical specificity when compared to a histologically defined true microarchitecture [24, 171]. There are known critical issues of the inter-scanner diffusion harmonization that go beyond noise effects [51, 52, 174].

Recently, it has become feasible to apply a data-driven approach to estimate tis-sue microarchitecture from in vivo diffusion weighted MRI using deep learning. This approach relied on a histologically defined truth with correspondingly paired voxels with diffusion weighted magnetic resonance imaging data. Yet, no approaches to date have addressed inter-scanner variation and scan-rescan reproducibility. Moreover, traditional deep learning architecture do not specifically create models that have these necessary characteristics for clinical translation. Here, we propose a new learning architecture, the null space deep network (NSDN), to address the short comings of precision and reproducibility across scanners. Within the NSDN framework, we use inter-scanner paired in vivo human data to stabilize the data driven approach linking pre-clinical DW-MRI with histological data. Using a withheld dataset, the NSDN

method is compared against a previously published fully connected network and the leading model-based approach, super resolved constrained spherical deconvolution (CSD) [175] in terms of the precision with which the model captures histologically defined truth from DW-MRI data, the reproducibility of the approach on in vivo human data, and the generalizability of the model to in vivo data acquired on an additional MRI scanner.

The remainder of this manuscript is organized as follows. Section 2 presents the acquisition and processing of all the data that has been used for the study. Section 3 presents the design and the parameters of the proposed network architecture. Section 4 presents the results. Section 5 presents the conclusion.

8.2. Data Acquisition and Processing

Three ex-vivo squirrel monkey brains were imaged on a Varian 9.4T scanner (Figure VIII.52). A total of 100 gradient volumes were acquired using a diffusion-weighted EPI sequence at a diffusivity value of 6000 s/mm2, acquired at an isotropic resolution of 0.3mm. Once acquired, the tissue was sectioned and stained with fluorescent dil and imaged on a LSM710 Confocal microscope following procedures outlined in [24, 171]. A similar procedure is outlined by [176]. The histological fiber orientation distribution (HFOD) was extracted using 3D structure tensor analysis. A multi-step registration procedure was used to determine the corresponding diffusion MRI signal. A total of 567 histological voxels were processed. 54 voxels of these were labelled as outliers qualitatively and were rejected from the analysis. A hundred random rotations were applied to the remaining voxels for both the MR signal and the HFOD to augment the data and bringing the total to 51,813 voxels [166]. A withheld set of 72 test voxels was maintained for

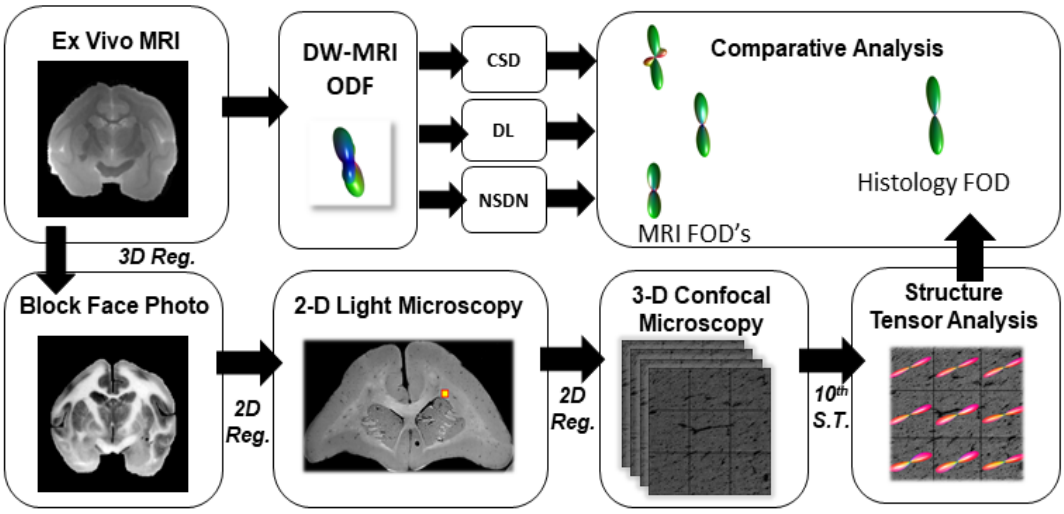

Figure VIII.52. Formation of the training dataset where 2D histology was performed on the squirrel monkey brains and FOD's were constructed per voxel basis using ensemble structure tensor analysis which correspond to *ex vivo* MRI acquisition of the squirrel monkey brains. Comparative analyses were performed between the reconstructed histology FOD's and FOD's from CSD, DN and NSDN.

validation. With rotations, these total to 7,272 voxels.

The in vivo acquisitions of the human subjects' data were acquired on three different sites, referred to as 'A', 'B' and 'C'. Three healthy human subjects were ac-quired with a scan each at the sites in the following manner. Subject 1: Site 'A' and Site 'B'. Subject 2: site 'A' and site 'B'. Subject 3: site 'B' and site 'C'. Structural T1 MPRAGE were acquired for all subjects at all sites. The diffusion acquisition protocol and scanner information are listed for each of the sites as follows.

Site 'A' was equipped with a 3T scanner with a 32-channel head coil. The scan was acquired at a diffusivity value of 2000 s/mm2 (approximating diffusion contrast of fixed ex-vivo scan at a b-value of 6000 s/mm2). 96 diffusion weighted gradient volumes were acquired with a 'b0'. Briefly the other parameters are: SENSE=2.5, partial Fourier=0.77, FOV=96x96, Slice=48, isotropic resolution: 2.5mm.

Site 'B' was equipped with a 3T scanner with a 32-channel head coil. All the parameters of the scan acquisition were as of scanner at site 'A' except for the isotropic resolution which was 1.9mmx1.9mmx2.5mm and up-sampled to 2.5mm isotropic.

Site 'C' was equipped with a 3T scanner with a 32-channel head coil. The scan acquisition

parameters were same as that of site ‘A’, except for the number of slices (n=50) and GRAPPA=2 (instead of SENSE).

All in vivo acquisitions were pre-processed with standard procedures eddy, topup, b0 normalization and then registered pairwise per subject [97, 98]. T1s were registered and transformed to the diffusion space. Brain extraction tool was used for skull stripping [177]. FAST white matter (WM) segmentation was performed using the T1 for the in vivo data [100]. Note that there were three pairs of pre-processed acquisitions in total.

The pair of data from Subject 1 along with the histology data set was used for the training of NSDN. The pairs of data from Subject 2 and 3 were used for quantitative and qualitative evaluation of the network. No site ‘C’ data were used in training.

8.3. Method: Network Design

Our proposed null space architecture is motivated by the linear algebra null spaces in that we need to design/constrain the aspect of the network that has no impact on the outcome. This work is inspired by [178] in which a person re-identification classification problem in computer vision was addressed using a Siamese architecture deep network. The novelty of our approach is that we use paired (but unlabeled) data to train the data-driven network to ignore potential features that would lead it to differentiate between the paired data.

The proposed network design takes three inputs of 8th order spherical harmonic (SH) coefficients (Figure VIII.53). Each input provides an orthonormal representation of the DW-MRI signal and is known to characterize the angular diffusivity signal well [33]. The network outputs a 10th order SH FOD. The base network consists of five fully connected layers; the numbers of neurons per layer are 45, 400, 66, 200 and 66 in the respective order. Activation functions of ‘ReLU’ have only been used for the first two layers. They have not been used for the remainder of the layers to allow for negativity in the network because SH coefficients need not be positive. The three outputs obtained by the network are merged with a common loss function which optimizes on the assumption that the pairwise difference should be zero given the

subject is the same and there should not be a difference in the FOD being predicted. For implementation simplicity, a modified weighted square loss function was defined as (here in λ=1):

$$\boldsymbol{L = \frac{1}{m}\sum_{i=1}^{m}(y_{true_i} - y_{pred_i})^2 + \lambda\,(P_{a_i} - P_{b_i})^2} \qquad \textbf{Eq. 25}$$

'm' is the total number of samples. P_a and P_b are paired in-vivo voxels. A sample size of 37,648 pairs of paired WM voxels were extracted from subject 1 using the acquisitions from site 'A' and site 'B'. A random selection of 37,648 data points was made from the training data set of the histology voxels. While training the network a K-fold cross-validation was used with K=5. The cross-validation set size was set to 0.2. 'RMSProp' was used as the optimizer of the network [167]. The number of iterations was determined at 3 using cross-validation. A batch size of 100 has been used.

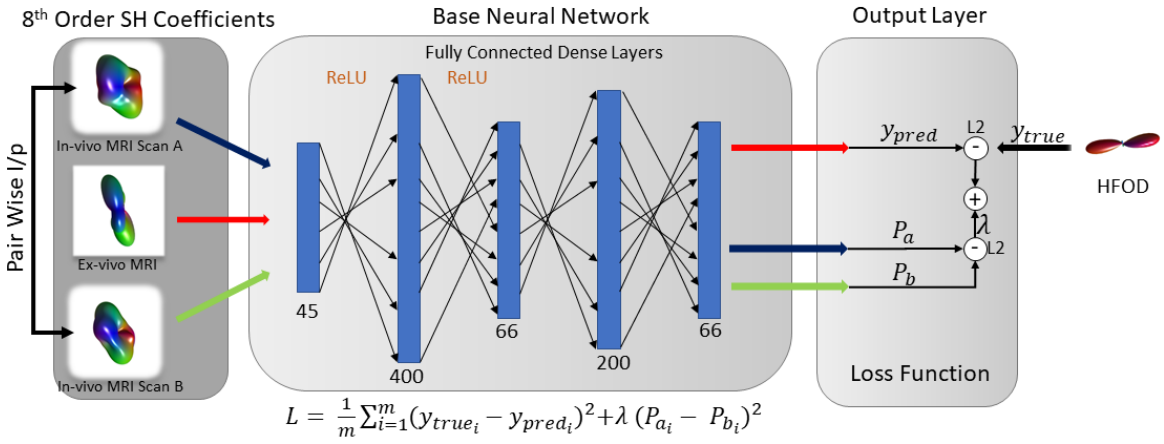

Figure VIII.53. Network design for the null space architecture. The architecture depicts how pairwise inputs of in-vivo voxels can be incorporated in a deep neural net architecture and can be added in the loss function as a noise enhancement/augmentation technique.

To evaluate the performance, we use Angular Correlation Coefficient (ACC) which describes the correlation between two FOD's on a scale of -1 to 1 [15], where '1' is the best outcome.

8.4. Results

The median of the ACC computed from the blind set of 7,272 augmented histology voxels for CSD, DN and NSDN were 0.7965, 0.8165, and 0.8281, respectively. Non-parametric signed rank test for all pairs of distributions were found to be $p < 0.01$. Qualitatively, we explore the results relative to the truth voxel in Figure VIII.54. At 25th percentile it can be observed that CSD and DN show a crossing fiber structure when compared to HFOD. NSDN is representative of more similar single fiber structure of histology. At 50th percentile CSD tends to show a crossing fiber structure, while DN and NSDN show a higher ACC and are like the structure of histology. At 75th percentile all three methods closely resemble the histology

For subject 2, the histogram distribution of ACC for NSDN is most skewed (to-wards higher ACC)

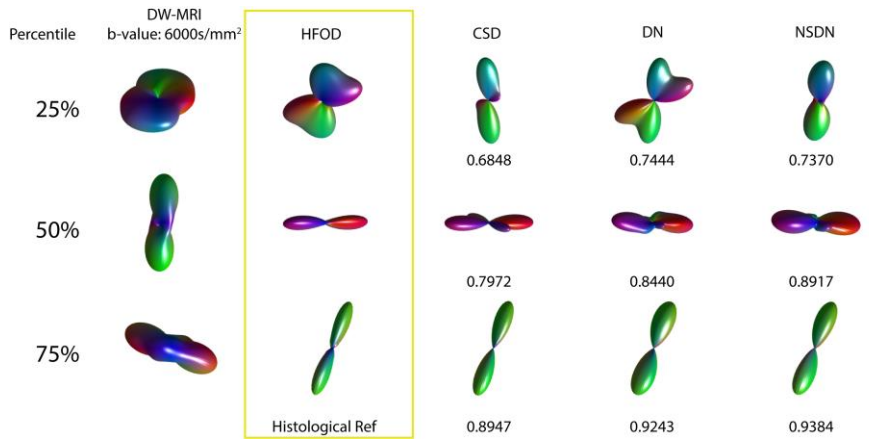

Figure VIII.54. Representative voxels are shown for the 25th, 50th, and 75th percentiles of CSD ACC along with corresponding DW-MRI, DN and NSDN glyphs and ACC's.

compared to DN and CSD (Figure VIII.55A). The median values for the ACC distributions of CSD, DN and NSDN are: 0.67, 0.74 and 0.82. The gain in performance is (calculated by the difference of the medians) is 21.19% for (CSD, NSDN) and 10.09% for (DN, NSDN). Non-parametric signed rank test for all pairs of ACC metrics per voxels for subject 2 resulted in $p < 0.001$.

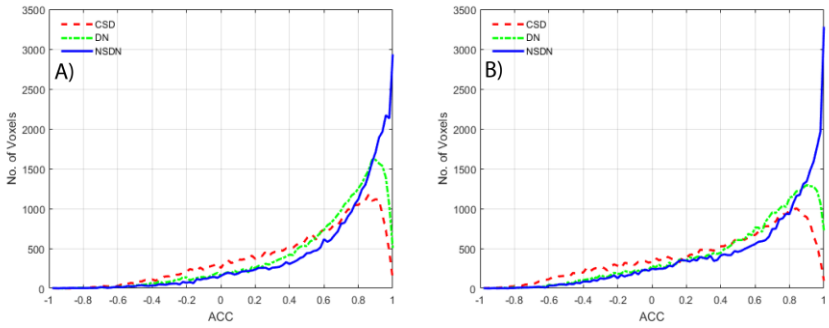

Figure VIII.55. A.) Histogram peaks of ACC per bin distributed over 100 bins for subject 2 of CSD, DN and NSDN. B) Histogram peak of ACC per bin distributed over 100 bins for subject 3 of CSD, DN and NSDN.

Qualitatively, we explore the spatial diffusion inferred structure in the WM of the frontal lobe of the middle axial brain slice (Figure VIII.56). CSD (A & D) show low correlation and spurious fibers in the crossing fiber regions. DN (B & E) improves correlation in crossing fiber regions however NSDN (C & F) shows the highest correlation for crossing fiber regions. For single fibers, all three methods show high correlation.

In the quantitative results for subject 3, we observe that the skewed distribution towards higher ACC for NSDN is the highest as compared to both the other methods (Figure VIII.55 B). The median for three distributions of CSD, DN, and NSDN are 0.62, 0.67 and 0.74. The performance gain of NSDN over CSD is 16.08% and DN is 10.41%. Non-parametric signed rank test for all pairs of voxels for subject 3 show $p < 0.001$.

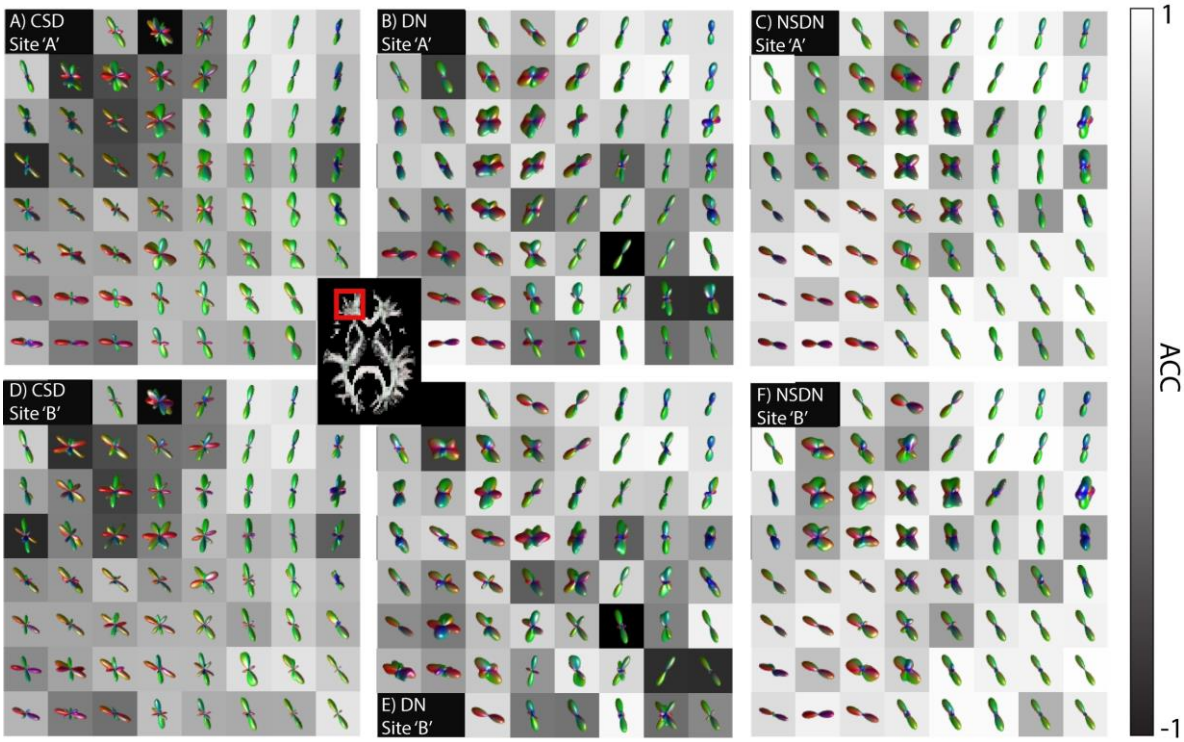

Figure VIII.56. The selected ROI shows internal capsule and a part of the global pallidus. The image underlay for the ROI's shown is Angular Correlation Coefficient (ACC) to indicate areas of agreement between scan-rescan (vertical pairs). CSD shows high correlation for core white matter where single fiber orientation exists (observer diagonal pattern of high ACC). DN shows increased correlation over a broader region which encompasses crossing fibers. NSDN shows a higher correlation across the most extended anatomical area.

8.5. Conclusion

The NSDN method for reconstructing fiber architecture is (1) more accurate when compared to histologically defined FODs, (2) more reproducible qualitatively and quantitatively on scan-rescan data, and (3) more reproducible on previously unseen scanners. While histological-MRI paired datasets are exceedingly rare, scan-rescan data are ubiquitous and often acquired as part of multi-site studies. The NSDN meth-od provides a natural framework for harmonization that can use already acquired scan-rescan data to ensure that analysis methods are as reproducible across all sites. A much wider comparative study with multiple different HARDI methods and using multiple scanners is warranted. It would be interesting to

explore the impact of including data from diffusion phantoms to enhance the diversity of signals captured in a data-driven approach.

While this work focused on DW-MRI, the NSDN approach can naturally be ap-plied to other deep learning-based networks with two relatively simple modifications. First, one needs to construct a multiple channel network graph of the same form as "base network," but with shared weights for all channels and without cross-connections between the channels. This will ensure that one input can be placed per channel and all inputs will see the "same" base network. Second, the loss function needs to be modified so it combines a traditional loss with a reproducibility loss. The traditional loss comes with (without loss of generality) from the first channel's output relative to a traditionally provided truth dataset. The reproducibility loss is then computed by a metric of reproducibility between the remaining channels (herein a weighted squared error metric, but Dice, surface distance, etc. could be used as ap-propriate for the datatype). The potential synergies with data augmentation and neighborhood information have yet to be explored.

9. Harmonizing 1.5T/3T Diffusion Weighted MRI Through Development of Deep Learning Stabilized Microarchitecture Estimators

This chapter has been adapted from the published work in [179].

9.1. Introduction

Diffusion-weighted magnetic resonance imaging (DW-MRI) is a non-invasive imaging modality that can provide information about microarchitecture of the brain at a millimetric scale [27, 168, 180]. Multiple DW-MRI approaches have been proposed to model the DW-MRI signal to estimate the 3D microstructure [22]. A common approach is diffusion tensor imaging (DTI) which can capture a single fiber population per voxel [35, 181, 182]. Although DTI has been established as reproducible and characterizes the DW-MRI signal well, it is limited that it cannot capture complex fiber microstructure like fanning, crossing, kissing fibers [93, 183]. Advanced approaches fall under the nomenclature of high angular resolution diffusion imaging (HARDI) methods [18]. HARDI methods are routinely applied to DW-MRI acquisition with more than 45 gradient volumes [18], as reviewed in [22]. While HARDI methods are quite reproducible (>0.7) in terms of fiber counts within a scanning protocol on a single device, but quite different across methods [19, 61, 94]. Challenge studies depict a similar story [22]. In the context of DW-MRI signal the measurements are affected by different scanners and HARDI method reconstruction is impacted due to these effects [184]. Hence, harmonization is a critical problem where the microarchitecture reconstruction is impacted by protocol effects, scanner effects, and site effects (Figure IX.57).

Harmonization has been pursued at for DTI using COMBAT to harmonize the metrics of fractional anisotropy (FA) and mean diffusivity (MD) [51]. While COMBAT is a method of harmonizing the metrics directly, [174] was proposed to harmonize the DW-MRI signal using rotational invariant spherical harmonic (RISH) features [54, 174]. Previously meta-analysis techniques have been used to quantify group differences for e.g., FA [50, 185]. Statistical covariates have been used to characterize site differences as well [186]. Recently, an approach to harmonize the microstructure was recently proposed to harmonize the microarchitecture directly in the space of spherical harmonics [187]. However, this approach relied on scanners of the same manufacturer and similar field strength and only relied on just an extra L2 loss for harmonization. The time separation between acquisitions were within ~2 weeks. This approach was built on the basis of sequential DNN approach. Herein, we extend the approach to inter-field strength harmonization and evaluate in a rather extreme case of 4 years mean separation. Moreover, we introduce a

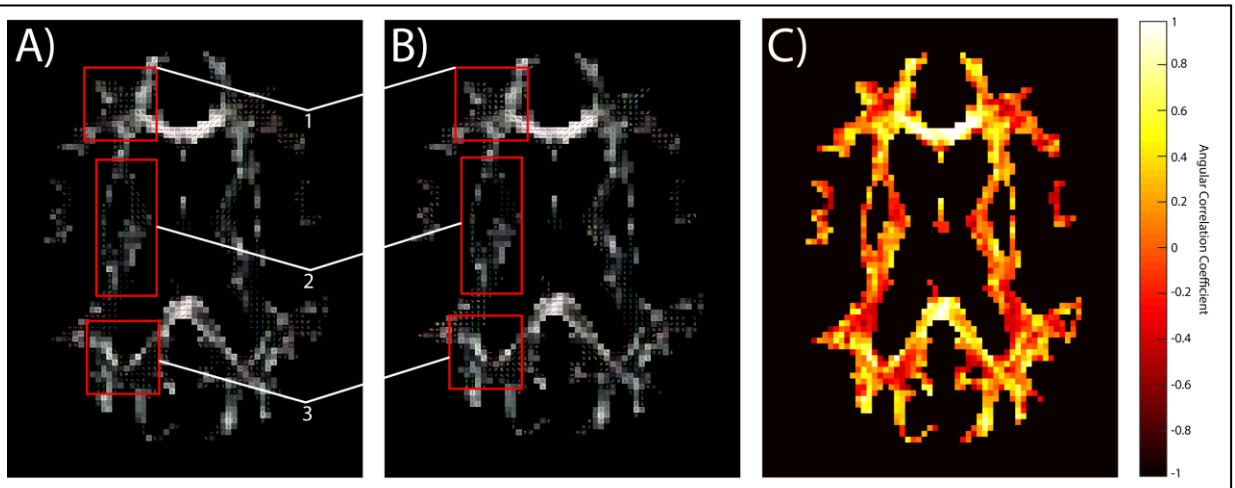

Figure IX.57. A) Middle axial slice on a 1.5T scanner depicting CSD FOD's across the WM. B) Middle axial slice on a 3T scanner depicting CSD FOD's across the WM. C) Angular Correlation Coefficient of the FOD's which is calculated on corresponding voxels of 1.5T and 3T scanner. Annotations 1,2 and 3 depict low correlation regions of the brain on the left side, same can be observed on the right side.

modification to loss function to improve consistency of scalar metric consistency while maintaining FOD consistency.

The NSDN approach is based on using a dataset where corresponding DW-MRI and ground truth FOD of microstructure was reconstructed using histology [171]. The dataset consists of white matter (WM)

voxels from three squirrel monkeys. A deep neural network (DNN) regression model could learn the histology using the inputs of DW-MRI on a voxel by voxel basis. The proposed NSDN approach takes paired inputs of voxels of 17 subjects (1.5T and 3T scanners). The hypothesis is that when predicted/estimated their outcome should be the same, this has been modelled using a custom loss function and with the help of identical dual networks. A total of 37 subjects from the Baltimore Longitudinal Study of Aging (BLSA) were identified with paired acquisitions on 1.5T and 3T scanners. 20 subjects were completely withheld for testing of the proposed approach. CSD and DNN as baseline.

9.2. Data & Methods

9.2.1. Squirrel Monkey Data Acquisition

Three squirrel monkeys ex-vivo acquisition were acquired on a Varian 9.4T scanner [23]. A total of 100 gradient volumes were acquired using a diffusion-weighted EPI sequence at a b-value of 6000 s/mm2. Voxel resolution was set at 0.3 mm. Once acquired the tissue was sectioned and stained with fluorescent dil and imaged on a LSM710 Confocal microscope with procedures outlined in [171]. The ground truth histological FOD (HFOD) was extracted with 3D structure tensor analysis. A total of 513 histological voxels were processed with corresponding DW-MRI after quality control. 100 random rotations were applied to each voxel thus augmenting the dataset to 51,813 voxels. A withheld set of 7,272 voxels were maintained for validation. The DW-MRI signal and HFOD were both represented in spherical harmonic (SH) coefficients of the 6th order using a regularized linear least square fit. 44,541 voxels were used for training DNN and NSDN.

9.2.2. BLSA Data Acquisition

37 control subjects were imaged after informed consent and institutional review board approval and retrieved in de-identified form from the BLSA based on paired acquisitions on 1.5T and 3T scanners [10]. All acquisitions were extracted with their corresponding MPRAGE. All scans were processed with topup and eddy for correction of susceptibility distortion and eddy currents [140, 159]. Skull stripping was

performed using FSL on all scans [177]. Multi-atlas segmentation was done on the MPRAGE using BrainCOLOR protocol [123]. The segmented image and the diffusion image were all co-registered using flirt [99, 117]. The DW-MRI was limited to only WM of the brain using the segmented image. All DW-MRI voxels were transformed to 6th order SH coefficients using a regularized linear least square [33] fit for training and testing the NSDN. A total of ~750,000 WM voxels were extracted from the 17 training subjects. 44,541 WM voxels were randomly selected from the BLSA ~750,000 WM voxels to down sample the unpaired training data for NSDN.

9.2.3. Null Space Deep Network

The NSDN architecture is built upon the base network of the DNN (Figure IX.58). The base network is composed of all fully connected dense layers with 28, 400, 66, 200 and 28 neurons per layer. The first three layers were activated using 'ReLU' while the remaining two outer layers were linearly activated to ensure that negativity as a SH coefficient prediction can be negative. Inputs to DNN are 6th order DW-MRI SH Coefficients and outputs are 6th order FOD SH Coefficients. Both are 1D vectors of 28

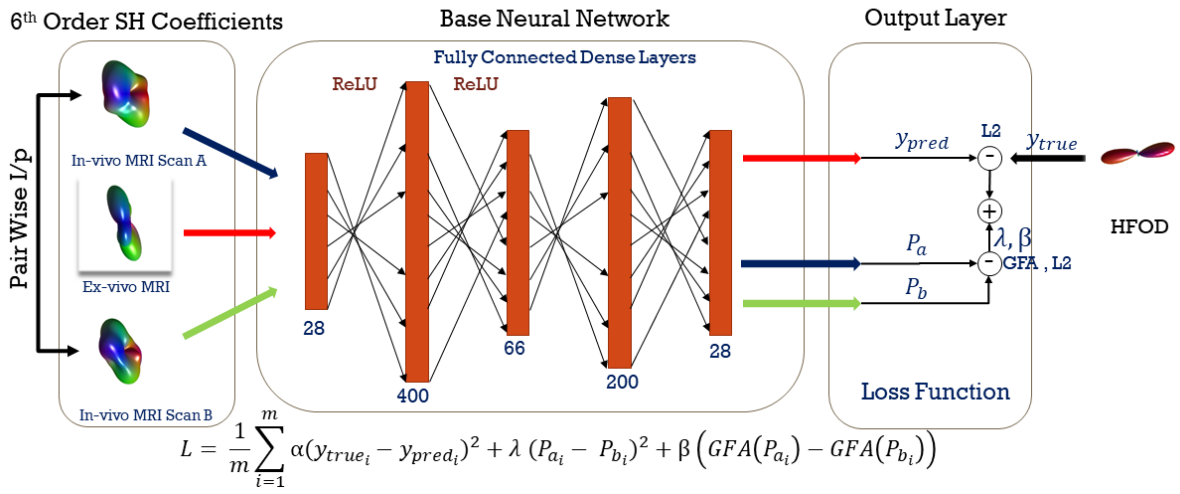

Figure IX.58. The base regression neural network is depicted in the center which describes the parameters of the fully-connected dense layers with respective activation functions. The plot at left: Depicts three inputs where the center input comes from corresponding DW-MRI with histology. The other two are pairwise inputs from corresponding voxels of scanner 1.5T and 3T. The plot at right: Depicts the loss function which uses the hypothesis that the outcome/prediction should be same irrespective of the scanner gradient strength.

coefficients.

NSDN is inspired by linear algebra null spaces where the motivation is that the network must be constrained in such a way that the outcome is left unaffected. A similar architecture was first proposed for person re-identification problem [188]. The novelty in this approach is that paired unlabeled data is used to drive the training of the network where it learns to ignore features of scanner noise and inter-scanner bias which would otherwise lead the network differentiate between the data. The NSDN architecture follows the DNN structure and has an identical dual network of the DNN. This implies the core network to have shared weights where it can learn and ignore potential scanner bias features. NSDN takes 3 inputs of 6th order DW-MRI SH coefficients where the first comes from the histology data and the other two are a corresponding/pair of voxels of the same subject from 1.5 Tesla and 3 Tesla scanners. The output of NSDN is a 6th order FOD SH coefficient.

The NSDN uses a custom loss function which is an additive combination of three L2 losses and all are governed by their individual hyper-parameters (1). The first is the L2 loss between the network's FOD prediction and the ground truth with the hyper-parameter 'α'. The second is the L2 loss between the paired voxels which has an expectation of 0 and the hyper-parameter is 'λ'. The third is the L2 loss between the GFA of the paired voxels which also has an expectation of 0 and the hyper-parameter is 'β'. The GFA is defined by std(ψ)/rms(ψ) where ψ is a set of SH Coefficients. GFA describes the anisotropic nature of the expansion of SH coefficients. The hyper-parameter 'α' was set to 2.75 in an ad-hoc manner. This was done to balance the relative energies of the other two loss function terms.

$$\boldsymbol{L = \frac{1}{m}\sum_{i=1}^{m} \alpha(y_{true_i} - y_{pred_i})^2 + \lambda\,(P_{a_i} - P_{b_i})^2 + \beta\left(GFA(P_{a_i}) - GFA(P_{b_i})\right)} \quad \textbf{Eq. 26}$$

All methods were evaluated using ACC (2), MSE and difference in GFA. ACC is a metric for calculating correlations of functions of all directions over a spherical harmonic expansion. It provides an estimate of how closely two FOD's relate to each other. 'u' and 'v' are the two sets of expansions of SH for corresponding voxels in all analysis. A value of '1' for ACC describes high correlation.

$$ACC = \frac{\sum_{j=1}^{\infty}\sum_{m=-j}^{j} u_{jm} v_{jm}^{*}}{\left[\sum_{j=1}^{\infty}\sum_{m=-j}^{j}|u_{jm}|^{2}\right]^{0.5}\cdot\left[\sum_{j=1}^{\infty}\sum_{m=-j}^{j}|v_{jm}|^{2}\right]^{0.5}} \quad \textbf{Eq. 27}$$

9.3. Results

The ACC of the NSDN has the most skewed distribution towards '1' as compared to both the baseline approaches (Figure IX.59A). DNN has the highest variance of MSE distribution with CSD having a lower variance and NSDN as the lowest (Figure IX.59 B). CSD and NSDN show an equally controlled distribution for difference in GFA while DNN having a higher variance as compared to the other two

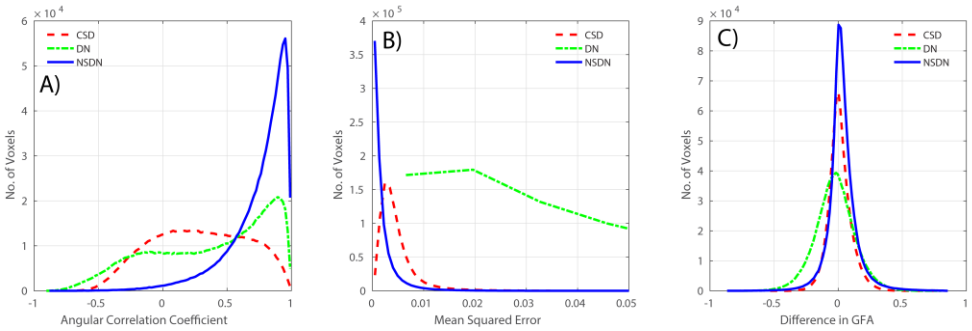

Figure IX.59. A) Shows ACC between 1.5T and 3T scanner acquisitions across all WM voxels of 20 withheld subjects for the three methods CSD, DNN and NSDN. B) Shows MSE between 1.5T and 3T scanner acquisitions across all WM voxels of 20 withheld subjects for the three methods CSD, DNN and NSDN. C) Shows difference in GFA between 1.5T and 3T scanner acquisitions across all WM voxels of 20 withheld subjects for the three methods CSD, DNN and NSDN.

(Figure IX.59 C). The median diff in GFA in terms of absolute values for CSD, DNN and NSDN are 0.05, 0.09 and 0.05. The median value of MSE for CSD, DNN and NSDN are 0.0037, 0.0326 and 0.0011. The median ACC of CSD, DNN and NSDN are 0.28, 0.46 and 0.81. Non-parametric signed rank test for all

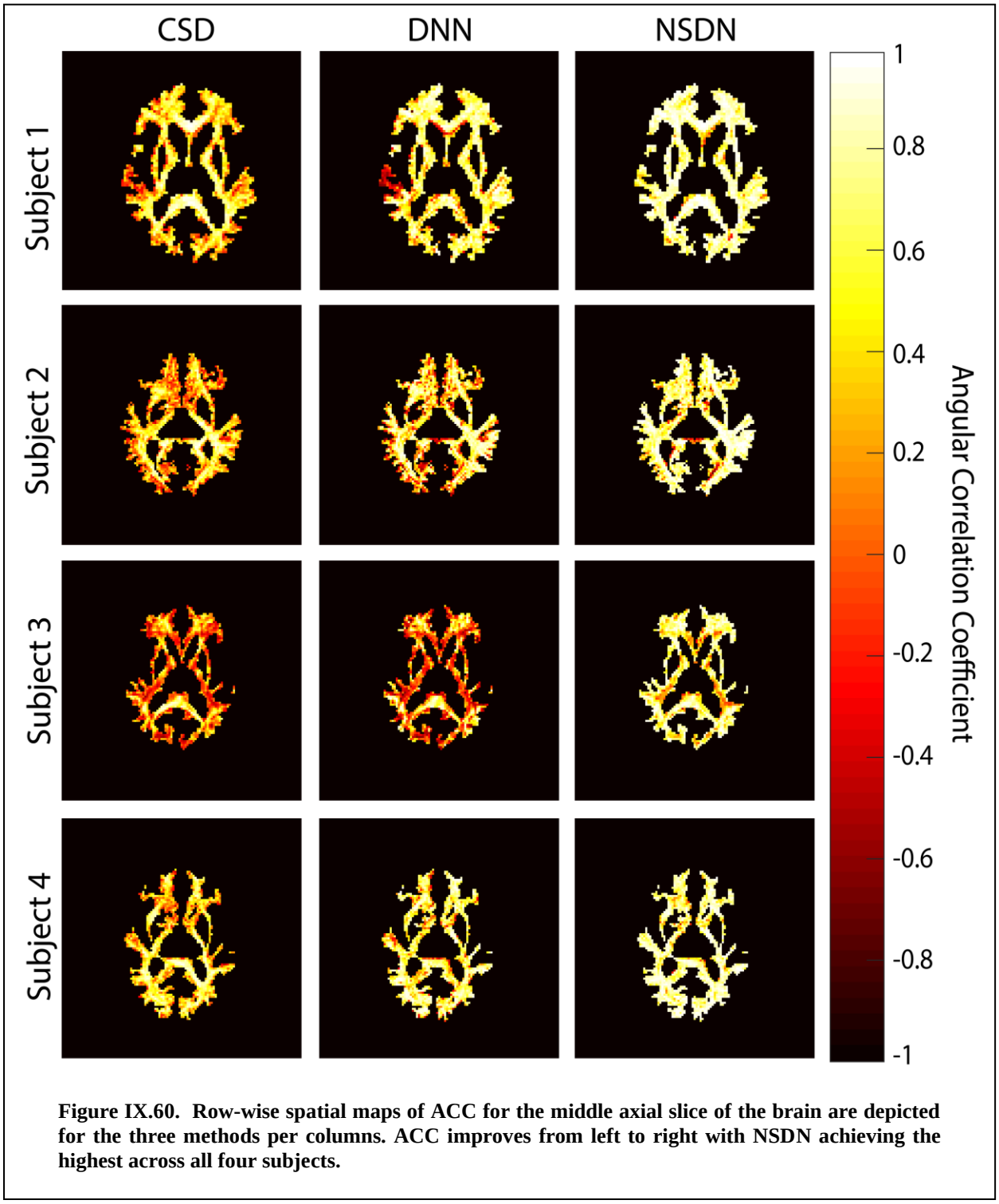

Figure IX.60. Row-wise spatial maps of ACC for the middle axial slice of the brain are depicted for the three methods per columns. ACC improves from left to right with NSDN achieving the highest across all four subjects.

pairs of distributions were found to be $p < 0.001$.

The mean ACC of CSD, DNN and NSDN are 0.28, 0.36 and 0.74 (Figure IX.60A, B & C orange and blue lines). CSD has the highest ACC variance while ACC variance decreases for DNN and is the most constrained for NSDN (Figure IX.60A, B, & C). The difference in GFA for pairs of WM voxels per subject has a similar constricted spread for CSD and NSDN while DNN has the widest variance (Figure IX.60 D, E & F). The distribution of MSE for all pairs of WM voxels per subjects shows that the DNN variance is the highest with CSD having quite a constrained distribution while NSDN having an even narrower one.

Figure IX.61 presents a qualitative comparison of spatial maps of ACC across four random subjects. Figure IX.62 examines the frontal lobe matter where we see low correlation for CSD, increased ACC with DNN, and the highest correlation with NSDN. Figure IX.63 presents a harmonized perspective of the subject presented in Figure IX.57.

9.4. Discussion

The NSDN architecture simultaneously (1) learns a mapping from ex vivo DW-MRI to a histology FOD, (2) ignores in-vivo scanner effects that occur between 1.5T and 3T, and (3) enforces consistency of GFA. This effort is a first attempt at direct harmonization of the estimated microstructure (FOD) through manipulation of data-driven tissue models. We believe this framework will be useful for harmonization of

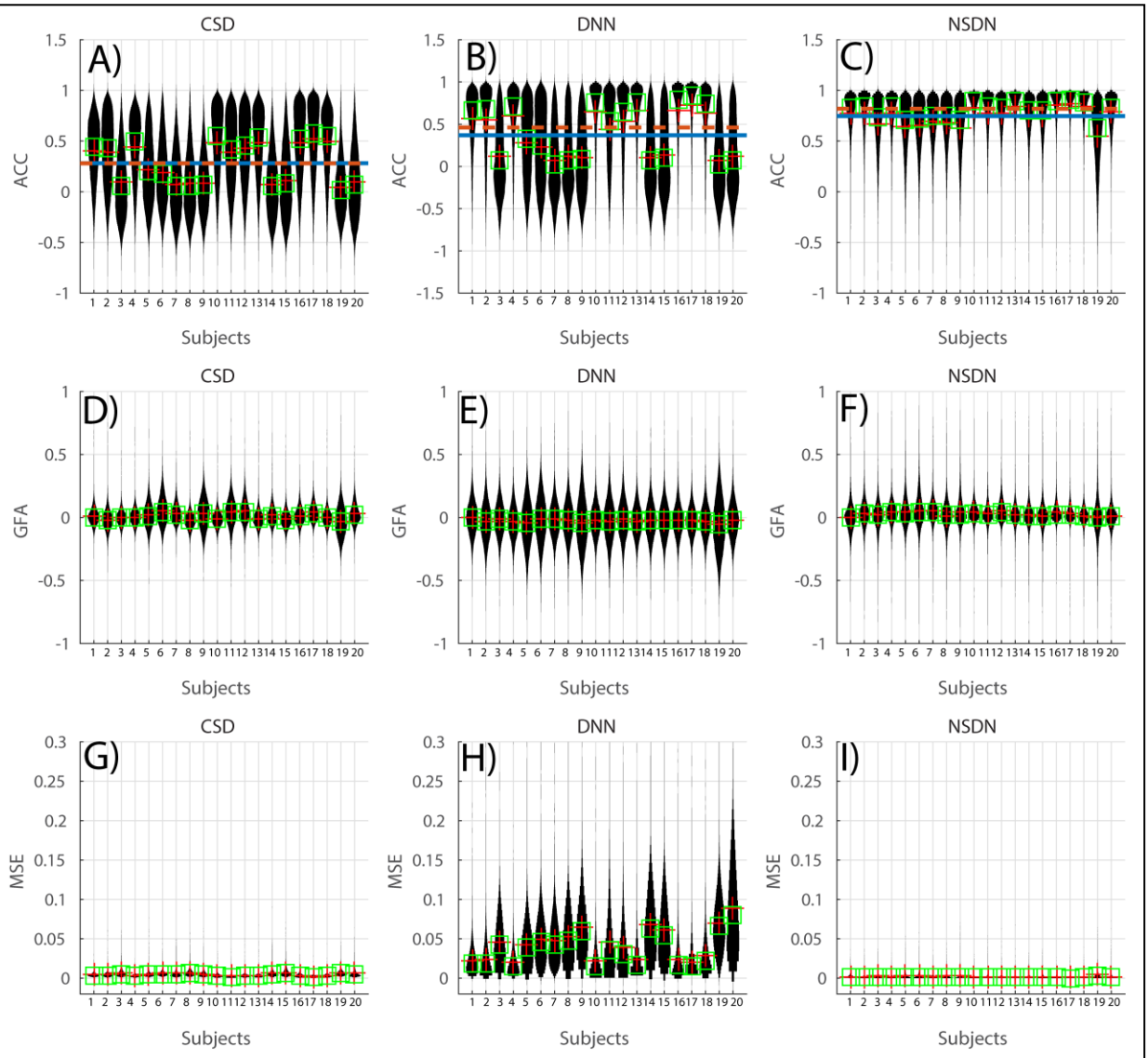

Figure IX.61. A, B and C) Depict the individual distribution per subject of ACC across WM voxels between 1.5T and 3T scanners. The blue and orange lines depict the mean and median of ACC across all 20 subjects. D, E and F) Depict the individual distribution per subject of difference in GFA across WM voxels between 1.5T and 3T scanners. G, H and I) Depict the individual distribution per subject of MSE across WM voxels between 1.5T and 3T scanners.

the microstructure in situations where scan-rescan data are available and have shown its efficacy in an extreme situation of non-contemporaneous 1.5T and 3T imaging. More broadly, the NSDN architecture could applied for other machine learning problems in different domains where there is paired unlabeled data.

Despite the success of the NSDN, there limitations of this study and additional validation is needed. First, the scan-rescan data were not contemporaneously acquired. Second, only two scanners / protocols

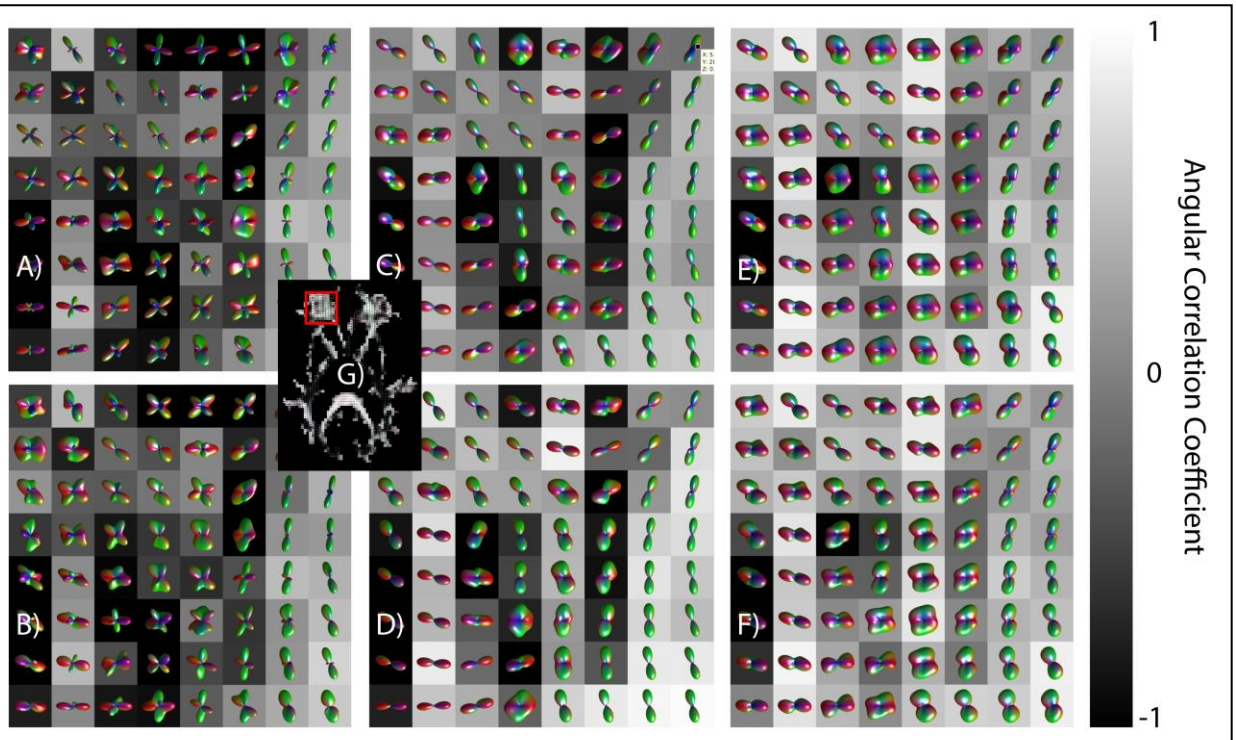

Figure IX.62. A & B) Depicts CSD FOD's with underlay of ACC on Scanner 1.5T and Scanner 3T data. C & D) Depicts DNN FOD's with underlay of ACC on Scanner 1.5T and Scanner 3T data. E & F) Depicts NSDN FOD's with underlay of ACC on 1.5T and 3T. G) The region of interest being observed is the left side frontal lobe WM.

were used, and these scanners were the same manufacturer. Third, subjects were all older than 67 years old. In terms of generalizability, NSDN has yet to be evaluated across scanners within a field strength, across b-values, or across gradient sample schemes.

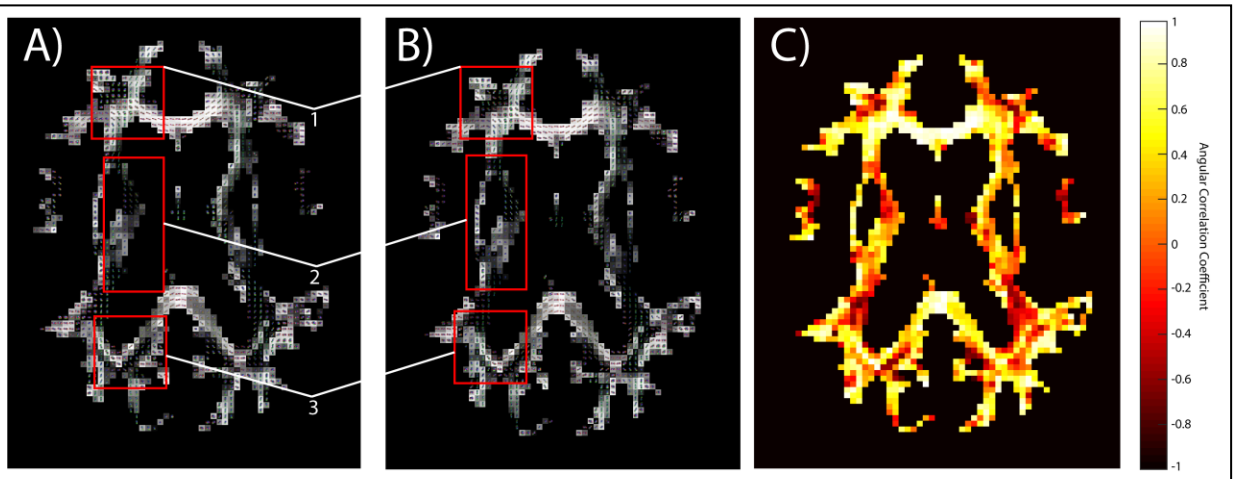

Figure IX.63. A) Middle axial slice on a 1.5T scanner depicting NSDN FOD's across the WM. B) Middle axial slice on a 3T scanner depicting NSDN FOD's across the WM. C) ACC of the FOD's which is calculated on corresponding voxels of 1.5T and 3T scanner. Annotations 1,2 and 3 depict higher correlation regions of the brain on the left side, same can be observed on the right side as compared to Figure 1.

10. Learning 3D White Matter Microstructure from 2D Histology

This chapter has been adapted from the published work in [189].

10.1. Introduction

Diffusion magnetic resonance imaging (dMRI) has proven a valuable tool in the neuroscience community due to its ability to infer tissue composition, microstructure, and structural connectivity of the brain [190, 191]. In the white matter, diffusion of water molecules is sensitive to the size, shape, and orientation of extra and intra-cellular tissue components [192], making it possible to infer the distribution of axonal fiber orientations in each dMRI voxel from a set of diffusion measurements, a model typically referred to as the fiber orientation distribution (FOD). These fiber orientation estimates can be used to reconstruct the structural connections between brain areas in a process known as fiber tractography [43, 45, 193]. However, tractography has been shown to have several fundamental limitations [194, 195], and validating the accuracy and reproducibility of these techniques is critical for them to become reliable medical and research tools.

Towards this end, the gold standard for validating dMRI (and a number of other contrasts) is in comparison to histology. For example, validation has been performed by (A) comparing tractography against histological tracers injected into the brain [44, 196, 197], (B) comparing estimated fiber orientations against myelin or axon stains [198-200], and (C) comparing diffusivity measures against cell or neuronal densities [201]. However, these studies have been limited to 2D analysis of tissue sections imaged with bright field microscopes. Very few studies have performed 3D validation [202, 203] due to limited field-of-view, time, and costs associated with 3D acquisitions. In contrast, 2D light microscopy is relatively inexpensive, can cover an entire tissue slice at high in-plane resolution, can be acquired on dozens to hundreds of sections of the same brain, and can be stained for a number of contrasts associated with tissue microstructure (for example, myelin, Nissl, tracers, etc.).

With this in mind, it would be of great interest to the neuroimaging community to be able to learn

the 3D tissue microstructure from 2D histology. Thus, in this study, we use ex vivo diffusion MRI of a squirrel monkey brain and corresponding myelin stained sections in combination with a convolutional neural network (CNN) to learn the relationship between the 3D diffusion estimated FOD and the 2D myelin stain. This network can be used to validate dMRI structural measurements in 3D. Additionally, this pre-trained network could be transferred to human stained sections to infer the 3D fiber distribution at resolutions currently unachievable with dMRI, allowing fiber tractography at unprecedented resolutions. We envision the use of similar networks to learn other 3D microstructural measures from an array of potential common 2D brightfield contrasts.

10.2. Methods

Here, we aim to use 2D myelin-stained micrographs, from which traditionally only 2D orientation information is extracted, in order to estimate the 3D fiber structure in these locations (Figure X.64).

10.2.1. Data Acquisition – MRI

MRI experiments, histological methods, and data registration were performed on an ex vivo squirrel monkey brain following the procedures described in [127, 204]. Briefly, ex vivo imaging was performed a Varian 9.4 T magnet, with diffusion weighted scans acquired using a PGSE multi-shot spin-warp imaging sequence (TR = 4.6 s, TE = 42 ms, 32 gradient directions, b ≈ 1000s/mm^2, 300 μm voxel, 192x128x115 matrix). Diffusion processing was performed in "histology" space after registration (see below) using constrained spherical deconvolution [205] for voxel-wise reconstruction, resulting in diffusion FODs reconstructed in histology space.

10.2.2. Data Acquisition – Histology

Following scanning, the brain was frozen and cut serially on a microtome in the coronal plane at 50 um thickness. Every 6th section (150um) was mounted and stained with a gallyas silver stain [206] to identify myelinated axons. Whole-slide brightfield microscopy was performed using a Leica SCN400 Slide Scanner at 20× magnification, resulting in a maximum in-plane resolution of 0.5 um/pixel. To ease computation and memory burdens images were down-sampled to 2 um/pixel to serve as input to the

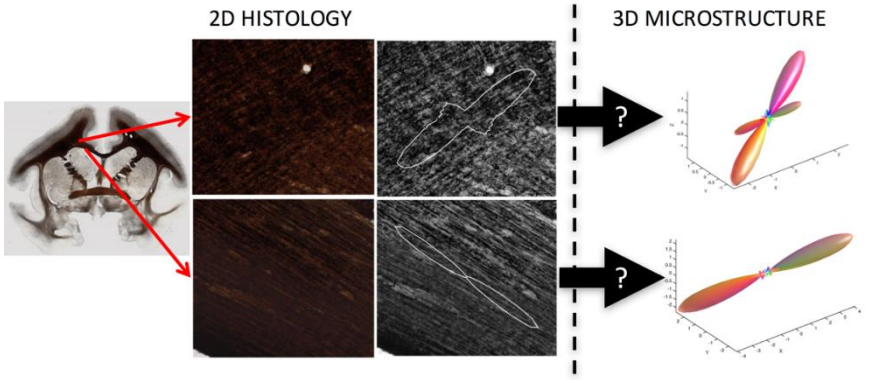

Figure X.64. Conventional micrographs are inherently 2D representations of the underlying tissue. Here, we aim to use Brightfield microscopy of myelin-stained tissue to estimate the 3D fiber orientation distribution

network. Finally, a multi-step registration procedure was utilized [207] that involves 2D affine and 2D non-rigid transformations to an intermediate "frozen tissue block space", followed by 3D affine and 3D non-rigid transformations to MRI data. Deformation fields produced by registration steps allow transfer of any set of data to any desired space for comparisons [204], in this study, we chose to process dMRI data in histology space, resulting in high resolution myelin images aligned with corresponding dMRI derived FODs.

10.2.3. Deep Learning Network

We use a CNN architecture that consists of four 2D convolutional layers (Figure X.65). The input is a "patch" of the high-resolution myelin image of size 256x256 pixels. The patch has a corresponding dMRI FOD, represented using spherical harmonic (SH) basis functions, in this case using 6^{th} order SH of

size 1x28. The C1, C2, C3, C4 layers consist 128x128, 64x64, 32x32 and 16x16 feature maps. For all 2D convolutional layers the kernel size was set to (3, 3) and strides at (2, 2). All 2D convolutional layers were activated by 'Relu' followed by batch normalization and 2D max pooling was used except for C1 where only max pooling was used. The C4 layer was flattened and a dropout of 0.5 was used while connecting to a fully connected layer of 28 neurons which is equivalent to the output of the network a set of 6^{th} order SH coefficients.

Ten myelin-stained micrographs were utilized in this study. Nine slices were used for the training and the validation of the CNN and a single slice was completely withheld for testing of the CNN. From these images, a total of 248,304 patches were generated with corresponding outputs. 30,000 of these were randomly selected for validation of the network to prevent over-fitting of the network. This leads to the training data being set at 218,304 patches. The CNN was trained with the loss function of 'mean squared error'. The optimizer of the network was used as 'RMSProp'. Batch size was set to 1000 as feeding more patches were not possible due to limitation of GPU memory. The network was trained for 50 epochs after which it attained convergence as per cross-validation criteria. A Nvidia 1080 Ti GPU was used to train the network.

10.3. Results

Figure X.66 shows the results of applying this network to the unseen myelin-stained section. The CNN iterated through 50 epochs, minimizing the mean squared error (MSE) of SH coefficients. MSE for the unseen slice (Figure X.66 A) are on the order of that seen during training. The converged network attained MSE of 0.025, which is close to the median shown for the hidden slice. Additionally, the estimated 3D FODs show a range of agreement with ground truth, with ACCs ranging from negative to nearly perfect (Figure X.66 B), with a median value of 0.48, indicating moderate predication ability. Example myelin-

stained patches, the ground truth FODs, and predicted FODs are shown for varying levels of reconstruction accuracy (Figure X.66 C-F).

Figure X.67 visualizes the predicted and ground truth FODs reconstructed across the full unseen slice. Visualizing across a larger FOV shows spatial consistency in predictions, in agreement with expected

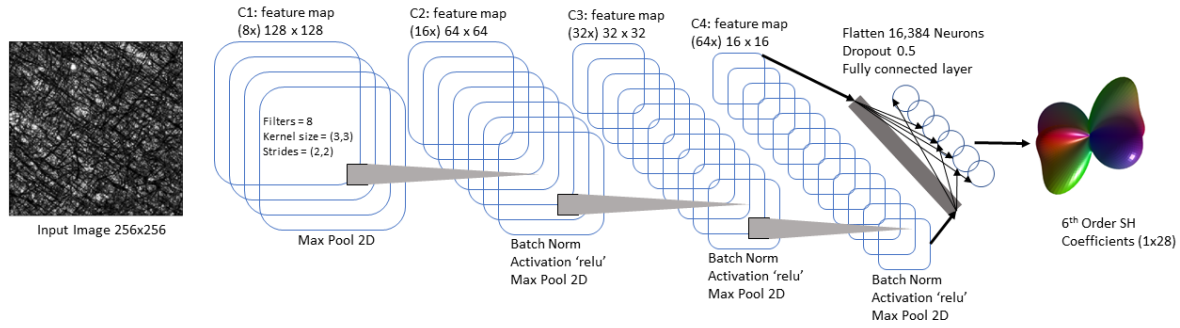

Figure X.65. This architecture begins with a single input patch of size 256x256x with a corresponding output of 6th order SH coefficients (1x28). The network consists of four convolutional layers with subsequent down-sampling and flattened towards the end to a fully connected dense layer. Relu activation, batch normalization and max pooling have been used for all convolutional layers.

anatomy and dMRI FODs. This suggests the possibility of utilizing this approach for fiber tractography on 2D images. While many regions show high prediction power (Figure X.67, left), this technique can result in largely inaccurate FOD reconstructions in others (Figure X.67, right).

10.4. Discussion

Recently, deep learning approaches have proven valuable in the field of diffusion MRI, demonstrating the ability to accurately predict FODs from the diffusion signal [208], and harmonizing signal across scanners [209, 210], lending insight into the relationship between the signal and tissue microstructure. In this study, we have used a CNN to estimate 3D fiber distributions from inherently 2D micrographs. To the best of our knowledge, this study is the first attempt to extract 3D features from 2D histology.

Although results varied across the brain, many predicted fiber geometries showed moderate to high agreement with the ground truth distributions. This analysis and network structure could potentially be used

to estimate other 3D tissue microstructure features from the large number of available 2D histological resources, atlases, and databases. For example, it may be possible to estimate cell or neuron densities, diameters, segment lengths, orientations, and undulations, among other microstructural features that may be clinical indicators of pathologies or diseases. This would be critical to diffusion MRI validation studies, in addition to a variety of other MRI contrasts.

Alternatively, this methodology could be used to reconstruct the 3D structural connections of a human brain, at spatial resolutions surpassing those feasible with current clinical and pre-clinical scanners. For example, an existing myelin-stained atlas of the human brain could be utilized, and processed using the current methodology, resulting in 3D fiber distributions at ~500um isotropic resolution, characterizing volumes much smaller than the typical 2.5mm resolution. Smaller voxels may alleviate partial volume affects and the crossing fiber problem [106, 211, 212], potentially improving the accuracy of tractography and improving our understanding of the human connectome.

There are several potential improvements to the current study, as well as future areas of exploration. The next step will be to include additional slices from the same monkey, as well as incorporate, test, or train on multiple subjects from existing atlases or databases [127, 204]. Successful generalization to new subjects would lend significant confidence in applying this to unseen brains or new species, for example the human brain. Finally, we recognize that the low ACC in many regions of the brain may be due to ambiguities associated with determining whether a fiber is going-into or coming-out-of the 2D plane, both of which may intuitively look similar in a 2D projection image. Implementing a loss function that is invariant to the through-plane orientation may lead to better reconstructions, although the network would not be able to identify whether the axon distribution is oriented into or out of plane. Successful implementation of this network would require some form of post-processing of FODs to ensure spatial

continuity. This could possibly be solved by “flipping” certain FODs in the through-plane direction so that they were consistent with their neighbors in a way similar to 2D phase unwrapping.

10.5. Conclusion

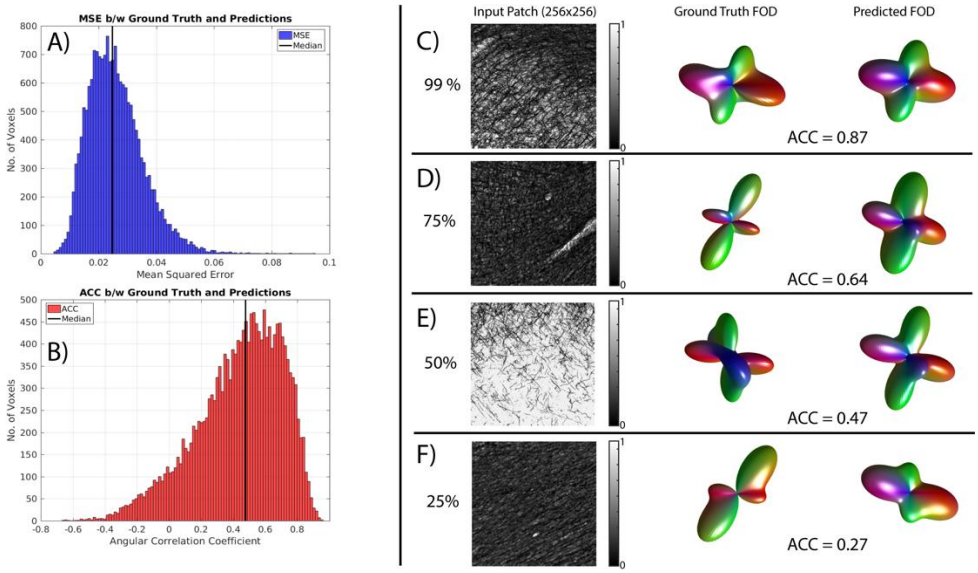

Figure X.66. The CNN minimizes the MSE of SH coefficients between ground truth and predictions (A), resulting in ACC values with median correlations of 0.49. Example myelin-stained patches, ground truth FODs and predicted FODs are shown for varying levels of reconstruction accuracy (C-F).

We have implemented a deep learning approach in order to extract 3D microstructural measures from inherently 2D microscopy images. Specifically, we trained a CNN to estimate the 3D fiber distribution from myelin-stain brightfield micrographs, with moderate to high accuracy throughout most of the brain. Spatial coherence suggests that this technique could be performed on consecutive 2D slices to perform 3D fiber tractography, potentially at resolutions much higher than that possible with current dMRI practices, and possibly on new unseen brains or specimens – for example the human brain. There is potential to use this, and similar, techniques to estimate a number of 3D metrics from 2D histological contrasts.

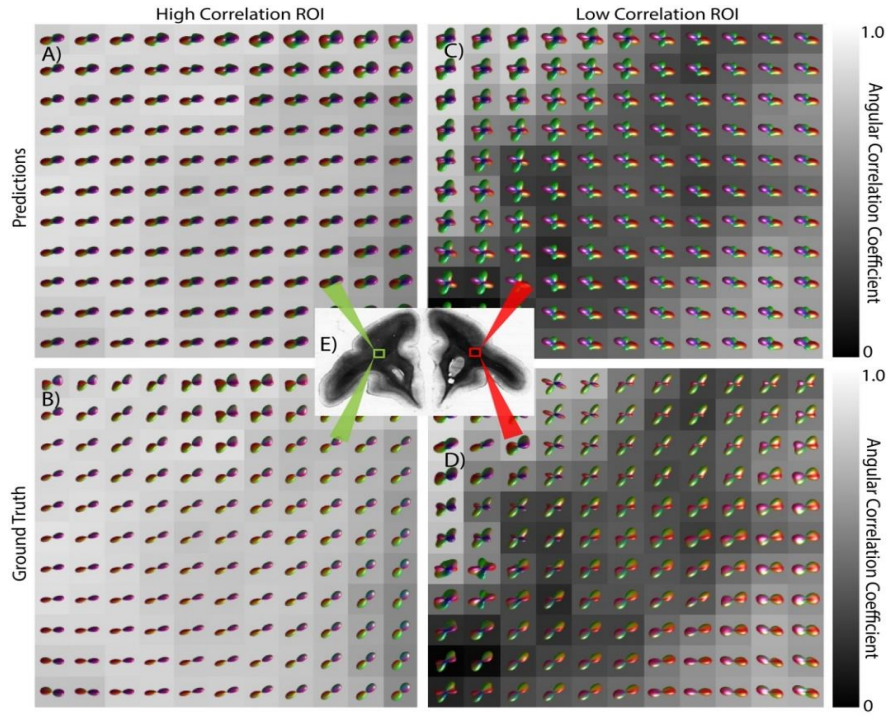

Figure X.67. Predicted FODs (top) from 2D histology are shown along with the ground truth fiber geometries (bottom) for two regions. One exhibiting high agreement (left) with the ground truth, the other with a lower agreement (right).

11. Deep Learning 3D White Matter Fiber Orientation from 2D Histology: Pulling 3D Rabbits Out of 2D Hats

This chapter has been adapted from the published work in [189].

11.1. Introduction

The gold standard for most measures of brain tissue microstructure, composition, pathology, and connectivity is typically histological analysis. For example, diffusion MRI validation of tractography and microstructure has been performed through comparisons against tracers [213, 214], against myelin or axons stains [199, 200], and against neuron stains [201]. However, these studies have been limited to 2-dimensional (2D) analysis of tissue sections with Brightfield microscopy. Very few studies have performed 3-dimensional validation due to limited time, fields-of-view, and costs associated with 3D acquisition and hardware. In contrast 2D light microscopy is inexpensive, can cover an entire tissue slice at high in-plane resolution, can be performed on dozens of sections of the brain, and can be stained with a number of contrasts to elucidate tissue microstructure. Thus, it would be of great interest to be able to learn 3D tissue structure from inherently 2D microscopy (Figure XI.68). In this study, we use ex vivo diffusion MRI of a squirrel monkey brain and corresponding myelin stained sections, in combination with a convolutional neural network (CNN) to learn the relationship between 3D diffusion estimated fiber orientation distributions (FOD) and the 2D myelin stain.

11.2. Methods

Two ex-vivo squirrel monkey brains were scanned, sectioned, and stained for this study. Diffusion MRI was performed (b=1000s/mm^2, 300μm isotropic resolution, 32 directions) on a 9.4T scanner, and FODs reconstructed using spherical deconvolution. Each brain was cut serially on a microtome, stained with Gallyas silver stain, and imaged on a whole-slide Brightfield microscope at 0.5um/pixel. Histology and MRI were registered following the procedures utilized in [207]. Thus, the input to the CNN is a 256x256 pixels "patch" of the high-resolution myelin image and a corresponding diffusion FOD, represented using 6th order spherical harmonic (SH) coefficients as a 28x1 vector of coefficients – the CNN then aims to estimate the FOD from a given patch. For learning, we utilized a CNN architecture with four 2D convolutional layers (Figure XI.69) and a fully connected dense layer.

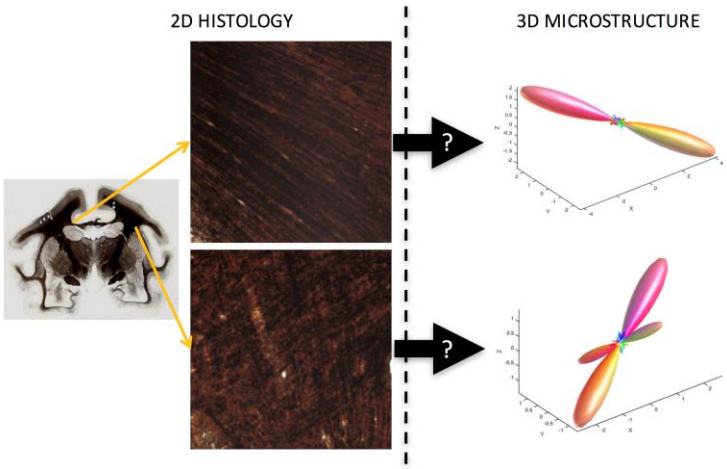

Figure XI.68. Conventional micrographs are inherently 2D representations of the underlying tissue. Here, we aim to use Brightfield microscopy of myelin-stained tissue to estimate the 3D fiber orientation distribution

One monkey was used for training/validation and the other for testing. For training, 9 myelin-stained micrographs from the first monkey were utilized, resulting in 248,304 patches (and corresponding FODs). For validation, a 10th slice from the same monkey brain was used to compare diffusion FODs and predicted FODs using the angular correlation coefficient (ACC) which describes the overall agreement between spherical functions. For testing, and to evaluate generalizability to a new brain, the network was

applied to two histological slices from the second monkey brain.

11.3. Results

Figure XI.70 shows the ACC results on the unseen validation slice from the same brain that training was performed on. The ACC is moderate to high in most white and gray matter regions, with a median value of 0.48 (for reference, a scan-rescan using spherical deconvolution has shown an ACC of 0.67 [88]).

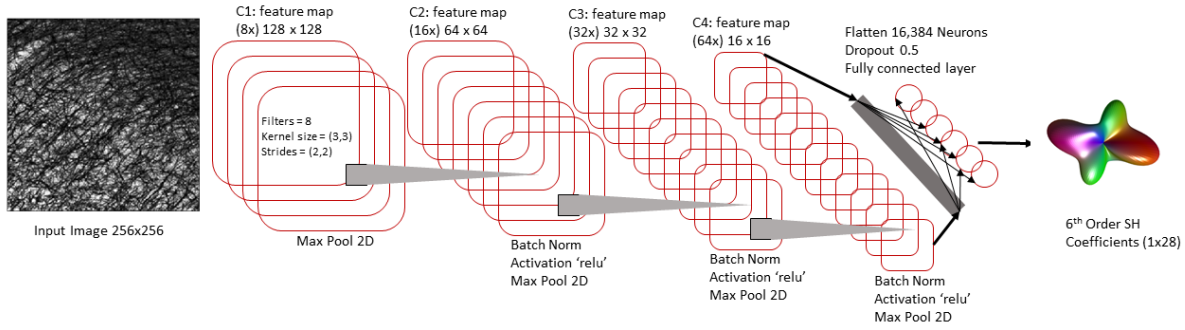

Figure XI.69. This architecture begins with a single input patch of size 256x256x with a corresponding output of 6th order SH coefficients (1x28). The network consists of four convolutional layers with subsequent down-sampling and flattened towards the end to a fully connected dense layer. Relu activation, batch normalization and max pooling have been used for all convolutional layers.

Comparisons of the true (diffusion MRI) FODs and predicted FODs show general agreement in shape and orientation, and importantly, the ability to predict through-plane orientation.

Figure XI.71 and Figure XI.72 show slices from the withheld second squirrel monkey brain, where 3D FODs were predicted from myelin-data alone. The FODs generally match expected fiber orientations and show the ability to extract through-plane orientation information.

11.4. Discussion

In this study, we have used a CNN to estimate 3D fiber distributions from inherently 2D

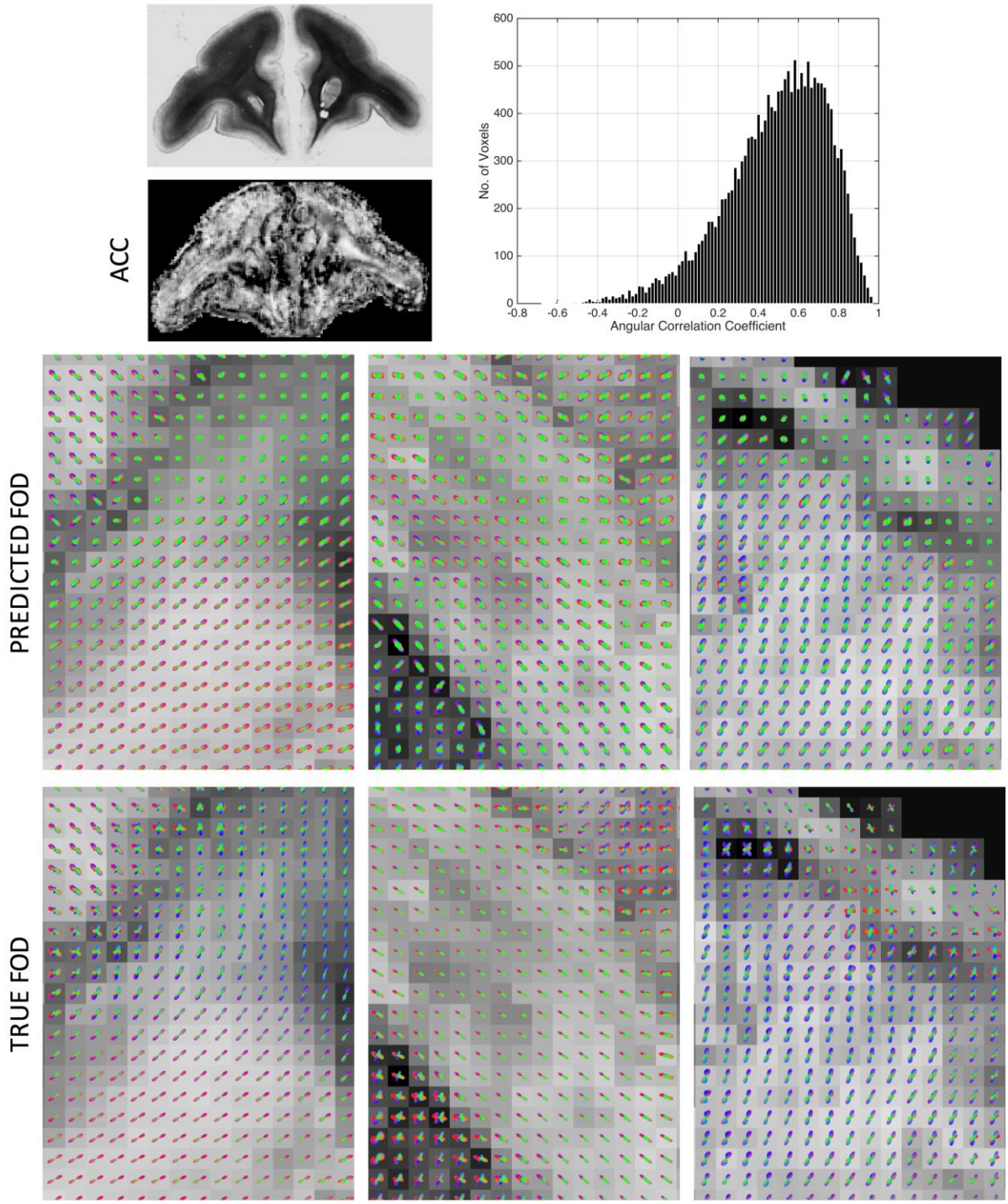

Figure XI.70. The CNN architecture can predict 3D FODs with moderate accuracy. The ACC map of the unseen slice is shown, with median ACC of 0.48. Predicted and True (diffusion MRI) FODs are shown in 3 regions of interest, where background color indicates ACC (brighter signifies greater agreement).

micrographs. We have shown that predicted FODs are in moderate to high agreement with FODs estimated

using diffusion MRI, the network can predict through-plane orientations, and that it can be generalized to stains of different brains. This analysis and network structure could potentially be used to estimate other 3D tissue features from the large number of available 2D histological atlases and databases. For example,

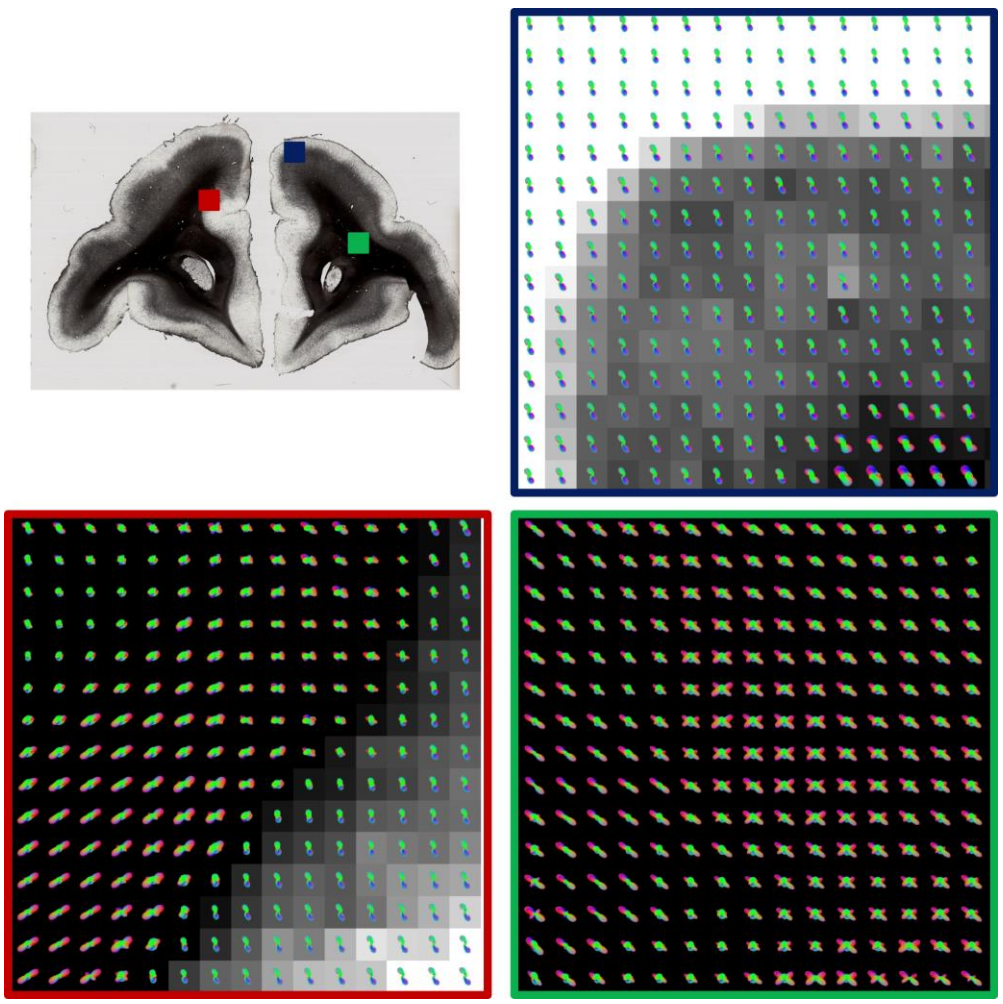

Figure XI.71. The network can predict 3D FODs from 2D histology of a different brain. Predicted FODs from slices on a different brain are shown zoomed-in for 3 white and gray matter regions (background image intensity is the myelin stain in the region).

3D estimates of cell and neuron diameters, lengths, orientations, undulations, and densities could be relevant for validating various MR contrasts.

One potential application of this methodology is for "myelin-stained" fiber tractography. Applying this network to a series of consecutive slices could be used to perform 3D tractography on brains that have not had a diffusion MRI scan. Future research should explore feasibility of myelin-tractography using

existing open-source atlases [204]. Alternatively, this methodology could be generalized to human myelin stains, and used to reconstruct the 3D structural connections of the human brain at spatial resolutions surpassing those feasible with current state-of-the-art acquisitions (<500 um histological resolution versus 2.5mm diffusion MRI), possibly alleviating partial volume and crossing fiber effects.

Future improvements to increase accuracy should include using additional slices for training, as well as training using multiple monkeys to increase robustness to variation in staining intensities.

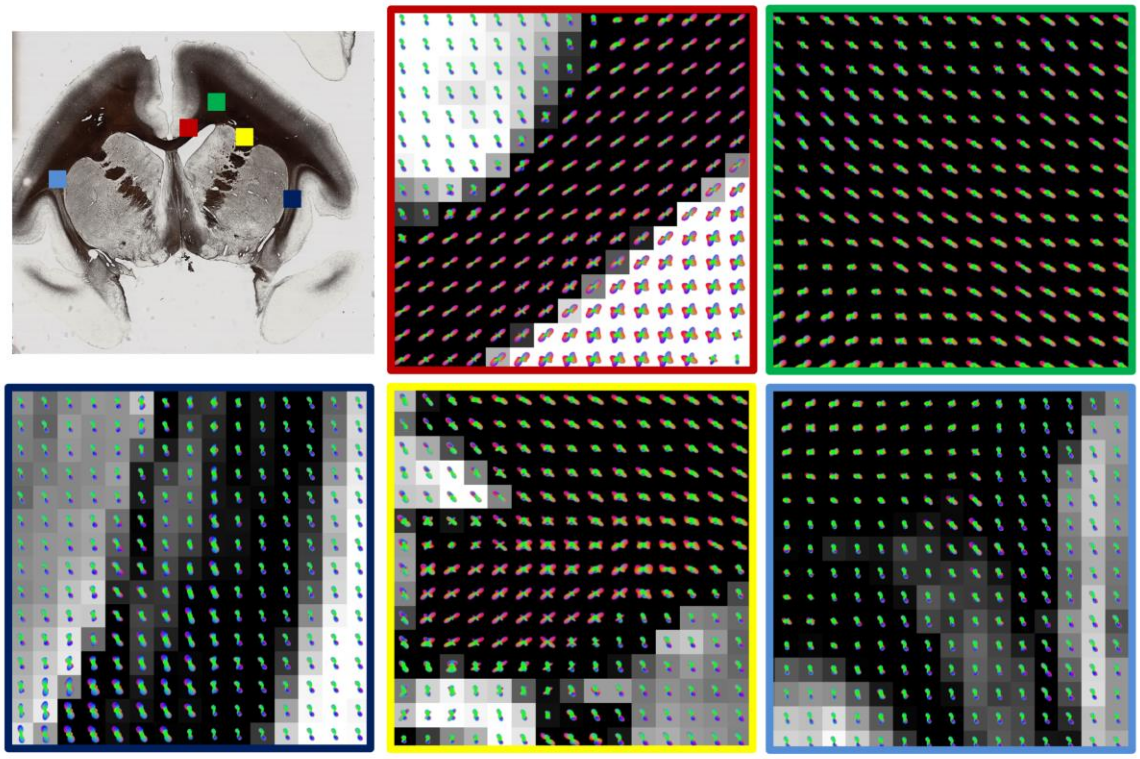

Figure XI.72. The network can predict 3D FODs from 2D histology of a different brain. Predicted FODs from slices on a different brain are shown zoomed-in for 5 white and gray matter regions.

12. Enabling Multi-Shell b-Value Generalizability of Data-Driven Diffusion Models with Deep SHORE

This chapter has been adapted from the published work in [26].

12.1. Introduction

Diffusion-weighted magnetic resonance imaging (DW-MRI) is essential for non-invasive reconstruction of the microstructure for the human in-vivo brain. These images are sensitized to the underlying organization of the tissue at a millimetric scale. Multiple approaches have been proposed that can model the non-linear relationship between the DW-MRI signal and biological microstructure with the most common being diffusion tensor imaging (DTI) [35]. Substantial efforts have shown that other advanced approaches can recover more elaborate reconstruction of the microstructure and these methods are collectively referred to as high angular resolution diffusion imaging (HARDI) [18]. HARDI methods have been broadly proposed in two categories of single shell acquisitions and multi-shell acquisitions (i.e., using multiple diffusivity values). A majority of single shell HARDI methods utilize spherical harmonics (SH) based modelling as in q-ball imaging (QBI) [13], super-resolved constrained deconvolution (sCSD) [175], and many others. However, SH based modelling cannot directly leverage additional information provided by multi-shell acquisitions. SH have been combined with other bases to represent multi-shell data, e.g., solid harmonics [31], simple harmonic oscillator reconstruction (SHORE) [37], and spherical polar Fourier imaging [215].

Methodological exploration has been driven through classical mathematical transforms while data-driven approaches have been limited (Figure XII.73) due to lack of external validation data. Prior work using data-driven approaches for DW-MRI has been shown in [216], however the primary application for their work is shown for outlier detection and low rank signal prediction. Validation through histology is critical to evaluate the precision of white matter (WM) reconstruction [23]. Prior work through machine learning approaches on reconstruction for single shell diffusion acquisitions has exhibited higher precision

and reproducibility. However, this has not been shown for multi-shell acquisitions due to lack of external validation data [187] (Figure XII.73).

To overcome these issues, we propose a novel approach, Deep SHORE, which incorporates the following key contributions: (1) an unsupervised hyper-parameter optimization for improved learning in the SHORE manifold, (2) representation of a microstructure model in the SHORE manifold to improve precision and reproducibility, and (3) a non-negativity constraint implementation for a deep learning model.

12.2. Data Acquisition

Three *ex-vivo* squirrel monkey brains were imaged on a Varian 9.4T scanner. A total of 100 gradient volumes were acquired using a diffusion-weighted EPI sequence at diffusivity values of 3000, 6000, 9000 and 12000 s/mm^2 at an isotropic resolution of 0.3mm. An observation is that approximation of b-values for the *ex-vivo* acquisition is equivalent to *in-vivo* b-values of 1000, 2000, 3000 and 4000 s/mm^2 [217]. After acquisition, the tissue was sectioned and stained with fluorescent dil and imaged on a LSM710 confocal microscope following procedures outlined in [23]. The histological fiber orientation distribution (HFOD)

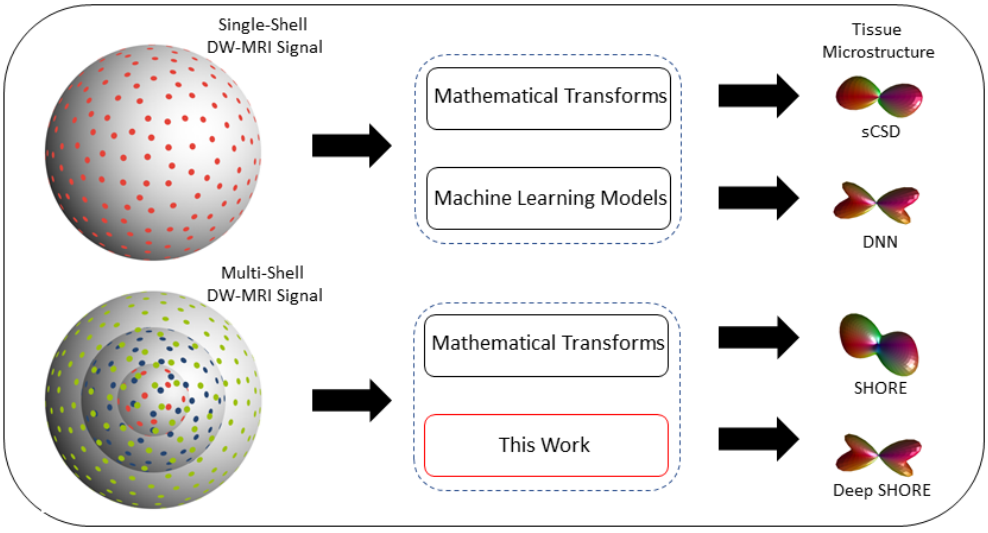

Figure XII.73. Different classes of methods have been used to infer tissue microstructure from single shell and multi shell DW-MRI data. The gap addressed herein is in data-driven machine learning models for multi-shell DW-MRI data.

was extracted using 3D structure tensor analysis. A multi-step registration procedure was used to determine

the corresponding diffusion MRI signal [23]. A similar procedure is outlined in [176]. A total of 567 histological voxels were processed. A hundred random rotations were applied to the remaining voxels for both the MR signal and the HFOD to augment the data, bringing the total to 57,267 voxels. As a limitation we acknowledge that there is a possibility of registration error approximately up to the size of MR voxels (up to 300 micrometers) [218].

The *in-vivo* acquisitions of the three human subjects were acquired on two sites 'A' and 'B'. Both sites were equipped with a 3T scanner with a 32-channel head coil. Structural T1 MPRAGE was acquired for all subjects on both the sites. The diffusion acquisition protocol and scanner information are listed on each site as follows:

Site 'A': The scan was acquired at a diffusivity values of 1000, 1500, 2000, 2500, 3000 s/mm^2. A total of 96 diffusion weighted gradient volumes were acquired per diffusivity value with a 'b0'. Briefly the other parameters are: SENSE=2.5, partial Fourier=0.77, FOV=96x96, Slice=48, isotropic resolution: 2.5mm.

Site 'B': All parameters of scan acquisition were same as that of the scanner at site 'A' except for the isotropic resolution which was 1.9x1.9x2.5mm^3 and down-sampled to 2.5mm iso.

The *in-vivo* acquisitions were pre-processed with standard procedures of eddy, topup and b0 normalization followed by pairwise registration per subject [99]. T1s were registered and transformed to the diffusion space. Brain extraction tool was used for skull stripping [177]. WM segmentation was performed using T1 for *in-vivo* data [100].

12.3. Methods

The SHORE basis function has been shown to capture the representation of multi-shell DW-MRI with minimal representation error [37] and ensure the same when modelling single shell DW-MRI. The DW-MRI normalized signal, *E(q)*, can be represented as:

$$\boldsymbol{E(q) = \sum_{n=0}^{N} \sum_{l=0}^{n} \sum_{m=-l}^{l} c_{nlm} G_{nl}(q,\zeta) Y_l^m(u)} \qquad \textbf{Eq. 28}$$

where *c* are the coefficients to be estimated, *G* depicts the radial basis combined with *Y*, the SH basis. The

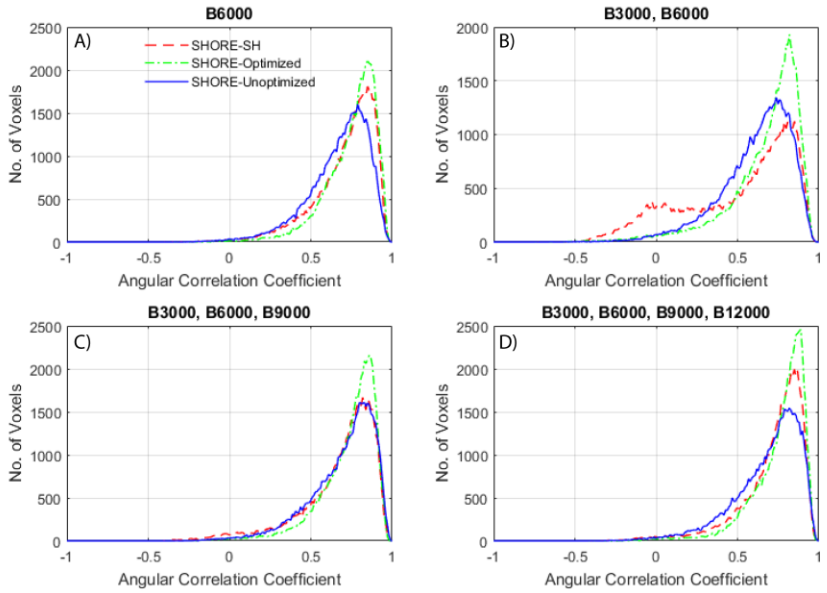

Figure XII.74. ACC of paired voxels between HFOD and predictions of the deep learning methods across three different manifolds (1) Optimized SHORE -> SH (2) Optimized SHORE -> SHORE (3) Unoptimized SHORE -> SHORE across all 57,267 voxels. Predictions across A) single shell of b-value 6000 s/mm^2. B) Two shells of 3000 and 6000 s/mm^2. C) Three shells of 3000, 6000 and 9000 s/mm^2. D) Four shells of 3000 -12000 s/mm^2.

radial basis *G* is represented as follows:

$$\boldsymbol{G_{nl}(q,\zeta) = \kappa_{nl}(\zeta) \left(\frac{q^2}{\zeta}\right)^{\frac{l}{2}} \exp\left(-\frac{q^2}{2\zeta}\right) L_{n-\frac{l}{2}}^{l+\frac{1}{2}}\left(\frac{q^2}{\zeta}\right)} \qquad \textbf{Eq. 29}$$

where ζ is the scale parameter, *q* is the radius of the diffusivity value, and *L* depicts the associated Laguerre polynomial. Eq (2) can be optimized using the BFGS [219] algorithm by iterative refitting of the coefficients *c*. BFGS is well-known for solving unconstrained non-linear optimization problems. The novelty that we introduce here is that, when functioning across several normalized datasets, an optimal ζ per dataset, will lead to learning on an optimized manifold. Additional parameters of SHORE include: radial

order: 6, and regularization constants: 1e-8 [37]. SHORE estimates 50 coefficients at 6^{th} order. Regularized linear least squares were used for the estimation of the coefficients. As one notes, l is even for diffusion spherical harmonics or SHORE due to symmetry of diffusion inference process. Essentially, SH at 8^{th} order and SHORE at 6^{th} order offer the same degree of freedom. SHORE at higher orders is known to suffer from overfitting effects [220].

The HFOD represented as SH coefficients can be fitted to the SHORE basis with the two considerations of (1) diffusivity value and (2) 'ζ' scaling parameter. First the SH coefficients from an 8^{th} order were sampled over a sphere of 100 gradient directions. The directions were ensured to be uniformly sampled on the sphere with minimized electrostatic repulsion. These directions were kept consistent at all times while predicting from the network as well. The diffusivity value was set to 2000 s/mm^2. After which, it was fitted to SHORE basis using the process described above.

Non-Negativity. The FOD, when modelled as SH, cannot exist with negative mean. If it does, then the microstructure exists in the imaginary part of the SH which does hold true when modelling with real even ordered SH. Hence, there was an existing gap to enforce non-negativity on a deep learning network while training and making predictions. We use a regularization value of 0.005 to truncate all values on a set of gradient directions where value is <=0 on both sides of input and output. Thereafter, log space is used instead of linear space: *ln(E(q))* and *ln(P(r))*, where E(q) is the normalized signal and P(r) is the FOD sampled over the gradient directions. Fitting of the representation method such as SH or SHORE follows after log transformation. After the predictions are made, the coefficients of a representation are transformed using exponential to recover them back to linear space.

Deep Network Design. We use a 5-layered deep network with the following number of neurons: x1:400, x2:45, x3:200, x4:45 and x5:200. A residual block was created for the layers x2, x3 and x4 and hence the number of neurons was kept equal for x2 and x4. All layers were activated with 'elu'. Additional parameters of the network: Loss function: mean squared error, batch size: 1000, optimizer: RMSProp. While training, only the input and output coefficients were modified for different subcases (discussed in next section). This was due to the fact that SHORE at 6^{th} radial order is defined by 50 coefficients and SH

at 8th order is defined by 45 coefficients. For training of the network, we used k-fold cross validation where k=5 for optimal training.

To evaluate on any withheld set of data we used the angular correlation coefficient (ACC) [15], which is a measure on a scale of -1 to 1 where 1 is the best correlation. ACC is defined using two sets of SH coefficients 'u' and 'v':

$$\boldsymbol{ACC} = \frac{\sum_{j=1}^{\propto}\sum_{m=-j}^{j} u_{jm}v_{jm}^{*}}{\left[\sum_{j=1}^{\propto}\sum_{m=-j}^{j}|u_{jm}|^{2}\right]^{0.5}\cdot\left[\sum_{j=1}^{\propto}\sum_{m=-j}^{j}|v_{jm}|^{2}\right]^{0.5}} \quad \textbf{Eq. 30}$$

Evaluation Strategies. From the total of 57,267, we create 8 testing sets of data where each set has 7,272 voxels except for the last one which has 6,363 voxels. The remaining data for each set were used as training. For all cross-validation experiments, blocks of 101 voxels were randomly allocated to testing/training cohorts ensuring that no synthetic rotations of training data were included in the testing phase. While fitting SHORE coefficients the diffusivity shell of 6000 s/mm^2 was withheld leading to four cases of evaluation in incrementing order of shells. For evaluation purposes, we used three sub-cases of deep learning manifolds 1.) Input of 'ζ' optimized SHORE DW-MRI and output SH-HFOD. 2) Input of unoptimized 'ζ' SHORE-DWMRI and output of SHORE-HFOD 3.) Input of optimized 'ζ' SHORE-DWMRI and output of SHORE-HFOD.

Furthermore, we make comparisons between single-shell and multi-shell approaches. For single shell, we show the comparison between the leading single shell approach sCSD, a prior proposed approach that utilizes deep learning [25, 187] (SHDNN) and SHORE derived FOD on the withheld shell and the same for multi-shell where sCSD and SHDNN were excluded.

For *in-vivo* reproducibility evaluation, we compare the ACC for all the pairs of WM voxels between the two sites 'A' and 'B' on a per subject basis. For SHORE based approaches, all the five shells of data were used while for sCSD we used b-value of 2000 s/mm^2 as the highest reproducibility was exhibited on the specific shell.

12.4. Results

Evaluation across the withheld shell using combinations with other shells shows that when learning

across the optimized SHORE manifold the ACC is most skewed towards higher correlation as compared to the other two approaches (Figure XII.74). The median of all distributions for each method is presented in Table XII.6. The median for optimized SHORE learning is the highest. We found significant improvements after non-parametric signed rank test for all pairs of distributions ($p << 0.001$, Wilcoxon signed rank test).

Table XII.6 Median & mean values of 4 dataset combinations for the deep learning approaches.

Deep Learning Approaches	One Shell		Two Shell		Three Shell		Four Shell	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
U SHORE → SHORE HFOD	0.70	0.67	0.65	0.61	0.74	0.69	0.73	0.68
O SHORE → SH HFOD	0.75	0.71	0.61	0.50	0.74	0.67	0.77	0.72
O SHORE → SHORE HFOD	**0.78**	**0.75**	**0.73**	**0.67**	**0.77**	**0.73**	**0.80**	**0.76**

U-Unoptimized, O-Optimized. All methods were intercompared per shell combination, using Wilcoxon signed rank test and corrected using Bonferroni correction. All combinations were found to be significant.

When comparing predictions across single shell methods (Figure XII.75 A), the trend in the following increasing order of correlation (median in parenthesis): sCSD (0.73), SHORE-FOD (0.74), SHDNN (0.76), NNSHORE-DL (0.77), SHORE-DL (0.78). Similarly, when making multi-shell comparisons (Figure XII.75 B) we can observe increasing order of correlation: SHORE-FOD (0.75), NNSHORE-DL (0.79) and SHORE-DL (0.80). Non-parametric signed rank test for all pairs of distributions were found to be $p << 0.001$.

Observing the distribution of ACC across all pairs of WM voxels per subject (Figure XII.76), increasing level of reproducibility from sCSD, SHORE-DL, SHORE-FOD and NNSHORE-DL. Non-parametric signed rank test for all pairs of distributions were found to be $p << 0.001$ (Table XII.7). The NNSHORE-DL exhibits highest reproducibility across all three subjects.

Table XII.7 Median and mean values of ACC for WM voxels across 3 subjects for the methods.

Method	Subject 1		Subject 2		Subject 3	
	Median	Mean	Median	Mean	Median	Mean
Super resolved CSD	0.37	0.37	0.49	0.47	0.31	0.31
SHORE-DL	0.49	0.41	0.59	0.49	0.48	0.42
SHORE-FOD	0.59	0.50	0.64	0.53	0.49	0.42
NNSHORE-DL	**0.63**	**0.52**	**0.67**	**0.56**	**0.61**	**0.52**

All methods were intercompared per subject, using Wilcoxon signed rank test and corrected using Bonferroni correction. All combinations were found to be significant.

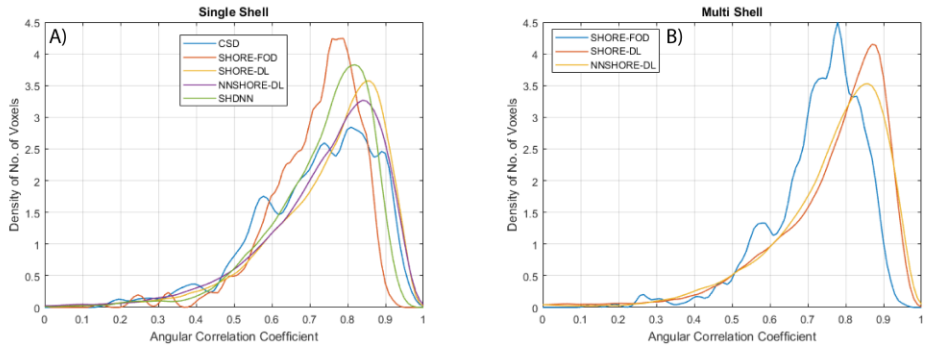

Figure XII.75. A) Comparison of single shell approaches on the diffusivity shell of 6000 s/mm^2 using ACC on all pairs of voxels of predictions of different methods with HFOD. B) Comparison of multi-shell approaches on all four shells between 3000 – 12000 s/mm^2 using ACC on all pairs of voxels of predictions for different methods.

Qualitatively we can observe that SHORE-FOD exhibits higher reproducibility than sCSD and NNSHORE-DL exhibits higher as compared to SHORE-FOD (Figure XII.77).

12.5. Conclusion

Deep SHORE is the first data-driven approach that generalizes diffusion microstructure estimation across multiple b-values, radial b-value sampling, and angular orientation sampling. Our approach enables direct comparison of data-driven diffusion analyses with model-based methods, e.g., sCSD, SHORE-FOD. Although Deep SHORE (NNSHORE-DL) compares favorably when subjected to quantitative cross-validation against histology data (Figure XII.75 & Figure XII.76, Table XII.6 & Table XII.7), the total amount of data available is a limitation of this study. As current and planned studies acquire more data, we

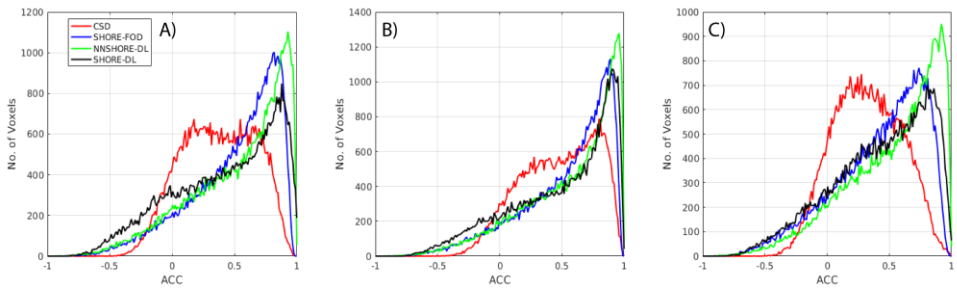

Figure XII.76. Comparison of proposed approaches with baselines of sCSD and SHORE-FOD across all pairs of WM voxels between the scans of site 'A' and 'B' for each subject. A) Subject 1 B) Subject 2 and C) Subject 3

will be able to better train and evaluate data-driven approaches.

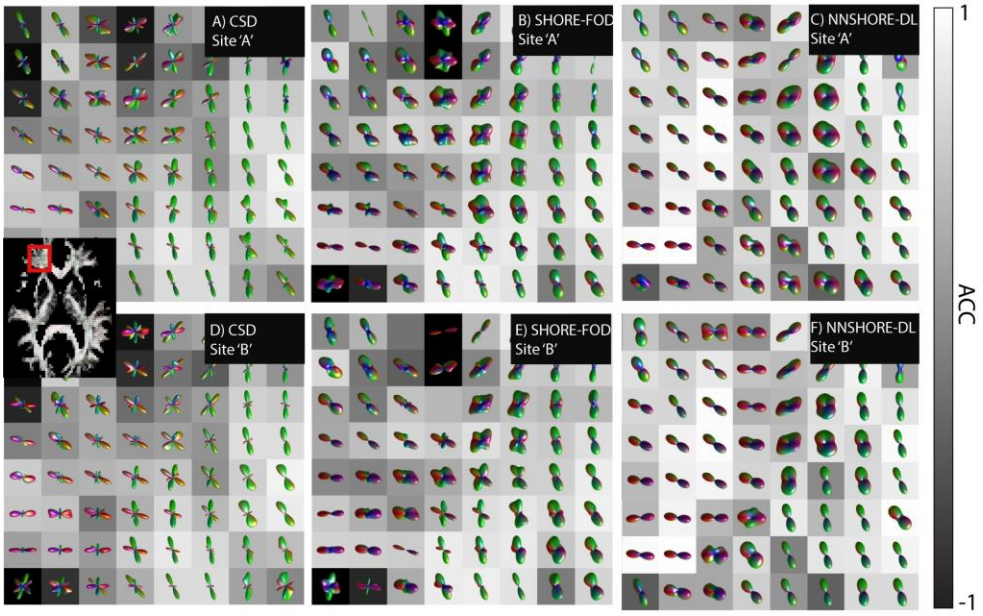

Figure XII.77. We focus on the ROI of left side frontal lobe of WM. The glyphs depict the FOD's derived from baselines of sCSD, SHORE-FOD and the proposed NNSHORE-DL. The underlay depicts the scalar measure of ACC. A & D) sCSD reconstructed on site 'A' and 'B'. B & E) SHORE-FOD reconstructed on site 'A' and 'B'. C & F) NNSHORE-DL reconstructed on site 'A' and 'B'.

13. Deep Learning Estimation of Multi-Tissue Constrained Spherical Deconvolution with Limited Single Shell DW-MRI

This chapter has been adapted from the published work in [221].

13.1. Introduction

Diffusion-weighted magnetic resonance imaging (DW-MRI), a non-invasive in-vivo MR imaging technique captures unique information regarding the microstructure of the human brain [222]. One of the first microstructure analysis techniques was diffusion tensor imaging (DTI) [35]. However, DTI has been limited by only recovery of fiber populations in a primary direction. Multiple advanced acquisition schemes with advanced reconstruction methods were proposed to detect crossing fiber populations [22]. The advanced reconstruction techniques are collectively referred to as high angular resolution diffusion imaging (HARDI) methods [18]. A primary application of reconstructed microstructure is for constructing the white matter (WM) neural pathways of the human brain also known as tractography [182]. Advanced tractography methods, such as high definition fiber tractography (HDFT) [40], have been applied for utilized for neurosurgery guidance. A caveat is that HDFT requires multi-shell DW-MRI (multiple diffusivity values) acquisitions which are expensive and take much more time as compared to a single shell acquisition [40]. This work is focused towards recovery of HDFT with single shell DW-MRI acquisitions (Figure XIII.78).

There are multiple methods that can be used as a prior for HDFT such as generalized q-sampling (GQI) [223] and multi-tissue constrained spherical deconvolution (MT-CSD) [224] both of which are microstructure reconstruction methods. This work tackles the problem of microstructure reconstruction only using the single shell DW-MRI acquisitions. There are a couple of existing approaches that have shown the possibility of recovery of tissue volume fraction from single-shell data. The first poses it as a non-negative factorization problem [225]. The second is a deep learning approach which directly takes the input of diffusion weighted images for fiber orientation distribution function (fODF) reconstruction [226]. The first approach has only been shown to reconstruct microstructure while the second one is restricted in terms

of broader applicability as it is tied to input of diffusion weighted images directly and does not include a joint estimation of tissue fraction. Herein, we propose to use deep learning techniques which ensure broader applicability by the usage of spherical harmonics and we explore the differences between using single voxel and cubic patches for reconstruction of MT-CSD.

Deep learning has become a powerful tool for learning non-linear mappings between a set of inputs and outputs where a non-linear mapping exists [166]. Although deep learning has been quite useful in other medical imaging domains, it is still in its nascent stages for DW-MRI. Recent work has been seen in microstructure estimation, harmonization and k-space reconstruction [85-87]. For this specific problem, we explore two different network architectures for recovery of MT-CSD microstructure. The first approach is a residual deep neural network (ResDNN) [25] of five layers with a residual block in between and takes input of a single voxel in the form of spherical harmonics (SH) while providing the output of fODF derived from MT-CSD which are also in the form of SH. The second network takes an input of a cubic patch of voxels and makes the prediction of the center voxel thus using spatial information as features for the deep learning network. This network has five convolutional layers with a residual block further consisting of convolutional layers (ResCNN). This network is inspired from previous work [66].

The methods proposed have been trained, validated, and tested on the human connectome dataset [227]. Deep learning networks were trained on 5 subjects and validated on 2. While 8 subjects were withheld for testing. The deep learning methods were also compared with the silver standard of super-resolved constrained spherical deconvolution (sCSD) [228] as a baseline. All comparisons were made with the MT-CSD being considered as the ground truth.

13.2. Data & Methods

13.2.1. Human Connectome Project Data

The human connectome project (HCP) dataset has an advanced acquisition scheme with three different diffusivity values 1000, 2000 and 3000 s/mm^2. All three diffusivity values are acquired with 90 gradient directions with interspersed b0's. The pre-processed dataset provided was used for this work. All

diffusion weighted volumes were normalized by the mean b0 as a standard pre-processing step. A total of 15 HCP subjects with the above acquisition scheme were used (Training: 5, Validation: 2, Testing: 8).

For training the proposed networks only the single shell of diffusivity value at 1000 s/mm^2 was used. The diffusion weighted volumes of that specific shell were fitted to 8th order SH which will be utilized as input training data for the network. Spherical harmonics in context of DW-MRI have become a standard way for representation of data with minimal representation error [33]. The output for the network which can be broken into two different parts 1.) SH coefficients of the fODF (8th order) which are reconstructed using MT-CSD on all three shells (1000, 2000 & 3000 s/mm^2) of DW-MRI data. 2.) Tissue volumes fractions which are scalar values for cerebrospinal fluid, apparent fiber density (white matter fraction), gray matter fraction.

13.2.2. Deep Learning Networks

The ResDNN is inspired from prior work and consists of five full connected dense layers (Figure XIII.79). The number of neurons used per layer are x1: 400, x2: 45, x3: 200, x4: 45, x5: 200. The residual block is formed using the addition of the x2 and x4 layer. All layers are activated using 'relu'. The inputs are a vector of 1x45 coefficients of 8th order DW-MRI SH. The outputs are 1x45 fODF SH coefficients at 8th order with an additional 1x3 scalar vector which represents the tissue fraction volume. To adapt the use of fractional volumes as output we use a modified loss function which is defined in (1) where m denotes the number of samples y_{true} is the set of fODF SH derived from MT-CSD y_{pred} is the set of SH predictions made by ResDNN P_{true} denotes the vector of tissue fraction value while P_{pred} denotes the predicted vector of tissue fractions.

$$L = \frac{1}{m}\sum_{i=1}^{m} \alpha(y_{true_i} - y_{pred_i})^2 + \beta\,(P_{true_i} - P_{pred_i})^2 \qquad \textbf{Eq. 31}$$

The ResCNN architecture stems from prior work where the network was originally intended for a harmonization problem. The network takes an input of 3x3x3x45 where the cubic patch consists of 8th order DWMRI SH coefficients for each voxel in the cubic patch (Figure XIII.79). The output is the same as defined for the ResDNN. This network architecture is divided into three parts, its core part being the residual block which consists of multiple functional units (each functional unit is dedicated to a specific order of

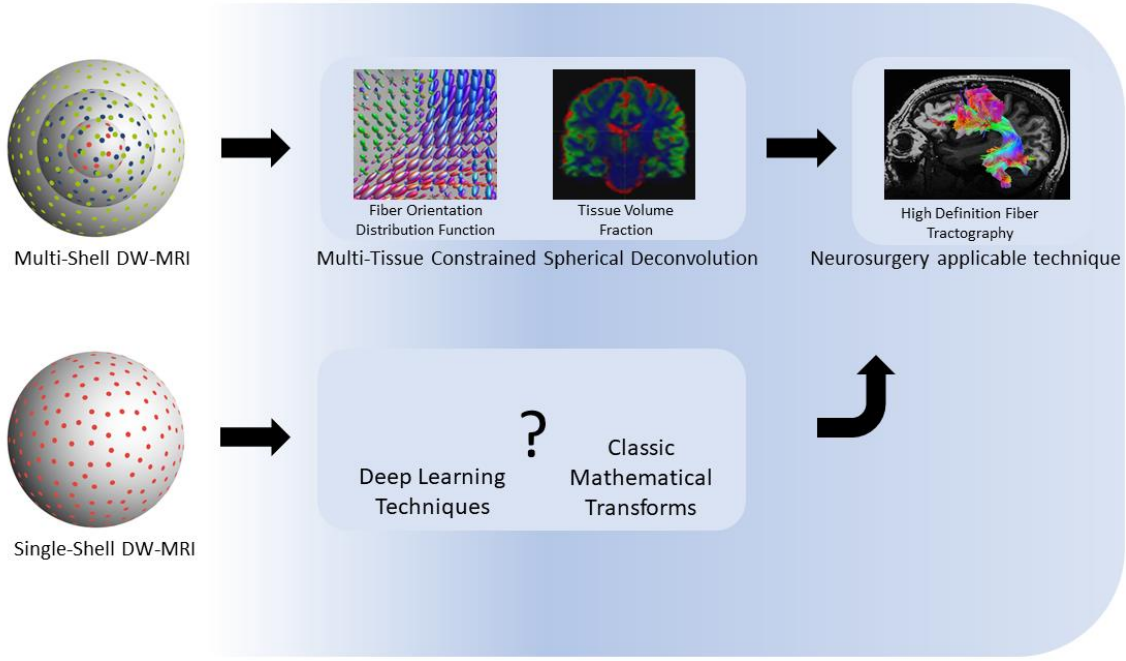

Figure XIII.78. The top row describes the pipeline of obtaining microstructure information from multi-shell DW-MRI which in turn can be used to perform high definition fiber tractography. The bottom row describes the problem that we tackle for this work where we explore of how to perform high definition fiber tractography using only single-shell DW-MRI data.

SH). The residual block can be stacked multiple times keeping the spatial dimensions intact. For our purpose we use a single block. The residual block is connected with two more convolutional kernels which is finally connected to a dense layer for predicting the center voxel of the cubic patch. All layers are 'relu' activated. We use the same modified loss as described above for ResDNN.

13.2.3. Evaluation Criteria

To compare the predictions of the proposed deep learning methods we use angular correlation

coefficient (ACC) [15] to evaluate the similarity of the prediction when compared with the ground truth estimate of MT-CSD. ACC is a generalized measure for all fiber population scenarios. It assesses the correlation of function of all directions over a spherical harmonic expansion. In brief, it provides the estimate of how closely a pair of fODF's are related on a scale of -1 to 1 where 1 is the best measure. Here 'u' and 'v' represent sets of SH coefficients.

$$ACC = \frac{\sum_{j=1}^{\infty}\sum_{m=-j}^{j} u_{jm} v_{jm}^{*}}{\left[\sum_{j=1}^{\infty}\sum_{m=-j}^{j}|u_{jm}|^{2}\right]^{0.5} \cdot \left[\sum_{j=1}^{\infty}\sum_{m=-j}^{j}|v_{jm}|^{2}\right]^{0.5}} \quad \textbf{Eq. 32}$$

The tissue fraction volumes are assessed using mean squared error (MSE) per voxel and per tissue fraction. MSE has been evaluated over the entire brain volume in encompassing all regions of the brain.

13.3. Results

We observe that ResCNN shows the most skewed distribution towards high correlation (Figure XIII.80 A) followed by ResDNN and sCSD. The mean ACC values across all 8 subjects were found to be 0.67, 0.72 and 0.64 for ResCNN, ResDNN and sCSD. Following subject wise ACC distributions (Figure XIII.80 B, C, D) for the entire brain volume we can see that ResCNN shows the most skewed distribution towards higher ACC for all subjects, followed by ResDNN and sCSD. Non-parametric signed rank test for all pairs of distributions were found to be $p < 0.001$. The root mean squared error across all pairs voxels for all subjects was found to be sCSD: 0.0323, ResDNN: 0.0168, ResCNN: 0.0124. Visually, the spatial map (Figure XIII.81) of the middle axial slice for a single subject indicates that sCSD has high ACC for WM, while ResDNN shows improvement in the circulatory regions between WM and GM with ResCNN showing the highest ACC. All three methods show low ACC for CSF and GM regions.

The RMSE for CSF and GM is lower for ResDNN as compared to ResCNN (Table XIII.8). While ResCNN only exhibits lower RMSE for only WM. The spatial maps for the middle axial slice of the same subject indicate higher error (MSE) of CSF and GM for ResCNN as compared to ResDNN. However lower MSE can be observed for WM for ResCNN as compared to ResDNN. Non-parametric signed rank test

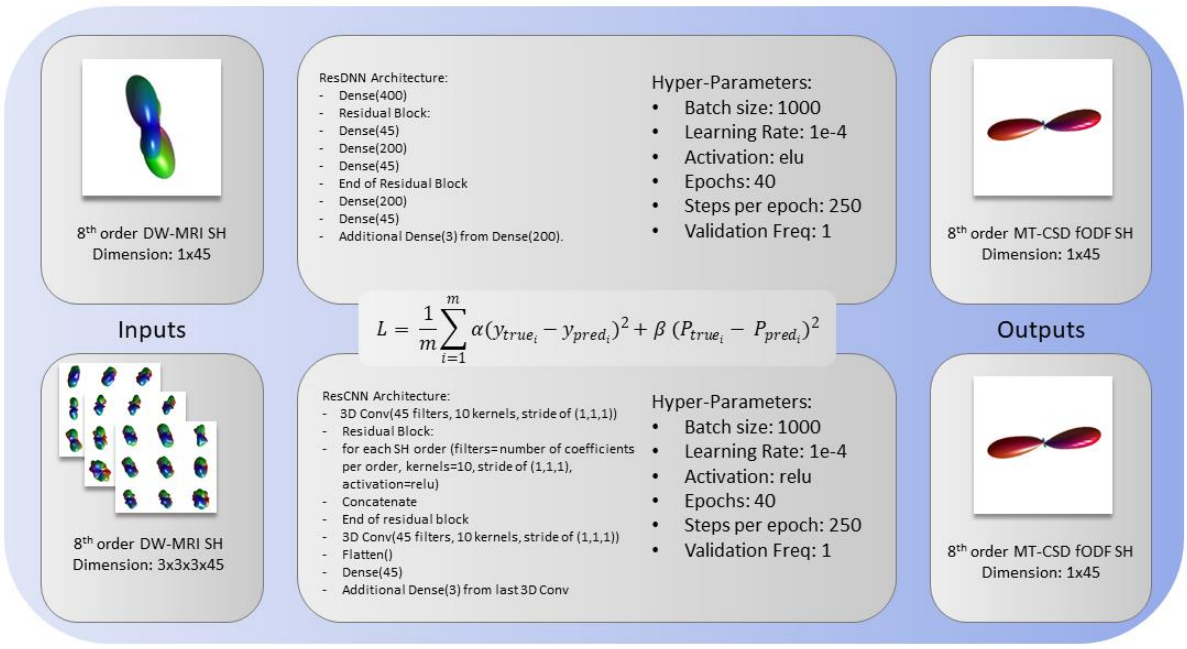

Figure XIII.79. Top row describes the input, the architecture and hyper-parameters of ResDNN and then the output of the network with the loss function. The bottom row descirbes the same for ResCNN architecture.

between distribution of MSE errors for the pair per tissue fraction were found to be $p < 0.001$.

Table XIII.8 RMSE of all tissue fraction volumes.

Method	RMSE CSF	RMSE GM	RMSE WM
ResDNN	**0.0135**	**0.0305**	0.0264
ResCNN	0.0173	0.0320	**0.0249**

13.4. Discussion

In this work, we depict a full pipeline that begins from scanner acquisition to HDFT reconstruction using only a single shell DW-MRI acquisition. Deep learning on SH coefficients allows for generalizability and applicability when other scanner acquisitions need to be used for HDFT reconstruction. The deep learning approaches proposed for this work well and could be further improved by exploration of more intricate deep learning architectures and larger training datasets. While we have shown the proposed method to be applicable at a clinical diffusivity value, further validation on reduced numbers of gradient directions is necessary.

To do a full 3 shell by 90 direction scan takes ~40 minutes. The deep learning can potentially provide usable data for HDFT quality scans at ~5 minutes for a diffusivity value of 1000 s/mm^2 with 32 gradient directions. Clinical MRI use of advanced methods is very limited due to longer scan acquisition time. Reducing scan time by a factor of 8 could set the conditions for clinical use of HDFT quality scanning

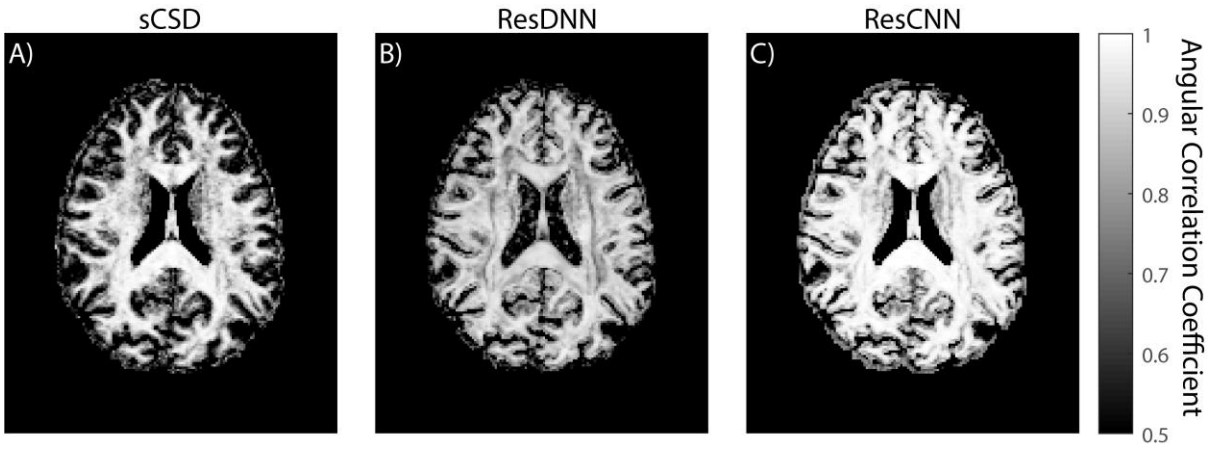

Figure XIII.80. Spatial angular correlation coefficient maps of the middle axial slice of the brain of a single subject. A) fODF predictions from sCSD. B) fODF predictions from ResDNN. C) fODF predictions from ResCNN

to grow dramatically perhaps several orders of magnitude.

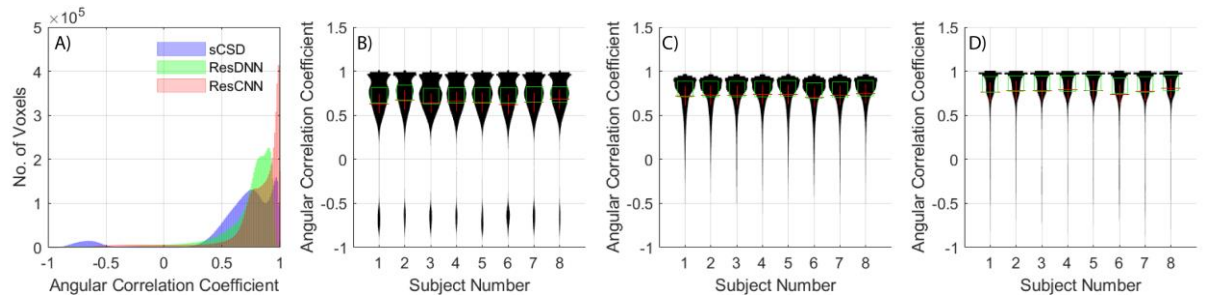

Figure XIII.81. A) Distribution of ACC for all voxels of withheld subjects using sCSD, ResDNN & ResCNN predicted fODF when compared with fODF of MT-CSD. B) Subject wise distribution of ACC for sCSD. C) Subject wise distribution of ACC for ResDNN. D) Subject wise distribution of ACC for ResCNN.

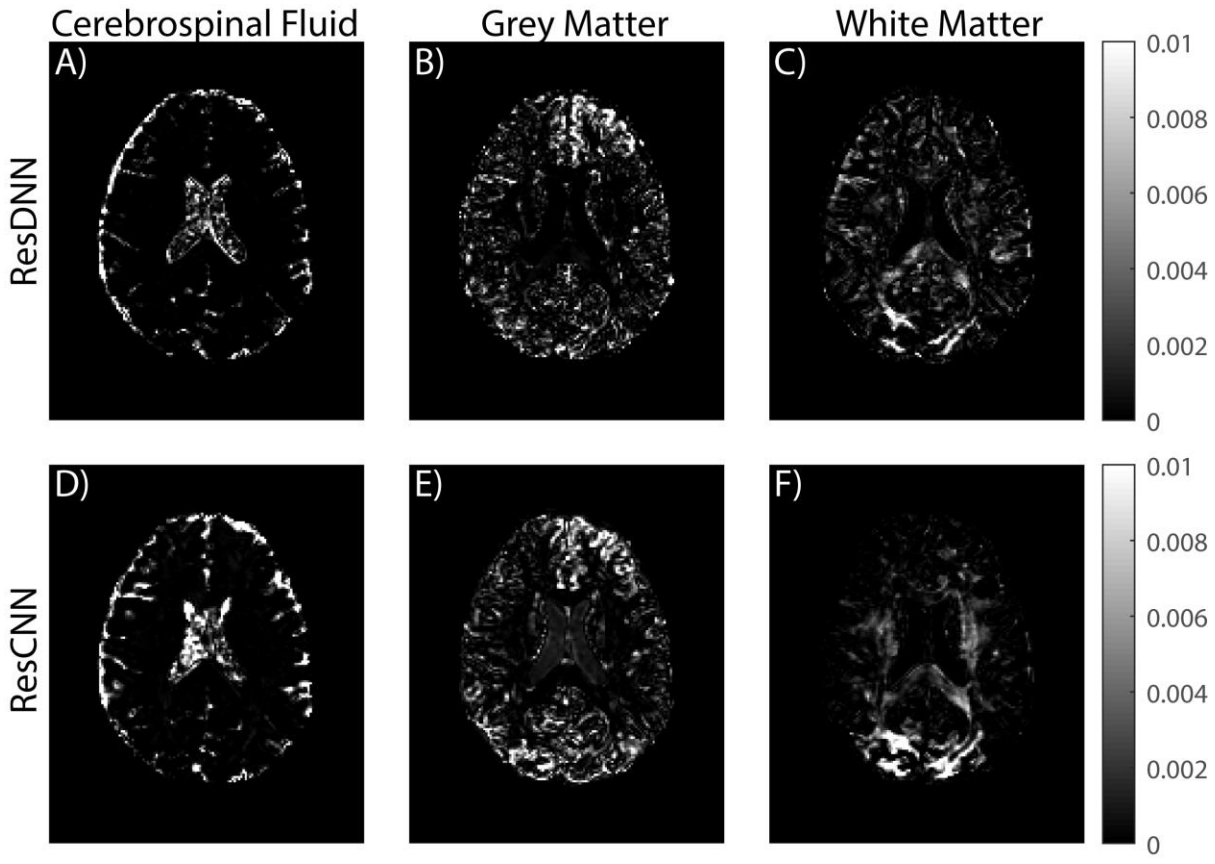

Figure XIII.82. Spatial maps of mean squared error between tissue fraction estimates of ResDNN and ResCNN with estimates of MT-CSD.

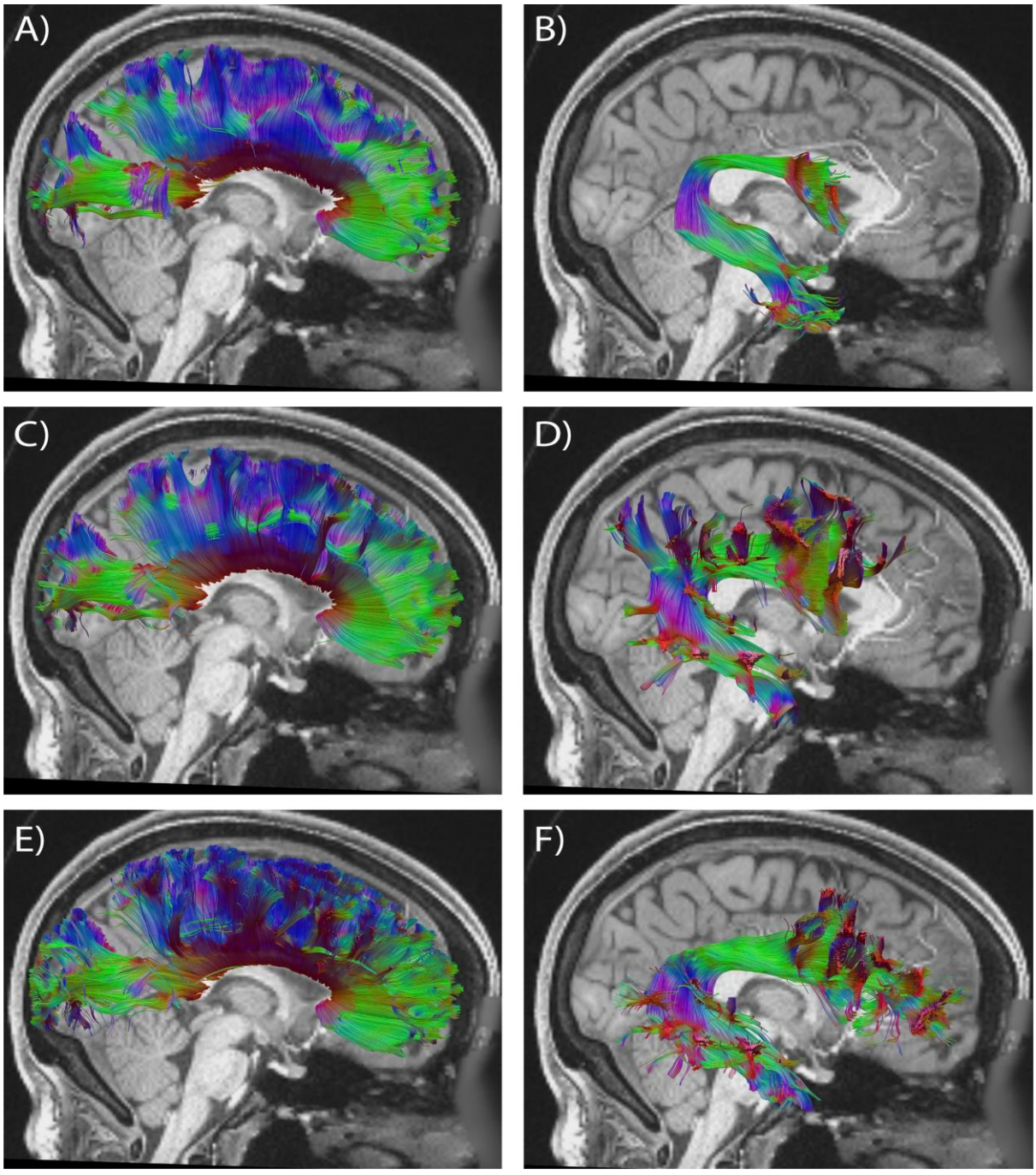

Figure XIII.83. Tractography reconstruction of HDFT for the tract corpus callosum using A) ResDNN C) ResCNN E) MT-CSD. Reconstruction of HDFT for tract arcuate using B) ResDNN D) ResCNN and F) MT-CSD.

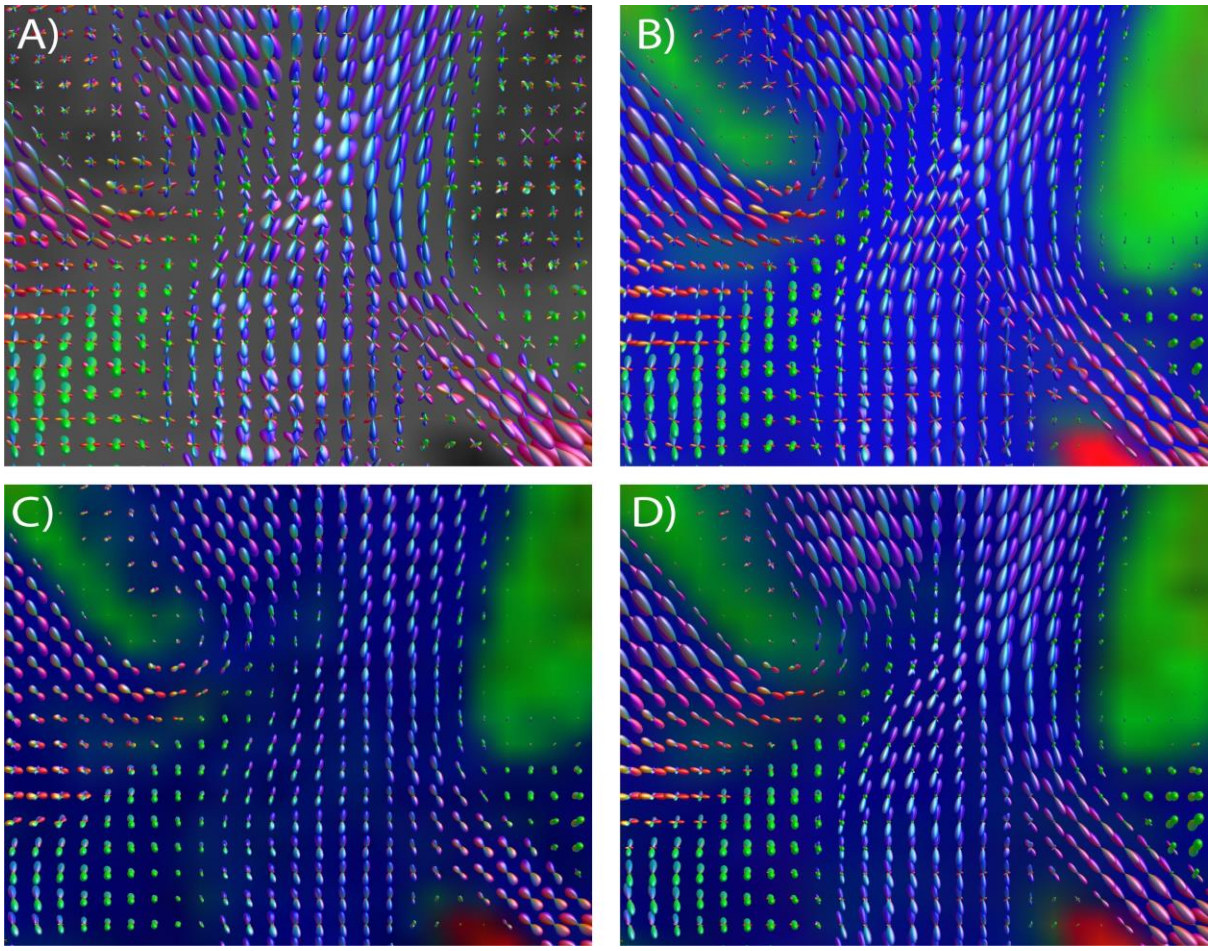

Figure XIII.84. Crossing fiber region of interest of fODF on a middle coronal slice reconstruction shown for A) sCSD, C) ResDNN, D) ResCNN and B) MT-CSD.

14. Recovery of Spherical Mean Technique for Single Shell DW-MRI

14.1. Introduction

Clinical use of diffusion-weighted magnetic resonance imaging (DW-MRI) has rapidly grown and its usage has increased for diagnosis of neurological diseases and surgical planning [40]. Diffusion imaging and microstructure models allow for estimation of tissue properties such as intra/extra cellular volume fraction (ICVF) and intra/extra diffusivities and structural connectivity for the in-vivo brain [222]. Recent development of scanner hardware enables high quality acquisitions of DW-MRI (high angular and radial sampling) [229]. Microstructure models are developed with the assumption of dense sampling of DW-MRI (for example at least 2 diffusivity shells).

Yet, typical clinical DW-MRI acquisition have only a single shell of diffusion measurement [10]. This is because the typical clinical scan is limited to a less than a ten-minute scan due to substantial reductions in reimbursements for the brain scan. A typical half our clinical slot involves structural, T2 and perhaps diffusion scanning, localizers and patient setup and removal within the 30 minutes slot. The Connectome scan was 40 minutes on a research dedicated scanner.

To have wide clinical use high resolution diffusion scanning must be done in a far shorter time. The primary motivation of this work to explore the possibility of estimating microstructure methods from a short duration single shell data and produce data through the use of deep learning methods to create near Connectome multi-shell quality data to set the stage for wide use of advanced diffusion tractography in clinical scans. In the paper we show that spherical mean technique (SMT) [17] a popular microstructure estimation technique, can be estimated using a single shell of data using machine learning.

Diffusion models can be segregated into two broad categories: 1) Geometric information such as fiber tracking [47] and tissue anisotropy. These are estimated using models such as Q-ball imaging (QBI), super-resolved constrained spherical deconvolution (sCSD), diffusion orientation transform (DOT), multi-tissue CSD [224]. 2.) Microstructural based models that estimate tissue properties such as neurite

orientation dispersion and density imaging (NODDI), SMT, AxCaliber, CHARMED, LEMANDE etc [16]. The geometric information in each voxel can be estimated using high angular sampling with single shell acquisitions. However, to estimate tissue microstructural properties, one requires multi-shell acquisitions with high b-values [16]. The classical diffusion models are derived based on mathematical model of diffusion process in biological tissue with complex geometries. Such models necessitate multiple b-values and therefore cannot be directly modelled using single shell (typically clinical) data. Yet, such a non-linear mathematical model can be learned using data-driven techniques specifically deep learning, which can enable the possibility to use these techniques clinically. Deep learning (DL) has most widely been used for semantic segmentation of different organs [230] for the human body and gradually it has trickled into inference of geometric tissue structure [25], surface parcellation [231, 232], registration [233], et cetera.

In brief, the method proposed in this paper is a pipeline that estimates MT-CSD and tissue volume fraction using a single shell. The MT-CSD tissue fraction volumes predicted from the trained models were used as inputs with SMT ground truth to train secondary networks. We present results for single voxel-based and patch-based DL regression techniques.

14.2. Methods

The section is segregated into two sub-sections of data pre-processing and the proposed method.

14.2.1. Data & Pre-processing

Preprocessed human connectome project (HCP) data were used [227]. A total of 15 subjects were utilized for the study of which 5 were used for training, 2 for validation and 8 for testing. The HCP acquisition is consistent across all subjects: 90 gradient directions per b-values: 1000, 2000 and 3000 s/mm^2.

MT-CSD [224] ground truth was reconstructed using all data (three shells). The MT-CSD jointly estimates fiber orientation distribution function (fODF) and tissue volume fractions for white matter (WM), grey matter (GM) and cerebrospinal fluid (CSF). SMT ground was reconstructed using all data as well. Spherical harmonics (SH) are a well-known representation of single shell DW-MRI data [33]. For input generalizability 8^{th} order even SH (45 coefficients) were fitted to b-value of 1000 s/mm^2 utilizing all 90 gradient directions. Regularized linear least squares were used for the fitting of SH coefficients.

14.2.2. Methods

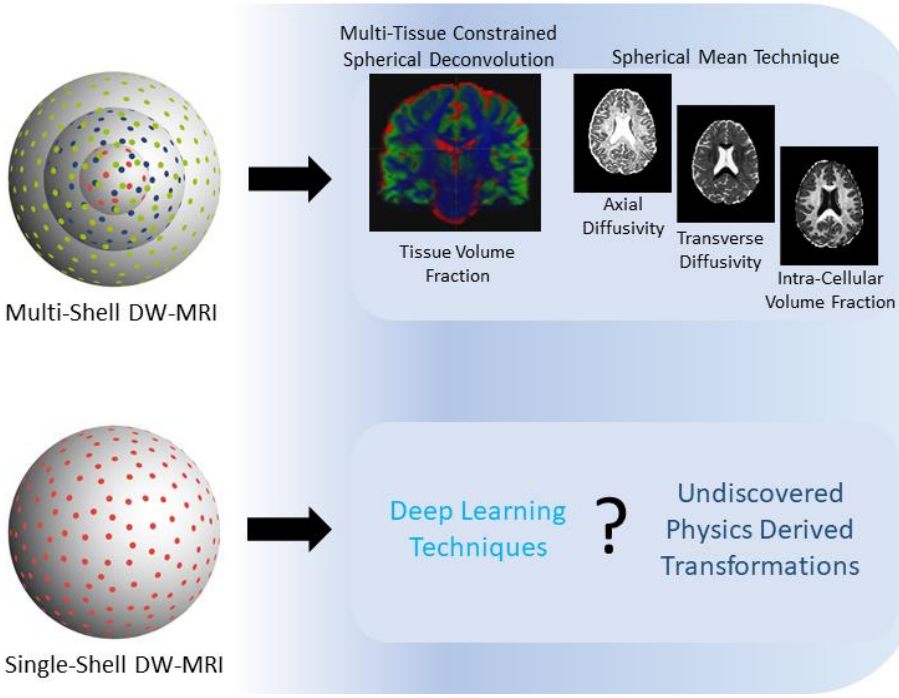

Figure XIV.85. Advanced microstructure methods rely on the assumption of the existence of multi-shell DW-MRI data. The proposed method is focused towards recovery of advanced measures such as spherical mean technique using single-shell DW-MRI at a clinical diffusivity value.

The proposed approach is segregated into two stages (Figure XIV.86ure 2). First, we train a model on single shell DW-MRI data to recover tissue volume fractions estimated via MT-CSD. Second, we train the secondary models to predict SMT metrics using predicted tissue volume fractions from stage 1.

14.2.3. Primary Deep learning Networks

The ResDNN network architecture from [25] was originally used to reconstruct histology derived fODF. For this work, the loss function was modified to incorporate a joint estimation of tissue volume fraction of MT-CSD. The loss function is defined as follows:

$$\boldsymbol{L1} = \frac{\mathbf{1}}{\boldsymbol{m}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \boldsymbol{\alpha}(\boldsymbol{F}_{\boldsymbol{t}_i} - \boldsymbol{F}_{\boldsymbol{p}_i})^{\mathbf{2}} + \boldsymbol{\beta}\,(\boldsymbol{P}_{\boldsymbol{t}_i} - \boldsymbol{P}_{\boldsymbol{p}_i})^{\mathbf{2}} \qquad \textbf{Eq. 33}$$

'm' represents batch size, 'F' is the MT-CSD fODF ground truth set of SH coefficients, 't' and 'p' represent true and predicted respectively, 'P' is the vector of tissue fraction volumes and α, β are hyper-parameters for the loss function terms. They were both set to 1 for this experiment. Inputs were a single voxel SH coefficients (1x45) and outputs were fODF (1x45) and tissue volume fractions (1x3).

ResCNN network follows the architecture from [66] and was originally proposed as network for harmonization of DW-MRI data across different sites. This network forms residual blocks per order of SH. Hence for our work it had 4 residual blocks dedicated to 2nd, 4th, 6th and 8th order SH coefficients. SH coefficients are orthogonal to each other and hence the above information was utilized to the advantage. The loss function was kept the same as above. Inputs were a cubic patch of 3x3x3x45 SH coefficients, and the outputs were tissue fraction volumes (1x3) and fODF (1x45) of a single voxel.

14.2.4. Secondary Deep learning Networks

The secondary deep learning networks were used for the estimation of SMT derived metrics of axial diffusivity (AD), transverse diffusivity (TD) and intra-cellular volume fraction (ICVF). They follow the architecture of the primary deep learning networks. For ResDNN only the inputs (1x3) and outputs (1x3) were modified. For ResCNN as there was no usage of SH coefficients only a single residual block was kept. The loss function that was used for both the secondary networks is defined below:

$$\boldsymbol{L2} = \frac{\mathbf{1}}{\boldsymbol{m}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \boldsymbol{\kappa}(\boldsymbol{x}_{\boldsymbol{t}_i} - \boldsymbol{x}_{\boldsymbol{p}_i})^{\mathbf{2}} + \boldsymbol{\nu}(\boldsymbol{y}_{\boldsymbol{t}_i} - \boldsymbol{y}_{\boldsymbol{p}_i})^{\mathbf{2}} + \boldsymbol{\rho}(\boldsymbol{z}_{\boldsymbol{t}_i} - \boldsymbol{z}_{\boldsymbol{p}_i})^{\mathbf{2}} \qquad \textbf{Eq. 34}$$

'x', 'y', 'z' represent AD, TD and ICVF respectively. 't' and 'p' represent true and predicted. κ, ν and ρ are weighted hyper-parameters. Due to the vast differences of scale for AD, TD vs ICVF there was a need to estimate κ, ν empirically. ICVF ranges on a scale of 0-1, while AD and TD range on a scale of 0-0.003. To ensure the physical nature of AD and TD we chose to estimate hyper-parameters instead of normalizing them on a scale of 0-1.

An empirical step wise search for weights of κ, ν was performed using a single training subject and validation subject in the range of (100 – 1100) at steps of 10. The best validation scores were estimated at 1020 for κ, ν both. The estimated hyper-parameter value was used for both ResDNN and ResCNN for the final training and overall results.

14.3. Results

Observing the quantitative results (Figure XIV.87), the ResCNN depicts lower RMSE for all three SMT metrics. Notably, the difference in the RMSE distribution is highest for TD. Non-parametric signed rank test between the two distributions showed statistical significance ($p< 0.001$). The RMSE for all subjects for ResDNN vs ResCNN are ICVF: 0.0893 vs 0.0768, AD: 0.0004 vs 0.0003 mm^2/s, TD: 0.0003 vs 0.0001 mm^2/s. Qualitatively (Figure XIV.88) ResCNN predictions look very similar to the ground truth upholding the spatial context. The RMSE for ICVF is low in CSF but higher in other WM regions. TD shows lower RMSE as compared to AD.

14.4. Discussion & Conclusion

The results indicate a recovery of SMT based metrics using only single-shell DW-MRI data. However, the RMSE is indicative of contextual anatomical based error for specific regions of interest. This can be indicative that more training data would be useful and also at the same time handling class imbalance when training such models. Such a recovery without radial sampling has so far been impossible with classical diffusion models. SMT has also been learned using linear short term memory models however they were limited by usage of two shell DW-MRI data as input [234]. An alternative explanation of the anatomical associations of error is that these regions represent the additional information that is present in multi-shell data that is not present (even when considering spatial context and prior information). Consideration of the differences between learned parameters with limited data and parameters estimated with all data may provide a fruitful area of biomarker exploration.

We have a proposed a novel approach to reconstruct SMT derived metrics using only a single shell of data at the cost of minimal to moderate error at specific locations. Relatively, the deep learning method

can recover microstructure information with a scan acquisition of 13 minutes approximately.

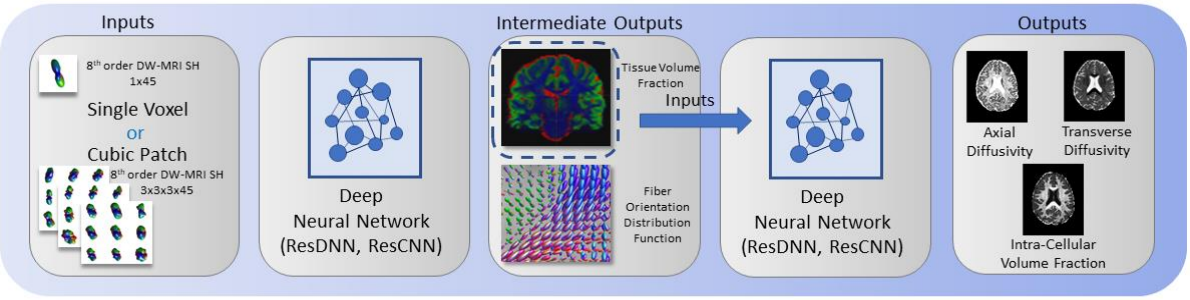

Figure XIV.86. Two stage deep learning framework where the primary stage is for learning the non-linear mapping of orientational fractional estimates. The fractional estimates are used as input for the second network to learn the non-linear mapping to predict spherical mean technique metrics.

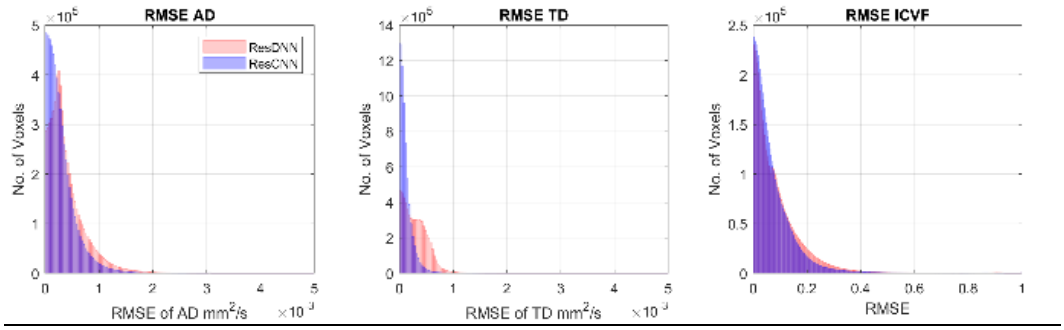

Figure XIV.87. Distribution of root mean squared error between predictions and ground truth of all voxels of all subjects for both methods ResDNN and ResCNN. Left: AD RMSE, Center: TD RMSE, Right: ICVF RMSE

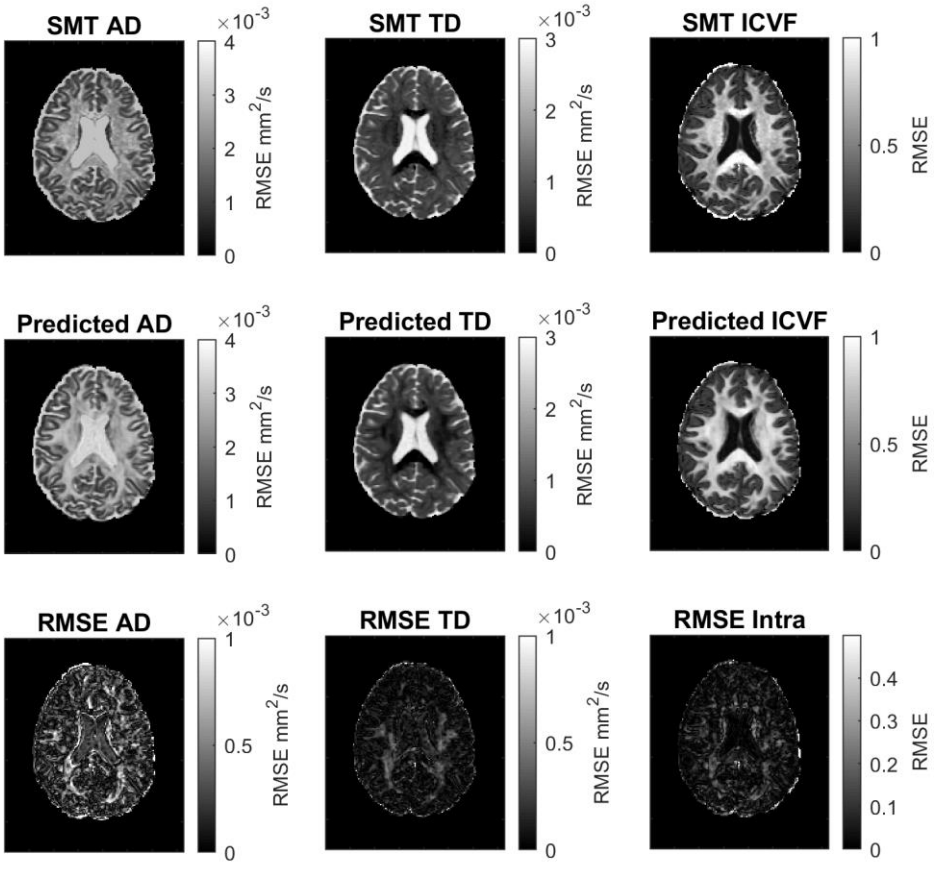

Figure XIV.88. **Spatial plots of a middle axial slice of a single subject. The columns depict AD, TD and ICVF respectively. The first row shows the ground truth derived from SMT using all data. The second row shows predictions made using only a single shell of data. The third row depicts spatial root mean squared error for the three metrics respectively.**

15. Conclusion & Future Work

15.1. Conclusion

The primary aspiration for this dissertation is to lay out the groundwork for the integration between DW-MRI and histology in terms of microstructure modelling. There are multiple rationales behind this primary aspiration of designing a wave of data-driven methodologies for DW-MRI microstructure tissue modelling. Firstly, the microstructure tissue models were primarily validated on phantom based acquisitions where even though the DW-MRI signal measurements are not 'synthetic' but the actual device as a phantom is 'synthetic' and it is biased to state that it can perfectly resemble the biological tissue properties of a human brain in-vivo. Secondly, lack of specificity and reproducibility among the many methods that model the tissue properties were always observable and there was a gap for improving sensitivity, specificity and reproducibility. This observation has been made through not only histology-based validation but also through multiple empirical reproducibility studies.

The empirical characterization performed as a part of this dissertation was unique in the way of large-scale single subject reproducibility datasets and the analysis performed upheld an observation of low inter-method agreement for microstructure modelling (Chapter II, III, IV and V). We utilized the aspect of histology-based validation as a forefront for establishment of data-driven machine learning methodologies for integration of DW-MRI with histology. The proposed methodologies paved the path for an increased specificity for histology driven validation which are a step closer to the biological tissue model complexity relative to phantom based validations. The new surge of data-driven methods holds key insights for a gain in information provided to neuro-clinicians, radiologists etc.

The gain in information from data-driven methodologies were proposed in different capacities for applicability to multiple domains of problems that exist across the spectrum of DW-MRI tissue microstructure modelling. The primary method is a generalized reconstruction of geometric microtissue structure that could map an in-vivo DW-MRI signal to ex-vivo microscopic based 3D histology

microstructure (Chapter VI, VII, XII). Due to higher specificity in terms of histology-based validation the information reconstructed is better validated and are more likely to reduce false positives for tractography based algorithms. Null-space variants of the same method (Chapter VI and VII) reconstruct harmonized microstructure across multiple scanner sites invariant to hardware specifications of the MRI scanners (Chapter VIII and IX). Harmonized reconstructions are more apt towards studies that are focused towards understanding effects on multiple populations where scanner hardware effects could lead to biased results, and usage of harmonized methodologies would cause a reduction in the bias. On a parallel note the data-driven methodologies promoted inter-modal learning between 2D microscopic imaging and 3D DW-MRI imaging where neighborhood information is being transformed to 3D microstructural information (Chapter X and XI). The inter-modal methodologies can provide 3D relevant information from microscopic imaging where there is no existence of DW-MRI information. The more advanced DW-MRI acquisition schemes where two or more shells of information are being acquired capture more detailed information. Due to increase in length of acquisition times novel data-driven methodologies are proposed that can reconstruct similar quality microstructural measures from single shell information (Chapter XIII, XIV).

In brief the data-driven methodologies have shown improvement for tract-based reconstructions and are applicable towards studying prognosis and diagnosis for neurological disorders and also for pre-surgical neurosurgery guidance information (Chapter VI, VII, XII, XIII, XIV). While the harmonizing and inter-modal data driven methodologies are applicable towards research studies for longitudinal, multi-site studies and when there is lack of DW-MRI acquisitions.

Analysis of brain microstructure is challenging for the white matter because of inconsistencies between different HARDI methods. The challenges are further elaborated by factors of multiple scanner sites, scanner hardware upgrades and variational longitudinal study designs. Moreover, because of inconsistencies of different HARDI methods tractography based analysis studies also get affected. Briefly, summarized are the contributions proposed in this dissertation and the successive section covers the new possible ventures for exploration based on the contributions.

Summarized contributions:

Contribution 1: Empirical characterization of reproducibility of microstructure

Contribution 2: Novel methods to increase precision and reproducibility of microstructure using single shell DW-MRI acquisitions

Contribution 3: Combining DW-MRI with 2D imaging to predict 3D measures from 2D imaging.

Contribution 4: Reconstruct reproducible and precise microstructure using multi-shell DW-MRI acquisitions.

Contribution 5: Multi-shell micro-architecture reconstruction using single shell DW-MRI imaging.

The successive sections describe the different kinds of future applicability on the grounds that this dissertation has developed.

15.2. Applicability of Proposed Contributions

15.2.1. Data-driven microstructural reconstructions

The aperture of data-driven modalities is a unique perspective that can drive reproducible microstructural reconstructions. The proposed idea that combines histology microstructural measures from a 3D perspective with DW-MRI is the first outlook that has been presented towards the scientific community engaged in DW-MRI research. Data-driven reconstruction methods for both single shell (Chapter VI, VII) and multi-shell DW-MRI (Chapter XII, XIII) have not only shown a higher reproducibility and specificity as compared to classical methodologies but also a wider preface of applicability towards tractography. Specifically, where the tractography is intended for pre-neurosurgical guidance information. Pilot studies have also indicated a similar observation.

15.2.2. Multi-site scanner harmonization

Multiple approaches have been proposed for harmonization of DW-MRI but in different contexts and specifically they have not been done so in terms of microstructural measures. The advocated contributions towards data-driven harmonization are the first of a kind of data-driven harmonization methodologies for DW-MRI microstructural measures. The harmonized reconstruction methods have been

shown to be generalizable to scan acquisitions from unseen/unknown scanner sites.

Harmonization methods such as these can be directly applied for a continued analysis of multiple populations of acquisitions that have been acquired at different scanner sites and enable a harmonized reconstruction of tracts based on microstructural reconstructions. Also, the harmonization methods were tested across scanners of different hardware capabilities such as 1.5T vs 3T. The methods are particularly useful for longitudinal analysis of data where the scanners often undergo hardware upgrades as the acquisitions are quite often separated by a few years in between.

15.2.3. Inter-modal data-driven methodologies

This dissertation advocates methods that can reconstruct 3D microstructural measures from 2D microscopic imaging by utilization of spatial neighborhood information from the 2D images. The methodology depicted is preliminary and has opened a crevice for the grounds of 3D reconstructions from 2D images. These kinds of methods can be used for studying microstructural measures where there is a lack of 3D microstructural measures due to lack of DW-MRI acquisitions. Although the results have been depicted on animal datasets, these methods can be extended towards applicability on cadaver-based tissue where there is a lack of DW-MRI acquisition.

Also, one might hypothesize to tackle the problem in the reverse direction, where the high resolution 2D image can be reconstructed from the in-vivo 3D DW-MRI imaging.

15.2.4. A Collection of data-driven methods

The primary resemblance for all the proposed methods is that they are of a data-driven nature and are inter-woven as such. They have collectively served as an inspiration for the successive contributions in this dissertation. The collective phase of all the data-driven methods under a common moniker have a high applicability in varied contexts and problems as mentioned in the prior sections. Nevertheless, at the same time based on these methods; also exists a wider cavity that thoroughly needs to be explored further for the methods to be applicable in a succinct and improved manner.

The consecutive section indicates of future ideas worthy exploring based on the work performed

as a part of this dissertation.

15.3. Future Ventures for Exploration

15.3.1. Ventures for data-driven methodologies for DW-MRI

Based on a thorough empirical validation for microstructural measures the work in this dissertation showed the tremendous applicability for a new wave of data-driven models for DW-MRI to model the microstructural measures (Contributions 1 and 2). However, the proposed contributions only show a tiny fraction of the upcoming possibilities for DW-MRI microstructure modelling. In brief, the following sub-areas of microstructure modelling still need to be addressed: 1.) Utility of spatial neighborhood information for modelling. 2.) The premises of for data-driven harmonization can be extended to multiple facets of clinical studies such as longitudinal studies, disease progression etc. 3.) The contributed data-driven methodologies are dependent upon 'unconstrained' deep learning networks, the term 'constrained' is an exciting area of exploration.

The primary approaches proposed in contribution 1 and 2 [19, 25, 41, 61, 129, 164, 179, 187] typically utilized a single 3D voxel-based input. Although data-driven, the methodology of a single voxel-based input is similar to most of the classical approaches [95, 105, 142, 216, 224, 235, 236] that model the DW-MRI signal to microstructural measures. The spatial neighborhood information generally has been quite aggressively explored for other imaging modalities such as segmentation for T1 MRI brain volumes [237-239] and even for deep learning based registration [233]. However, to a large extent the spatial information remains unexplored for DW-MRI microstructure modelling and currently there is no existence of an evaluation of how much spatial information gain is presented on incrementing or decrementing the spatial patch dimensions. Spatial information gain could also be directly utilized for addressing future problems for data-driven harmonization specifically for longitudinal studies where the neighborhood information could contain information about temporal artifacts or as such.

There are only a few data-driven harmonization techniques [66, 67, 187] so far. The prior proposed methods [66, 67] and including our contributions [164, 179, 187] only address harmonization in terms of

different scanner sites and varied acquisition schemes. It should also be noted that even the proposed methods have only been validated on a maximum 3-4 different scanning sites. Considering the real-world scenario, the problem is much wider as there more than 4 scanner manufacturers which manufacture various kinds of scanners with different hardware. Another lacking area of contribution is that these harmonization methodologies are not directly applicable towards harmonizing data of clinical studies for longitudinal aging or disease progression. Temporal affects are still widely unaccounted for when conducting statistical analysis for long duration clinical studies. The shift from traditional approaches to data-driven approaches is eminent and the current use of deep learning has been heavily incorporated in prior harmonization data-driven approaches [66, 67, 179, 187]. However, the deep learning networks in general have been framed as non-convex optimization objectives.

A typical deep learning model consists of multiple convolutional and fully connected layers. The optimization algorithms such as stochastic gradient descent, that channel the backpropagation is based on first order differentials of standard loss functions that are only 'indirectly' linked with the layers. However, the alternative hypothesis is that the layers of the deep learning network can be constrained to the domain-specific data. The constraining of deep learning layers is a recent development [240] and has not still been explored in most of the wide fields of NLP, Genomics and medical imaging etc. In general, medical imaging data has its own idiosyncrasies and domain specific constraints. These specific constraints need to be discovered and applied to push for an overall effective data-driven field of medical imaging in general and specifically for DW-MRI.

15.3.2. Ventures for 2D to 3D microstructural measures

The contribution 3 in this dissertation showed the wide possibilities of the recovery of 3D microstructural measures from super-high resolution 2D microscopic histology imaging. Although still data-driven approach at its core, the contribution spawns' multiple other problems that can be utilized for information gain from DW-MRI. However, the discovery that 3D reconstruction is possible in brief lays out the following other grounds: 1.) Reconstruction of tractography based streamlines from high resolution

2D microscopic imaging 2.) Reconstruction of high-resolution microscopic histology from 3D DW-MRI brain volumes.

There is sparse work on reconstruction of 3D microstructural measures from 2D microscopic histology imaging. The proposed work in this dissertation (contribution 3) is probably one of the first few proposed approaches in the literature [189]. Achieving tractography based reconstruction from 2D histology has so far been an unachievable feat. Yet, the validation for the 3D microstructural measures are closely a direct factor on tract-based reconstruction. Variation of spatial information patch for 3D microstructure reconstruction has also not been validated till date. Vice-versa, this open problem can also be framed the opposite way (3D DW-MRI to 2D High resolution microscopic imaging) for clinical information feasibility. High resolution microscopic imaging holds key information for pathology related studies and such a reconstruction has not been performed till date.

15.3.3. Ventures for DW-MRI Multi-shell learning

The work shown as a part of this dissertation (contribution 4 and 5) discovered the learning manifolds for data-driven methods of multi-shell DW-MRI. Although, multi-shell scanning acquisitions are moderately found in clinical studies; they cannot be declared a standard norm (for e.g single shell clinical acquisitions). The current progress is indicative of the fact that acquisition time for multi-shell scans is being reduced with advanced scanner hardware. The multi-shell acquisitions allow for more advanced and complex microstructural measures that are more sensitive to the underlying tissue properties of a voxel internally. The contribution 4 and 5 opened up the possibilities for the following in brief: 1.) Tractography based on multi-shell DW-MRI data driven based reconstruction. 2.) The possibility of recovery of the 'many many' DW-MRI multi compartment models from a single-shell of information.

15.4. Concluding Notes on Future Work

There are multifarious possibilities for applicability when utilizing data-driven methodologies. DW-MRI as a modality is the only one so far that holds information to model intra-voxel tissue properties. This dissertation is a critical step towards making the DW-MRI measures more specific, reproducible and

information towards pre-neurosurgical guidance information.

List of Publications

1. **V. Nath**, K. G. Schilling, P. Parvathaneni, C. B. Hansen, A. E. Hainline, C. Bermudez, Y. Huo, I. Lyu, J. A. Blaber, Y. Gao, V. A. Janve, I. Stepniewska, A. W. Anderson, B. A. Landman, "*Deep learning reveals untapped information for local white-matter fiber reconstruction in diffusion-weighted MRI*" MRI (2019)
2. **V. Nath**, K. G. Schilling, P. Parvathaneni, A. E. Hainline, Y. Huo … B. A. Landman, "*Tractography Reproducibility Challenge with Empirical Data (TraCED), 2017*", JMRI (2019)
3. Parvathaneni P, **Nath V**, McHugo M, Huo Y, Resnick SM, Woodward ND, Landman BA, Lyu I. "*Improving Human Cortical Sulcal Curve Labeling in Large Scale Cross-Sectional MRI using Deep Neural Networks.*" Journal of neuroscience methods. (2019) Jun 12:108311.
4. Schilling KG, Blaber J, Huo Y, Newton A, Hansen C, **Nath V**, Shafer AT, Williams O, Resnick SM, Rogers B, Anderson AW. *"Synthesized b0 for diffusion distortion correction (Synb0-DisCo)."* MRI (2019) May 7.
5. Schilling KG, Daducci A, Maier-Hein K, Poupon C, Houde JC, **Nath V**, Anderson AW, Landman BA, Descoteaux M. *"Challenges in diffusion MRI tractography–Lessons learned from international benchmark competitions.*" MRI 2019 Apr 1;57:194-209.
6. **V. Nath**, K. G. Schilling, P. Parvathaneni, A. E. Hainline, J. A. Blaber, Z. Ding, A. W. Anderson, B. A. Landman, "*Empirical Estimation of Intra-Voxel Structure with Persistent Angular Structure and Q-ball Models of Diffusion-Weighted MRI*", JMI (2017)
7. A. E. Hainline, **V. Nath**, P. Parvathaneni, J. A. Blaber, K. G. Schilling, A. W. Anderson, H. Kang, B. A. Landman, "*Empirical single sample quantification of bias and variance in Q-ball imaging*", MRM
8. P. Parvathaneni, **V. Nath**, K. G. Schilling, A. E. Hainline, J. A. Blaber, E. Mojahed, A. W. Anderson, B. A. Landman, "*Empirical reproducibility, sensitivity, and optimization of acquisition protocol, for Neurite Orientation Dispersion and Density Imaging using AMICO*", MRI
9. K. G. Schilling, **V. Nath**, J. A. Blaber, P. Parvathaneni, A. W. Anderson, B. A. Landman"*Empirical consideration of the effects of acquisition parameters and analysis model on clinically feasible q-ball imaging*", MRI
10. A. E. Hainline, **V. Nath**, P. Parvathaneni, K. G. Schilling, J. A. Blaber, A. W. Anderson, H. Kang, B. A. Landman, "*A deep learning approach to estimation of bias and variance in high angular resolution diffusion imaging*" In MRI
11. C. B Hansen, **V. Nath**, A. E. Hainline, K. G. Schilling, P. Parvathaneni, R. G. Bayrak, J. A. Blaber, O. Irfanoglu, C. Pierpaoli, A. W. Anderson, B. P. Rogers, B. A. Landman, "*Characterization and Correlation of Signal Drift in Diffusion Weighted MRI*", MRI
12. K. G. Schilling, **V. Nath**, C. B. Hansen, P. Parvathaneni, J. A. Blaber, Y. Gao, … B. A. Landman, "*Limits to anatomical accuracy of diffusion tractography using modern approaches*", Neuroimage (Submitted)

13. P. Parvathaneni, I. Lyu, K. G. Schilling, **V. Nath,** J. A. Blaber, A. E. Hainline, A. W. Anderson, N. D. Woodward, B. A. Landman "*Improved Gray Matter Surface based Spatial Statistics (N-GSBSS) in Neuroimaging Studies*", Neuroimage
14. Y. Huo, J. Blaber, S. M. Damon, B. D. Boyd, S. Bao, P. Parvathaneni,C. B. Noguera, S. Chaganti, **V. Nath**, J. M. Greer, et al., "*Towards portable large-scale image processing with high-performance computing*", Journal of Digital Imaging, 2018
15. K. G. Schilling, F. Yeh, **V. Nath**, C. B. Hansen, O. Williams, S. Resnick, A. W. Anderson, B. A. Landman. "*A fiber coherence index for quality control of B-table orientation in diffusion MRI scans*". MRI
16. **V Nath**, Sudhir K. Pathak, Kurt G. Schilling, Walter Schneider, Bennett A. Landman *"Deep Learning Estimation of Multi-Tissue Constrained Spherical Deconvolution with Limited Single Shell DW-MRI"*, SPIE 2020, Medical Imaging, Houston, TX, USA
17. **Nath V**, Lyu I, Schilling KG, Parvathaneni P, Hansen CB, Huo Y, Janve VA, Gao Y, Stepniewska I, Anderson AW, Landman BA. *"Enabling Multi-shell b-Value Generalizability of Data-Driven Diffusion Models with Deep SHORE."* In International Conference on Medical Image Computing and Computer-Assisted Intervention (2019) Oct 13 (pp. 573-581). Springer, Cham.
18. K. G. Schilling*, **V. Nath***, S. Remedios, R. G. Bayrak, Y. Gao, J. A. Blaber, Y. Huo, B. A. Landman, A. W. Anderson "*Learning 3D White Matter Microstructure from 2D Histology*" ISBI (2019), Venice, Italy
19. **V. Nath**, S. Remedios, P. Parvathaneni, C. B. Hansen, R. G. Bayrak, C. Bermudez, J. A. Blaber, K. Ramadass, K. G. Schilling, V. A. Janve, Y. Gao, Y. Huo, I. Lyu, O. Williams, S. Resnick, L. Beason-Held, B. P. Rogers, I. Stepniewska, A. W. Anderson, B. A. Landman, "*Harmonizing 1.5T/3T Diffusion Weighted MRI through Development of Deep Learning Stabilized Microarchitecture Estimators*" , SPIE (2019)
20. Parvathaneni P, Bao S, **Nath V**, Woodward ND, Claassen DO, Cascio CJ, Zald DH, Huo Y, Landman BA, Lyu I. *"Cortical Surface Parcellation using Spherical Convolutional Neural Networks."* In International Conference on Medical Image Computing and Computer-Assisted Intervention (2019) Oct 13 (pp. 501-509). Springer, Cham.
21. Hansen CB, **Nath V**, Hainline AE, Schilling KG, Parvathaneni P, Bayrak RG, Blaber JA, Williams O, Resnick S, Beason-Held L, Irfanoglu O. *"Consideration of cerebrospinal fluid intensity variation in diffusion weighted MRI.*" In Medical Imaging 2019: Physics of Medical Imaging 2019 Mar 1 (Vol. 10948, p. 109482G). International Society for Optics and Photonics.
22. Huo Y, Terry JG, Wang J, **Nath V**, Bermudez C, Bao S, Parvathaneni P, Carr JJ, Landman BA. *"Coronary calcium detection using 3D attention identical dual deep network based on weakly supervised learning."* In Medical Imaging 2019: Image Processing 2019 Mar 15 (Vol. 10949, p. 1094917). International Society for Optics and Photonics.
23. KG Schilling*, **V Nath***, Samuel Remedios, Roza G Bayrak, Yurui Gao, Justin A Blaber, Adam W Anderson, Bennett A Landman. *"Deep learning 3D white matter fiber orientation from 2D histology: pulling 3D rabbits out of 2D hats"*. In Proceedings of the 27th Annual Meeting of the International Society for Magnetic Resonance in Medicine. Abstract #5316. (2019)

24. Colin B. Hansen, **V Nath**, Kurt G. Schilling, Justin A. Blaber, Baxter P. Rogers, Bennett A. Landman. *"Compensation of Gradient Field Non-linearity and Signal Drift in Diffusion Weighted MRI".* ISBI 2019, Venice, Italy.
25. **V. Nath**, K. G. Schilling, J. A. Blaber, Z. Ding, A. W. Anderson, B. A. Landman, "*Comparison of multi-fiber reproducibility of PAS-MRI and Q-ball with empirical multiple b-value HARDI*", SPIE 2017
26. **V. Nath**, K. G. Schilling, A. E. Hainline, P. Parvathaneni, J. A. Blaber, I. Lyu, A. W. Anderson, H. Kang, A. T. Newton, B. P. Rogers, B. A. Landman, "*SHARD: spherical harmonic-based robust outlier detection for HARDI methods*", SPIE 2018
27. **V. Nath**, K. G. Schilling, P. Parvathaneni, C. B. Hansen, A. E. Hainline, C. Bermudez, S. Remedios, J. A. Blaber, V. A. Janve, Y. Gao, I. Stepniewska, B. P. Rogers, A. T. Newton, T. Davis, J. Luci, A. W. Anderson, B. A. Landman, "*Inter-Scanner Harmonization of High Angular Resolution DW-MRI using Null Space Deep Learning*" MICCAI-CDMRI, 2018
28. **V. Nath**, K. G. Schilling, P. Parvathaneni, A. E. Hainline, C. B. Hansen, C. Bermudez, A. J. Plassard, J. A. Blaber, V. A. Janve, Y. Gao, I. Stepniewska, A. W. Anderson, and B. A. Landman "*Deep Learning Captures More Accurate Diffusion Fiber Orientations Distributions than Constrained Spherical Deconvolution*" ISMRM 2018
29. A. E. Hainline, **V. Nath**, P. Parvathaneni, J. A. Blaber, K. G. Schilling, B. P. Rogers, A. T. Newton, J. Luci, H. Edmonson, H. Kang, B. A. Landman, "*Evaluation of inter-site bias and variance in diffusion-weighted MRI*" SPIE 2018
30. K. G. Schilling, **V. Nath**, J. A. Blaber, R. L. Harrigan, Z. Ding, A. W. Anderson, B. A. Landman, "*Effects of b-value and number of gradient directions on diffusion MRI measures obtained with Q-ball imaging*", SPIE 2017
31. S. Remedios, S. Roy, J. A. Blaber, C. Bermudez, **V. Nath**, M. B. Patel, J. A. Butman, D. L. Pham, B. A. Landman, "*Distributed deep learning for robust multi-site segmentation of CT imaging after traumatic brain injury*", SPIE 2019

1. Tamnes, C.K., et al., *Brain maturation in adolescence and young adulthood: regional age-related changes in cortical thickness and white matter volume and microstructure.* 2009. **20**(3): p. 534-548.
2. Crimmins, E.M.J.T.G., *Lifespan and healthspan: past, present, and promise.* 2015. **55**(6): p. 901-911.
3. Gunn-Moore, D., et al., *Alzheimer's disease in humans and other animals: A consequence of postreproductive life span and longevity rather than aging.* 2018. **14**(2): p. 195-204.
4. Alzheimer's, A.J.A.s. and d.t.j.o.t.A.s. Association, *2015 Alzheimer's disease facts and figures.* 2015. **11**(3): p. 332.
5. Koch-Henriksen, N. and P.S.J.T.L.N. Sørensen, *The changing demographic pattern of multiple sclerosis epidemiology.* 2010. **9**(5): p. 520-532.
6. Alonso, A. and M.A.J.N. Hernán, *Temporal trends in the incidence of multiple sclerosis: a systematic review.* 2008. **71**(2): p. 129-135.
7. Elsabbagh, M., et al., *Global prevalence of autism and other pervasive developmental disorders.* 2012. **5**(3): p. 160-179.
8. Van Essen, D.C., et al., *The WU-Minn human connectome project: an overview.* 2013. **80**: p. 62-79.
9. Van Essen, D.C., et al., *The Human Connectome Project: a data acquisition perspective.* 2012. **62**(4): p. 2222-2231.
10. Shock, N.W., *Normal human aging: The Baltimore longitudinal study of aging.* 1984.
11. Kawas, C., et al., *Age-specific incidence rates of Alzheimer's disease: the Baltimore Longitudinal Study of Aging.* 2000. **54**(11): p. 2072-2077.
12. Jansons, K.M. and D.C. Alexander, *Persistent angular structure: new insights from diffusion magnetic resonance imaging data.* Inverse problems, 2003. **19**(5): p. 1031.
13. Tuch, D.S., *Q-ball imaging.* Magnetic resonance in medicine, 2004. **52**(6): p. 1358-1372.
14. Özarslan, E., et al., *Resolution of complex tissue microarchitecture using the diffusion orientation transform (DOT).* 2006. **31**(3): p. 1086-1103.
15. Anderson, A.W., *Measurement of fiber orientation distributions using high angular resolution diffusion imaging.* Magnetic Resonance in Medicine, 2005. **54**(5): p. 1194-1206.
16. Zhang, H., et al., *NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain.* 2012. **61**(4): p. 1000-1016.
17. Kaden, E., F. Kruggel, and D.C.J.M.r.i.m. Alexander, *Quantitative mapping of the per-axon diffusion coefficients in brain white matter.* 2016. **75**(4): p. 1752-1763.
18. Descoteaux, M., *High angular resolution diffusion imaging (HARDI).* Wiley Encyclopedia of Electrical and Electronics Engineering, 2015.
19. Nath, V., et al. *Comparison of multi-fiber reproducibility of PAS-MRI and Q-ball*

with empirical multiple b-value HARDI. in *Medical Imaging 2017: Image Processing*. 2017. International Society for Optics and Photonics.
20. Schilling, K.G., et al., *Empirical consideration of the effects of acquisition parameters and analysis model on clinically feasible q-ball imaging.* Magnetic Resonance Imaging, 2017. **40**: p. 62-74.
21. Côté, M.-A., et al., *Tractometer: towards validation of tractography pipelines.* Medical image analysis, 2013. **17**(7): p. 844-857.
22. Daducci, A., et al., *Quantitative comparison of reconstruction methods for intra-voxel fiber recovery from diffusion MRI.* IEEE transactions on medical imaging, 2014. **33**(2): p. 384-399.
23. Schilling, K., et al., *Comparison of 3D orientation distribution functions measured with confocal microscopy and diffusion MRI.* Neuroimage, 2016. **129**: p. 185-197.
24. Schilling, K., et al., *Reproducibility and variation of diffusion measures in the squirrel monkey brain, in vivo and ex vivo.* 2017. **35**: p. 29-38.
25. Nath, V., et al., *Deep learning reveals untapped information for local white-matter fiber reconstruction in diffusion-weighted MRI.* Magnetic Resonance Imaging, 2019.
26. Nath, V., et al., *Enabling Multi-Shell b-Value Generalizability of Data-Driven Diffusion Models with Deep SHORE.* arXiv preprint arXiv:1907.06319, 2019.
27. Le Bihan, D., *Looking into the functional architecture of the brain with diffusion MRI.* Nature Reviews Neuroscience, 2003. **4**(6): p. 469.
28. Mitchell, H., et al., *The chemical composition of the adult human body and its bearing on the biochemistry of growth.* 1945. **158**(3): p. 625-637.
29. Le Bihan, D., et al., *MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders.* Radiology, 1986. **161**(2): p. 401-407.
30. Stejskal, E.O. and J.E. Tanner, *Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient.* The journal of chemical physics, 1965. **42**(1): p. 288-292.
31. Descoteaux, M., et al., *Multiple q-shell diffusion propagator imaging*. 2011. **15**(4): p. 603-621.
32. Ozarslan, E., et al., *Simple harmonic oscillator based reconstruction and estimation for three-dimensional q-space MRI.* 2009.
33. Descoteaux, M., et al., *Apparent diffusion coefficients from high angular resolution diffusion imaging: Estimation and applications.* Magnetic Resonance in Medicine, 2006. **56**(2): p. 395-410.
34. Ozarslan, E., et al., *Simple harmonic oscillator based reconstruction and estimation for three-dimensional q-space MRI.* 2009.
35. Basser, P.J., J. Mattiello, and D. LeBihan, *MR diffusion tensor spectroscopy and imaging.* Biophysical journal, 1994. **66**(1): p. 259-267.
36. Mori, S. and J.J.N. Zhang, *Principles of diffusion tensor imaging and its applications to basic neuroscience research.* 2006. **51**(5): p. 527-539.
37. Cheng, J., T. Jiang, and R. Deriche. *Theoretical analysis and practical insights on eap estimation via a unified hardi framework*. in *MICCAI Workshop on Computational Diffusion MRI (CDMRI)*. 2011.
38. Basser, P.J., et al., *In vivo fiber tractography using DT-MRI data.* 2000. **44**(4): p.

625-632.
39. Abhinav, K., et al., *Advanced diffusion MRI fiber tracking in neurosurgical and neurodegenerative disorders and neuroanatomical studies: a review.* Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 2014. **1842**(11): p. 2286-2297.
40. Fernandez-Miranda, J.C., et al., *High-definition fiber tractography of the human brain: neuroanatomical validation and neurosurgical applications.* Neurosurgery, 2012. **71**(2): p. 430-453.
41. Nath, V., et al., *Tractography reproducibility challenge with empirical data (traced): The 2017 ismrm diffusion study group challenge.* Journal of Magnetic Resonance Imaging, 2020. **51**(1): p. 234-249.
42. Schilling, K.G., et al., *Challenges in diffusion MRI tractography–Lessons learned from international benchmark competitions.* 2018.
43. Schilling, K.G., et al., *Limits to anatomical accuracy of diffusion tractography using modern approaches.* bioRxiv, 2018.
44. Schilling, K.G., et al., *Anatomical accuracy of standard-practice tractography algorithms in the motor system - A histological validation in the squirrel monkey brain.* Magnetic Resonance Imaging, 2019. **55**: p. 7-25.
45. Mori, S., et al., *Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging.* Ann Neurol, 1999. **45**(2): p. 265-9.
46. Mori, S., et al., *Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging.* Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society, 1999. **45**(2): p. 265-269.
47. Schilling, K.G., et al., *Limits to anatomical accuracy of diffusion tractography using modern approaches.* 2019. **185**: p. 1-11.
48. Baker, M., *Reproducibility crisis?* Nature, 2016. **533**(26): p. 353-66.
49. Beer-Borst, S., et al., *Dietary patterns in six European populations: results from EURALIM, a collaborative European data harmonization and information campaign.* European journal of clinical nutrition, 2000. **54**(3): p. 253-262.
50. Jahanshad, N., et al., *Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: A pilot project of the ENIGMA–DTI working group.* Neuroimage, 2013. **81**: p. 455-469.
51. Fortin, J.-P., et al., *Harmonization of multi-site diffusion tensor imaging data.* Neuroimage, 2017. **161**: p. 149-170.
52. Mirzaalian, H., et al. *Harmonizing diffusion MRI data across multiple sites and scanners*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2015. Springer.
53. Mirzaalian, H., et al., *Inter-site and inter-scanner diffusion MRI data harmonization.* NeuroImage, 2016. **135**: p. 311-323.
54. Mirzaalian, H., et al., *Multi-site harmonization of diffusion MRI data in a registration framework.* Brain imaging and behavior, 2018. **12**(1): p. 284-295.
55. Messiou, C., et al., *Optimising diffusion weighted MRI for imaging metastatic and myeloma bone disease and assessing reproducibility.* European radiology, 2011. **21**(8): p. 1713-1718.
56. Morbach, A.E., et al., *Diffusion-weighted MRI of the lung with hyperpolarized*

helium-3: a study of reproducibility. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2005. **21**(6): p. 765-774.
57. Rosenkrantz, A.B., et al., *Diffusion-weighted imaging of the abdomen at 3.0 Tesla: Image quality and apparent diffusion coefficient reproducibility compared with 1.5 Tesla.* Journal of Magnetic Resonance Imaging, 2011. **33**(1): p. 128-135.
58. Jian, B. and B.C. Vemuri. *Multi-fiber reconstruction from diffusion MRI using mixture of wisharts and sparse deconvolution*. in *Biennial International Conference on Information Processing in Medical Imaging*. 2007. Springer.
59. Jian, B., B.C. Vemuri, and E. Özarslan, *A mixture of wisharts (mow) model for multifiber reconstruction*, in *Visualization and Processing of Tensor Fields*. 2009, Springer. p. 39-56.
60. Jian, B., *Mathematical modeling for multi-fiber reconstruction from diffusion-weighted magnetic resonance images*. Vol. 70. 2009: Citeseer.
61. Nath, V., *Empirical Estimation of Intra-Voxel Structure with Persistent Angular Structure and Q-ball Models of Diffusion Weighted MRI.* 2017.
62. Farrell, J.A.D., et al., *Effects of signal-to-noise ratio on the accuracy and reproducibility of diffusion tensor imaging–derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5 T.* 2007. **26**(3): p. 756-767.
63. Landman, B.A., et al., *Effects of diffusion weighting schemes on the reproducibility of DTI-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5 T.* Neuroimage, 2007. **36**(4): p. 1123-1138.
64. Alexander, D. *A comparison of q-ball and PASMRI on sparse diffusion MRI data*. in *Proc. Int. Soc. Magn. Res. Med*. 2004.
65. Blumberg, S.B., et al. *Multi-stage prediction networks for data harmonization*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2019. Springer.
66. Koppers, S., et al. *Spherical harmonic residual network for diffusion signal harmonization*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2018. Springer.
67. Tax, C.M., et al., *Cross-scanner and cross-protocol diffusion MRI data harmonisation: A benchmark database and evaluation of algorithms.* NeuroImage, 2019. **195**: p. 285-299.
68. Rosenblatt, F., *The perceptron: a probabilistic model for information storage and organization in the brain.* Psychological review, 1958. **65**(6): p. 386.
69. Widrow, B. and M.E. Hoff, *Associative storage and retrieval of digital information in networks of adaptive "neurons"*, in *Biological Prototypes and Synthetic Systems*. 1962, Springer. p. 160-160.
70. Schmidhuber, J., *Deep learning in neural networks: An overview*. Neural networks, 2015. **61**: p. 85-117.
71. McCulloch, W.S. and W. Pitts, *A logical calculus of the ideas immanent in nervous activity.* The bulletin of mathematical biophysics, 1943. **5**(4): p. 115-133.
72. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* nature, 2015. **521**(7553): p. 436.

73. LeCun, Y., *LeNet-5, convolutional neural networks*. URL: http://yann. lecun. com/exdb/lenet, 2015. **20**: p. 5.
74. Szegedy, C., et al. *Inception-v4, inception-resnet and the impact of residual connections on learning*. in *Thirty-first AAAI conference on artificial intelligence*. 2017.
75. Ronneberger, O., P. Fischer, and T. Brox. *U-net: Convolutional networks for biomedical image segmentation*. in *International Conference on Medical image computing and computer-assisted intervention*. 2015. Springer.
76. Milletari, F., N. Navab, and S.-A. Ahmadi. *V-net: Fully convolutional neural networks for volumetric medical image segmentation*. in *2016 Fourth International Conference on 3D Vision (3DV)*. 2016. IEEE.
77. Shi, W., et al. *Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
78. Zhao, Z.-Q., et al., *Object detection with deep learning: A review.* IEEE transactions on neural networks and learning systems, 2019. **30**(11): p. 3212-3232.
79. Rawat, W. and Z. Wang, *Deep convolutional neural networks for image classification: A comprehensive review.* Neural computation, 2017. **29**(9): p. 2352-2449.
80. Goodfellow, I., et al. *Generative adversarial nets*. in *Advances in neural information processing systems*. 2014.
81. Young, T., et al., *Recent trends in deep learning based natural language processing.* ieee Computational intelligenCe magazine, 2018. **13**(3): p. 55-75.
82. Eraslan, G., et al., *Deep learning: new computational modelling techniques for genomics.* Nature Reviews Genetics, 2019. **20**(7): p. 389-403.
83. Ivezić, Ž., et al., *Statistics, data mining, and machine learning in astronomy: a practical Python guide for the analysis of survey data*. 2019: Princeton University Press.
84. Greenspan, H., B. Van Ginneken, and R.M. Summers, *Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique.* IEEE Transactions on Medical Imaging, 2016. **35**(5): p. 1153-1159.
85. Yang, G., et al., *DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction.* IEEE transactions on medical imaging, 2017. **37**(6): p. 1310-1321.
86. Seitzer, M., et al. *Adversarial and perceptual refinement for compressed sensing MRI reconstruction*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2018. Springer.
87. Schlemper, J., et al. *Stochastic Deep Compressive Sensing for the Reconstruction of Diffusion Tensor Cardiac MRI*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2018. Springer.
88. Nath, V., et al., *Inter-Scanner Harmonization of High Angular Resolution DW-MRI using Null Space Deep Learning.* arXiv preprint arXiv:1810.04260, 2018.
89. Koppers, S., C. Haarburger, and D. Merhof. *Diffusion mri signal augmentation: from single shell to multi shell with deep learning*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2016. Springer.
90. Tournier, J.-D., et al., *Direct estimation of the fiber orientation density function*

from diffusion-weighted MRI data using spherical deconvolution. NeuroImage, 2004. **23**(3): p. 1176-1185.
91. Özarslan, E., et al., *Resolution of complex tissue microarchitecture using the diffusion orientation transform (DOT).* NeuroImage, 2006. **31**(3): p. 1086-1103.
92. Farrell, J.A., et al., *Effects of signal-to-noise ratio on the accuracy and reproducibility of diffusion tensor imaging–derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5 T.* Journal of Magnetic Resonance Imaging, 2007. **26**(3): p. 756-767.
93. Landman, B.A., et al., *Effects of diffusion weighting schemes on the reproducibility of DTI-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5 T.* Neuroimage, 2007. **36**(4): p. 1123-1138.
94. Gorczewski, K., S. Mang, and U. Klose, *Reproducibility and consistency of evaluation techniques for HARDI data.* Magnetic Resonance Materials in Physics, Biology and Medicine, 2009. **22**(1): p. 63.
95. Cho, K.-H., et al., *Evaluation of the accuracy and angular resolution of q-ball imaging.* 2008. **42**(1): p. 262-271.
96. Alexander, D.C. *Maximum entropy spherical deconvolution for diffusion MRI*. in *Biennial International Conference on Information Processing in Medical Imaging*. 2005. Springer.
97. Andersson, J.L., S. Skare, and J.J.N. Ashburner, *How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging.* 2003. **20**(2): p. 870-888.
98. Andersson, J.L. and S.N. Sotiropoulos, *An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging.* Neuroimage, 2016. **125**: p. 1063-1078.
99. Jenkinson, M., et al., *Improved optimization for the robust and accurate linear registration and motion correction of brain images.* Neuroimage, 2002. **17**(2): p. 825-841.
100. Zhang, Y., M. Brady, and S. Smith, *Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm.* IEEE transactions on medical imaging, 2001. **20**(1): p. 45-57.
101. Descoteaux, M., et al., *Regularized, fast, and robust analytical Q-ball imaging.* Magnetic resonance in medicine, 2007. **58**(3): p. 497-510.
102. Cook, P., et al. *Camino: open-source diffusion-MRI reconstruction and processing*. in *14th scientific meeting of the international society for magnetic resonance in medicine*. 2006. Seattle WA, USA.
103. Sweet, A. and D. Alexander. *Reduced encoding persistent angular structure*. in *International society for magnetic resonance in medicine*. 2010.
104. Nath, V., et al., *Empirical estimation of intravoxel structure with persistent angular structure and Q-ball models of diffusion weighted MRI.* Journal of Medical Imaging, 2018. **5**(1): p. 014005.
105. Tournier, J.D., et al., *Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution.* NeuroImage, 2004. **23**(3): p. 1176-1185.
106. Alexander, D.C. and K.K. Seunarine, *Mathematics of crossing fibers*, in *Diffusion*

MRI: theory, methods, and application, D.K. Jones, Editor. 2010, Oxford University Press: Oxford ; New York. p. 451-464.
107. Parker, G.J.M. and D.C.J.P.t.o.t.R.S.B.B.s. Alexander, *Probabilistic anatomical connectivity derived from the microscopic persistent angular structure of cerebral tissue.* 2005. **360**(1457): p. 893-902.
108. Seunarine, K.K., et al. *Exploiting peak anisotropy for tracking through complex structures*. in *2007 IEEE 11th International Conference on Computer Vision*. 2007. IEEE.
109. Yo, T.-S., et al. *Quantifying brain connectivity: a comparative tractography study*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2009. Springer.
110. Descoteaux, M. and R. Deriche, *From local Q-ball estimation to fibre crossing tractography*, in *Handbook of Biomedical Imaging*. 2015, Springer, Boston, MA. p. 455-473.
111. Sweet, A. and D.C. Alexander. *Reduced encoding persistent angular structure*. in *International society for magnetic resonance in medicine*. 2010.
112. Xie, S., et al., *How does B-value affect HARDI reconstruction using clinical diffusion MRI data?* 2015. **10**(3): p. e0120773.
113. Perrin, M., et al., *Validation of q-ball imaging with a diffusion fibre-crossing phantom on a clinical scanner.* 2005. **360**(1457): p. 881-891.
114. Dayan, M., S. Kreutzer, and C.A. Clark, *Tractography of the optic radiation: a repeatability and reproducibility study.* NMR in Biomedicine, 2015. **28**(4): p. 423-431.
115. Andersson, J.L.R., S. Skare, and J. Ashburner, *How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging.* Neuroimage, 2003. **20**(2): p. 870-888.
116. Andersson, J.L.R. and S.N. Sotiropoulos, *Non-parametric representation and prediction of single-and multi-shell diffusion-weighted MRI data using Gaussian processes.* Neuroimage, 2015. **122**: p. 166-176.
117. Jenkinson, M. and S. Smith, *A global optimisation method for robust affine registration of brain images.* Medical image analysis, 2001. **5**(2): p. 143-156.
118. Holmes, C.J., et al., *Enhancement of MR images using registration for signal averaging.* 1998. **22**(2): p. 324-333.
119. Tustison, N.J., et al., *N4ITK: improved N3 bias correction.* 2010. **29**(6): p. 1310.
120. Avants, B.B., N. Tustison, and G.J.I.j. Song, *Advanced normalization tools (ANTS).* 2009. **2**: p. 1-35.
121. Avants, B.B., et al., *Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain.* 2008. **12**(1): p. 26-41.
122. Asman, A.J. and B.A.J.I.t.o.m.i. Landman, *Formulating spatially varying performance in the statistical fusion framework.* 2012. **31**(6): p. 1326-1336.
123. Asman, A.J. and B.A. Landman, *Non-local statistical label fusion for multi-atlas segmentation.* Medical image analysis, 2013. **17**(2): p. 194-208.
124. Freund, Y., R.E.J.J.o.c. Schapire, and s. sciences, *A decision-theoretic generalization of on-line learning and an application to boosting.* 1997. **55**(1): p. 119-139.

125. Klein, A., et al. *Open labels: online feedback for a public resource of manually labeled brain images*. in *16th Annual Meeting for the Organization of Human Brain Mapping*. 2010.
126. Cook, P.A., et al. *Camino: open-source diffusion-MRI reconstruction and processing*. in *14th scientific meeting of the international society for magnetic resonance in medicine*. 2006. Seattle WA, USA.
127. Schilling, K., et al., *Reproducibility and variation of diffusion measures in the squirrel monkey brain, in vivo and ex vivo.* Magn Reson Imaging, 2017. **35**: p. 29-38.
128. Reeder, S.B., et al., *Practical approaches to the evaluation of signal-to-noise ratio performance with parallel imaging: application with cardiac imaging and a 32-channel cardiac coil.* 2005. **54**(3): p. 748-754.
129. Nath, V., et al. *SHARD: spherical harmonic-based robust outlier detection for HARDI methods*. in *Medical Imaging 2018: Image Processing*. 2018. International Society for Optics and Photonics.
130. Tournier, J., F. Calamante, and A. Connelly. *How many diffusion gradient directions are required for HARDI*. in *Proc. Intl. Soc. Mag. Reson. Med*. 2009.
131. Henkelman, R.M., *Erratum: Measurement of signal intensities in the presence of noise in MR images [Med. Phys. 12, 232 (1985)].* Medical physics, 1986. **13**(4): p. 544-544.
132. Sharman, M.A., et al., *Impact of outliers on diffusion tensor and Q-ball imaging: Clinical implications and correction strategies.* Journal of Magnetic Resonance Imaging, 2011. **33**(6): p. 1491-1502.
133. Chang, L.C., D.K. Jones, and C. Pierpaoli, *RESTORE: robust estimation of tensors by outlier rejection.* Magnetic resonance in medicine, 2005. **53**(5): p. 1088-1095.
134. Chang, L.C., L. Walker, and C. Pierpaoli, *Informed RESTORE: a method for robust estimation of diffusion tensor from low redundancy datasets in the presence of physiological noise artifacts.* Magnetic resonance in medicine, 2012. **68**(5): p. 1654-1663.
135. Collier, Q., et al., *Iterative reweighted linear least squares for accurate, fast, and robust estimation of diffusion magnetic resonance parameters.* Magnetic resonance in medicine, 2015. **73**(6): p. 2174-2184.
136. Elhabian, S., et al. *Compressive sensing based Q-space resampling for handling fast bulk motion in hardi acquisitions*. in *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*. 2016. IEEE.
137. Pannek, K., et al., *HOMOR: higher order model outlier rejection for high b-value MR diffusion data.* Neuroimage, 2012. **63**(2): p. 835-842.
138. Andersson, J.L., et al., *Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images.* NeuroImage, 2016. **141**: p. 556-572.
139. Andersson, J.L. and S.N. Sotiropoulos, *Non-parametric representation and prediction of single-and multi-shell diffusion-weighted MRI data using Gaussian processes.* Neuroimage, 2015. **122**: p. 166-176.
140. Andersson, J.L., S. Skare, and J. Ashburner, *How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging.* Neuroimage, 2003. **20**(2): p. 870-888.

141. Jeurissen, B., et al., *Diffusion MRI fiber tractography of the brain.* NMR in Biomedicine, 2017.
142. Jeurissen, B., et al., *Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging.* Human brain mapping, 2013. **34**(11): p. 2747-2766.
143. Schilling, K.G., et al., *Empirical consideration of the effects of acquisition parameters and analysis model on clinically feasible q-ball imaging.* 2017. **40**: p. 62-74.
144. Smith, S.M., et al., *Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data.* Neuroimage, 2006. **31**(4): p. 1487-1505.
145. Lawes, I.N.C., et al., *Atlas-based segmentation of white matter tracts of the human brain using diffusion tensor tractography and comparison with classical dissection.* Neuroimage, 2008. **39**(1): p. 62-79.
146. Schmahmann, J.D., et al., *Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography.* Brain, 2007. **130**(3): p. 630-653.
147. Dauguet, J., et al., *Comparison of fiber tracts derived from in-vivo DTI tractography with 3D histological neural tract tracer reconstruction on a macaque brain.* Neuroimage, 2007. **37**(2): p. 530-538.
148. Dyrby, T.B., et al., *Validation of in vitro probabilistic tractography.* Neuroimage, 2007. **37**(4): p. 1267-1277.
149. Donahue, C.J., et al., *Using diffusion tractography to predict cortical connection strength and distance: a quantitative comparison with tracers in the monkey.* Journal of Neuroscience, 2016. **36**(25): p. 6758-6770.
150. Thomas, C., et al., *Anatomical accuracy of brain connections derived from diffusion MRI tractography is inherently limited.* Proceedings of the National Academy of Sciences, 2014. **111**(46): p. 16574-16579.
151. Azadbakht, H., et al., *Validation of high-resolution tractography against in vivo tracing in the macaque visual cortex.* Cerebral Cortex, 2015. **25**(11): p. 4299-4309.
152. Knösche, T.R., et al., *Validation of tractography: comparison with manganese tracing.* Human brain mapping, 2015. **36**(10): p. 4116-4134.
153. Calabrese, E., et al., *A diffusion MRI tractography connectome of the mouse brain and comparison with neuronal tracer data.* Cerebral Cortex, 2015. **25**(11): p. 4628-4637.
154. Maier-Hein, K.H., et al., *The challenge of mapping the human connectome based on diffusion tractography.* Nature communications, 2017. **8**(1): p. 1349.
155. Neher, P.F., et al., *Fiberfox: facilitating the creation of realistic white matter software phantoms.* Magnetic resonance in medicine, 2014. **72**(5): p. 1460-1470.
156. Ning, L., et al., *Sparse Reconstruction Challenge for diffusion MRI: Validation on a physical phantom to determine which acquisition scheme and analysis method to use?* Medical image analysis, 2015. **26**(1): p. 316-331.
157. Pujol, S., et al., *The DTI challenge: toward standardized evaluation of diffusion tensor imaging tractography for neurosurgery.* Journal of Neuroimaging, 2015. **25**(6): p. 875-882.
158. Wakana, S., et al., *Reproducibility of quantitative tractography methods applied to cerebral white matter.* 2007. **36**(3): p. 630-644.

159. Andersson, J.L. and S.N.J.N. Sotiropoulos, *Non-parametric representation and prediction of single-and multi-shell diffusion-weighted MRI data using Gaussian processes*. 2015. **122**: p. 166-176.
160. Smith, S.M., et al., *Advances in functional and structural MR image analysis and implementation as FSL*. Neuroimage, 2004. **23**: p. S208-S219.
161. Leemans, A. and D.K. Jones, *The B-matrix must be rotated when correcting for subject motion in DTI data.* Magnetic resonance in medicine, 2009. **61**(6): p. 1336-1349.
162. Huo, Y., et al. *Combining multi-atlas segmentation with brain surface estimation*. in *Proceedings of SPIE--the International Society for Optical Engineering*. 2016. NIH Public Access.
163. Huo, Y., et al., *Consistent cortical reconstruction and multi-atlas brain segmentation.* NeuroImage, 2016. **138**: p. 197-210.
164. Nath, V., et al. *Deep Learning Captures More Accurate Diffusion Fiber Orientations Distributions than Constrained Spherical Deconvolution*. in *Proceedings of International Society for Magnetic Resonance in Medicine*. 2018. Paris, France.
165. Tournier, J.-D., F. Calamante, and A. Connelly, *Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution.* Neuroimage, 2007. **35**(4): p. 1459-1472.
166. Krizhevsky, A., I. Sutskever, and G.E. Hinton. *Imagenet classification with deep convolutional neural networks*. in *Advances in neural information processing systems*. 2012.
167. Hinton, G., N. Srivastava, and K. Swersky, *Neural Networks for Machine Learning-Lecture 6a-Overview of mini-batch gradient descent*. 2012, Coursera Lecture slides.
168. Le Bihan, D., et al., *Artifacts and pitfalls in diffusion MRI.* Journal of magnetic resonance imaging, 2006. **24**(3): p. 478-488.
169. Jones, D.K. and M. Cercignani, *Twenty-five pitfalls in the analysis of diffusion MRI data.* NMR in Biomedicine, 2010. **23**(7): p. 803-820.
170. Lenglet, C., et al., *Mathematical methods for diffusion MRI processing.* Neuroimage, 2009. **45**(1): p. S111-S122.
171. Schilling, K., et al., *Comparison of 3D orientation distribution functions measured with confocal microscopy and diffusion MRI.* 2016. **129**: p. 185-197.
172. Helmer, K., et al. *Multi-site study of diffusion metric variability: effects of site, vendor, field strength, and echo time on regions-of-interest and histogram-bin analyses*. in *Medical Imaging 2016: Biomedical Applications in Molecular, Structural, and Functional Imaging*. 2016. International Society for Optics and Photonics.
173. Huo, J., et al., *Between-scanner and between-visit variation in normal white matter apparent diffusion coefficient values in the setting of a multi-center clinical trial.* Clinical neuroradiology, 2016. **26**(4): p. 423-430.
174. Mirzaalian, H., et al., *Inter-site and inter-scanner diffusion MRI data harmonization.* NeuroImage, 2016. **135**: p. 311-323.
175. Tournier, J.-D., F. Calamante, and A.J.N. Connelly, *Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-*

resolved spherical deconvolution. 2007. **35**(4): p. 1459-1472.
176. Stolp, H., et al., *Voxel-wise comparisons of cellular microstructure and diffusion-MRI in mouse hippocampus using 3D Bridging of Optically-clear histology with Neuroimaging Data (3D-BOND).* Scientific reports, 2018. **8**(1): p. 4011.
177. Smith, S.M., *Fast robust automated brain extraction.* Human brain mapping, 2002. **17**(3): p. 143-155.
178. Li, S., et al. *A discriminative null space based deep learning approach for person re-identification*. in *Cloud Computing and Intelligence Systems (CCIS), 2016 4th International Conference on*. 2016. IEEE.
179. Nath, V., et al. *Harmonizing 1.5 T/3T diffusion weighted MRI through development of deep learning stabilized microarchitecture estimators*. in *Medical Imaging 2019: Image Processing*. 2019. International Society for Optics and Photonics.
180. Le Bihan, D., et al., *Diffusion tensor imaging: concepts and applications.* Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2001. **13**(4): p. 534-546.
181. Basser, P.J. and D.K. Jones, *Diffusion-tensor MRI: theory, experimental design and data analysis–a technical review*. NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In Vivo, 2002. **15**(7-8): p. 456-467.
182. Basser, P.J., et al., *In vivo fiber tractography using DT-MRI data*. Magnetic resonance in medicine, 2000. **44**(4): p. 625-632.
183. Farrell, J.A., et al., *Effects of signal-to-noise ratio on the accuracy and reproducibility of diffusion tensor imaging–derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5 T.* 2007. **26**(3): p. 756-767.
184. Zhan, L., et al., *Magnetic resonance field strength effects on diffusion measures and brain connectivity networks.* Brain connectivity, 2013. **3**(1): p. 72-86.
185. Kochunov, P., et al., *Multi-site study of additive genetic effects on fractional anisotropy of cerebral white matter: comparing meta and megaanalytical approaches for data pooling.* Neuroimage, 2014. **95**: p. 136-150.
186. Forsyth, J.K., et al., *Reliability of functional magnetic resonance imaging activation during working memory in a multi-site study: analysis from the North American Prodrome Longitudinal Study.* Neuroimage, 2014. **97**: p. 41-52.
187. Nath, V., et al., *Inter-Scanner Harmonization of High Angular Resolution DW-MRI using Null Space Deep Learning.* 2018.
188. Chung, D., K. Tahboub, and E.J. Delp. *A two stream siamese convolutional neural network for person re-identification*. in *The IEEE international conference on computer vision (ICCV)*. 2017.
189. Nath, V., et al. *Learning 3D White Matter Microstructure from 2D Histology*. in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. 2019. IEEE.
190. Basser, P.J. and C. Pierpaoli, *Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI.* J Magn Reson B, 1996. **111**(3): p. 209-19.
191. Jones, D.K., *Diffusion MRI: Theory, Methods, and Applications*. 2010: Oxford University Press, USA.

192. Beaulieu, C., *The basis of anisotropic water diffusion in the nervous system - a technical review.* NMR Biomed, 2002. **15**(7-8): p. 435-55.
193. Mori, S. and P.C.M. van Zijl, *Fiber tracking: principles and strategies – a technical review.* NMR in Biomedicine, 2002. **15**(7-8): p. 468-480.
194. Maier-Hein, K.H., et al., *The challenge of mapping the human connectome based on diffusion tractography.* Nat Commun, 2017. **8**(1): p. 1349.
195. Thomas, C., et al., *Anatomical accuracy of brain connections derived from diffusion MRI tractography is inherently limited.* Proc Natl Acad Sci U S A, 2014. **111**(46): p. 16574-9.
196. Donahue, C.J., et al., *Using Diffusion Tractography to Predict Cortical Connection Strength and Distance: A Quantitative Comparison with Tracers in the Monkey.* J Neurosci, 2016. **36**(25): p. 6758-70.
197. Knosche, T.R., et al., *Validation of tractography: Comparison with manganese tracing.* Hum Brain Mapp, 2015. **36**(10): p. 4116-34.
198. Seehaus, A.K., et al., *Histological validation of DW-MRI tractography in human postmortem tissue.* Cereb Cortex, 2013. **23**(2): p. 442-50.
199. Choe, A.S., et al., *Validation of diffusion tensor MRI in the central nervous system using light microscopy: quantitative comparison of fiber properties.* NMR Biomed, 2012. **25**(7): p. 900-8.
200. Leergaard, T.B., et al., *Quantitative histological validation of diffusion MRI fiber orientation distributions in the rat brain.* PLoS One, 2010. **5**(1): p. e8595.
201. Jespersen, S.N., et al., *Determination of axonal and dendritic orientation distributions within the developing cerebral cortex by diffusion tensor imaging.* IEEE Trans Med Imaging, 2012. **31**(1): p. 16-32.
202. Schilling, K.G., et al., *Histological validation of diffusion MRI fiber orientation distributions and dispersion.* Neuroimage, 2018. **165**: p. 200-221.
203. Schilling, K., et al., *Comparison of 3D orientation distribution functions measured with confocal microscopy and diffusion MRI.* Neuroimage, 2016. **129**: p. 185-97.
204. Schilling, K.G., et al., *A Web-Based Atlas Combining MRI and Histology of the Squirrel Monkey Brain.* Neuroinformatics, 2018.
205. Tournier, J.D., F. Calamante, and A. Connelly, *MRtrix: Diffusion tractography in crossing fiber regions.* International Journal of Imaging Systems and Technology, 2012. **22**(1): p. 53-66.
206. Gallyas, F., *Silver staining of Alzheimer's neurofibrillary changes by means of physical development.* Acta Morphol Acad Sci Hung, 1971. **19**(1): p. 1-8.
207. Choe, A.S., et al., *Accuracy of image registration between MRI and light microscopy in the ex vivo brain.* Magn Reson Imaging, 2011. **29**(5): p. 683-92.
208. Nath, V., et al. *Deep Learning Captures More Accurate Diffusion Fiber Orientations Distributions than Constrained Spherical Deconvolution*
in *Proceedings of International Society for Magnetic Resonance in Medicine*. 2018. Paris, France.
209. Vishwesh Nath, S.R., Prasanna Parvathaneni, Colin B. Hansen, Roza G. Bayrak, Camilo Bermudez, Justin A. Blaber, Karthik Ramadass, Kurt G. Schilling, Vaibhav A. Janve, Yurui Gao, Yuankai Huo, Ilwoo Lyu, Owen Williams, Susan Resnick, Lori Beason-Held, Baxter P. Rogers, Iwona Stepniewska, Adam W. Anderson, Bennett A. Landman. *Harmonizing 1.5T/3T Diffusion Weighted MRI*

through Development of Deep Learning Stabilized Microarchitecture Estimaors. in *SPIE Medical Imaging*. 2019. San Diego, CA.
210. Vishwesh Nath, K.G.S., Prasanna Parvathaneni, Colin B. Hansen, Allison E. Hainline, Camilo Bermudez, Samuel Remedios, Justin A. Blaber, Vaibhav Janve, Yurui Gao, Iwona Stepniewska, Baxter P. Rogers, Allen T. Newton, Taylor Davis, Jeff Luci, Adam W. Anderson, Bennett A. Landman. *Inter-Scanner Harmonization of High Angular Resolution DW-MRI using Null Space Deep Learning*. in *MICCAI-CDMRI*. 2018. Granada, Spain.
211. Schilling, K., et al., *Can increased spatial resolution solve the crossing fiber problem for diffusion MRI?* NMR Biomed, 2017.
212. Jeurissen, B., et al., *Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging.* Hum Brain Mapp, 2013. **34**(11): p. 2747-66.
213. Gao, Y., et al., *Validation of DTI tractography-based measures of primary motor area connectivity in the squirrel monkey brain.* PLoS One, 2013. **8**(10): p. e75065.
214. Dyrby, T.B., et al., *Validation of in vitro probabilistic tractography.* Neuroimage, 2007. **37**(4): p. 1267-77.
215. Cheng, J., et al. *Model-free and analytical EAP reconstruction via spherical polar Fourier diffusion MRI*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2010. Springer.
216. Christiaens, D., et al., *Learning Compact ${q} $-Space Representations for Multi-Shell Diffusion-Weighted MRI.* IEEE transactions on medical imaging, 2019. **38**(3): p. 834-843.
217. Dyrby, T.B., et al., *An ex vivo imaging pipeline for producing high-quality and high-resolution diffusion-weighted imaging datasets.* 2011. **32**(4): p. 544-563.
218. Choe, A.S., et al., *Accuracy of image registration between MRI and light microscopy in the ex vivo brain.* 2011. **29**(5): p. 683-692.
219. Liu, D.C. and J.J.M.p. Nocedal, *On the limited memory BFGS method for large scale optimization.* 1989. **45**(1-3): p. 503-528.
220. Özarslan, E., et al., *Mean apparent propagator (MAP) MRI: a novel diffusion imaging method for mapping tissue microstructure.* 2013. **78**: p. 16-32.
221. Nath, V., et al., *Deep Learning Estimation of Multi-Tissue Constrained Spherical Deconvolution with Limited Single Shell DW-MRI.* arXiv preprint arXiv:2002.08820, 2020.
222. Le Bihan, D.J.N.R.N., *Looking into the functional architecture of the brain with diffusion MRI.* 2003. **4**(6): p. 469.
223. Yeh, F.-C., V.J. Wedeen, and W.-Y.I. Tseng, *Generalized ${q} $-sampling imaging.* IEEE transactions on medical imaging, 2010. **29**(9): p. 1626-1635.
224. Jeurissen, B., et al., *Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data.* NeuroImage, 2014. **103**: p. 411-426.
225. Christiaens, D., et al., *Unsupervised multi-tissue decomposition of single-shell diffusion-weighted imaging by generalization to multi-modal data*. Proceedings ISMRM 2016, 2016: p. 3038.
226. Lin, Z., et al., *Fast learning of fiber orientation distribution function for MR tractography using convolutional neural network.* Medical physics, 2019.
227. Van Essen, D.C., et al., *The WU-Minn human connectome project: an overview.*

2013. **80**: p. 62-79.
228. Tournier, J.D., F. Calamante, and A.J.N. Connelly, *Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution.* 2007. **35**(4): p. 1459-1472.
229. Ye, W., et al., *An efficient interlaced multi-shell sampling scheme for reconstruction of diffusion propagators.* IEEE transactions on medical imaging, 2012. **31**(5): p. 1043-1050.
230. Shin, H.-C., et al., *Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning.* IEEE transactions on medical imaging, 2016. **35**(5): p. 1285-1298.
231. Parvathaneni, P., et al., *Cortical Surface Parcellation using Spherical Convolutional Neural Networks.* arXiv preprint arXiv:1907.05395, 2019.
232. Parvathaneni, P., et al., *Improving Human Cortical Sulcal Curve Labeling in Large Scale Cross-Sectional MRI using Deep Neural Networks.* Journal of neuroscience methods, 2019: p. 108311.
233. Zhu, W., et al., *NeurReg: Neural Registration and Its Application to Image Segmentation.* arXiv preprint arXiv:1910.01763, 2019.
234. Ye, C., X. Li, and J. Chen, *A deep network for tissue microstructure estimation using modified LSTM units.* Medical image analysis, 2019. **55**: p. 49-64.
235. Tournier, J.D., F. Calamante, and A. Connelly, *Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution.* Neuroimage, 2007. **35**(4): p. 1459-1472.
236. Tournier, J.D., F. Calamante, and A. Connelly. *How many diffusion gradient directions are required for HARDI*. in *Proc. Intl. Soc. Mag. Reson. Med*. 2009.
237. Huo, Y., et al. *Spatially localized atlas network tiles enables 3D whole brain segmentation from limited data*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2018. Springer.
238. Huo, Y., et al., *3D whole brain segmentation using spatially localized atlas network tiles.* NeuroImage, 2019. **194**: p. 105-119.
239. Roth, H.R., et al. *A new 2.5 D representation for lymph node detection using random sets of deep convolutional neural network observations*. in *International conference on medical image computing and computer-assisted intervention*. 2014. Springer.
240. Agrawal, A., et al. *Differentiable convex optimization layers*. in *Advances in Neural Information Processing Systems*. 2019.