Multi -Atlas Guided Automatic Tract Reconstruction of White Matter Fibers

By

Xuan Wang

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Electrical Engineering

May 31, 2020

Nashville, Tennessee

Approved:

Bennett A. Landman, Ph.D.
Richard Alan Peters, Ph.D.

ACKNOWLEDGEMENTS

I would first like to thank my thesis advisor Bennett Landman. The door to Prof. Landman office was always open whenever I ran into a trouble spot or had a question about my research or writing. He consistently allowed this paper to be my own work, but steered me in the right direction whenever he thought I needed it.

I want to thank the folks at Vanderbilt University who have helped me. Roza Gunes Bayrak, thank you for taking care of me during TracEM project. You are such a humble, smart and nice friend, which makes everyone like you. Shunxing Bao, Riqing Gao, and Yucheng Tang thank you for helping me for both my career and study. I am grateful to my lab folks, Karthik Ramadass, Justin Blaber, Colin Hansen, Kurt Schilling. I am grateful and thank you for your help.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS ii

LIST OF ABBREVIATIONS v

LIST OF TABLES vi

LIST OF FIGURES vii

Chapter

1. Introduction 1

2. Materials 3

    2.1. Subject 3
    2.2. Image Acquisition 3
    2.3. Pre-Process 3
    2.4. Atlas Generation 4
    2.5. Data Scale 4

3. Methodology 5

    3.1. Registration 6
    3.2. Label Propagation 6
        3.2.1. Coordinate Transformation 6
        3.2.2. Label Propagation 6
    3.3. Label Fusion 7
        3.3.1. Normalize Images 7
        3.3.2. First Branch 7
        3.3.3. Second Branch 7
    3.4. Voting on Averaged Labels 8
    3.5. Fiber Tracking 8

4. Result 9

    4.1. Qualitative Analysis 9
    4.2. Quantitative Analysis 16
        4.2.1. Inter-Rater Tract Reliability 16
        4.2.2. Tract Reproducibility 18

5. General Discussion 21
5.1. Threshold Selection and Weight Assignment 21
5.2. ROA Regions 22
5.3. Time Consumption 22
6. Conclusion 23
REFERENCE 24
Appendix A 28

LIST OF ABBREVIATIONS

WM	White Matter
MRI	Magnetic Resonance Imaging
DWI	Diffusion-Weighted Magnetic Resonance Imaging
DTI	Diffusion Tensor Imaging
ROI	Regions of Interest
ROA	Region of Avoidance
MNI	Montreal Neurological Institute and Hospital
BLSA	Baltimore Longitudinal Study of Aging
HCP	Human Connectome Project
DICE	Dice Correlation Coefficient
HARDI	High Angular Resolution Diffusion Imaging
QBI	Q-Ball Image
GQI	Generalized Q-sampling Image

LIST OF TABLES

Table **Page**

1 . The Quantitative Result of The First Method Branch 17

2 . The Quantitative Result of The Second Method Branch 18

3 . Tract Reproducibility 19

LIST OF FIGURES

Figure **Page**

1 . Summary of The Method. 5

2 . Visualization Result of All Tracts 12

3 . Tracts in The Brainstem. 13

4 . Tracts in Projection Fibers..... 14

5 . Association fibers. 15

6 . Commissural fibers..... 16

7 . Complex tract (cerebral peduncle) with complex ROA regions 19

8 . Simple tract (genu corpus callosum) with simple ROA regions 20

9 . Small tract (superior cerebellar peduncle) with complex ROA regions 20

Chapter 1. Introduction

White matter (WM) contains millions of nerve axons that connect different parts of the brain. Axons have similar destinations and can be grouped in large fiber bundles called white matter tracts. The size of prominent tracts can be as large as a few centimeters in the human brain, which has been recognized in previous studies by using postmortem samples (Dejerine, 1895). Because fiber bundles play a very important role in the human brain, mapping these tracts can help us reveal the connectivity abnormality that indicate some diseases and/or behavioral disorders in humans. Recently, it has been shown that many of these tracts can be reconstructed noninvasively and three-dimensionally based on pixel-by-pixel diffusion orientation information obtained from diffusion tensor imaging (DTI) (Basser et al., 2000; Conturo et al., 1999; Jones et al., 1999b; Lazar et al., 2003; Mori et al., 1999, 2005; Parker et al., 2002; Poupon et al., 2000; Wakana et al., 2004). In fact, this 3D reconstruction technique, often called tractography (three-dimensional tract reconstruction) based on diffusion tensor imaging (DTI) is the only current tool which allows us to visualize white matter tracts in vivo and noninvasively. Diffusion-weighted magnetic resonance imaging (DWI) (Basser et al., 1994) is a powerful non-invasive brain imaging technique introduced in (Le Bihan et al., 1986; Merboldt et al., 1985; Taylor et al., 1985). DWI measures water diffusion and allows water diffusion to be visualized by sensitizing images in the direction of the field gradients. Therefore, it provides biologically and clinically relevant information about white matter connectivity and integrity that is not available from other imaging modalities. Diffusion tensor imaging is calculated from DWI and gives information about the principal direction of diffusion. By combining the directional information and magnitude of anisotropic diffusion of the individual voxels, the white matter tracts can be reconstructed.

It is important to accurately map white matter fibers and structures from whole-brain tractography. When fibers are grouped into bundles, the bundles can offer valuable insight about how disease affects the particular white matter tracts (Price et al., 2007, 2008). Clustering methods can group fibers computed from tractography into organized bundles or tracts, enabling large population studies of disease and tract shapes. One simple and practical strategy is to select anatomically particular white matter tracts that interconnect anatomical regions of interest (ROI) (Wakana et al., 2007; Zhang et al., 2010). However, the process always requires manual intervention to screen out false positive fibers.

For this project, I used the software DSI-Studio to group fibers as a clustering method. DSI-Studio (http://dsi-studio.labsolver.org) is a software for diffusion tensor image computation, region of interest based deterministic diffusion fiber tracking and 3D visualization. In simple terms DSI-Studio can be used to calculate and reconstruct characteristic fiber bundles between anatomical regions of interest (ROI). DSI-Studio is an ROI-based tractography tool. Regions are defined as a set of voxels stored by their coordinates. Deterministic tractography is a maximum likelihood estimation approach for the fiber tracts, with high sensitivity parameters (high SNR, high anisotropy, small step size) it generates reliable fiber pathway trajectories (Côté et al., 2013; Christidi et al., 2016). There are three different types of regions that inform tract shapes. The combination of multiple regions can be used to control grouping fibers when identifying certain tract bundles or complicated white matter structures. Regions can be created either by hand drawing or by using segmentation methods. Types of regions includes:

Seed Region: Seed regions serve as starting points for the clustering. Seeding voxels then are converted to starting points by the tracking algorithm.

Region of Interest (ROI Region): These regions lead the fibers to where anatomical fields are. The streamlines come from the starting points and are filtered by ROI regions. If fields are not passing through these regions,

they will be eliminated from the bundle.

Region of Avoidance (ROA Region): In the literature, these regions are also commonly referred to as NOT region or exclusion region. It is used to eliminate unwanted streamlines.

For quantitative analysis of brain magnetic resonance (MR) images, one of the first and most common steps is to define corresponding structures across subjects. One of the most widely used approaches is the voxel-based approach, in which shapes and positions of all subjects are aligned such that anatomical locations can be referenced by x, y, z coordinates in a standardized space such as Montreal Neurological Institute (MNI) coordinates. An alternative way is based on structural segmentation. For example, we can define a region by grouping all voxels that belong to this region in each subject. Then, we can compare the relevant information, such as the number of voxels and the average voxel intensity, across different subjects. The second approach requires human knowledge about the definition of this region, since the regions of interest have to be determined manually. For instance, one approach is the automated image segmentation approach, called multi-atlas segmentation (MAS) (Iglesias et al., 2015). This approach also requires a large number of pre-segmented atlases which were not readily available before this project. Our team in the MASI Lab of Vanderbilt University built an available dataset culminating in the TractEM project.

The TractEM is a fast protocol set for whole brain deterministic tractography with reference to the Eve labeling protocol (Oishi et al., 2010). TractEM is composed of datasets that include real human acquisitions, along with careful manual selection and definitions of ground truth streamlines. These protocols define 61 unique pathways on multiple subjects, process these pathways multiple times, and provide not only white matter volumes (or labels), but also fiber bundles themselves, which is not provided in other open source datasets (Wasserthal et al., 2018; Zhang et al., 2018). For my project, the TractEM protocols are effective and valid guidance for region drawing. TractEM protocols define 61 unique pathways (53 unique pathways, 22 of which have bilateral pairs of homologous tracts (22×2=44)), 5 commissural pathways, 3 of the brainstem tracts and the optic tract) and divide these pathways into simple and complex tracts according to the complexity of the protocols. TractEM protocols provide guidance not only on the number and approximate location of ROI and ROA regions, but also the tracking parameters used in DSI-studio. By following TractEM protocols on multiple subjects and processing them multiple times, a dataset will then contain manually drawn regions (ROI and ROA), density files(the number of streamlines go through each voxel) and streamlines that themselves can that can be produced by DSI studio.

The TractEM protocols is efficient and takes less than 6 hours for one subject. However, extracting fibers manually is still time consuming and requires related knowledge about brain anatomy. Therefore, the demand for a larger dataset creates a need for automated clustering methods. Additionally, a large manual dataset has inevitably lots of human errors. It means that cleansing this kind of dataset would be costly. This study is an extension of the TractEM project. By applying multi-atlas segmentation (MAS) on 58 whole brain pre-segmented atlases, I tried to find a fully automated method to reconstruct 31 white matter tracts defined by TractEM protocols.

Chapter 2. Materials

2.1. Subject

DTI data from a total of 38 subjects were collected from the Baltimore Longitudinal Study of Aging (BLSA) (Ferrucci, 2008) and the Human Connectome Project (HCP) (Van Essen et al., 2012). There are two BLSA datasets used in this project. The BLSA dataset contains 28 subjects with ages ranging from 57 to 77 years old. The HCP dataset used in this project includes 10 subjects with an age group of 26 to 36 year olds.

2.2. Image Acquisition

Each BLSA subject has a T1-weighted structural MP-RAGE with following parameters: number of slices = 170, voxel size = 1.0×1.0×1.2mm^3, reconstruction matrix = 256×256, flip angle = 8 degrees and Repetition Time (TR)/Echo Time (TE) = 6.5ms/3.1ms and two DWI acquisitions. For each diffusion-weighted imaging (DWI) data, a 3D spin-echo diffusion-weighted EPI sequence was used with these following parameters: Repetition Time (TR)/Echo Time (TE) = 7454 ms/75 ms, a total of 170 slices for 32 directions using b value = 700s/mm2 per sequence, voxel size = 0.81×0.81×2.2mm3, reconstruction matrix=320×320, flip angle = 90 degrees, acquisition matrix=116×115, field of view = 260×260mm.

Each HCP subject has a T1-weighted structural MP-RAGE as well. For each diffusion-weighted imaging (DWI) data, a 3D spin-echo diffusion-weighted EPI sequence was used with Repetition Time (TR)/Echo Time (TE) = 5520 ms/89.50 ms. Each sequence consists of six b0 and 90 32 directions all using 3 shells of b=1000, 2000, and 3000s/mm2 interspersed with an approximately equal number of slices on each shell within each run with the following parameters: number of slices = 111, voxel size = 1.25×1.25×1.25mm3, flip angle = 79 degrees, field of view = 210mm. During the preprocessing step, susceptibility correction (Andersson et al., 2003), eddy current (Andersson and Sotiropoulos, 2016) correction techniques, and b0 signal normalization were applied to the diffusion data.

2.3. Pre-Process

Head motion, eddy current effects, distortion (Van Essen et al., 2012), and b-table rotations (Schilling et al., 2019) already have been already corrected on the raw diffusion data. The T1 weighted images I will use in registration are co-registered to b=0, and aligned to the same space (Talairach space) and stored in the same coordinate. two high angular resolution diffusion imaging (HARDI) models for voxel-wise reconstruction were applied on each dataset. HARDI is the most efficient protocol that can be used to obtain the orientation information required to resolve crossing fiber issues (Susumu et al., 2013).

In this project, 4th order spherical harmonic Q-ball applied on the BLSA dataset. Q-ball imaging (QBI) uses a single shell (single b-value) to estimate the dODF to represent the fiber orientation distribution. Generalized Q-sampling (GQI) reconstruction method was used for multi-shell HCP dataset. GQI can utilize multi-shell (multi-b-value) data and is the most effective way to similarly estimate the dODF with fairly dense sampling of directions (Yeh et al., 2010).

2.4. Atlas Generation

As I mentioned in the introduction, DSI-studio is a ROI based deterministic diffusion fiber tracking and 3D visualization software. In this project, the atlas was produced by DSI-Studio with manual ROI, Seed and ROA regions guided by TractEM protocols. TractEM protocols were developed on 10 subjects from the HCP dataset and 10 subjects from BLSA dataset. The manual dataset based on these 20 subjects were used to evaluate TractEM protocols. For my project this dataset is used as a training dataset. And 18 BLSA dataset was a expansion of the 10 BLSA dataset. 10 BLSA subjects have been manually labeled by two raters and the other have been manually labeled by only one rater. Overall there are 58 atlases were used in my study.

2.5. Data Scale

The dataset provided by TractEM includes 3660 manually identified tracts reduced to 35 tracts ((anterior commissure(ac), anterior corona radiata (acr left and right), anterior limb internal capsule (aic left and right), body corpus callosum (bcc), cerebral peduncle (cp left and right), cingulum cingulate gyrus (cgc left and right), cingulum hippocampal (cgh left and right), corticospinal tract (cst left and right), fornix (fx left and right), fornix stria terminalis (fxst left and right), genu corpus callosum (gcc), inferior cerebellar peduncle (icp left and right), inferior fronto-occipital fasciculus (ifo left and right), inferior longitudinal fasciculus (ilf left and right), medial lemniscus (ml left and right), midbrain(m), middle cerebellar peduncle (mcp), olfactory radiation (olfr left and right), optic tract (opt), pontine crossing tract (pct), posterior corona radiata (pcr left and right), posterior limb internal capsule (pic left and right), posterior thalamic radiation (ptr left and right), sagittal stratum (ss left and right), splenium corpus callosum (scc), superior cerebellar peduncle (scp left and right), superior corona radiata (scr left and right), superior fronto-occipital fasciculus (sfo left and right), superior longitudinal fasciculus (slf left and right), tapetum corpus callosum (tap left and right), uncinate fasciculus (unc left and right), frontal lobe (fl left and right), temporal lobe (tl left and right), parietal lobe (pl left and right), occipital lobe (ol left and right)). However, in my project I did not focus on lobe tracts, because the drawing process of 3D lobar masks does not involve human subjective judgment. The aim of my project was to find a valid method to draw the region of interest to fit into its anatomical definition for the pathways.

Chapter 3. Methodology

As I mentioned, all data I used in this project came from the TractEM project (https://my.vanderbilt.edu/tractem/). The TractEM is a fast protocol set for fiber tracking and takes less than 6 hours for each subject with minimal neuroanatomical expertise required. However, it is a manual labeling protocol. For future work based on TractEM, we need more data requiring lower cost. Therefore, my project aims to find a valid basic method to label ROI regions automatically without neuroanatomical expertise and raters. The density image we gained by DSI studio contains not only the density information of each voxel, but also the shape of the whole fiber bundle. The binary density image is a mask covering the whole fiber bundle and can work as a 3D ROI region (or label) in DSI studio. For this project, I designed a basic multi-atlas framework to automatically draw 3D ROI regions based on the density image of the fiber bundle. In traditional image segmentation, the deformable may be non-rigidly registered to the target image to be labeled. The final label can be obtained by mapping the training label onto the target image. For this multi-atlas case, the same process makes sense. Multiple atlases and registrations can be used to transfer multiple training labels onto one target image. The final label can be obtained by assigning different weights to multiple atlases during a process called label fusion. Label fusion has two advantages: 1) it is easier to accommodate large individual variations in anatomy if one does not have to rely on a single atlas; 2) multiple registrations improve the robustness and reduces occasional registration failures and non-global minima of the registration cost function. The same idea can also improve voxel-based or tensor-based morphometry (Lepore et al., 2008). Therefore, the main idea of my project has four sections: 1) registration, 2) label propagation, 3) label fusion (merging and voting), and 4) fiber tracking.

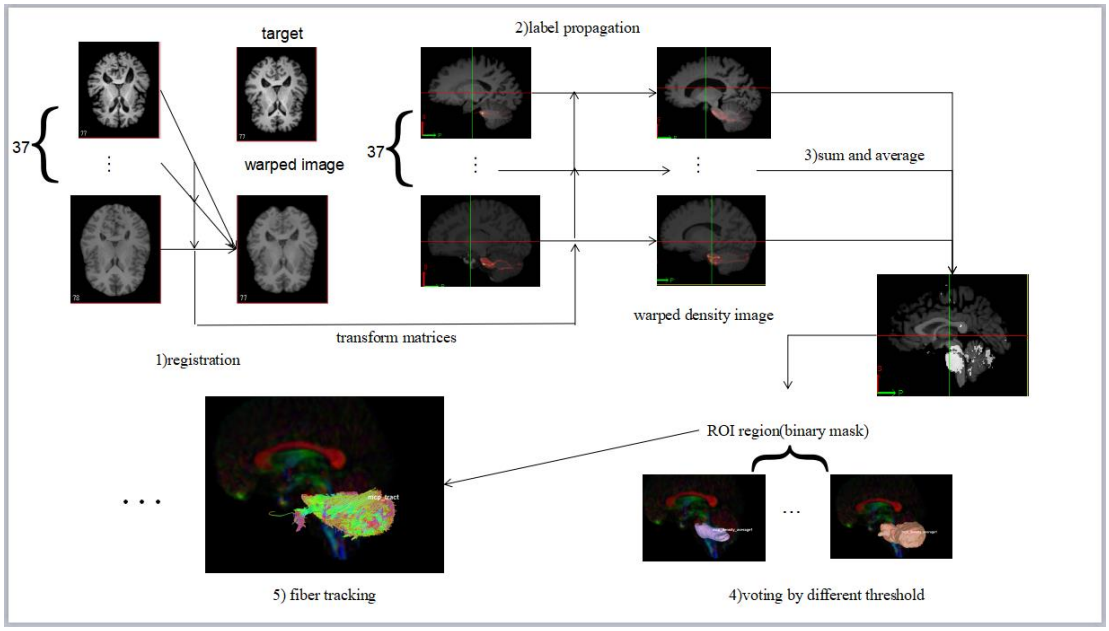


Figure 1. 1) Warp T1 weight structure images of all subjects by ANTs registration (symmetric image normalization method (SyN)) and then obtain transform matrices between each pair of them. 2) Map labels of atlas onto target subject. 3) Sum all warped labels and then average them. In this step all labels have the same weight. 4) Vote the probability map. After the average process the label matrix becomes a probability map. 5) Use binary labels as a 3D ROI region in fiber tracking.

3.1. Registration

Registration is a computationally complex task that aims to deform one of the images until it is similar to the other. Registration is the task of establishing spatial correspondence between images. During the registration each background of labeled image is registered with the target image. The labeled-to-target image transformation can then be concatenated with the labels-to-target transformations in order to propagate labels from the atlases to the target image. In biomedical image processing, the registration algorithm is always applied to structure images. In this step, I use ANTs registration (the symmetric image normalization method (SyN)) to warp a T1 image belongs to one subject to the other subject' s T1 image in each pair of subjects. The SyN algorithm also produces the inverse transform matrix simultaneously. Therefore, for each pair of subjects two times registration is unnecessary. I executed 703 registrations in a total of 38 subjects (10 BLSA, 10 HCP and 18 BLSA).

3.2. Label Propagation

When the spatial correspondence is established, the classical multi-atlas segmentation strategy proceeds by propagating the atlas labels to the novel image coordinates (Iglesias et al., 2015). In this process, I use density images as atlas labels, where each voxel sorts the density information of fiber bundles. The density image not only the density information of each voxel, but also the shape of the whole fiber bundle. In fact, they are not conventional labels because they are not binary. Since the T1 weighted image is not binary, I assumed that the transform matrix applies to warp one T1 image to the other T1 image also could be applied to density images.

3.2.1. Coordinate Transformation

Before label propagation, coordinate transformation should apply to the density images produced by DSI-Studio. In the registration process the moving image and fixed image are both in the LAS (L left, R right, A anterior, P posterior, I inferior, & S superior) coordinate with center point (-78,112,7). However, the density, ROA, ROI and seed images produced by DSI-Studio are all in RAS coordinate with center point (0, 0, 77). When I used the ANTs transform (antsApplyTransforms) to warp atlas labels (density images) onto the target image by the transform matrix produced by the SyN algorithm, I gained a zero image. Because of the mismatching orientations and mismatching centre point, ANTs registration cannot rotate the subject's image 180 degrees to match the template orientation. The registration starts by matching the center of mass (default) and then later will run the other registration steps (rigid, affine, SyN). Therefore, in the ANTs space these two images under different coordinates have no overlap. In this step I transfer all atlas labels to LAS coordinates with center point (-78,112,70) by NiBabel library in Python.

3.2.2. Label Propagation

In this part, I followed the leave one out method warping all atlas labels from 37 atlas to one target subject. I repeated the same transformation process 38 times for each subject. The ANTs transform command antsApplyTransforms, is the algorithm I used to warp labels. The main idea of this label propagation is letting the

labels go through the way that the atlas went before.

3.3. Label Fusion

Label fusion is the step of combining propagated atlas labels. There are two branches of my plan with different weight assignments. 10 BLSA and 10 HCP subjects have been manually labeled (ROI, seed and ROA) by two raters and the other have been manually labeled by only one rater. The first branch is assigning equal weight to each rater. In that way the weights assigned to 37 subjects are different. For subjects in 10 BLSA or 10 HCP the weight is 2/56, while for each subject of BLSA 18 the weight is 1/57. Another branch is assigning equal weight (1/37) to each subject. In that way, I merged labels from two rates before assigning weight to multiple atlas.

3.3.1. Normalize Images

In density images, the values of each voxel are on drastically different scales. The maximum may be over 50000 while the minimum is 0. The goal of normalization is to make sure the value in each voxel has the same scale so that the normalized image can work as a probability map. In the fiber tracking process, all streamlines that pass through the ROI region are filtered in. However, when the 3D ROI region is larger and covers more voxels than the 2D ROI region defined by TracEM, more unwanted and unrelated streamlines are clustered into the fiber bundle. In that case, I through that selecting the principle (densest) part of the fiber bundle might be a reasonable plan. Setting a threshold for a probability map can effectively select the densest part of the fiber bundle. Min-max normalization (Formula (1)) is one of the most common ways to normalize data. After min-max normalization the density in each voxel is scaled to have a value between 0 to 1.

$$\frac{X_i - \min(X)}{\max(X) - \min(X)} \qquad (1)$$

3.3.2. First Branch

First, I created the binary label by applying the threshold to the normalized density image. Second, I summed all atlas judged by different raters respectively and then average the summed label. For subjects in 10 BLSA group or 10 HCP group the weight is 2/56, while for each subject of BLSA 18 the weight is 1/57. The threshold of binary label is 0.05 and 0.0.

3.3.3. Second Branch

First, I averaged two normalized density images of the same subject from two raters. Next, I create a binary label by applying the threshold to the averaged normalized density image. Finally, I summed these binary labels and then averaged the merged label. In this way, due to each subject having a unique merged atlas, the weight assignment is uniform (1/37) for each target subject. the threshold of binary label is 0.05 and 0.0.

3.4. Voting on Averaged Labels

The output (averaged labels) from the last step is not binary. Since the ROI regions I would use in the fiber tracking processing must be binary masks, I voted on these probability maps during the binary process. For the first branch, while I gave the same weight to each atlas, there are 56 or 57 people taking part in each election. In that case, when the value of one voxel is greater than 0.5, it means that half of the voters agree on this voxel. In this step, participants have the same number of votes and voting weights. I selected several thresholds ranging from 0 to 1 to filter eligible voxels that can form binary masks.

3.5. Fiber Tracking

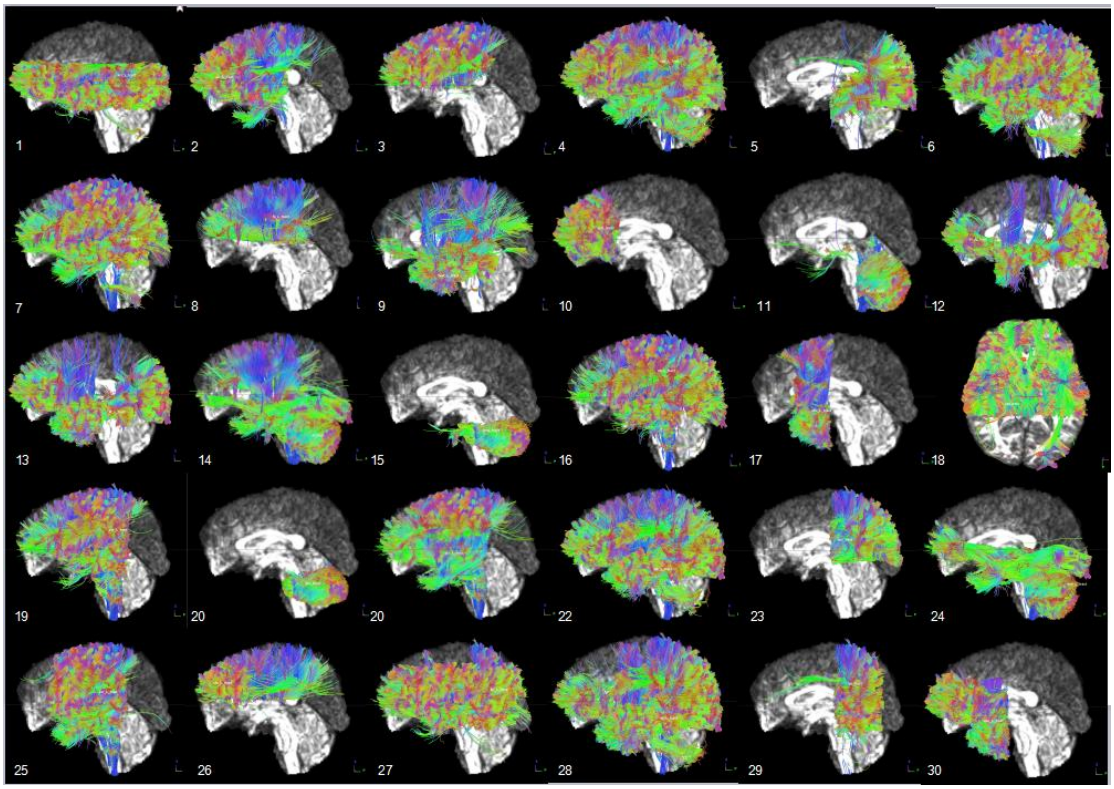
As I mentioned in the above part, because the 3D regions are larger and cover more voxels than the 2D ROI regions, ROA regions are necessary during the fiber tracking process. I used 3D masks created by the label fusion process as ROI regions to filter whole-brain tractography. If no seed regions are set, DSI-Studio would use the whole brain as the seed region. The ROA regions I used in the fiber tracking process were produced by the first branch of the label fusion process. On account of low agreement between raters, the ROA created by multi-atlas segmentation is not perfect and does not work well for all 31 tracts. I will discuss this situation in the result section. In this part I used DSI-Studio commands to automatically execute fiber tracking. In this project I set the tracking parameters in accordance with TracEM protocols. In the TractEM project the tracking parameters are termination index normalized QA (nqa), smoothing is 1, tract count is 100K, track length is 30-300mm, and FA threshold is 0.1 which is used to ensure the maximum number of fibers are produced (Bayrak, et al., 2019) This is a suggested minimum while lower values may cause false fibers (http://dsi-studio.labsolver.org). Angular and FA thresholds are the determining parameters of how the fibers are tracked from the region of interest. Angular threshold is set to zero, therefore stopping criteria are not determined by this parameter; instead the length constraint or end regions serve as an end condition. The tracking algorithm is streamline (Euler), seed orientation is primary, and the direction interpolation is set to trilinear. Randomize seeding is off. This parameter sets the random seed generator to a constant to ensure tracking results are identical, and reproducible. Because the time cost of each tract is an important parameter in fiber tracking, I set a threshold for timeout. If 100000 streamlines cannot be extracted in fifteen minutes for each individual tract, the fiber tracking process of this tract will be killed.

Chapter 4. Result

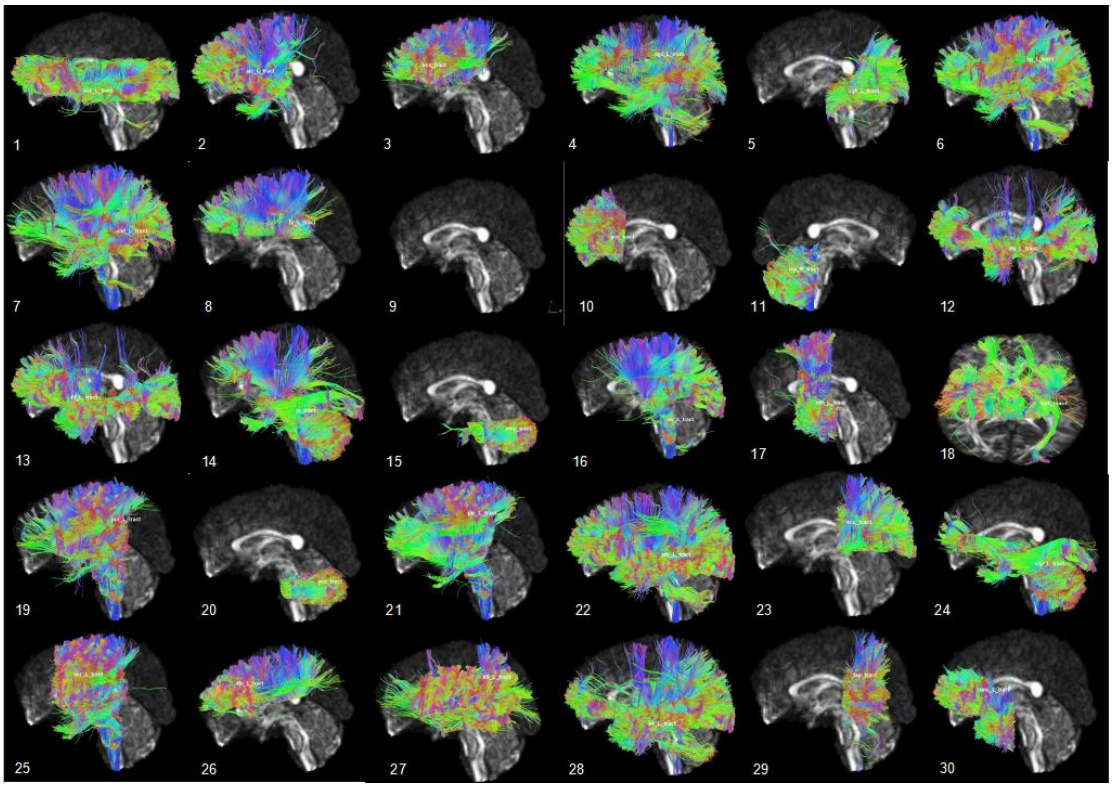
By testing the two branches proposed above, I created a new dataset based on 38 subjects. For each weighting approach I applied two thresholds 0.05 and 0.0 to the normalized density images and three voting thresholds 0.1, 0.3, and 0.5 to the averaged labels. Evaluation of these two branches was performed in 38 subjects where each subject was traced twelve times by different thresholds and reproducibility was assessed both qualitatively and quantitatively. This section shows the evaluation procedure through-out the development and validation of my methods.

4.1. Qualitative Analysis

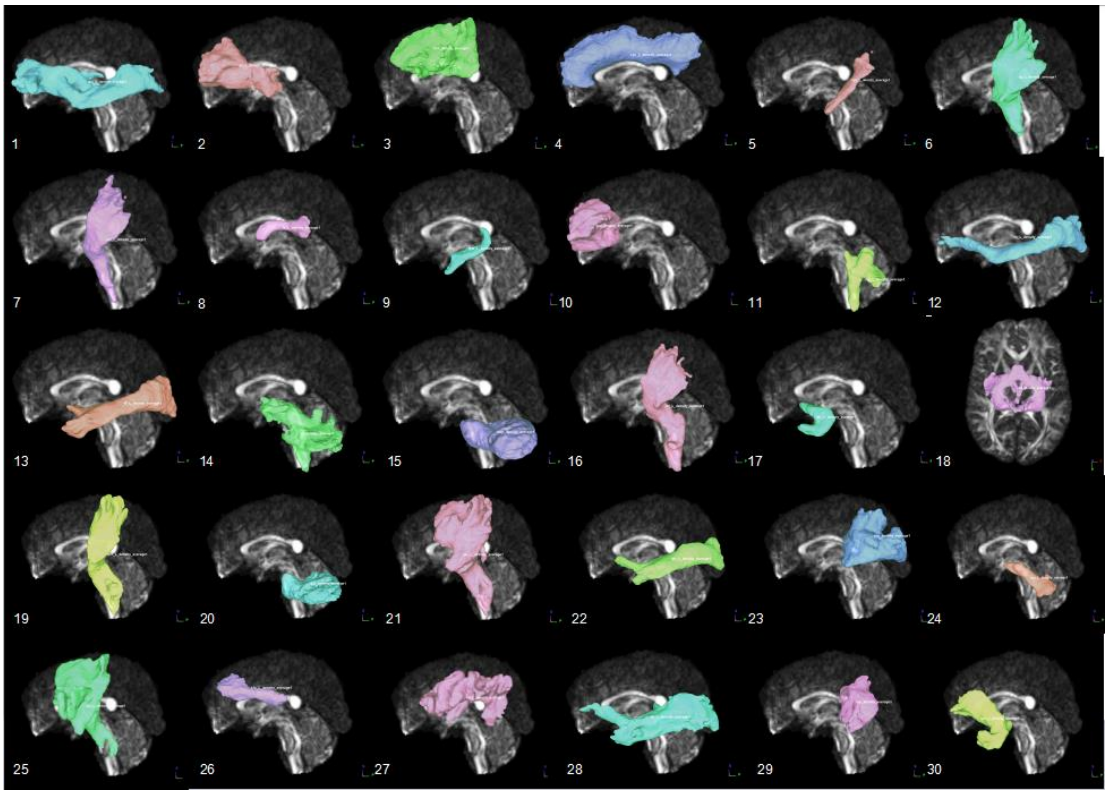
In qualitative analysis, I put the resulting fiber bundles into the corresponding background (subject). The characterization of individual fiber bundles and the difference between automatic fiber bundles and raters' judgement can be inferred from these images, and the tracts can be investigated by their behavior. I also show the ROA regions (Figure2 (e)) produced by multi-Atlas segmentation (first-branch label fusion) in the form of graphical interface through DSI-Studio. The validation of ROA regions can be judged by eye through comparing new ROA regions with hand-drawn ROA regions. Figure2 (e) shows that the ROA regions created through the multi-Atlas segmentation method is 3D and rough. The standard ROA regions required by TractEM protocols are planes and located on certain slices. However, there are many discontinuities with large voids in my ROA regions. ROA regions are used to eliminate unwanted streamlines, meaning that these large voids in ROA regions allow the streamline to extend out of the exclusion region. It creates more unwanted overlap and confusion between multiple tracts. Figure2 (a) shows streamline representations of individual tracts for the all 30 tracts of a single BLSA subject. The aim of this project is to create a fully automated tool for labeling valid regions of interest which can be used in DSI-Studio. Additionally, the streamlines of individual tracts produced by DSI-Studio have to correspond to the 31 manual labeling tracts defined by TractEM protocols. Thus, I also demonstrate the tract in the form of visualization through DSI-studio in four structural categories (1) tracts in the brainstem (Figure3), (2) fibers that connect cerebral cortex to the central regions and to the other end of the spinal cord (Figure4), (3) association fibers of cerebral cortex (Figure5), (4) commissural fibers that connect the two hemispheres (Figure6). Figure2 (b) and (d) show that small thresholds create smaller ROI regions and then smaller ROI regions produce smaller pathways. However, Figures 3-6 demonstrate that although smaller thresholds (th=0.05) create smaller pathways with less unwanted overlap and confusion between multiple tracts, there are many excrescent streamlines which do not belong to the particular tract defined by TractEM.


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
20
22
23
24
25
26
27
28
29
30

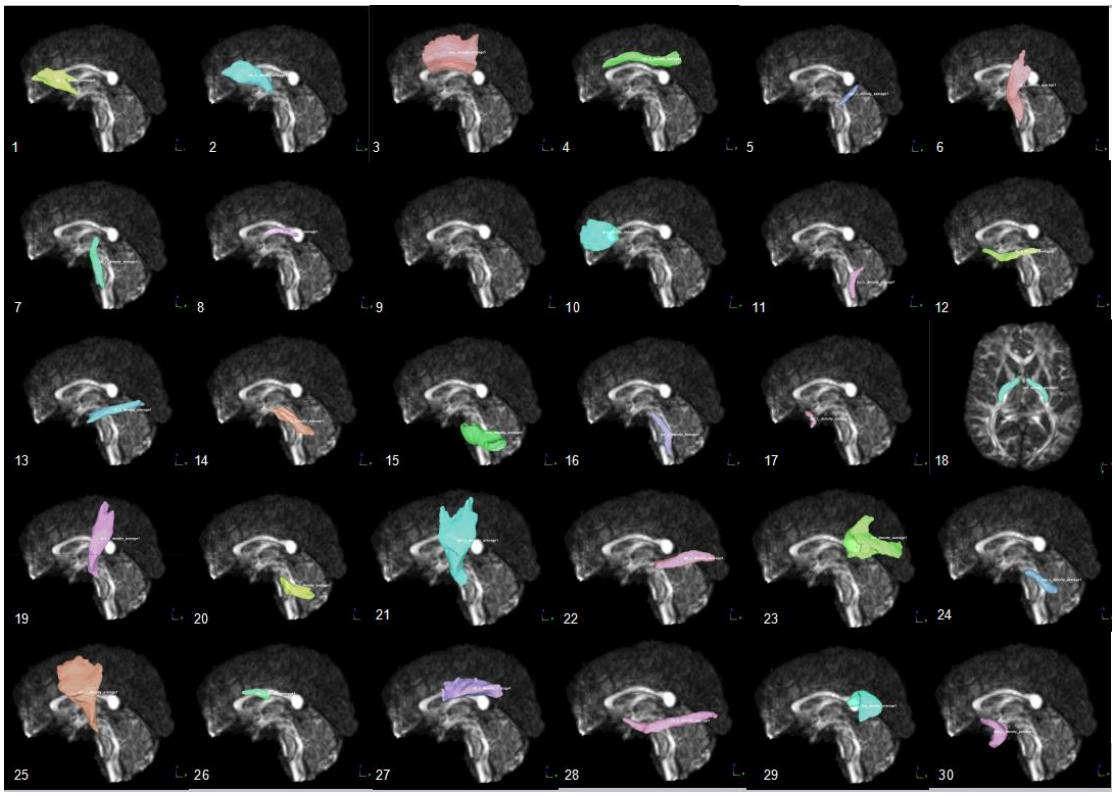
(a)


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

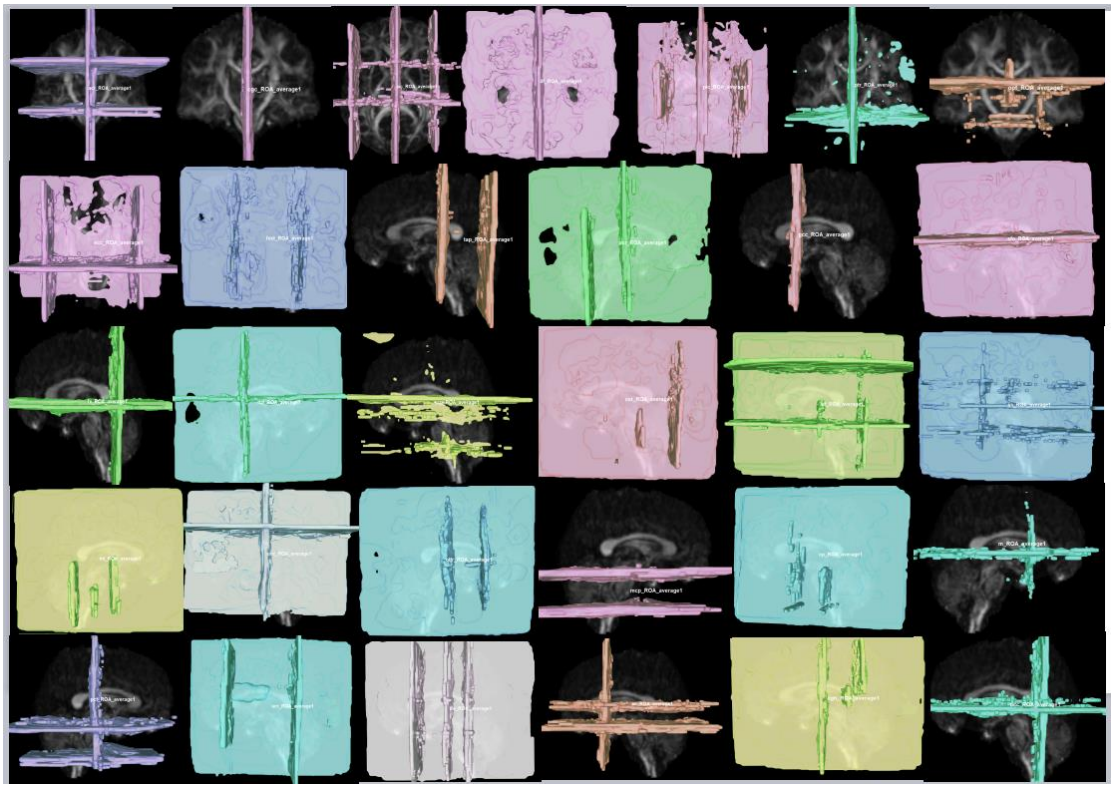
(b)


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

(c)


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

(d)



(e)

Figure 2. (a) Complete protocol streamline visualization for a single BLSA subject (th=(0.0,0.5)), (b) Complete protocol streamline visualization for the same BLSA subject (th=(0.05,0.5)), (c) Complete protocol ROI regions used for fiber tracking visualization for the same BLSA subject (th=(0.0,0.5)), (d) Complete protocol ROI regions used for fiber tracking visualization for the same BLSA subject (th=(0.05,0.5)), (e) Complete protocol ROA regions used for fiber tracking visualization for the same BLSA subject (th=(0.1)). (1) anterior corona radiata, (2) anterior limb internal capsule, (3) body corpus callosum, (4) cingulum cingulate gyrus, (5) cingulum hippocampal, (6) cerebral peduncle, (7) corticospinal tract, (8) fornix, (9)fornix stria terminalis, (10) genu corpus callosum, (11) inferior cerebellar peduncle, (12) inferior fronto-occipital fasciculus, (13) inferior longitudinal fasciculus, (14) midbrain, (15) middle cerebellar peduncle, (16) medial lemniscus,(17) olfactory radiation, (18) optic tract, (19) posterior corona radiata, (20) pontine crossing tract , (21) posterior limb internal capsule, (22) posterior thalamic radiation, (23) splenium corpus callosum , (24) superior cerebellar peduncle , (25) superior corona radiata , (26) superior fronto-occipital fasciculus , (27) superior longitudinal fasciculus, (28) sagittal stratum , (29) tapetum corpus callosum, (30) uncinate fasciculus

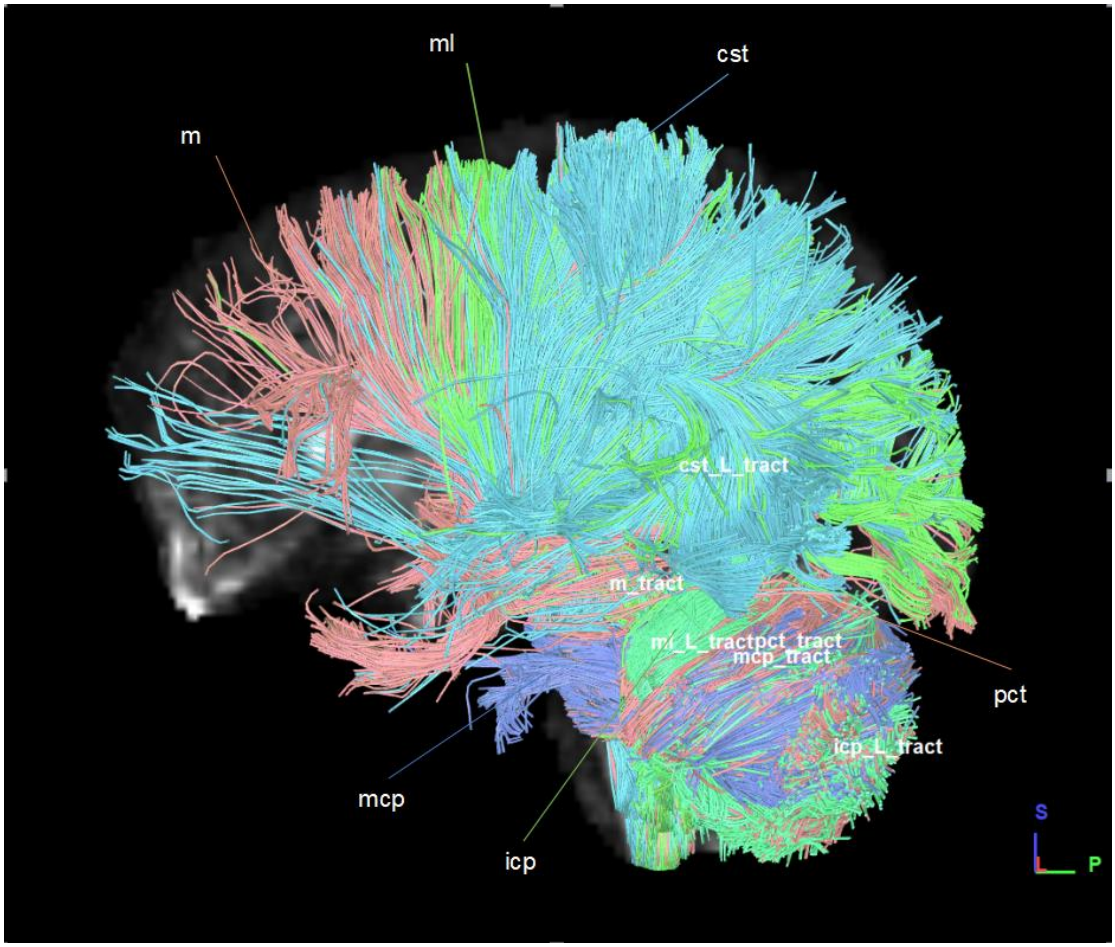


Figure 3. Tracts in the brainstem; corticospinal tract (CST), medial lemniscus (ML), midbrain (M), pontine crossing tract (PCT), inferior cerebellar peduncle (ICP), middle cerebellar peduncle (MCP), and superior cerebellar peduncle (SCP). The fibers of different colors coalesce into independent tract. According to the anatomical structure of white matter, these characteristic tracks should cover different parts with little overlap. However, this figure shows the opposite pattern.

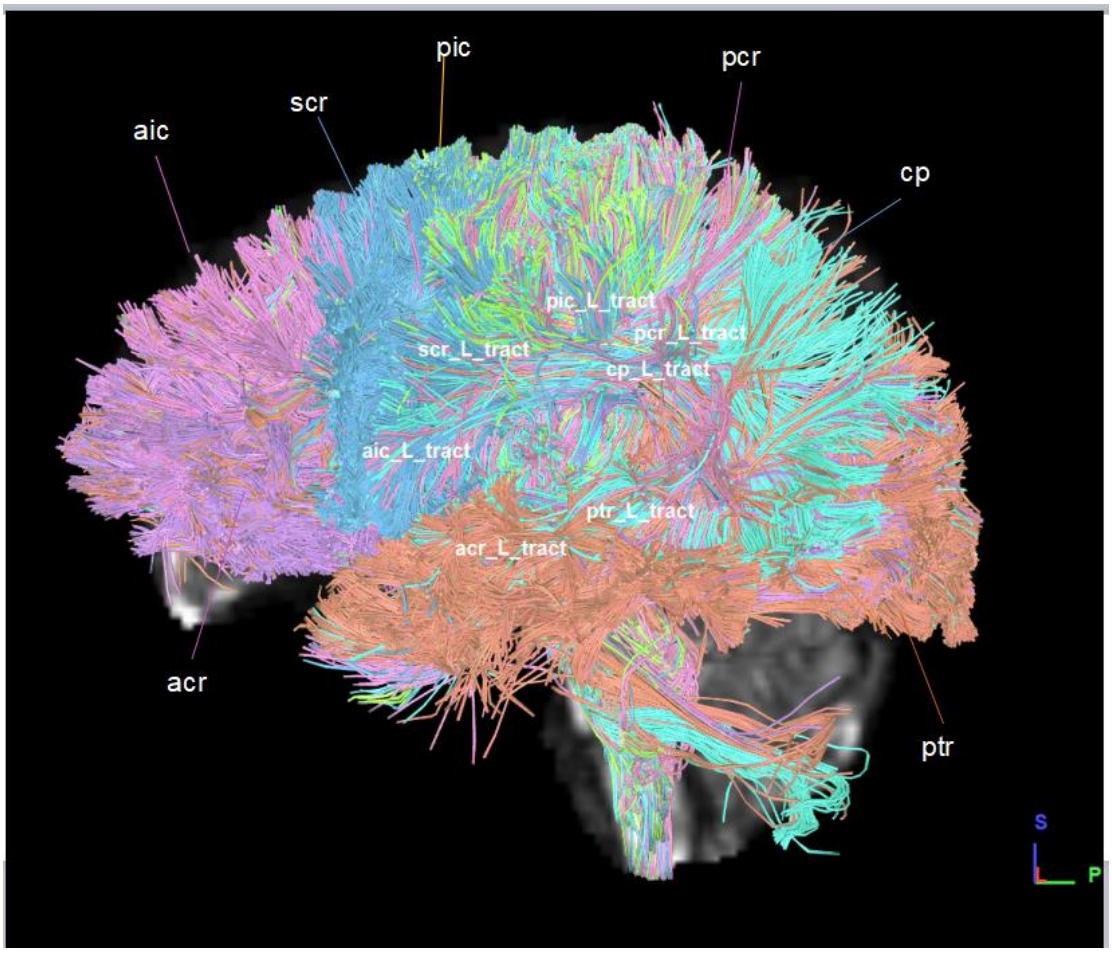


Figure 4. Projection fibers; anterior corona radiata (ACR), superior corona radiata (SCR), posterior corona radiata (PCR), posterior thalamic radiation (PTR), anterior limb of internal capsule (AIC), posterior limb of internal capsule (PIC), cerebral peduncle (CP). The fibers of different colors coalesce into independent tract. According to the anatomical structure of white matter, these characteristic tracks should cover different parts with a certain degree of overlap. However, this figure shows a more complicated pattern. It is difficult to differentiate these tracts in most areas and they also cover areas they are not supposed to.

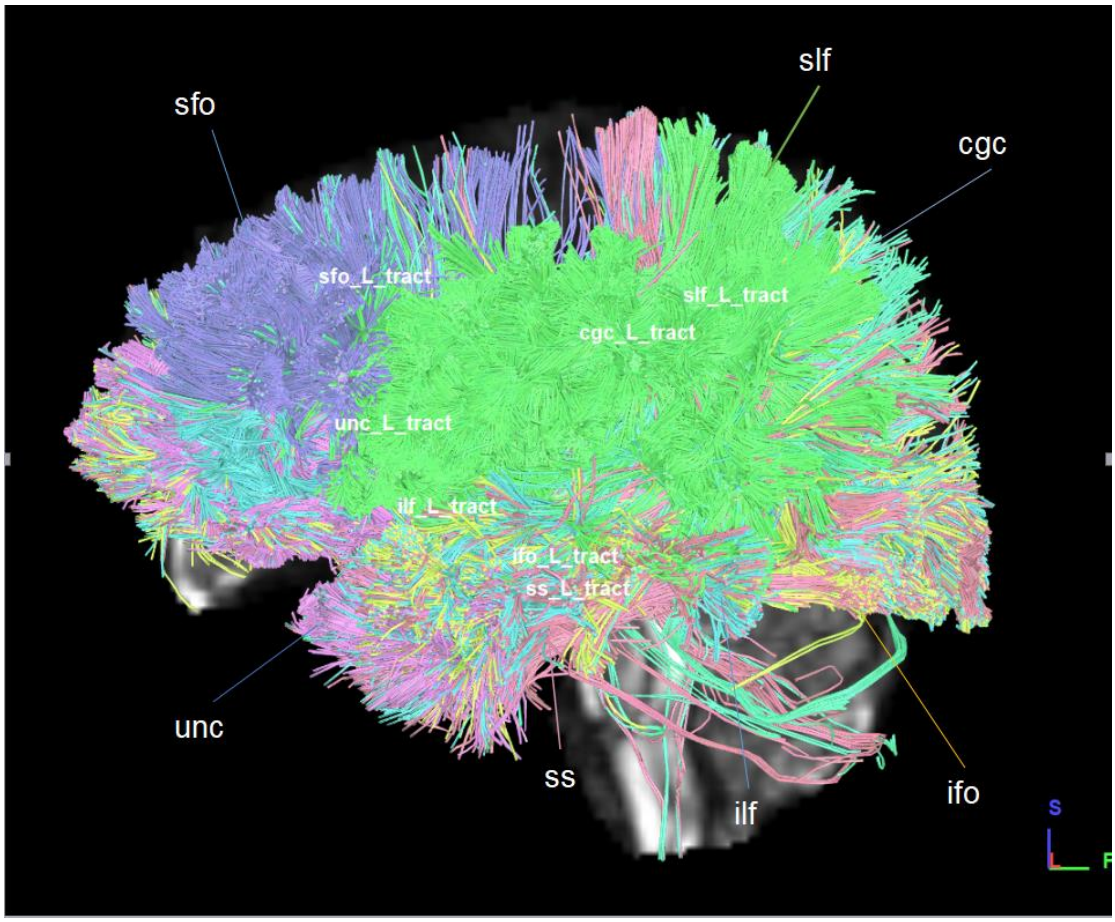


Figure 5. Association fibers; superior fronto-occipital fasciculus (SFO), inferior fronto-occipital fasciculus (IFO), superior longitudinal fasciculus (SLF), inferior longitudinal fasciculus (ILF), sagittal stratum (SS), uncinate fasciculus (UNC), cingulum cingulate gyrus (CGH). The fibers of different colors coalesce into independent tract. According to the anatomical structure of white matter, these characteristic tracks should cover different parts with a certain degree of overlap. However, this figure shows a more complicated pattern. It is difficult to differentiate these tracts in most areas and they also cover areas they are not supposed to.

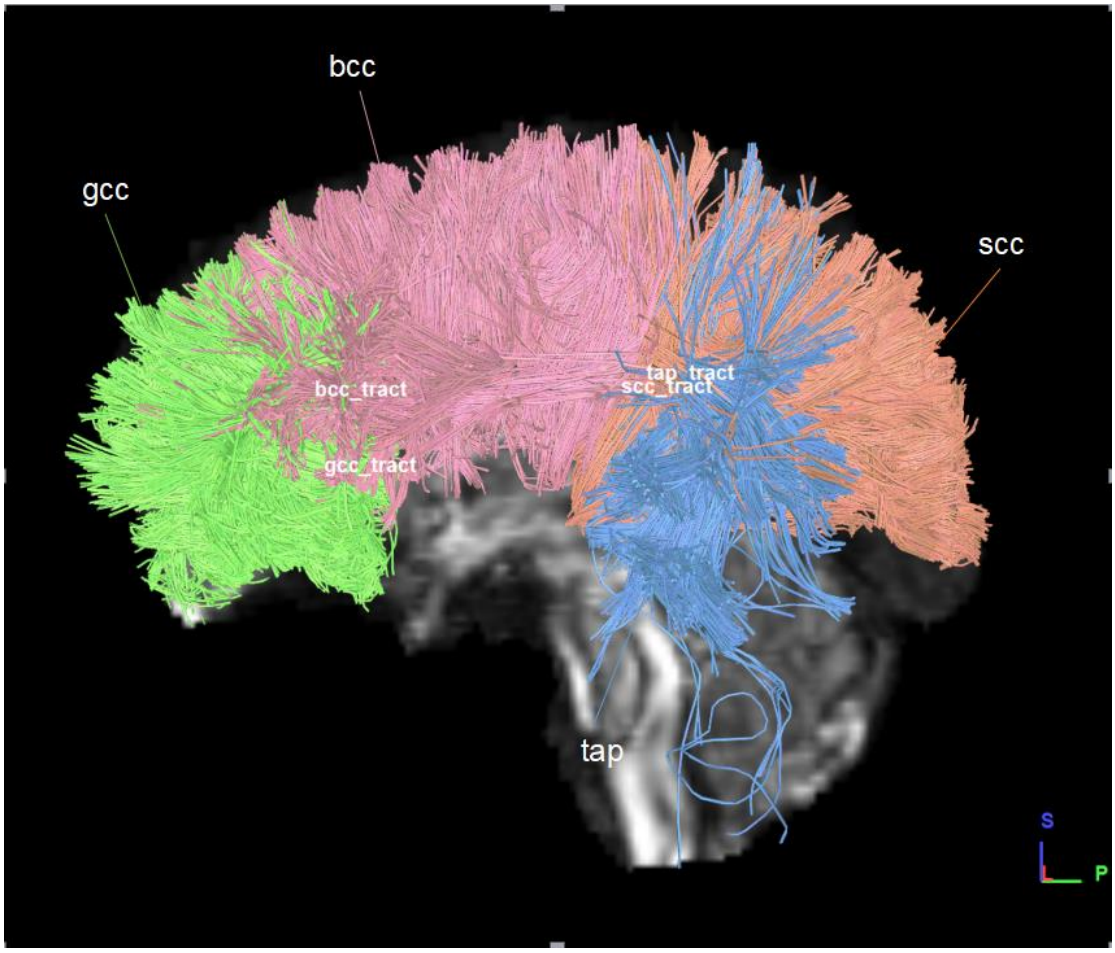


Figure 6. Commissural fibers; genu of the corpus callosum (GCC), body of the corpus callosum (BCC), splenium of the corpus callosum (SCC), and tapetum (TAP). The fibers of different colors coalesce into independent tract. According to the anatomical structure of white matter, these characteristic tracks should cover different parts with little overlap. This figure shows a similar pattern.

4.2. Quantitative Analysis

For this analysis, the fiber bundles filtered by the automatically merged ROI regions are regarded as produced by a new rater. To evaluate the new rater' s job, I calculate DICE between new rater and old rater1, then old rater2 respectively. And I also calculate DICE between old rater1 and old rater2 as a reference. And under a different threshold group I also describe the reproducibility of each branch.

4.2.1. Inter-Rater Tract Reliability

This analysis aims to evaluate rater agreement for all individual tracts. The score is calculated by calculating similarity degrees across all subjects for each tract. I use Dice correlation coefficient (DICE) as a valid metric in this analysis.The Dice correlation coefficient is a spatial overlap matrix. The Dice coefficient is defined as

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \qquad (2)$$

where A and B represent binary matric with values 0 and 1 for each voxel and computed as follows:

$$|A \cap B| = \sum_i a_i b_i \qquad (3)$$

where |A|, |B| are the cardinalities of the two sets, and a_i, b_i is the values of each voxel of image A and B. Therefore, the result of the formula shows the ratio of overlap between two mask images.

Each boxplot (Appendix A, Figure A1-A27) shows the median, maximum, minimum and quantile of DICE. In order to assess inter-rater tract reliability under different threshold groups, I list two tables (Table 1 and Table 2) to group all tracts based on medians of DICE scores corresponding to each tract. Table 1 describes the quantitative result of the first method branch and Table 2 shows the result of the second method branch. According to these two tables, these two weighting approaches have similar reliability on fiber tracking. When the first threshold is equal to 0.05 almost all tracts have higher reliability, and both HCP and BLSA show the same trend. However, it also shows that the lower threshold allows fewer reproducible region-specific fiber bundles to reproduce within fifteen minutes. When the second threshold is equal to 0.5, the reproducible tracts also perform higher reliability. At the same time, the lower second threshold also shows the same trend, that is fewer tracts are reproduced with fifteen minutes. Tracts which cannot be reproduced within fifteen minutes are anterior commissure (ac), cingulum hippocampal (cgh left and right), fornix (fx left and right), fornix stria terminalis (fxst left and right), inferior fronto-occipital fasciculus (ifo left and right), optic tract (opt), inferior fronto-occipital fasciculus (iof left and right), superior fronto-occipital fasciculus (sfo left and right) and olfactory radiation (olfr left and right). These tracts can be divided into two types, complex tract (opt, ifo, cgh) defined by TractEM protocols and small tracts (sfo, fx, fxst, ac, olfr). Sometimes, for small tracts, manually drawn regions cannot extract valid 100000 streamlines within fifteen minutes. For complex tracts, there are more than two ROA, seed and ROI regions that work together during the fiber tracking process. As I mentioned in the qualitative analysis process, the ROA regions created using a multi-atlas segmentation method are not valid for all tracts because raters could hardly reach a consensus on selecting ROA planes even for the same subject. Moreover, simple tracts showed generally higher reliability. Only simple tracts defined by TractEM protocols (mcp, aic, acr, bcc, gcc, scc and so on) can perform high reproducibility (DICE>=0.7). But this trend is not stable for all subjects across simple tracts. HCP subjects have better performance on high reliability ranges (DICE>=0.7).

Table 1. The quantitative result of the first method branch

number of tracts	median=0	median>=0.5	median>=0.6	median>=0.7	median>=0.8
HCP (th= (0.05,0.5))	13	28	17	12	1
HCP (th= (0.05,0.3))	5	19	13	4	0
HCP (th= (0.05,0.1))	1	13	5	0	1
BLSA (th= (0.05,0.5))	12	27	14	5	0
BLSA (th= (0.05,0.3))	7	21	8	2	0
BLSA (th= (0.05,0.1))	0	6	2	0	0
BLSA18 (th= (0.05,0.5))	15	28	9	4	0
BLSA18 (th= (0.05,0.3))	6	10	5	0	0
BLSA18(th= (0.05,0.1))	0	3	1	0	0
HCP (th=(0.0,0.5))	0	12	2	0	0

HCP (th= (0.0,0.3))	0	7	0	0	0
HCP (th= (0.0,0.1))	0	4	0	0	0
BLSA (th= (0.0,0.5))	0	6	0	0	0
BLSA (th= (0.0,0.3))	0	3	0	0	0
BLSA (th= (0.0,0.1))	0	2	0	0	0
BLSA18 (th= (0.0,0.5))	0	3	0	0	0
BLSA18 (th= (0.0,0.3))	0	3	0	0	0
BLSA18 (th= (0.0,0.1))	0	2	0	0	0

Table 2. The quantitative result of the second method branch

number of tracts	median=0	median>=0.5	median>=0.6	median>=0.7	median>=0.8
HCP(th=(0.05,0.5))	11	28	18	12	1
HCP (th= (0.05,0.3))	7	18	14	5	0
HCP (th= (0.05,0.1))	1	12	5	0	1
BLSA (th= (0.05,0.5))	11	27	15	5	0
BLSA (th= (0.05,0.3))	4	21	8	2	0
BLSA (th= (0.05,0.1))	0	6	2	0	0
BLSA18 (th= (0.05,0.5))	13	28	9	3	0
BLSA18 (th= (0.05,0.3))	6	12	5	1	0
BLSA18 (th= (0.05,0.1))	1	5	0	0	0
HCP (th= (0.0,0.5))	0	11	0	0	0
HCP (th= (0.0,0.3))	0	8	0	0	0
HCP (th= (0.0,0.1))	0	3	0	0	0
BLSA (th= (0.0,0.5))	0	5	0	0	0
BLSA (th= (0.0,0.3))	0	3	0	0	0
BLSA (th= (0.0,0.1))	0	3	0	0	0
BLSA18 (th= (0.0,0.5))	0	3	0	0	0
BLSA18 (th= (0.0,0.3))	0	2	0	0	0
BLSA18 (th= (0.0,0.1))	0	2	0	0	0

4.2.2. Tract Reproducibility

This analysis is made possible by comparing the reliability of each characteristic tract reconstracted by small regions with voting threshold 0.3 and 0.5. According to Appendix A, Figure A.28 I define two types of tracts: complex tracts and simple tracts. The DICE value of complex tracts is over 0.5 while the DICE value of simple tracts is lower than 0.5. By correspond this two types of tracts (Table3) with TracEM protocols I could draw the following results: 1) Most of complex tracts defined by complex ROA regions (the ROA regions cover more than two slices or have to be drawn by rater’ s subjective judgement), while most of simple tracts defined by simple ROA regions (the ROA regions cover one slice or lie in certain slice). 2) All small tracts (fxst, scp, fx and ifo) are complex tracts, since it is difficult to reconstruct these small tracts by manual regions.

Table 3. Tract reproducibility

Complex tracts （Complex ROA）	Simple tracts (Simple ROA)
ac, cgh, fxst, sfo, olfr, scp, ptr, ifo, ml, icp, cst, opt, cp, ss, ilf, m, fx	pct, acr, unc, tap, acr, aic, slf, scr, pcr, cgc, pic, scc, bcc, gcc, mcp

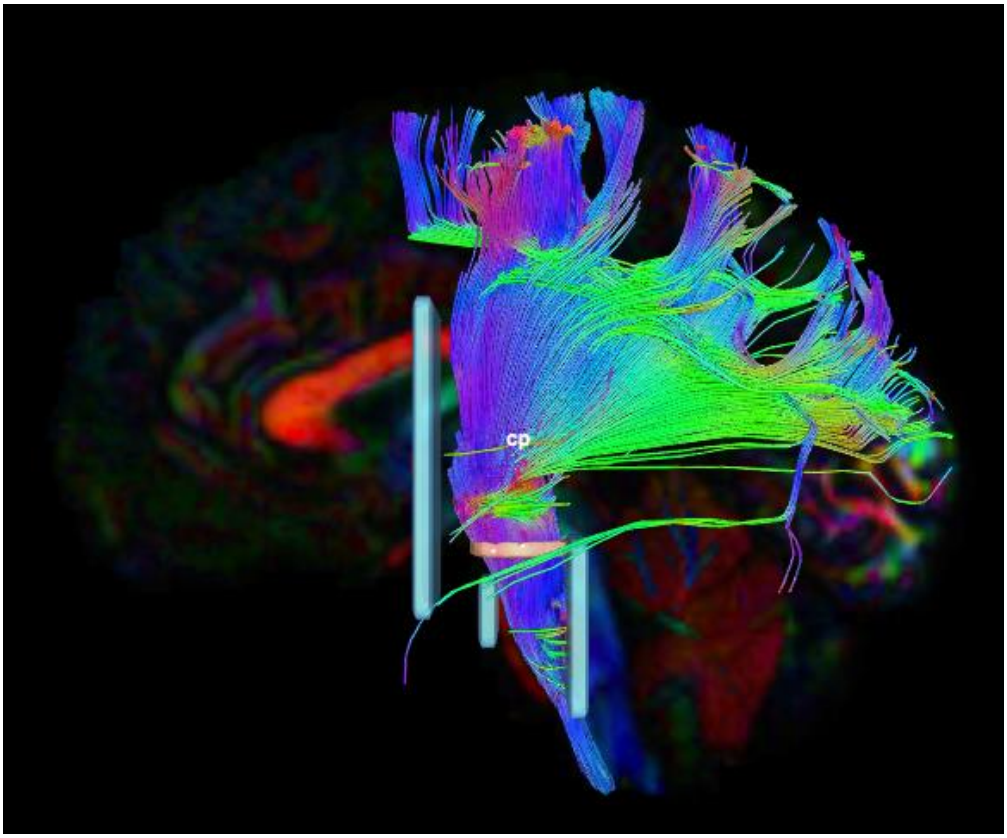


Figure 7. Complex tract (cerebral peduncle) with complex ROA regions

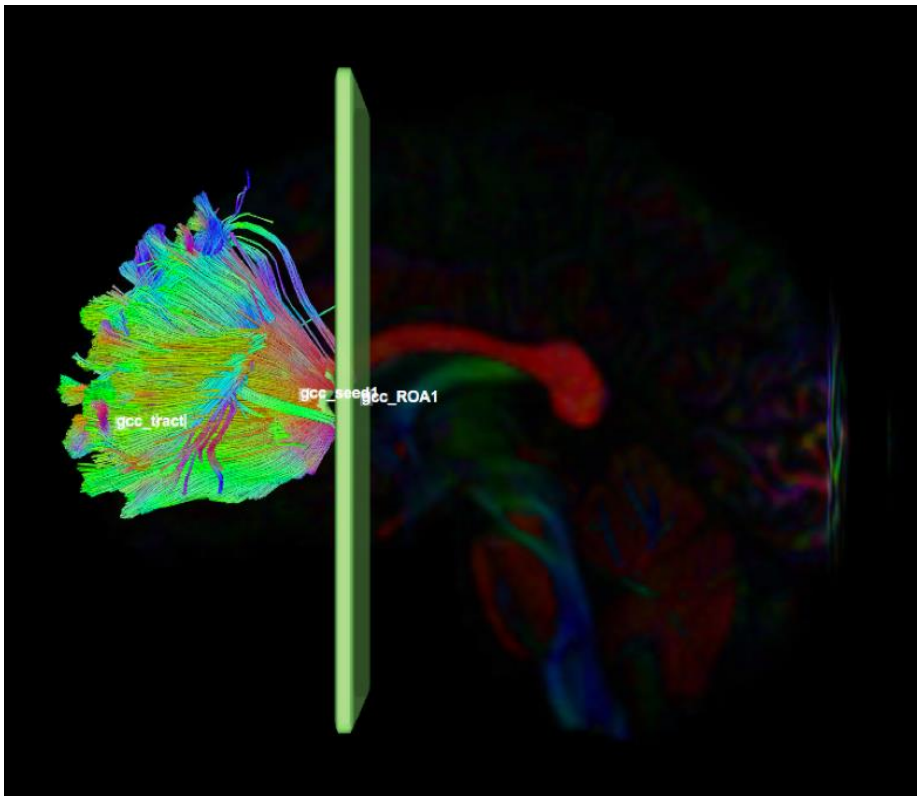


Figure 8. Simple tract (genu corpus callosum) with simple ROA regions

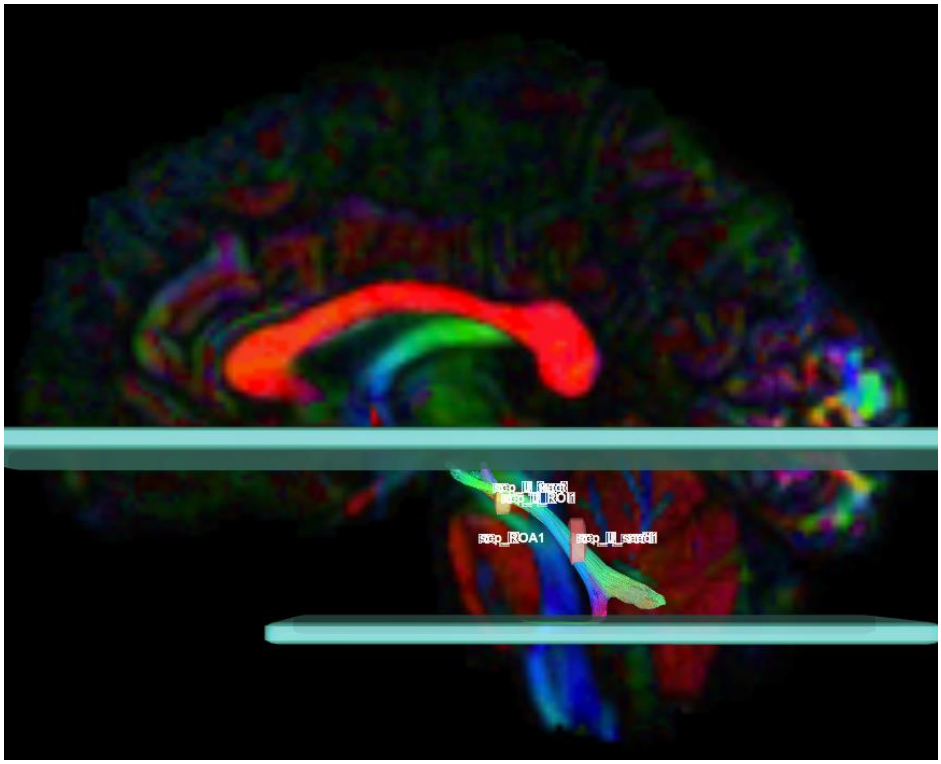


Figure 9. Small tract (superior cerebellar peduncle) with complex ROA regions

Chapter 5. General Discussion

Diffusion tractography offers tremendous potential for studying white matter architecture. Tractography is the only tool that we currently have that can be used to visualize white matter architecture in vivo and noninvasively. Qualitative and quantitative analysis of the estimated trajectories of various white matter tracts is crucial for advancing our understanding of abnormal brain anatomy. Reproducible tractography are essential in a backbone to the group-wise analyses and longitudinal testing studies for a comprehensive modern atlas are essential. Despite the importance, there are a limited number of reproducibility studies with a limited number of fiber pathways (Heiervang et al., 2006; Wakana et al., 2007; Besseling et al., 2012). The TractEM project proposed 35 protocols valid for 61 unique tracts. Additionally, there are low time costs required by TractEM protocols for example, it takes less than six hours each subject for whole brain tractography. TractEM protocols are manual labeling protocols that require raters to have technical and anatomical knowledge. Even people who do not have any technical and anatomical knowledge can draw regions of interest by following the protocols. However, humans make mistakes. Raters make unintended mistakes and subjective mistakes in the manual process. Because of these diverse errors we have to spend much more than six hours for each subject cleaning data. Therefore, it is important to find a automatic way to replace human raters.

TractEM project is relying on DSI-Studio which is a program for diffusion tensor image computation, region of interest (ROI) based deterministic diffusion fiber tracking and 3D visualization. Therefore, my job aims to produce a valid binary mask for each unique pathway which can be used as a ROI region in DSI-Studio. Furthermore, each ROI region has to work with a corresponding ROA region. For this kind of problems multi-atlas segmentation (MAS) might be an effective method to produce masks based on 38 atlases I have. Even if my results demonstrate that my method is not a effective way to reproduce fibers automatically, it does not mean multi-atlas segmentation (MAS) is a bad idea. I prefer to discuss the result of my method and future work in three aspects: through threshold selection and weight assignment, through ROA regions, and through time consumption.

5.1. Threshold Selection and Weight Assignment

In this project, I selected thresholds two time in different steps: 1) I selected a threshold to extract the densest part of each tract and 2) in voting process I used threshold to measure raters' agreement on each voxel. For the original protocol ROI regions are 2D, small and local on certain slice. Therefore, there is a reasonable prediction that is a larger and 3D ROI regions would allow much more streamlines to cross. A mask covers complete pathway might not be a good ROI, since a large 3D mask has some overlap with multiple anatomical landmarks. I use threshold=0.05 to cut down the overlaps of 3D mask. In my prediction the densest part of each tract only belongs to this tract. However, the result (Figure1 (b)) demonstrates that the fiber produced by the densest mask also contain many overlaps and unwanted streamlines. One possible reason causes this result is that the multi-atlas segmentation (MAS) method I used in this project did not produce correct masks in the right location. To evaluate this MAS result I use DICE (th=0.0) to calculate similarity between ROI regions (threshold=0.0) created by the MAS method and the tract density file produced by manual regions of same subject (Appendix A). Since BLSA and HCP dataset have similar trend I did not show all boxplots in this paper. Both FigureA.10 and FigureA.1 have similar trend of reproducibility for each individual tract. In this project I did not considered asymmetrical weights assignment method during lab fusion process. Although I simply assigned weight according to which dataset these

subjects came from, a result this kind of weight assignment method did not make any sense. Because of that I think for future work, more complicated label fusion algorithm should be used in MAS method. For instance, it would be more accurate to assign weights according to the degree of similarity between pairs of subjects.

5.2. ROA Regions

The main challenge of this project was that it is difficult to automatically draw ROA regions without human subjective judgement. Additionally, the tract reproducibility analysis shows that the localization and the shape of ROA regions play crucial roles in tract reconstruction, since there no strict rules about the position of ROA regions in TractEM protocols, raters could hardly reach a consensus on selecting a ROA slice. Raters adjust ROA slices according to their subjective judgement. For future work, inverse ROA regions might be a good idea. The inverse ROA region is a 3D region covering all voxels except the ROI regions. Because inverse ROA regions are very strict exclusion regions, the ROI masks produced by the MAS method have to be extremely precise.

5.3. Time Consumption

In this project I used 3D ROI regions and the whole brain as a seed region in the fiber tracking process. The speed of fiber tracking of this way is significantly slower than using a 2D manually drawn seed region on a particular slice in fiber tracking. For TractEM protocols, if we calculate the time consumption by loading all regions directly and computing tracts by command lines, fiber tracking for one subject would take roughly two hours. However, for 3D ROI regions and whole brain seed regions, fiber tracking for one subject would take roughly 4 to 8 hours (it is up to machine computation and the size of ROI regions). Therefore, I think using particular seed regions instead of whole brain seed regions would make the process more efficient. The other seed regions have to be limited by more strict ROA regions.

Chapter 6. Conclusion

In this project, I applied a basic multi-atlas segmentation (MAS) method on 58 atlases, and tried to find a fully automated method to reconstruct 31 white matter tracts defined by TractEM protocols (https://my.vanderbilt.edu/tractem/). The result shows that the densest part of the characteristic fiber bundles has better reliability in fiber tracking but it requires more time cost. Due to ROA regions play a crucial role in fiber tracking process, defining a precise ROA region is as important as defining a valid ROI region. Unfortunately, the method I proposed in this paper cannot be used as a effective fully automated reconstruction method. For future work more complicated label fusion algorithm can be considered on the dataset: 1) globally weighted voting method (Artaechevarria et al. 2008; Cao et al., 2011b; Langerak et al., 2010; Langerak et al., 2011;), 2) Local weighted voting method (Isgum et al., 2009; Iglesias and Karssemeijer,2009), 3) computing weights at each voxel (Coupe et al., 2011; Fonov et al., 2012), and 4) computing the label fusion weights by seeking sparse linear combinations from a patch dictionary to reconstruct the novel image (Liao et al., 2013; Zhang et al., 2012; Wang et al., 2014d).

REFERENCE

Artaechevarria, X., Muñoz-Barrutia, A., & Ortiz-de-Solórzano, C. (2008). Efficient classifier generation and weighted voting for atlas-based segmentation: two small steps faster and closer to the combination oracle. *Medical Imaging 2008: Image Processing*. https://doi.org/10.1117/12.769401

Basser, P. J., Mattiello, J., & Lebihan, D. (1994). Estimation of the Effective Self-Diffusion Tensor from the NMR Spin Echo. *Journal of Magnetic Resonance, Series B*. https://doi.org/10.1006/jmrb.1994.1037

Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., & Aldroubi, A. (2000). In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine*. https://doi.org/10.1002/1522-2594(200010)44:4<625::AID-MRM17>3.0.CO;2-O

Bayrak, R. G., Schilling, K. G., Greer, J. M., Hansen, C. B., Greer, C. M., Blaber, J. A., … Landman, B. A. (2019). TractEM: Fast Protocols for Whole Brain Deterministic Tractography-Based White Matter Atlas. BioRxiv. https://doi.org/10.1101/651935

Besseling, R. M. H., Jansen, J. F. A., Overvliet, G. M., Vaessen, M. J., Braakman, H. M. H., Hofman, P. A. M., … Backes, W. H. (2012). Tract specific reproducibility of tractography based morphology and diffusion metrics. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0034125

Cao, Y., Yuan, Y., Li, X., & Yan, P. (2011). Putting images on a manifold for atlas-based image segmentation. *Proceedings - International Conference on Image Processing, ICIP*. https://doi.org/10.1109/ICIP.2011.6116265

Christidi, F., Karavasilis, E., Samiotis, K., Bisdas, S., & Papanikolaou, N. (2016). Fiber tracking: A qualitative and quantitative comparison between four different software tools on the reconstruction of major white matter tracts. *European Journal of Radiology Open*. https://doi.org/10.1016/j.ejro.2016.06.002

Conturo, T. E., Lori, N. F., Cull, T. S., Akbudak, E., Snyder, A. Z., Shimony, J. S., … Raichle, M. E. (1999). Tracking neuronal fiber pathways in the living human brain. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.96.18.10422

Côté, M. A., Girard, G., Boré, A., Garyfallidis, E., Houde, J. C., & Descoteaux, M. (2013). Tractometer: Towards validation of tractography pipelines. *Medical Image Analysis*. https://doi.org/10.1016/j.media.2013.03.009

Coupé, P., Manjón, J. V., Fonov, V., Pruessner, J., Robles, M., & Collins, D. L. (2011). Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2010.09.018

Ferrucci, L., 2008. *The baltimore longitudinal study of aging (blsa): a 50-year-long journey and plans for the future*.

Fonov, V., Coup, P., Eskildsen, S.F., Manjon, J.V., Collins, L., et al., 2012. Multi-atlas labeling with population-specic template and non-local patch-based label fusion, in: MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling, pp. 63-66.

G. J.M, P., K.E, S., G.J, B., J.B, R., D.G, M., C.A.M, W.-K., … R., T. (2002). Initial demonstration of in vivo traccing of axonal projections in the macaque brain and comparison with the human brain using diffusion tensor imaging and fast marching tractography. *Neuroimage*.

Heiervang, E., Behrens, T. E. J., Mackay, C. E., Robson, M. D., & Johansen-Berg, H. (2006). Between session reproducibility and between subject variability of diffusion MR and tractography measures. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2006.07.037

Iglesias, J. E., & Karssemeijer, N. (2009). Robust initial detection of landmarks in film-screen mammograms using multiple ffdm atlases. *IEEE Transactions on Medical Imaging*. https://doi.org/10.1109/TMI.2009.2025036

Iglesias, J. E., Sabuncu, M. R., Aganj, I., Bhatt, P., Casillas, C., Salat, D., Boxer, A., Fischl, B., & Van Leemput, K. (2015). An algorithm for optimal fusion of atlases with different labeling protocols. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2014.11.031

Iglesias, J. E., & Sabuncu, M. R. (2015). Multi-atlas segmentation of biomedical images: A survey. *Medical Image Analysis*. https://doi.org/10.1016/j.media.2015.06.012

Išgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M. A., & Van Ginneken, B. (2009). Multi-atlas-based segmentation with local decision fusion-application to cardiac and aortic segmentation in CT scans. *IEEE Transactions on Medical Imaging*. https://doi.org/10.1109/TMI.2008.2011480

Jones, D. K., Simmons, A., Williams, S. C. R., & Horsfield, M. A. (1999). Non-invasive assessment of axonal fiber connectivity in the human brain via diffusion tensor MRI. *Magnetic Resonance in Medicine*. https://doi.org/10.1002/(SICI)1522-2594(199907)42:1<37::AID-MRM7>3.0.CO;2-O

Langerak, T. R., Van Der Heide, U. A., Kotte, A. N. T. J., Viergever, M. A., Van Vulpen, M., & Pluim, J. P. W. (2010). Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). *IEEE Transactions on Medical Imaging*. https://doi.org/10.1109/TMI.2010.2057442

Langerak, T. R., Van Der Heide, U. A., Kotte, A. N. T. J., Berendsen, F. F., & Pluim, J. P. W. (2011). Local atlas selection and performance estimation in multi-atlas based segmentation. *Proceedings - International Symposium on Biomedical Imaging*. https://doi.org/10.1109/ISBI.2011.5872495

Lazar, M., Weinstein, D. M., Tsuruda, J. S., Hasan, K. M., Arfanakis, K., Meyerand, M. E., … Alexander, A. L. (2003). White matter tractography using diffusion tensor deflection. *Human Brain Mapping*. https://doi.org/10.1002/hbm.10102

Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., & Laval-Jeantet, M. (1986). MR imaging of intravoxel incoherent motions: Application to diffusion and perfusion in neurologic disorders. *Radiology*. https://doi.org/10.1148/radiology.161.2.3763909

Lepore, A. C., Rauck, B., Dejea, C., Pardo, A. C., Rao, M. S., Rothstein, J. D., & Maragakis, N. J. (2008). Focal transplantation-based astrocyte replacement is neuroprotective in a model of motor neuron disease. *Nature Neuroscience*. https://doi.org/10.1038/nn.2210

Liao, S., Gao, Y., Lian, J., & Shen, D. (2013). Sparse patch-based label propagation for accurate prostate localization in CT images. *IEEE Transactions on Medical Imaging*. https://doi.org/10.1109/TMI.2012.2230018

M. , A. (1895). Anatomie des Centres Nerveux, par J. Dejerine, professeur agrégé à La Faculté de Médecine de Paris, avec la collaboration de Madame Dejerine-Klumpke. With 401 figures. Paris: Rueff et Cie, 1895. Price, 32 frcs . American Journal of Psychiatry. https://doi.org/10.1176/ajp.51.3.402

Merboldt, K. D., Hanicke, W., & Frahm, J. (1985). Self-diffusion NMR imaging using stimulated echoes. *Journal of Magnetic Resonance (1969)*. https://doi.org/10.1016/0022-2364(85)90111-8

Mori, S., Crain, B. J., Chacko, V. P., & Van Zijl, P. C. M. (1999). Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of Neurology*. https://doi.org/10.1002/1531-8249(199902)45:2<265::AID-ANA21>3.0.CO;2-3

Oishi, K., Faria, A. V., Van Zijl, P. C., Mori, S., 2010. *MRI atlas of human white matter*. Academic Press.

Poupon, C., Clark, C. A., Frouin, V., Régis, J., Bloch, I., Le Bihan, D., & Mangin, J. F. (2000). Regularization of diffusion-based direction maps for the tracking of brain white matter fascicles. *NeuroImage*. https://doi.org/10.1006/nimg.2000.0607

Price, G., Cercignani, M., Parker, G. J. M., Altmann, D. R., Barnes, T. R. E., Barker, G. J., … Ron, M. A. (2007). Abnormal brain connectivity in first-episode psychosis: A diffusion MRI tractography study of the corpus callosum. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2006.12.019

Price, G., Cercignani, M., Parker, G. J. M., Altmann, D. R., Barnes, T. R. E., Barker, G. J., … Ron, M. A. (2008). White matter tracts in first-episode psychosis: A DTI tractography study of the uncinate fasciculus. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2007.09.012

Schilling, K. G., Yeh, F. C., Nath, V., Hansen, C., Williams, O., Resnick, S., … Landman, B. A. (2019). A fiber coherence index for quality control of B-table orientation in diffusion MRI scans. *Magnetic Resonance Imaging*. https://doi.org/10.1016/j.mri.2019.01.018

Taylor, D. G., & Bushell, M. C. (1985). The spatial mapping of translational diffusion coefficients by the NMR imaging technique. *Physics in Medicine and Biology*. https://doi.org/10.1088/0031-9155/30/4/009

Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E. J., Bucholz, R., … Yacoub, E. (2012). The Human Connectome Project: A data acquisition perspective. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2012.02.018

Wakana, S., Caprihan, A., Panzenboeck, M. M., Fallon, J. H., Perry, M., Gollub, R. L., … Mori, S. (2007). Reproducibility of quantitative tractography methods applied to cerebral white matter. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2007.02.049

Wakana, S., Jiang, H., Nagae-Poetscher, L. M., Van Zijl, P. C. M., & Mori, S. (2004). Fiber Tract-based Atlas of Human White Matter Anatomy. *Radiology*. https://doi.org/10.1148/radiol.2301021640

Wang, L., Shi, F., Gao, Y., Li, G., Gilmore, J. H., Lin, W., & Shen, D. (2014). Integration of sparse multi-modality representation and anatomical constraint for isointense infant brain MR image segmentation. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2013.11.040

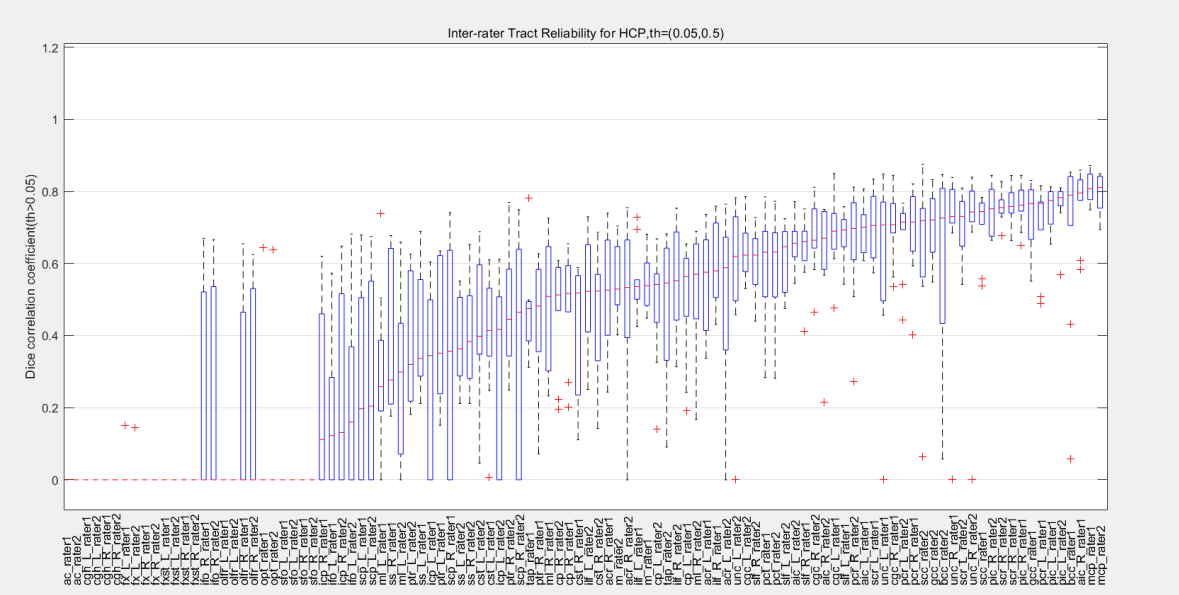
Wasserthal, J., Neher, P., & Maier-Hein, K. H. (2018). TractSeg - Fast and accurate white matter tract segmentation. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2018.07.070

Zhang, D., Guo, Q., Wu, G., & Shen, D. (2012). Sparse patch-based label fusion for multi-atlas segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-642-33530-3_8

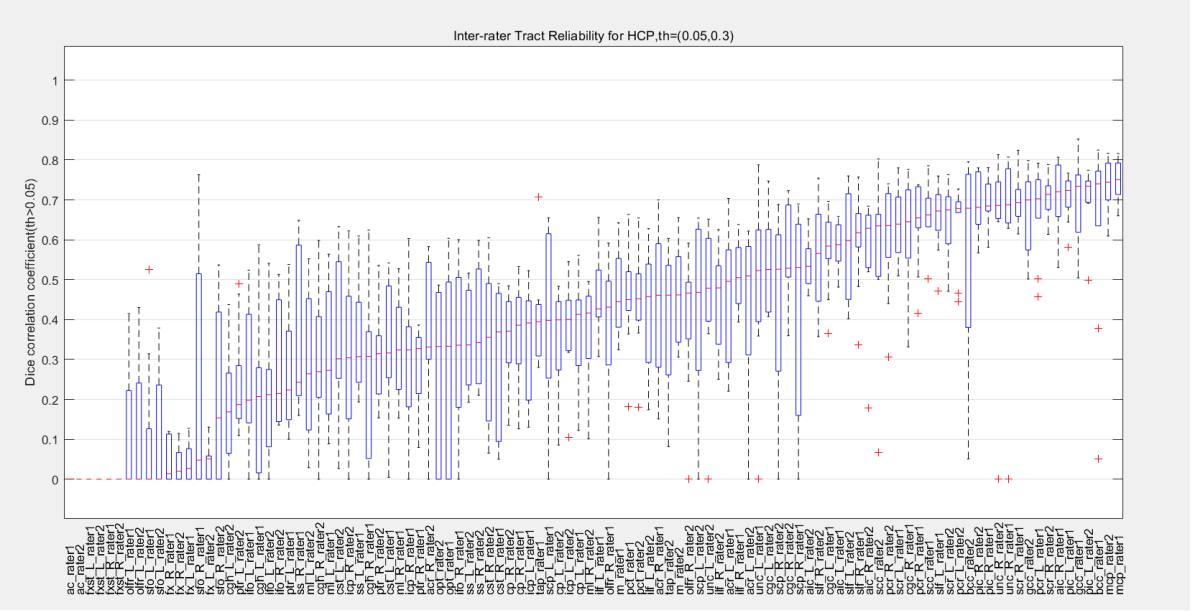
Zhang, S., & Arfanakis, K. (2018). Evaluation of standardized and study-specific diffusion tensor imaging templates of the adult human brain: Template characteristics, spatial normalization accuracy, and detection of small inter-group FA differences. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2018.01.046

Zhang, Y., Zhang, J., Oishi, K., Faria, A. V., Jiang, H., Li, X., … Mori, S. (2010). Atlas-guided tract reconstruction for automated and comprehensive examination of the white matter anatomy. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2010.05.049

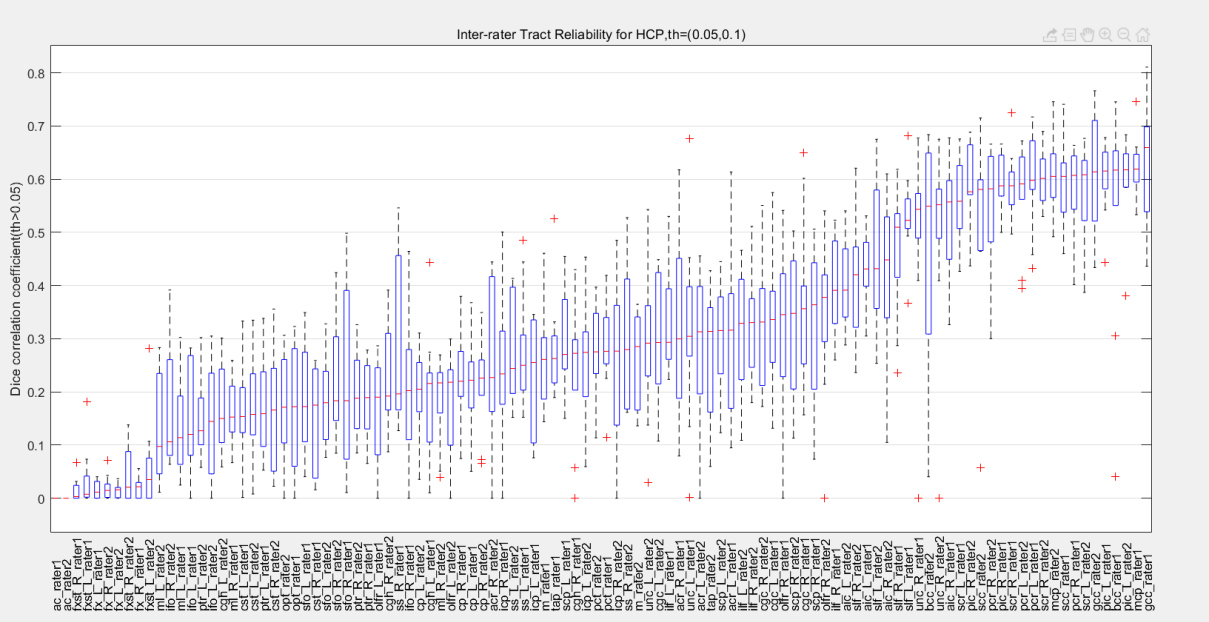
Appendix A



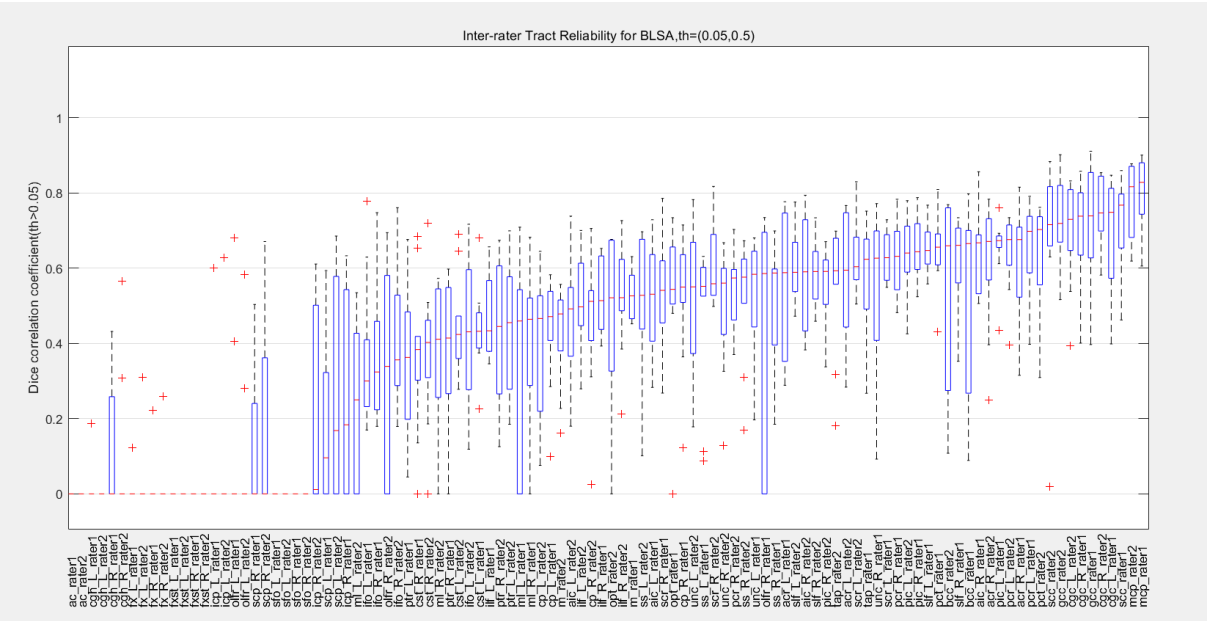
A.1. Tract reliability for all 53 fiber bundles is shown.These tracts are reconstruct by branch1. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 10 subjects from the HCP dataset with first threshold=0.05 and voting threshold=0.5. The tracts are sorted from low to high reliability.



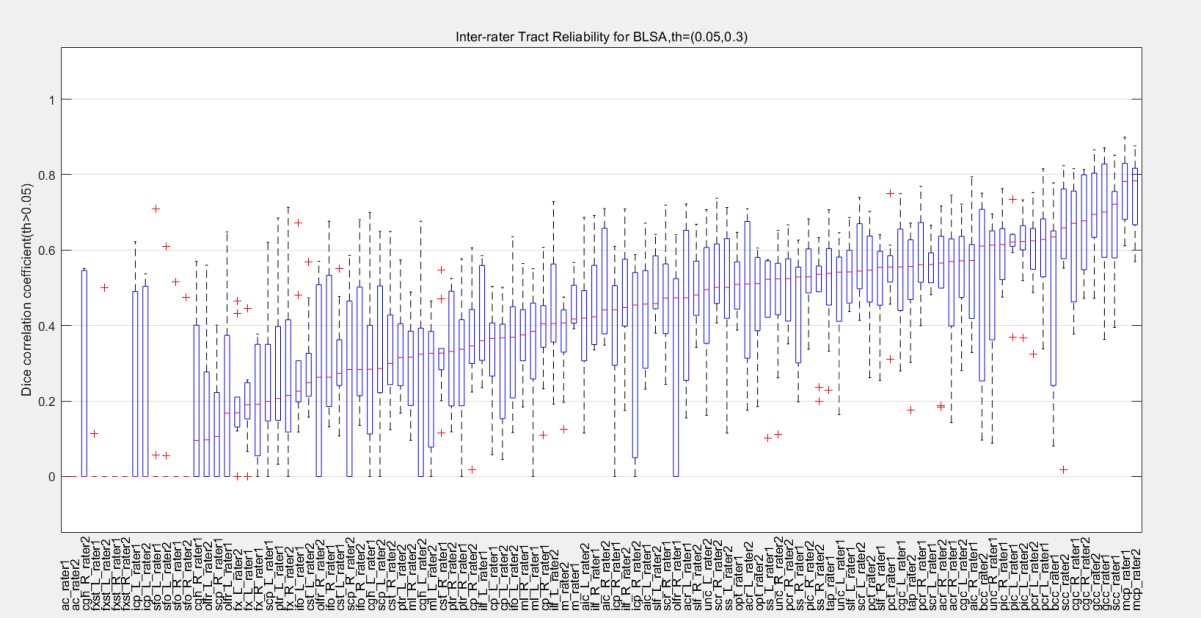
A.2. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch1. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 10 subjects from the HCP dataset with first threshold=0.05 and voting threshold=0.3. The tracts are sorted from low to high reliability.



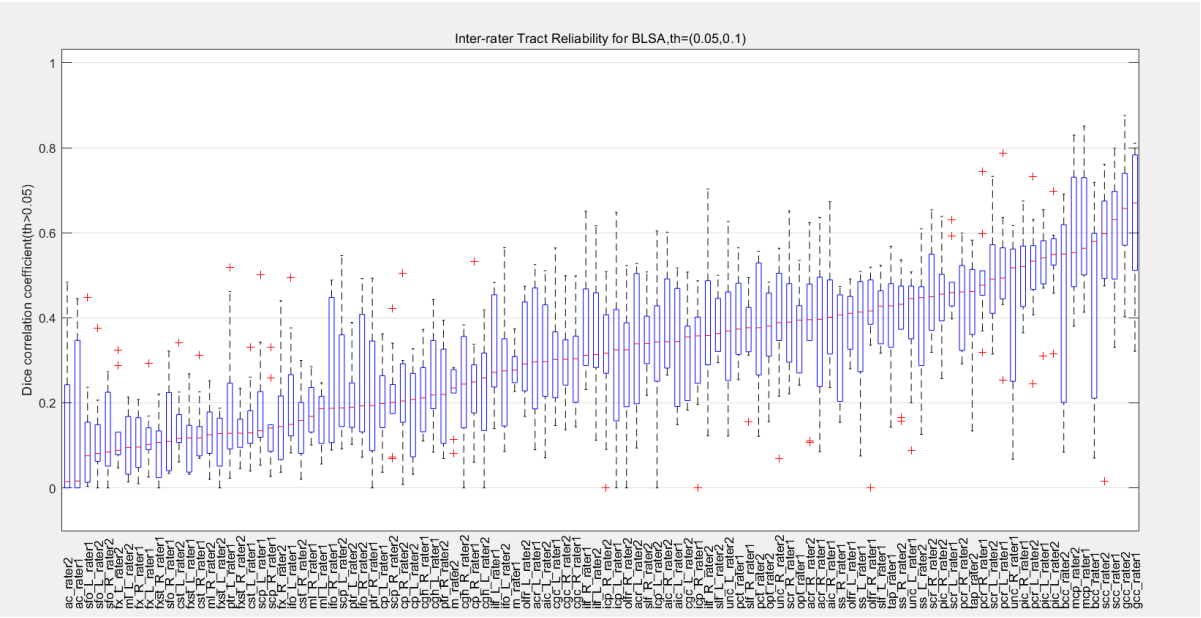
A.3. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch1. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 10 subjects from the HCP dataset with first threshold=0.05 and voting threshold=0.1. The tracts are sorted from low to high reliability.



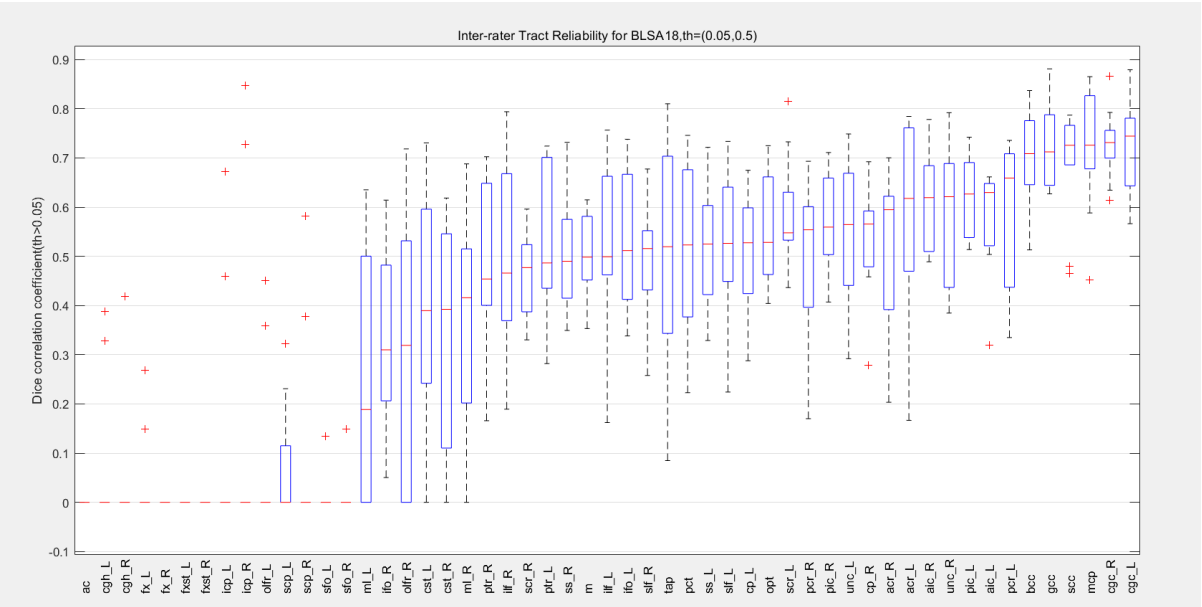
A.4. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch1. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 10 subjects from the BLSA dataset with first threshold=0.05 and voting threshold=0.5. The tracts are sorted from low to high reliability



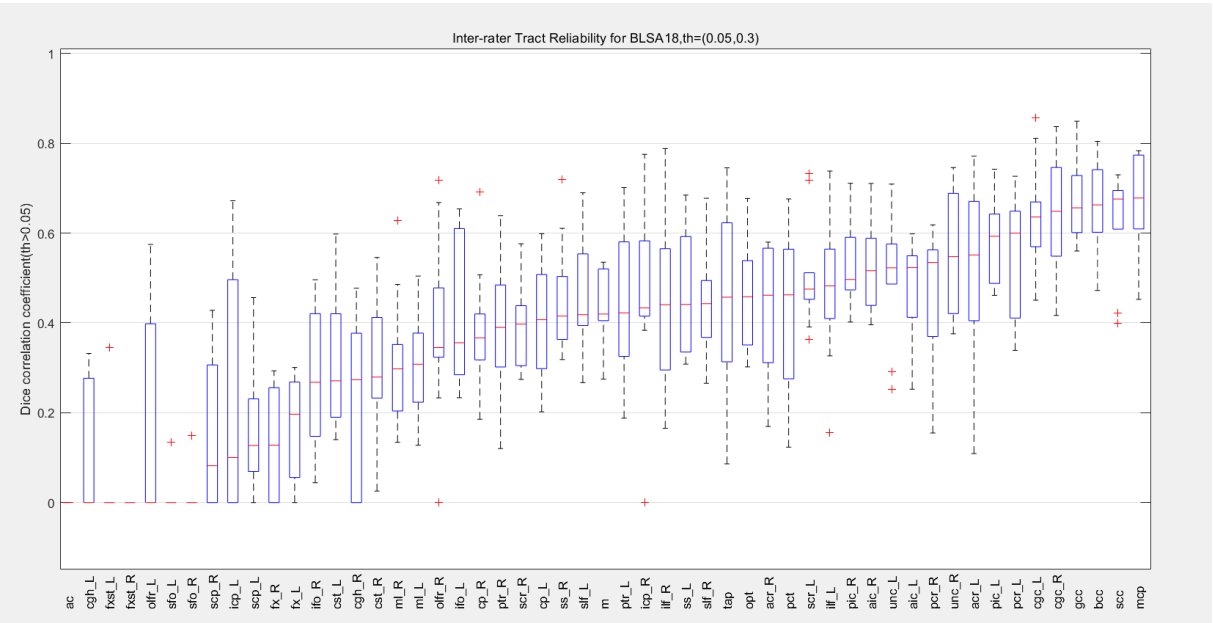
A.5. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch1. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 10 subjects from the BLSA dataset with first threshold=0.05 and voting threshold=0.3. The tracts are sorted from low to high reliability.



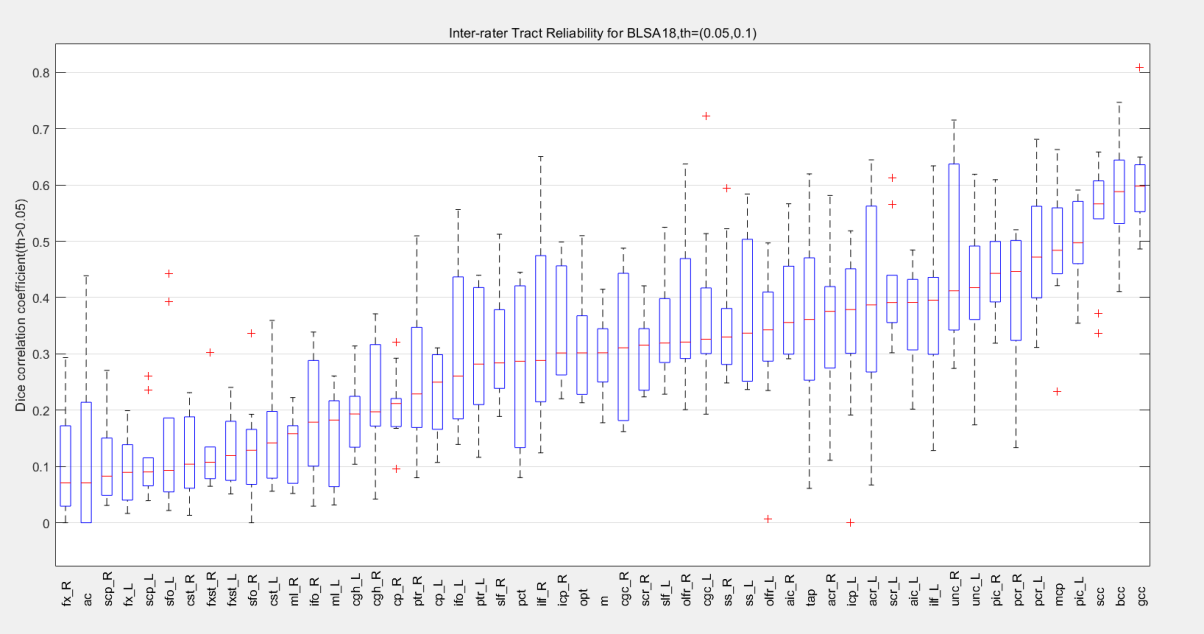
A.6. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch1. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 10 subjects from the BLSA dataset with first threshold=0.05 and voting threshold=0.1. The tracts are sorted from low to high reliability.



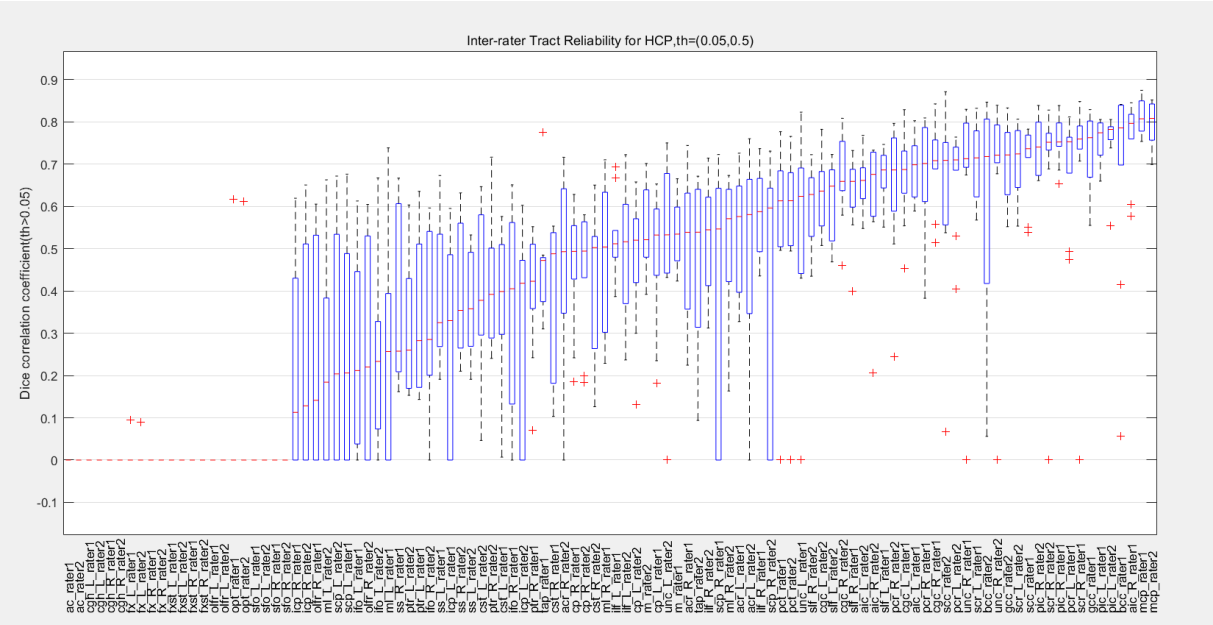
A.7. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch1. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 18 subjects from the BLSA18 dataset with first threshold=0.05 and voting threshold=0.5. The tracts are sorted from low to high reliability.



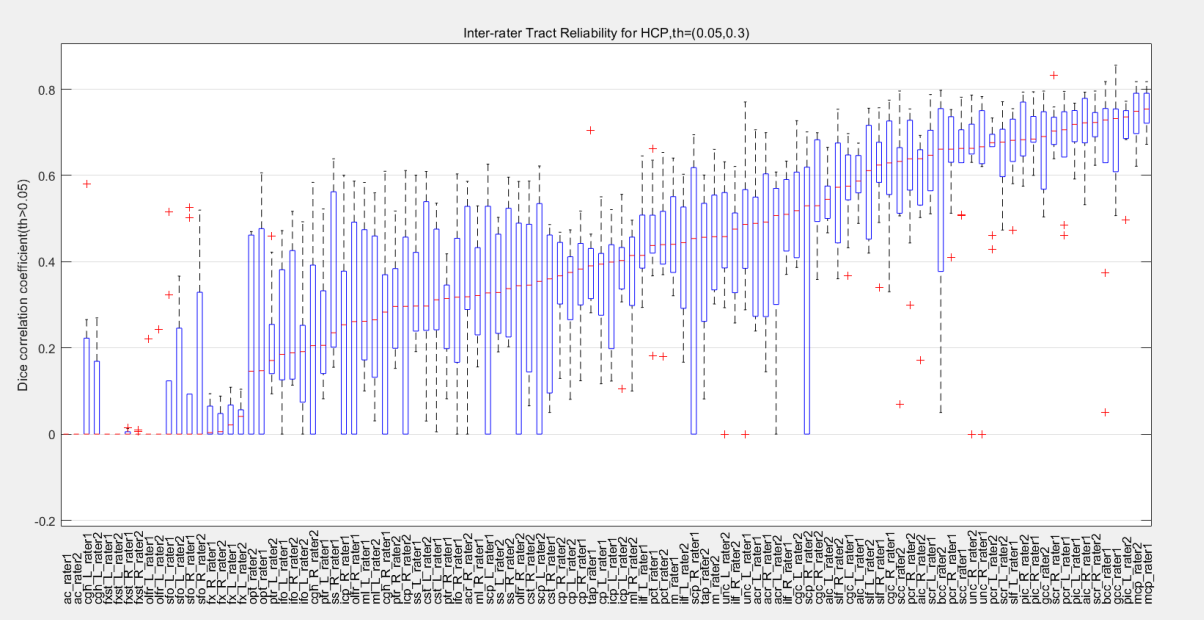
A.8. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch1. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 18 subjects from the BLSA18 dataset with first threshold=0.05 and voting threshold=0.3. The tracts are sorted from low to high reliability.



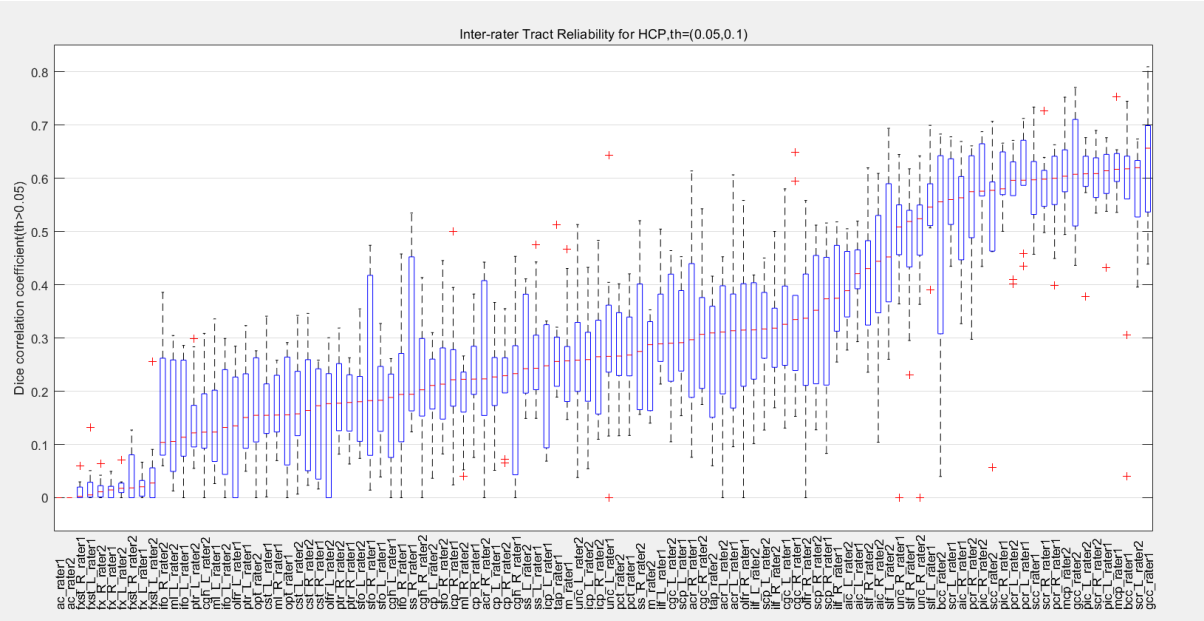
A.9. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch1. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 18 subjects from the BLSA18 dataset with first threshold=0.05 and voting threshold=0.1. The tracts are sorted from low to high reliability.



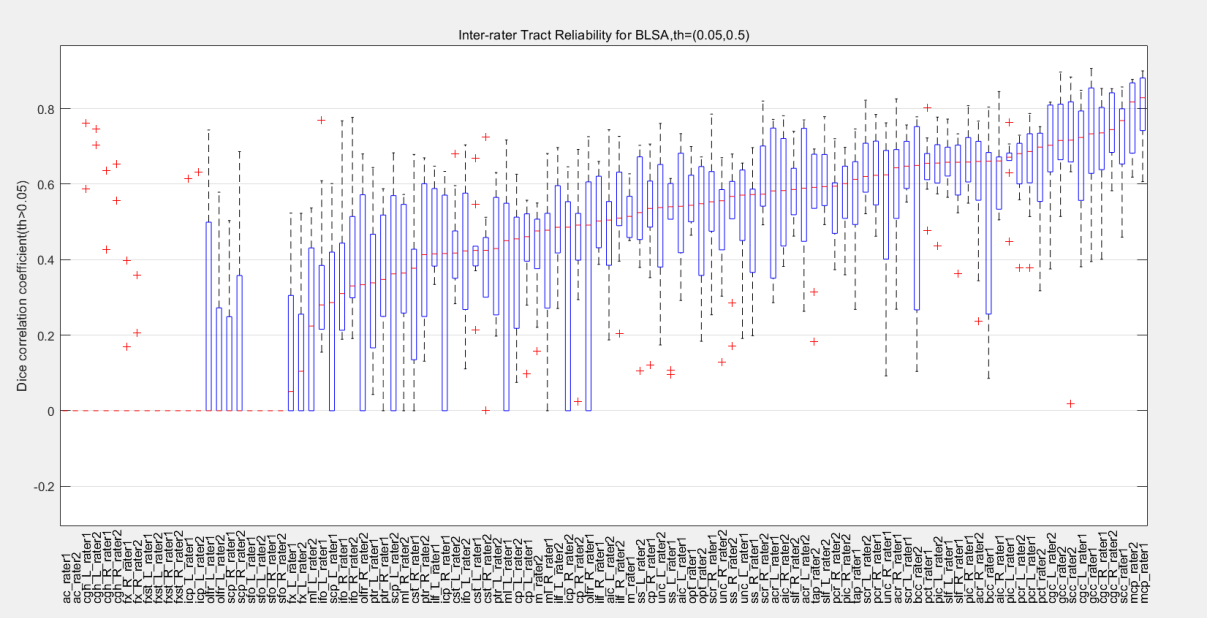
A.10. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch2. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 10 subjects from the HCP dataset with first threshold=0.05 and voting threshold=0.5. The tracts are sorted from low to high reliability.



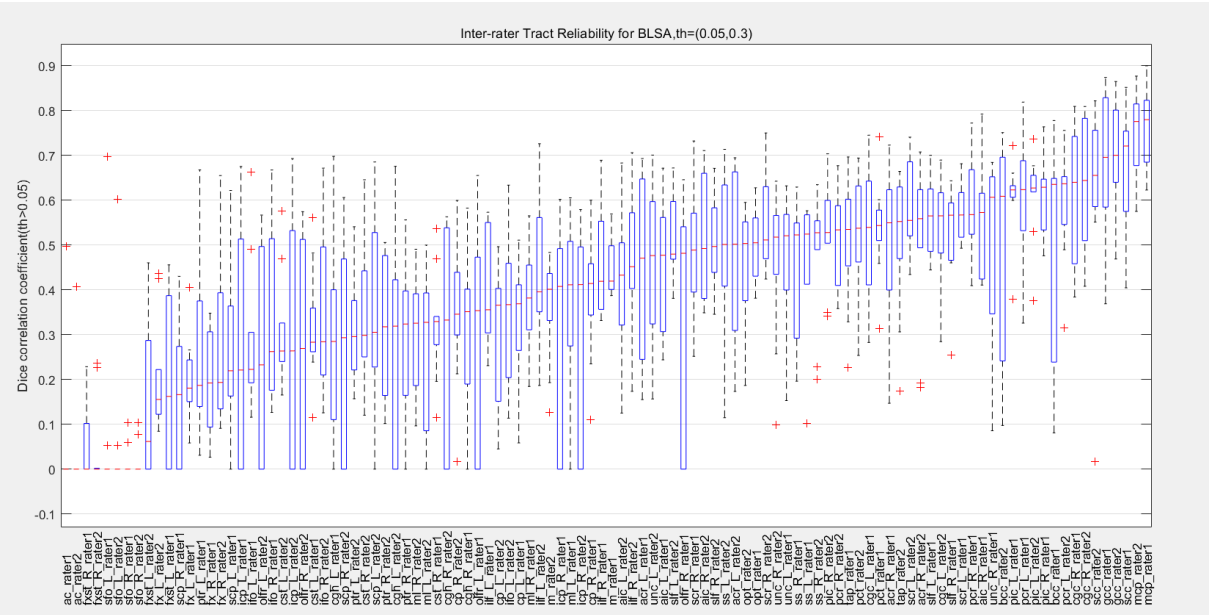
A.11. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch2. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 10 subjects from the HCP dataset with first threshold=0.05 and voting threshold=0.3. The tracts are sorted from low to high reliability.



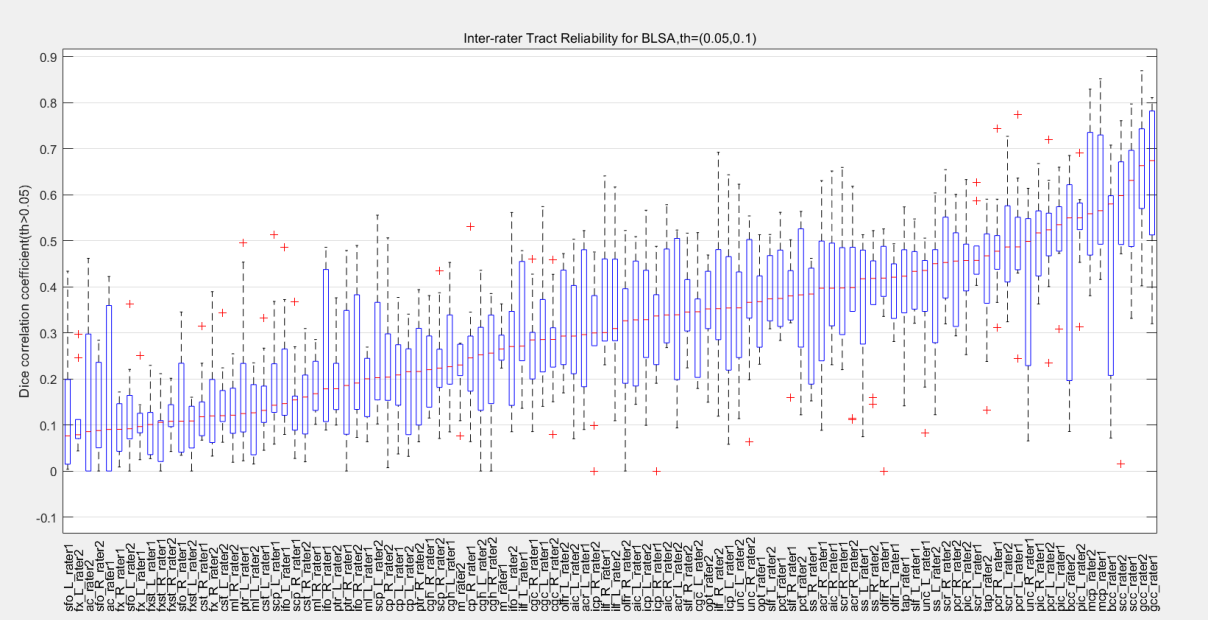
A.12. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch2. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 10 subjects from the HCP dataset with first threshold=0.05 and voting threshold=0.1. The tracts are sorted from low to high reliability.



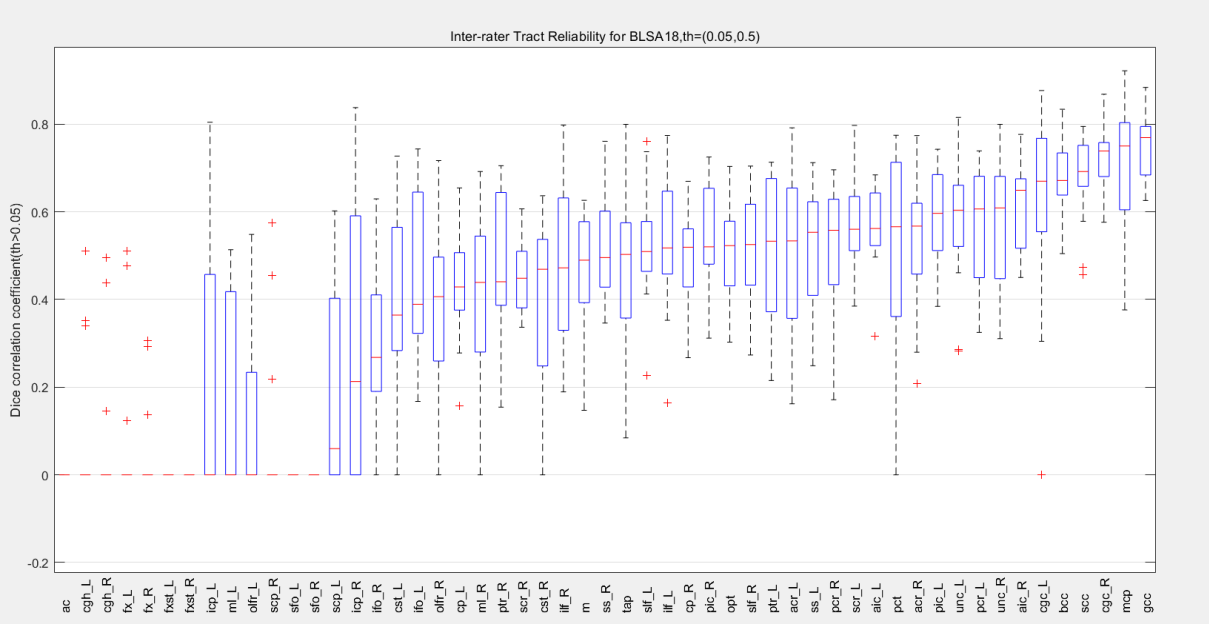
A.13. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch2. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 10 subjects from the BLSA dataset with first threshold=0.05 and voting threshold=0.5. The tracts are sorted from low to high reliability.



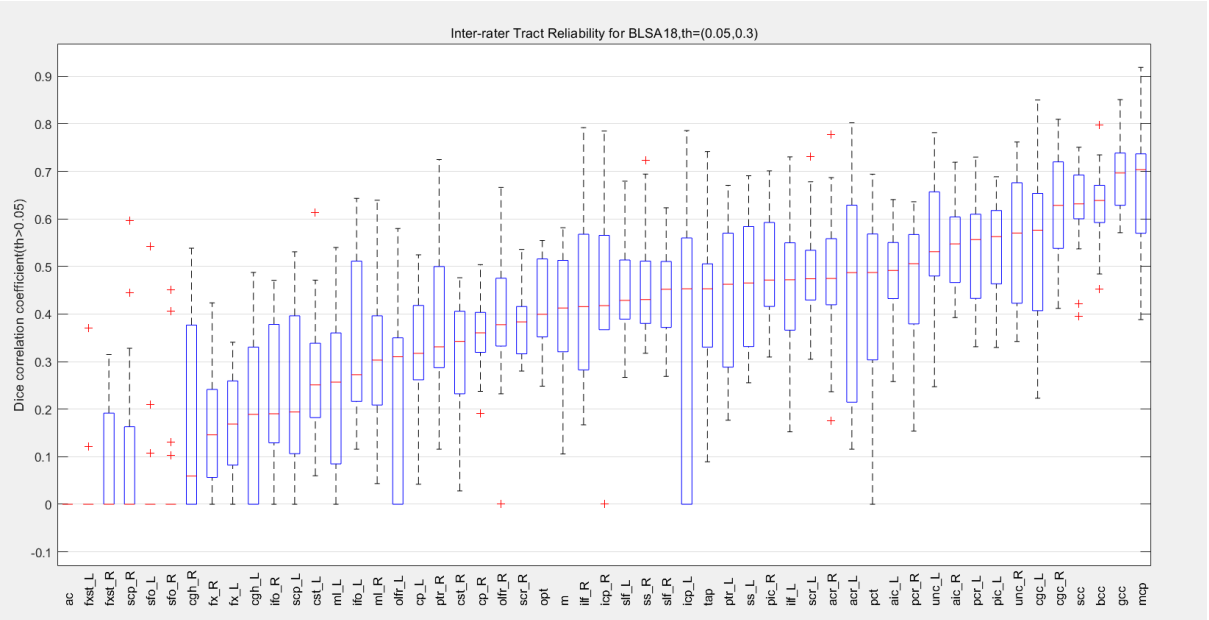
A.14. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch2.Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 10 subjects from the BLSA dataset with first threshold=0.05 and voting threshold=0.3. The tracts are sorted from low to high reliability.



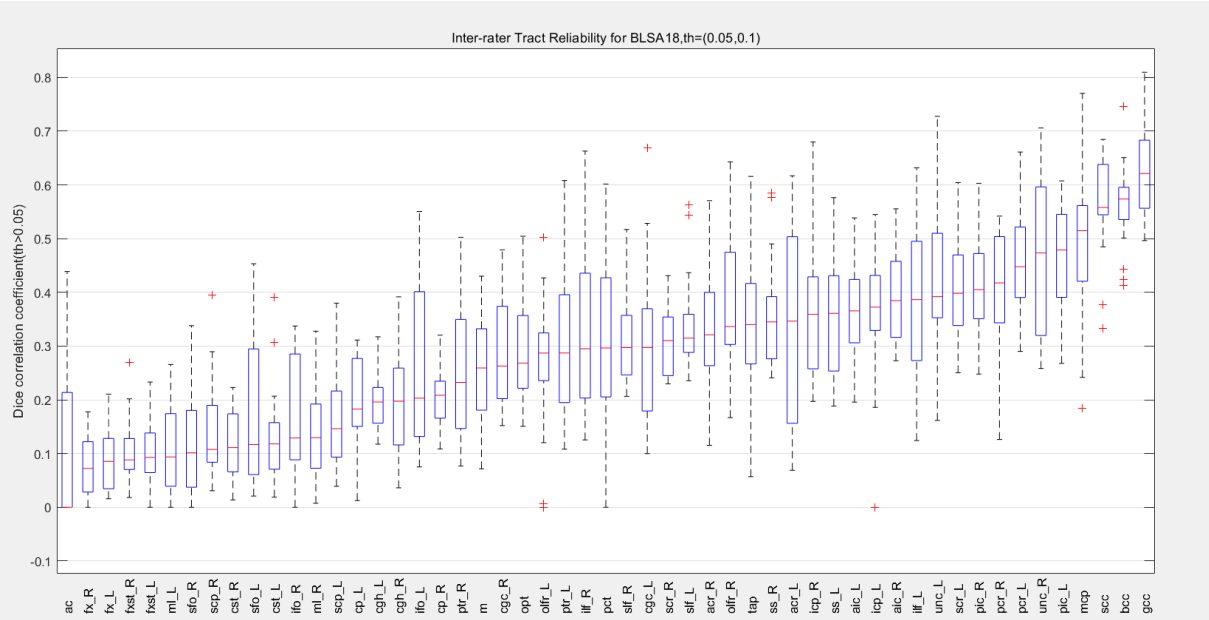
A.15. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch2. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 10 subjects from the BLSA dataset with first threshold=0.05 and voting threshold=0.1. The tracts are sorted from low to high reliability.



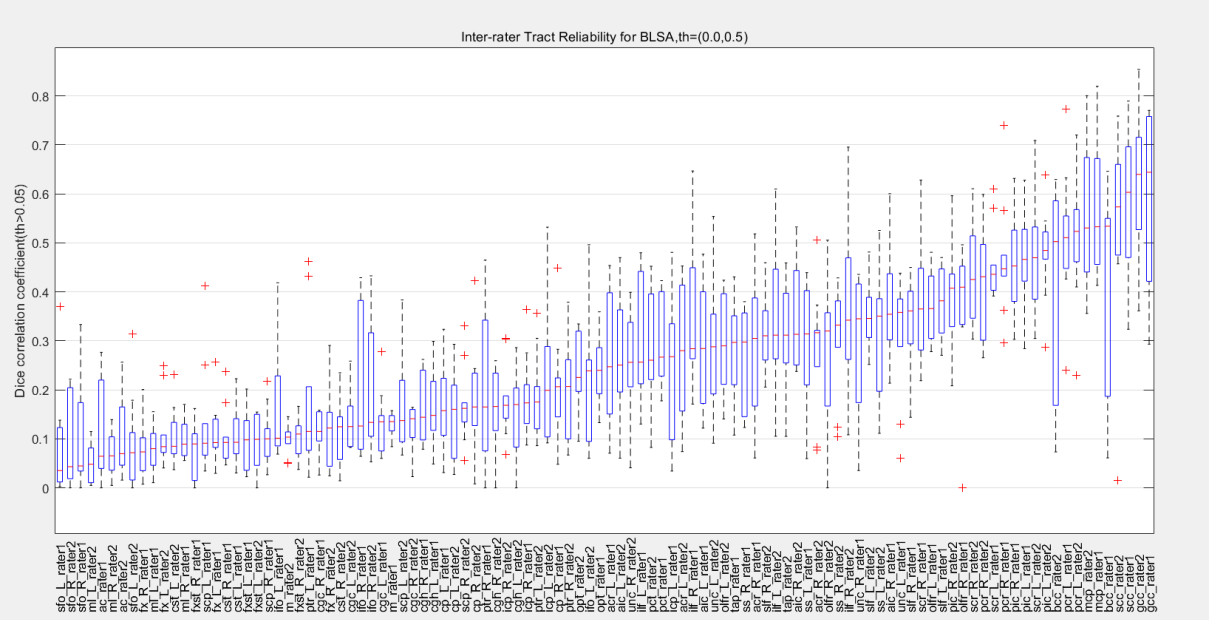
A.16. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch2. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 18 subjects from the BLSA18 dataset with first threshold=0.05 and voting threshold=0.5. The tracts are sorted from low to high reliability.



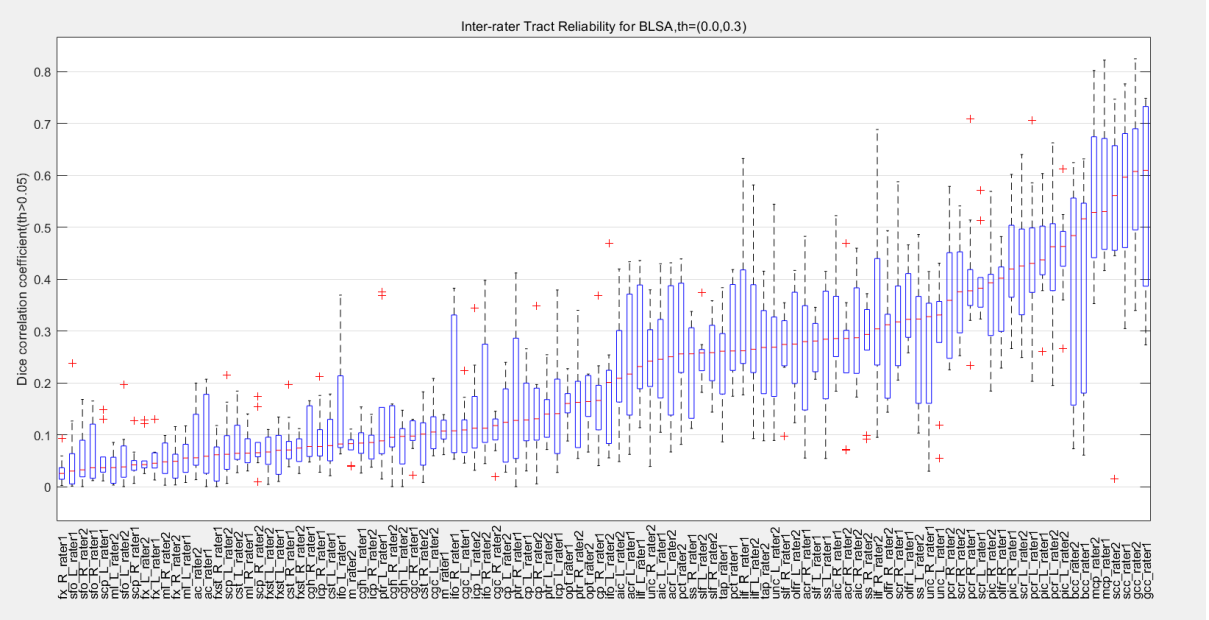
A.17. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch2. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 18 subjects from the BLSA18 dataset with first threshold=0.05 and voting threshold=0.3. The tracts are sorted from low to high reliability.



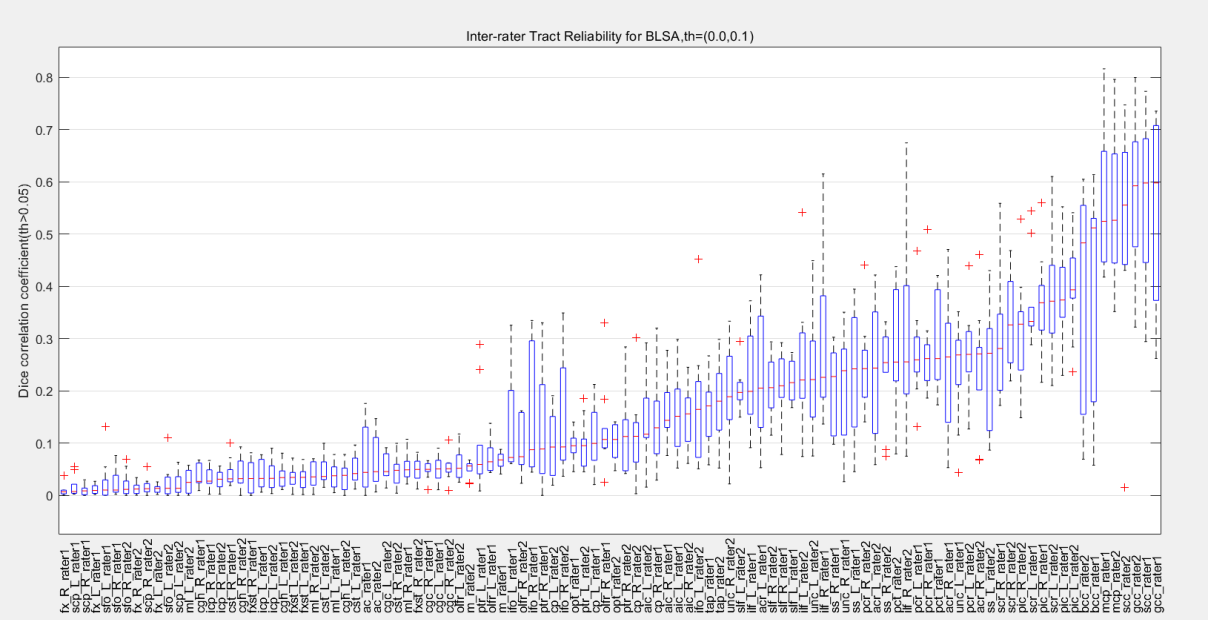
A.18. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch2. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 18 subjects from the BLSA18 dataset with first threshold=0.05 and voting threshold=0.1. The tracts are sorted from low to high reliability.



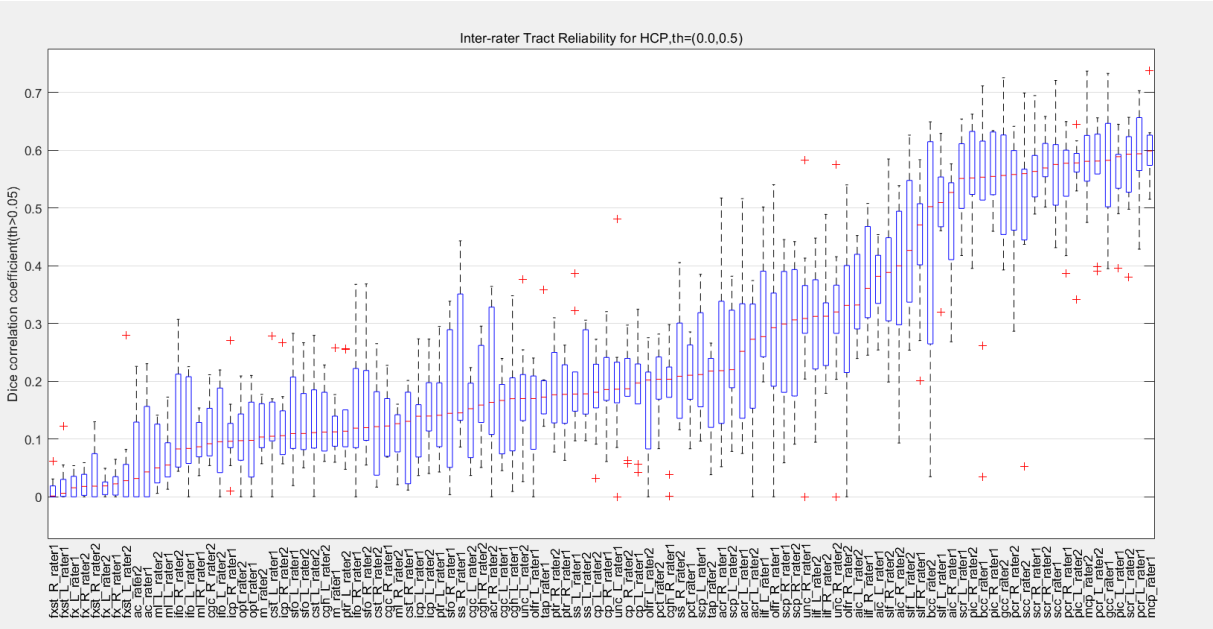
A.19. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch1. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 10 subjects from the BLSA dataset with first threshold=0.0 and voting threshold=0.5. The tracts are sorted from low to high reliability.



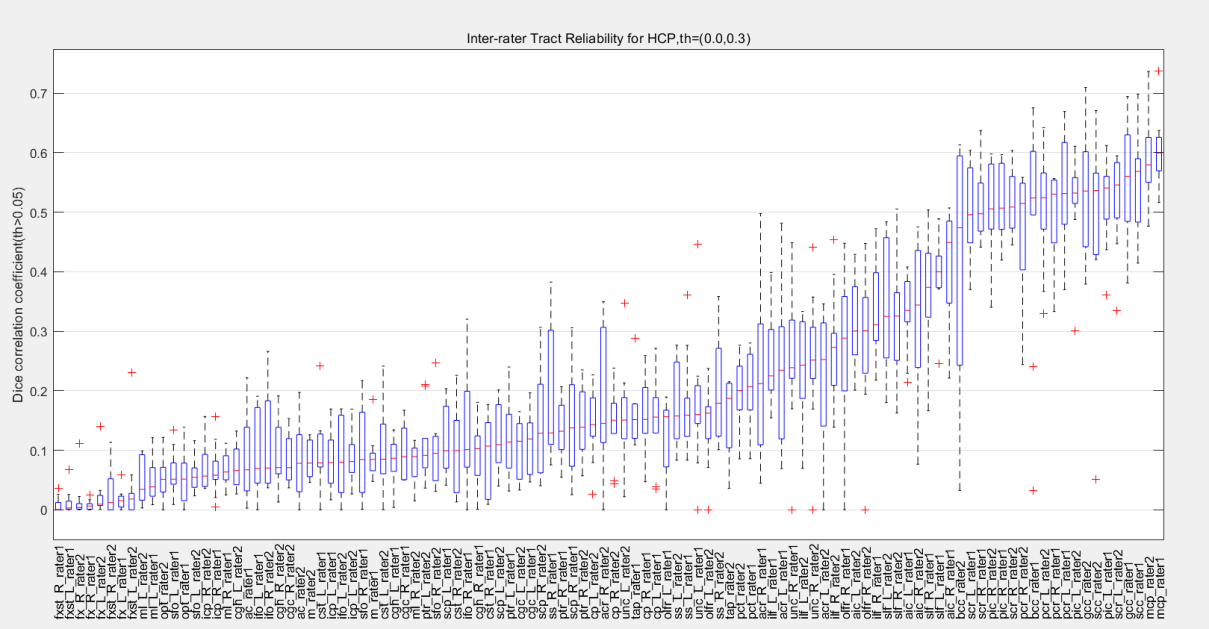
A.20. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch1. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 10 subjects from the BLSA dataset with first threshold=0.0 and voting threshold=0.3. The tracts are sorted from low to high reliability.



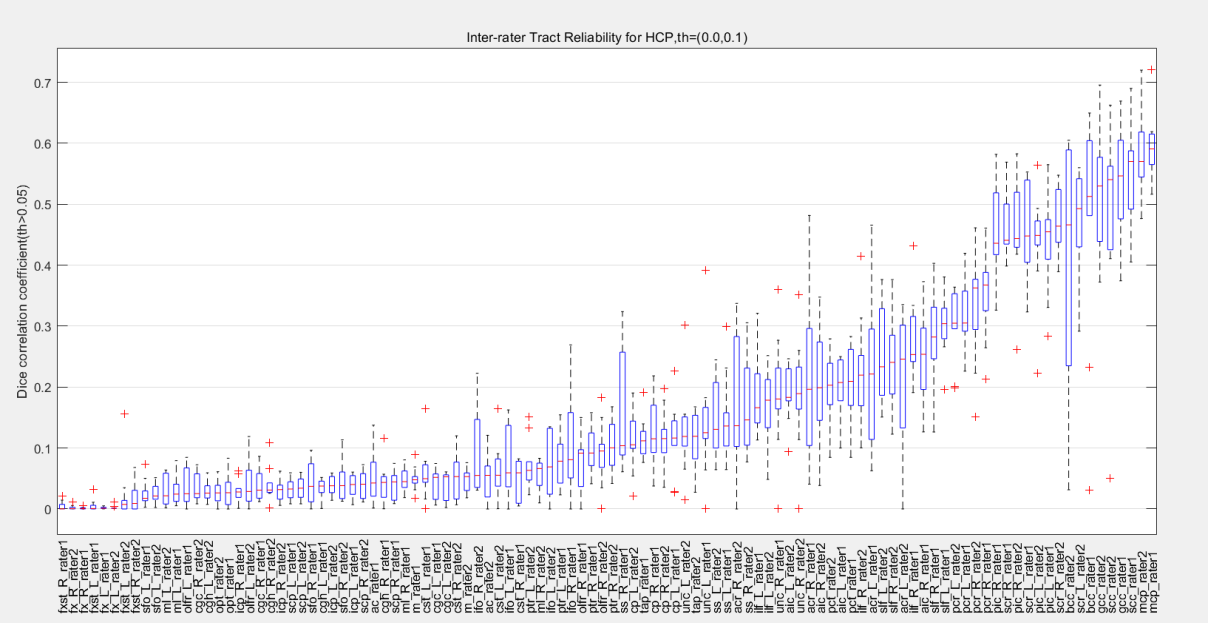
A.21. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch1. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 10 subjects from the BLSA dataset with first threshold=0.0 and voting threshold=0.1. The tracts are sorted from low to high reliability.



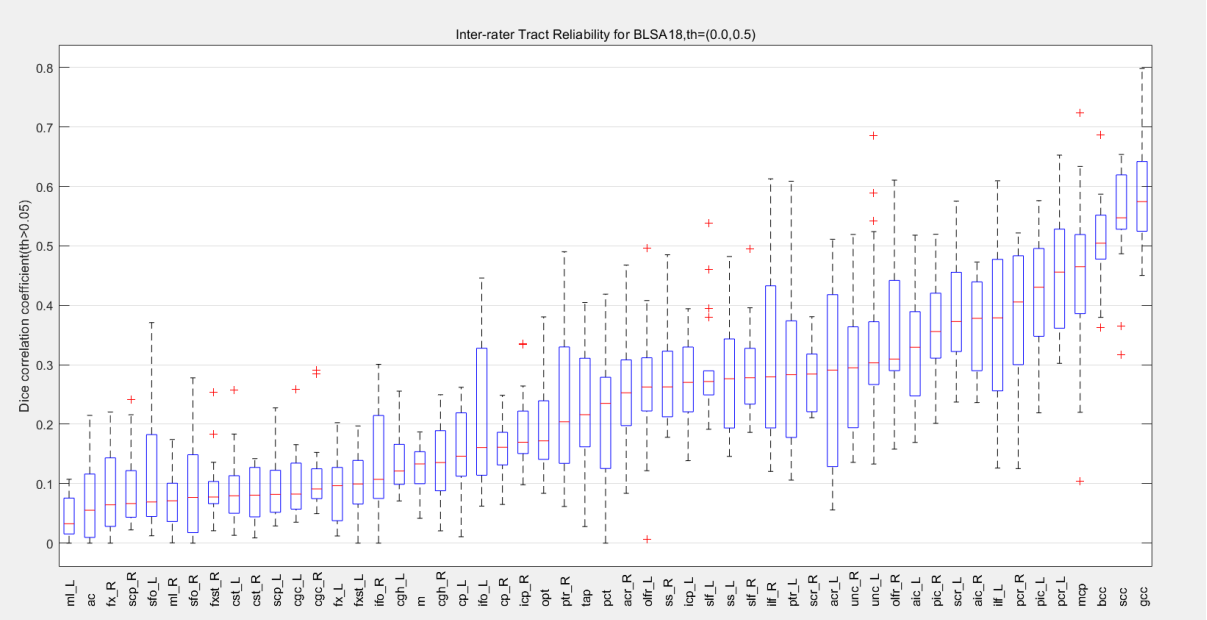
A.22. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch1. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 10 subjects from the HCP dataset with first threshold=0.0 and voting threshold=0.5. The tracts are sorted from low to high reliability.



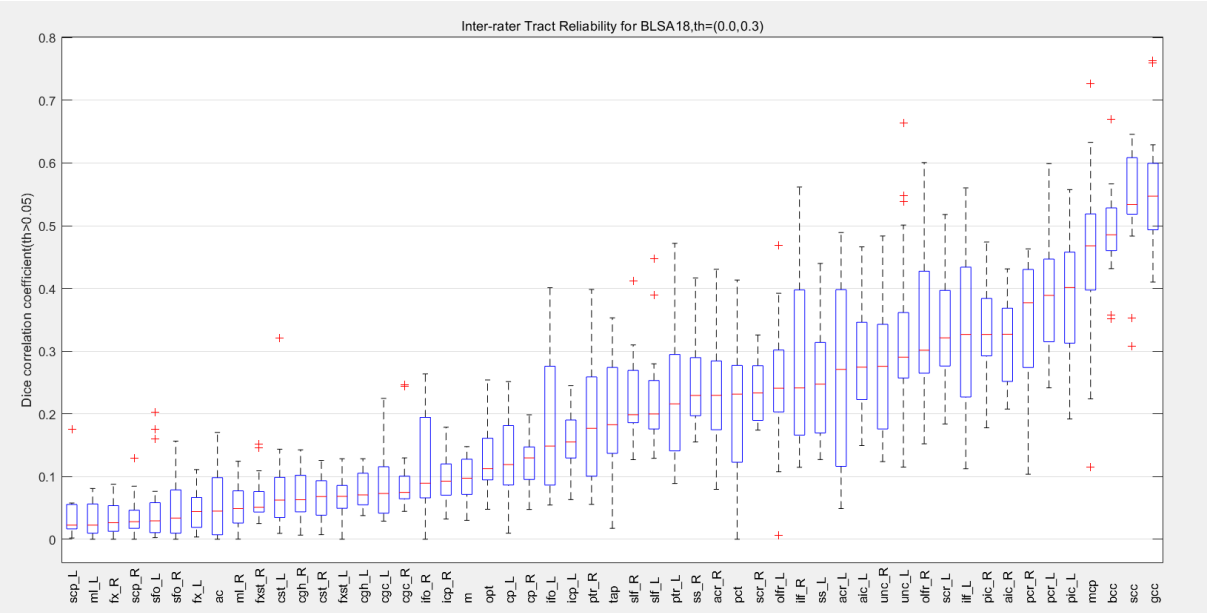
A.23. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch1. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 10 subjects from the HCP dataset with first threshold=0.0 and voting threshold=0.3. The tracts are sorted from low to high reliability.



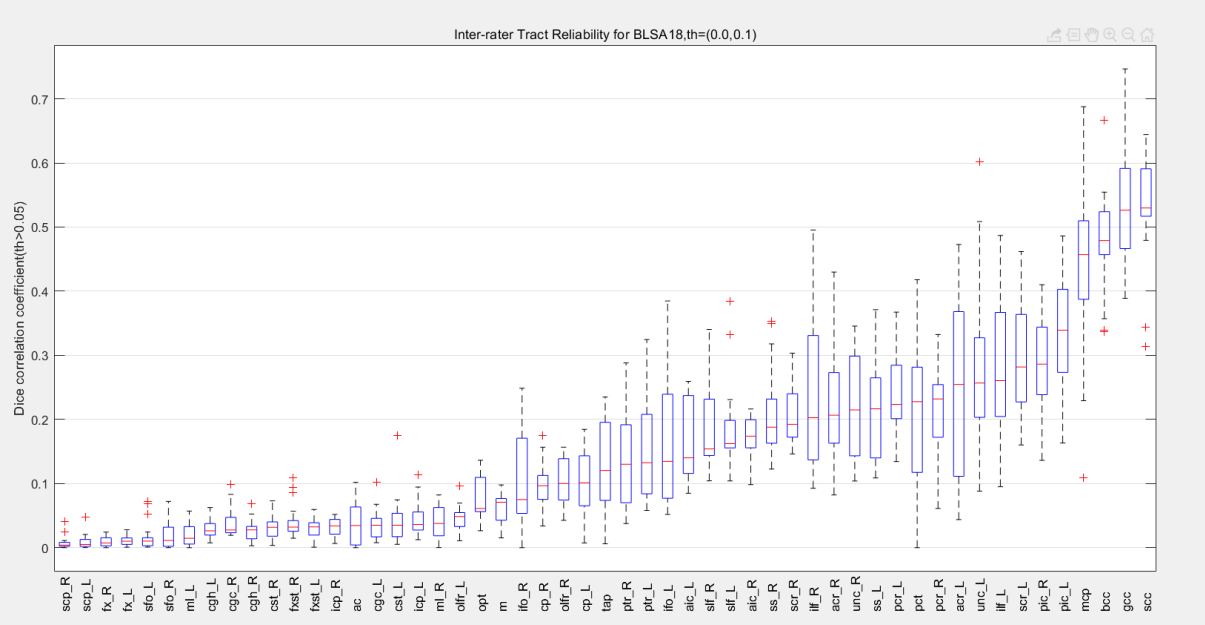
A.24. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch1. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 10 subjects from the HCP dataset with first threshold=0.0 and voting threshold=0.1. The tracts are sorted from low to high reliability.



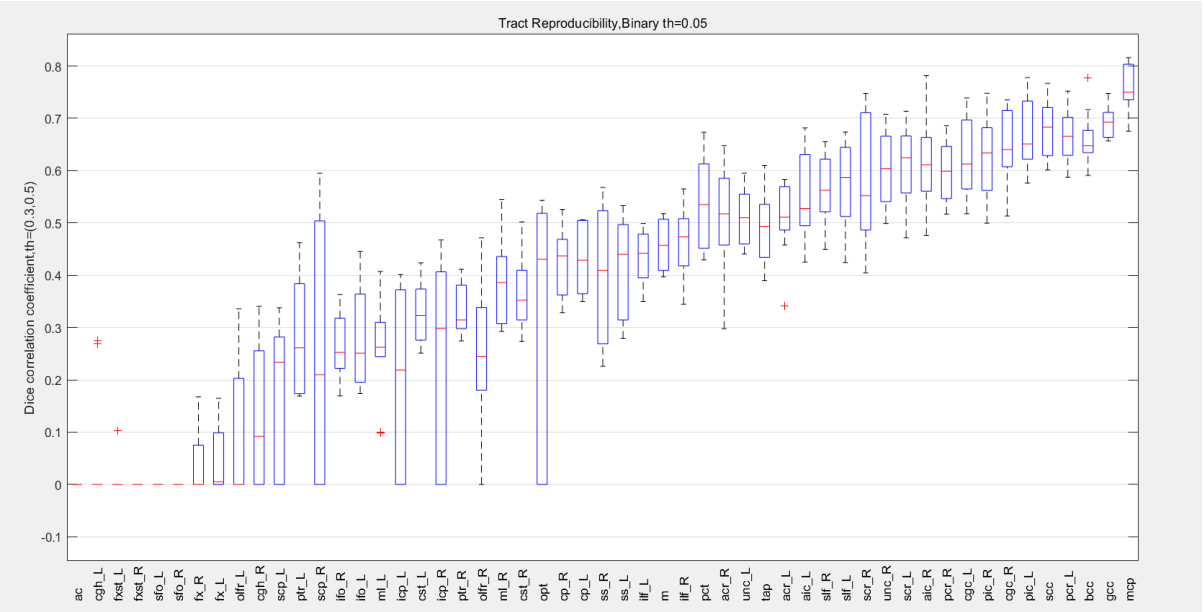
A.25. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch1. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 18 subjects from the BLSA18 dataset with first threshold=0.0 and voting threshold=0.5. The tracts are sorted from low to high reliability.



A.26. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch1. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 18 subjects from the BLSA18 dataset with first threshold=0.0 and voting threshold=0.3. The tracts are sorted from low to high reliability.



A.27. Tract reliability for all 53 fiber bundles is shown. These tracts are reconstruct by branch1. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the similarity of same-subject different-rater pairs for 18 subjects from the BLSA18 dataset with first threshold=0.0 and voting threshold=0.1. The tracts are sorted from low to high reliability.



A.28. Tract reproducibility for all 53 fiber bundles is shown. These tracts are reconstruct by both branch1 and branch2. Density images were used for similarity analysis and similarity was calculated using DICE. Each boxplot represents the DICE values of independent tract reconstructed from all the 38 subjects with voting threshold=0.3, 0.5.