Optical Flow-Based Image Registration in FLAIR MRI

by

Sergiu Mocanu
Bachelor of Engineering, Ryerson University, Toronto, 2017
Bachelor of Science, Dalhousie University, Halifax, 2012

A thesis
presented to Ryerson University
in partial fulfillment of
the requirements for the degree of
Master of Applied Science (M.A.Sc.)
in the program of
Biomedical Engineering

Toronto, Ontario, Canada, 2020

Author's Declaration For Electronic Submission Of A Thesis

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

OPTICAL FLOW-BASED IMAGE REGISTRATION IN FLAIR MRI

Sergiu Mocanu

Master of Applied Science (M.A.Sc.)

Biomedical Engineering

Ryerson University, 2020

Abstract

Medical imaging is one of the most common areas of computer vision research and algorithm development. FLAIR-MRI is particularly useful in highlighting damaged and necrotic tissue in brain images due to high contrast and resolution. Image registration is a method of warping images to the same geometric space to quantify tissue changes with accuracy. With advances in deep-learning via convolutional neural networks, complex problems can now move closer to some semblance of a solution with purpose-built and domain specific models. To overcome the non-learnable nature of current registration algorithms, ideas are adapted from video processing solutions of calculating optical flow between temporally spaced frames using unsupervised CNN-based methods to warp moving medical images to a fixed image space. The proposed total network loss combines pixelwise photometric differences, flow smoothness, and intensity correlation. Registration accuracy of the proposed and four other registration algorithms is measured by examining tissue integrity, pixelwise alignment, orientation, and global intensity similarity. The results, tested on two large FLAIR-MRI datasets consisting of 700 and 4000 brain volumes, show that the optical-flow registration technique is able to obtain maximal alignment while maintaining structural tissue integrity.

Acknowledgments

I would like to thank my supervisor Dr. April Khademi for teaching me new skills, unwavering guidance and support throughout, and showing me the importance of high-quality work in medicine and engineering.

I also owe great thanks to the committee members, Dr. Androutsos, Dr. Alirezaie, and Dr. Bagheri for their participation and input in the thesis defence process. Additionally, I'd like to thank them for being the first source of inspiration in the signals and image processing courses they teach.

I'd like to thank my BME friends, Sepehr, Faraz, and Sharif for their advice and help since second year undergrad and for being great company in and out of the classroom. I'd like to thank Sehar for helping me make the right choices and sharing all the ups and downs of grad school. Finally, I'd like to thank my mom for always believing in me and pushing me to be the best version of myself. I dedicate this work to her.

Table of Contents

List of Tables

List of Figures

List of Appendix Figures

Chapter 1

Introduction

According to a study by United Nations, 16% of the total population will be over the age of 65 by 2050 [1]. As the aging population increases, there will be higher need for late-age medical care due to an increased prevalence of age-related diseases. Multiple Sclerosis prevalence has increased by 69% from 1991 to 2013 in Canada [2] and about 5.5 million people in the United States and 24 million worldwide are affected by Alzheimer's that leads to dementia [3]. These staggering numbers have a significant burden on the healthcare system of every country. Much of the analysis for these patients involves medical image processing prior to any given treatment. Magnetic Resonance Imaging (MRI) offers a high resolution and specifically Fluid Attenuated Inversion Recovery (FLAIR-MRI) is particularly useful in detecting age-related lesions (Figure 1.1).

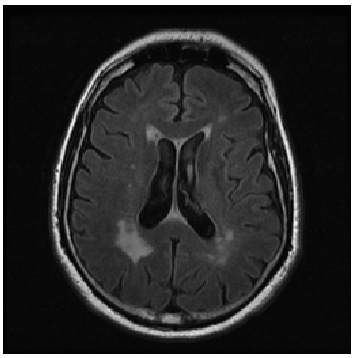


Figure 1.1: FLAIR-MRI image is able to show lesions as bright hyper-intense regions of the brain

In medical image analysis, reliable data usage is of utmost importance when it comes to algorithm design and clinical applications. As major medical centers are switching their imaging methods to the digital domain, the amount of data can increase to unmanageable levels. Much of it has to be normalized or placed in the same geometric domain so that meaningful information can

be obtained. An important part of all image processing workflow is geometric normalization via image registration. Registration is the process of aligning two images, one fixed and one moving so that they are in the same spatial domain (Figure 1.2). Image registration is particularly useful in automatic image segmentation pipelines as a pre-processing tool and can be used on two-dimensional images or three dimensional volumes. Since image components are non-uniform, deformations can be done globally via affine transformations or on a per-pixel level through the use of non-uniform deformation fields.

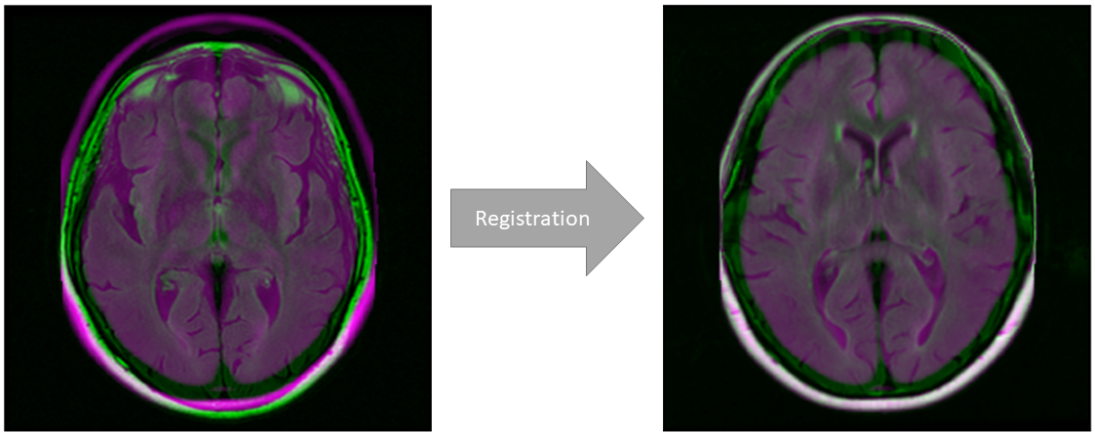


Figure 1.2: Registration involves transforming a moving image $M(x, y)$ to the fixed image $F(x, y)$ space.

This work proposes a registration method using Convolutional Neural Networks (CNNs) to predict the deformation field needed to warp a moving image to the fixed image space. Recent advancements in hardware development of graphic processing units (GPUs) have allowed a wide-spread application and uses of CNNs in the image processing field [4, 5]. In the video processing field, many have proposed methods to calculate velocities between temporally spaced frames via optical flow [6, 7]. Recent unsupervised methods have shown promising flow calculation between subsequent frames [8]. The implementation has been adapted to medical images to calculate the deformation field between the moving and fixed image and thus perform registration.

Alongside, this piece of writing will attempt to examine different image registration algorithms via several validation metrics. The focus is to evaluate the efficacy of each registration modality via intensity-based metrics, orientation metrics, and information theory metrics. From a clinical perspective, the integrity of structures in the images will be assessed prior to and after each registration technique. This chapter will serve as the foundation for the proposed work and its clinical relevance, a description of medical imaging techniques in MRI, and a brief description of the main technical parts.

1.1 Background

In neuronal tissue, specifically the cerebrum, the structural organization signifies underlying functional significance. One method of classification is by subdividing the tissue into white and gray matter. White matter consists of myelinated axons of the neurons while gray matter contains the cell bodies of the neurons [9]. Neurodegenerative diseases such as Multiple Sclerosis [10], Dementia [11, 12], and Alzheimers disease [13] are typically seen in the white matter of the brain. While the etiology of White Matter Lesions (WMLs) remains largely unknown, they are an indicator of brain disease and neuroimaging research offers new insights and possible explanations for their origination [14, 15]. A common method to observe these anatomical structures is to use Magnetic Resonance Imaging (MRI), which is known for its high resolution when it comes to examining soft tissue in the human body.

1.1.1 MRI, FLAIR

Magnetic Resonance Imaging (MRI) is an often-used medical imaging technique especially in brain tissue. From an imaging perspective, the main advantage of MRI over radiographic imaging such as X-Ray and CT is that it provides higher resolution and better contrast images of soft tissues, like brain matter, due to the underlying physics involved in the process.

The principle of MRI is based on the abundance of hydrogen atoms in the human body. An atom is composed of an inner core called the nucleus and is surrounded by fast spinning electrons. The

nucleus contains the protons and electrons of an atom. Hydrogen, atomic number 1, contains one proton and one electrons and zero neutrons. For that reason, it is common nomenclature to refer to hydrogen as a proton.

In its resting state, the hydrogen nucleus is imbalanced due to the odd number of protons. As the atom rotates, this imbalance causes the atom to have an angular momentum. This momentum on a large scale can be measured as each atom spins and creates a small magnetic field. The magnetic field can be measured by the following equation:

$$\vec{\mu} = \gamma \vec{J}, \tag{1.1}$$

where $\vec{\mu}$ is the magnetic moment, γ is the gyromagnetic constant, and $\vec{J}$ is the spin angular moment. The gyromagnetic constant is measured in $megaHertz/Tesla$ and is a known value for most elements [16].

The protons spin with a specific energy and the axis about which it spins is oriented completely randomly. The spinning, also known as precessing, is similar to a spin-top that is swaying as it spins (Figure 1.3). In nature, due to random orientation of the spinning axis vectors, the net energy is close to zero. In the presence of a magnetic field, the axis about which the precessing occurs, become aligned and they are unidirectional. As the nuclei spin-axis becomes aligned, the net magnetic energy of the nuclei in the direction of the field is higher than those that are aligned transverse [16].

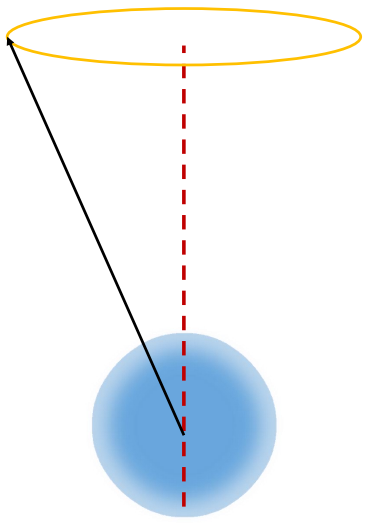


Figure 1.3: The angular momentum of a hydrogen atom.

A strong magnetic field can be applied to align the protons in the direction of the magnetic field. The spin-axes of the protons is in the same direction making the net magnetic vector point in the same direction as the applied magnetic field. At this point, the nuclei are not in phase with each other, meaning they are all spinning with different frequencies. One method to calculate the precession frequency, is to find the Larmor frequency by differentiating Equation 1.1 to obtain:

$$\omega_0 = \gamma H_0, \tag{1.2}$$

where γ is the same gyromagnetic ratio and H_0 is the strength of the external magnetic field. By finding the ω_0, the Larmor frequency can be used to apply the RF pulses and change the spinning orientation of the hydrogen atoms [16], [17].

Once the hydrogen atoms are all spinning with the same angular momentum, the net magnetic vector now contains a y-component also known as a 90° pulse. This means that net magnetization vector which was previously in the same direction as the magnetic field and had no y-component, now will be pointing in the transverse direction of the applied magnetic field. This vector is now at a higher energy state than previously. By applying another 90° RF pulse, the net magnetic vector is

now pointing in the opposite direction than the applied magnetic field and thus being at the highest possible energy state [16].

As the protons return to the initial resting state prior to the RF pulse induction, they emit RF pulses themselves. These pulses can be detected by RF detector coils in the MR machine. The two types of relaxation, T1 and T2 are longitudinal and transverse relaxation, respectively. If a proton flips from a high-energy state to a low-energy state, this is known as the T1 relaxation or spin-lattice. As the precessing protons move out of phase, the net magnetization vector returns from transverse to zero xy-plane and this energy is known as T2 or spin-spin. Different tissues types vary in terms of their relaxation times with respect to magnetization. In T1 water molecules return faster to the initial magnetic field direction, while fatty tissue is slower. Similarly, T2 relaxation times also vary with tissues (Table 1.1).

Table 1.1: T1 and T2 relaxation times for different tissue types [16].

Tissue	$T_1(ms)$	$T_2(ms)$
Fat	150	150
Liver	250	44
White Matter	300	133
Gray Matter	475	118
Blood	525	261
CSF	2000	250

As the net magnetic vector experiences relaxation, the RF coils are able to measure the decay of the magnetic moment called Free Induction Decay (FID). The movement from transverse to longitudinal axis, the coils are able to detect the FID and observe the reduction in energy over time. This electromagnetic signal is received by the computer system and is converted from analog to digital. The signal is temporarily stored in an imaging space called the k-space. The k-space is a two- or three-dimensional representation of the frequency intensities and phase localization of the signal. By performing an inverse Fourier Transform of the frequencies in the k-space, the image is obtained. To obtain the localization of different slices, a magnetic gradient is used. One can think of the gradient as a wedge that attenuates the applied external magnetic field. Since, the amount

of attenuation is a known value, as the signal is being received by the RF coils, the location of where the signal is coming from the target space is also known due to the attenuation factor of the magnetic gradient. The magnetic gradients are located in all x, y, and z directions [16].

One method of improving the image quality such as contrast and resolution is by altering the strength of the pulses and the sequencing. The spin-echo pulse usually consists of applying a $90°$ pulse, waiting, and then applying a $180°$ pulse. This effect cancels the external inhomogeneity (T2*) of the external applied magnetic field. The result is an Echo that resembles a sinc function and the time between the $90°$ pulse and the Echo is known as the time to echo (TE). At this point, the protons return to relaxation and another $90°$ pulse can be applied which is measured as the time to relaxation (TR) (Figure, 1.4). TE and TR are key factors to obtaining high contrast images. By adjusting the TE and TR values, different pulse sequences are produced. Shorter TE and TR results in T1 weighted images, long TR and short TE results in proton-density, and both long TE and long TR results in T2 weighted images [18] [19].

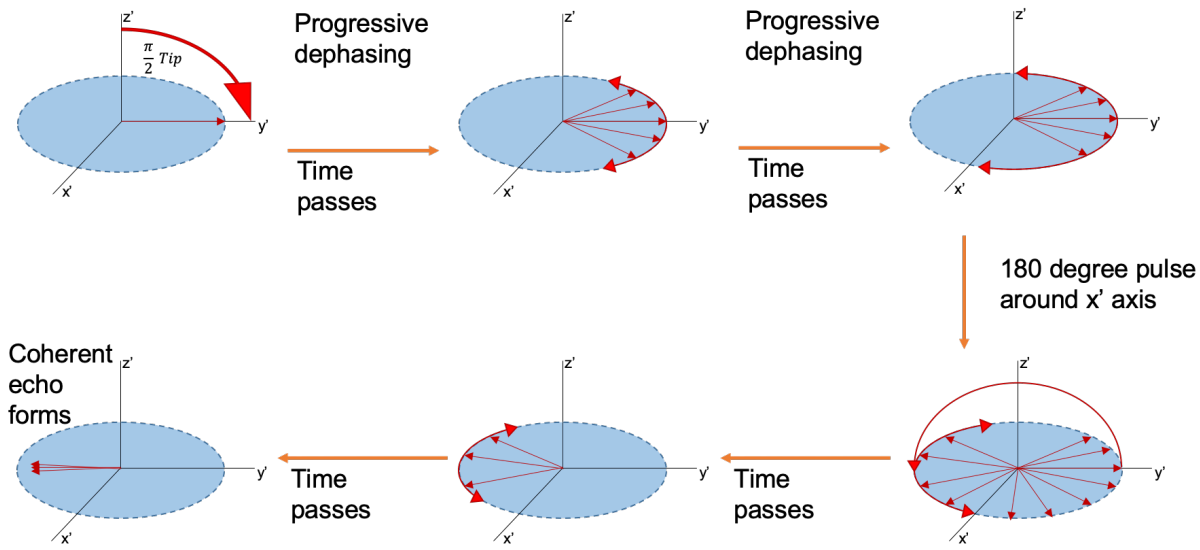


Figure 1.4: Formation of a spin-echo [19].

Inversion Recovery is the application of a prepratory pulse of $180°$. As the nuclei begin to relax and the net magnetization vector passes the transverse plane, a $90°$ pulse is applied and the normal spin-echo sequence is applied. The time between the $180°$ and $90°$ is known as inversion time (TI). This type of sequencing causes water, which has a low magnetization factor after the $90°$ pulse,

appears dark during TI [20].

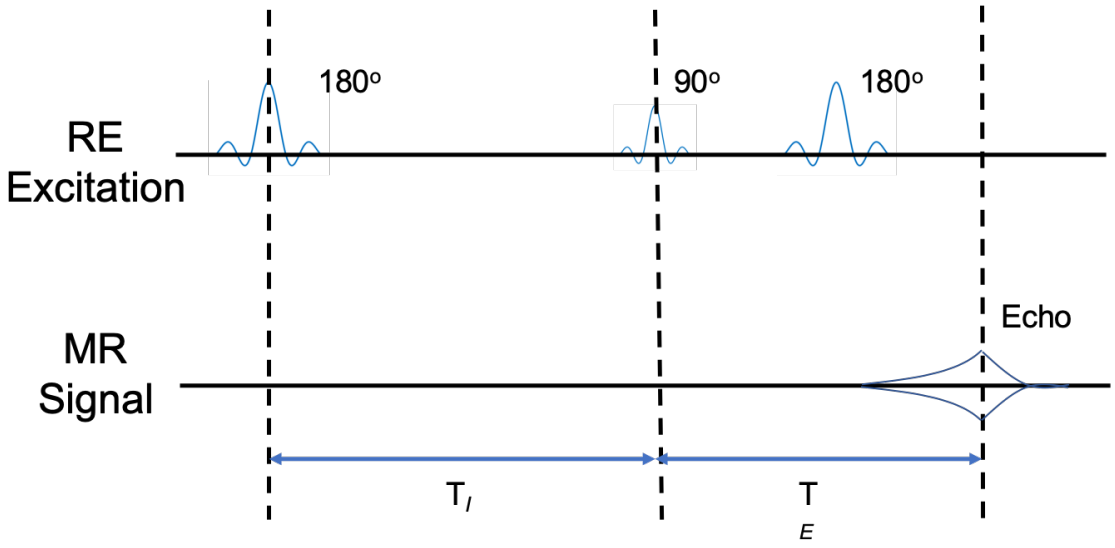


Figure 1.5: Inversion Recovery pulse sequencing [19].

Fluid Attenuated Inversion Recovery or FLAIR takes advantage of this technique to accurately delineate brain lesions that are near or close to cerebrospinal fluid (CSF) locations like the ventricles. This is because as the net magnetization vector that received the $180°$ preparatory pulse passes the null transverse vector, water-containing-CSF, there is little to no longitudinal net magnetization vector that is typical of water. Therefore, no signal is acquired by the RF coils of the machine and true to its name, the fluid becomes fully attenuated (Figure 1.6).

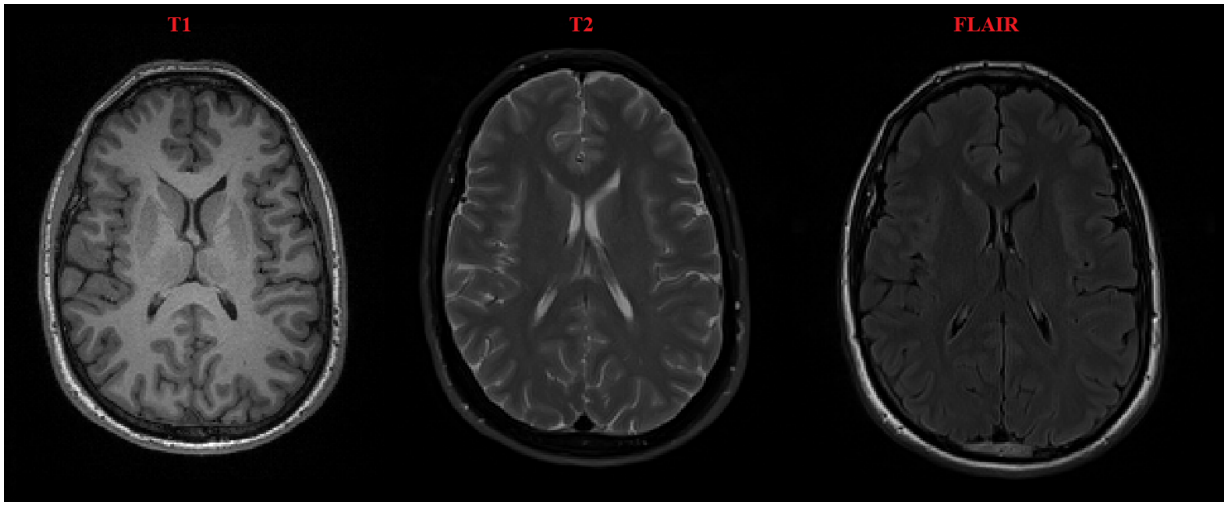


Figure 1.6: Examples of images recovered using T1, T2, and FLAIR sequencing.

1.1.2 White Matter Lesions

FLAIR imaging is particularly useful in examining White Matter Lesions (WML) in neurodegenerative and in aging patients. As mentioned previously, this is particularly useful because fatty tissue along with CSF is suppressed in FLAIR. In the brain, the white matter, which contains the axons of the surrounding cell-body neurons is suppressed because the axons are surrounded by a fatty tissue called the myelin sheath. The myelin sheath is particularly useful in neurons because it acts like a blanket that insulates and allows for faster signal transmission from the cell body to the axon terminal via saltatory conduction [9]. This suppression is similar as in T2 imaging seen above in Figure 1.6, however, one other benefit of FLAIR is the CSF suppression. Many neurodegenerative disorders consist of white matter hyperintensities that are next to the ventricles also known as periventricular, which are well defined in FLAIR imaging [21].

FLAIR MRI is becoming a widely-used imaging modality among clinicians involved in nervous disease treatment and research [21, 22, 23, 24], , which creates a need for robust and easy to use tools to aid with diagnosis, treatment, disease monitoring, prognosis, and outcome. Image processing algorithms are being developed for modality-specific purposes as part of a clinician's daily workflow [21]. Many tasks begin with a pre-processing step that can standardize all the incoming data for a more robust algorithm result. Some of the most important tasks in algorithm design such as WML segmentation or disease prediction lay in the early stages of data augmentation and standardization. These steps can significantly improve algorithm performance due to the fact that the raw data is now all on the same playing field and there aren't any portions that have a higher impact while others have less of an impact. One of the most important steps is to transform images to the same geometric space via image registration.

1.2 Problem Statement

Typical registration algorithms involve an iterative-based approach that calculates an adequate deformation parameter to warp a moving image to the fixed geometric space [25,26,27,28]. The algorithms typically try to minimize a dissimilarity metric at every iteration through a process known

as gradient descent. The main downside of such approaches is that the process is computationally expensive and any calculated parameters are not saved or learned so that the following registration procedure may be more efficient or accurate. This means that after calculating a transformation parameter for a specific moving-fixed image pair, the parameters are discarded for the next pair of images. The non-transferable nature of traditional image registration algorithms has created a need for a learning-based registration method. These models should learn any transformation parameter and accurately warp images from one pair to the next and thus be able to generalize to larger data and other image pairs.

Image registration has become integrated in most medical image processing workflows as a method of transforming images to be in the same space. Algorithms in white matter lesion segmentation (WML) are highly reliant on registration to improve their performance [24]. Researchers and clinicians often work side-by-side during algorithm development [29]. Specifically, there is a high reliance on experts to provide an opinion in terms of disease progression, lesion annotation, or efficacy of image processing algorithms [30]. The clinicians then have to take crucial time outside of their practice to evaluate and be active participants in research. At the same time, if several clinicians are asked to perform a task, such as annotating WMLs, there may be some variation in the opinions of what pixels to include and which ones to exclude. While there are measures to assess the different raters reliability in such segmentation results [31, 32], there is a need for consistent measures of algorithm performance. This is especially true when it comes to image registration.

As there is no Ground Truth (GT) when it comes to quality assessment of registration results, most clinicians typically visually inspect for any major anomalies as the only means of evaluation; this can create inconsistencies throughout different results. In a clinical setting, registration is typically used to assess the evolution of a diseased tissue or of any major anatomic structures. Similarly, researchers and algorithm developers typically register images to a certain space and once a certain similarity metric reaches an optimum value, they may move forward with the segmentation algorithm. In both cases, subtle changes in structure often times can signify major developments in function, tissue evolution, or treatment efficacy. To accurately review these changes, the images have to be warped so that they are in the same geometric space therefore

meaningful conclusions can be made. However, even before that, it is important to know whether this warping or registration has been accurately done. To do that, there is a need for a consistent method of evaluating registration quality and validating that the results are as accurate as possible [33].

The challenges with image registration can come from the variation of the data. As scanners present intensity inhomogeneity across different manufacturers and models this can give a large variation of image quality metrics in terms of sharpness, contrast, and noise levels. Similar variations in brain anatomy between and within patients can give way to larger differences that have to be accounted for when applying algorithms that involve segmentation and registration. Finally, factors like hyperintense and hypointense regions of the brain can lead to drastic changes in local intensities [33].

Since visual inspection of image registration quality is a daunting task, tools have to be designed to accurately and consistently measure the alignment of images in registration tasks. The tool should be fully automated and measure several aspects of image registration, from intensity similarity to geometric and volumetric similarity. The tool should also be robust so that it can be used with the results of any registration method and be able to provide these metrics instantly to a researcher or a clinician so that adjustments and tunings can be done at the time of experiment and reduce the need for repeat trials.

1.2.1 Objective

The aim of the work presented in this thesis is to adapt a model used in natural images to predict optical flow to design an unsupervised CNN-based registration tool. Further, image registration quality validation metrics algorithms from several similarity perspectives based on structural integrity, volumetric consistency, intensity levels, information theory, and geometric orientation. Secondary objectives include:

- Generation of a FLAIR-specific atlas generation
- Longitudinal registration of repeat scan subjects

The methods developed in this thesis can be used during algorithm development to ensure that image registration is accurate prior to any implementation steps. The optical flow-based registration technique is based on deep learning with the use of convolutional neural networks (CNN) to register volumes and generate necessary warp fields. The proposed registration validation techniques were used to evaluate the most optimal registration for the generation of an atlas. Longitudinal registration allows for quick evaluation of patient brain tissue change over time.

1.3 Proposed Techniques

This section will highlight the two major themes of image registration in FLAIR MRI.

The first portion proposes a new fast medical image registration technique via neural networks and deep-learning methods based on optical flow. The network is trained on FLAIR MRI images and is able to accurately predict the deformation fields via a dense per-pixel deformation field.

The second portion is related to a proposal of new validation metrics for measuring image similarity and the efficacy of image registration techniques. Two metrics are related to evaluating the volumetric and structural changes of brain tissue. The first being percent volume (PV) change which looks at the proportional change of brain structures as a function of total brain volume. The second measure is a structural change to see how it warps with respect to the brain surface, named Surface to Surface Distance (SSD). Both PV and SSD are implemented using binary masks that are warped according to the same field that is generated for warping a moving image to the same space as the fixed image. Image alignment is measured as a Head Angle (HA) measure. The typical image similarity metrics were also implemented to confirm the efficacy of the aforementioned techniques. Four common image registration algorithms were tested and evaluated according to the proposed metrics.

1.4 Contributions

The main contributions of this work is related to developing an optical-flow based registration technique and evaluating and validating image registration techniques in FLAIR MRI according to the following:

1. Propose an adapted optical flow-based fast FLAIR-specific registration technique.
2. Inclusion of a correlation loss component.
3. Image registration validation via volumetric, structural, and alignment metrics.
4. Test four existing algorithms with the proposed validation measures and existing measures.
5. Validate novel registration against existing ones with the use of the proposed validation measures and existing measures.

The thesis is organized in the following manner: Chapter 2 will explore the methods and techniques of image registration from the most basic concepts to most current methods, Chapter 3 will explore the main contributions of this thesis, optical flow-based registration and registration validation metrics, Chapter 4 will describe the performed experiments and results, Chapter 5 explores the clinical implications of image registration, and Chapter 6 will present a high-level overview summary of the proposed methods along with the limitations and suggestions for future works.

Chapter 2

Image Registration

Image registration is an iterative image processing technique used for matching one moving image to a stationary or fixed one. The purpose is to find the optimal geometric transformation that indicates the highest level of similarity between the two images. Registration can be *multimodal*, images obtained via two different methods, Computer Tomography (CT) to MRI, T2 MRI to T1 MRI or they can be *monomodal* where images take via the same method are registered to each other. Monomodal image registration is typically done for longitudinal measures where a patient may receive multiple scans and to accurately analyze them, they have to be warped to the same geometric space [34] [35].

Earlier methods of registration required input from a user to specify corresponding points between two images defined as control points. Based on these control points, the images would align to minimize the distance between the control points. This technique is seen often times, in older software that is used to stitch multiple images together to form a panorama. The work presented, will focus on automatic image segmentation, where no user input is required. Particularly it will be focused on monomodal image registration in MRI. The contents of this chapter are organized in the following manner: the types of transformations used, the method of image warping and pixel sampling from the moving space to the new fixed space, a deep dive in the most common and some newer image registration techniques, an explanation of the theory behind optical flow estimation and Convolutional Neural Networks (CNNs).

2.1 Transformation Matrix

To accurately understand the basics of image transforms, one must first understand the basics of image composition. At the most simple level, a gray-scale image can be thought of as a matrix of pixels, where each pixel has an intensity value, between a certain range, and a coordinate based on its location in the image. A pixel is the smallest functional unit of an image, and together with all the neighboring pixels makes an image. A voxel is a volumetric pixel that has a depth component. Voxels are seen in medical images where three-dimensional scans of body tissue is obtained.

Transforming images involves moving a pixel from one pair of coordinates to another. Image coordinates are different than typical coordinates, where the origin, (0, 0) is located at the top left corner. In the description of an image as a matrix, the rows become the Y coordinates, and the columns are the X coordinates in the (X, Y) form. Knowing this, images can now be transformed via matrix operations. If an image I is multiplied with a transformation matrix T, the new transformed image is I':

$$I' = I \times T. \tag{2.1}$$

Using matrix multiplication one can now see mapping of (X, Y) coordinates to the new (X', Y') space according to the following operation:

$$\begin{bmatrix} X' \\ Y' \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}. \tag{2.2}$$

Through matrix multiplication, pixel coordinates in one image can be mapped to new location. It is important to note, that in matrix multiplication the order of operations matter and to achieve adequate results, the vector inputs of the pixel coordinates should always follow the transform matrix. In the following section, the different types of transformations will be examined.

2.2 Types of Transformations

Image transformations are typically classified into two types: rigid and non-rigid. Rigid transformation preserve distances, straightness of lines, and all non-zero angles between straight lines. Non-rigid transformations do not preserve lengths of straight lines or non-zero angles. Affine transformations describe how many transformations can be combined and projective transforms involve the warping of angles in an image.

2.2.1 Rigid (Translation, Rotation)

If all the coordinates of an image I are stored in vectors (X, Y), a rigid transformation can be defined as follows:

$$T(v) = R(v) + t, \tag{2.3}$$

where v is the one of the coordinate vectors, R is the rotation orthogonal transformation, and t is the translation. The rigid transformation can perform both rotation and translation simultaneously. One method of determining that indeed the transformation is a rigid transform is showing that:

$$det(R) = 1, \tag{2.4}$$

if $det(R) = -1$ then a reflection has occurred. For example, if a pixel is located at the image coordinate $(1, 1)$ a rigid transformation may be represented by a rotation of $\theta = 15°$ and translation of $tx, ty = 10$ in both vertical and horizontal axis. The rotation matrix is seen below and is based on orthogonal transformations where the shape is preserved (Figure 2.1).

$$\begin{bmatrix} X' \\ Y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}. \tag{2.5}$$

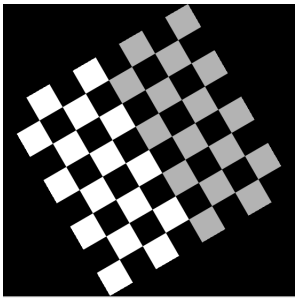


Figure 2.1: Rotation by 30 degrees.

Translation is a linear remapping of pixels to a new coordinate system by simple movement of fixed value and can be done using the transformation matrix in the following implementation (Figure 2.2):

$$\begin{bmatrix} X' \\ Y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & tx \\ 0 & 1 & ty \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}. \tag{2.6}$$

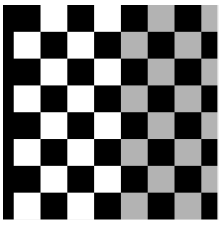


Figure 2.2: Translation to the right by 20 pixels.

One can see from the matrix operations defined in 2.5 and 2.6 that the rigid transformation can be done in one step by combining the two matrices and thus achieving the rigid transform in 2.3.

$$\begin{bmatrix} X' \\ Y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & tx \\ \sin\theta & \cos\theta & ty \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}. \tag{2.7}$$

Returning to the example above, the mapping of the image coordinates is located at the new pixel

location as described by the following expression:

$$\begin{bmatrix} 10.7071 \\ 11.2247 \\ 1 \end{bmatrix} = \begin{bmatrix} \cos 15 & -\sin 15 & 10 \\ \sin 15 & \cos 15 & 10 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \tag{2.8}$$

If the coordinates of all pixels are stored in vector arrays, the matrix multiplication can quickly relocate all pixels to the new vector space. One way to verify that a transformation is indeed rigid is to measure the Euclidean distance between any two pixels in the original image and again between the same two points in the transformed image. The Euclidean Distance between two points p, q, is based on the Pythagorean formula and is described as follows:

$$d(p, q) = \sqrt{(p_X - q_X)^2 + (p_Y - q_Y)^2}. \tag{2.9}$$

2.2.2 Non-rigid (Scaling, Shear)

Non-rigid transformations are especially useful in image registration problems where there may have been some image distortion during acquisition. The simplest non-rigid transforms are scaling and shear. Scaling can be defined by the following operation (Figure 2.3):

$$I' = S \times I. \tag{2.10}$$

where S is a diagonal matrix where each component along the diagonal is a scaling factor of the corresponding coordinates:

$$\begin{bmatrix} X' \\ Y' \\ 1 \end{bmatrix} = \begin{bmatrix} sx & 0 & 0 \\ 0 & sy & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}. \tag{2.11}$$

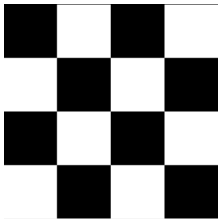


Figure 2.3: Scaling by a factor of 2.

If the scaling is isotropic, the same in all directions, the transformation is defined as a *similarity transformation*, also known as a dilation, even for values less than one [36]. Shear is defined as a deformation by ”pushing” one portion of the image along an axis, such that parallel lines remain parallel, but angles are not preserved 2.4. The transform matrix for shear is as follows:

$$\begin{bmatrix} X' \\ Y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & shx & 0 \\ shy & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}. \tag{2.12}$$

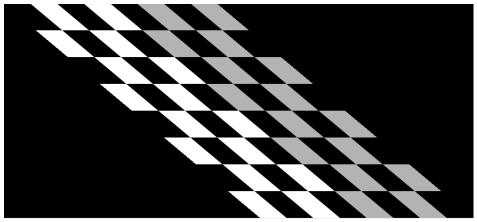


Figure 2.4: Shear to the right.

2.2.3 Affine

All the transform matrices mentioned above can be combined through matrix multiplication to create one matrix that performs all transformations in one operation. Affine transformations inherit all the properties of the rigid and non-rigid transforms. The straightness of lines, parallelism, and the planes are all preserved, however angles are not (due to shear transform). A combination of all

transforms may be as follows [1]:

$$\begin{bmatrix} X' \\ Y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & tx \\ \sin\theta & \cos\theta & ty \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} sx & shx & 0 \\ shy & sy & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}. \tag{2.13}$$

One important aspect of the affine transformation is the order that transformations are performed since matrix multiplication is not commutative ($AB \neq BA$). If we take our earlier example of the pixel coordinate $(1, 1)$ and perform the following operations, a rotation by $15°$, translation by 10 pixels, scaling of 1.5, and no shear, in this order, the output is as follows:

$$\begin{bmatrix} 11.0607 \\ 11.8371 \\ 1 \end{bmatrix} = \begin{bmatrix} \cos 15 & -\sin 15 & 10 \\ \sin 15 & \cos 15 & 10 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1.5 & shx & 0 \\ shy & 1.5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \tag{2.14}$$

If the order of the two transform matrices are switched the result is significantly different:

$$\begin{bmatrix} 16.0607 \\ 16.8371 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.5 & shx & 0 \\ shy & 1.5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos 15 & -\sin 15 & 10 \\ \sin 15 & \cos 15 & 10 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \tag{2.15}$$

Therefore, it is important to choose a method that aligns with the scope of the problem at hand. One convention for example, is to translate the center of the image to the origin of the coordinate system, perform a rotation, translate back, perform the translation parameter, then scaling (*RTS*).

[1] Note: translation and rotation are combined in one matrix and shear and scaling are combined in another matrix.

2.2.4 Projective

The same affine matrix can be used for projective transforms. Projective transformation preserves straight lines and relates the image to a flat scene of the environment. One way to think of projective transformations is that opposing lines of the image intersect on a flat plane in the background at a point as seen in Figure 2.5.

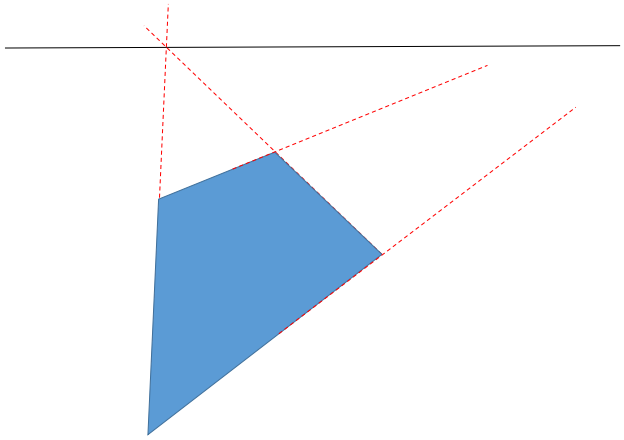


Figure 2.5: Projective transformation visualization. The blue figure is an image for which the two vertical edges are transformed so that their projections meet a point on the plane represented by the black line.

This type of projective transform is called a *perspective* transformation. Perspective transformations use homogenous coordinates. Homogenous coordinates are the input vectors as we saw in all the transformation matrix operations, where the first two vectors are the *(X,Y)* coordinates and the third coordinate is 1. In an affine matrix (Equation: 2.2) the g and h denote the projective transform vectors:

$$\begin{bmatrix} X' \\ Y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ px & py & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}. \tag{2.16}$$

As such, one can see that that projective transforms can be added to any affine matrix since affine matrices have zero values for the projective components so that parallel lines can be preserved.

2.3 Deformation Fields

Deformation fields can be easily understood as a field where each pixel in an image has a vector representation of where it should be mapped in the transformed grid space. The vector magnitude represents the amount of pixel displacement while the angle shows the direction. This type of transformation is particularly useful in non-uniform image registration problems. Most commonly, deformation vector fields are used in modern image registration algorithms due to the fact that they can perform fine image adjustment (Figure 2.6), whereas affine matrix-based registration is coarse adjustment.

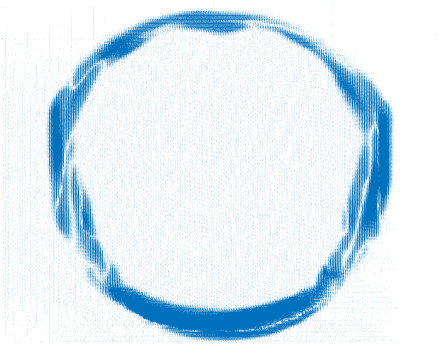


Figure 2.6: Example of a deformation field used to warp a moving image.

2.4 Image Warping

After a transformation is applied to an image, the individual pixels are now remapped to a different coordinate system. In an image, coordinates are integers, however after transformation is applied, pixels are remapped to non-integer values (Figure 2.7).

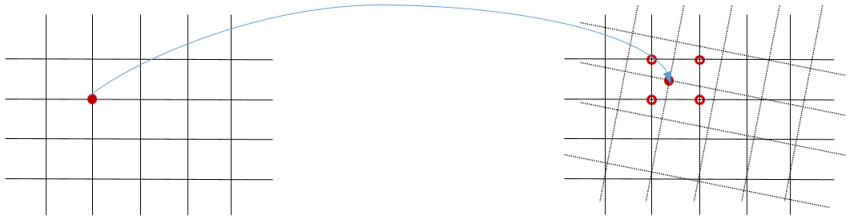


Figure 2.7: Forward mapping of a pixel from the image on the left to a new location in the image on the grid of the new image on the right.

Calculating the new pixel values is computationally expensive, as such, to speed up the process of pixel remapping, the backwards or inverse mapping is performed. For the known homogeneous coordinates in the new warped space, every pixel is backwards sampled from the original image according to the following equation:

$$I = A^{-1} I'_g, \tag{2.17}$$

where I is the original image, A^{-1} is the inverse of the transformation matrix A, and I'_g is the grid coordinate system of the transformed image. This method has the advantage of not having to create new pixel values, but rather sampling from the original image and placing into the new grid space (Figure 2.8). To overcome the floating pixel coordinates, interpolation methods are applied which will be discussed in the coming section (Section 2.4.1).

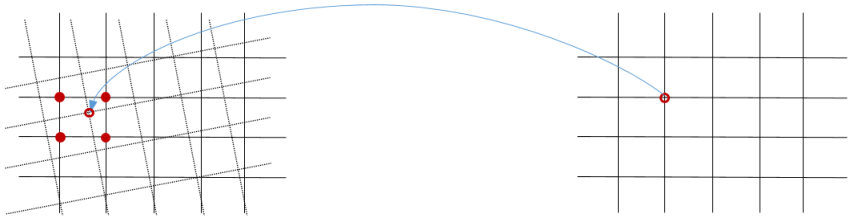


Figure 2.8: Backward mapping by sampling from the original image into the new image grid space.

2.4.1 Interpolation (Nearest-Neighbour, Bilinear, Bicubic)

The quickest way to interpolate is via nearest neighbour interpolation. If the new pixel mapping is located between four pixels, the Euclidean distance can be found and the new pixel value will take the intensity of the closest pixel (Figure 2.9).

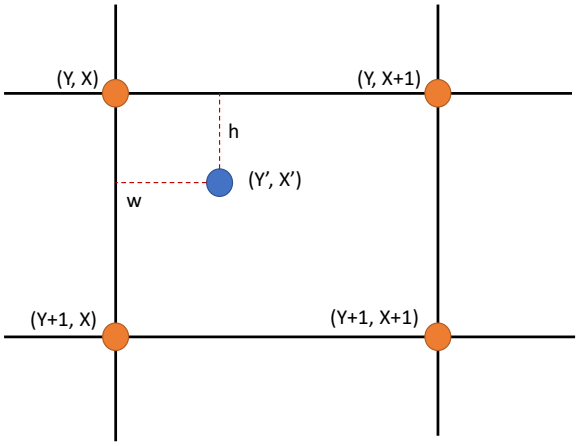


Figure 2.9: Pixel intensity mapping of non-integer values of pixel (Y', X').

Since pixel coordinates are integers, the distance between two pixels is taken to be 1 and any number in between is a floating point value. As seen, the mapping often falls at these non-integer values. The smallest distance is determined only from a pool of the four nearest pixels, and whichever is smallest, that corresponding pixel's intensity is assigned to the new mapping calculated by:

$$d = \sqrt{a^2 + b^2}, \tag{2.18}$$

where a is the distance in the X direction and b is the distance in the Y direction[2].

Bilinear interpolation is a weighted average between pixels to create new pixels with new interpolated values. If the location of a new pixel is known, then the new image intensity can be calculated

[2]Note that distance from X+1 pixel coordinate to X' is $1 - w$ and from Y+1 to Y' is $1 - h$.

as follows:

$$I'(X',Y') = W_{X,Y}I(X,Y) + W_{X+1,Y}I(X+1,Y) + W_{X,Y+1}I(X,Y+1) + W_{X+1,Y+1}I(X+1,Y+1), \tag{2.19}$$

where the $W_{X,Y}$ is the area made by $w \times h$, $W_{X+1,Y}$ is the area made by $(1-w) \times h$, $W_{X,Y+1}$ is the area made by $w \times (1-h)$, and $W_{X+1,Y+1}$ is the area made by $(w-1) \times (h-1)$. The advantage of bilinear interpolation over nearest neighbour interpolation is that the pixel values are not repeated and the image is smoother and edges are preserved.

Bicubic interpolation involves fitting a bicubic spline to a linear gradient of two points. The idea is that if a pixel is located in between four pixels, the interpolation is now based not only on the four pixels surrounding the new location, but also the further out area. This is because when the gradient of the neighbouring pixels is found, they rely on the neighbouring pixels to produce the gradient. Based on that curve, one can now calculate the intensity value of the new pixel. This method creates even smoother images than bilinear interpolation, but because of the derivative calculations and increased sampling area, it is significantly more computationally intensive. The bicubic interpolation summation formula can be seen below:

$$I(X,Y) = \sum_{i=0}^{3}\sum_{j=0}^{3} a_{ij}x^i y^j. \tag{2.20}$$

The scope of the equation is to solve for the a values. This can be done via matrix multiplication since, and the $I(X,Y)$ intensity values form a 4×4 area of 16 pixels, from which the derivatives have to be calculated. The matrix approximation is as follows:

$$I'(X',Y') = \begin{bmatrix} X'^3 & X'^2 & X' & 1 \end{bmatrix} \begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{02} \\ a_{10} & a_{11} & a_{12} & a_{13} \\ a_{20} & a_{21} & a_{22} & a_{23} \\ a_{30} & a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} Y'^3 \\ Y'^2 \\ Y' \\ 1 \end{bmatrix}. \tag{2.21}$$

Solving for the a values can be done through the following matrix operation:

$$\begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{02} \\ a_{10} & a_{11} & a_{12} & a_{13} \\ a_{20} & a_{21} & a_{22} & a_{23} \\ a_{30} & a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -3 & 3 & -2 & -1 \\ 2 & -2 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} f(0,0) & f(0,1) & f_y(0,0) & f_y(0,0) \\ f(1,0) & f(1,1) & f_y(1,0) & f_y(1,1) \\ f_x(0,0) & f_x(0,1) & f_{xy}(0,0) & f_{xy}(0,1) \\ f_x(1,0) & f_x(1,1) & f_{xy}(0,0) & f_{xy}(0,1) \end{bmatrix} \begin{bmatrix} 1 & 0 & -3 & 2 \\ 0 & 0 & 3 & -2 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \tag{2.22}$$

where the f are the derivatives of the four corners in four directions for each corner around the pixel of interest. The derivatives in an image are calculated as differences in intensity of one pixel to the previous pixel in the coordinate system. Since cubic interpolation create new pixels, sometimes, overshooting of the spline curves occurs and new intensity values are created that are higher than the maximum an image format can display (Ex: 256) so the new values may be clipped and this may create unnatural fuzzy edges around areas of high contrast. Further interpolations exist via bicubic convolution and cubic splines which will not be explored, for the sake of brevity, since bicubic interpolation is the most common method of resampling an image in most image processing algorithms [37].

2.5 Registration Algorithms

Although there are point-based registration techniques, where fiducial landmarks between two images are known, the focus of this body of work is established on automatic registration techniques. Automatic image registration techniques are based on intensity measures between two images. These types of registrations have become some of the most widely used methods, particularly

in medical images. The advantage of intensity-based registration measures is that the similarity problem can be solved for images that are taken via different modalities. For example images that are obtained via CT can be registered to MR, PET to MR, and even Ultrasound to MR.

Intensity-based registration involves finding a deformation field between images based only on intensities. This is typically an iterative process that involves calculating a similarity metric based on intensity values. While, all pixels can be used for image registration, the implementation is done on a portion of the image. This can significantly improve the speed of the algorithm. The following sections describe the most commonly used registration algorithms in research and clinical environments. Most novel registration algorithms are compared to one or two of the mentioned registration modalities.

2.5.1 Demons

In image processing, researchers often take inspiration from other fields and incorporate those concepts in a novel way of producing desirable results. An interesting application of this type of transfer learning is described by Thirion's demons [25]. The idea is based on thermodynamics and Maxwell's demons. In thermodynamics of diffusion, typically the entropy is decreased by having high energy molecules move to a lower energy state and achieve equilibrium. Maxwell introduced some demons that act as controllers of gated channels that allow only high energy molecules on one side of a container and low energy on another side. From a palpable perspective, this would mean that a warm object would get warmer and a cold object would get colder [38].

In terms of registration, the same concept was first proposed by Thirion. He suggested that if demons are placed at certain locations along the perimeter of the moving image, they would act as the controllers that sort particles or pixels in one side or another. The particles would move the image inward if it is larger than the static reference image and move it outward if it is contained within the static reference image (Figure 2.10) [26].

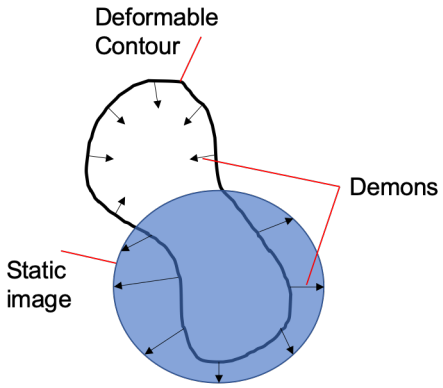


Figure 2.10: The action of the Demons to deform the contour of a moving a image according to a static image.

Thirion's registration method is based on a gradient descent algorithm. Specifically, if there are two images R and T and there is a demon **d** at an (x, y) position which exerts a force p. This force depends on the gradient of R and the image difference $R(d) - T(d)$. The force deforms the image with a positive gradient if $R(d) < T(d)$ and with a negative gradient if $R(d) > T(d)$.

One of the most apparent properties of images registered with the demon's algorithm is that it is an iterative process which requires several considerations. One of the considerations is the number and types of demons. This is determined via the optical flow equation which is as follows:

$$0 = \frac{\mathrm{d}}{dt}R(x,t) = \partial_t R(x,t) + (\nabla R(x,t))^T \partial_t x. \tag{2.23}$$

Since a pixel intensity does not change with time, the equation is equated to zero and can be easily solved. By setting $v = \partial_t x$ and rearranging the equation 2.23, Thirion suggests adjusting the equation such that the value of v is reduced. This is because if there are any small perturbations they may lead to amplified errors [26], [34]. Due to these constraints, the algorithm becomes susceptible to small errors and is somewhat computationally expensive in terms of operational time (1-5 minutes runtime). The calculated displacement is smoothed with a Gaussian filter:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}, \tag{2.24}$$

where σ is the standard deviation of the distribution and is what is used to adjust the amount of blurring or smoothing.

2.5.2 ANTs

Diffeomorphisms are differentiable map for which the inverse is also differentiable [39]. Avants [28] has proposed a diffeomrphic registration technique that is based on Lagrangian diffeomorphic registration. They proposed that when an image I is being deformed to match another static image J the deformation field path from one to another is the same. As one image moves along the diffeomorphism, the image warps to match the other. The same is true for the static image moving along the deformation path to match to the moving path. In a sense, this path is symmetric and therefore it is not dependent on the optimizer or similarity metric. This is advantageous over typical registration processes because the direction of the path does not matter and backward or forward transforms can be calculated without having any interpolation errors. Yet another advantage of this symmetric approach is that the difference between the two images can be thought of as a linear timescale of how one image warps to become another. This symmetry means that the path between the two images can be divided in half and the deformation can be calculated from both directions simultaneously (Figure 2.11).

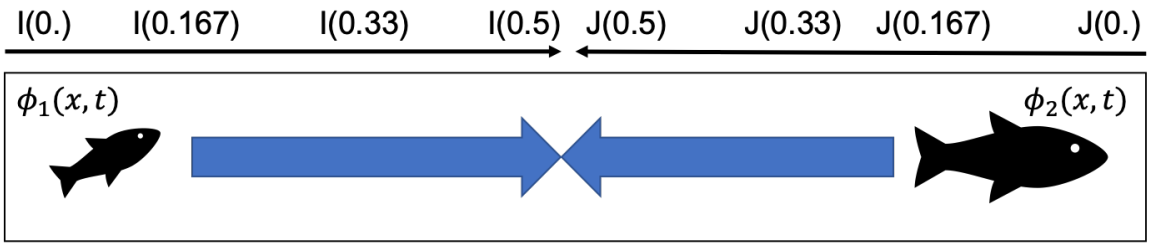


Figure 2.11: Illustration of the symmetric normalization path between images I and J.

The diffeomorphism φ is in domain Ω and used to transform image I according to

$$\varphi I = I \circ \varphi(\mathbf{x}, t = 1), \tag{2.25}$$

where t is time, $\mathbf{x}$ is a spatial coordinate, $v(\mathbf{x}, t)$ on Ω is a velocity field that is square-integrable and continuous. The deformation fields are found by performing an integration with respect to t:

$$\phi(\mathbf{x}, 1) = \phi(\mathbf{x}, 0) + \int_0^1 v(\phi(\mathbf{x}, t), t)dt, \tag{2.26}$$

and the distance is found by:

$$D(\phi(\mathbf{x},0),\phi(\mathbf{x},1)) = \int_0^1 ||v(\mathbf{x},t)||_L dt, \tag{2.27}$$

where L is the linear operator required to regularize the velocity. It is known that diffeomorphisms can be decomposed into two portions φ_1 and φ_2. This fact allows the transformation path to be divided in half and have the image I have equally as much contribution as the image J as previously seen in Figure 2.11.

The registrations is performed such that $t \in [0,1]$ and is indexed in both φ_1 and φ_2 but in opposite directions and to find the similarity metric $\varphi_1(\mathbf{x},1)I = J$. Using the property that diffeomorphisms are symmetric, the result of derivation of the above equation as two fields is as follows:

$$\phi_1(\mathbf{x},t)I = \phi_2(\mathbf{z},1-t)J, \tag{2.28}$$

where $\mathbf{x}$ and $\mathbf{z}$ are the spatial coordinates of the image I and J, respectively. For example if two anatomical landmarks $\mathbf{x}$ are known, then the calculation from two different directions along t takes half the time and the error becomes the minimization of the distance predicted by the two warp field, also known as *symmetric normalization* (SyN). The symmetry distance formula can be calculated for $t = 0.5$ for the two respective fields, therefore the integration in 2.27 only has to be done from 0 to 0.5 for the two fields due to the following constraint:

$$D\left(\varphi_1(\mathbf{x},0.5)\right) = D\left(\varphi_2(\mathbf{x},0.5)\right). \tag{2.29}$$

Avants [28] also proposes to use Cross-Correlation (CC) as opposed to Mutual Information (MI) as a similarity metric. Their argument states that MI works quite well for rigid problems, however, CC is more suitable for multi-modal registration because it can adapt well to local intensity variance. The images are first transformed according to the respective fields:

$$\begin{aligned} I_1 &= I\left(\varphi_1(\mathbf{x},0.5)\right), \\ J_2 &= J\left(\varphi_2(\mathbf{x},0.5)\right). \end{aligned} \tag{2.30}$$

The local mean μ is then found over a local window n^D where typically $n = 5$ and D is the image dimension, centered at position x. The mean is subtracted from the warped images I_1 and J_2 to obtain $\overline{I}(\mathbf{x})$ and $\overline{J}(\mathbf{x})$. With this information, the CC can be calculated which is dependent on φ_1 and φ_2 as follows:

$$CC(\overline{I}, \overline{J}, \mathbf{x}) = \frac{< \overline{I}, \overline{J} >^2}{< \overline{I} >< \overline{J} >} = A^2/BC. \tag{2.31}$$

The optimization can then be calculated via Error measurement of the gradients for both deformation fields[3]. The values A, B, and C are image representations of the products of mean differences seen in the CC equation (Equation 2.31). These three values can be computed prior to any deformation changes and be stored for any subsequent derivative calculations that may be necessary. Because the diffeomorphisms and velocity fields at time 0.5 are known, as the CC is being calculated, the velocity fields and diffeomorphism become updated by calculating the gradient between the current estimation of φ_i and the inverse φ_i^{-1} one. This way the gradients can be calculated and convergence can be achieved.

The symmetric normalization model steps can be summarized by the following steps:

1. Initialize $\varphi_1 = \mathbf{Id} = \varphi_1^{-1}$ and $\varphi_2 = \mathbf{Id} = \varphi_2^{-1}$.
2. Repeat the following steps until convergence.
3. Deform I by $\varphi_1(0.5)$ and J by $\varphi_2(0.5)$, calculate $\overline{I}$ and $\overline{J}$, calculate CC and store images representing A, B, C.
4. Calculate each $\mathbf{v}_i$ by smoothing results in step 3.
5. Change φ_i by $\mathbf{v}_i$ through ordinary differential equations.
6. Generate the time 1 solutions from $\phi_1(1) = \phi_2^{-1}\left(\phi_1(\mathbf{x}, 0.5), 0.5\right)$ and $\phi_1^{-1}(1) = \phi_2\left(\phi_1^{-1}(\mathbf{x}, 0.5), 0.5\right)$.

The results of the ANTs algorithm has shown significant improvement and faster implementation of image registration algorithms, particularly being useful in being able to simultaneously predict forward and backward images mapping [28].

[3]For a detailed derivation of all error and derivative formulas, the author directs the reader to the Avants paper [28]

2.5.3 Elastix

The Elastix toolkit is an open-source software that has been developed as a means to test several registration strategies [27]. It works via command-line interface and is built on the popular Insight Toolkit [40]. The registration framework consists of several components that can be adjusted accordingly. As with any registration problem, the main idea is to minimize a cost-function $\mathcal{C}$ that denotes the difference between a fixed image and warped moving image:

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \mathcal{C}\left(\boldsymbol{T}_{\boldsymbol{\mu}}; I_F, I_M\right). \tag{2.32}$$

In the equation above, $\boldsymbol{\mu}$ shows that the transformation has been parametrized, which can be applied with the transformation $\boldsymbol{T}_{\boldsymbol{\mu}}$. The fixed image is represented as I_F and the moving image is I_M. Several strategies are employed when choosing the components of the registration framework that can affect the results of registration. The cost function $\mathcal{C}$ is used to determine the similarity between the warped moving image and the fixed image. The following cost functions are available in Elastix:

1. Mean Squared Difference (MSD):

$$\mathrm{MSD}\left(\boldsymbol{T}_{\boldsymbol{\mu}}; I_F, I_M\right) = \frac{1}{N} \sum_{\boldsymbol{x} \in \Omega_F} \left(I_F(\boldsymbol{x}) - I_M\left(\boldsymbol{T}_{\boldsymbol{\mu}}(\boldsymbol{x})\right)\right)^2, \tag{2.33}$$

   where Ω_F is the fixed image space from which the pixels $\boldsymbol{x}$ are sampled.

2. Normalized Correlation:

$$\gamma = \frac{\sum_{I_F, I_M} (I_F - \overline{I_F})(I_M - \overline{I_M}) p(I_F, I_M)}{\left\{\left(\sum_{I_F, I_M} p(I_F, I_M)(I_M - \overline{I_M})^2\right)\left(\sum_{I_F, I_F} p(I_F, I_M)(I_F - \overline{I_F})^2\right)\right\}^{1/2}}, \tag{2.34}$$

   where $p(I_F, I_M)$ is the probability of the fixed image have a pixel correspond to a pixel in the moving image, $\overline{I_F}$ is the local mean of the fixed image, and $\overline{I_M}$ is the local mean of the moving image [41].

3. Mutual Information (MI) as seen in Equation 3.23.

4. Normalized Mutual Information (NMI) as seen in equation 3.24.

5. α-MI defined by:

$$\alpha I(X;Y) = \frac{1}{\alpha - 1} log \left(\sum_{y \in Y} \sum_{x \in X} p^{\alpha}_{(X,Y)}(x,y) \ \left(p_X(x) p_Y(y)\right)^{1-\alpha} \right), \tag{2.35}$$

   where $p_{(X,Y)}(x,y)$ is the joint probability distribution and $p_X(x)$ and $p_Y(y)$ are the marginal probabilities. [42], [43].

6. κ - statistic defined by:

$$\kappa = \frac{p_o - p_c}{1 - p_c}, \tag{2.36}$$

   which measures similarity between raters [44].

For monomodal images, the MSD similarity metric typically works well, while multimodal registration problems are best described by MI-based similarity metrics. The κ-statistic can be used for registering binary masks. For non-rigid transforms, a regularizer term is used as penalty term for any unwanted deformations.

An example of an affine transformation using the coordinate system $\boldsymbol{T_\mu}$ using the following equation:

$$\boldsymbol{T_\mu}(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{t}, \tag{2.37}$$

where A is the transform matrix and t is the translation vector. The components of μ is composed of six degrees of freedom for 2D and and nine degrees of freedom in 3D. B-spline representations are used from physics-based approaches, which benefit from user-defined control points. These B-spline function control points ensure that no overtly large deformations are learned. One other advantage of Elastix is that it allows for user specific transformations. For example, if a user has a pre-defined transformation matrix, then it is applied according to the following equation:

$$\boldsymbol{T_\mu}(\boldsymbol{x}) = \sum_i w_i \boldsymbol{T}_i(\boldsymbol{x}). \tag{2.38}$$

The equation is a weighted sum of the predefined user transforms where the weights w_i are the

components of the parameter vector μ. This approach is particularly useful because it offers the ability to perform several levels of transforms. As such, coarse rigid and non-rigid transforms can be applied prior to a finer registration field is learned.

The optimizer is used to solve the iterative problem of finding the most optimal transform parameters according to the following:

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k - a_k \boldsymbol{d}_k, \tag{2.39}$$

where $\boldsymbol{\mu}_k$ is the transform parameter at iteration k in the direction $\boldsymbol{d}_k$ at a step size a_k. One method of calculating direction is to find the derivative of the cost function at the current position. The optimization can be done via several methods: gradient descent, quasi-Newton, and stochastic gradient descent. The Robbins-Monro stochastic gradient descent method has been shown as the best performing due to its fast computation time per iteration and it uses a small portion of the fixed image voxels to compute the derivatives of the cost function [45], [46].

The cost function requires a sampling strategy such that not the entire fixed image is used. This approach would significantly reduce computation time. One proposed method is to use salient features such as edges. A sampler can also simply just pick random points in an image. Picking random points has shown to improve the smoothness of the cost function $\mathcal{C}$. A user-supplied mask can force the sampling algorithm to look at the areas specified by the mask. For every iteration of the gradient descent, if the algorithm samples from a new area, or if using localized MI, it picks random samples from within a neighbourhood as opposed from the entire fixed image.

All types of interpolations are supported including nearest-neighbour, bilinear, and N-th order B-spline interpolation. With the use of hierarchical approaches a pyramid of multiresolutions is implemented. The types of pyramids available are Gaussian with or without downsampling. The purpose of the pyramid strategy is that lower resolution images can be used for coarse registration via affine transforms. Gradually as the resolution increases, finer control is regained over the images. Therefore, elastix offers a complete toolbox with many user-adjustable parameters and techniques that can create custom and user-defined methods of image registration for a desired task [27].

2.5.4 Voxelmorph

Advances in computational power has led to development of new algorithms using deep learning and convolutional neural networks (CNNs). One such paper and algorithm has been suggested to significantly reduce the computation time for image registration, namely Voxelmorph [47]. The overarching idea is the authors try to match the learning process to that of a human learning. An analogy serves that if someone sees enough pairs of shoes in a house enough times they are then able to learn to match any unmatched shoes together. The authors suggest that a similar approach can be performed using CNNs. Through a learning process, a model can be created that allows for fast generation of deformation fields to warp a moving image to match a fixed one. Similar to ANTs, Voxelmorph works via diffeormophisms that have a zero-velocity field representation. If ϕ is the deformation field that maps coordinates from the fixed image to the moving, then using ordinary differential equations (ODE) one can define the deformation field as:

$$\frac{\partial \phi^{(t)}}{\partial t} = \boldsymbol{v}\left(\boldsymbol{\phi}^{(t)}\right), \tag{2.40}$$

where $\phi^{(0)} = Id$ is the identity transform and t is time, v is the stationary velocity field integrated over $t = [0, 1]$. One portion of the Voxelmorph algorithm is the scaling and squaring. Integration of stationary ODE is a one-parameter subgroup of diffeomorphisms. Using group theory that can be used to represent vector spaces, particularly useful in deformation fields. Group theory is particularly useful in incorporating many geometric transformations such as rotation matrices and reflections. One particular group, known as a Lie group has a differentiable property and can be used to calculate object symmetry [48]. To better understand the generation of the diffeomorphisms in Voxelmorph, some of the derivation of the obtained fields are described. Taking the exponent of $\boldsymbol{v}$ one can obtain $\phi^{(1)} = \exp(\boldsymbol{v})$. As such, applying scalars to the exponentiated equation such as t and t' can be used to expand into a composition map via pointwise operation $\exp((t + t')\boldsymbol{v}) = \exp(t\boldsymbol{v}) \circ \exp(t'\boldsymbol{v})$. Using the previously mentioned method of decomposing the exponential function, one can break down $\phi^{(1)} = \phi^{(1/2)} \circ \phi^{(1/2)}$. If $\phi^{(1/2^T)} = \boldsymbol{p} + \boldsymbol{v}(\boldsymbol{p})$ where $\boldsymbol{p}$ is a map of spatial locations the parameter T can be chosen such that $\boldsymbol{v} \approx 0$.

Generating a transform map ϕ_z from image x and y is done via a generative model based on a variational inference method using CNNs. The unsupervised model learns the network parameters by using a fixed and moving image as the input and generating a deformation field that is used to warp the moving image. The difference between the fixed and warped moving image is used as a similarity metric. The details of the generative model are described via the following equation:

$$p(z) = \mathcal{N}(z; 0, \Sigma_z), \tag{2.41}$$

where $\mathcal{N}$ is a multivariate normal distribution with mean μ and covariance Σ. The velocity field is defined by the ODE in Equation 2.40. Using a Laplacian L of a neighbourhood graph defined on a voxel grid, it can be described by $L = D - A$, where D is the graph degree matrix (a diagonal matrix describing the degree of each vertex) and A is the voxel neighbourhood adjacency matrix (a square matrix used to represent a finite graph). Spatial smoothness is achieved via $\Sigma_z^{-1} = \Lambda_z = \lambda \mathrm{L}$ where Λ_z is the precision matrix (inverse of the covariance matrix). Noise is added to the x image based on the warped image y, with σ^2 variance noise using the following equation:

$$p(x|z; y) = N\left(x; y \circ \phi_z, \sigma^2\right). \tag{2.42}$$

The deformation field is approximated by finding the posterior probability $p(z|x; y)$ and maximizing a posteriori estimation for a new image pair (x, y). Although, computing this posterior probability is unmanageable, a variation is introduced via $q_\psi(z|x; y)$ with parameter ψ. The learning is performed by minimizing the KL divergence according to the following:

$$\begin{aligned} \min_\psi \mathrm{KL}\left[q_\psi(z|x; y) \| p(z|x; y)\right] &= \min_\psi E_q\left[\log q_\psi(z|x; y) - \log p(z|x; y)\right] \\ &= \min_\psi \mathrm{KL}\left[q_\psi(z|x; y) \| p(z)\right] - E_q[\log p(x|z; y)]. \end{aligned} \tag{2.43}$$

The approximate posterior is then modelled using the same noise equation as 2.42 as follows:

$$q_\psi(z|x; y) = \mathcal{N}\left(z; \mu_{z|x,y}, \Sigma_{z|x,y}\right), \tag{2.44}$$

where $\Sigma_{z|x,y}$ is a diagonal matrix. Both $\mu_{z|x,y}$ and $\Sigma_{z|x,y}$ are estimated via the CNN computation

according to parameter ψ. Stochastic gradient descent used to minimize the KL-div equation in 2.43 and obtain the loss given by:

$$\mathcal{L}(\psi; \boldsymbol{x}, \boldsymbol{y}) = \mathrm{KL}\left[q_\psi(z|x;y)\|p(z)\right] - E_q[\log p(x|z;y)]. \tag{2.45}$$

The deformation field is learned via a 3D U-net style architecture [5] with the first arm being an encoding side to obtain low-level features and the last portion is the decoder that upsamples and concatenates previous layers for learning higher-level features and obtaining a deformation field that is the same size as the moving and fixed images. The network is composed of a convolutional layer with 16 filters, four downsampling layers with 32 convolutional filters and a stride of 2, and three upsampling layers with 32 filters. LeakyReLu are the activation functions for all convolutional layers with 3x3x3 kernels. The learned field deforms the moving image such that it matches the fixed image and then the squared difference is found. Voxelmorph is implemented using Keras [49] with Tensorflow backend [50] using the ADAM optimizer [51].

Chapter 3

FlowReg: Optical flow-based Image Registration

This chapter will describe the main contributions of the thesis. It begins with an explanation of the basics of optical flow and how traditional algorithms operate. Modern deep-learning models, namely *FlowNet*, are explored. The next portion of the chapter describes the *FlowReg* model architecture along with the improved loss function. Key differences between the original *FlowNet* and *FlowReg* models are explored. Finally, the several registration validation metrics are shown that measure registration quality by analyzing tissue integrity and pixel alignment.

3.1 Optical Flow

Optical flow is a vector that shows the movement or displacement of a specific pixel between two images. A video is composed of a number of frames. As each frame is shown on the screen at a certain frame rate per second, $\frac{F}{s}$, motion is perceived by the human visual system. Frames are just images that are captured by a video camera in sequence and thus can be analyzed from an image processing perspective using pre-existing tools. Hence, optical flow is often used in video processing to calculate velocity and acceleration of objects in a scene [6]. The next subsections will describe the basics of optical flow and the methods by which optical flow is calculated. Some aspects of optical flow that are used in real-world applications include motion-based segmentation, obtaining three-dimensional representation from motion, and global alignment such as compensation for camera motion. The global alignment property was the reason why optical flow was selected as a possible method of image alignment in medical image registration.

3.1.1 Basics of Optical Flow

To calculate the optical flow between two frames, let t be the time variable, then $I(\vec{x}, y)$ is the intensity of the image I warped by vector $\vec{x} = (x, y)T$ [52].

$$I(\vec{x}, t) = I(\vec{x} + \vec{u}, t + 1), \tag{3.1}$$

where $\vec{u}$ is the 2D velocity vector $\vec{u} = (u_1, u_2)$. One assumption in the calculation of optical flow is *brightness constancy*. Brightness constancy is a concept from perceptual psychology that says an object appears the same to the human eye regardless of the lighting conditions. In optical flow, brightness constancy does not hold because the difference in brightness between two pixels is very small, thus the intensity of pixels is conserved. As such, the derivative of the same pixel with respect to time is equal to zero. Estimation of the optical flow can be done by Taylor series expansion of the following equation:

$$f_1(x - d) = f_1(x) - df_1'(x) + O\left(d^2 f_1''\right), \tag{3.2}$$

where $f' = df(x)/dx$ and d is the displacement [52]. Generalizing to a two-dimensional image, the equation becomes:

$$I(\vec{x} + \vec{u}, t + 1) \approx I(\vec{x}, t) + \vec{u} \cdot \nabla I(\vec{x}, t) + I_t(\vec{x}, t), \tag{3.3}$$

where $\nabla I = (I_x, I_y)$, I_t is the spatial and temporal partial derivatives of the image I, and $\vec{u} = (u_1, u_2)^T$ is the two-dimensional velocity vector. Substituting using Equation 3.1 the derivation gives the *gradient constraint equation*:

$$\nabla I(\vec{x}, t) \cdot \vec{u} + I_t(\vec{x}, t) = 0. \tag{3.4}$$

One can now use two frames to approximate the temporal and spatial gradient by simple subtraction $\delta I(\vec{x}, t) = I(\vec{x}, t + 1) - I(\vec{x}, t)$. Equation 3.4 cannot be easily solved because $\vec{u}$ is a vector with two components u_1 and u_2. One method to overcome the problem of one equation and two unknowns is to place more constraints. This can be done by observing the movement of nearby pixels within a certain region [52], [53]. As it is unfeasible to have a single velocity vector to

describe the movement of multiple pixels within a region, Least Squares Estimation (LSE) can be used:

$$E(\vec{u}) = \sum_{\vec{x}} g(\vec{x}) \left[\vec{u} \cdot \nabla I(\vec{x}, t) + I_t(\vec{x}, t)\right]^2, \tag{3.5}$$

where $g(\vec{x})$ is a weighting function that describes the amount of influence from the nearby pixels, which is commonly Gaussian and thus the center pixels have a higher influence than others. The lowest value for $E(\vec{u})$ can be found where its derivative with respect to $\vec{u}$ is zero. Equation 3.5 can be converted to matrices for ease of computation:

$$\mathbf{M}\vec{u} = \vec{b}, \tag{3.6}$$

where $\mathbf{M} = \begin{bmatrix} \sum gI_x^2 & \sum gI_xI_y \\ \sum gI_xI_y & \sum gI_y^2 \end{bmatrix}$ and $\vec{b} = -\begin{pmatrix} \sum gI_xI_t \\ \sum gI_yI_t \end{pmatrix}$. As such, the optical flow estimate is calculated by rearranging to obtain:

$$\hat{u} = \mathbf{M}^{-1}\vec{b}. \tag{3.7}$$

Therefore, $\mathbf{M}$ should be of order two or higher and thus invertible so that it can be used for optical flow estimation $\hat{u}$. In the coming section, the established Lucas-Kinade iterative algorithm will be discussed along with its implementation [53].

3.1.2 Classic Algorithms

Assuming the displacement of the pixels d is relatively small, iteratively the optical flow can be calculated through Equation 3.5. The calculated displacement is then used to undo or warp the pixel motion in the second image until the two images are relatively similar:

$$I^0(\vec{x}, t + \delta t) = I\left(\vec{x} + \vec{u}^0\delta t, t + \delta t\right), \tag{3.8}$$

where I^0 is the warped image according to the estimated displacement field $\vec{u}^0$. Since the warp vector is made up by the current vector plus the residual $\vec{u} = \vec{u}^0 + \delta\vec{u}$, the residual can then be calculated as the temporal derivative. As such, the difference between two frames, let's say I^0

and I^1, is the displacement vector described by their corresponding displacement vectors plus the residual as shown in:

$$\vec{u}^1 = \vec{u}^0 + \delta\hat{u}. \tag{3.9}$$

Every iteration j through the LSE updates the residual $\delta\hat{u}$ until it is zero and the following holds true $\vec{u}^1 = \vec{u}^0$ as seen in the following:

$$E(\delta\vec{u}) = \sum_{\vec{x}} g(\vec{x}) \left[I^j(\vec{x}, t) - I^j(\vec{x} + \delta\vec{u}, t + 1) \right]^2, \tag{3.10}$$

which is the iterative version of LSE described in Equation 3.5.

Temporal aliasing may occur due to images having a lower sampling frequency than the established Nyquist frequency. This is often seen in images as the banding effect or the squaring lego-effect of round edges when the resolution of the image is higher than the display resolution, thus being sampled at a much lower frequency [54].

To address the aliasing problems, a coarse to fine estimation has been proposed [55], [56]. If an image is sufficiently blurred, then the motion vector size is larger because the neighbourhood contents of the central pixel become similar to one another. Gaussian blurring pyramids are then used to blur the images and calculate the displacements from a coarse of the most blurred image to a fine estimation until the original image resolution is obtained. This is similar to the iterative method, except that the calculated displacement field has to be upsampled from a blurred level of the pyramid to a finer resolution level. One disadvantage of the coarse to fine estimation is that if a vector is calculated inaccurately at a low resolution level, the error will propagate through all of the layers.

One assumption of the neighbourhood pixel approach is that the estimated optical flow in the local neighbourhood is smooth by means of non-parametric estimation. Horn and Schunck [6] proposed global smoothing according to the following energy equation:

$$E(\vec{u}) = \int \left(\nabla I \cdot \vec{u} + I_t\right)^2 + \lambda \left(\|\nabla u_1\|^2 + \|\nabla u_2\|^2 \right) dxdy. \tag{3.11}$$

The proposed equation is especially useful in calculating large distances due to the global smoothing such as camera movement. It examines an even surface like a wall because global approaches can interpolate movement in between known displacement values[1]. The main disadvantage of Equation 3.11 lies in the expensive computation. The following section will examine some of the modern approaches via Convolutional Neural Networks.

3.1.3 Deep Learning

In recent years, due to advancements in hardware development, CNNs have garnered attention and have become widely used in attempts to provide solutions to various problems throughout many fields in arts, science, and engineering [4]. CNNs have become particularly useful in image processing fields and one of the regions that has seen some major advantages and a disruption of established methods is in the field of optical flow calculation [7]. Convolutions in the discrete two dimensional domain involve using a kernel across an image to generate feature maps. Typically several kernels are used to generate numerous features across multiple layers. Graphics processing units (GPUs) in modern computers are particularly useful in performing the fast simple operations of multiplication and addition due to their fast memory and parallel computing power.

One major drawback of CNN-based approaches is that there is a need for ground truth labels. A CNN model learns the response to these labels and then generalizes when new data is introduced to the model without labels. Optical flow has very limited ground truth data and most of the current methods have been implemented via the Lucas-Kanade method previously mentioned [53]. One way to overcome the lack of data is to use subsequent frames in a video to calculate the optical flow of the pixels from one frame to the next [8]. The approach is similar to reducing the residual displacement vector value via an iterative process, except the features are learned via convolutional layers.

One suggested approach, named FlowNet, has shown significant progress in predicting optical flow from subsequent frames. The network structure (seen in Figure 3.1) is composed of an encoder and

[1] A more detailed explanation of the iterative algorithms, Euler-Langrange solutions, and implementation of the Horn-Schunck equation can be found in Fleet and Weiss paper [52]

a decoder arm similar to the U-net style. The encoder arm is composed of several convolutional layers that reduce the resolution of the image, but are able to encode smaller features with each level of convolution. This is a coarse to fine approach, where there is a need for many more features maps to be able to encode small features, as opposed to large features which have a more general structure. The decoder arm is able to upscale the obtained features so the output is the same size as the input. Through skip connections, the model, termed *FlowNet-Simple* can then concatenate the large-scale features with the small upscaled features. This technique provides a more accurate representation of the learned feature maps.

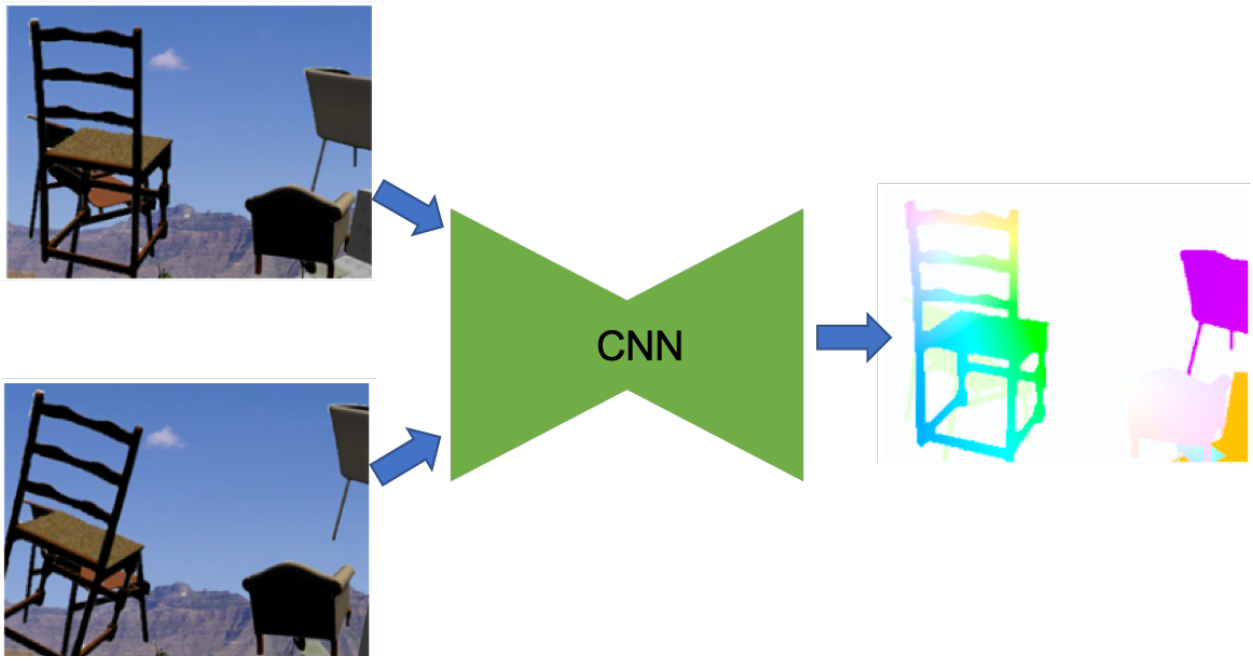


Figure 3.1: FlowNet convolutional model to learn optical flow between two moving frames [7].

Given enough labeled data, a network can eventually learn the optical flow between two frames. Dosovitskiy et al. [7], suggest a parallel feature map learning path for the two images. This approach produces some meaningful representations of the two images prior to a combination at a higher level. To further aid the network to learn, a correlational layer has been proposed. Patches from each image are compared via the following equation:

$$c\left(\mathbf{x}_1, \mathbf{x}_2\right) = \sum_{\mathbf{o} \in [-k,k] \times [-k,k]} \left\langle \mathbf{f}_1\left(\mathbf{x}_1 + \mathbf{o}\right), \mathbf{f}_2\left(\mathbf{x}_2 + \mathbf{o}\right) \right\rangle, \tag{3.12}$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are patches centered in feature map $\mathbf{f}_1$ and feature map $\mathbf{f}_2$, respectively, and $K := 2k + 1$ is the size of the square patch. This implementation, termed *FlowNet-Corr* is a similar technique as performing convolution with a kernel, but instead of the random kernel, the convolution is being performed with a patch of another image. Since the values of the correlation are significantly larger than convolutions with kernels, the displacement of $\mathbf{x}_2$ is constrained in a neighborhood $D := 2d + 1$. The rest of the network is similar to *FlowNet-Simple*. The experiments with optical flow in this thesis have been performed the *FlowNet-Simple* architecture.

3.1.4 Model Architecture - FlowReg

The major drawback of the *FlowNet* implementation is that there is a requirement for ground truth labels. The authors have created a synthetic database, *Flying Chairs*, for which they artificially move the chairs around via pre-defined flow vectors [7]. Unsupervised methods have been proposed to calculate optical flow between image pairs by warping subsequent video frames and generating the pixel-flow between the two time points [8]. The implementation by Yu et. al [8] was adapted to medical images. The original loss function (Eqn: 3.13) is composed of photometric and smoothness loss components as seen below. Further details are described in the *FlowReg* desription.

$$\begin{aligned}\mathcal{L}(\mathbf{u}, \mathbf{v}; F(x, y), M_w(x, y)) =& \gamma \ell_{photometric}(\mathbf{u}, \mathbf{v}; F(x, y), M_w(x, y)) + \\ & \lambda \ell_{smoothness}(\mathbf{u}, \mathbf{v}).\end{aligned} \tag{3.13}$$

The proposed network structure is identical to *FlowNet-Simple* [7] composed of an encoder and decoder arm with skip-connections (Figure 3.2), similar to the *U-net* structure [5]. The encoder arm is composed of several convolutional layers that reduce the resolution of the image, but are able to encode smaller features with each level of convolution. This is a *coarse-to-fine* approach, where there is a need for many more features maps to be able to to encode smaller features, as opposed to the large features which have a more general structure. The decoder arm is able to upscale the obtained features through strided convolutions such that the output is the same size as the input.

The key difference between the proposed implementation and *FlowNet-Simple* and the proposed registration method lies in the loss function. As *FlowNet-Simple* is trained on a synthetic database, the ground-truth is known and thus is able to learn the optical flow from temporally spaced video frames and dissimilarities are calculated as a simple endpoint-error (EPE). To employ a fully unsupervised optical flow generation method, the strategy of Yu et al [8] was employed. At various resolutions of the decoder arm, the generated flow, is used to warp the moving image and compared to the fixed image. The recently proposed Spatial Transformer Network (STN) [57] is used to warp the moving image with the predicted flow. The differentiable nature of STNs are particularly useful to be able to create the sampling grid and warp the moving image according to the learned deformation field.

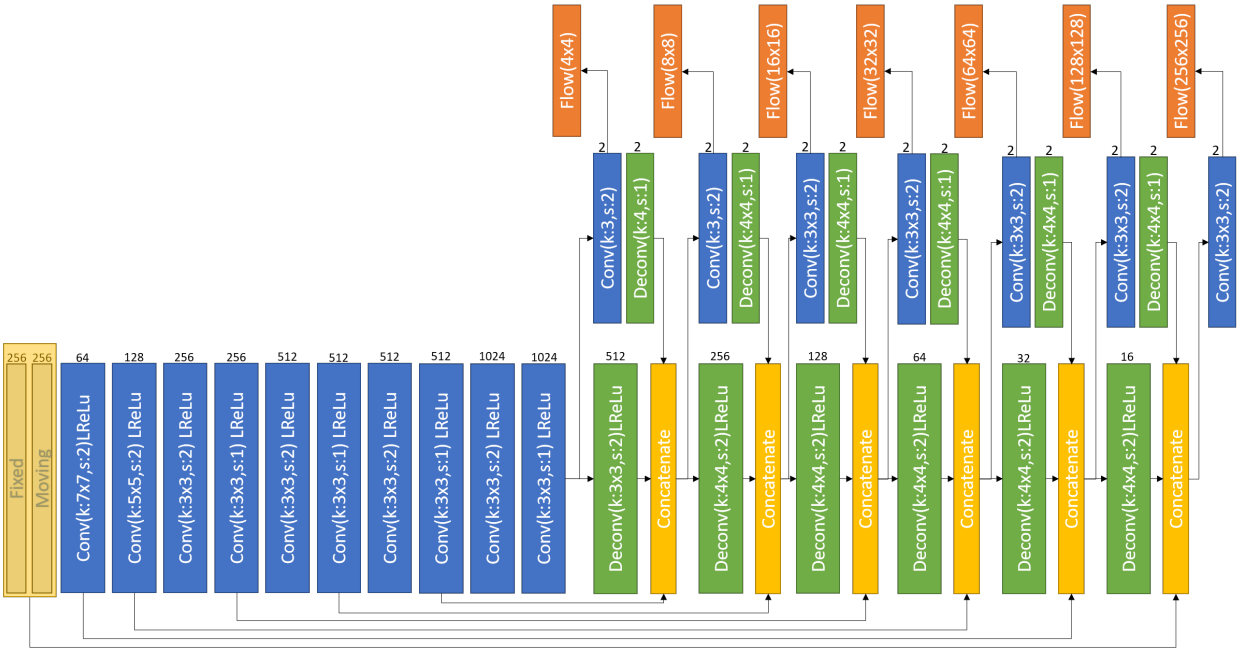


Figure 3.2: Model architecture used to predict optical flow from a moving and a fixed image. The input to the network is two images of size $H = 256$, $W = 256$. The number of feature maps is indicated above each convolution operation. The kernel size and stride are indicated in the block of each convolution operation along with the activation function (Leaky ReLu). Upsampling is done via strided Deconvolution operations. The flow at seven resolution levels is indicated in the output of the network.

Similar to Yu et al.'s [8] work, the loss function is a combination of comparisons between the warped moving image and the fixed image and a flow loss function component. The total loss function is a weighted summation of three components, photometric loss, correlation loss, and

smoothness loss as described below:

$$\begin{aligned}\mathcal{L}(\mathbf{u},\mathbf{v};F(x,y),M_w(x,y)) =& \gamma\ell_{photometric}(\mathbf{u},\mathbf{v};F(x,y),M_w(x,y))+\\ & \zeta\ell_{correlation}(\mathbf{u},\mathbf{v};F(x,y),M_w(x,y))+\\ & \lambda\ell_{smoothness}(\mathbf{u},\mathbf{v}),\end{aligned} \tag{3.14}$$

where $F(x,y)$ is the fixed image, $M_w(x,y) = M(x+u, y+v)$ is the warped moving image, $\mathbf{u},\mathbf{v}$ are the horizontal and vertical vector fields, and γ, ζ, and λ are regularizing parameters that can be tuned to control the amount of influence or weight of the various loss components.

The photometric loss is a simple difference measure between the fixed and the warped moving image. This loss component evaluates to what degree the predicted optical flow is able to warp the moving to match the fixed image on a pixel-by-pixel basis:

$$\ell_{photometric}(\mathbf{u},\mathbf{v};F(x,y),M_w(x,y)) = \frac{1}{N}\sum_{i,j}\rho_D(F(i,j) - M_w(i,j))). \tag{3.15}$$

The Charbonnier penalty function ρ_D is used to reduce any overtly large displacements that may occur due to outliers of pixels and is defined by:

$$\rho(x) = (x^2 + \epsilon^2)^\alpha, \tag{3.16}$$

where ϵ is a small number and a α is a tunable parameter. To encourage an overall global alignment of the mean intensities between the fixed and warped moving images, the correlation loss was implemented as an additional loss component:

$$\ell_{correlation}(\mathbf{u},\mathbf{v};F(x,y),M_w(x,y)) = \frac{cov(F(x,y),M_w(x,y))}{\sigma(F(x,y))\times\sigma(M_w(x,y))}, \tag{3.17}$$

where the covariance is calculated by:

$$cov(X,Y) = \frac{1}{N-1}\sum(X-\bar{x})\times(Y-\bar{y}), \tag{3.18}$$

and σ is the standard deviation of the respective image. An ablation study was performed to observe

the effect of the correlation loss $\ell_{correlation}$ by removing the component from the total loss function. The outcome is described in the Results Chapter (Ch. 4. The smoothness loss is implemented to reduce discontinuity of optical flow at object boundaries. The loss component encourages small differences between neighbouring flow vectors in the height and width directions:

$$\ell_{smoothness}(\mathbf{u},\mathbf{v}) = \sum_{j}^{H}\sum_{i}^{W}[\rho_S(u_{i,j} - u_{i+1,j}) + \rho_S(u_{i,j} - u_{i,j+1}) + \rho_S(v_{i,j} - v_{i+1,j})] + \rho_S(v_{i,j} - v_{i,j+1})]. \tag{3.19}$$

A summary of the loss components is seen in Figure 3.3.

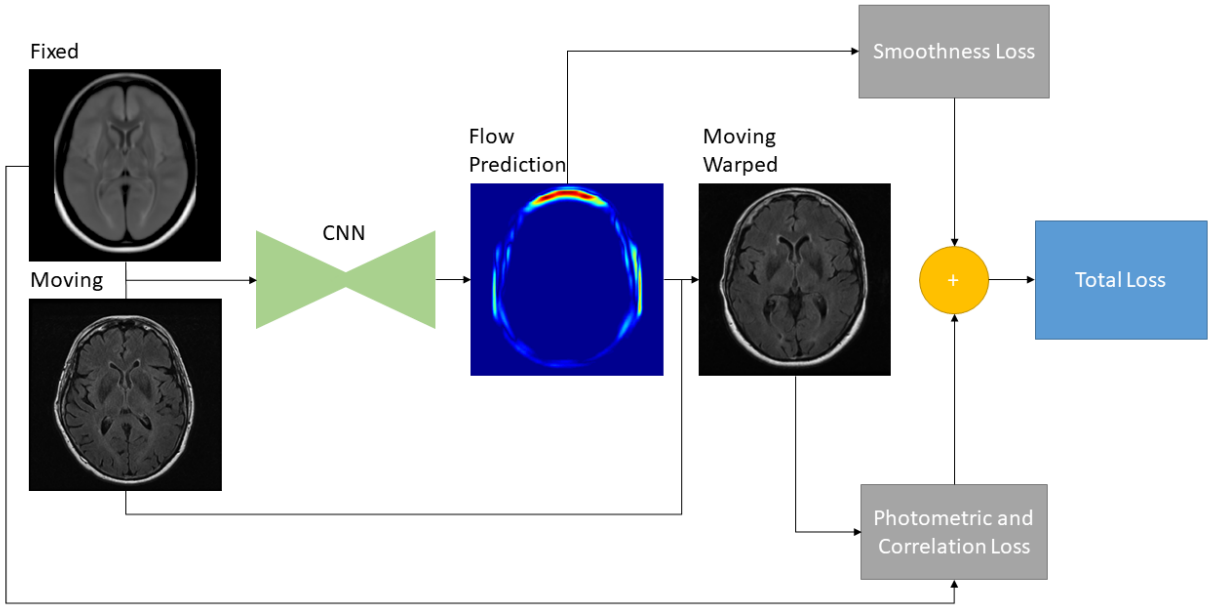


Figure 3.3: Overview of the loss components calculations.

3.2 Registration Validation Metrics

As medical images are typically organized in volumetric stacks of two-dimensional images, there is a need for volumetric and structural measures of anatomic integrity. The suggested metrics are focused on quantifying registration quality based on volumetry, structural deformation, and orientation. Volumetry is measured by calculating a structure's volume proportional to the whole

brain, named proportional volume (PV). Structural measures are defined as a surface-to-surface distance (SSD) between a brain structure and the surface of the brain. Head Angle (HA) is an orientation measure calculated by finding the angle of rotation with respect to the saggital plane. For the sake of completion, the typical registration quality measures are also evaluated such as MI and CC.

An anatomical atlas provides a representation of key structures with corresponding coordinates. Current neurological FLAIR-specific atlases are of low resolution and do not account for brain WMLs. One of the secondary goals of this work is to propose a FLAIR-specific atlas as a fourth measure of registration by averaging registered volumes.

3.2.1 Validation of Image Registration

In image registration, the calculated transformations can typically be extracted to be used to warp the moving image to the fixed space or to deform the warped image back to the original space. These transformations can typically be applied to binary masks. A binary mask is defined as pixels that are included in an area of interest. The included pixels are typically labeled with a value of 1 and the remainder are 0. These areas of interest can be obtained via manual annotations or automatic segmentations.

The calculated transformations can also be used to warp corresponding binary masks of the moving image. This is particularly useful, because now different metrics can be calculated to measure how different areas in the image have changed after registration. One method of measuring similarity of two areas is to measure the overlap between the ground truth and a predicted value. There are several methods of measuring overlap. One method mentioned derived by Dice [58] the Dice Similarity Coefficient:

$$DSC = \frac{2TP}{2TP + FP + FN}, \tag{3.20}$$

where TP is the True Positive, FP is the False Positive, and FN is the False Negative area of the overlap as seen in Figure 3.4.

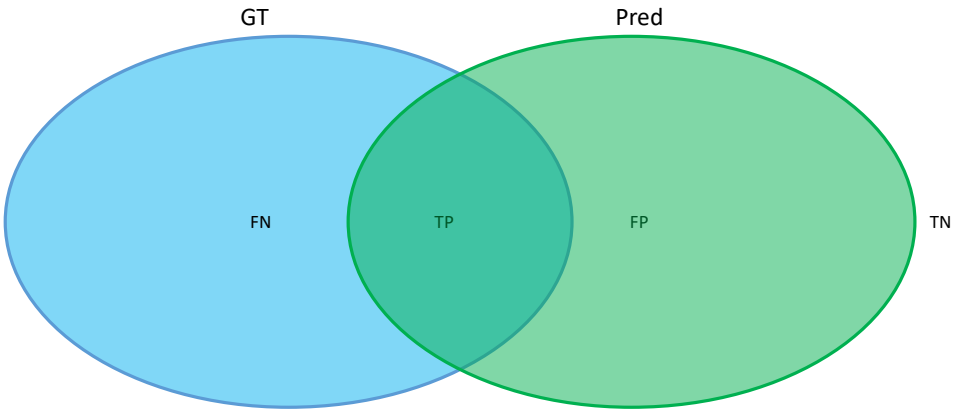


Figure 3.4: Overlap visualization. TP = True Positive, FP = False Positive, TN = True Negative, FN = False Negative.

Other methods of measuring overlap is by using the Jaccard Similarity or Overlap Coefficient. The Jaccard Similarity is described by the intersect divided by the union of two sample sets as seen in the equation below:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}. \tag{3.21}$$

The intersect $(A \cap B)$ describes only the area where two data sets overlap as seen in Figure 3.4 marked by TP. The union $(A \cup B)$ is the area in two sets that includes all of Set A and all of Set B, except the union. Figure 3.4 is all of GT and all Pred but not TP so as to not repeat TP twice. Another way to interpret is $FN + TP + FP$. The overlap similarly uses the intersect, divided by the smallest of the two sets:

$$OC(A,B) = \frac{|A \cap B|}{argmin(|A|,|B|)}. \tag{3.22}$$

Information theory can be used to also describe the similarity between two images. For every image a probability density function (PDF) can be calculated via histogram binning. For a number of bins, the proportional number of pixels from all pixels in an image X falling within a certain intensity range can be easily calculated $(p_X(x))$. The joint PDF between two images can be found by finding the probability that each pixel in one image X falls within a histogram bin range and also falls within the histogram bin range of another image Y denoted by $(p_{X,Y}(x,y))$.

Since images are considered as discrete random variables one can use the joint probability and the marginal probabilities to calculate the similarities between images by using Mutual Information. This metric is defined as the amount of information shared between two random variables. It can also be thought of as the similarity between the joint probability and the product of the marginal probabilities of two random variables. In the discrete domain, the variables can be images X and Y, and the double sums are calculated as seen below:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p_{(X,Y)}(x,y)\ log\left(\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\right). \tag{3.23}$$

For two images that are exactly the same, the mutual information should be a high number (typically between 2 and 3 for MR volumes, depending on the number of pixels and intensity ranges). As images are more dissimilar, the MI value drops. Since there is no upper bound, normalized variants have been proposed that are based on the marginal Entropy Measure of each random variable as seen below:

$$NMI(X;Y) = \frac{2 \times I(X;Y)}{H(X) + H(Y)}, \tag{3.24}$$

where $I(X;Y)$ is the mutual information between two random variables X and Y, the $H(X)$ is the marginal entropy of the random variable X, $H(Y)$ is the marginal entropy of the random variable X as seen in Figure 3.5.

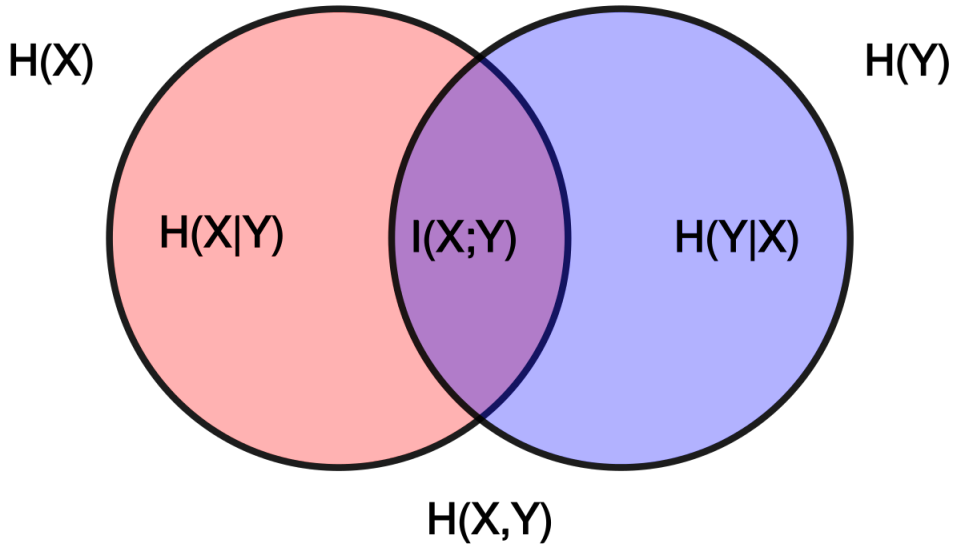


Figure 3.5: Entropy representation of Mutual Information. The violet area is the mutual information, the entire pink circle is the entropy of random variable X, denoted by H(X) and the entire blue circle is the entropy of Y, denoted by H(Y).

To find the entropy measures of each random variable, let's say random variable X, one can calculate the sum of the logarithmic probabilities that a pixel will fall within an intensity range.

$$H(X) = \sum \log(p_X(x)). \tag{3.25}$$

The same equation can be applied to random variable Y. In Figure 3.5, one can see that the conditional entropy $H(X|Y)$ can now be calculated by subtracting the $I(X;Y)$ from marginal entropy $H(X)$. The Normalized Mutual Information can now be easily calculated and will give values that are bound between 0 and 1 [59]. This is particularly useful because it provides a more relative scale for many variables to be compared with a particular one, not unlike registered images and an atlas template.

Some aspects of image registration that have to be considered consist of how different areas of an image are affected during registration. Volumetric changes are not accounted for when measuring overlap or when using any of the information theoretic metrics. Since deformation fields are not

uniform, some areas of brain tissue may be affected more than others. Ventricles are filled with CSF and as was discussed above (Figure 1.6), in FLAIR, these water-filled cavities are nulled and therefore are some of the hypointense or darkest areas of a brain. Hyperintense regions of the brain are the WMLs and they can also be affected by registration. This presents a need to observe how registration affects these particular regions after registration from a volumetric and structural perspective.

Measures of alignment are needed to assess the level of agreement between the transformed moving image and the fixed one. Correlation measures such as Pearson's correlation determine the pixel-wise relationship agreement between two images.

3.2.2 Percent Volume (PV)

In MRI, volumetric measurements are calculated voxel-wise. Each voxel V has a real-world dimensionality defined by the MRI scanner typically measured in mm^3. The V_x and V_y dimensions correspond to the pixel width and height, respectively, while the V_z dimension is found from the slice-thickness measure. With all the known parameters, the volume of a structure can be easily calculated by counting the number of total voxels n_v in a structure and the following multiplication:

$$vol = V_x \times V_y \times V_z \times n_v. \tag{3.26}$$

Pre-registration volumetric measures are calculated through the use of manually segmented binary masks and then summating across n_v, then translating to real-world measures. The volumes of the WMLs, ventricles, and whole brain are calculated via the same method as described above. The proportional volume (PV) is then obtained by dividing the volume of the structure (ventricles or WML) vol_S by the total brain volume vol_B as seen below:

$$PV = \frac{vol_s}{vol_b} \times 100. \tag{3.27}$$

Since all brain structures are contained within the brain, the proportional volume should remain constant no matter which three-dimensional space it is located in. This measure should then be

able to assert the amount that proportional volumes of hyperintense regions, such as WMLs, or hypointense CSF-containg regions, such as ventricles, change before and after registration. The finest performing registration algorithms should not drastically change this ratio.

3.2.3 Surface to Surface Distance (SSD)

Structural measures should be able to measure how the overall formation of brain tissue warps. This is a difficult measure, due to the non-uniformity of human anatomy. One method to examine this structural change is based on the surface to surface distance. Based on the Hausdorff distance [60], the idea is to find the closest point of a surface to another surface. The distance was found from the surface of the ventricles to the surface of the brain (Figure 3.6). The ventricles were selected as the structure of choice due to their consistent presence within a brain at approximately the same location. The formula for the SSD measure is as follows:

$$SSD = \frac{1}{N}\left[\sum_{i=1}^{N} argmin\left(\sqrt{S_{xyz}^2 - B_{xyz}^2}\right)\right]. \tag{3.28}$$

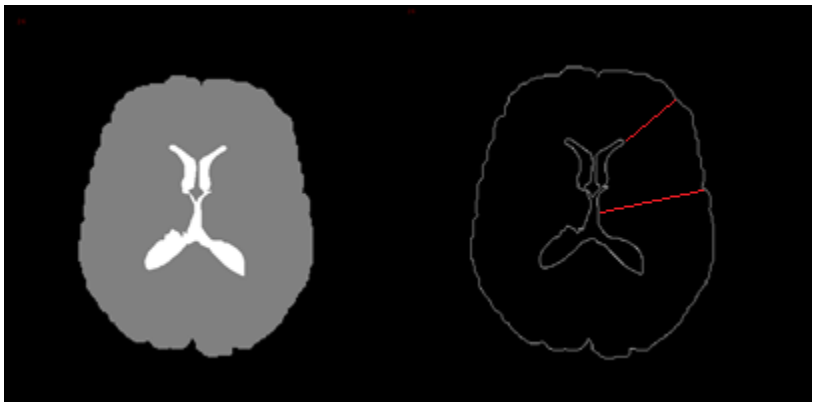


Figure 3.6: Two dimensional representation of SSD computation.

For every voxel in the ventricle structure S, the shortest distance to the surface of the brain B, is found and the average of all the minimum distances provides one numerical value for the SSD metric. In theory, the overall structural integrity of the ventricles should be maintained with respect to the brain structure post-registration. Any differences in this metric can amount to significant

structural changes in the shape and organization of the ventricles with respect to the surface of the brain.

3.2.4 Head Angle

The Head Angles (HA) measure is an orientation measure that examines to what extent the head is rotated with respect to the midsaggital plane. The process involves a combination of Principle Component Analysis(PCA) and an angle sweep as described in [61].

An MRI volume is first binarized using a combination of adaptive thresholding and opening and closing techniques. Otsu's method was chosen for adaptive thresholding due to rapid and accurate performance [62]. This adaptive method is able to divide the intensities in an image into two classes by minimizing the intra-class variance according to the following equation:

$$\sigma_w^2(t) = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t). \tag{3.29}$$

where σ^2 are the corresponding class variances and w are the corresponding weight classes. The algorithm is dependent on the PDF of the image and iterates through all intensity probabilities until the the intraclass variance is minimized [62].

Image closing and opening are two techniques used to more accurately define the outlines of images and remove noise. Specifically, image closing is defined by a dilation and then erosion [63]. Dilation can be applied using two-dimensional kernels, for example 3×3 and widening any areas of an image such that structures that were separated, can now be clustered together (Figure 3.7). Erosion essentially does the opposite where images that contain an undesired noise outside the kernel, are ignored, and thus the noise or structure is removed (Figure 3.8).

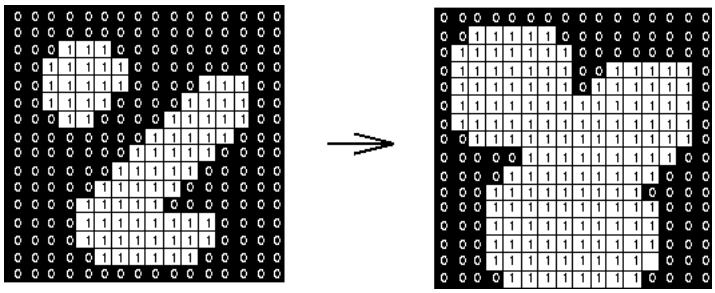


Figure 3.7: Binary image dilation example [64].

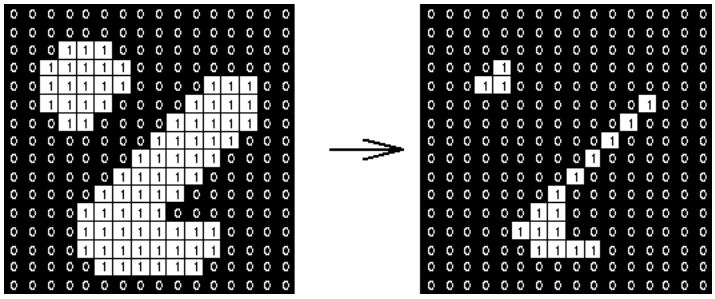


Figure 3.8: Binary image erosion example [64].

Image opening is essentially the reverse order of closing, first an erosion and then a dilation is performed [63]. After the binary head-masks are obtained, the coordrinates of each non-zero pixel are stored into two X and Y arrays. PCA analysis on these two arrays allows for efficient eigenvector calculation with the use of a covariance matrix which is calculated with the following equation:

$$P_x = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - m_x\right)\left(X_i - m_x\right)^T, \tag{3.30}$$

where m_x is the row-vector mean. The results is a symmetric matrix with a size 2×2. The values in the matrix are the variance of each column, and the joint variance denoted by f, h, and g:

$$P_x = \begin{bmatrix} f & g \\ g & h \end{bmatrix}. \tag{3.31}$$

Solving for the two eigenvectors can now be done by solving the following quadratic equation:

$$\lambda^2 - \lambda(f+h) + \left(fh - g^2\right) = 0, \tag{3.32}$$

for which the λ_1 and λ_2 values can be used to calculate the two eigenvectors according to:

$$e_j = \frac{1}{\sqrt{g^2 + (\lambda_j - f)}} \begin{bmatrix} g \\ \\ \lambda_j - f \end{bmatrix}, \tag{3.33}$$

where j is 1 or 2 specifying the two eigenvectors and two λ values. Once the two eigenvectors are known, the angle between them can be easily found through an inverse tangent operation:

$$\theta_1 = \tan^{-1}\left(\frac{v_y}{v_x}\right). \tag{3.34}$$

This θ_1 value represents the coarse angle and it is the first step in finding the HA rotation. PCA angle calculation was performed on every second slice from the middle transverse slice to the top of the head. These slices are some of the closest approximations of an ellipse and most useful in angle of rotation calculation. The lowest three angles are taken as candidates due to the fact that head irregularities and errors in segmentation may produce large numbers that cannot be corrected in the second step.

The fine angle calculation consists of an angle sweep and determines the final θ_2 angle calculation of the three candidate slices. Each of the three slices are inversely rotated using their corresponding θ_1 values. The images are then mirrored and rotated at 0.5° angle increments according to the following angle sweep strategy:

$$-2 \times \theta_1 < \theta_2 < 2 \times \theta_1. \tag{3.35}$$

At every increment of rotation, the cross-correlation between the mirrored rotating and the original coarsely rotated image is calculated and a score is recorded. The angle at which the highest score is found, denotes the θ_2 value [65]. The final angle is found by addition $\theta = \theta_1 + \theta_2$. Of the three final angles, the lowest angle is chosen because large rotations are typically indicative of poor headmask binarizing.

3.2.5 Mutual Information

Mutual information is calculated according to Equation 3.23 between a fixed atlas and a moving volume.

3.2.6 Correlation

Pearson's correlation coefficient r was used as the second intensity-based measure between two volumes according to the following equation:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{3.36}$$

where n is the number of voxels, x_i and y_i are the corresponding voxels, and $\overline{x}$ and $\overline{y}$ are the volume mean intensities.

3.2.7 Pixelwise Agreement

Pixelwise Agreement (PWA) is a measured across every slice in the registered volumes compared to the atlas template. From slice to slice there can be major structural differences especially at the extremities (Figure 3.9). PWA measures how similar the pixel values are at corresponding slices to the atlas by calculating the Mean-Squared Error per slice.

$$MSE(x, y) = \frac{1}{N}\sum(x - y)^2. \tag{3.37}$$

Once the slice PWA is obtained for each registered, the average across the entire database is obtained as one error value for each slice.

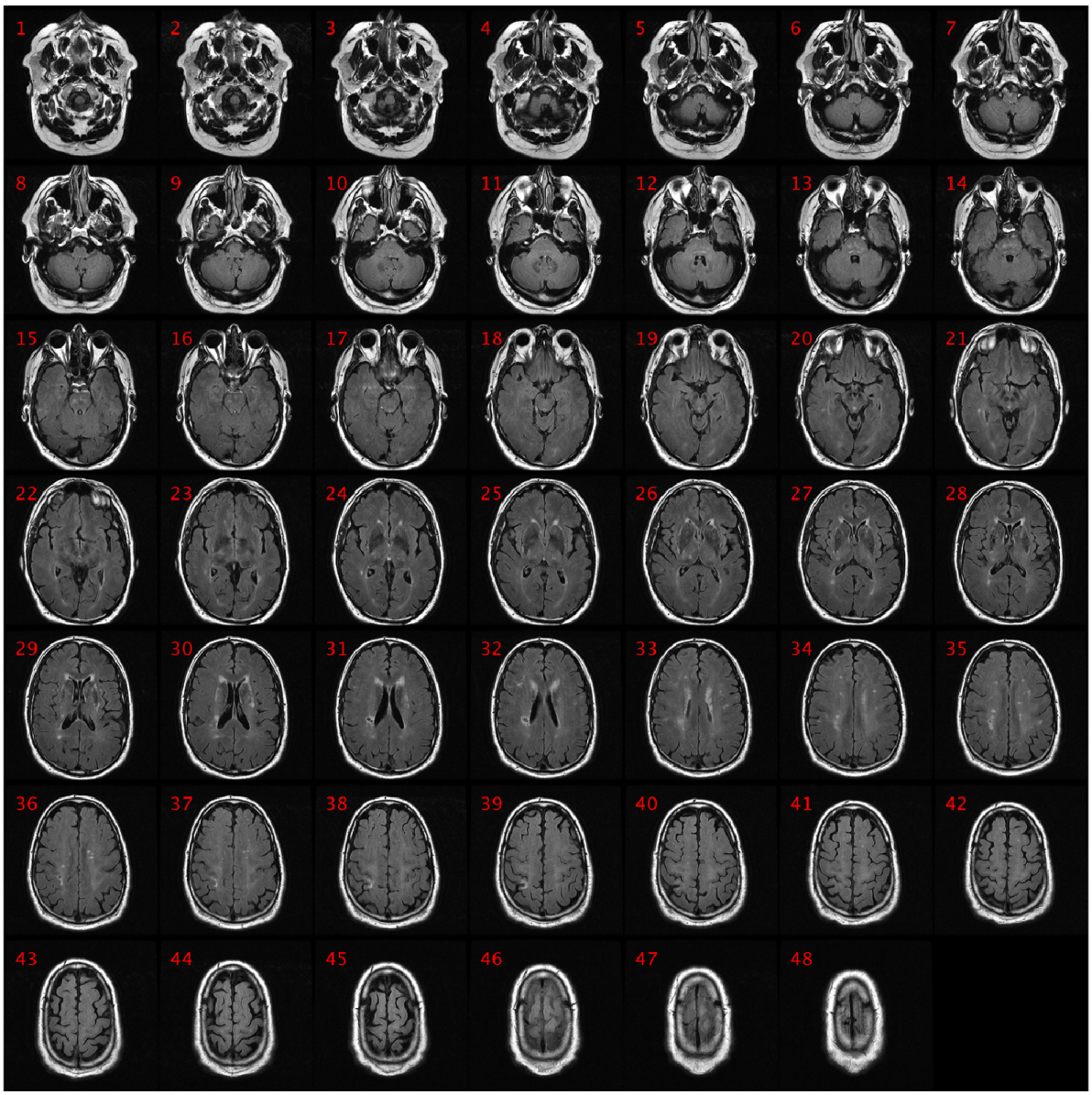


Figure 3.9: A typical brain volume is composed of many tissue-types that vary from slice to slice.

Chapter 4

Experiments and Results

This portion of the thesis presented will detail the experiments performed and results obtained. The discussion will begin with an analysis of the data used, experimental setup, the different registration settings for FLAIR registration, an overview of the Optical Flow models tested, and finally the results obtained.

4.1 Data

The FLAIR MRI scans were obtained from two databases. The first is the Canadian Atherosclerosis Imaging Network (CAIN) and second being the Alzheimer's Disease Neuroimaging Initiative. Both databases consist of aging population. CAIN is made up of 700 multi-patient and longitudinal patient volumes. All 700 volumes were used for Head Angle, Mutual Information, Correlation, and Synchronized Averaging metrics. PV and SSD calculations require manual segmentations of the brain, ventricles, and WMLS. Fifty volumes were segmented by trained medical students according to a predefined protocol. The mean age for CAIN was 73.87 ± 8.29 of which 410 were Male (74.47 ± 8.26 years old), 266 were Female (72.96 ± 8.25, and 24 had anonymized information. The fifty volumes were randomly selected from the CAIN database. The age of 47 of the participantrs ranged from 60 to 85 (74.6 ± 6.1) years old, of which 15 were female (75.5 ± 5.5) and 32 were male (74.2 ± 6.4). Three participants had anonymized age and gender. The scans for the entire CAIN set were obtained from February 2010 until January 2016 from various centers across Canada.

A second dataset from the ADNI database was used, consisting of 4263 volumes [66]. The total mean age was 73.48 (± 7.37) of which 1981 were Female (72.5 ± 7.4), 2280 were Male (74.34 ± 7.24), and 2 were unknown (82.76). From the ADNI database, 78 volumes were excluded due to severe artifacts sucah as hyperintense noise regions and missing or blank slices. Twenty volumes had their ventricles and white matter lesions annotated by the same experts and according to the same protocol used in the CAIN annotations.

For CAIN, the FLAIR MRI data came from several centers throughout North America that used a various number of scanner models from three manufacturers, General Electric Medical Systems (Discovery MR750, n=181), Phillips Medical Systems (Achieva, n=230), and Siemens (Skyra, n=162; IteraMR, n=14; TrioTim, n=113). The corresponding scanner settings are summarized in the Table 4.1.

Table 4.1: FLAIR MRI acquisition criteria for the CAIN database.

$Manufacturer$		$TR(ms)$	$TE(ms)$	$TI(ms)$	$VoxelSize(mm^3)$
GE	Discovery MR750	9700	140.84 - 150.24	2200	2.2157
Phillips	Achieva	1100	125	2800	1.1020
Siemens	Skyra	9000	119	2500	3.0
	IteraMR	9000	119	2500	3.0
	TrioTim	9000	122	2500	3.0

The ADNI database consisted of three MR scanner manufacturers with 15 total models. The three manufacturers were Philips Medical Systems (Intera, n=191; Achieva, n=520; Ingenia, n=83; Ingenuity, n=19; Gemini, n=35; total=848), Siemens (Verio, n=633; TrioTim, n=1401; Skyra, n=286; Prisma, n=5; Biograph-MR, n=13; SymphonyTim, n=2; total=2340), and General Electric Medical Systems (SignaHDx, n=23; SignaHDxt, n=438; Discovery MR750, n=600; Discovery MR750w, n=14; total=1075). The acquisition settings are shown in table 4.2.

Table 4.2: FLAIR MRI acquisition criteria for the ADNI database.

$Manufacturer$		$TR(ms)$	$TE(ms)$	$TI(ms)$	$VoxelSize(mm^3)$
GE	SignaHDx	11002	153.98	2250	3.6927
	SignaHDxt	11002	147.93-192.17	2250	3.6927
	DiscoveryMR750	11000	149.67	2250	3.6927
	DiscoveryMR750	11000	151.40	2250	3.6927
Phillips	Intera	9000,6000	90, 140	2500,2000	3.6927,4.3945
	Achieva	9000	90	2500	3.6927
	Ingenia	9000	90	2500	3.6927
	Ingenuity	9000	90	2500	3.6927
	Gemini	9000	90	2500	3.6927
Siemens	Verio	9000	91	2500	3.6927
	TrioTim	9000	90	2500	3.6927
	Skyra	9000,11900	91	2500,2753.4	3.6927
	Prisma	9000	122	2500	3.6927
	Biograph-MR	9000	122	2500	3.6927
	SymphonyTim	10000	125	2200	4.3945

The standardized volumes had the intensity histograms aligned and the background was nulled according to [67].

4.2 Validation Metrics Experimental Setup

The FLAIR volumes from both databases were registered to the FLAIR template space provided by Brainder via the four registration methods [68]. An overview of the metrics extraction pipeline is seen in Figure 4.1.

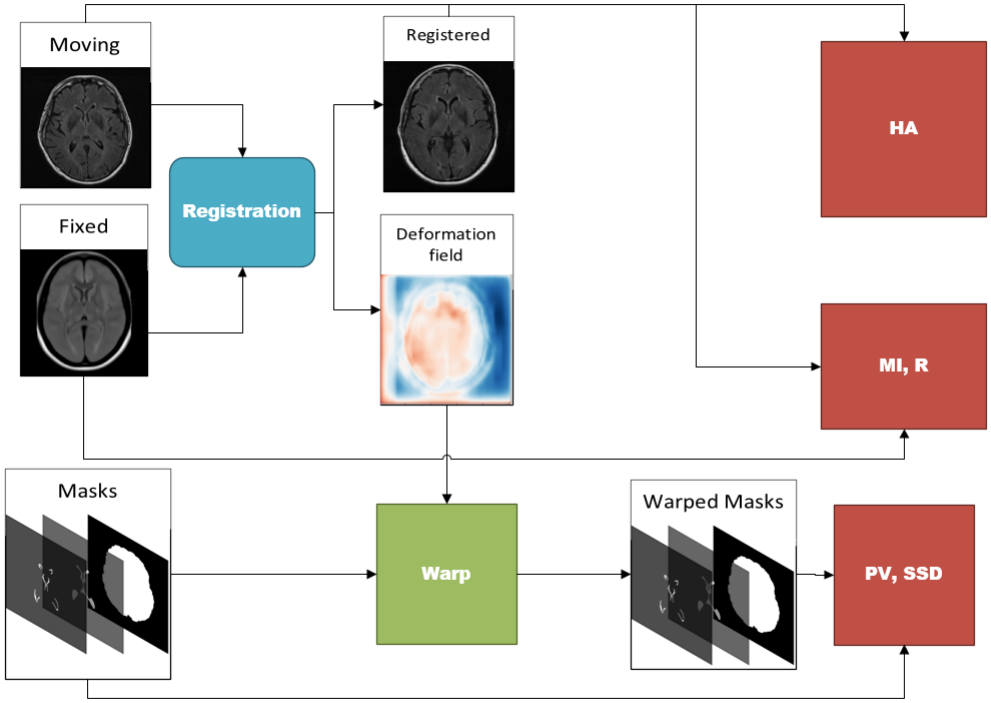


Figure 4.1: Registration metrics extraction process. HA = Head Angle, MI = Mutual Information, R = Correlation Coefficient, PV = Percent Volume, SSD = Surface to Surface Distance

For the corresponding brain, ventricle, and WML masks, the resultant deformation field from each registration method is applied to warp to the same space as the registered volume. The PV calculations of the ventricles and WMLs are with respect to the whole brain volume. The SSD is computed between the ventricle surface and the brain surface. Finally, for all registration methods, the large-scale metrics are computed: Head Angle with respect to the saggital plane, MI and R were both calculated between the registered volumes and atlas. For a baseline comparison, all metrics are computed prior to any registration on the volumes and masks. No pre-processing was done to any of the volumes other than resizing to the atlas (resolution $256, 256, 55$). The warped masks were binarized with a threshold of 0.1 so as to avoid the non-integer values obtained from interpolation.

For the Voxelmorph training procedure, both the atlas and the moving volumes were resized to

$256, 256, 64$. This particular resolution was chosen to be a binary multiple of 2^n to ensure that the decoder arm of the U-net network is able to rebuild the learned feature maps to the original input size and warp the volumes accordingly.

4.3 Optical Flow Implementation Details

The FlowReg model was trained solely on the CAIN database, using 400 volumes of 55 slices for a total of $22,000$ images. As a pre-processing step, the volumes were first affinely registered using ANTs [28]. The fixed volume is the the template from Brainder database [68]. The optimizer used was Adam [51] with a $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and a learning rate of $lr = 10^-4$. Training was done for 100 epochs using a batch size of 32 image pairs at seven resolutions (Figure 3.2). In the loss function, the Charbonnier penalty function (3.16) played a major role in both reducing overfitting and overtly large deformations (Figure 4.2). For both smoothness and photometric loss, the chosen α value was 0.2.

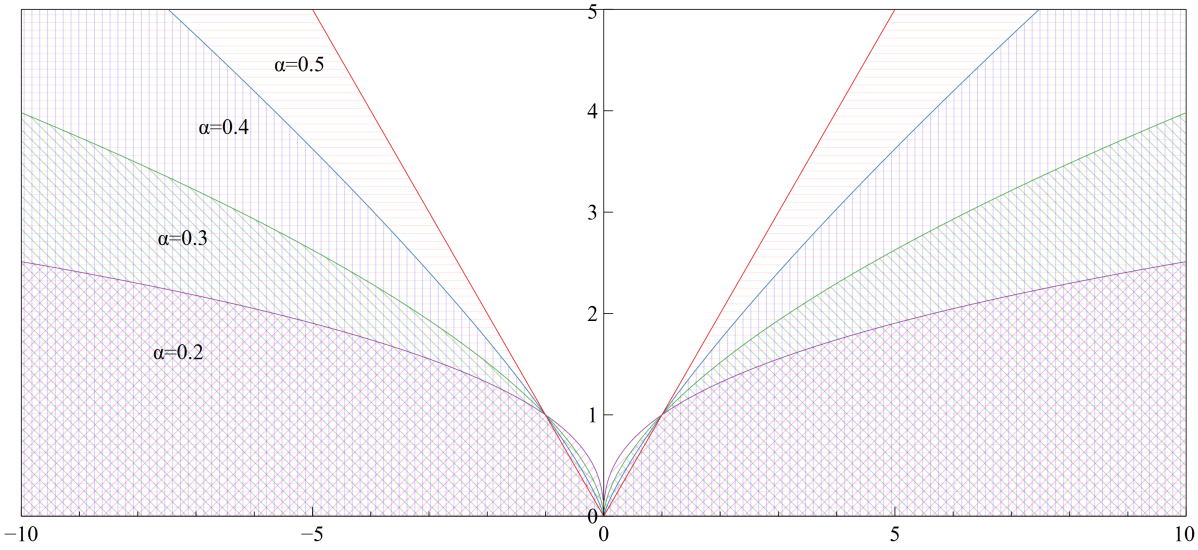
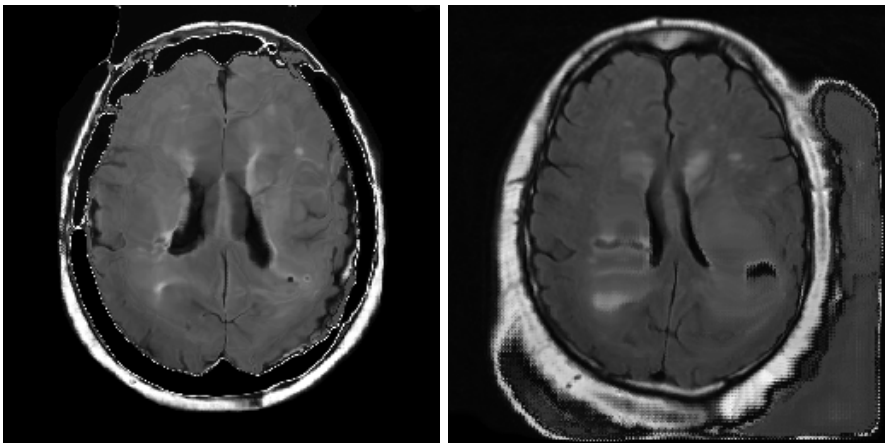
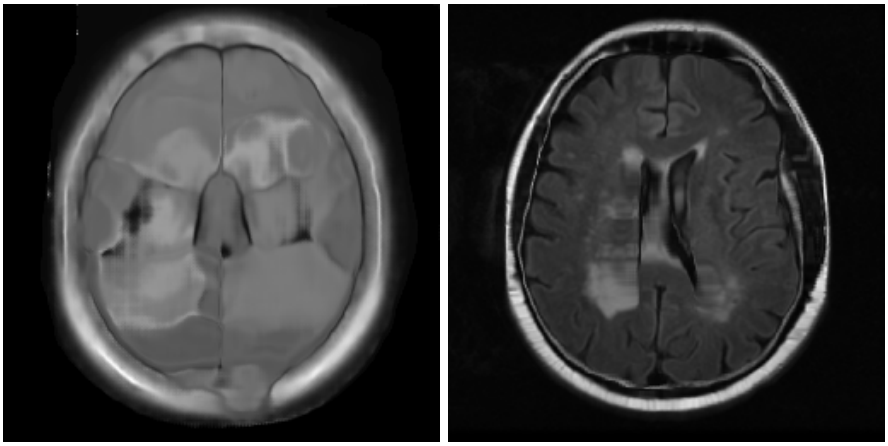


Figure 4.2: Charbonnier function (Eqn 3.16) showing the effect of various α values of 0.5, 0.4, 0.3, and 0.2.

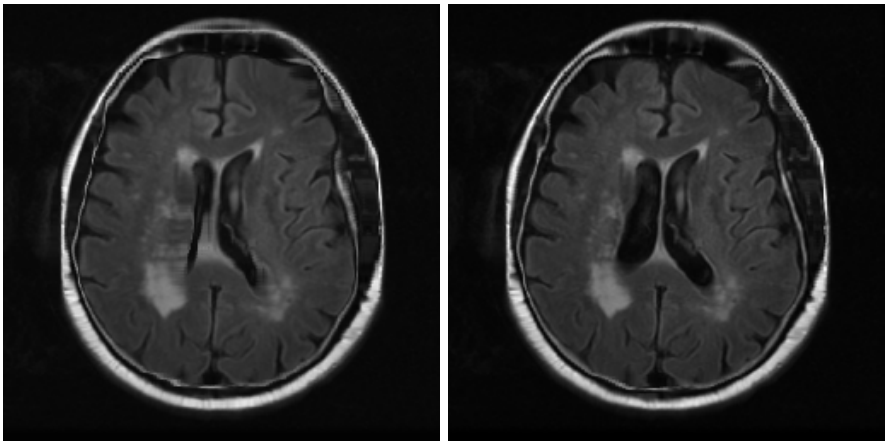
The training procedure consisted of hyper-parmater tuning by varying the α value and compounding the correlation loss (Eqn: 3.17) to the total loss and some results are seen in Figure 4.3).



(a) $\alpha = 0.45$. (b) $\alpha = 0.4$.



(c) ℓ_{corr} only with smoothness. (d) $\alpha = 0.3$.



(e) $\alpha = 0.25$. (f) $\alpha = 0.2$.

Figure 4.3: The results of testing different α values in the Charbonnier penalty function during the training phase (Eqn: 3.16).

Once the desired model was attained, an ablation study was performed by removing key portions of the loss function 3.14 to observe what effect the included components played. The first experiment was to remove the correlation loss 3.17 and keep only the photometric and smoothness loss. The major result that was noticed is that there was subpixel level of vector movement. This was evident by examining the vector field magnitude at various slices as seen in Figure 4.4. Essentially, the moving image looked exactly as it was prior to registration as the highest vector magnitudes are in the 10^{-3} magnitude range.

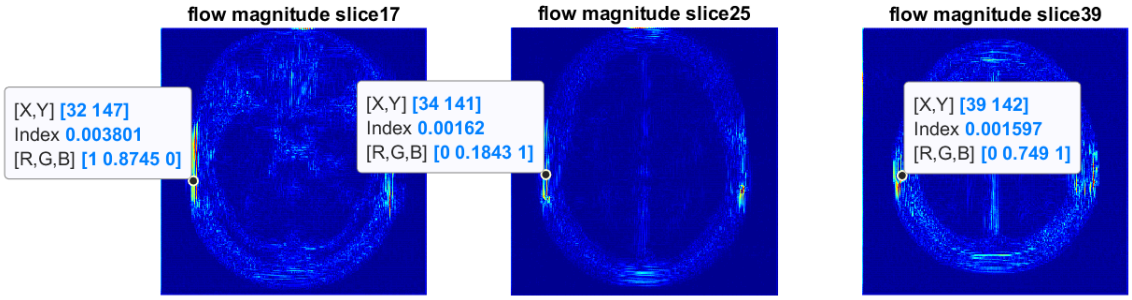
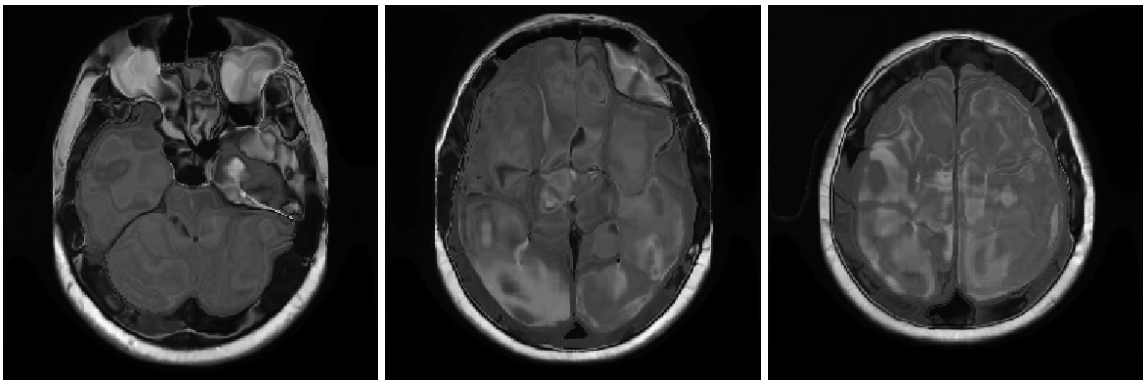


Figure 4.4: Vector field magnitudes. The index value represents the magnitude or how far that pixel should move.

The last ablation experiment was to train using only correlation without photometric or smoothness. One example was already shown in Figure 4.3c and the same was confirmed in Figure 4.5. Wheras in the



(a) Slice 15. (b) Slice 25. (c) Slice 35

Figure 4.5: Ablation study using only the correlation component of the loss function.

The regularization parameters for the total loss function (Eqn: 3.14) were set as follows: $\gamma = 0.5$, $\zeta = 1.0$, and $\lambda = 1.0$. The specific values were chosen because of the inherent nature of the warping. It was desired to obtain the similarly distributed intensity distributions, with low variance between the fixed and moving warped and high covariance (Eqn: 3.17). As a balancing act, the smoothness loss (Eqn: 3.19) implements a weight factor of $\lambda = 1.0$ to reduce overfitting and any large distortions. The photometric loss (Eqn: 3.15) has a weight of 0.5 to also reduce any large distortions and smoothing as seen in Figure 4.3a.

4.4 Registration Settings

Four validation metrics were used to asses the efficacy of FLAIR MRI registration via all registration methods. All four registration algorithms use different strategies to achieve warp moving images to a fixed image space, thus giving an insight into their real-world practicality. Registration was performed in the Python programming language, while the metrics were calculated in Matlab [69]. The data was stored as three dimensional arrays and saved as .MAT files which allows quick interchange in both programming languages for registration and metric extraction. Three of the four registration methods were performed on a Windows 64-bit system, that has a 3.2GHz Intel Core i7, 32Gb of RAM, and an nVidia GTX1080-TI with 11Gb of video-RAM. ANTs registration currently only works on Unix-based systems (although workarounds are available via Ubuntu command line and Virtual machines, but are highly inefficient since resources are taken by these pseudo-Unix machines on top of the actual registration algorithms) and thus was performed on a Macbook Pro 13" with macOS Mojave (10.14.5), 3.5GHz Intel Core i7, 16Gb DDR3 RAM, and integrated Intel graphics.

4.4.1 Demons

Demon's registration was done using SimpleITK's implementation of the original Demon's algorithm based on Thirion's paper [25] [26] [70] [40]. This method uses intensity histogram matching to make the intensities similar prior to registration. The settings used were as follows:

```
matcher = sitk.HistogramMatchingImageFilter()
matcher.SetNumberOfHistogramLevels(1024)
matcher.SetNumberOfMatchPoints(7)

demons = sitk.DemonsRegistrationFilter()
demons.SetNumberOfIterations(50)
demons.SetStandardDeviations(1.0)
```

The number of bins for joint histogram and marginals were 1024, 50 iterations, and the σ for the Gaussian smoothing of the deformation field was set to 1.0 as seen in Equation 2.24. Once the displacement field is calculated, the binary masks are warped into the fixed image space.

4.4.2 ANTs

The Advanced Normalization Tools (ANTs) algorithm was used via the Python implementation and using the classic (Symmetric Normalization) option. These options include affine and a deformable transformation. The optimization metric is Mutual Information (MI).

```
mytx = ants.registration(fixed=fixedVol, moving=movingVol, type_of_transform='SyN',
    verbose=1)
regvol = mytx['warpedmovout']
fwdtform = mytx['fwdtransforms']
invtform = mytx['invtransforms']
```

where the $regvol$ is the registered volume, $fwdtform$ is the transform to move from moving to fixed space, and $invtform$ is the inverse transform from fixed to moving (Figure 2.11) [28].

4.4.3 Elastix

The Elastix algorithm was implemented using the Python API designed for SimpleITK [40]. A Python package was designed by Marstal et al [71] to allow ease of implementation. The algorithms uses a four level Gaussian smoothing pyramid. The deformations are rigid, affine, and b-spline transforms with the following similarity measures for each, respectively: Normalized Cross-Correlation, Mutual Information, and Mutual Information with a transform energy penealty that constrains any large erroneos transforms. For all three transforms the Adaptive Stochastic Gradient Descent (ASGD) optimizer and a B-spline image interpolator were used.

4.4.4 VM

Voxelmorph (VM) is a Convolutional Neural Network (CNN) based approach implemented in Tensorflow using the Keras wrapper [50] [49] [47]. The U-net style network consists of Convolutional Layers followed by Rectilinear Activation functions which learns the vector flow of individual voxels [5]. The last layer consists of a Dense Deformation to warp the moving image according to the learned deformation vector field to the fixed image space. The loss function between the transformed and fixed volumes is a sum of the 3D cross-correlation and *l2* gradient loss using the Adam optimizer [51]. Training was done on all 700 volumes, with a $80/20$ training validation split for 10 epochs. The weights were saved at every epoch, noticeable errors were noticed after the seventh epoch, as such, all experiments trained for 7 epochs, similar to the author's experiments [47].

4.5 Manual Masks Registration Results

All metrics were calculated in MATLAB prior and post registration for comparison purposes. The Mean-Error (ME) metric was used to assess the amount of difference between the *Before* and *after* registration. For *FlowReg* as *AffineANTs* was used as a preparatory step, the difference was found with respect to the *Before* and to *AffineANTs*. This is due to the additive effect that may be perceived after the affine registration with *ANTs*. To test the true effect of *FlowReg* on PV and SSD, it is important to isolate it from the error that may be induced by the affine transform that was done via *AffineANTs*. The equation is as follows:

$$\varepsilon(A + B) = \varepsilon(A) + \varepsilon(B), \tag{4.1}$$

where A is *AffineANTs* and B is *FlowReg* registration errors in the PV of the structure in question.

4.5.1 PV

The proportional volume calculations for both ventricles and white matter lesions in the CAIN database were calculated as seen in Figure 4.6 and Figure 4.7.

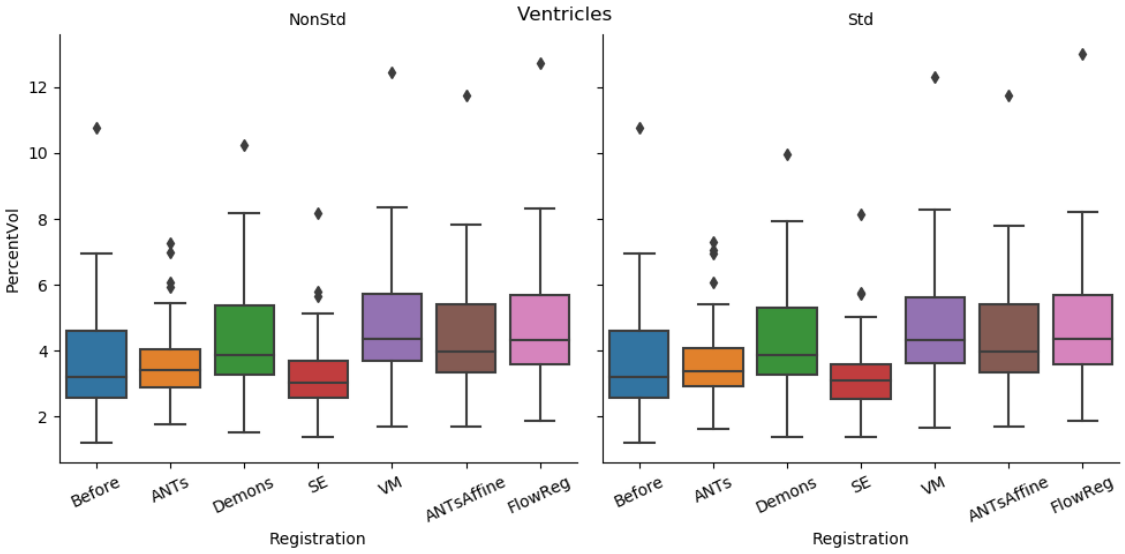


Figure 4.6: Proportional ventricular volume of the pre- and post-registered binary masks for CAIN ($n = 50$).

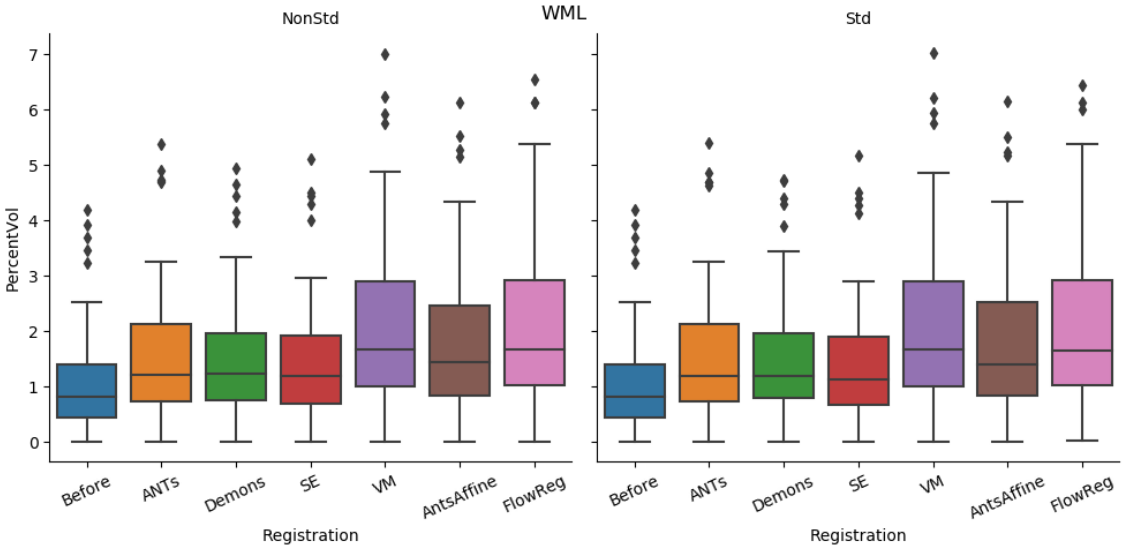


Figure 4.7: Proportional ventricular volume of the pre- and post-registered binary masks for CAIN ($n = 50$).

Difference magnitutdes between registered and non-registered volumes are found and shown in Figure 4.8 and Figure 4.9.

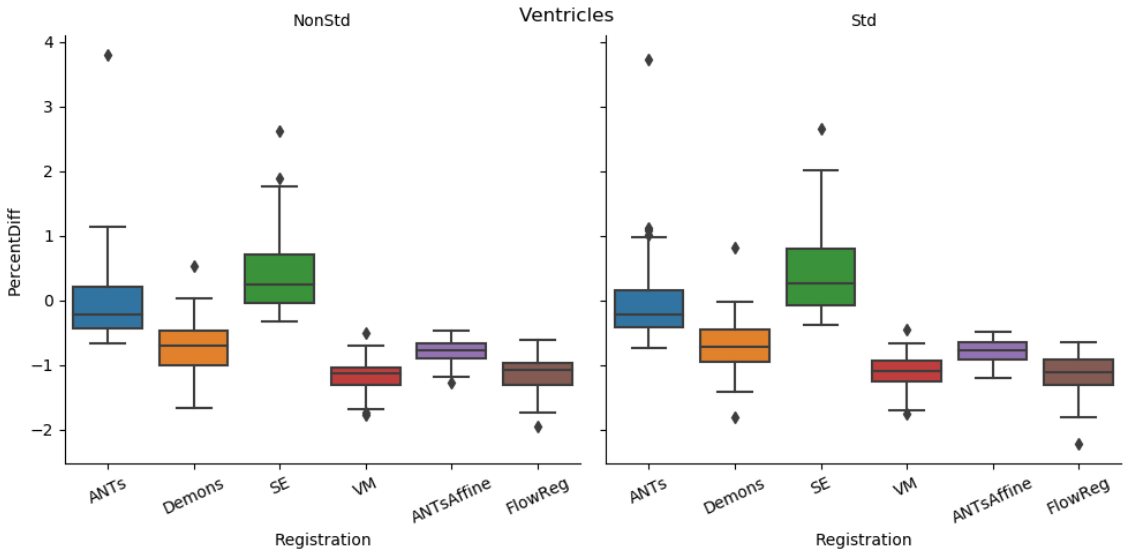


Figure 4.8: Proportional ventricular volume difference between pre- and post-registered binary masks for CAIN ($n = 50$).

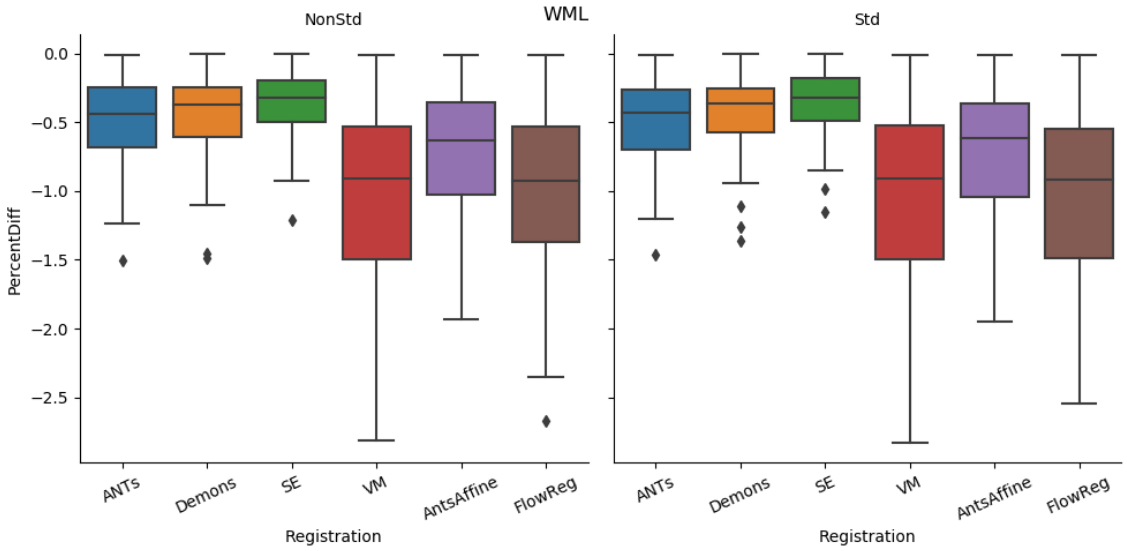


Figure 4.9: Proportional WML volume between pre- and post-registered binary masks for CAIN ($n = 50$).

The ME errors are shown in the figure below (Figure 4.10).

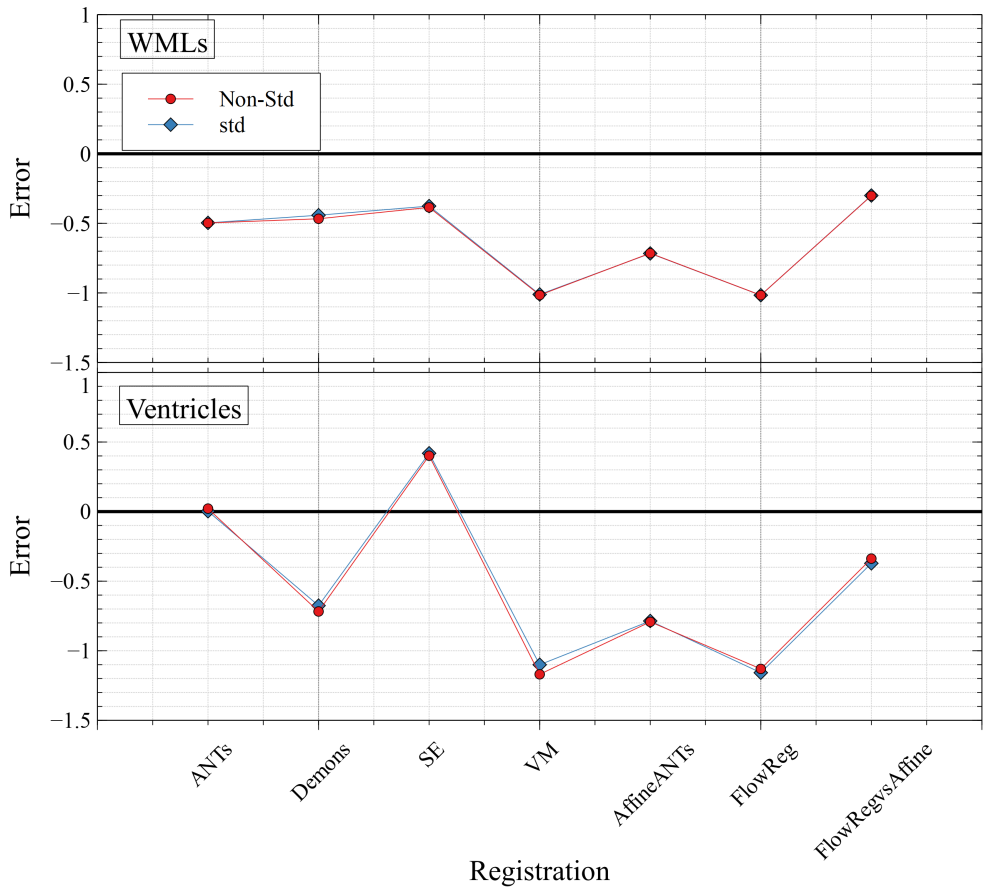


Figure 4.10: Error of PV from *Before* registration for the CAIN database. *FlowRegvsAffine* is a comparison of the PV metric with respect to the affine registration.

Overall, it can be seen that, on a general basis, the PV of both WMLs and ventricles are being overestimated for most methods. Taking a closer look at the PV-WML, *ANTs*, *Demons*, and *SE* are the closest to the *Before* for both Raw and Standardized data. *VM* has the highest difference. It is important to keep in mind that *FlowReg* uses *ANTsAffine* as a preparatory step, thus any observed effects are likely cumulative from the affine transformation. This effect is especially visible when examining Figure 4.10 where the error is minimal when compared to *ANTsAffine* and thus equation 4.1 hold true where $\varepsilon_{FlowReg} = \varepsilon_{AffineANTS} + \varepsilon_{FlowRegvsAffine}$. The same trend can be observed when comparing proportional ventricular volume where *ANTs* and *SE* perform well, and *FlowReg* has a low error compared to *ANTsAffine*.

The ADNI manual segmentation were treated to the same metrics as the CAIN. Proportional volume

calculations were performed for both ventricles and WMLs for all four registration modalities. For the raw and standardized volumes the MSE compared to prior registration is observed. (Figures 4.11, 4.13, 4.15). The WML proportional volumes, similarly show no difference to pre-registration except for Voxelmorph (Figures 4.12, 4.14, 4.15). This remains true for both non-standardized and standardized volumes.

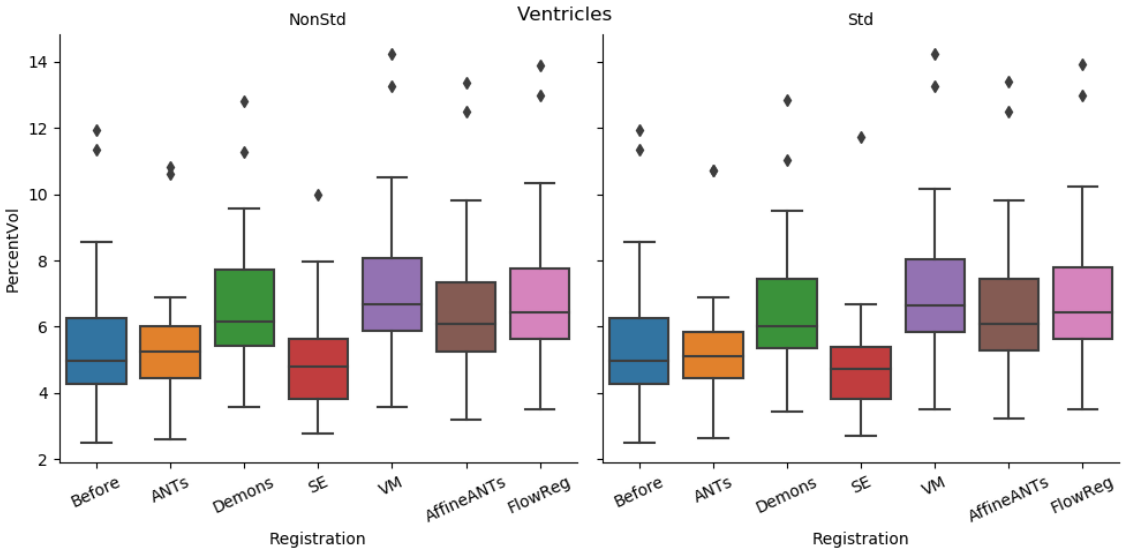


Figure 4.11: Proportional ventricular volume of the pre- and post-registered binary masks for ADNI (n=20).

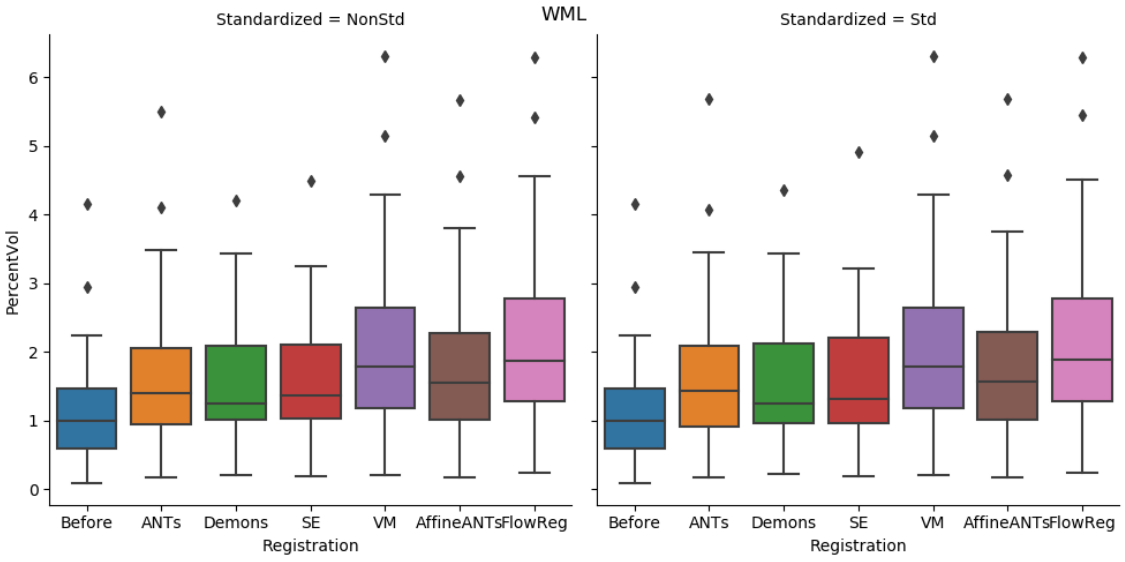


Figure 4.12: Proportional WML volume of the pre- and post-registered binary masks for ADNI (n=20).

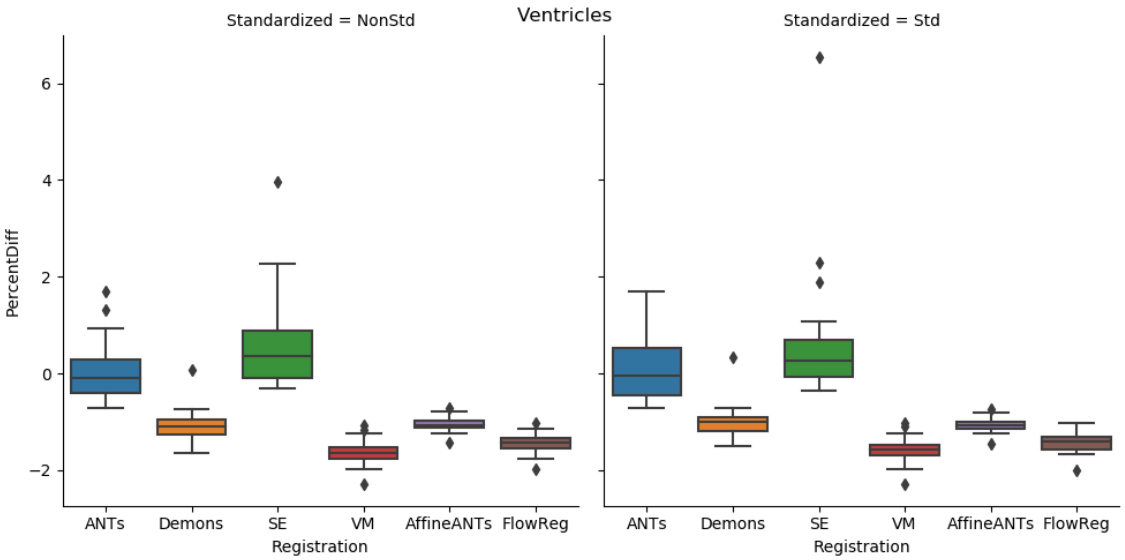


Figure 4.13: Proportional ventricular volume difference between pre- and post-registered binary masks for ADNI (n=20).

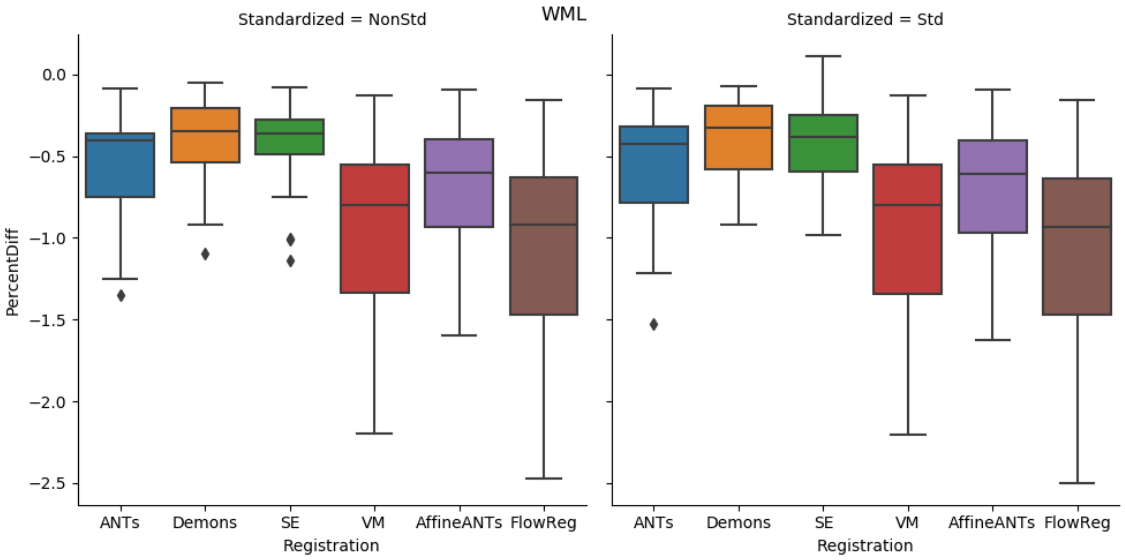


Figure 4.14: Proportional WML volume between pre- and post-registered binary masks for ADNI (n=20).

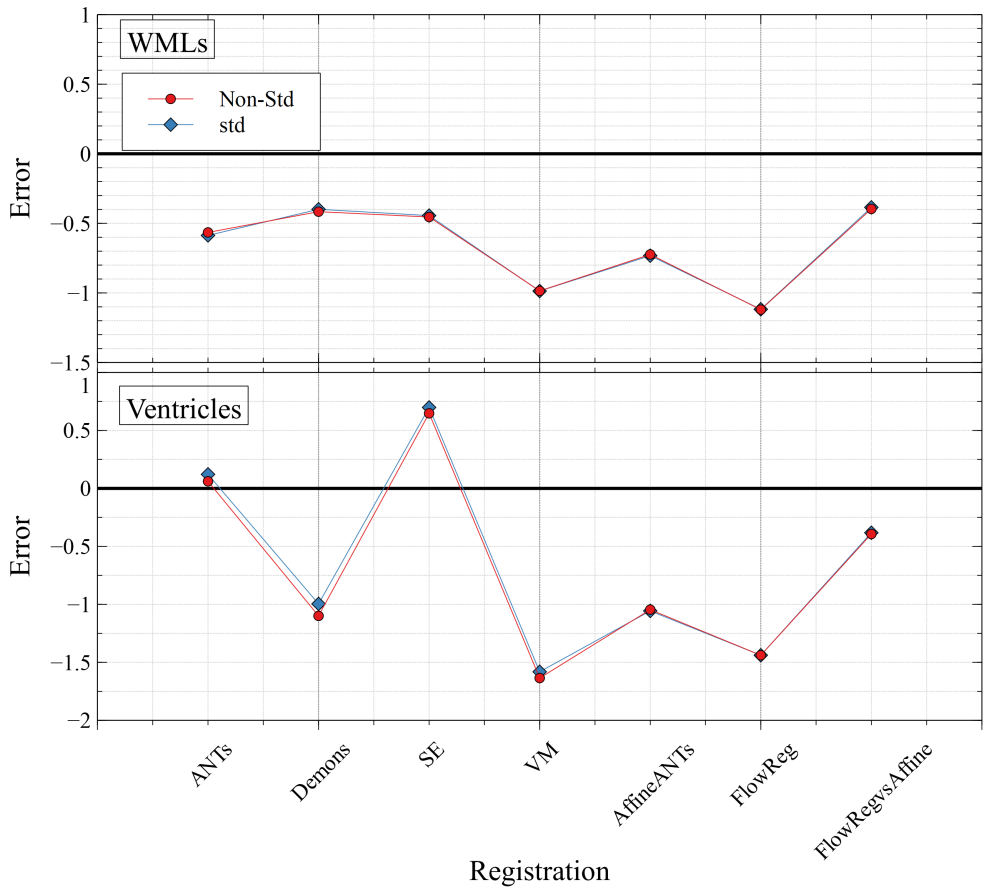


Figure 4.15: Error of PV from *Before* registration for the ADNI database. *FlowRegvsAffine* is a comparison of the PV metric with respect to the affine registration.

4.5.2 SSD

Surface to surface distance (SSD) was calculated from the surface of the ventricles to the surface of the brain. The average (Figure 4.16) and standard deviation SSD (Figure 4.17) are calculated. Significant changes indicate that the overall surface of the ventricles has changed with respect to the brain surface (Figure 4.18). Interestingly, *FlowReg* with *ANTsAffine* performs best when comparing to *FlowReg* on its own as defined by equation 4.1.

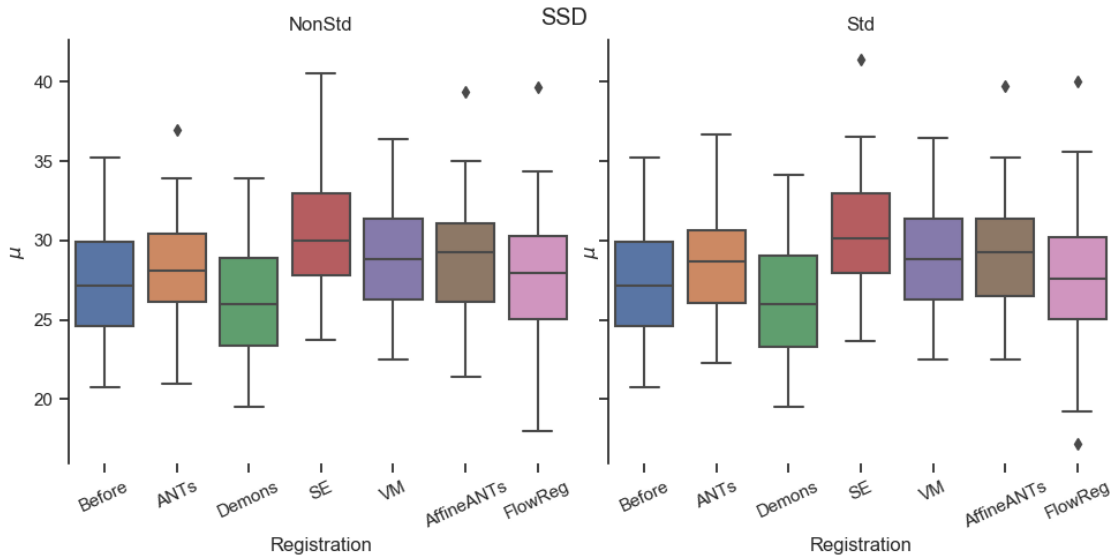


Figure 4.16: Mean Surface to Surface distance registered via the four methods of the CAIN volumes (n=50).

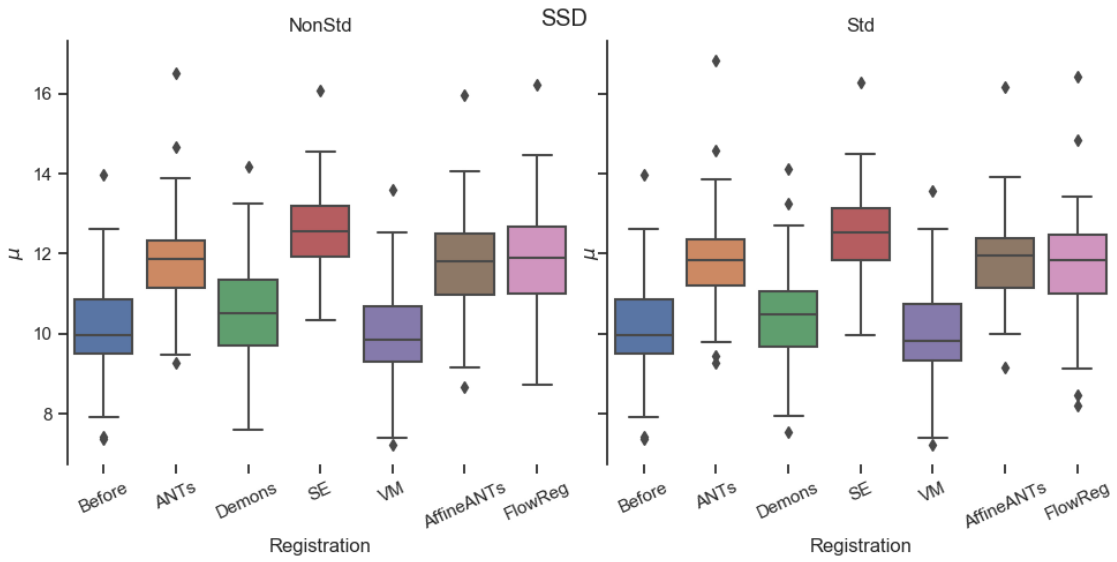


Figure 4.17: Standard Deviation Surface to Surface distance registered via the four methods of the CAIN volumes (n=50).

Examining the mean SSD, *SE* and *FlowReg* with *AffineANTs* registration show the largest differences while *ANTS*, *Demons*, and *VM* have the lowest errors. Standardization does not change the outcome showing that *ANTs* and *Demons* are still similar to pre-registration while *SE* and *FlowReg* are different. The standard Deviation SSD metric shows *Demons* and *VM* as being the closest

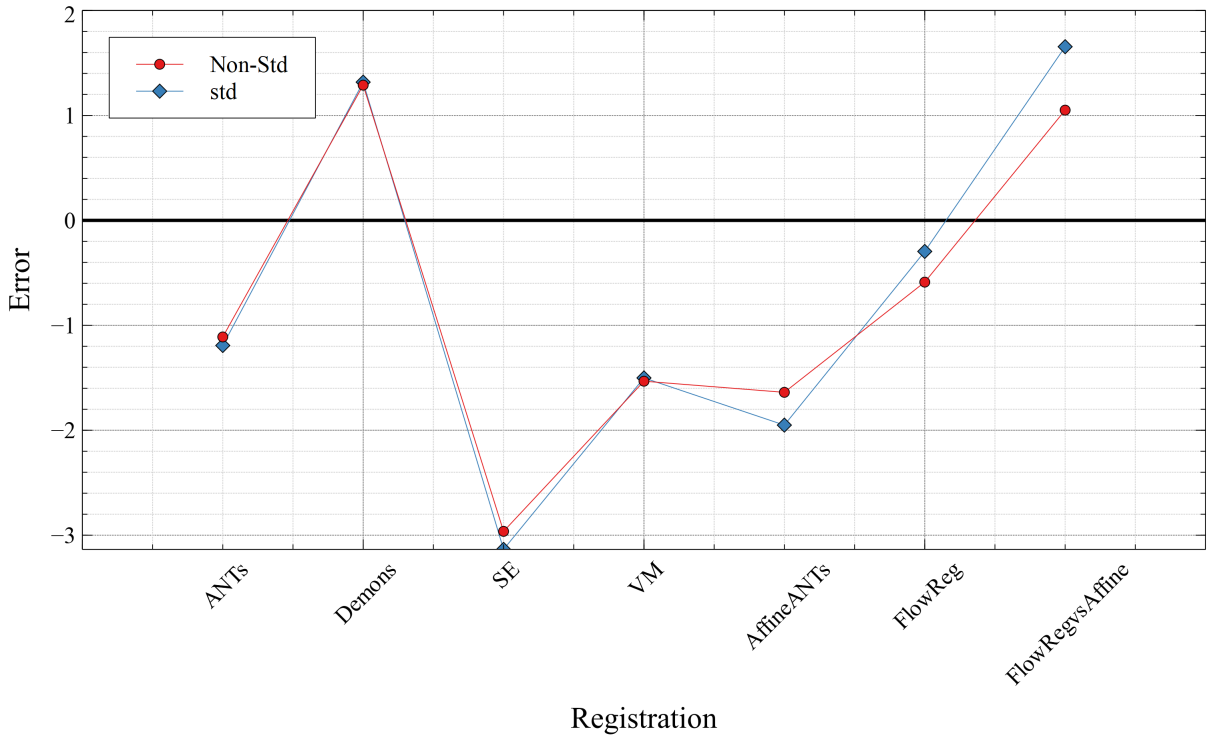


Figure 4.18: The Error of the SSD metric compared with prior to registration of the CAIN database. *FlowRegvsAffine* is a comparison of the SSD metric with respect to the affine registration.

ADNI results show simlar trends as CAIN(Figures 4.19, 4.20). ANTs and SE both consistently have higher mean SSD than prior to registration and Standardization does not change the results. The standard deviation SSD is also higher for ANTs and SE which describes higher variation in the distance between the ventricles and the brain in the ADNI database. The same is true when examining the standardized volumes. Observing the mean error in 4.21 it can be seen that the error effect as defined by 4.1 is especially apparent for this metric where *FlowReg* has a near null error.

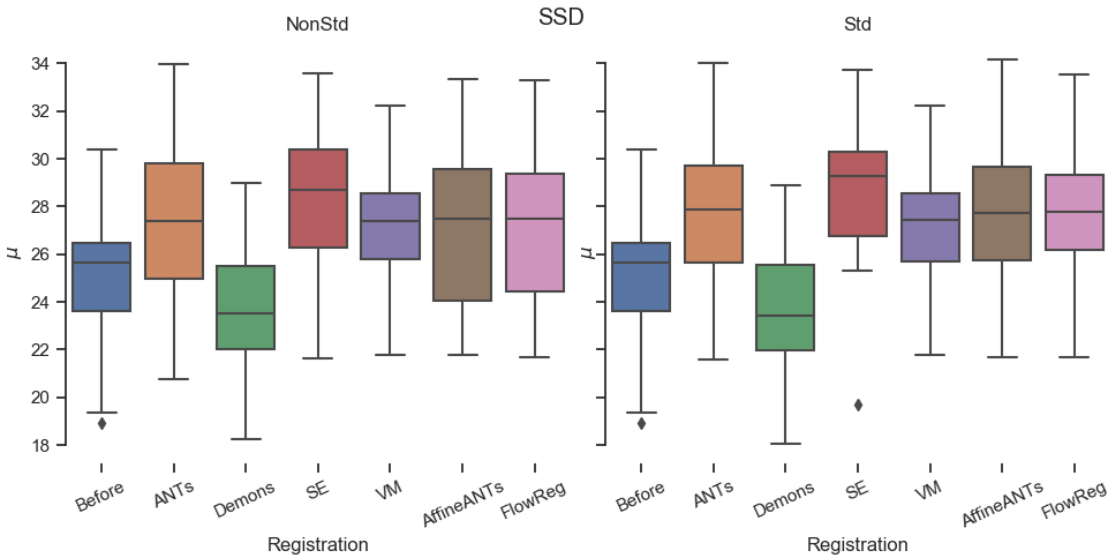


Figure 4.19: Mean Surface to Surface distance registered via the four methods of the ADNI volumes (n=20).

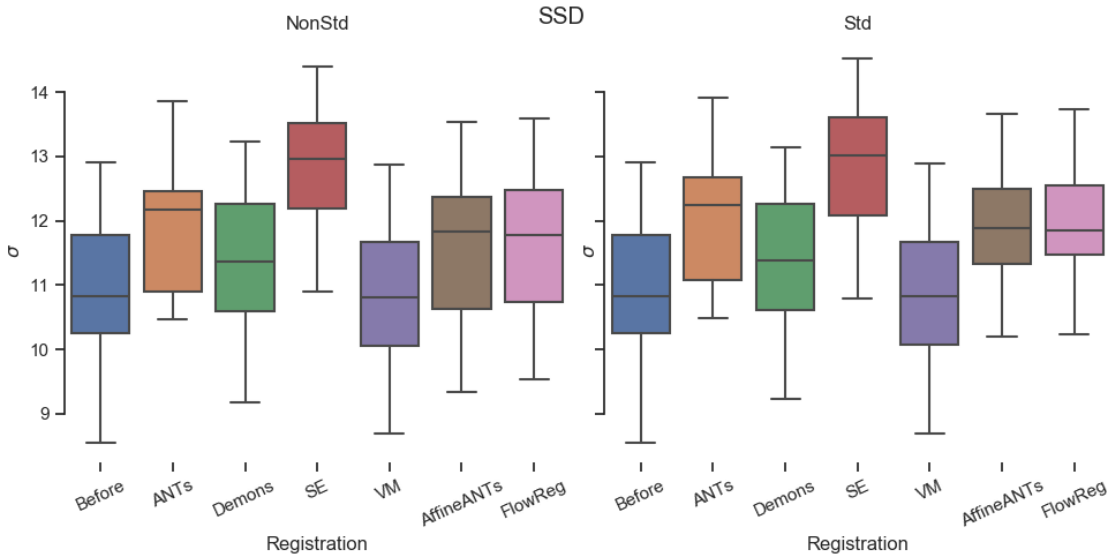


Figure 4.20: Standard Deviation Surface to Surface distance registered via the four methods of the ADNI volumes (n=20).

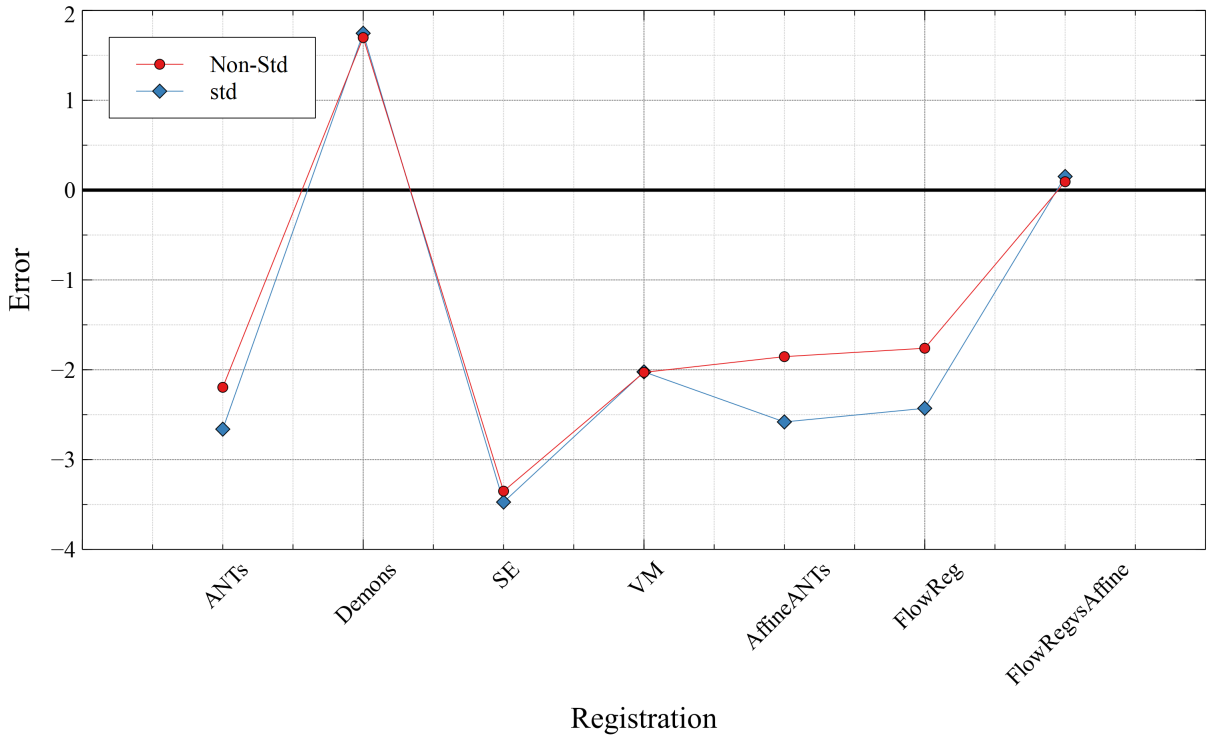


Figure 4.21: The Error of the SSD metric compared with prior to registration of the ADNI database. *FlowRegvsAffine* is a comparison of the SSD metric with respect to the affine registration.

4.6 Large Scale Registration Metrics

Three metrics were used to assess the quality of registration on a large scale without the use of binary masks. These metrics are the Head Angle, Pixelwise Agreement, Mutual Information, and Correlation. Synchronized averaging has been used to create atlases based on the registration modality.

4.6.1 Head Angle, Pixelwise Agreement, Mutual Information, Correlation

The Head Angle calculations were performed on the entire dataset for both databases, CAIN and ADNI. The plots indicate the angle of all the volumes in the database with respect to the midsaggital plane. Calculations were done for both Raw and Standardized volumes. For this metric, it is important to reduce the angle with respect to the vertical. Thus, tightly clustered angles to the 0° line indicates highest performes in terms of vertical alignment (Figure 4.22). To measure the

variability in angles we define spread, ς, as three times the standard deviation σ away from the mean μ. Prior to registration of the CAIN database, the mean average angle is found to be $\mu = 0.37°$ with a spread $\varsigma = 15.64°$. *FlowReg* has the lowest spread of all the metrics $\varsigma = 1.52°$, followed by *AffineANTs* $\varsigma = 2.52°$, and finally *ANTs* with a spread of $\varsigma = 6°$. Standardization significantly improves the performance of the registration methods on this metric. Specifically, we notice *ANTs* has a mean of $\mu = -0.21°$ with a spread of $\varsigma = 1.62°$. *FlowReg* again has the lowest spread at $\varsigma = 1.6°$ for the standardized dataset.

The ADNI non-standardized pre-registered volumes have a mean angle of $\mu = 0.83°$ and a spread of $\varsigma = 7.38°$ (Figure 4.23). The volumes registered via *FlowReg* have the tightest region of spread in terms of angle via visual inspection, the mean is $\mu = -0.12°$ and the spread of $\varsigma = 1.68°$ followed by *SE* with a spread of $\varsigma = 1.94°$, and *ANTs* $\varsigma = 2.26°$.. Voxelmorph is the sole registration modality that has the highest measure of spread at $\varsigma = 7.53°$. Standardization reduces the spread of the head angle measurement for all registration methods.

Pixelwise Agreement (PWA) is measured by calculating the per-slice MSE when compared to the original atlas template and the average error at each slice is reported in Figure 4.24 for CAIN and Figure 4.25. As MRI images vary considerably across a single volume due to structural differences from one slice to the next, if registration is to be performed accurately, the level of agreement has to be measured accordingly, which is the intent behind PWA. The highest error is seen by *Demons*, *SE*, *VM*, while *FlowReg* holds the lowest by a large margin for the non-standardized dataset across all slices. Interestingly, again we see that standardization significantly improve the *ANTs* registration, bringing the error values in the *FlowReg* territory.

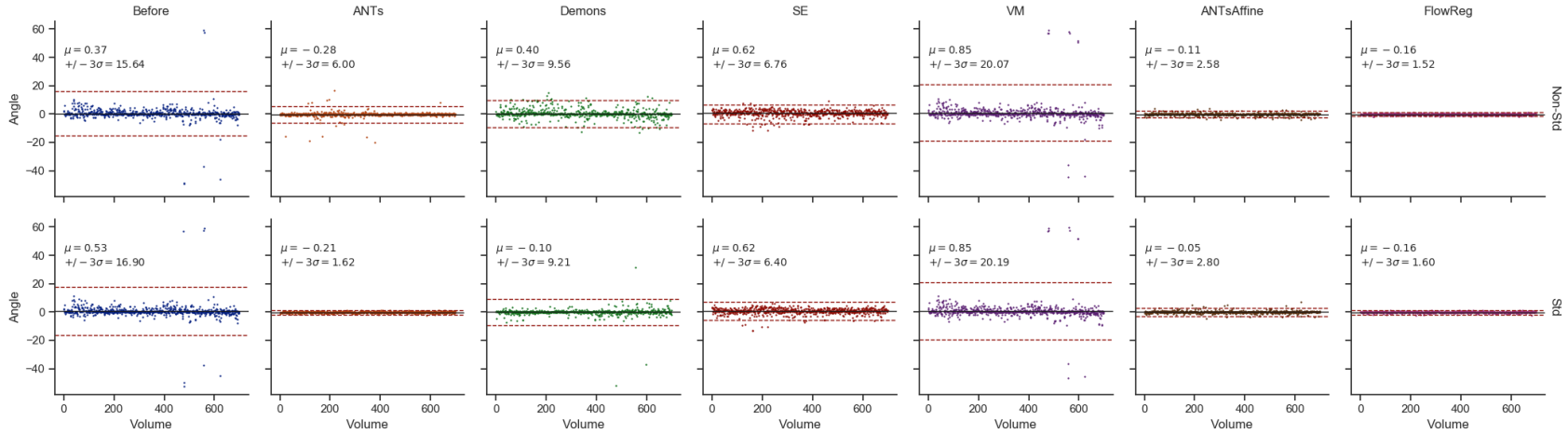


Figure 4.22: CAIN head angle measurement. The black line indicates the sample mean, while the dotted line is the spread measure ς away from the mean, with the actual values reported in the figures (n=700).

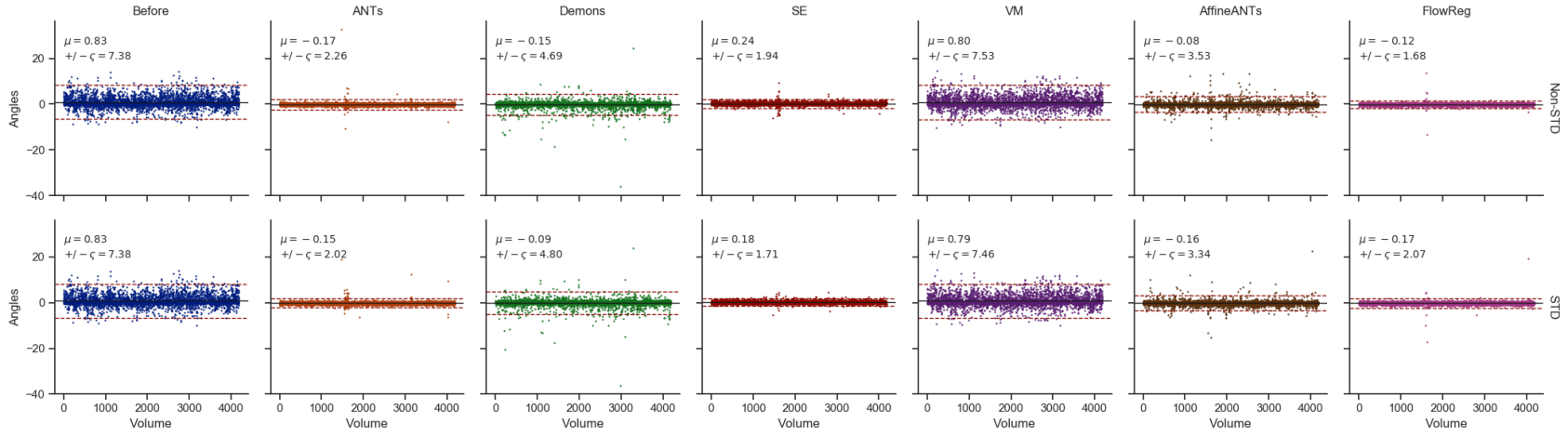


Figure 4.23: ADNI head angle measurement. The black line indicates the sample mean, while the dotted line is the spread measure ς away from the mean, with the actual values reported in the figures (n=4263).

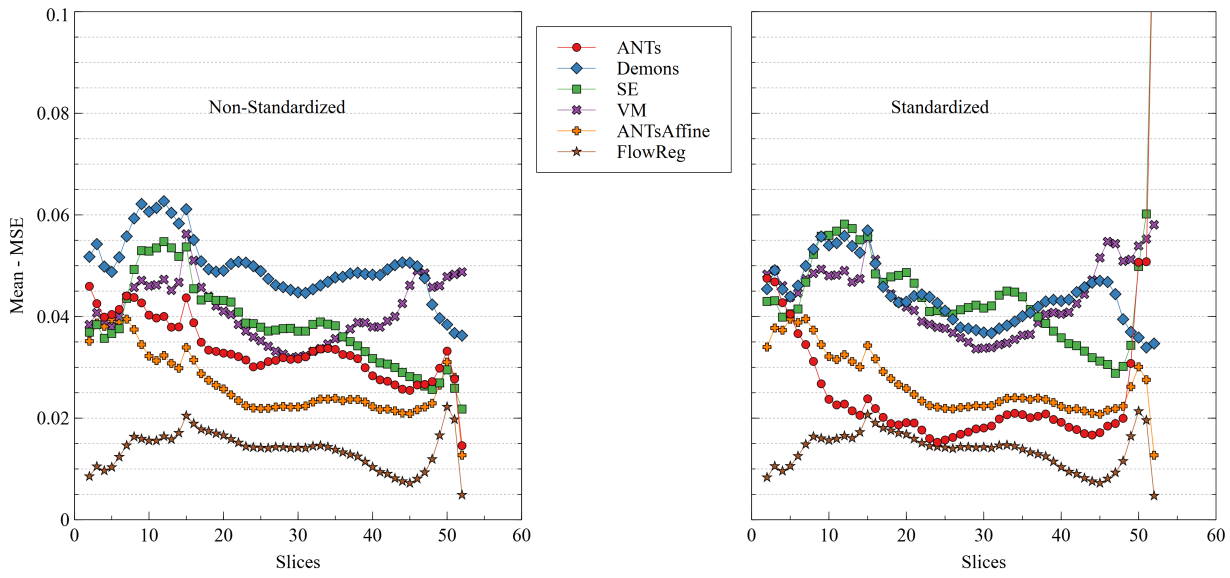


Figure 4.24: PWA as a measure of alignment after various registrations for CAIN. The error is reported as an average across all volumes on a per-slice basis.

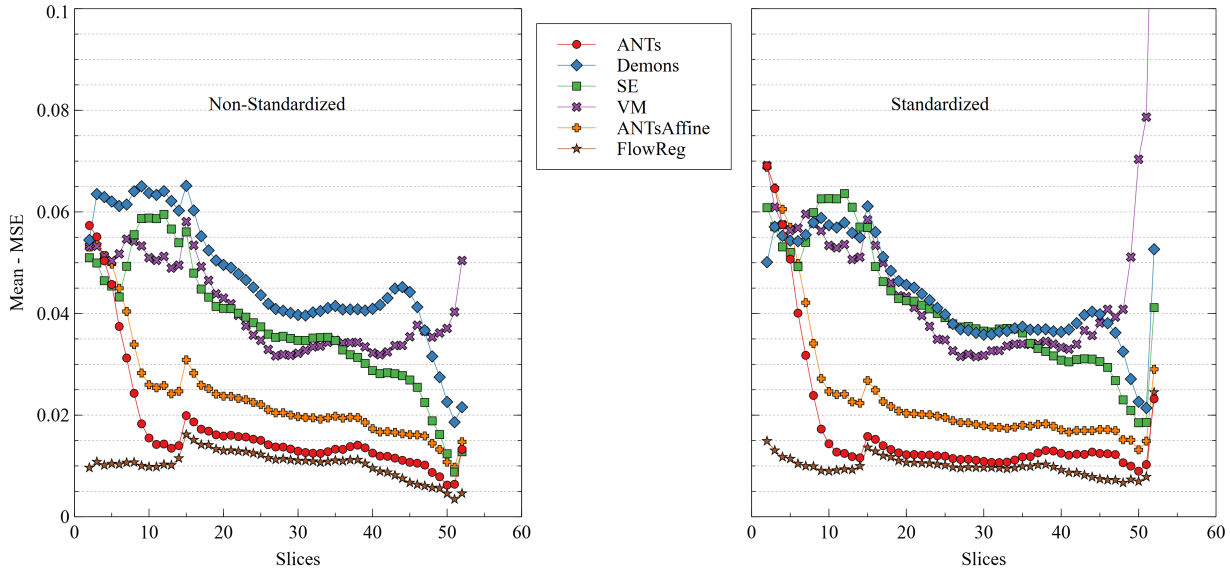


Figure 4.25: PWA as a measure of alignment after various registrations for ADNI. The error is reported as an average across all volumes on a per-slice basis.

Mutual Information (MI) is an information theoretic metric that is often used as a similarity metric in many registration metrics. Thus, it is expected that those modalites should perform best. Compared to pre-registration, *FlowReg*, *ANTs*, and *ANTsAffine* have the highest MI for the CAIN database. Standardization improves the ANTs Registration (Figure 4.26). For the ADNI dataset, the highest MI is seen by *FlowReg*, *ANTs*, and *SE* (Figure 4.27).

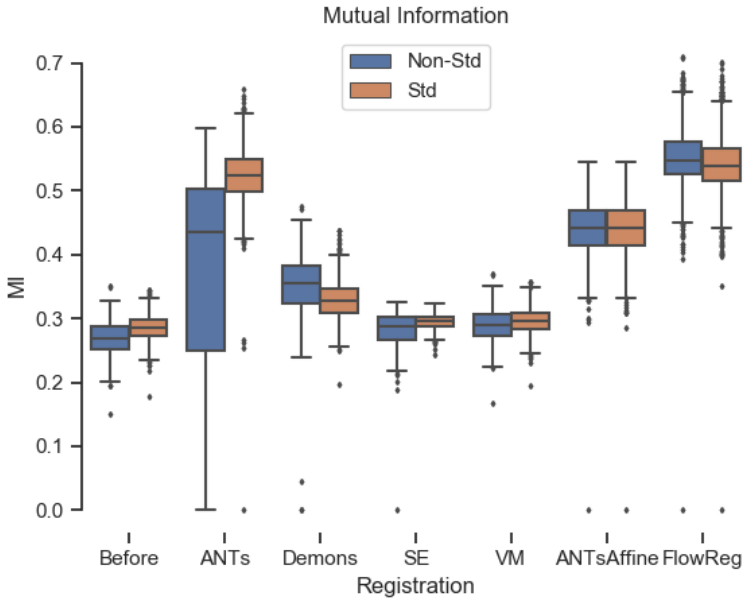


Figure 4.26: CAIN Mutual Information metric. Higher number indicate better correspondence between the probability distribution functions.

For the ADNI volumes, the results are seen in Figure 4.27. *FlowReg*, *ANTs*, *SE*, and Demons have the highest MI values. Standardization increases the results of all registration metrics.

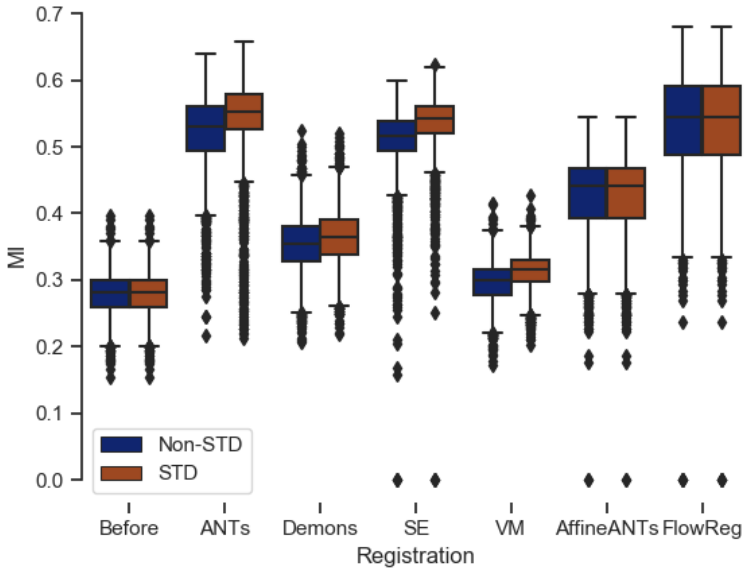


Figure 4.27: ADNI Mutual Information metric. Higher number indicate better correspondence between the probability distribution functions.

Correlation coefficients are found for both CAIN and ADNI and are seen in Figure 4.28 and 4.29. For CAIN, the highest correlation values are seen for *FlowReg* and *ANTs*. Standardization improves *ANTs* significantly, and significantly reduces the spread of *Demons* and *SE*.

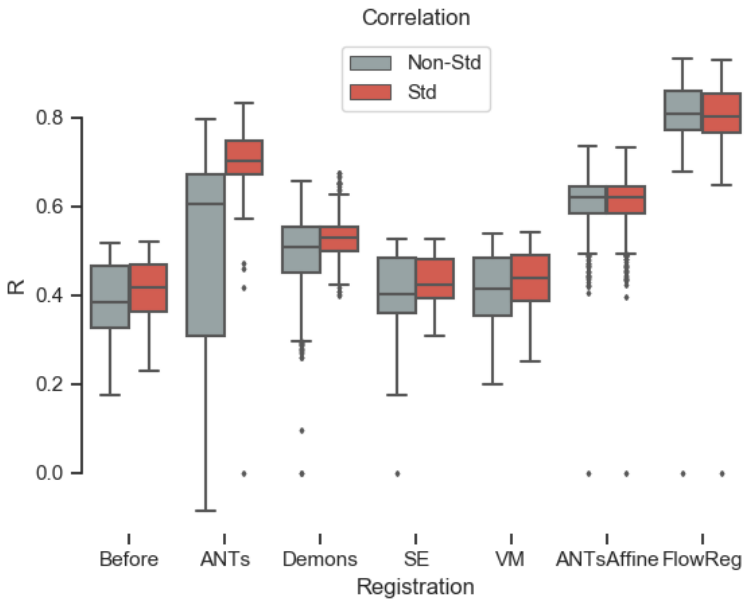


Figure 4.28: CAIN correlation coefficients are found by finding the Pearson's r.

Similar trends are observed in the ADNI database. The highest correlation are seen in *FlowReg* and *ANTs* and standardizing the volumes improves *Demons* and *SE* by reducing the spread significantly.

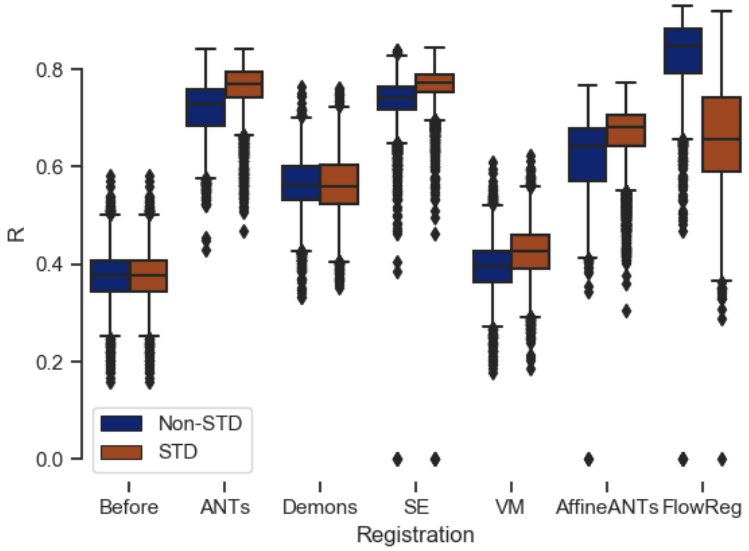
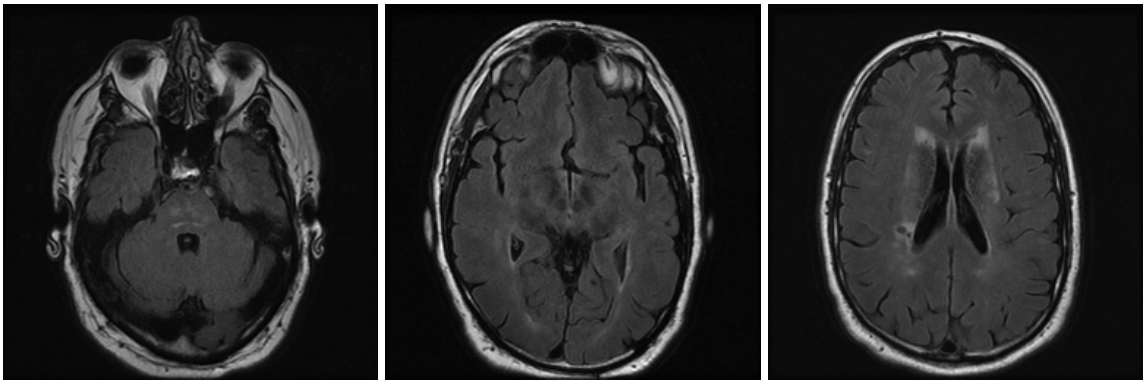
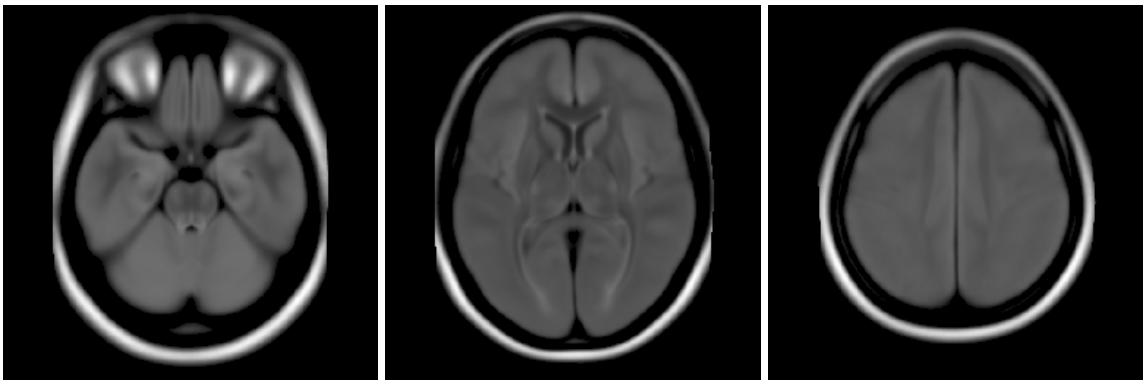


Figure 4.29: ADNI correlation coefficients are found by finding the Pearson's r.

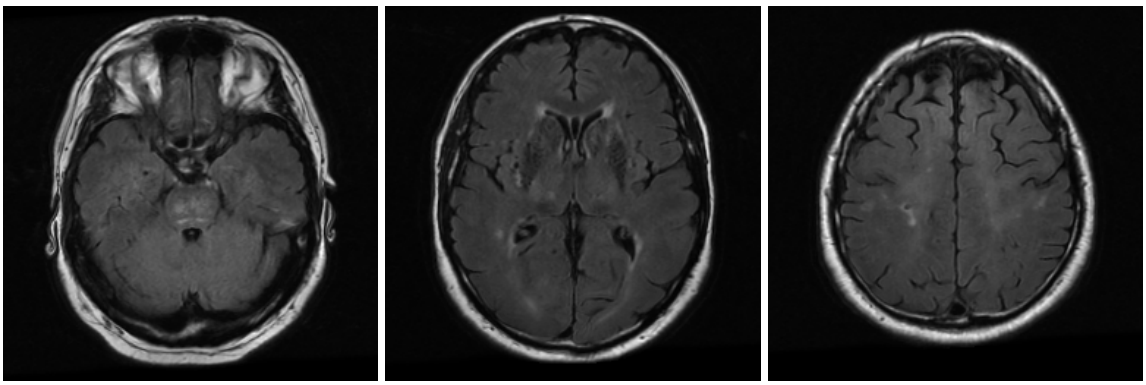
To provide a visual representation of image registration quality, one moving volume's registration journey is shown below as it was registered across all registration modalities (Figure 4.30, Figure 4.31, Figure 4.32). One should begin examining the different visual characteristics of the moving volume and the fixed volume. It is important to remember that for example looking at slice-wise comparisons of these two volumes in Figure 4.30 there are major structural differences between the moving and the fixed at slices 35. The most striking visual anomalies are seen in the *Demons* registration with *VM* being a close second. While both of these registration perform poorly, they are on the opposite extreme ends of the visual quality registration measures. *Demons* can be seen (Figure 4.31) that there are some major distortions especially in the shape of the skull, eye orbits, and brain tissue at the longitudinal fissure. *VM* shows that there was virtually no movement of the tissue and this can be seen by examining slice 35. *ANTs*, *SE*, and *FlowReg* are the most similar images to the fixed volume while still maintaining tissue integrity. The *FlowReg* registration effect can be observed comparing each registered slice to the *Affine* pre-registered images.



(a) Moving volume, slice 15. (b) Moving volume, slice 25. (c) Moving volume, slice 35.

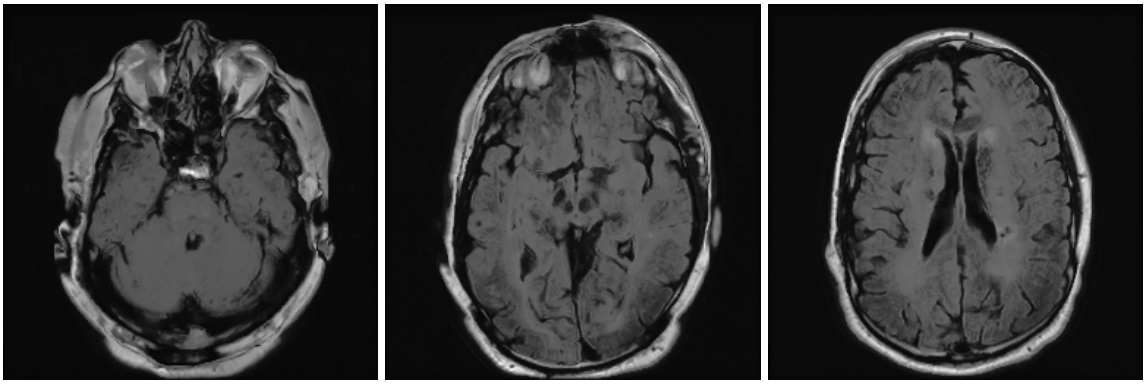


(d) Fixed volume, slice 15. (e) Fixed volume, slice 25. (f) Fixed volume, slice 35.

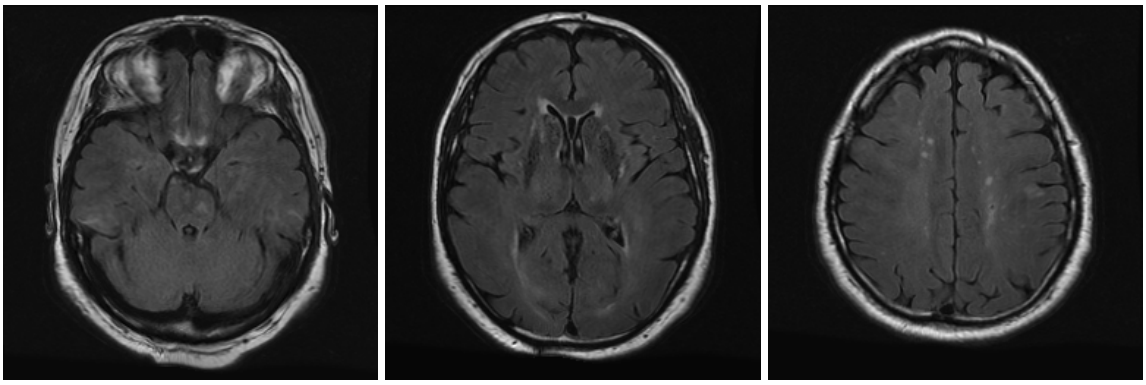


(g) ANTs registered, slice 15. (h) ANTs registered, slice 25. (i) ANTs registered, slice 35.

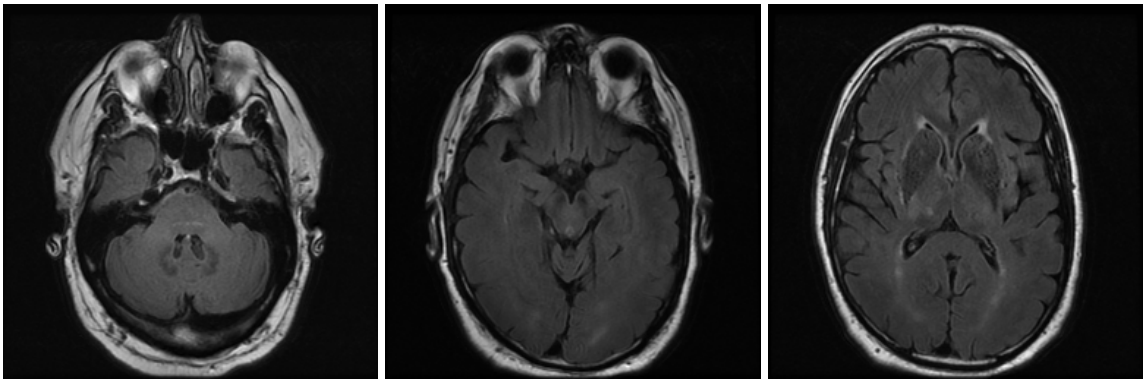
Figure 4.30: One moving volume example across all registration methods.



(a) Demons registered, slice 15. (b) Demons registered, slice 25. (c) Demons registered, slice 35.

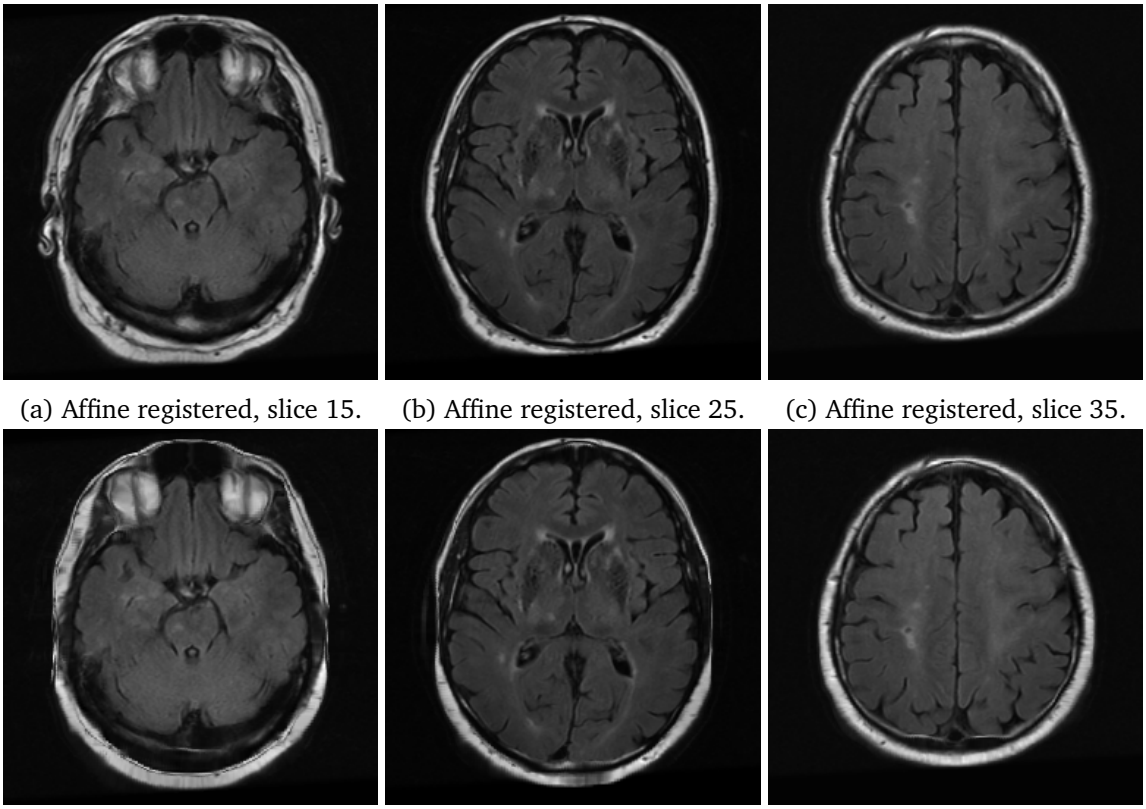


(d) SE registered, slice 15. (e) SE registered, slice 25. (f) SE registered, slice 35.



(g) VM registered, slice 15. (h) VM registered, slice 25. (i) VM registered, slice 35.

Figure 4.31: One moving volume example across all registration methods.



(a) Affine registered, slice 15. (b) Affine registered, slice 25. (c) Affine registered, slice 35.

(d) FlowReg registered, slice 15. (e) FlowReg registered, slice 25. (f) FlowReg registered, slice 35.

Figure 4.32: One moving volume example across all registration methods.

Chapter 5

Clinical Applications

5.1 Image Registration in Clinical Applications

At its essence, image registration is a method of aligning one moving image to the same geometric space as a fixed image. This is typically done via rigid and non-rigid transformations.

From a clinical perspective, the purpose of registration is associated with warping images so that they can be analyzed with software tools and eliminate some of the variations that are associated with the image acquisition process also known as the Multi-Center Effect [67]. As different centers and scanner manufacturers have settings that can be tuned during a procedure, there are others that are unchanged. For this reason, a patient may have several scans throughout his or her lifetime that have to be analyzed for disease progression and treatment efficacy. To get an accurate measurement it is of utmost importance to have the images be as similar as possible, even with all the variance in manufacturer and scanner model setting. Registration is one of the methods to obtain consistency by being able to warp these images so that they are in the same space and an objective and accurate diagnosis can be made. This reduction in variation before any type of analysis is performed, can aid in reduction of errors and misdiagnosis, provide more accurate treatment regimens, and reduce the number of repeat scans. For these reasons, registration in medical image processing is an important cogwheel in the machine of many healthcare systems throughout the world by reducing the patient wait times and overall economic burden.

5.2 Atlas Generation

A fourth measure of registration accuracy was done by synchronized averaging. All volumes were registered and synchronized voxel-wise averaging was performed to obtain an averaged representation of image alignment. This is particularly useful in observing rigid and affine transformations calculated between the two images. It is also representative of images that are outliers and have outright failed registration, which will form a ghosting phenomenon and thus lose sharpness.

The resulting volume is a novel brain atlas specific to the registered dataset and a single volume representation of the database contents (Figure 5.1). Brain atlases such as the Talaraich [72] and the Montreal Neuroimaging Institute (MNI) [73] are widely used template spaces to which many volumes are warped to prior to any segmentation approaches. The Talraich atlas, first introduced in 1988 is known as a coordinate system where a grid is proposed that denote particular structural locations. The MNI atlas was suggested in 1992 as a replacement to the Talaraich for use in minimal invasion surgery, namely stereotaxy [73]. The MNI space has templates for T1, T2, and Proton Density weighted (PDw) MRI. Currently, there are no widely used atlases for FLAIR MRI other than [68]. The *Brainder* template space has been co-registered to each other and then to the MNI space using the Statistical Parametric Mapping (SPM) [74] toolbox developed to work in Matlab [69] (Figure 5.2).

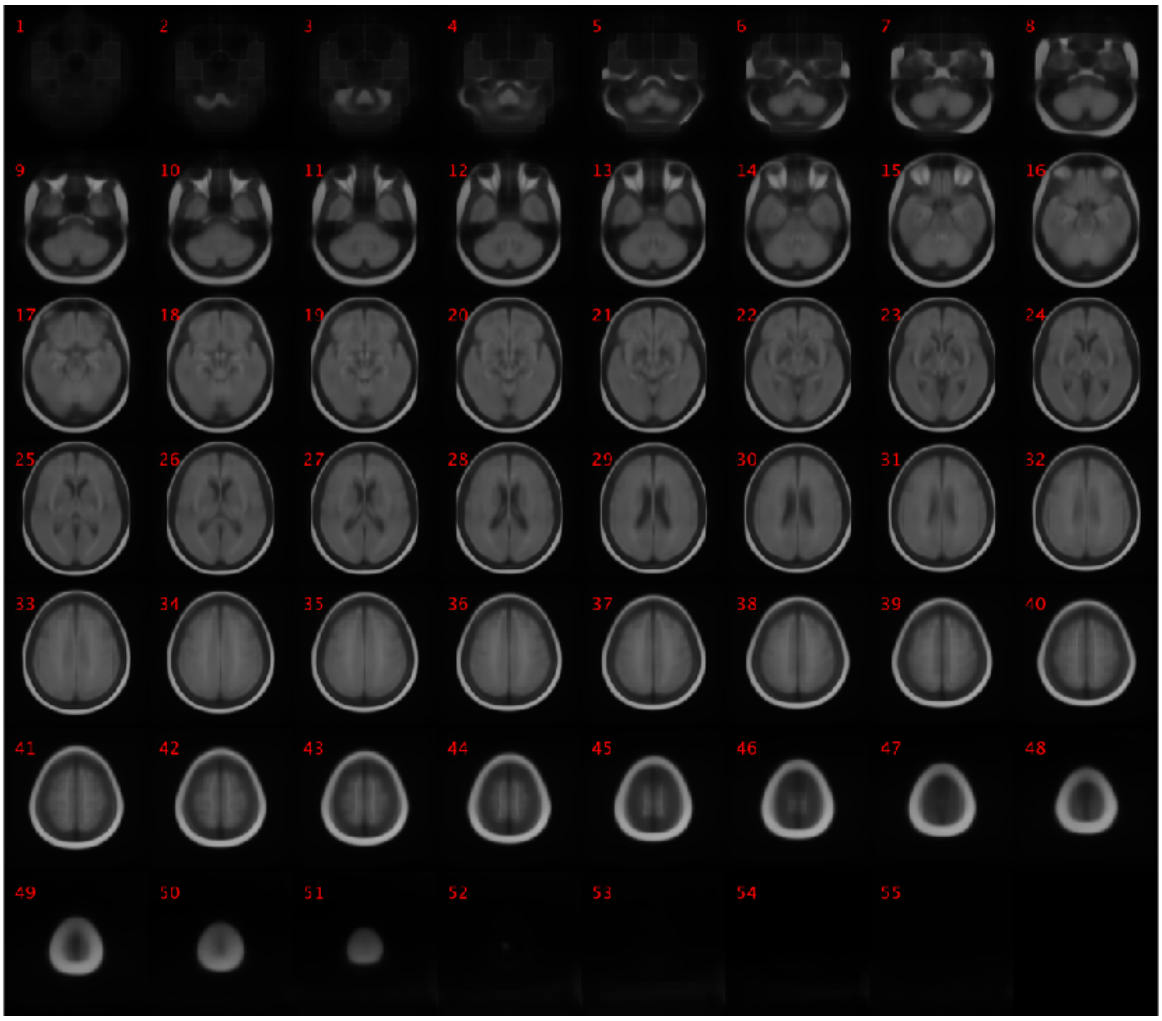


Figure 5.1: FlowReg generated atlas via synchronized averaging of registered volumes.

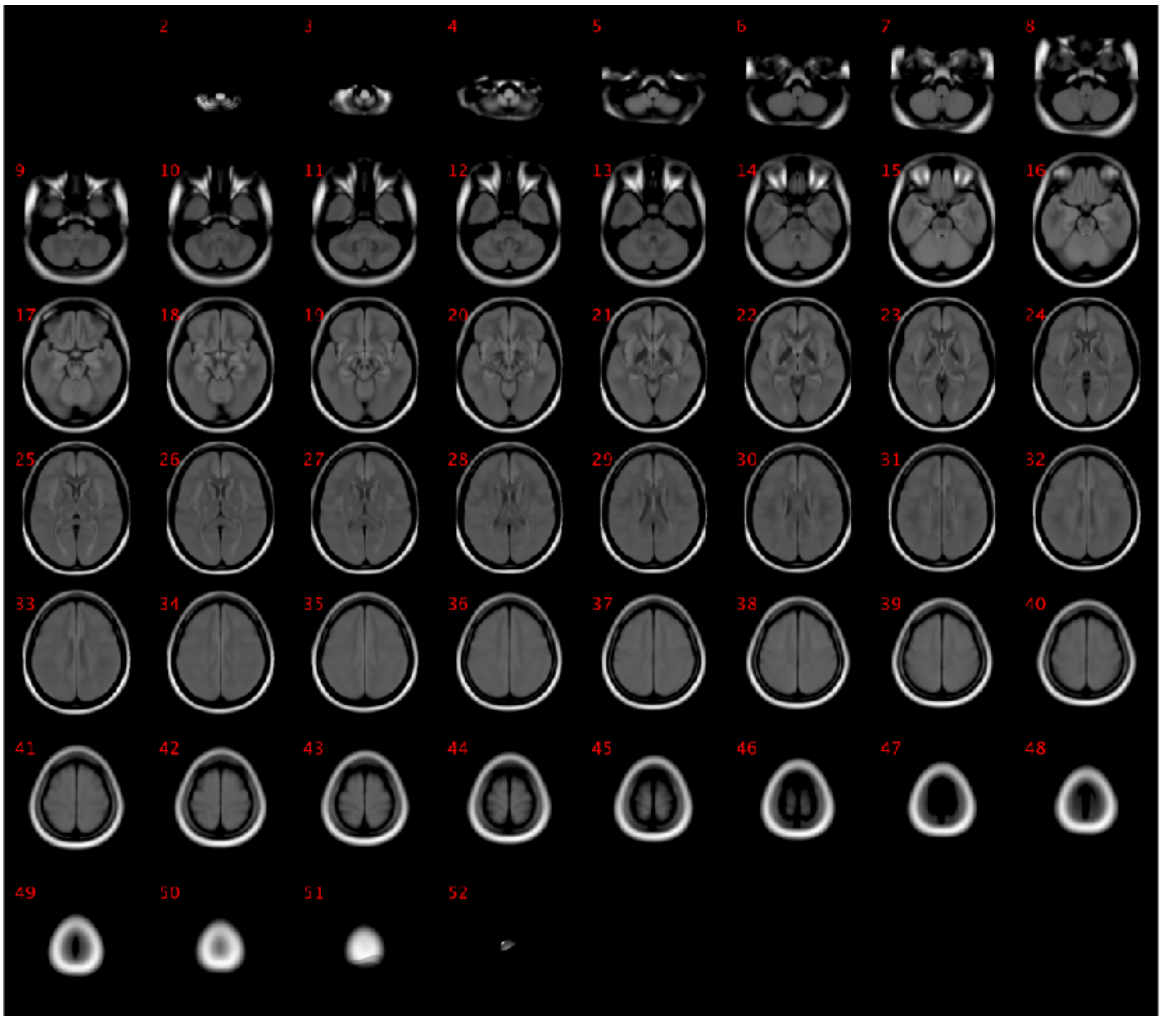


Figure 5.2: The FLAIR atlas provided by [68] *Brainder*.

5.2.1 Atlas Histograms

The synchronized averaging atlases were analyzed by examining the produced histograms and probability distribution functions (PDF) (Figures 5.3, 5.4). In FLAIR MRI, histogram peaks can typically represent significant and separable areas of the brain such as a brain matter and CSF.

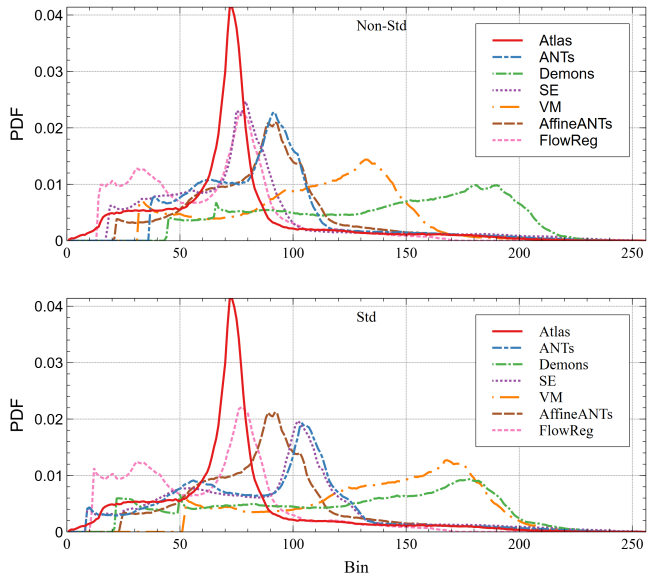


Figure 5.3: CAIN averaged registered volume atlas histograms.

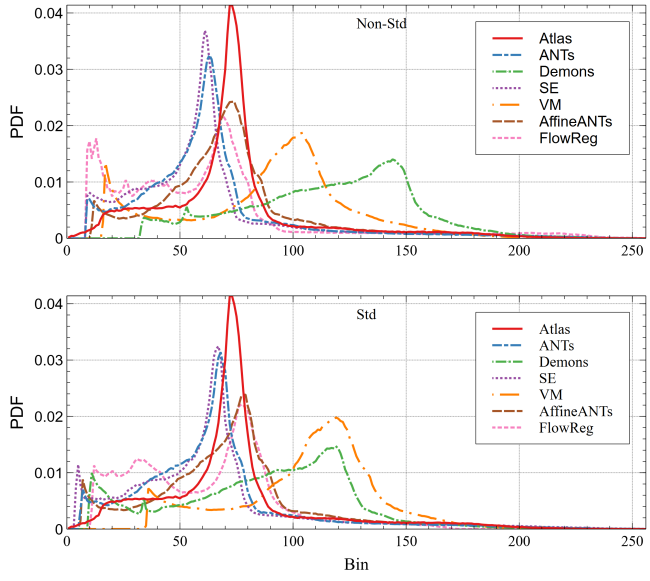


Figure 5.4: ADNI averaged registered volume atlas histograms.

The MSE was found between the *atlas* and calculated atlases for the respective registrations as seen in Figures 5.5 and 5.6. The lowest error and thus closest to the original atlas are observed by the obtained atlases via *ANTs*, *SE*, and *FlowReg*. These results signify that the overall registered data is similar to the original fixed atlas according to a low-reported error.

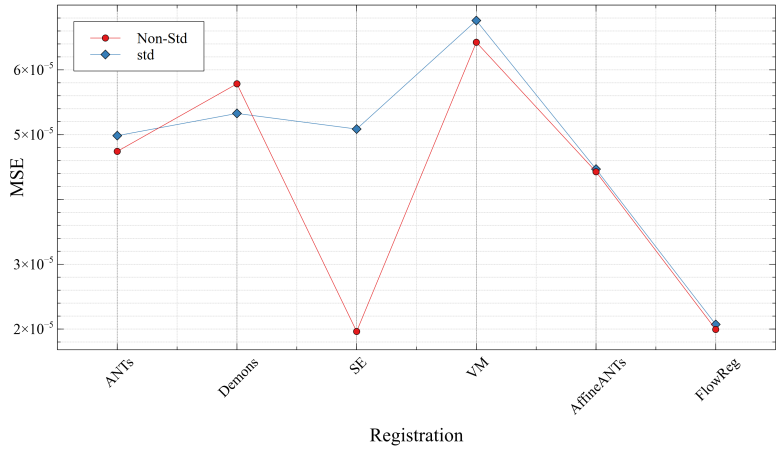


Figure 5.5: MSE between the registration-specific atlas PDFs and original atlas for the CAIN database.

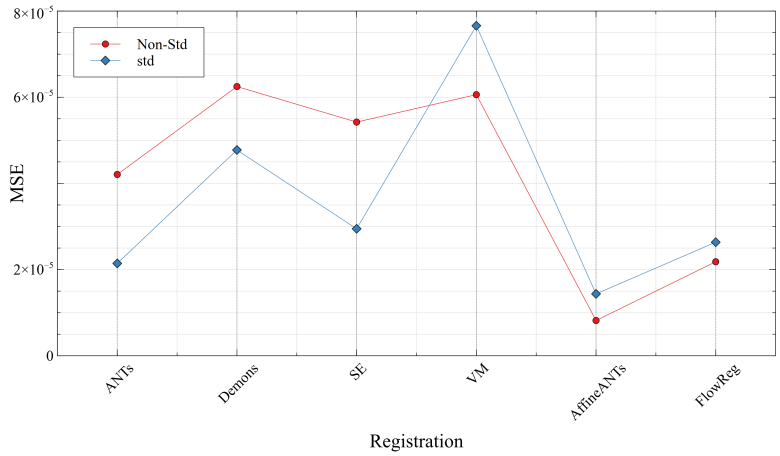
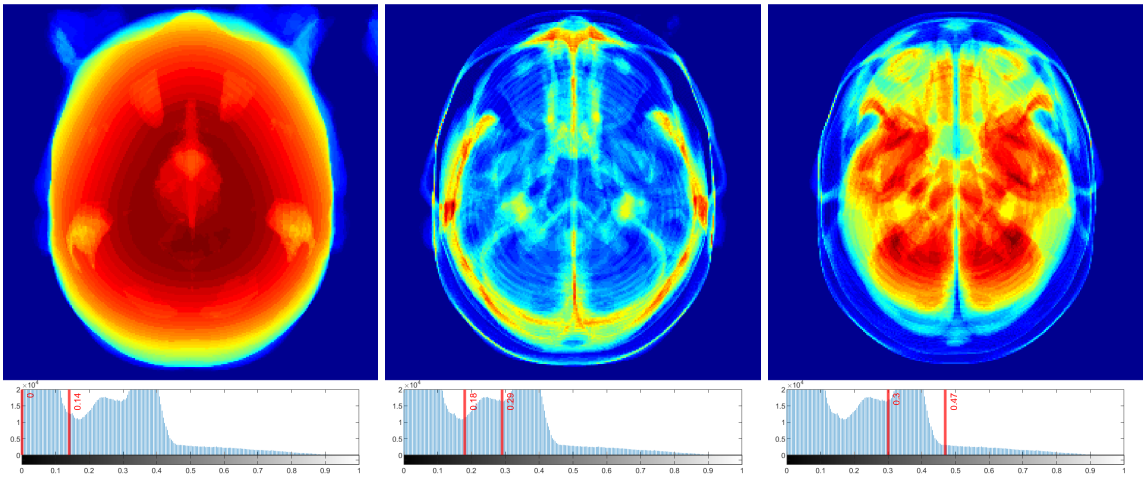


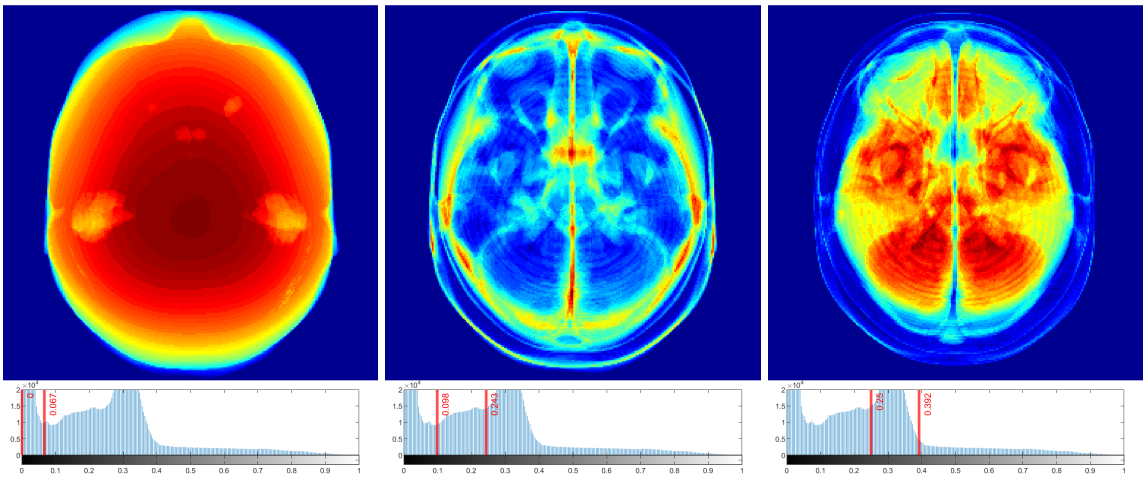
Figure 5.6: MSE between the registration-specific atlas PDFs and original atlas for the ADNI database.

Qualitative observation of averaged volumes was done by generating a heatmap of the several anatomical areas that can be separated via thresholding from the atlas histograms. The background was subtracted and all slices were summed to obtain the heatmaps observed in Figure 5.7. More examples of all registration methods across both databases both regular and standardized are found in the appendix A.

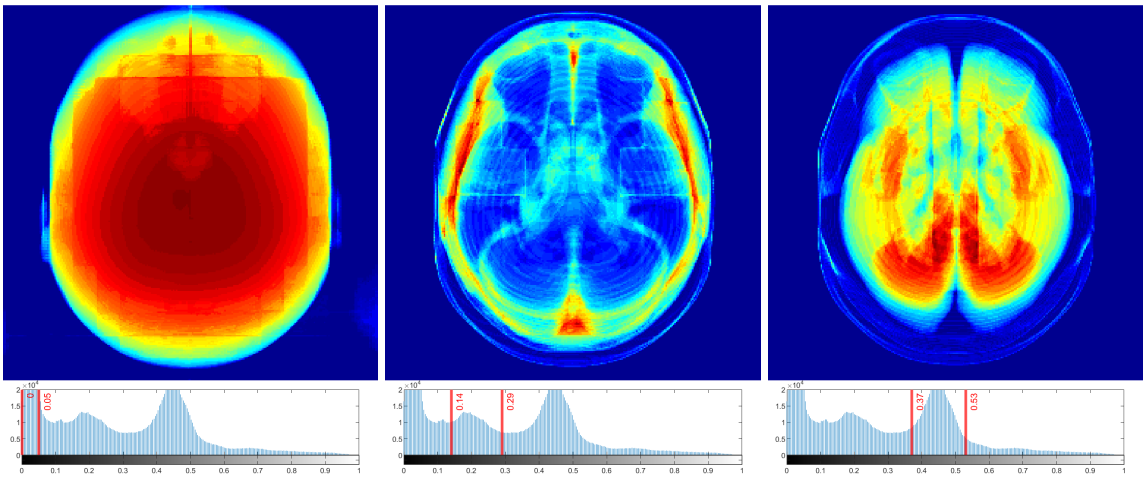
Similarly via thresholding techniques the two major separable areas that are denoted by the peaks in the atlas histograms can be separated into CSF and brain tissue as seen in the figures below. The reader's attention is brought to the intensity histogram below each image that shows the upper and lower bounds of each peak that can be used to separate the whole head, CSF, and brain tissue.



(a) ANTs head (b) ANTs CSF (c) ANTs brain



(d) SE head (e) SE CSF (f) SE brain



(g) FlowReg head (h) FlowReg CSF (i) FlowReg brain

Figure 5.7: Heatmaps showing the separability of tissue for the non-standardized CAIN data.

5.3 Longitudinal Registration

Longitudinal registration was performed on patients who had multiple scans. As mentioned previously, longitudinal scans are one of the main clinical applications for image registration. As images are placed in the same geometric space, quantifiable metrics can be extracted with respect to changes in brain volume, tissue loss, tumour evolution, and treatment efficacy. The coming sections will explore the experimental setup for longitudinal registration and present the results as a function of lesion change.

5.3.1 Setup

Longitudinal registration was performed on the CAIN database on 361 patients. Each of them had a *time-0* scan and then at least one other subsequent scans. 236 patients had just two scans and 103 patients had three scans. All scans following *time-0* were registered to *time-0*. To reiterate, if a patient had a scan at the present time, and then a second scan six months from now, then a third scan one year from now, the two subsequent scans were registered to the volume obtained today. This approach allows to have a timeline of different brain scans that can be compared to the baseline via metric extraction according to the time passed. These metrics can then be compared to each other without having to register volumes from *time-2* to *time-1* and still provide salient results of tissue change, but significantly reduce computation time.

5.3.2 Lesion Change

Lesion change has been chosen as a metric of longitudinal registration. It can be observed in Figure 5.8 the WML tissue change in longitudinally registered volumes. As time progresses the volume and overall area has increased. To more adequately observe the amount of change, a voxel-wise subtraction was performed (Figure 5.9).

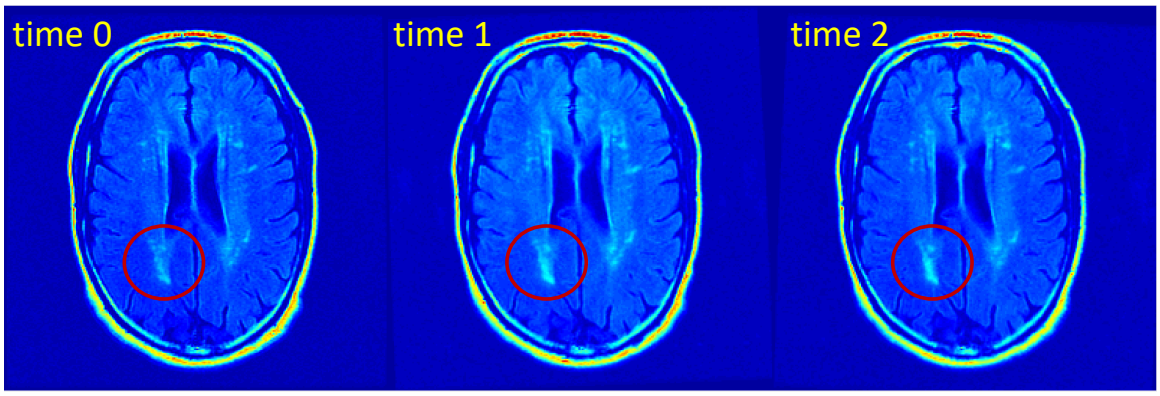


Figure 5.8: Longitudinally registered volumes. Time 1 was registered to time 0 and time 2 was also registered to time 0.

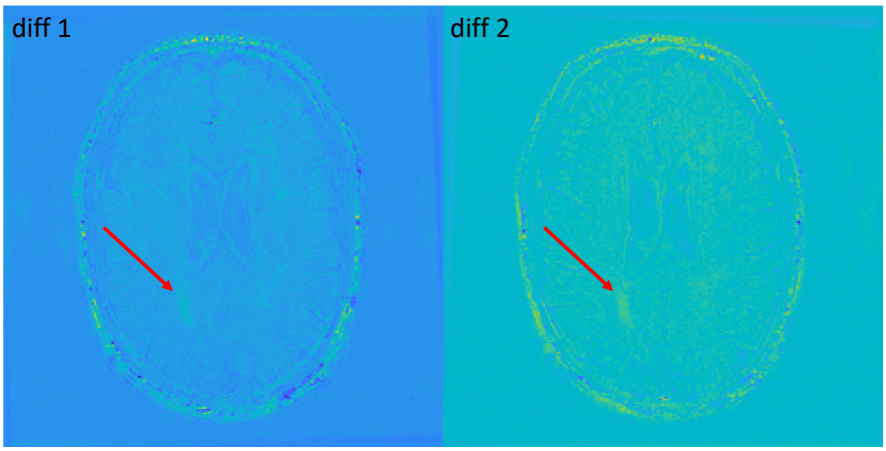


Figure 5.9: Differences between longitudinally registered volumes. Diff 1 = time1 - time0; diff 2 = time2 - time0.

The most significant difference in tissue change is seen in the largest time difference. Quantification of lesion change can be done via several methods, via manual or automatic segmentation. As automatic WML segmentation is a problem that is an active and ongoing research topic [75] a

combination of automatic and manual segmentation was adopted for this work. The total lesion volume was calculated prior to registration, after affine registration, and after *FlowReg* as seen in Figure 5.10.

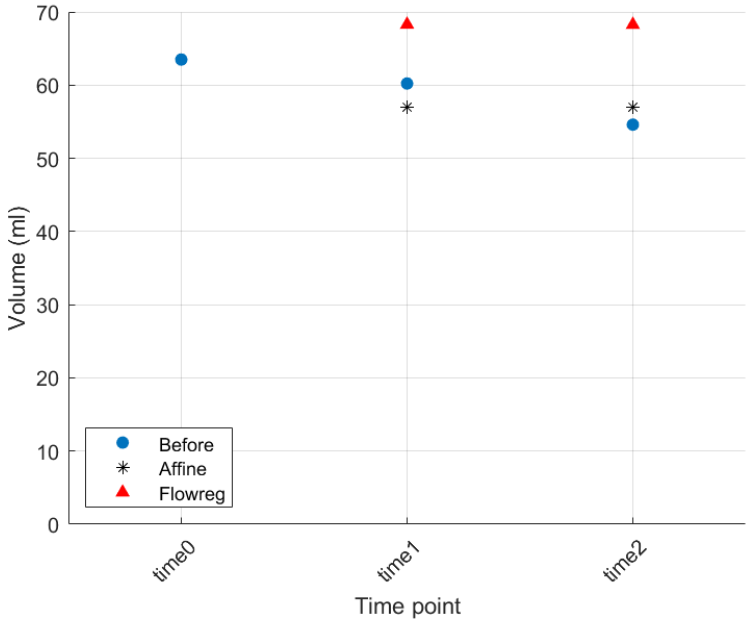
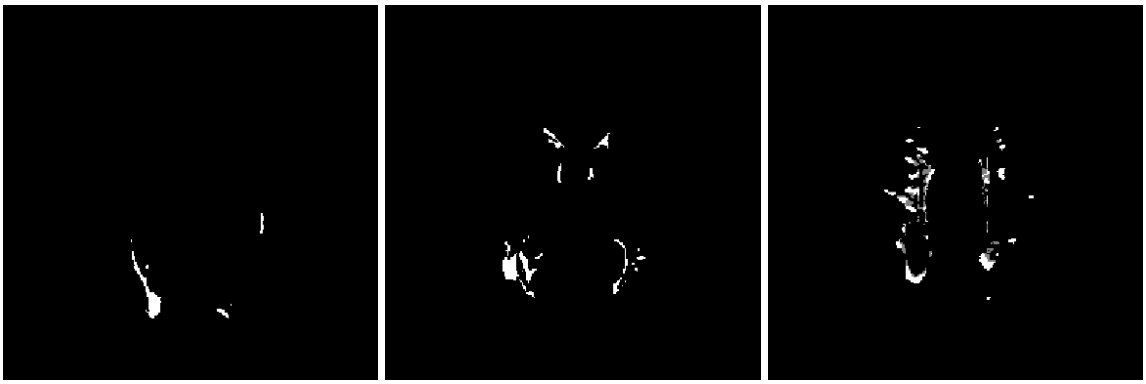
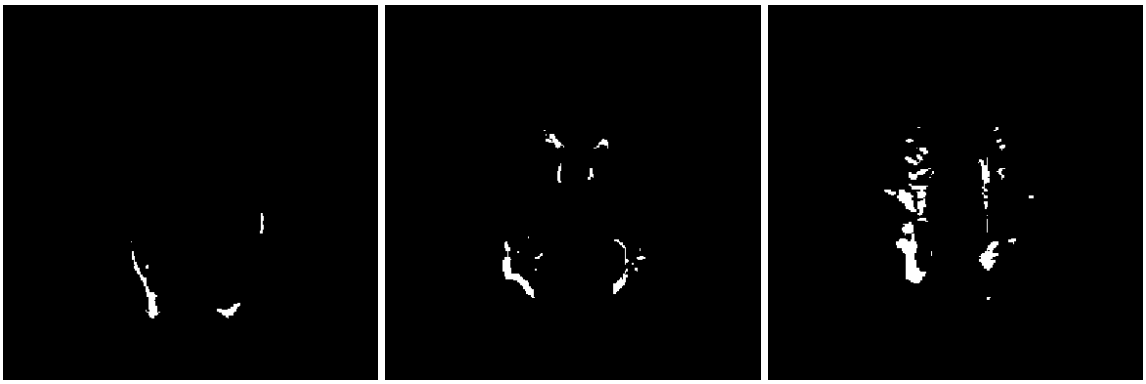


Figure 5.10: Longitudinal WML volume progression of one patient.

The overall trend suggest that this patient's WML volume is decreasing with subsequent scans based on the obtained manual segmentations. One curious aspect to analyze is examining *FlowReg* it seems the overall volume does not change from time point 1 to time point 2. This could be due to lesions being present in some places but not in others and due to issue during image binarization. It is important to also remember that the training procedure did not include any longitudinal training. The fixed volume was the atlas provided by *Brainder*. Thus, the fixed volume features learned by the network during atlas registration cannot be transferred to longitudinal registration where the fixed volume is the time point 0. To observe the actual effect one may subtract the registered time point 1 from the original fixed volume at time point 0. The same can be done for time point 2 with respect to time point 0. Three slices of each difference are shown below prior to registration (Figure 5.11).



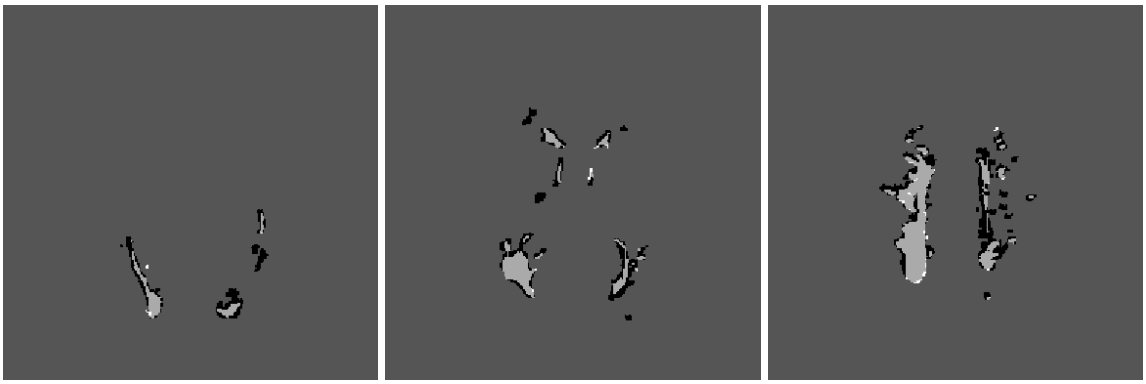
(a) Slice 21. (b) Slice 26. (c) Slice 33



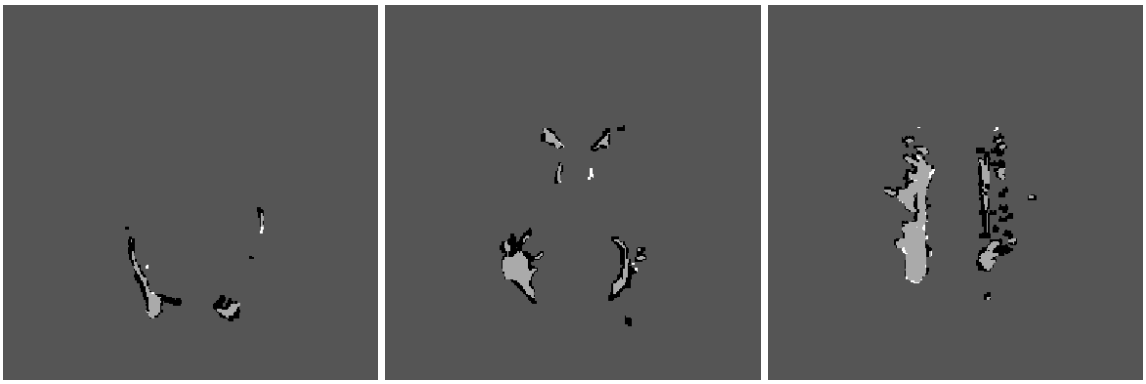
(d) Slice 21. (e) Slice 26. (f) Slice 33

Figure 5.11: Difference of lesions prior to registration. The lesion at subsequent time points was subtracted from time 0. (Ex: $time0 - time1$.)

It can be noted that prior to registration, because there is no correspondence between the volumes, there are large difference at almost every slice level. It is for this reason that the images shown in Figure 5.11 appear as binary mask. The subtraction shows that at one time point the lesions are not lined up and thus the entire lesion is beign subtracted from an area where there may no other lesion. After registration, one can now notice more variation in lesion change 5.12. Since the lesion are now in correspondence with one another, it is apparent that lesions vary. There are areas where the lesion has decreased in size and other areas where it has increased in size.



(a) Slice 21. (b) Slice 26. (c) Slice 33



(d) Slice 21. (e) Slice 26. (f) Slice 33

Figure 5.12: Difference of lesions after *FlowReg* registration. The lesion at subsequent time points was subtracted from time 0. (Ex: $time0 - time1$.)

In the figure above, the dark areas represent areas where the lesion has decreased in size, while bright areas where it has increased in size. The gray area is a net-zero change and can be matched with the background for a qualitative assessment of lesion change.

Chapter 6

Conclusion

The concluding portion of this thesis describes the primary and secondary contributions. Potential clinical applications of the proposed validation metrics and optical flow registration model are outlined. Finally, the limitations of the proposed methods along with possible future improvements on the design of the metrics and registration model are suggested.

6.1 Summary of Contributions

The work proposed in this thesis is related to proposing an image registration method that has been adapted from natural images using optical flow. An unsupervised model that is able to generate a deformation field for a moving image to the fixed image space was designed using deep learning convolutional neural networks with custom loss functions altered to suit medical images. To assess the accuracy of image registration, validation metrics were proposed. The validation metrics are based on evaluating volumetric and structural changes in tissue morphology. A third metric, measured orientation with respect to the saggital plane. Intensity-based and pixel alignment measures explored the amount of correspondence between the warped moving image and the fixed. Global alignment was measured via synchronized averaging and generation of an atlas. Longitudinal intrapatient registration was performed and lesion change was observed via a pixel-wise subtraction.

Chapter 1 presented an introduction to WMLs in the context of neurodegenerative disorders. The

concepts of MRI, in particular related to FLAIR MRI were inspected from a physics perspective and the advantage of imaging neurological tissue were discussed. The importance of image registration in the medical imaging workflow pipeline was analyzed from a clinical and research viewpoint. The objective of the thesis with respect to validation of image registration algorithms were examined and problems with existing measures were outlined. Finally the major and secondary contributions of the thesis were described.

Image registration algorithms were analyzed in Chapter 2. The chapter began with a gentle introduction to image transforms with the use of a transformation matrix. Rigid, non-rigid, and affine transforms were described. Image interpolation techniques were introduced to overcome the non-integer pixel coordinates after a transformation has been applied. Nearest-neighbour, bilinear interpolation, and bicubic interpolation techniques were described. Image registration processes were examined in detail from the most basic methods of aligning images to iterative processes that is inherent to most registration algorithms. Four registration algorithms were examined in detail, Thirion's Demons, Symmetric Normalization by ANTs, Elastix, and finally Voxelmorph.

Chapter 3 described the main contribution by describing optical flow registration methods and the validation metrics for medical image registration. The chapter examined optical flow as a means of warping images via a deformation field generated from subsequent video frames. The basics of optical flow generation were described via the Lucas-Kanade algorithm and ending with modern fast optical flow generation via deep learning methods such as FlowNet. The *FlowReg* model architecture is described along with the proposed loss functions. The registration metrics were described and how they analyze the quality of image registration from a volumetric, structural, orientation, pixelwise alignment, and intensity standpoint. Volumetry changes of brain tissue was examined by finding the proportional volume change of structures after registration. To observe structural changes the closest distance from the surface of the ventricles to the surface of the brain was observed and how it changes after image registration. Saggital orientation was measured as the Head Angle metric. Overall alignment was measured via synchronized averaging of generating atlases and pixel-wise alignment. MI and CC were calculated for obtaining intensity measures of image registration quality.

The experimental setup and results are described in Chapter 4. The chapter begins with a description of the data used from a demographics and acquisition perspective. Registration metrics were obtained via a pipeline described in Figure 4.1. For the four tested registration algorithms the relevant metrics were detailed. The optical flow model structure was described along with the training procedure. The final portion of the chapter describes the results of all four metrics for the four registration metrics used on the two datasets. The ventricular and WML PV change show that the best performing algorithms do not alter both structures after registration. Similar performance was seen for the SSD metric. For both metrics the best performing algorithms are Demons and Ants. For the large scale measures, HA, MI, and CC the top performing are ANTs and SE with the lowest spread ς, and highest MI and CC values.

Chapter 5 examines the clinical applications of medical image registration. Particularly, the newly generated atlases provide a true representation of a FLAIR-FLAIR atlas. These atlases were made with the intent to be made public for use in future FLAIR registration projects. Heatmaps of The generated atlases were then analyzed using Mean Error to the original target atlas. Longitudinal intrapatient registration was assessed by visual inspection of lesion change.

6.2 Applications

The proposed optical flow registration process is particularly useful because it is based on deep learning conventions. The major advantage of this approach over conventional registration algorithms is that the deformation fields have already been learned and the model is trained to predict the deformation field in one single iteration. The model This significantly reduces the computation time and can thus perform large-scale registrations at a rate of one volume per 10 seconds as opposed to several minutes that typical registration algorithms require. Further, as the model is deep learning based, further improvements can be added by retraining and improving the accuracy of the model. A naive analogy states that once a person has seen a number of pairs of shoes enough times, when given a large cluster of unpaired images, they would be able to find the correct pairs quickly based on features.

From a clinical perspective, the assessment of image registration has several major advantages were highlighted. As more and more image processing algorithms are developed in medical imaging, there is a need for constancy of the data used. Image registration is one method of normalizing data in a geometric way. The transformations used can place the images in the same space so that algorithms and analysis can extract salient information. An expert can make more informed decisions in regards to a patient's scan if the images are in the same geometric space and there are as few variations as possible. Tumour evolution and treatment efficacy can be assessed more accurately with image registration. Similarly image registration is an essential tool in most automatic WML segmentation algorithms. For this reason it is important to assess whether the registration is as accurate as possible and if it may adversely affect these hyperintense regions of interest. The validation metrics offer a measurement to see whether the volume or structure has changed after registration with respect to the whole brain. Any undesired effects can then be accounted for when designing any subsequent segmentation tools that use registration as a geometric normalization factor.

Similarly, the HA measurement can be used as a metric at the image acquisition time to ensure that a scan is as accurate and in line with all other scans taken. The orientation metric can also reaffirm the point that images have been correctly aligned during the affine registration process. Pixel-wise averaging is especially useful because ultimately new atlases are produced through these image registration processes. The atlases are database specific and can be used in the future for further registration to a more accurate and representative atlas.

6.3 Pitfalls and Limitations

The optical flow deformation field estimation has been performed on a per slice basis. The reason for this choice has been because of the large number of over 3 million feature parameters. As such, the deformation fields predicted by the optical flow algorithm are limited to two dimensions. This may create some unwanted deformations at the extremities of the brain. For example, if the moving image has a slice close to the bottom, the nose and orbits of the eye, but the fixed image is mostly empty, the deformation predicted may be of poor quality. The same can be said about the top

portions where little of the skull and brain are represented in both the moving and fixed images. A further limitation is the training procedure. Convolutional neural networks require an extensive training regimen for the model to learn meaningful features. Powerful computers equipped with modern high-RAM GPUs are required. The time to train is significant from several hours to multiple days for some models. This is not only computationally inefficient, but environmentally unfavourable due to high power consumption of said GPUs [76]. The result, or the weights learned by the model can be saved and then used in further experiments which are not as computationally intensive, however the path to reach that stage typically requires multiple iterations of the same model, testing of varying epoch lengths, and loss functions, all of which will have a significant impact on power consumption. One method to overcome this, is to used pre-trained networks and through transfer learning researchers can build on existing weights to obtain desired results in novel applications such as medical imaging [77].

Image registration algorithms are inherently a computationally intensive process and to analyze such complex algorithms to ensure that structural and volumetric cohesiveness is maintained, the metrics themselves should be as robust and easy to use as possible. One limitation of the proposed metrics is their inherent dependability on manual structural segmentations. Volumetric and structural changes rely on warping binary masks with the calculated deformation field, thus obtaining accurate and consistent annotations of relevant structures is of the utmost importance.

Further, the orientation measure, HA, is designed to measure affine transforms. Thus for images at which the affine transform has been the most accurate the measure of HA may be high, but it is unknown whether the registration itself is accurate on the pixel-wise warping when the deformation field is been used. HA is calculated with respect to the vertical saggital plane, however it is possible to also calculate the rotation in the two other planes for a more robust three-dimensional representation. A counter-argument could be proposed that rotation in the transverse or coronal plane are less probable due to the anatomy of the body. In a laying down, supine resting state, it is more likely to slightly rotate the head left or right in the transverse plane as opposed to front or back in the saggital plane, or tilting left or right in the coronal plane.

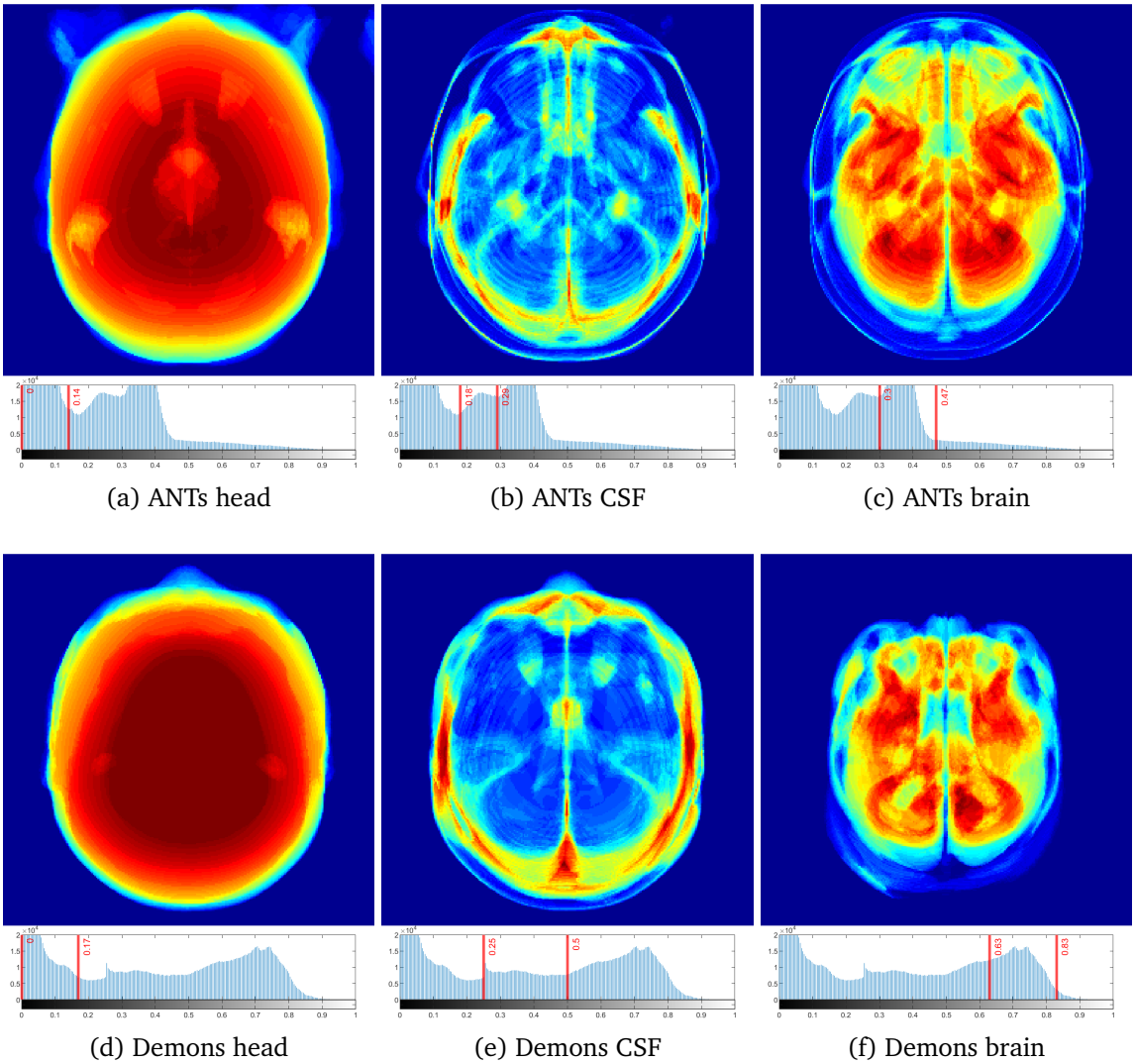
6.4 Future Works

The work presented in this paper is far from complete when it comes to proposing novel image registration methods and measuring the quality of registrations in FLAIR MRI. Some of the biggest improvements can be made in the areas of image registration speed and computational efficiency. Through experiments of different model architectures, loss functions, and data aughmentation, an efficient and fast performing FLAIR-specific registration method can be achieved in 3D.

Appendices

Chapter A

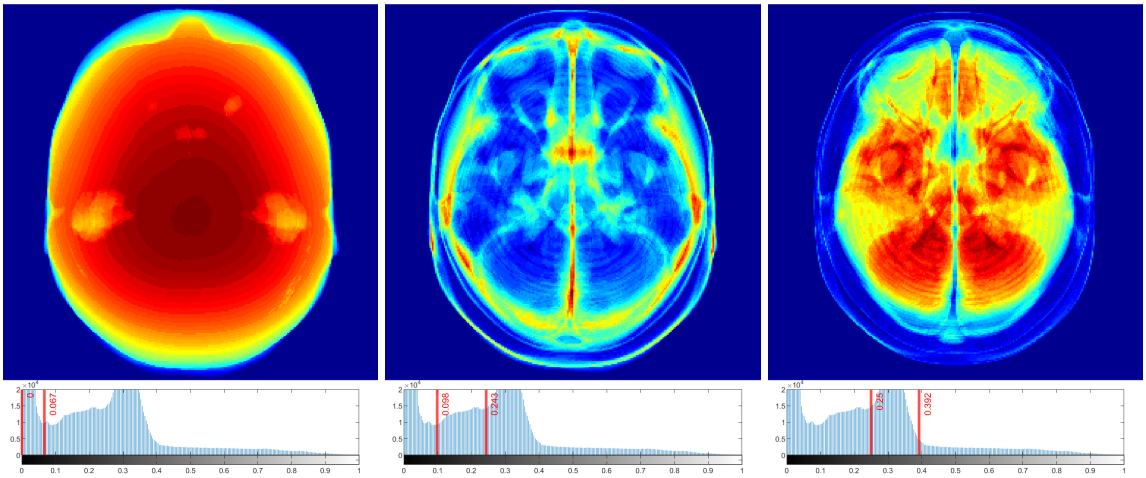
Tissue Heatmaps



(a) ANTs head (b) ANTs CSF (c) ANTs brain

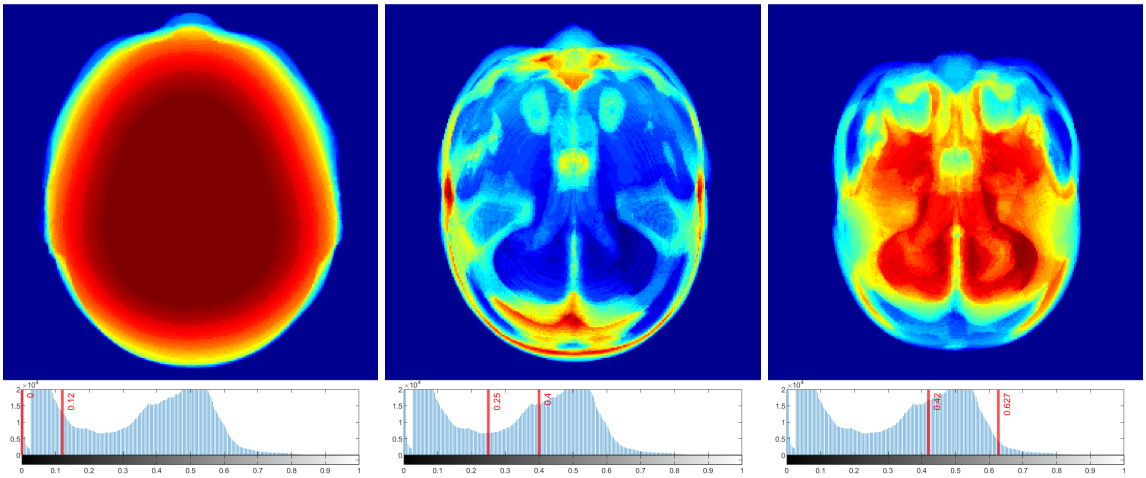
(d) Demons head (e) Demons CSF (f) Demons brain

Figure A.1: Heatmaps showing the separability of tissue for the non-standardized CAIN data.

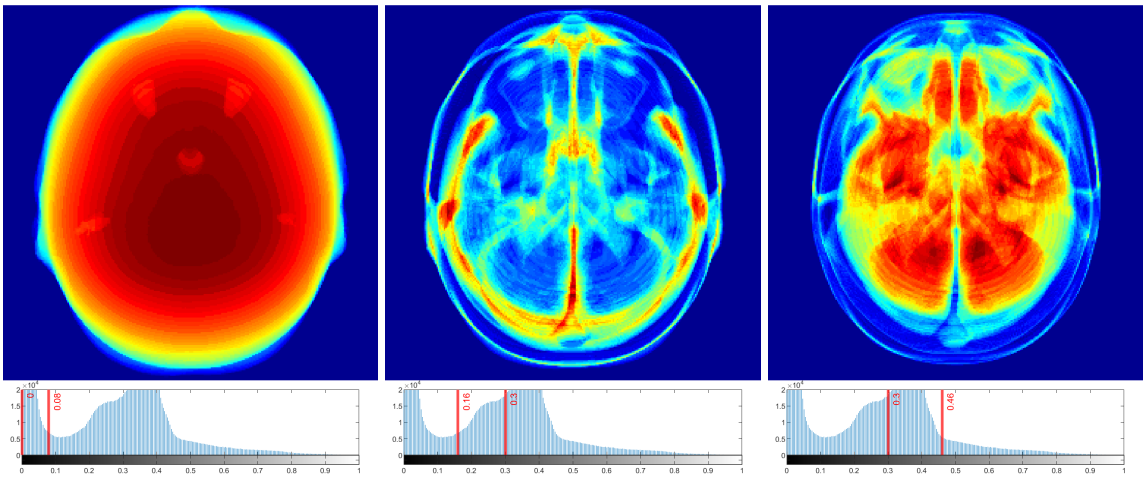


(a) SE head (b) SE CSF (c) SE brain

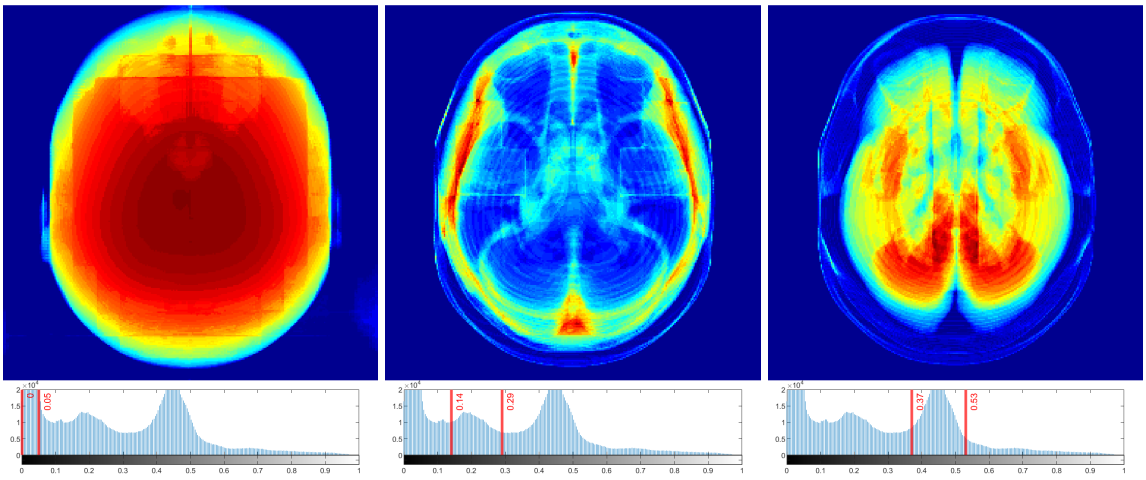
Figure A.2: Heatmaps showing the separability of tissue for the non-standardized CAIN data.



(a) VM head (b) VM CSF (c) VM brain

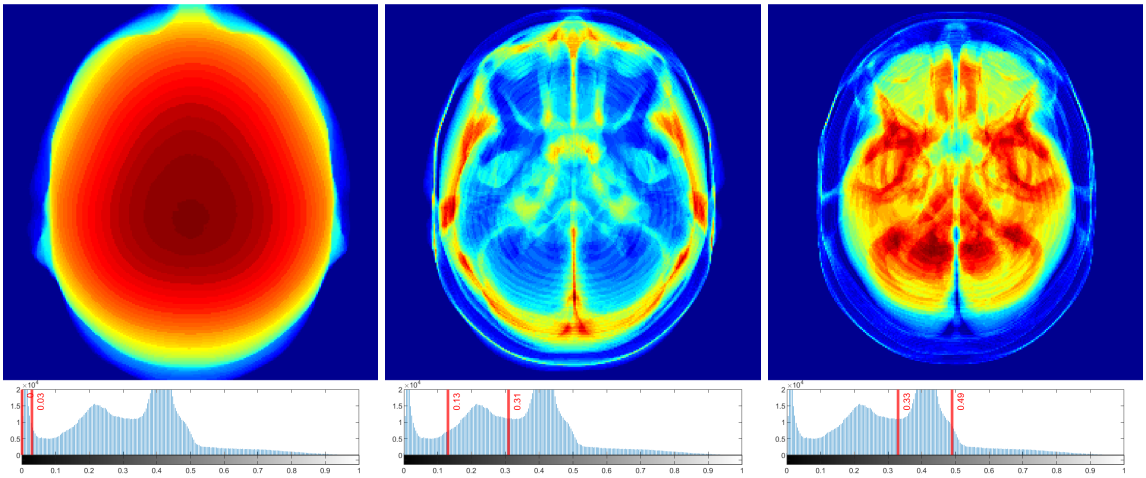


(d) Affine head (e) Affine CSF (f) Affine brain

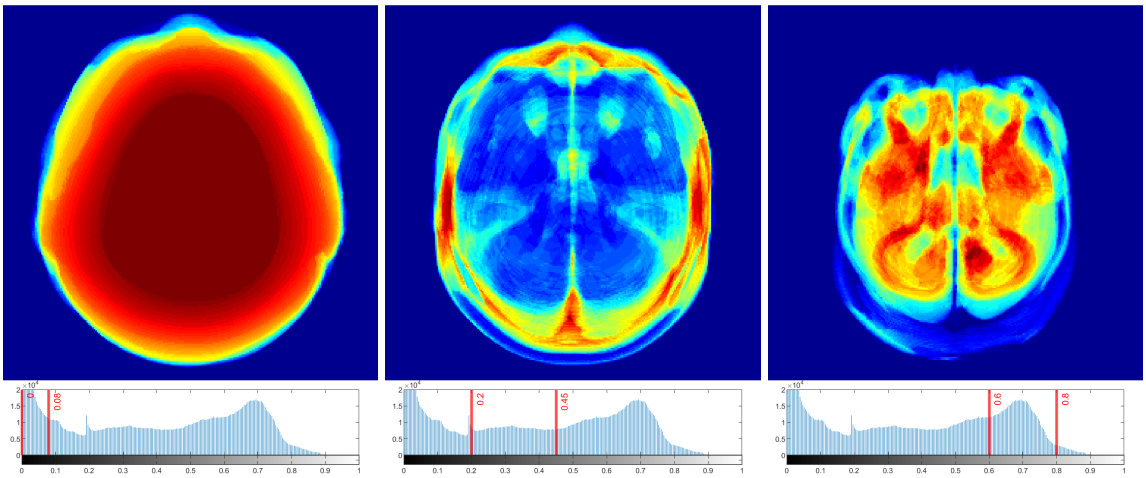


(g) FlowReg head (h) FlowReg CSF (i) FlowReg brain

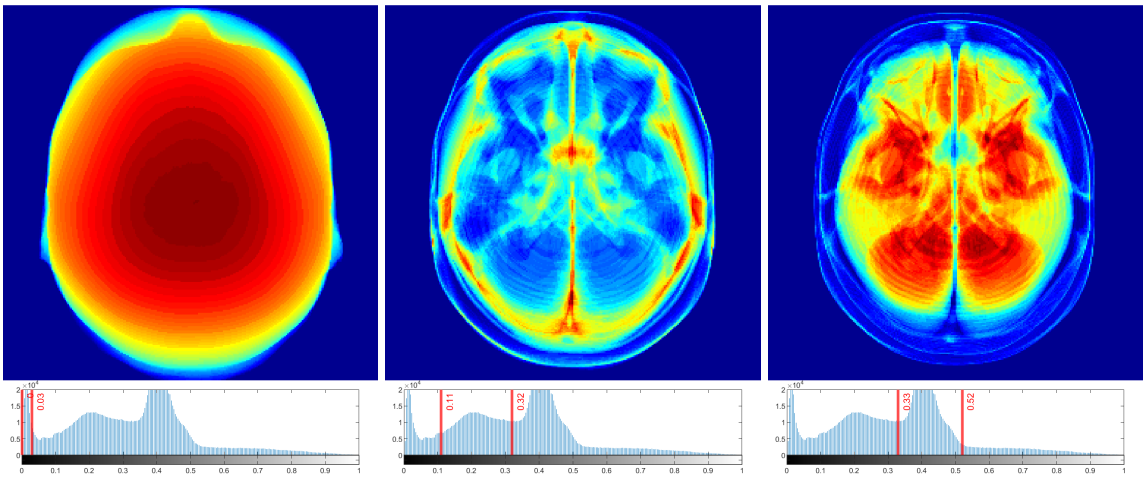
Figure A.3: Heatmaps showing the separability of tissue for the non-standardized CAIN data.



(a) ANTs head (b) ANTs CSF (c) ANTs brain

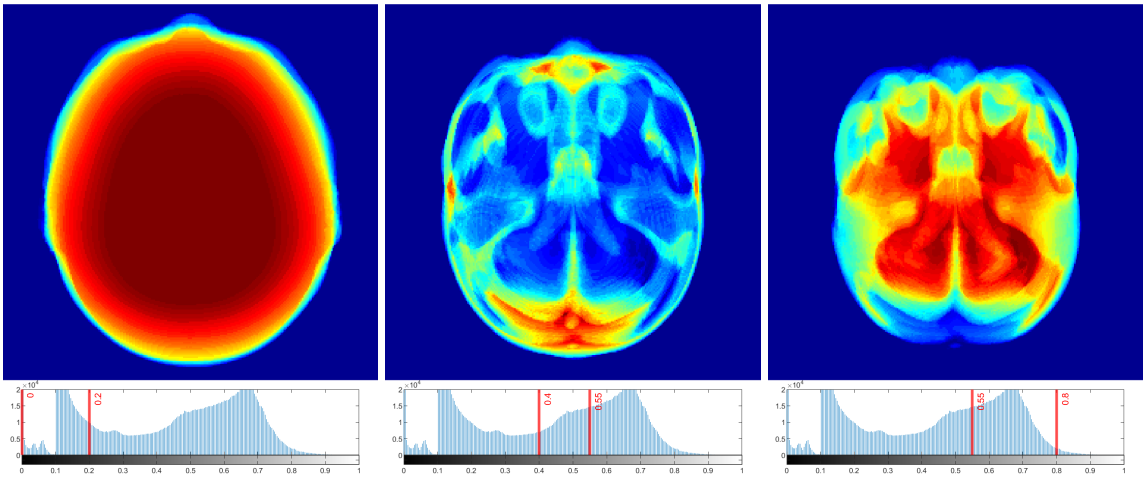


(d) Demons head (e) Demons CSF (f) Demons brain

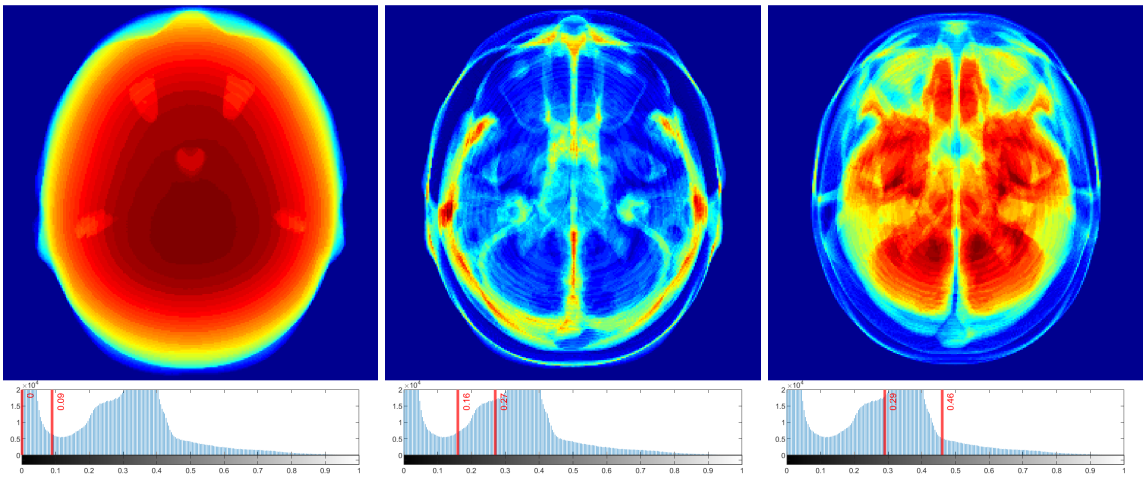


(g) SE head (h) SE CSF (i) SE brain

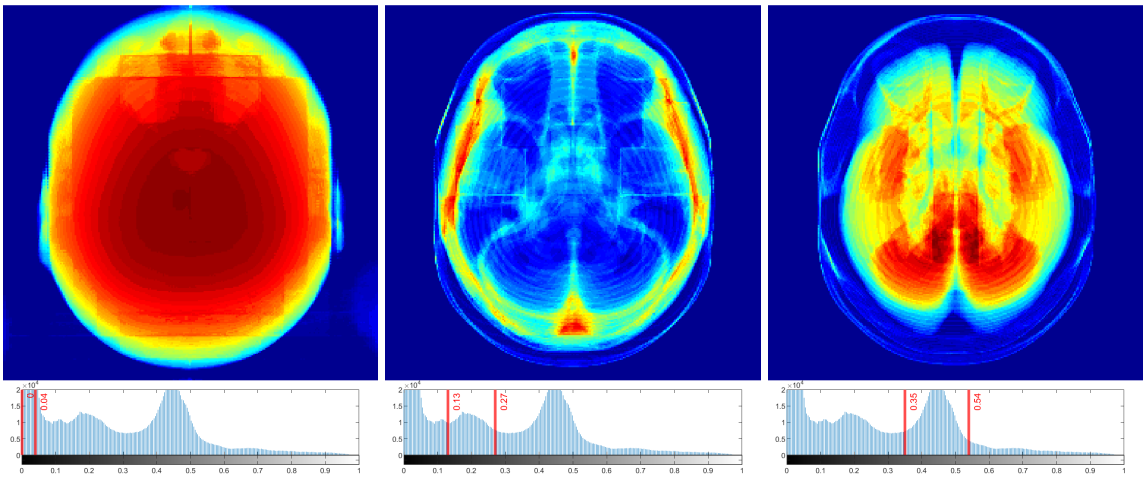
Figure A.4: Heatmaps showing the separability of tissue for the standardized CAIN data.



(a) VM head (b) VM CSF (c) VM brain

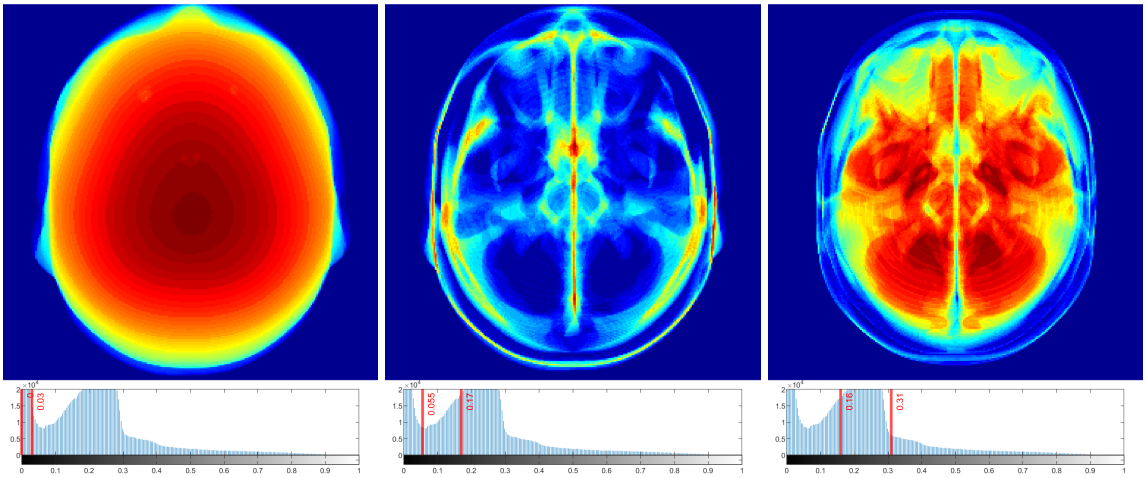


(d) Affine head (e) Affine CSF (f) Affine brain

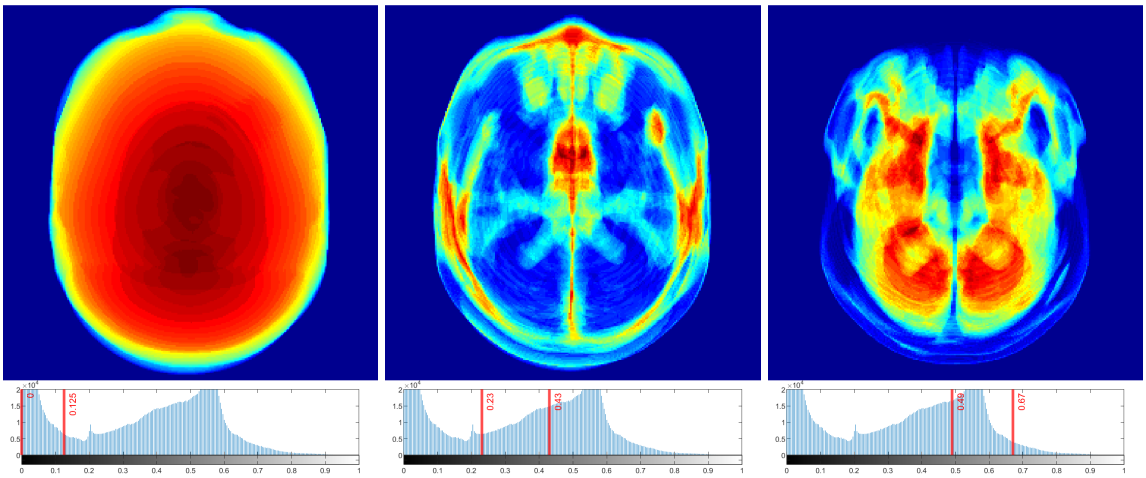


(g) FlowReg head (h) FlowReg CSF (i) FlowReg brain

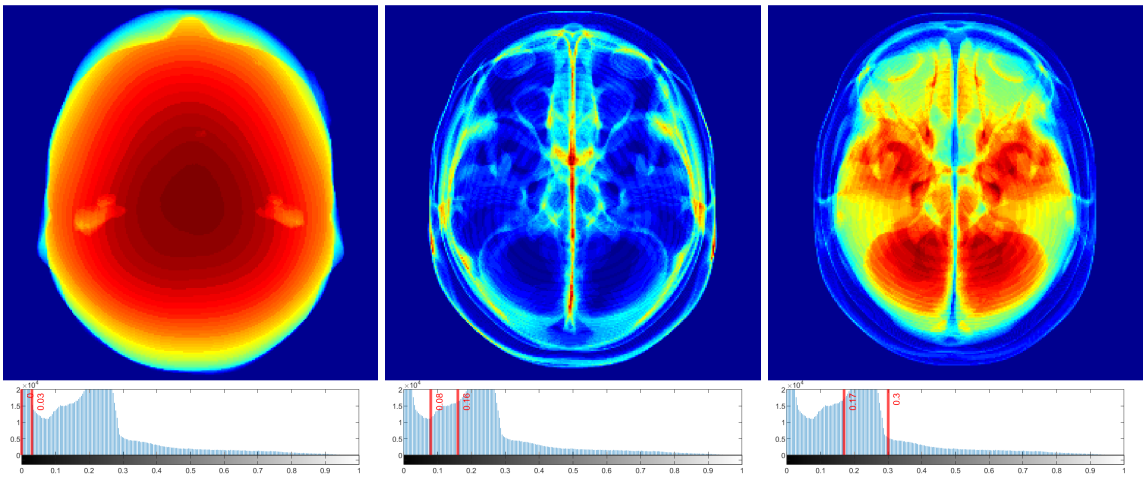
Figure A.5: Heatmaps showing the separability of tissue for the standardized CAIN data.



(a) ANTs head (b) ANTs CSF (c) ANTs brain

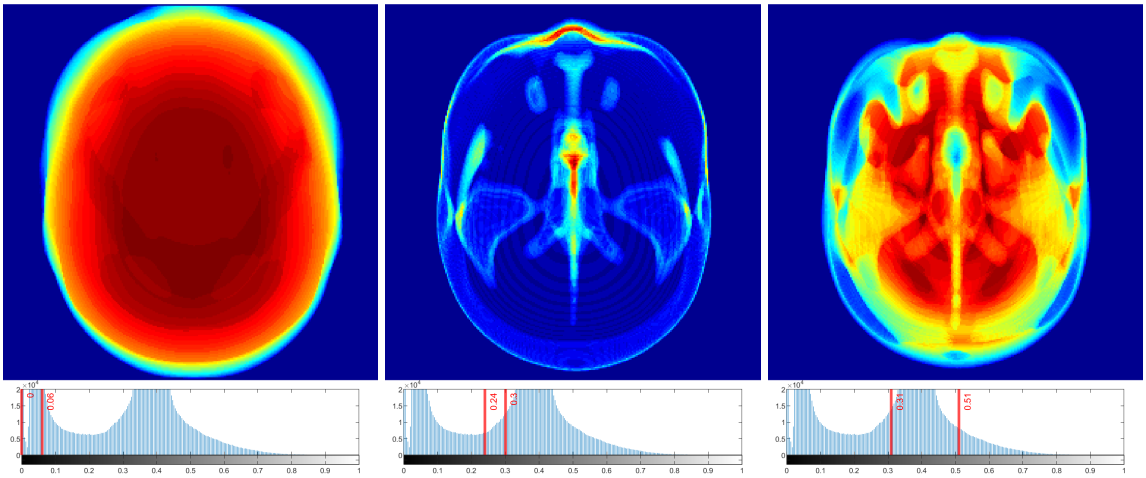


(d) Demons head (e) Demons CSF (f) Demons brain

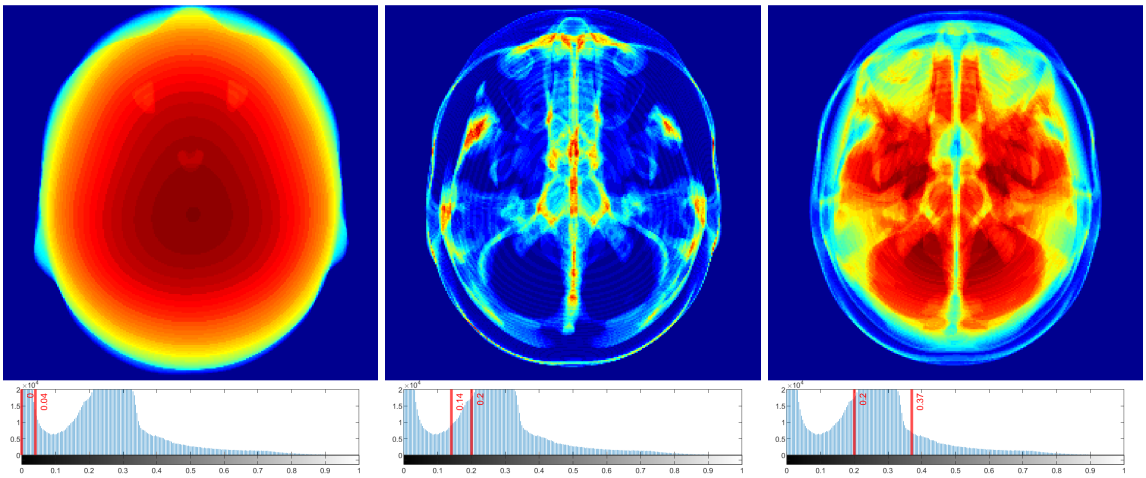


(g) SE head (h) SE CSF (i) SE brain

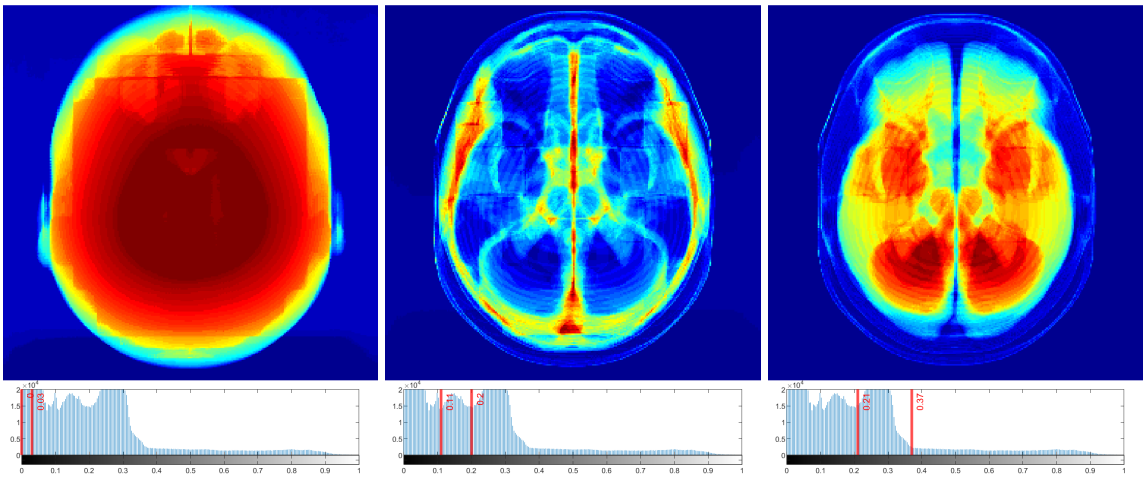
Figure A.6: Heatmaps showing the separability of tissue for the non-standardized ADNI data.



(a) VM head (b) VM CSF (c) VM brain

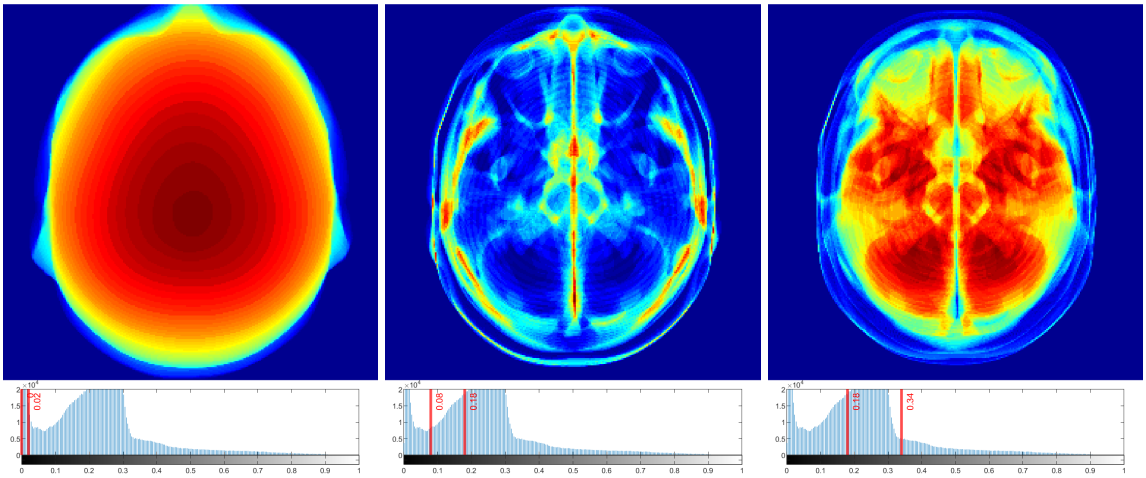


(d) Affine head (e) Affine CSF (f) Affine brain

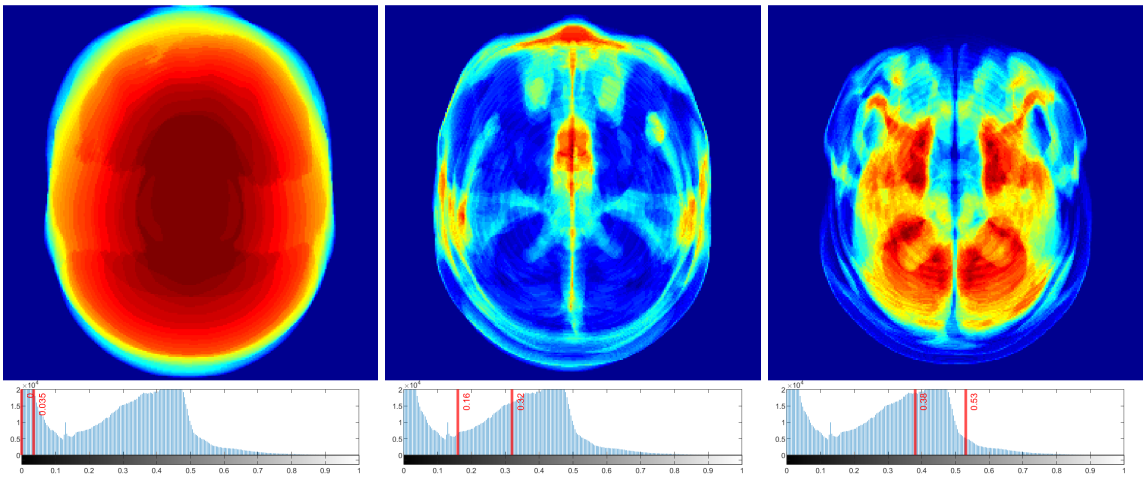


(g) FlowReg head (h) FlowReg CSF (i) FlowReg brain

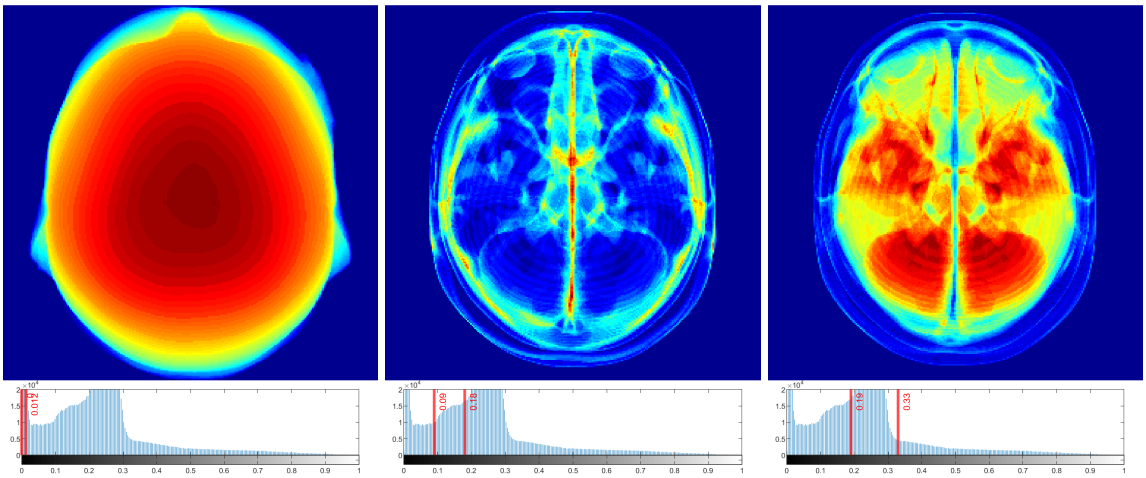
Figure A.7: Heatmaps showing the separability of tissue for the non-standardized ADNI data.



(a) ANTs head (b) ANTs CSF (c) ANTs brain

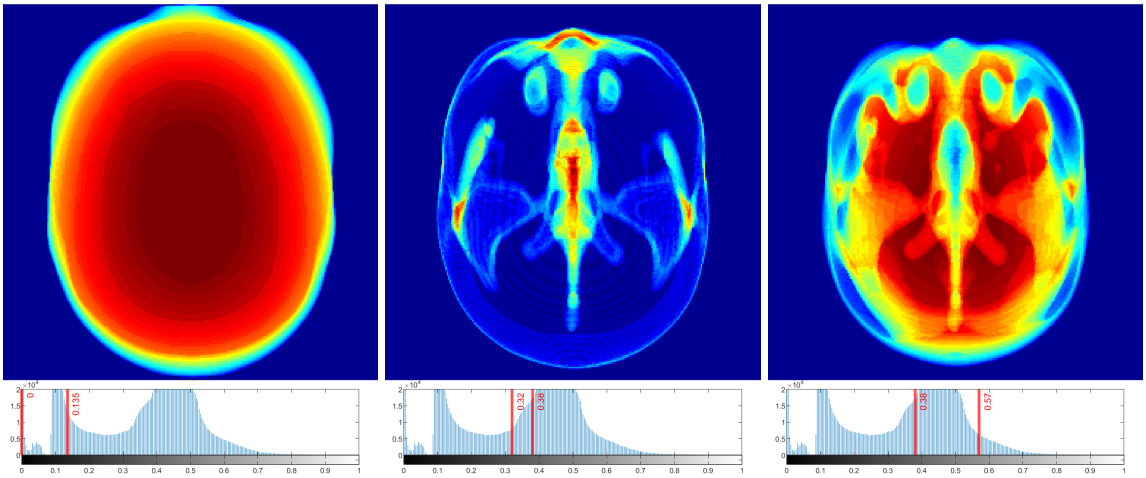


(d) Demons head (e) Demons CSF (f) Demons brain

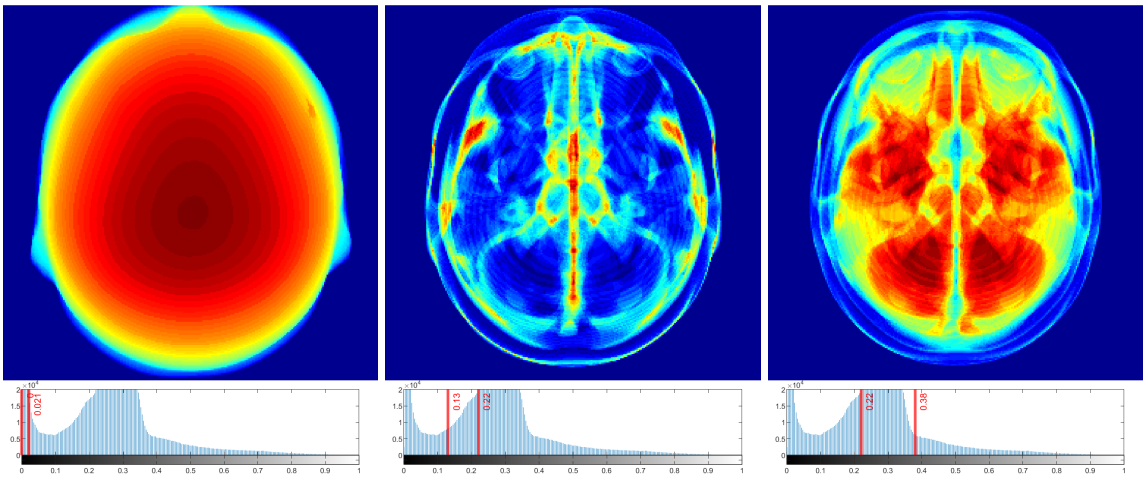


(g) SE head (h) SE CSF (i) SE brain

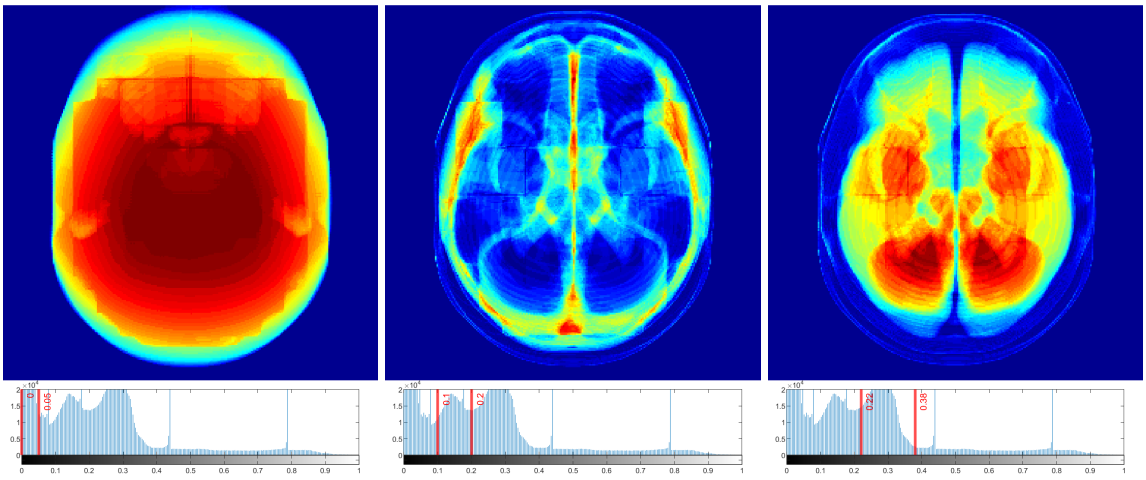
Figure A.8: Heatmaps showing the separability of tissue for the standardized ADNI data.



(a) VM head (b) VM CSF (c) VM brain



(d) Affine head (e) Affine CSF (f) Affine brain



(g) FlowReg head (h) FlowReg CSF (i) FlowReg brain

Figure A.9: Heatmaps showing the separability of tissue for the standardized ADNI data.

Bibliography

[1] A. E. Raftery, N. Li, H. Ševčíková, P. Gerland, and G. K. Heilig, "Bayesian probabilistic population projections for all countries," *Proceedings of the National Academy of Sciences*, vol. 109, no. 35, pp. 13 915–13 921, 2012.

[2] D. L. Rotstein, H. Chen, A. S. Wilton, J. C. Kwong, R. A. Marrie, P. Gozdyra, K. M. Krysko, A. Kopp, R. Copes, and K. Tu, "Temporal trends in multiple sclerosis prevalence and incidence in a large population," *Neurology*, vol. 90, no. 16, pp. e1435–e1441, 2018.

[3] R. Mayeux and Y. Stern, "Epidemiology of alzheimer disease," *Cold Spring Harbor Perspectives in Medicine*, vol. 2, no. 8, p. a006239, 2012.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich, Germany: Springer, 2015, pp. 234–241.

[6] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.

[7] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. Van Der Smagt, D. Cremers, and T. Brox, "Flownet: Learning optical flow with convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 2758–2766.

[8] J. Y. Jason, A. W. Harley, and K. G. Derpanis, "Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness," in *European Conference on Computer Vision*. Springer, 2016, pp. 3–10.

[9] F. H. Martini and E. F. Bartholomew, *Essentials of Anatomy & Physiology: Pearson New International Edition*. Pearson Higher Ed, 2017.

[10] R. I. Grossman and J. C. McGowan, "Perspectives on multiple sclerosis." *American Journal of Neuroradiology*, vol. 19, no. 7, pp. 1251–1265, 1998.

[11] K. Oppedal, T. Eftestøl, K. Engan, M. K. Beyer, and D. Aarsland, "Classifying dementia using local binary patterns from different regions in magnetic resonance images," *Journal of Biomedical Imaging*, vol. 2015, pp. 1–14, 2015.

[12] P. Malloy, S. Correia, G. Stebbins, and D. H. Laidlaw, "Neuroimaging of white matter in aging and dementia," *The Clinical Neuropsychologist*, vol. 21, no. 1, pp. 73–109, 2007.

[13] K. Kobayashi, M. Hayashi, H. Nakano, Y. Fukutani, K. Sasaki, M. Shimazaki, and Y. Koshino, "Apoptosis of astrocytes with enhanced lysosomal activity and oligodendrocytes in white matter lesions in alzheimer's disease," *Neuropathology and Applied Neurobiology*, vol. 28, no. 3, pp. 238–251, 2002.

[14] E. M. Castaño, C. L. Maarouf, T. Wu, M. C. Leal, C. M. Whiteside, L.-F. Lue, T. A. Kokjohn, M. N. Sabbagh, T. G. Beach, and A. E. Roher, "Alzheimer disease periventricular white matter lesions exhibit specific proteomic profile alterations," *Neurochemistry International*, vol. 62, no. 2, pp. 145–156, 2013.

[15] N. Altaf, P. S. Morgan, A. Moody, S. T. MacSweeney, J. Gladman, and D. P. Auer, "Brain white matter hyperintensities are associated with carotid intraplaque hemorrhage," *Radiology*, vol. 248, no. 1, pp. 202–209, 2008.

[16] A. P. Dhawan, "Medical imaging modalities: magnetic resonance imaging," pp. 99–136, 2011.

[17] S. C. Bushong and G. Clarke, *"Fundamentals", in Magnetic Resonance Imaging-E-Book: Physical and Biological Principles*. Elsevier Health Sciences, 2015.

[18] M. A. Bernstein, K. F. King, and X. J. Zhou, *Handbook of MRI pulse sequences*. Elsevier, 2004.

[19] J. L. Prince and J. M. Links, *Medical imaging signals and systems*. Pearson Education Prentice Hall Upper Saddle River, 2006.

[20] R. Bitar, G. Leung, R. Perng, S. Tadros, A. R. Moody, J. Sarrazin, C. McGregor, M. Christakis, S. Symons, A. Nelson *et al.*, "Mr pulse sequences: what every radiologist wants to know but is afraid to ask," *Radiographics*, vol. 26, no. 2, pp. 513–537, 2006.

[21] A. Khademi, A. Venetsanopoulos, and A. R. Moody, "Robust white matter lesion segmentation in flair mri," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 3, pp. 860–871, 2011.

[22] R. Bakshi, S. Ariyaratana, R. H. Benedict, and L. Jacobs, "Fluid-attenuated inversion recovery magnetic resonance imaging detects cortical and juxtacortical multiple sclerosis lesions," *Archives of Neurology*, vol. 58, no. 5, pp. 742–748, 2001.

[23] A. Khademi, A. Venetsanopoulos, and A. R. Moody, "Image enhancement and noise suppression for flair mris with white matter lesions," *IEEE Signal Processing Letters*, vol. 17, no. 12, pp. 989–992, 2010.

[24] J. Knight, G. W. Taylor, and A. Khademi, "Voxel-wise logistic regression and leave-one-source-out cross validation for white matter hyperintensity segmentation," *Magnetic Resonance Imaging*, vol. 54, pp. 119–136, 2018.

[25] J.-P. Thirion, "Fast non-rigid matching of 3d medical images," 1995.

[26] ——, "Image matching as a diffusion process: an analogy with maxwell's demons," *Medical Image Analysis*, vol. 2, no. 3, pp. 243–260, 1998.

[27] S. Klein, M. Staring, K. Murphy, M. A. Viergever, J. P. Pluim *et al.*, "Elastix: a toolbox for intensity-based medical image registration," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 196–205, 2010.

[28] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," *Medical Image Analysis*, vol. 12, no. 1, pp. 26–41, 2008.

[29] D. Hawkes, "Algorithms for radiological image registration and their clinical application," *The Journal of Anatomy*, vol. 193, no. 3, pp. 347–361, 1998.

[30] D. Rueckert and J. A. Schnabel, "Medical image registration," in *Biomedical Image Processing*. Springer, 2010, pp. 131–154.

[31] M. Visser, D. Müller, R. van Duijn, M. Smits, N. Verburg, E. Hendriks, R. Nabuurs, J. Bot, R. Eijgelaar, M. Witte *et al.*, "Inter-rater agreement in glioma segmentations on longitudinal mri," *NeuroImage: Clinical*, vol. 22, p. 101727, 2019.

[32] G. M. Cattaneo, M. Reni, G. Rizzo, P. Castellone, G. L. Ceresoli, C. Cozzarini, A. J. M. Ferreri, P. Passoni, and R. Calandrino, "Target delineation in post-operative radiotherapy of brain gliomas: interobserver variability and impact of image registration of mr (pre-operative) images on treatment planning ct scans," *Radiotherapy and Oncology*, vol. 75, no. 2, pp. 217–223, 2005.

[33] M. Dadar, V. S. Fonov, D. L. Collins, A. D. N. Initiative *et al.*, "A comparison of publicly available linear mri stereotaxic registration techniques," *Neuroimage*, vol. 174, pp. 191–200, 2018.

[34] J. Modersitzki, *Numerical methods for image registration*. Oxford University Press on Demand, 2004.

[35] A. A. Goshtasby, *Image registration: Principles, tools and methods*. Springer Science & Business Media, 2012.

[36] J. M. Fitzpatrick and M. Sonka, "Handbook of medical imaging, volume 2. medical image processing and analysis (spie press monograph vol. pm80)," in *SPIE–The International Society for Optical Engineering*, vol. 1, 2004, pp. 447–513.

[37] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 6, pp. 1153–1160, 1981.

[38] C. G. Knott, *Life and scientific work of Peter Guthrie Tait*. , 1911, vol. 1.

[39] V. I. Arnold and B. A. Khesin, "Topological methods in hydrodynamics," *Annual Review of Fluid Mechanics*, vol. 24, no. 1, pp. 145–166, 1992.

[40] H. J. Johnson, M. McCormick, L. Ibáñez, and T. I. S. Consortium, *The ITK Software Guide*, 3rd ed., Kitware, Inc., 2013, *In press*, http://www.itk.org/ItkSoftwareGuide.pdf, last accessed on 2019-12-29.

[41] C. Studholme, D. L. Hill, and D. J. Hawkes, "Automated 3-d registration of mr and ct images of the head," *Medical Image Analysis*, vol. 1, no. 2, pp. 163–175, 1996.

[42] H. Neemuchwala and A. Hero, "Entropic graphs for registration," *Multi-Sensor Image Fusion and Its Applications*, pp. 185–235, 2005.

[43] M. Staring, U. A. Van Der Heide, S. Klein, M. A. Viergever, and J. P. Pluim, "Registration of cervical mri using multifeature mutual information," *IEEE Transactions on Medical Imaging*, vol. 28, no. 9, pp. 1412–1421, 2009.

[44] A. P. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer, "Morphometric analysis of white matter lesions in mr images: method and validation," *IEEE Transactions on Medical Imaging*, vol. 13, no. 4, pp. 716–724, 1994.

[45] S. Klein, M. Staring, and J. P. Pluim, "Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2879–2890, 2007.

[46] H. Robbins and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, pp. 400–407, 1951.

[47] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "An unsupervised learning model for deformable medical image registration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, Utah, USA, 2018, pp. 9252–9260.

[48] I. B. Frenkel and V. Kac, "Basic representations of affine lie algebras and dual resonance models," *Inventiones Mathematicae*, vol. 62, no. 1, pp. 23–66, 1980.

[49] F. Chollet *et al.*, "Keras," https://keras.io, 2015, last accessed on 2019-12-29.

[50] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org last accessed on 2019-12-29. [Online]. Available: https://www.tensorflow.org/

[51] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[52] D. Fleet and Y. Weiss, "Optical flow estimation," in *Handbook of Mathematical Models in Computer Vision*. Springer, 2006, pp. 237–257.

[53] B. D. Lucas, T. Kanade *et al.*, "An iterative image registration technique with an application to stereo vision," 1981.

[54] D. P. Mitchell and A. N. Netravali, "Reconstruction filters in computer-graphics," in *ACM Siggraph Computer Graphics*, vol. 22, no. 4. ACM, 1988, pp. 221–228.

[55] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," *International Journal of Computer Vision*, vol. 2, no. 3, pp. 283–310, 1989.

[56] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," in *European conference on computer vision*. Springer, 1992, pp. 237–252.

[57] M. Jaderberg, K. Simonyan, A. Zisserman *et al.*, "Spatial transformer networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 2017–2025.

[58] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[59] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

[60] H. Attouch, G. Buttazzo, and G. Michaille, *Variational analysis in Sobolev and BV spaces: applications to PDEs and optimization*. SIAM, 2014.

[61] H. Rehman and S. Lee, "An efficient automatic midsagittal plane extraction in brain mri," *Applied Sciences*, vol. 8, no. 11, p. 2203, 2018.

[62] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[63] R. C. Gonzalez, R. E. Woods *et al.*, "Digital image processing," 2002.

[64] R. Fisher, P. S, W. A, and W. E, "Morphology," https://homepages.inf.ed.ac.uk/rbf/HIPR2/morops.htm, last accessed on 2019-07-21.

[65] Y. Liu, R. T. Collins, and W. E. Rothfus, "Robust midsagittal plane extraction from normal and pathological 3-d neuroradiology images," *IEEE Transactions on Medical Imaging*, vol. 20, no. 3, pp. 175–192, 2001.

[66] S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett, "The alzheimer's disease neuroimaging initiative," *Neuroimaging Clinics*, vol. 15, no. 4, pp. 869–877, 2005.

[67] B. Reiche, A. Moody, A. Khademi, A. D. N. Initiative *et al.*, "Pathology-preserving intensity standardization framework for multi-institutional flair mri datasets," *Magnetic Resonance Imaging*, 2019.

[68] A. Winkler, P. Kochunov, and D. Glahn, "FLAIR Templates—Brainder," last accessed 2019-12-29. [Online]. Available: http://brainder.org

[69] MATLAB, *version 9.6.0 (R2019a)*. Natick, Massachusetts: The MathWorks Inc., 2019.

[70] X. Pennec, P. Cachier, and N. Ayache, "Understanding the demons algorithm: 3d non-rigid registration by gradient descent," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 1999, pp. 597–605.

[71] K. Marstal, F. Berendsen, M. Staring, and S. Klein, "Simpleelastix: A user-friendly, multi-lingual library for medical image registration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Athens, Greece, 2016, pp. 134–142.

[72] J. Talairach and P. Tournoux, "Co-planar stereotaxic atlas of the human brain.3-dimensional proportional system: An approach to cerebral imaging," vol. 39, p. 145, 01 1988.

[73] A. C. Evans, "An mri-based stereotactic atlas from 250 young normal subjects," *Soc. neurosci. abstr, 1992*, 1992.

[74] K. J. Friston, A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith, and R. S. Frackowiak, "Statistical parametric maps in functional imaging: a general linear approach," *Human Brain Mapping*, vol. 2, no. 4, pp. 189–210, 1994.

[75] A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, Eds., *Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV*, ser. Lecture Notes in Computer Science, vol. 11073. Springer, 2018. [Online]. Available: https://doi.org/10.1007/978-3-030-00937-3

[76] E. Strubell, A. Ganesh, and A. McCallum, "Energy and policy considerations for deep learning in nlp," *arXiv preprint arXiv:1906.02243*, 2019.

[77] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.