# M.Sc.  Thesis

# Nonrigid Image Registration using 3D Convolutional Neural Network with Application to Brain MR images

### Chenhong Ji

### Abstract

Image registration is a vital tool in medical image analysis with a large number of applications assisting the medical experts. Currently, conventional image registration approach with predefined dissimilarity metric and iterative optimization, is widely used. In this thesis, we proposed a method to solve medical image registration problem using a deep learning approach. A convolutional neural network architecture is proposed, named as "RegNet", applied on monomodality image registration problem. The proposed RegNet does not require any dissimilarity metric and is capable of directly estimating the spatial relationship between two images. The training is based on the pseudo-real-world displacement vector field, created by the conventional registration tool and artificial deformation simulation, resembling a deformation similar with the real-world deformation case. Multi-stage framework is also implemented to increase the capture range of RegNet. This thesis evaluates the performance of "RegNet" for an intrasubject magnetic resonance brain images registration problem.

**TUDelft**

# Nonrigid Image Registration using 3D Convolutional Neural Network with Application to Brain MR images

THESIS

submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

in

ELECTRICAL ENGINEERING

by

Chenhong Ji
born in Dalian, China

This work was performed in:

Circuits and Systems Group
Department of Microelectronics & Computer Engineering
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology

**Delft University of Technology**

# Abstract

Image registration is a vital tool in medical image analysis with a large number of applications assisting the medical experts. Currently, conventional image registration approach with predefined dissimilarity metric and iterative optimization, is widely used. In this thesis, we proposed a method to solve medical image registration problem using a deep learning approach. A convolutional neural network architecture is proposed, named as "RegNet", applied on monomodality image registration problem. The proposed RegNet does not require any dissimilarity metric and is capable of directly estimating the spatial relationship between two images. The training is based on the pseudo-real-world displacement vector field, created by the conventional registration tool and artificial deformation simulation, resembling a deformation similar with the real-world deformation case. Multi-stage framework is also implemented to increase the capture range of RegNet. This thesis evaluates the performance of "RegNet" for an intrasubject magnetic resonance brain images registration problem.

# Acknowledgments

First of all, I would like to express my deep gratitude to everyone who have offered me great help in this master project.

I would like thank my supervisor dr.ir. R.F. Remis for the opportunity and his assistance and support during the thesis process.

I am grateful to dr.ir. M. Staring, my supervisor at LUMC, for offering me an opportunity to have this master project in LUMC. He gives me a lot of support and guidance.

I also want to express my deepest appreciation to my daily supervisor drs. Hessam Sokooti for his valuable guidance during the thesis project. I really appreciate his patience and feedback.

Chenhong Ji
Delft, The Netherlands
09-10-2019

# Contents

# List of Figures

# List of Tables

# Introduction

<div style="text-align: right">**1**</div>

## 1.1 Background

Image Registration is a process to find the spatial relationship between a pair of images. In clinical studies, medical image registration supports the medical experts analyzing the correspondence between two images obtained at different times or using different imaging modality[18]. It is a vital tool for diverse medical applications, including disease progression monitoring, atlas registration, and medical image fusion. Image registration is applied to find two different kinds of transformation, "rigid transformation", representing the movements that simply rotate and translate the image to align with the reference; "non-rigid transformation", adjusting images with local deformations[12] to match with the reference. This thesis focuses on non-rigid registration.

In a pair-wise registration, two images are involved. One is defined as a fixed image $I_F$, to be used as the reference. The other one is defined as the moving image $I_M$. The transformation represents the mapping from the fixed image domain to the moving image domain[41]. Following the obtained transformation, the moving image is then deformed. The deformed moving image should be aligned with the fixed image[24].

Conventional registration method is widely used in medical image analysis. Conventional registration methods try to compute the spatial relationship between images by solving an optimization problem aiming at minimizing the dissimilarity of two images through iterative optimization. However, conventional registration methods can be time-consuming[36] when a large amounts of iterations required. The optimization is also possible to get stuck in local optima.

In 2012, AlexNet achieved a huge success in the ImageNet challenge[18]. Since then, deep learning method is introduced into the computer vision domain, including image registration. Deep learning is a type of machine learning method that uses a neural network, such as, convolutional neural network (CNN), to accomplish a pre-defined task by learning data[18]. Neural network is suitable to be applied in highly non-linear problem [36], such as image registration problem. Comparing with the traditional methods, neural network is in general more robust against local optimum[13]. Moreover, due to highly parallel architecture, solving the registration problem using CNN when GPU is utilized is much faster.

Deep learning approach has enabled real-time, robust medical image registration[18]. Based on the design of the training process, most deep learning methods can be classified into two types, supervised learning and unsupervised learning. In supervised learning approach, the target output is used during the training process, while in unsupervised learning method there is no need to provide corresponding output during the training.

Recently, there are some unsupervised approaches proposed to solve the image registration

problem. In [14], an end-to-end unsupervised CNN is proposed. It receives the fixed image and moving image, predicting the parameters for a spatial transformer yielding the displacement vector field. During the training the intensity-based similarity metric is optimized. In [45], Large Deformation Diffeomorphic Metric Mapping (LDDMM) model is applied, also intending to minimize the dissimilarity. A recent paper [2] also applies a CNN in an unsupervised method, with a segmentation-based metric for CNN training. Generative Adversarial Network (GAN) [17] is applied in [29] to directly learn the plausible deformation represented by density function.

With simulated deformation, it is also possible to use supervised learning approach in medical image registration. In [36], mesh segmentations are applied to create the ground-truth deformation for the training of U-net[37] like architecture. Random deformation is used in [41] as the target output for CNN training. Statistical appearance models (SAMs) is used in [44] to generate deformation for CNN training.

## 1.2 Motivation of this thesis

This thesis proposes a supervised deep learning approach that applies CNN and a multistage framework to predict the deformation between the $I_M$ and $I_F$. The CNN applied is a modified version of the CNN originally designed for lung CT image registration in [41]. The main challenge for applying supervised learning approach to solve the medical registration problem is creating the realistic ground-truth deformation for the training. One method is to simulate deformation, based on known transformation[27][45]. Deformations created using models are in general arbitrary and diverse, but it is hard for simulated data to resemble real underlying deformation with inherent noises. The other method is to create deformation by applying conventional registration tool on a pair of images[9]. The deformation created in this method is more realistic. However limited by the conventional registration method, only specific type of deformation can be learned.

In this project, the CNN is trained with a combination of deformation generated by both methods mentioned above. Ground truth deformation applied in training are created by synthetic simulation and by running conventional registration methods on pairs of images.

## 1.3 Problem definition

This thesis aims at answering questions below,

1. What is the performance of applying the proposed CNN, RegNet, on the intrasubject brain registration problem?

2. How could the training data influence the registration performance?

3. How could the multistage framework influence the registration performance?

## 1.4    Thesis outline

This thesis is organized as below, Chapter 2 presents the dataset used in this thesis and the selection criteria. In Chapter 3, the proposed network is given, alongside with the multistage framework and details related to the ground-truth creation. Chapter 4 presents the registration result. Chapter 5 analyzes the result and suggests future work. Chapter 6 concludes the project.

# 2

# Dataset

## 2.1 Magnetic resonance imaging

Magnetic Resonance Imaging (MRI) is a non-invasive imaging technology based on nuclear magnetic resonance (NMR phenomenon). It allows patients to be scanned without exposing to radiation[33].

NMR describes the movement of nuclei under a high magnetic field, after an excitation of RF pulse. Placed in a strong magnetic field B0, nuclei in the human body start to spin around the static magnetic field B0 at the Larmor frequency, which is proportional to the magnitude of B0 and the gyromagnetic ratio. Gyromagnetic ratio is a constant unique for all kinds of nuclei. When additional RF pulses are added, an outer magnetic field B1 perpendicular to B0 will be brought by the RF pulses. The aligned nuclei then start rotating around B1. After the external RF excitation is switched off, the nuclei begin to rotate back to its previous status. "Relaxation" is used to describe the process of realignment. The spin of nuclei will trigger changes in net magnetism, which will be captured by receiver coil near the patient[33]. The duration of relaxation is depending on the different amount of activated nuclei, which is tissue-dependent. Thus the difference of the corresponding MRI signals measured by receiver coil indicate different tissues.

The relaxation can be decomposed into two types, T1 and T2. T1 is the process that takes the nuclei gradually recover to its previous status before the external RF pulse is added. T1 time describes the time that the recovering magnetic field reaches 63% of its maximum value[5]. While T2 relaxation describes the process of the gradual reduction of magnetic field in the transverse direction. T2 relaxation follows the exponential decay. T2 time describes the time that the rest magnetic field reaches 37% of its initial value[16].

In clinical MRI, hydrogen nuclei are used for imaging. Because they can easily absorb energy and emit electromagnetic waves when exposed in the magnetic field[32], and they are also widely residing in the human body. In most organs, hydrogen nuclei from water are the only source for magnetic resonance.

Besides, hydrogen nuclei from lipid hydrogen can also be used as a hydrogen source excited in magnetic field. Lipid hydrogen is another type of hydrogen activated in the human body during the scanning. Lipid hydrogen mainly reside in aliphatic (-$CH_2$-) chains of triglycerides, which can be commonly found in adipose tissues[15]. Lipid hydrogen has complex motion mode when placed in the magnetic field, including some of them rotate at the Larmor frequency together with the water hydrogen nuclei. The motion of lipid hydrogen accelerate relaxation and reduce the T1 time[34][35].

Changing the pulse sequence, the impact of either T1 or T2 time on the images can be adjusted to be dominating. T1-weighted images are the images where the T1 time will play

a key role in imaging. Since T1 is affected significantly by the amount of lipid hydrogen in tissues, T1-weight images are also recognized as a map of adipose tissue, where the region with a higher portion of fatty tissue will be brighter[3].

On the other hand, T2 images are created by pulse sequences where tissues with different signal decays can be differentiated[6]. T2, the signal decay time depends on the amount of water and adipose tissue[26]. Different from T1-weighted image, the tissue only with a higher amount of hydrogen nuclei will also be bright in T2-weighted images. When comparing the T1-weighted and T2-weighted images taken from the same region, the tissue abundant with fat is bright on both the T1-weighted and T2-weighted images. Though the water-based tissue will only be bright on T2-weighted images.

By applying different pulse sequences, it is easy to differentiate soft tissues in MR images. Thus, the boundary between soft tissues in the MR images can be more easily observed. That is why MR imaging is powerful in neural imaging. It also explains why MR images are more sensitive to abnormal brain tissue deformation[19].

To better differentiate tissues, a uniform magnetic field works as a key for high-quality MR images. Because only with a uniform magnetic field, the MR signals measured by the receiver coils will be dominated by the amount of the same nuclei type. Inhomogeneous magnetic filed can result in different artifacts, including shading, spatial distortion, blurring and other issues. To tackle this issue, in this project, images are processed with the non-parametric non-uniform intensity normalization (N3)[7].

There are also drawbacks of using MRI images. One of the main disadvantages is that the MR image intensity scale is pulse sequence dependent[4]. Exposed in different pulse sequences, the same tissue will be drawn in different intensities, which makes the intensity-based registrations more difficult.

## 2.2 ADNI database

In this project, brain MR images from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (available from www.loni.ucla.edu/ADNI) are chosen. Alzheimer's disease is an irreversible neurodegenerative disease[22]. Moreover, it is also the most common cause of dementia, with symptoms such as memory loss and other cognitive skills declines. The early diagnosis is very challenging. Originally, ADNI was initiated to combine MR images, positron emission tomography (PET), (18F)-fluorode-oxyglucose positron emission tomography (FDG PET), cerebrospinal fluid biomarkers (CSF), and other biological markers for Alzheimer's disease early diagnosis and progression estimation related projects[22].

In ADNI database, to keep the consistency of the MR images, MR acquisition was deliberately done using the ADNI acquisition protocol as defined in [10]. In this way, similar image qualities, including contrast-to-noise ratio, spatial resolution, can be maintained[22]. Changes of ADNI MR protocol can be divided into 3 phases, ADNI-1 (2004-2009) focused on longitudinal imaging. 1.5T scanners were used with T1 and dual-echo T2-weighted sequences[22]. From 2010-2016, the ADNI-GO/ADNI-2 study applied protocol similar to ADNI-1, 3T scanners with T1-weighted imaging parameters were used. 2D FLAIR and T2-weighted imaging were

also added in ADNI-GO/ADNI-2 study[22]. Different from the previous study, ADNI-3 study is implemented using 3T scanners[22].

### 2.2.1   N3 technique

To better exclude undesirable affect of inhomogeneous magnetic field, non-parametric non-uniform intensity normalization (N3)[40] is applied. In MR scans, inhomogeneous magnetic field reduces the high frequency component of the coil-measured signal, resulting into the blurring of the probability density of the signal. Thus restoring the signal under a uniform magnetic field can be interpreted as a problem to maximize the frequency content of the intensity distribution[7].

The probability density of the measured signal is the convolution of the probability density of the true signal and that of the bias field[40]. Without any information of the bias field, it is not possible to restore the true signal. N3 assumes that the unknown bias magnetic field is following Gaussian distribution. Since any Gaussian distribution can be decomposed into convolution of several narrower Gaussian distribution[7], N3 gradually estimates the distribution of the biased non-uniform magnetic field by iteratively deconvolving narrower Gaussian distribution on the subsequent estimated target signal.

To reduce the affect of the non-uniformity brought by inhomogeneous magnetic field so that the best registration performance can be achieved, T1 MR images from ADNI database with N3 correction are chosen for this experiment.

### 2.2.2   Hippocampus mask

Hippocampus is a grey matter structure in the brain, which plays an important role in memory processing[10]. It is known that the progression of Alzheimer disease affects the hippocampus, even at the early stage. Significant hippocampus volume reduction can be observed in Alzheimers Disease patients[10]. Figure 2.1 is giving a visualization of hippocampus.

In the ADNI 1 study, semi-automated hippocampus masks are available. Those hippocampal masks in ADNI1 study were created using Medtronic Surgical Navigation Technologies (SNT). At first local landmarks for individual hippocampus and global landmarks for MR image are placed, then fluid image transformation is applied to match landmarks with landmark from template[10]. Some further refinement of the segmentation are carried out manually afterwards in case of poor quality segmentation. These hippocampus masks are directly applied for the evaluation of this project.

## 2.3   Data selection

The dataset used in this experiment is composed of two subsets, disease progression (DP) dataset and 1-Year dataset, both chosen from the ADNI database. This section will further explain the selection scenario. All of the selected images are in the NIfTI format.
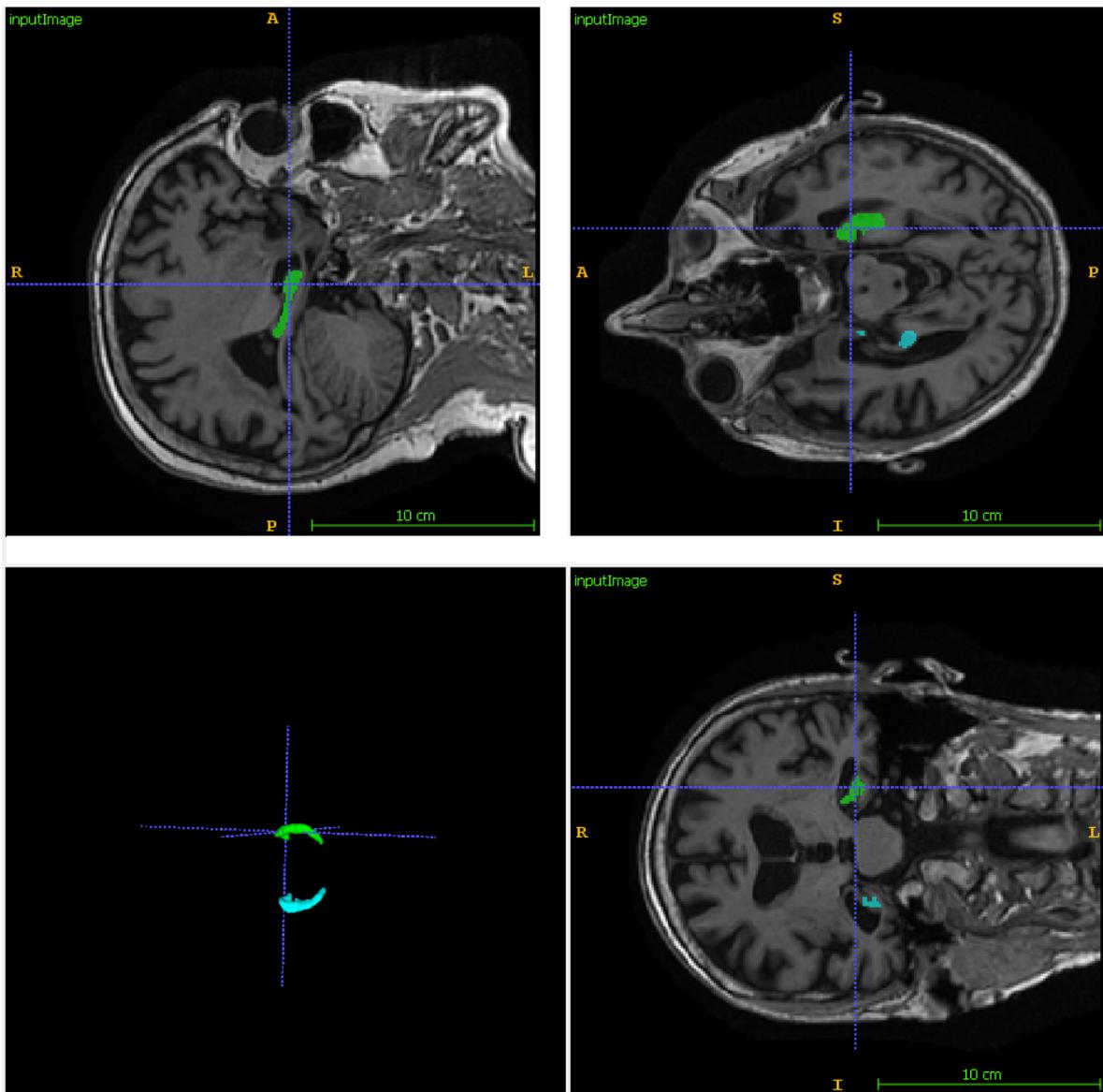
Figure 2.1: ADNI Brain MR image with Hippocampus mask

### 2.3.1  1-Year dataset

A set of subjects with two or more scans including hippocampus masks are selected from the ADNI-1 study. The histogram of the interval between the baseline and follow-up for subjects with hippocampus mask is shown in fig. 2.2. Among all of these pairs, only those with an interval over one year between the baseline and follow-up are chosen. T1-weighted images in those scans are selected in this project. The filtered dataset contains 157 pairs of brain MR T1 images for the experiment. 26 out of the total 157 pairs are used as test images. ADNI indexes for all subjects can be found in Appendix.

The dataset contains 7 disease progressive types, as shown in fig. 2.2. Bars with different colors
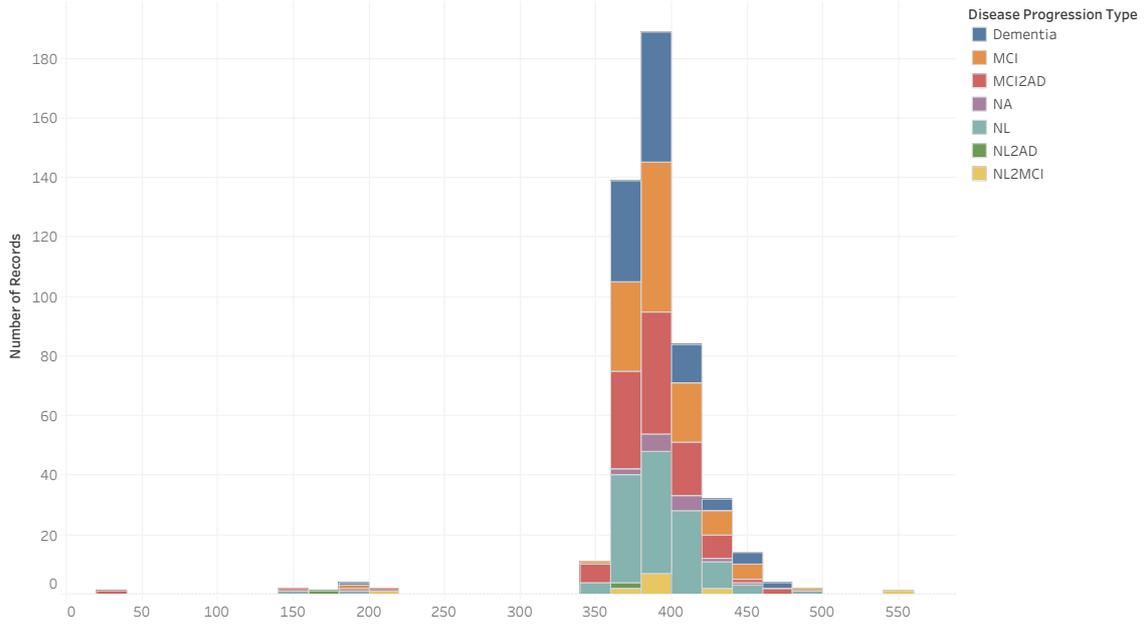
Figure 2.2: Interval between intrasubject image pairs with hippocampus mask, MCI, MCI2AD, NA, NL, NL2AD, NL2MCI denote mild cognitive impairment, mild cognitive impairment to Alzheimer's, None, Normal, Normal to Alzheimer's, Normal to Mild Cognitive Impairment respectively.

indicate different disease progression type, including dementia, mild cognitive impairment (MCI), Alzheimer's disease (AD), Normal (NL), MCI changing to AD (MCI2AD), and a few cases marked as subject developing from Normal to MCI (NL2MCI) or Normal to AD (NL2AD). The 1-Year dataset is composed of images with different disease progressive types.

### 2.3.2 Disease progression dataset

The other selection DP contains images from subjects with different disease stages in baseline and follow-up. The subject status changes in this data set include MCI to AD, Normal to AD, Normal to MCI. Comparing with the subjects in the 1-Year dataset, the subjects in DP dataset in general have longer interval from the baseline scan to the follow-up scan as shown in fig. 2.3. All in all 151 pairs of T1-weighted images from different subjects are chosen for this experiment, including 26 pairs only used in the evaluation. ADNI indexes for all subjects can be found in Appendix. It should be noted that there are no overlapping subjects between DP and 1-Year dataset.
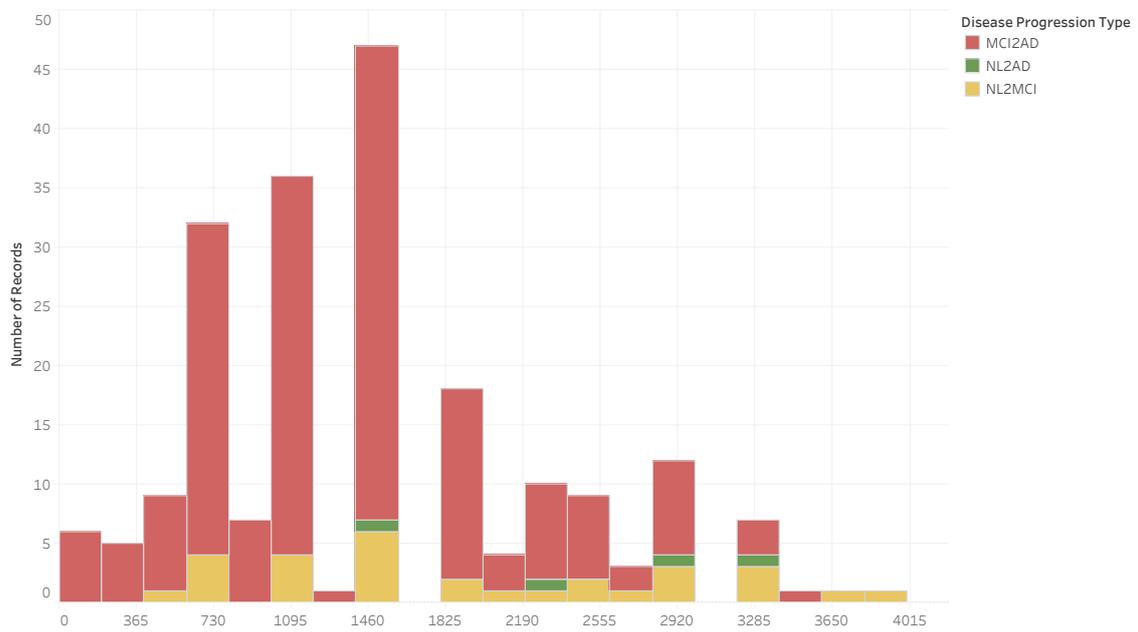
Figure 2.3: Interval between intrapatient image pairs for subject in disease progression dataset, MCI2AD, NL2AD, NL2MCI denote mild cognitive impairment to Alzheimer's, Normal to Alzhimer's, Normal to Mild Cognitive Impairment respectively

# Methodology

<span style="font-size:3em">3</span>

## 3.1 Registration framework

### 3.1.1 Registration problem

Image registration problem aims at finding a coordinate transformation that aligns the moving image $I_M$ and the fixed image $I_F$. In other words, the target of registration is to find the transformation $T(x) = x + \mu$ assuring that $I_M(x + \mu)$ and $I_F(x)$ have high similarity. In this thesis the transformation $T(x)$ is defined as the coordinate mapping from the fixed image to the moving image[41]. This confusing definition is because if the deformation is from the moving image coordinate to the fixed image coordinate, when the transformation is applied onto the moving image, there is no guarantee that all voxels from the fixed image can be linked with voxels in the moving image coordinate, and holes may appear in the deformed moving image[24].

The quality of image registration is evaluated by the distance or the similarity metric $S$, for example, the sum of squared differences (SSD), the correlation ratio, and the mutual information (MI)[23]. With the help of these metrics, registration can be interpreted as an optimisation problem to minimize the dissimilarity,

$$\hat{T} = \arg\min_{T} C(T; I_F, I_M), \tag{3.1}$$

$$C(T; I_F, I_M) = -S(T; I_F, I_M) + \gamma P(T) \tag{3.2}$$

where C denotes the cost function to be minimised. Since the registration problem is an ill-posed problem[23], a penalty term P is needed. $\gamma$ is added as the weight for regularity.

There are two approaches to define the transformation, one is the parametric approach and the other one is the non-parametric approach. In the parametric method, the deformation is composed of known deformation models, such as, B-spline model, thin-plate spline model, etc. Then, during the optimization, a set of parameters within the known model is optimized based on the dissimilarity metric[30]. The original optimization problem can then be rephrased as,

$$\hat{T}_\mu = \arg\min_{T_\mu} C(T_\mu; I_F, I_M) \tag{3.3}$$

where $\mu$ denotes the set of parameters based on the applied model.

Using the non-parametric method to solve the registration problem is to directly estimate the displacement vector field (DVF) without parameters involved[28]. DVF is a set of vectors pointing out the movement of individual pixel or voxel. DVF in 3D space is a 3D set of vectors with the length of 3. Our proposed CNN is a non-parametric method with more details in section 3.2.
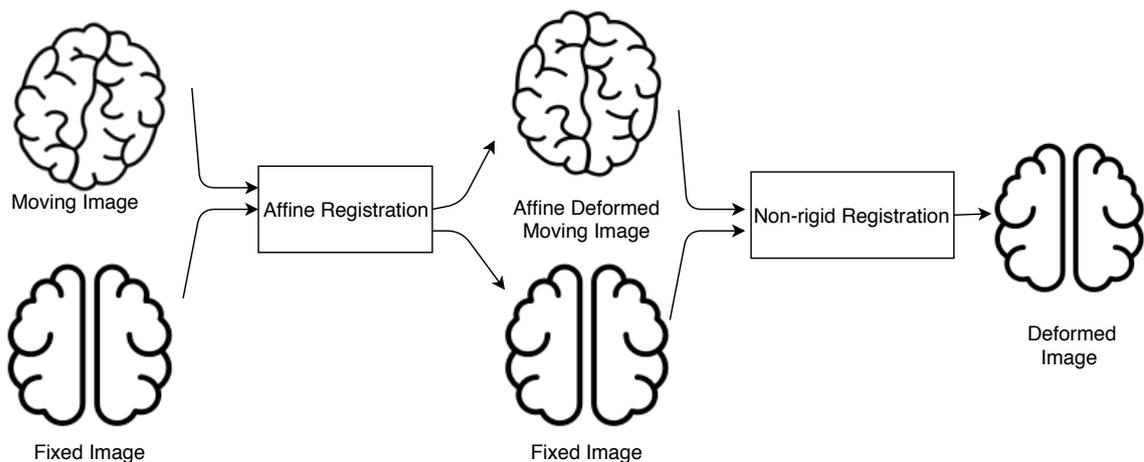
Figure 3.1: Registration Framework

### 3.1.2 Framework

As illustrated in fig. 3.1, this thesis applies a registration framework consisting of affine registration and subsequent non-rigid registration. Affine registration is a global image registration method. It estimates the deformation brought by subject position movement and orientation adjustment. Using affine registration for pre-alignment also simplifies subsequent image registration steps[13]. In this project, mutual information metric is used in affine registration to measure the dissimilarity. With the help of the affine transformation, the moving image can be pre-aligned with the fixed image. The affine transformation model is defined as 3.4

$$T_\mu(x) = A(x - c) + t + c \tag{3.4}$$

where the transformation matrix A is an arbitrary matrix, $c$ and $t$ denote the center of rotation and the translation respectively. The parameter set $\mu$ is composed of matrix components in $A$ and the vector $t$ in this affine registration model. In 3D registration, $\mu$ is a vector of length 12. Using the affine registration, the image can be scaled, rotated, translated and sheared with only the parallelism kept.

After affine registration, the pre-aligned moving image will be further refined using a non-rigid registration. Receiving the pre-aligned moving image and fixed images as inputs, the non-rigid image transformation model is used to estimate local transformation. In the proposed method, a CNN architecture is trained to be used as the non-rigid registration model, which will be further explained in Section 3.2.

## 3.2 RegNet

### 3.2.1 Supervised learning registration framework

The non-rigid registration method we proposed in this project is implemented by a supervised learning framework, shown in fig. 3.2. There are two phases of the deep learning method,
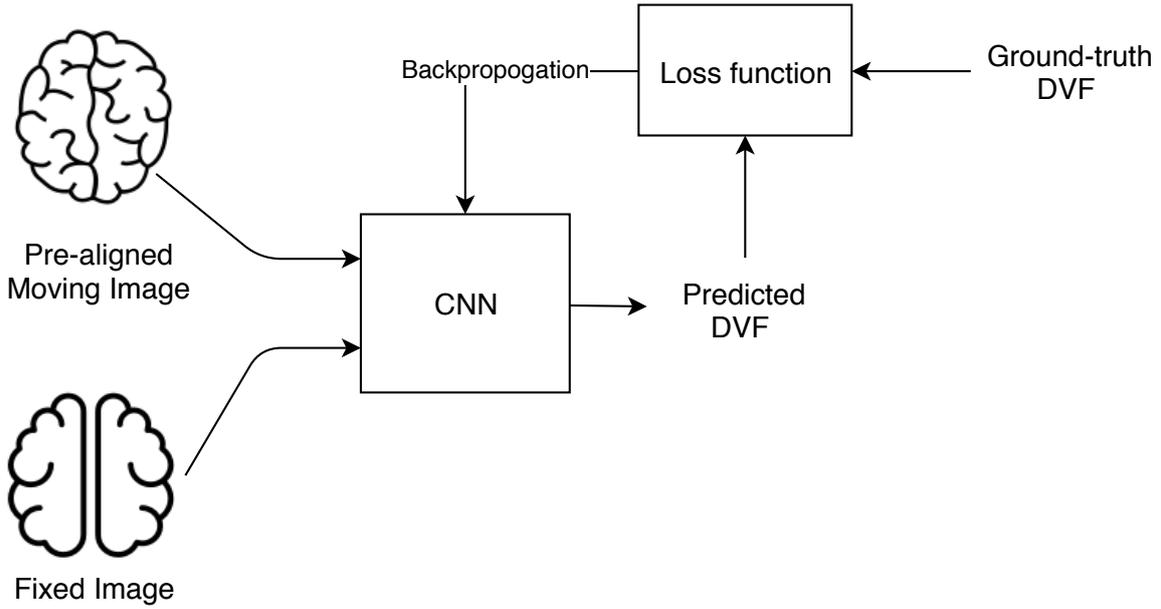
Figure 3.2: A supervised learning registration framework

training phase and testing phase. In the training phase, the CNN receives pairs of images as input and yields a displacement vector field (DVF). During the training, the predicted DVF will be compared with the ground truth DVF using a loss function. The gradient of loss will then be propagated back into the CNN. This feedback will help the CNN to learn how to reduce the difference between the target and the predicted DVF.

In the testing phase, receiving the input of unseen input image pairs, the trained CNN will directly predict the DVF. This predicted DVF is supposed to transform the moving image to be similar to the fixed image.

### 3.2.2   RegNet architecture

The proposed architecture, called RegNet, is a modified version of the convolutional neural network architecture as defined in [41]. It analyzes 3D input patches at two resolution, with the architecture as shown in fig. 3.3. This end-to-end architecture is designed to estimate 3D non-rigid displacement vector field with fixed and moving images given.

In this new network architecture, early fusion technique is applied, which means patches from the moving and fixed image will be concatenated first and then processed by the convolutional layers. Patches with a size of $91 \times 91 \times 91$ at the same location from the input images are stacked into a 4D array to be processed by the CNN.

This new 4D array will be processed by two pipelines. In one pipeline, the central $57 \times 57 \times 57$ arrays will enter the RegNet as the input. In the other pipeline, the $91 \times 91 \times 91$ patch is down-sampled into $43 \times 43 \times 43$ patch first as input. In these two pipelines, two input patches are first processed by a convolutional layer with $3 \times 3 \times 3$ kernel. Using $3 \times 3 \times 3$ kernel as well, 1 convolutional layer with dilation of 2, and 2 convolutional layers with a dilation of 4
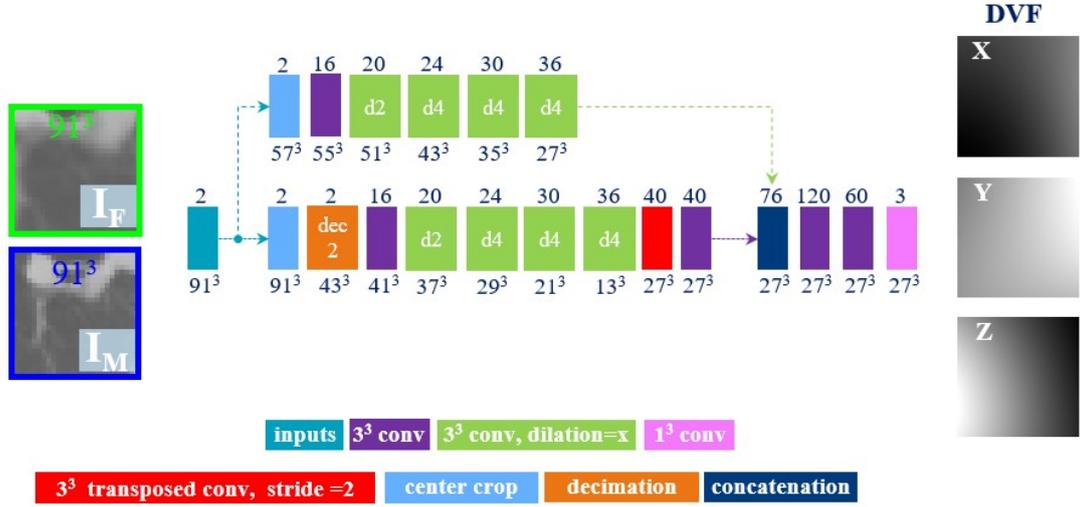
Figure 3.3: Network structure

are followed. After these convolutional layers, the output will be a $27 \times 27 \times 27 \times 36$ array, originated from the original resolution patch, and a $13 \times 13 \times 13 \times 36$ array, originated from the down-sampled patch.

Before concatenating the output of convolutional layers from different pipelines, the impact of downsampling should be eliminated. For the convolutional layer output coming from the down-sampled patch, a transposed convolution layer is added to upsample the result. Another convolutional layer with $3 \times 3 \times 3$ kernel and same padding is added afterwards.

The result of those convolutional layer pipelines can then be merged into one. The concatenated result will be processed by two $3 \times 3 \times 3$ convolutional layers following by a $1 \times 1 \times 1$ convolutional layer. The result will be a $27 \times 27 \times 27 \times 3$ array, predicting the deformation of the center $27 \times 27 \times 27$ voxels in the input patches. All convolutional layers use batch normalization and ReLu activation.

### 3.2.3 Loss function

The loss function applied in this project is composed of Huber error and a bending energy. Huber Error is introduced to measure the mismatch between the predicted result and the target output. The bending energy is added to penalize the undesired transformation. During the training, the predicted DVF is compared with the ground truth, to compute the loss as

14

defined in the eq. (3.5). $L_\delta$ denotes the Huber loss in eq. (3.5). Bending energy is shown as $p_{BE}$, with a weight of $\gamma$, 0.1.

$$L = L_\delta + \gamma p_{BE} \tag{3.5}$$

Huber loss [20] is defined in eq. (3.6), where $x$ denotes the difference between each element in predicted DVF and the ground-truth DVF. $d$ denotes delta, the point where the Huber loss function changes from a quadratic function to a linear function. In our experiment, $d$ is chosen to be 1. Comparing with the squared error loss, Huber loss is robust against the outliers.

$$L_\delta = \begin{cases} 0.5 \cdot x^2 & |x| \leq d \\ 0.5 \cdot d^2 + d \cdot (|x| - d) & \text{otherwise} \end{cases} \tag{3.6}$$

Bending energy is applied in this experiment as a penalty term to solve the ill-posed registration problem. Bending energy shows the energy stored in the shape. The sharper the deformation, the larger the bending energy will be. Thus adding the bending energy as a penalty term will force the displacement vector field to be smooth. The bending energy[39] is defined as eq. (3.7), where V defines the volume of the image, $x, y, z$ points to the three orthogonal coordinates, T is the displacement vector field that RegNet predicts.

$$p_{BE}(\mu) = \frac{1}{V} \iiint_V [\frac{\partial^2 T}{\partial x^2}^2 + \frac{\partial^2 T}{\partial y^2}^2 + \frac{\partial^2 T}{\partial z^2}^2 + 2 * \frac{\partial^2 T}{\partial x \partial y} + 2 * \frac{\partial^2 T}{\partial x \partial z} + 2 * \frac{\partial^2 T}{\partial y \partial z}] dx dy dz$$

(3.7)

### 3.2.4 Multistage framework

In conventional registration method, a hierarchical multi-stage technique is used for optimization because it is less sensitive to local optima and avoid imaging folding[39]. We also adopted this technique into this project, by coarse-to-fine image registration using RegNet[39].

In this framework, there are two RegNet models used, one is the stage-1 model, the other is the stage-2 model. The training procedure of the stage-1 model is introduced in section 3.2.1. For the training of the stage-2 model, the original input patches and corresponding DVF are downsampled first. Following the same procedure shown in fig. 3.2, the stage-2 model is trained using the downsampled image pairs as input and the downsampled DVF as the ground truth.

At test time, input patches also need to be downsampled first. Downsampled patches will flow through the stage-2 model yielding a stage-2 DVF. The output stage-2 DVF will be applied to deform the moving image, creating the pre-deformed moving image. Together with the fixed image, the pre-deformed image will be entering the stage-1 model, the RegNet model at the original resolution. The stage-1 model will predict the DVF at the original resolution to further refine the result. The whole procedure is depicted in fig. 3.4.
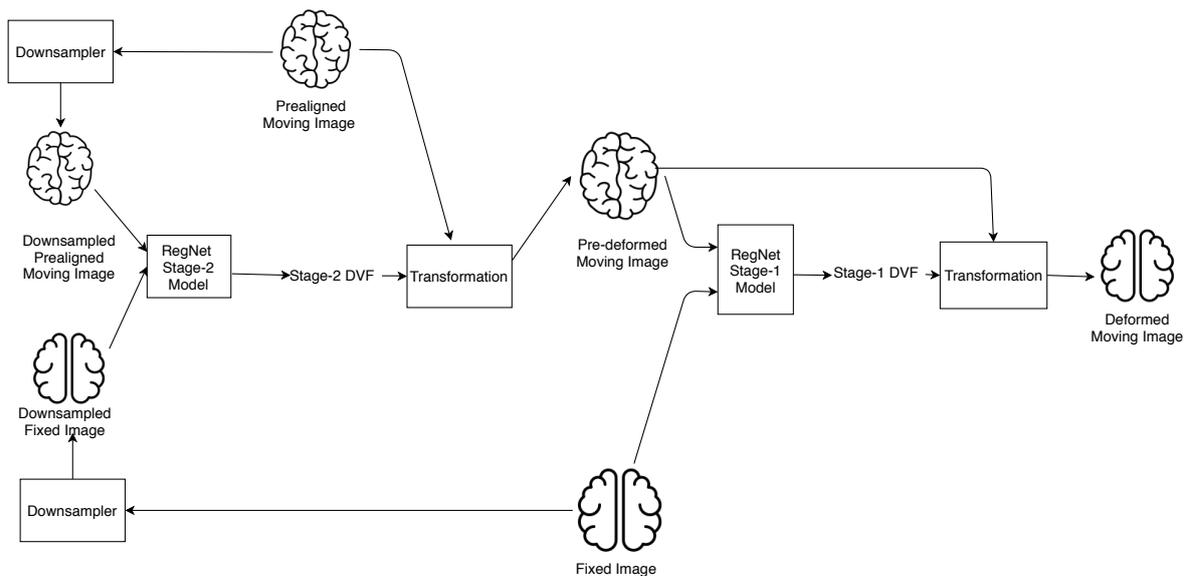
Figure 3.4: Testing phase for multistage framework

## 3.3 Image preparation

Images used in this experiment are obtained from Alzheimers Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI) as discussed in Chapter 2. This section elaborates more details of the data preprocessing techniques, details for DVF generation methods and other data preparation steps.

### 3.3.1 Image preprocessing

The challenge that the MR images analysis facing is coming from non-universal intensity ranges and contrast, and the noises originated from equipment, field strength and operator performance[1]. Thus, preprocessing steps are required before performing image registration. In this experiment, images are preprocessed as follows, standardization, resampling and world matrix adjustment. This section further explains these steps.

#### 3.3.1.1 Image standardization

Intensity standardization is the step of adjusting the intensities of images into a standard scale[1]. Because MR image intensity is pulse-sequence dependent, the same tissue may be visualized in different intensity under different pulse sequence, as mentioned in section 2.1. Without an adjustment, the loss function can not correctly indicate the similarity of two images. It is important to standardize the image intensity for a better registration performance.

In this experiment, z-score is applied for image standardization. Z-score of each voxel is computed by the original intensity value subtracting the mean image intensity obtained from
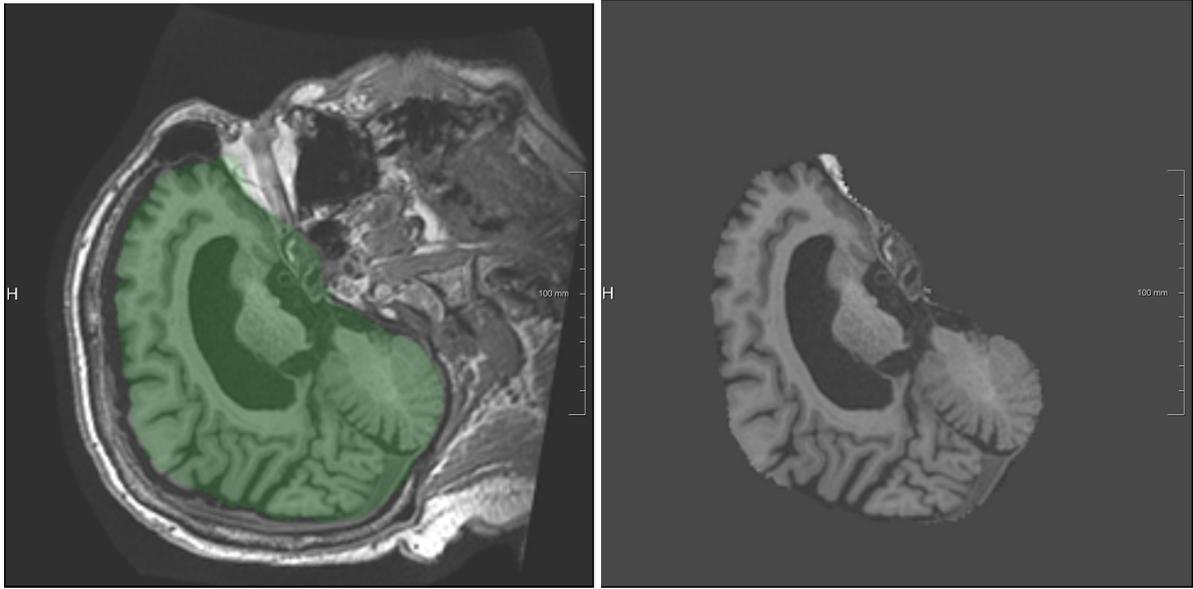
(a) Original brain image         (b) ROBEX stripped image

Figure 3.5: Example of ROBEX skull strip technique

voxels in the whole image, and divided by the standard deviation of the intensity value of the whole image[1].

#### 3.3.1.2   Image resampling and world matrix adjustment

When RegNet is implemented for the image registration, all images are directly transformed into arrays first, regardless of the world matrix and resolution. During the training, RegNet is not taking the resolution and world matrix into account. However, these two are both key factors in registration problem. This means that the proposed method is only capable of solving the registration problem for the images with the same resolution and the same world matrix.

To eliminate the additional affect brought to the registration performance due to the inconsistent voxel size and world matrix. In this experiment, voxel size of all images entering RegNet is resampled to $1mm \times 1mm \times 1mm$. Training images and test images are selected only with the world matrix as,

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

### 3.3.2   Skull stripping technique

During registration, to better focus on intracranial tissues, the region of interests, skull stripping technique is introduced. Robust Brain Extraction (ROBEX)[21] technique is chosen for

17

skull stripping considering its robust performance. *ROBEX* is composed of a discriminative and a generative model for robust brain skull stripping. The discriminative model is first applied on the input image to estimate the brain boundary. A generative model will be following, ensuring that the segmentation contour to be in a plausible shape[21].

The discriminative model applied in *ROBEX* is a Random Forest classifier[8]. The random forest classifier is built by a group of decision trees. During the training, bootstrap aggregation technique is applied, which means each tree is trained separately with different combination of samples[21]. During test time, unknown voxels are travelling down into individual decision tree. Each tree will assign a label either "In the skull region" or "Outside the skull region" to the voxel. Collecting the result from all decision trees, the probability of a voxel belonging to each label is calculated. Till now the voxels probable to be in the skull region are found.

The generative model is built by a point distributive model (PDM), to ensure that the final segmentation result will be in a plausible shape within the probable space defined by the discriminative model. To define the intracranial region, PDM is constructed by landmarks representing the brain shape. The PDM is then deformed using the active shape models technique[11], to detect a plausible shape within the space composed of voxels probable to be in the skull region. A cost function is built by the landmark position and the normalized rate of "In the skull region" voted by decision trees. This cost function evaluates the quality of the predicted shape. Minimizing the cost function, the most probable and plausible shape can be found[21].

As shown in fig. 3.5, with the help of *ROBEX* technique, a whole-brain map can be generated (the green map in fig. 3.5a), thus the whole brain region can be extracted for further usage. In this experiment, the background of the image is adjusted as a default value to exclude impact of the background noise in the registration. In this project, the background region in skull stripped images is set to be 0.

### 3.3.3 Ventricle segmentation

This project uses ventricle maps for registration performance evaluation. Ventricle is a set of cavities in the brain, in which system the cerebrospinal fluid (CSF) is produced[31]. The ventricle system comprises four parts, including lateral ventricles right and left, third ventricle and fourth ventricle. An example of the ventricle segmentation using this method is shown as fig. 3.6. Only the lateral ventricle is segmented for this project.

A semi-automatic segmentation technique, called Snake Evolution, is applied in this experiment[46]. The term snake is describing a closed curve in 3D images or a closed surface in 2D[47]. This semi-automatic segmentation procedure starts with a manual rough segmentation. It defines the initial shape of the snake. The boundary of the snake will continuously grow till a desired status, and the growing velocity of snake is decided by the combination of image-depending velocity and shape-depending velocity. Both types of velocities are perpendicular to the snake.

Two types of image depending velocities are involved in computing the snake growing velocity, propagation velocity and advection velocity. The propagation velocity is depending on the feature map of the region to be segmented [46]. We apply the edge-based feature map in
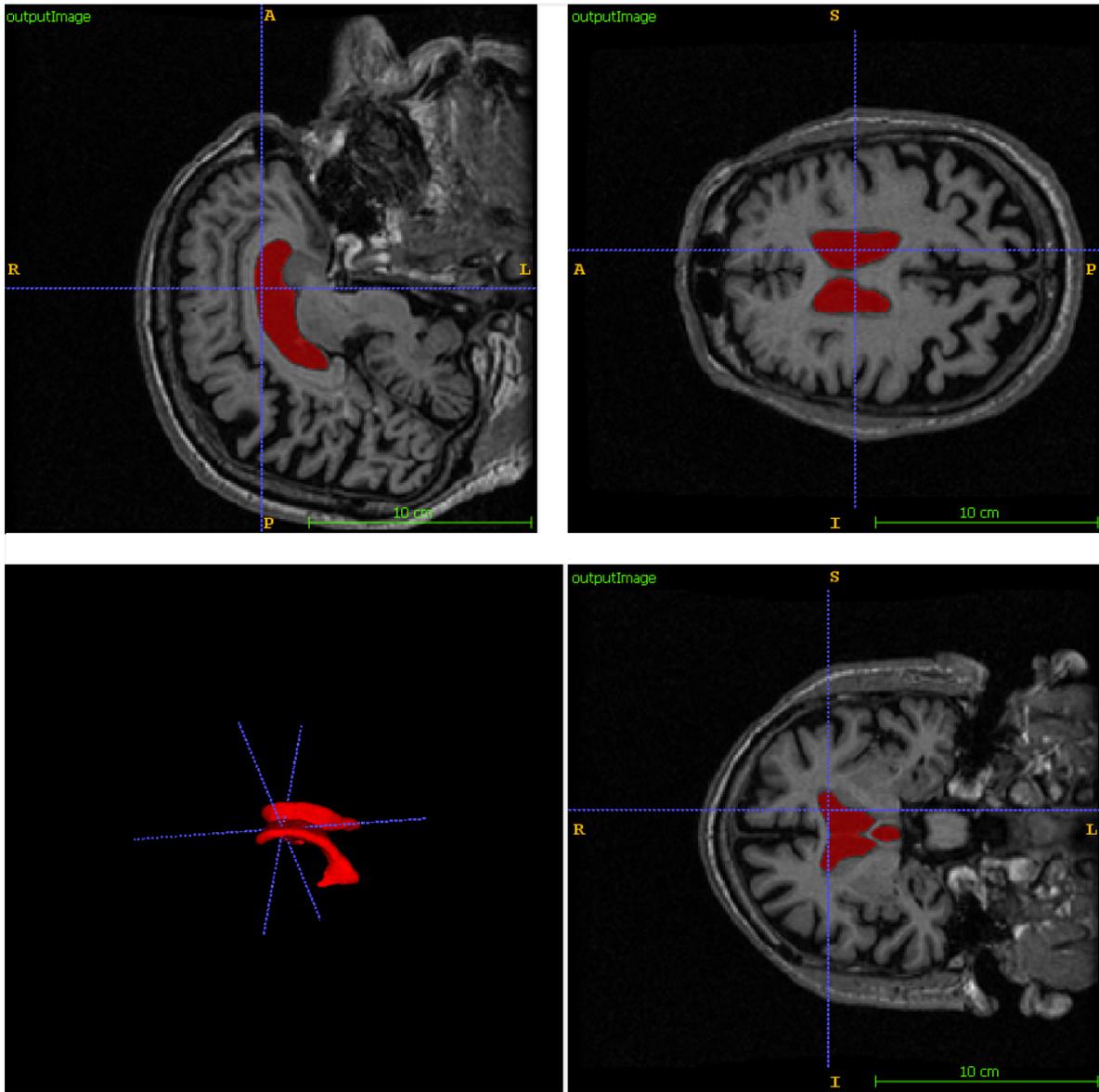
Figure 3.6: Brain MR image with ventricle mask

this project. Because there are relatively large difference on intensity between ventricle and adjacent area, using the edge map can correctly estimate the boundary of the ventricle. Besides propogation velocity, advection velocity is another type of image depending velocity applied. It slows down the snake when the snake is approaching the edge.

The shape-depending velocity is depending on the curvature of the snake. It is working inwards and proportional to the curvature. The shape-depending velocity can effectively prevent the snake leaking into adjacent area[46].
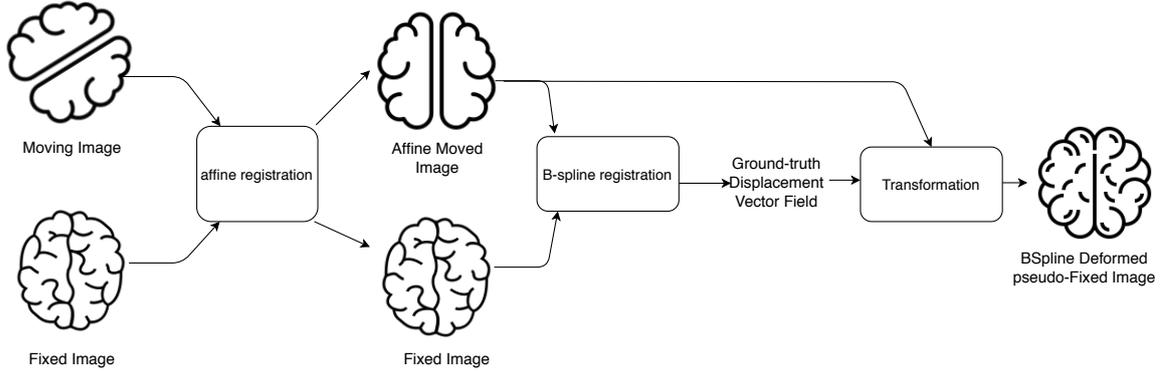
Figure 3.7: B-spline model based DVF generation

### 3.3.4 Ground-truth displacement vector field generation

The challenge of applying supervised learning approach into image registration problem is to generate ground-truth DVF to train the model. In this project, we apply two methods for ground-truth DVF generation, one is to use the DVF created using the B-spline model, the other one is to simulate artificially.

#### 3.3.4.1 B-spline based DVF generation

In this method, the DVF is created using pairs of intrasubject images. Affine registration is first performed between the image pairs. Using the affine registration result, the moving image will then be deformed, to roughly align with the fixed image. B-spline model is then applied to estimate the deformation between the pre-aligned image and the fixed image.

B-spline transformation model is chosen considering its local support and smoothness of the predicted DVF. In B-spline registration model[38], a regular grid is put overlaying the fixed image, this grid is named as the control point grid. Intersections between the control point grid and the fixed image are named as control points. The displacement of individual voxel is computed based on the displacement of adjacent control points.

The B-spline transformation[23] is defined as,

$$T_\mu(x) = x + \sum_{x_k \in N} p_k \beta^3 \left( \frac{x - x_k}{\sigma} \right) \tag{3.8}$$

where $x_k$ denotes control points, $\beta^3$ denotes the multidimensional B-spline polynomial [43] as,

$$\beta^3(a) = \begin{cases} \frac{2}{3} - |a|^2 + \frac{|a|^3}{2} & 0 \le |a| < 1 \\ \frac{(2-|a|)^3}{6} & 1 \le |a| < 2 \\ 0 & 2 \le |a| \end{cases} \tag{3.9}$$

(a) Input image

(b) B-spline deformed image

(c) Difference image

(d) Estimated displacement vector field
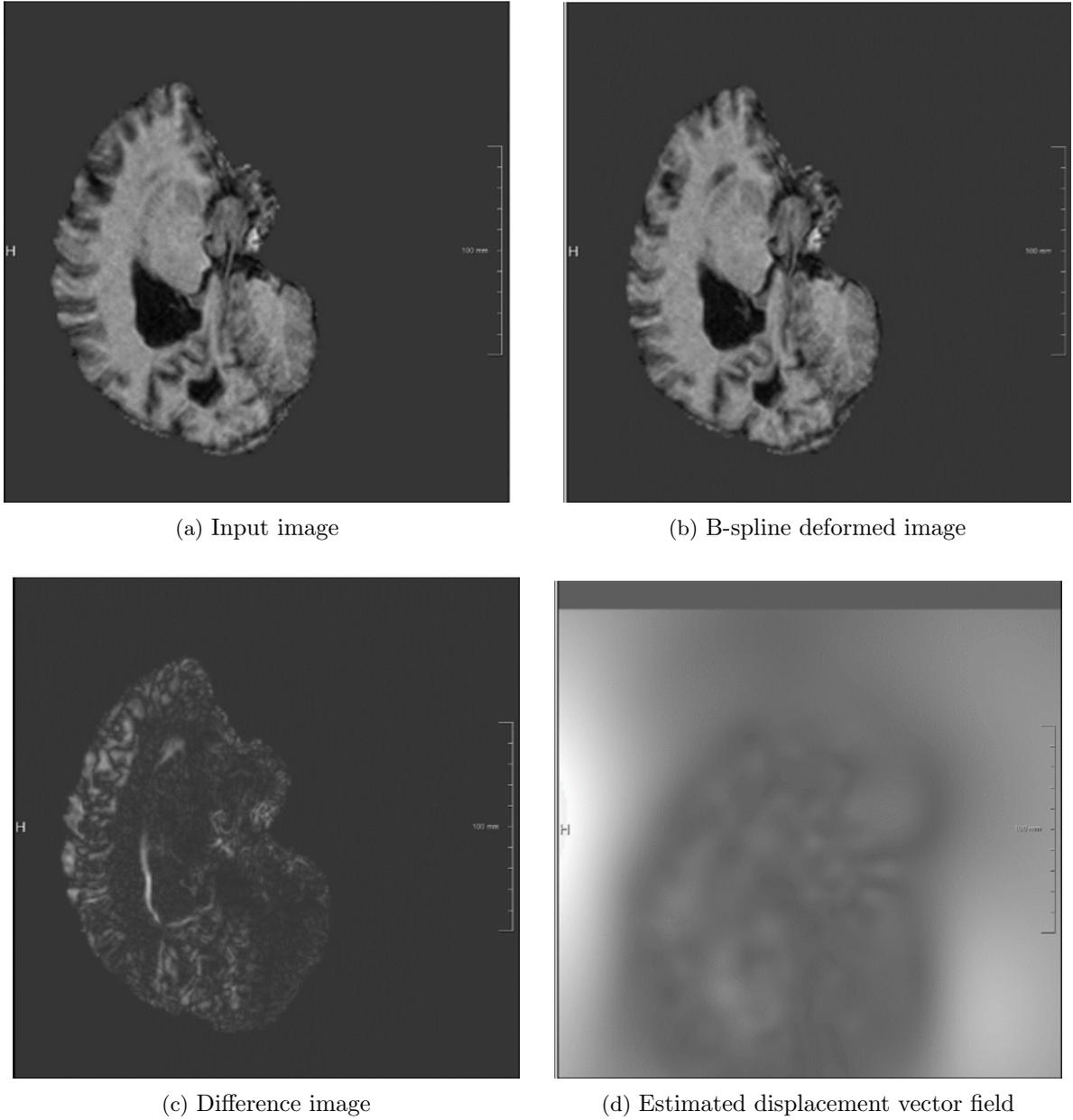
Figure 3.8: Example of the B-spline created DVF

$p_k$ is the B-spline coefficient parameter, composing the $\mu$ with parameters to be optimized, $\sigma$ indicates the B-spline control point spacing and $N$ composes of all control points surrounding $x$ to support predict the displacement of $x$[25][23].

To obtain the optimal B-spline parameters, mutual information is applied as the cost function. Adaptive stochastic gradient descent optimizer is applied for optimization. After obtaining the optimal parameter set $\mu$, the B-spline estimated DVF can be constructed.

For each pair of images in the training set, the DVF between the pre-aligned moving image and the fixed image is estimated using the B-spline model. The B-spline estimated DVF will
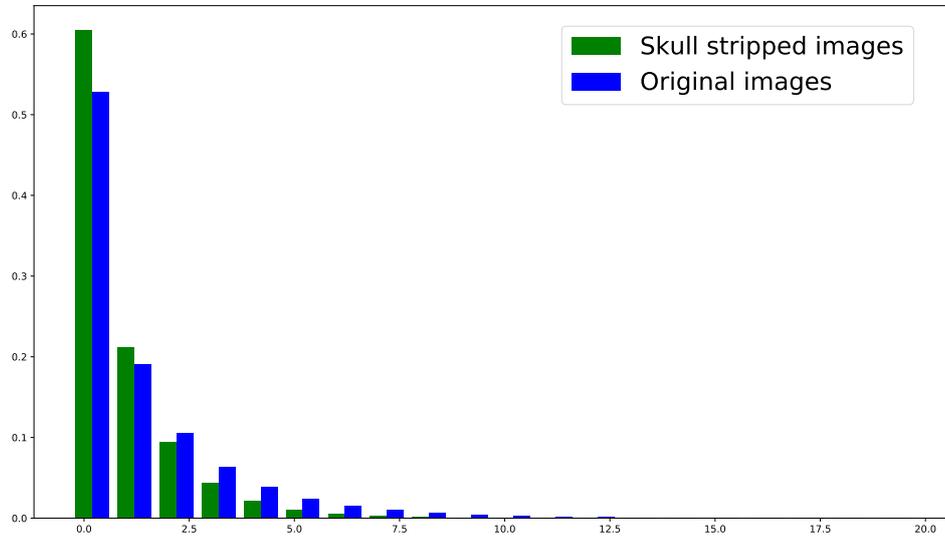
Figure 3.9: Histogram of magnitude of displacement in original images and the skull stripped images
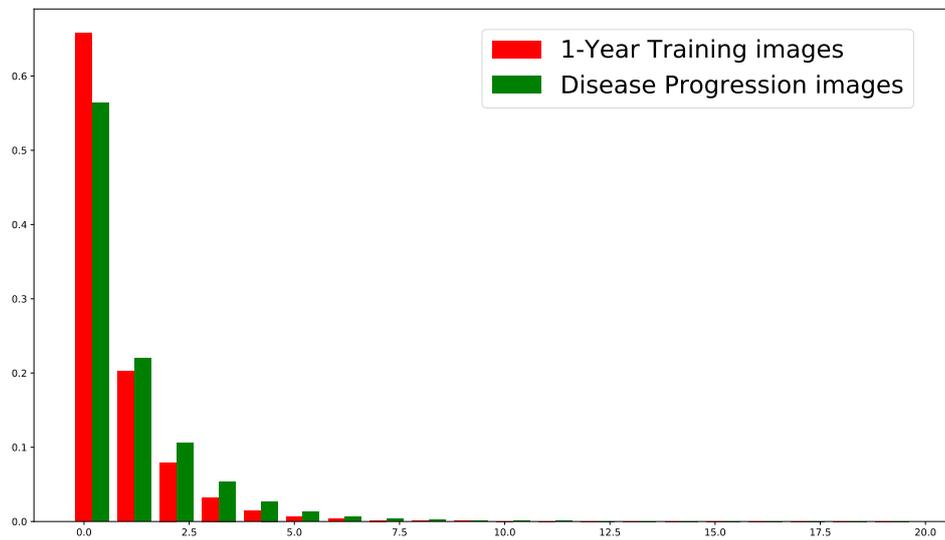


Figure 3.10: Histogram of magnitude of DVF in DP and 1-Year skull-stripped image training set

be utilized to transform the pre-aligned moving image, to create a new pseudo-fixed image being used during the training. With the help of this process, as depicted in fig. 3.7, we can create synthetic brain images that are similar to the original reference image. The B-spline estimated DVF will be used as ground-truth during the training. For every subject in the training set, a pre-aligned moving image, a deformed image and a ground truth DVF are utilized during the training. Examples of the synthetic new reference is shown in fig. 3.8.

Figure 3.9 is depicting the histogram of the magnitude of DVF obtained using the B-spline model of the skull stripped images and the original images in the DP and 1-Year dataset. When estimating the DVF for skull stripped images, only the deformation for the voxels inside the skull region in the fixed image is computed. As can be seen, the average magnitude of the ground-truth deformation of the skull stripped image are smaller than that of the original image. This is because in the region outside the skull, there are organs like neck or shoulder, where position shift may have large impact on the DVF obtained which is out of interest of the brain registration.

We also compare the difference between the DVF from the DP dataset and from the 1-Year dataset. The histogram of the magnitude of B-spline estimated DVF is shown in fig. 3.10. As can be seen, the 1-Year dataset and the DP dataset can be easily differentiated from their different displacement scale, as shown in fig. 3.10, because the deformation between the baseline image and the follow up image in the DP dataset is much larger than that in the 1-Year dataset. This is caused by the development of subject status and relatively longer intervals from the baseline scan to the follow-up scan for subjects in the DP dataset.

Above all, histograms of both datasets show similar right-skewed distribution. Around half of the voxels in the image are deformed with the distance smaller than 1 mm, and the majority of the others have been deformed within 10 mm.


### 3.3.4.2    Artificial DVF generation

To create synthetic DVF, a control point grid is first created. Random values following uniform distribution with a magnitude larger than $m$ are given to each control point in every direction, to create a set of control point deformation parameters. Control points in this method resemble the points with local maximum deformation.

In this method, the parameter $m$ is set to define the maximum displacement used in artificial DVF generation. Since almost all of the displacement between the baseline image and the follow-up image in both dataset is smaller than 10 as shown in fig. 3.10, $m$ is chosen to be 10 in this experiment.

Values at the control points will then be smoothed using a multi-dimensional Gaussian filter. After this, the smoothed parameter set is rescaled into the range from $-m$ to $m$ to keep the maximum displacement smaller than $m$. The obtained normalized parameter set can then be changed into a DVF using the same transformation defined in eq. (3.8). An example of the artificial simulated DVF is shown in Figure 3.11. Figure 3.12 is depicting the histogram of the DVF magnitude. Similar to fig. 3.10, the histogram of the artificial simulated DVF also shows a right-skewed distribution.

(a) Input image



(b) Synthetic deformed image



(c) Difference brought by artificial deformation



(d) Artificial displacement vector field
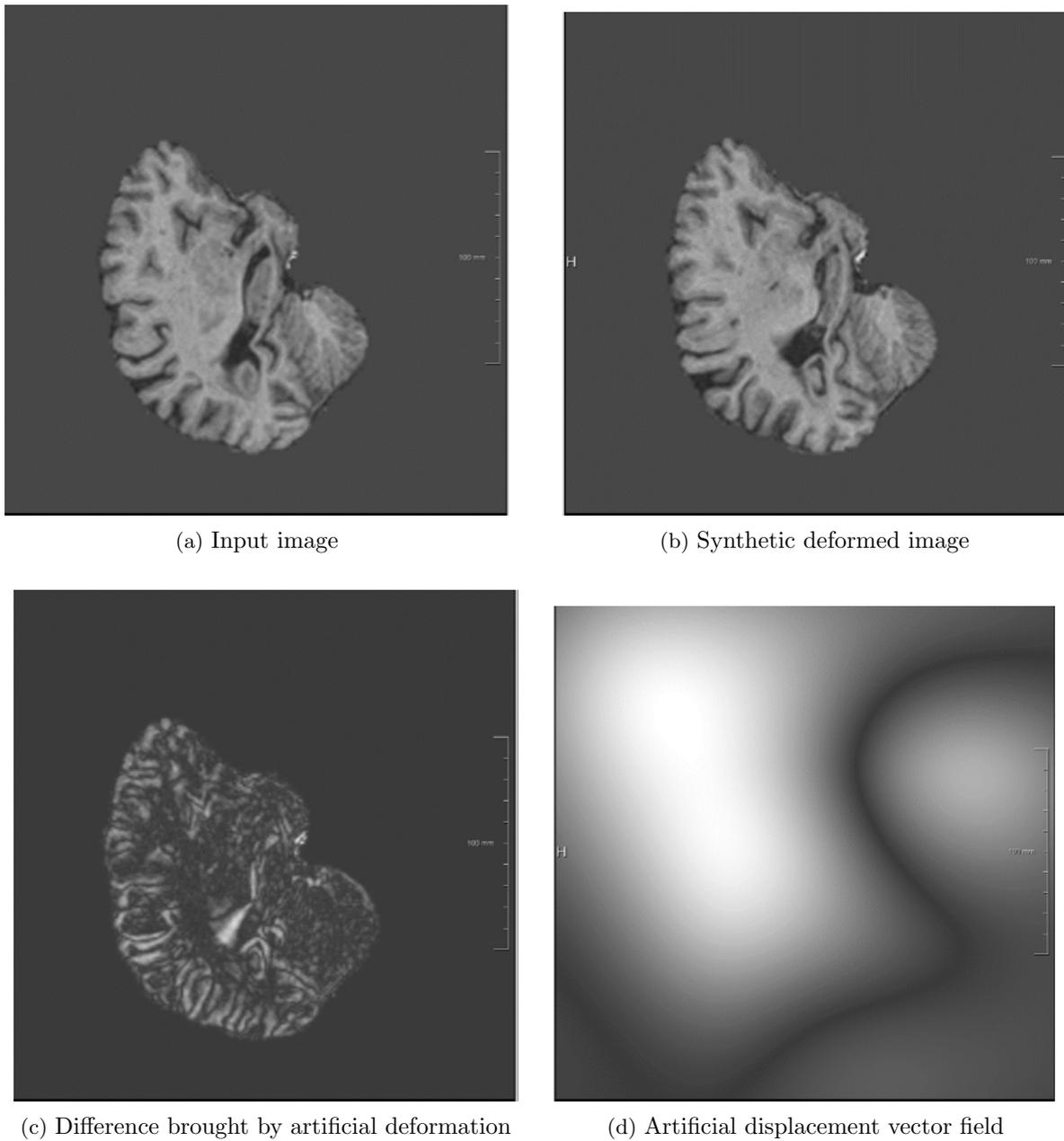
Figure 3.11: Example of artificial displacement vector field creation

### 3.3.5 Data augmentation

Image flipping is applied in this experiment for data augmentation to avoid overfitting. Images are flipped around the sagittal plane, as shown in fig. 3.13. This method takes advantage of brain asymmetry. Normal anatomical and functional asymmetry widely occur in biological anatomical systems. Human brain hemispheres also show some degree of asymmetry[42]. Thus with the help of this flipping technique, it is possible to double the amount of images in the training set with pseudo-natural brain MR images for RegNet training.
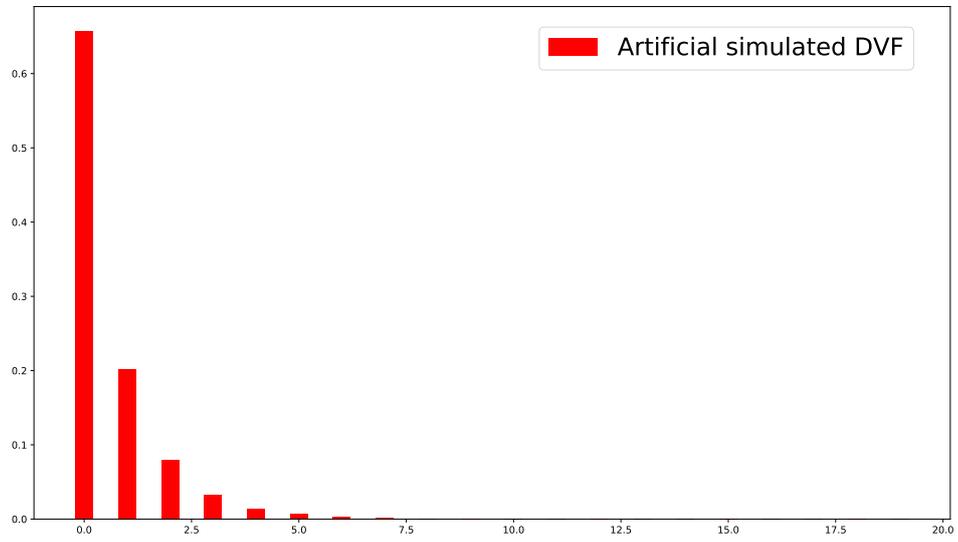
Figure 3.12: Histogram of magnitude of artificial simulated DVF
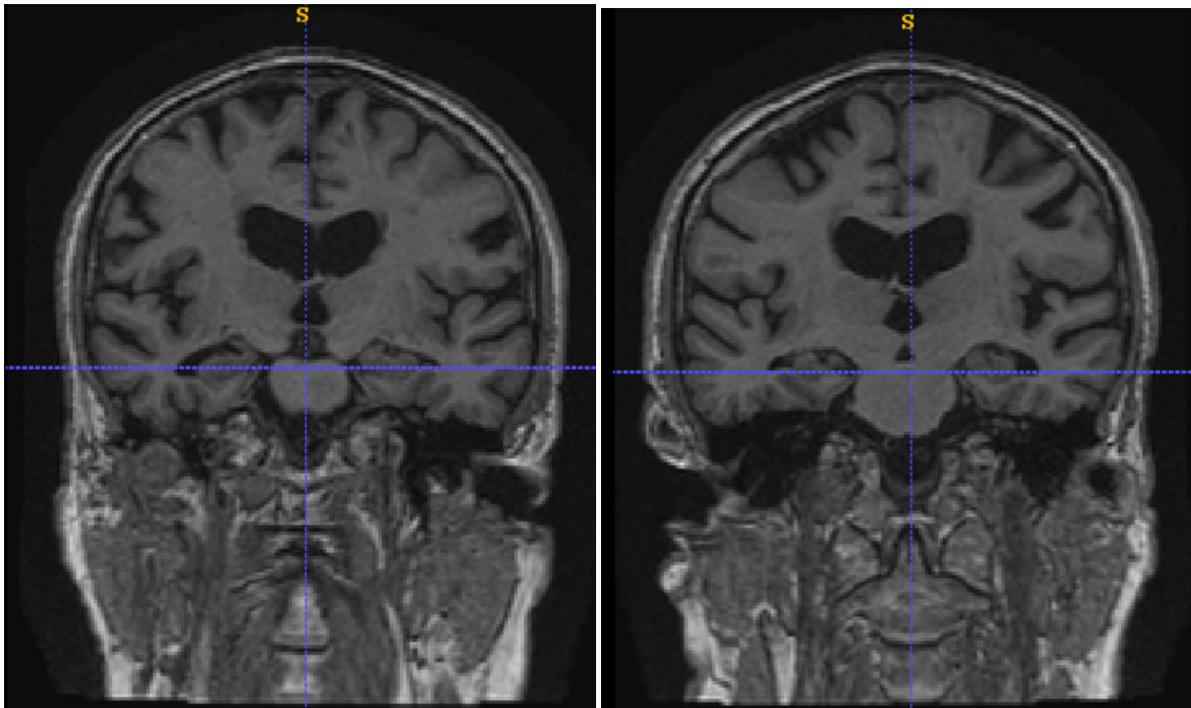


(a) Original brain image

(b) Flipped brain image

Figure 3.13: Image flipping technique example

# Experiment and Result

<span style="font-size:4em;">4</span>

## 4.1 Evaluation criteria

The performance of the proposed registration method is evaluated with intra-patient registration experiments, using segmented anatomical maps, including the hippocampus map from the ADNI database and the semi-manual delineated ventricle map. Then the maps of the moving images are deformed with the obtained DVF from RegNet and then compared with the reference map for the same anatomical structure of the fixed image. The registration performance is assessed by measuring the overlapping using the following 3 evaluation metrics.

### 4.1.1 Dice coefficient

Dice Coefficient (DSC) measures the overlap as defined in eq. (4.1), where X and Y indicate the binary image of the deformed segmentation and the reference segmentation respectively, and $|\cdot|$ indicates the number of voxels that equal to 1. The Dice coefficient is between 0 and 1, which indicates no overlapping or full overlapping respectively.

$$DSC(X,Y) = \frac{2|X \cap Y|}{|X| + |Y|} \tag{4.1}$$

### 4.1.2 Mean surface distance

Mean surface distance, MSD, as defined in eq. (4.2), is the average surface distance between the deformed map surface and the reference map surface. The smaller the MSD, the better the registration performance.

$$MSD_{mean} = \frac{1}{2}[\overline{d}(S_X, S_Y) + \overline{d}(S_X, S_Y)] \tag{4.2}$$

The surface distance $\overline{d}$ is defined in eq. (4.3). It computes the average of the smallest distance from each voxel on the propagated surface to every voxel on the reference surface. X and Y respectively indicate the binary map of the same anatomical structure in the deformed moving image and the fixed image. $S_X$ and $S_Y$ are the set of voxels on the surface of $X$ and $Y$. $d$ computes the Euclidean distance between two voxels.

$$\overline{d}(S_X, S_Y) = \frac{1}{S_X} \sum_{x \in S_X} \min_{y \in S_Y} d(x,y) \tag{4.3}$$

27

### 4.1.3 Hausdorff distance with 99 percentage

Hausdorff distance (HD) is measuring the maximum of the smallest distance from each voxel on the propagated surface $S_X$ to every voxel on the reference surface $S_Y$. The smaller the HD, the better the registration performance. HD is defined as,

$$HD = \max d_H(S_X, S_Y), d_H(S_Y, S_X) \tag{4.4}$$

$$d_H(S_X, S_Y) = \max_{x \in S_X} \min_{y \in S_Y} d(x, y) \tag{4.5}$$

The 99 percentile of the largest surface distance is chosen in evaluation, this is more robust comparing with the original Hausdorff distance definition because Hausdorff distance is sensitive to noise and outliers in segmentation[21].

## 4.2 Experiment

### 4.2.1 Training data

Images used in this experiment are chosen according to 2 criteria, as defined in section 2.3, 1-year training dataset (131 subjects for training, including 12 subjects used for validation) and the DP training dataset (125 subjects for training, including 17 subjects for validation). For each subject in the training dataset, there are one B-spline estimated DVF and one artificial simulated DVF generated. For image pairs created by the same DVF creating method , there are original images and skull stripped images available.

In total there are six different combinations of training data. Among that six training dataset, four sets are only using the B-spline estimated DVF as ground truth. One of these four is composed of original images from DP and 1Y dataset, another one contains the skull stripped images from DP and 1Y dataset, the last two contain skull stripped images either only from the 1-Year dataset or only from the DP dataset. Besides, there is one model trained only with the artificial simulated DVF, also another model trained with a combination of the artificial simulated DVF and the B-spline estimated DVF. Detailed information regarding the training data can be found in table 4.1. One-stage and multistage framework are utilized for all of the training set. In total there are twelve models to be evaluated.

### 4.2.2 Training setup

During the RegNet training, there are 20 images loaded into the memory in order to handle the memory efficiently. 15 pairs of patches with size of $91 \times 91 \times 91$ are truncated from these 20 images. To ensure that the training data are balanced in the DVF magnitude, 3 labels (class 0, class 1, class 2) are given to those patches, based on the magnitude of DVF of the central voxel in the patch. Class 0 contains patches with the central voxel DVF smaller than 1. Class 1 contains patches with the central voxel DVF smaller than 3 but larger than 1. Class 2 contains patches with the central voxel DVF smaller than 10 but larger than 3. For

| Experiment | Training Image | Ground-Truth DVF Type | Framework | Test Image |
|---|---|---|---|---|
| DP-S-B-One | DP Skull Images | B-spline Estimated DVF | One-stage | DP test Skull Images |
| DP-S-B-Mul | DP Skull Images | B-spline Estimated DVF | Multistage | DP test Skull Images |
| 1Y-S-B-One | DP Skull Images | B-spline estimated DVF | One-stage | 1Y test Skull Images |
| 1Y-S-B-Mul | DP Skull Images | B-spline estimated DVF | Multistage | 1Y test Skull Images |
| DP-1Y-O-B-One | DP and 1Y Original Images | B-spline estimated DVF | One-stage | DP and 1Y test Skull images |
| DP-1Y-O-B-Mul | DP and 1Y Original Images | B-spline estimated DVF | Multistage | DP and 1Y test Skull images |
| DP-1Y-S-B-One | DP and 1Y Skull Images | B-spline estimated DVF | One-stage | DP and 1Y test Skull images |
| DP-1Y-S-B-Mul | DP and 1Y Skull Images | B-spline estimated DVF | Multistage | DP and 1Y test Skull images |
| DP-1Y-S-A-One | DP and 1Y Skull Images | Artificial simulated DVF | One-stage | DP and 1Y test Skull images |
| DP-1Y-S-A-Mul | DP and 1Y Skull Images | Artificial simulated DVF | Multistage | DP and 1Y test Skull images |
| DP-1Y-S-A-B-One | DP and 1Y Skull Images | Artificial simulated DVF and B-spline estimated DVF | One-stage | DP and 1Y test Skull images |
| DP-1Y-S-A-B-Mul | DP and 1Y Skull Images | Artificial simulated DVF and B-spline estimated DVF | Multistage | DP and 1Y test Skull images |

Table 4.1: RegNet model list, models are trained using Disease Progression (DP) dataset and 1-Year (1Y) dataset, some trained with the original (O) images and some with the skull stripped images (S), using ground truth DVF created via B-spline model (B) or by artificial simulation (A), either with the one-stage framework (One) or the multistage framework (Mul)

the skull stripped training set, only when the center voxel is inside the skull region the patch will be chosen for training.

We augment the data with image flipping technique, as described in section 3.3.5, and noises. One Gaussian noise with the standard deviation of 0.05 is added individually to each voxel. Another random constant value is added onto every input training patch. This is to resemble the difference of intensity scale between the fixed image and the moving image during test time.

### 4.2.3  Test setup

In total, 52 pairs of images from the 1-year image set (26 subjects, with delineated ventricle and hippocampus map) and the DP image set (26 subjects, with delineated ventricle maps) are kept for registration evaluation.

The DVF that the RegNet predicts at once can only cover the voxels in the central region of the input patches. To create DVF for the whole image, during test time, sliding window is shifting over the training images, and the result DVF are stitched into one complete DVF for the whole image. To avoid sharp changes in the output DVF, the movement of the sliding window is designed deliberately so that the output DVF for each patch pairs will be overlapping with the DVF estimated by adjacent input patch pairs.

## 4.3   Software

This project is implemented in *Python* using *TensorFlow* backend . The optimization is performed by Adam, with a base learning rate starting at 0.001, and a decay factor of 0.9 in each epoch, in this way the convergence rate will be improved.

Registration Toolbox *elastix* is applied for ground-truth displacement vector field generation and prior affine registration. Software *ITK-SNAP* is applied for ventricle segmentation. Toolbox *ROBEX* is utilized for skull stripping. Software *MeVisLab* is used for image visualization in this report.

## 4.4   Result

The proposed RegNet model is compared with the conventional registration method using the B-spline model. Table 4.2, table 4.3 and table 4.4 give an overview of average Dice coefficients, MSD and HD obtained with the proposed RegNet models, B-spline model, and a baseline of affine registration. Table 4.2 evaluates the method on DP test images, using ventricle maps. Table 4.3 and table 4.4 report the result based on 1-Year test images and evaluated by ventricle and hippocampus. Boxplots plotted in fig. 4.1, fig. 4.2 and fig. 4.3 give a quantitative analysis for those tests mentioned above. Examples of the registration result and the difference between the result and the fixed images are illustrated in fig. 4.4, fig. 4.5, fig. 4.6 and fig. 4.7. In all RegNet models, the image flipping technique is applied due to its performance improvement. In an early experiment DP-1Y-S-B , the Dice coefficient for ventricle is increased from 0.82 to 0.844 (p-value =0.004) when the image flipping technique is applied, for an evaluation based on DP test images.

In evaluation based on DP test images, the best RegNet result is achieved by the model DP-1Y-S-B-Mul. It achieves a Dice coefficient of 0.87, with an MSD of 0.54 and a 99%HD of 2.893 for ventricle, on par with the B-spline performance. DP-1Y-S-A-B-Mul, the RegNet model trained with the artificial simulated DVF combined with the B-spline estimated DVF also achieves a comparable good result. The superior performance of DP-1Y-S-B-Mul is also visible on fig. 4.5g. When the registration is not properly done, the edge of ventricle can be pointed out, as shown in fig. 4.5f for the result of DP-1Y-O-B-Mul model. In contrast, this mismatch is almost unnoticeable in fig. 4.5g. The best model, DP-1Y-S-B-Mul is trained using DP and 1-year skull stripped images and using B-spline estimated DVF as ground truth. We also observe that other models trained with the skull stripped images outperform those trained with original images; models using the B-spline estimated DVF as ground truth during the training can obtain better performance comparing with the ones using artificial

| DP-Test Result - Ventricle | | | |
| --- | --- | --- | --- |
| | Dice -Ventricle | MSE-Ventricle | HD-Ventricle |
| Before non-rigid Registration | 80.0% ( 0.084 ) | 0.87 ( 0.332 ) | 2.76 ( 1.065 ) |
| Elastix Bspline | 90.9% ( 0.024 ) | 0.39 ( 0.060 ) | 2.23 ( 1.042 ) |
| DP-1Y-O-B-One | 85.1% ( 0.043 ) | 0.62 ( 0.102 ) | 2.56 ( 1.121 ) |
| DP-1Y-O-B-Mul | 84.5% ( 0.043 ) | 0.63 ( 0.113 ) | 2.65 ( 1.103 ) |
| DP-1Y-S-B-One | 86.8% ( 0.040 ) | 0.55 ( 0.091 ) | 2.37 ( 0.978 ) |
| DP-1Y-S-B-Mul | 87.6% ( 0.032 ) | 0.52 ( 0.059 ) | 2.40 ( 1.110 ) |
| DP-S-B-One | 84.8% ( 0.046 ) | 0.64 ( 0.120 ) | 2.56 ( 1.002 ) |
| DP-S-B-Mul | 86.8% ( 0.033 ) | 0.55 ( 0.058 ) | 2.42 ( 1.036 ) |
| DP-1Y-S-A-One | 84.3% ( 0.055 ) | 0.67 ( 0.170 ) | 2.47 ( 0.991 ) |
| DP-1Y-S-A-Mul | 85.2% ( 0.051 ) | 0.62 ( 0.147 ) | 2.45 ( 0.955 ) |
| DP-1Y-S-A-B-One | 86.5% ( 0.040 ) | 0.56 ( 0.089 ) | 2.53 ( 1.104 ) |
| DP-1Y-S-A-B-Mul | 87.0% ( 0.036 ) | 0.55 ( 0.081 ) | 2.52 ( 1.208 ) |

Table 4.2: Dice, MSD and HD Result of registration on Disease Progression Test set using ventricle for evaluation

| 1-YEAR-Test Result Ventricle | | | |
| --- | --- | --- | --- |
| | Dice Ventricle | MSE Ventricle | HD Ventricle |
| Before non-rigid Registration | 86.3% ( 0.049 ) | 0.51 ( 0.159 ) | 2.02 ( 0.616 ) |
| Elastix Bspline | 89.9% ( 0.025 ) | 0.36 ( 0.043 ) | 1.90 ( 0.637 ) |
| DP-1Y-O-B-One | 86.5% ( 0.032 ) | 0.48 ( 0.078 ) | 2.04 ( 0.532 ) |
| DP-1Y-O-B-Mul | 85.2% ( 0.045 ) | 0.52 ( 0.125 ) | 2.19 ( 0.572 ) |
| DP-1Y-S-B-One | 87.7% ( 0.030 ) | 0.45 ( 0.073 ) | 1.94 ( 0.584 ) |
| DP-1Y-S-B-Mul | 87.6% ( 0.030 ) | 0.45 ( 0.067 ) | 1.95 ( 0.586 ) |
| 1Y-S-B-One | 87.3% ( 0.031 ) | 0.46 ( 0.074 ) | 1.97 ( 0.565 ) |
| 1Y-S-B-Mul | 87.1% ( 0.031 ) | 0.47 ( 0.066 ) | 1.99 ( 0.585 ) |
| DP-1Y-S-A-One | 87.2% ( 0.037 ) | 0.47 ( 0.099 ) | 1.98 ( 0.594 ) |
| DP-1Y-S-A-Mul | 87.4% ( 0.034 ) | 0.46 ( 0.090 ) | 1.98 ( 0.601 ) |
| DP-1Y-S-A-B-One | 87.5% ( 0.031 ) | 0.45 ( 0.077 ) | 1.96 ( 0.579 ) |
| DP-1Y-S-A-B-Mul | 87.6% ( 0.031 ) | 0.45 ( 0.075 ) | 1.95 ( 0.581 ) |

Table 4.3: Dice, MSD and HD Result of registration on 1-year Test set using ventricle for evaluation

simulated DVF. In fig. 4.1, it is not hard to notice that the models using multistage framework can clearly outperform the models with one-scale framework for test using DP images with ventricle for evaluation.

In 1-Year test using ventricle for evaluation, the DP-1Y-S-B-Mul model also gives good result, following by DP-1Y-S-A-B-Mul and DP-1Y-S-B-One as shown in table 4.3. Table 4.3 reveals that the RegNet model using the B-spline estimated DVF as ground truth can outperform the one using artificial stimulated DVF. For instance, DP-1Y-S-B-One and DP-1Y-S-A-One are both trained with the same amount of skull-stripped images and one-stage framework. Comparing with DP-1Y-S-A-One, DP-1Y-S-B-One uses B-spline DVF as ground truth, and

| 1-YEAR-Test Result Hippocampus | | | |
|---|---|---|---|
| | Dice Hippocampus | MSD Hippocampus | HD Hippocampus |
| Before non-rigid Registration | 66.9% ( 0.047 ) | 0.55 ( 0.054 ) | 1.75 ( 0.389 ) |
| Elastix Bspline | 67.6% ( 0.039 ) | 0.54 ( 0.039 ) | 1.73 ( 0.381 ) |
| DP-1Y-O-B-One | 66.6% ( 0.046 ) | 0.56 ( 0.052 ) | 1.80 ( 0.345 ) |
| DP-1Y-O-B-Mul | 65.5% ( 0.055 ) | 0.57 ( 0.076 ) | 1.86 ( 0.339 ) |
| DP-1Y-S-B-One | 66.9% ( 0.045 ) | 0.55 ( 0.050 ) | 1.74 ( 0.385 ) |
| DP-1Y-S-B-Mul | 66.7% ( 0.047 ) | 0.55 ( 0.045 ) | 1.81 ( 0.381 ) |
| 1Y-S-B-One | 66.2% ( 0.046 ) | 0.56 ( 0.060 ) | 1.83 ( 0.385 ) |
| 1Y-S-B-Mul | 65.5% ( 0.048 ) | 0.57 ( 0.060 ) | 1.88 ( 0.380 ) |
| DP-1Y-S-A-One | 66.7% ( 0.049 ) | 0.55 ( 0.055 ) | 1.76 ( 0.339 ) |
| DP-1Y-S-A-Mul | 66.8% ( 0.047 ) | 0.55 ( 0.051 ) | 1.78 ( 0.335 ) |
| DP-1Y-S-A-B-One | 66.9% ( 0.045 ) | 0.55 ( 0.051 ) | 1.76 ( 0.383 ) |
| DP-1Y-S-A-B-Mul | 66.8% ( 0.046 ) | 0.55 ( 0.054 ) | 1.77 ( 0.410 ) |

Table 4.4: Dice, MSD and HD Result of registration on 1-year Test set using hippocampus for evaluation

obtains lower MSD and 99%HD for ventricle and hippocampus in 1-Year test set. Similar to the result based on DP images, DP-1Y-S-B-One also performs slightly better than the DP-1Y-O-B-One with the only difference of using skull stripped images for training. We also compare the result of the multistage framework and the one-stage framework. However, there is no evident difference between the one-stage and multistage models as shown in fig. 4.3 in the test based on 1-Year test images.

In the test using 1-Year images, when the hippocampus is used for evaluation, all RegNet models can only obtain a Dice coefficient almost equal with the pre-alignment. There are also little differences between models as shown in table 4.4 using hippocampus for evaluation.

Moreover, comparing with the conventional method, it takes RegNet much shorter time to finish the registration problem. On average it takes the conventional method around 7 minutes to finish the task and it only takes less than 2 minutes for RegNet.
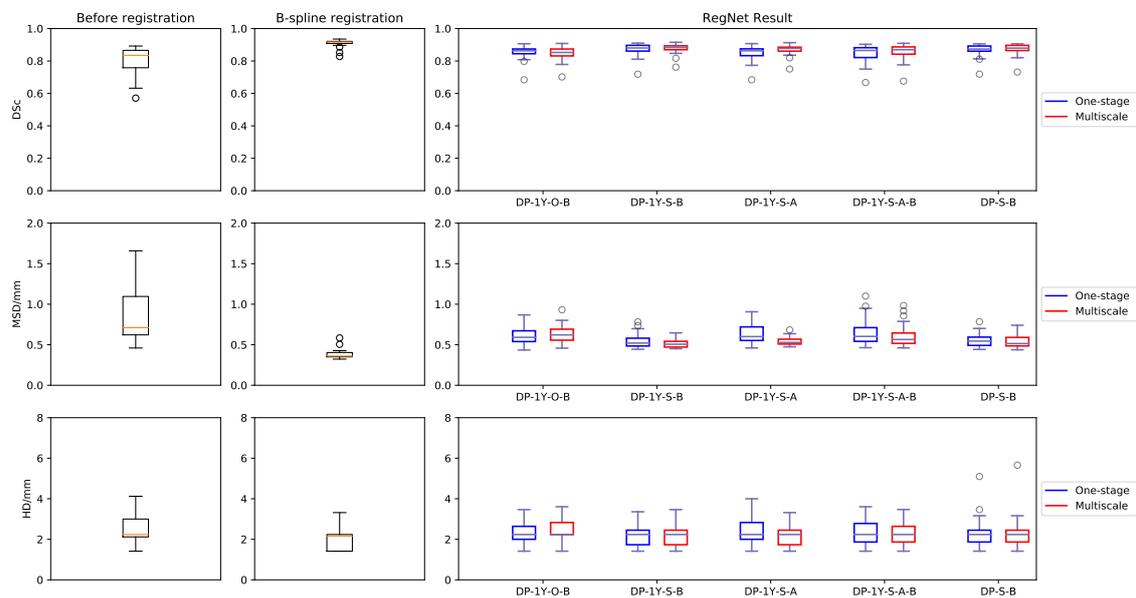
Figure 4.1: Box-plot result of test using DP test images and ventricle for evaluation
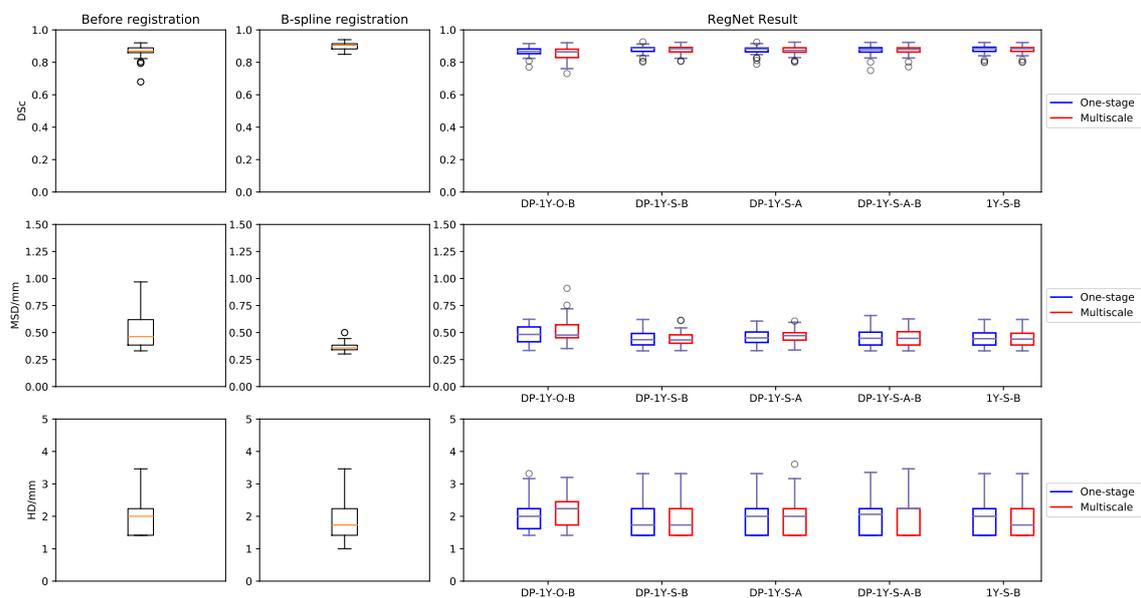


Figure 4.2: Box-plot result of test using 1-Year test images and ventricle for evaluation
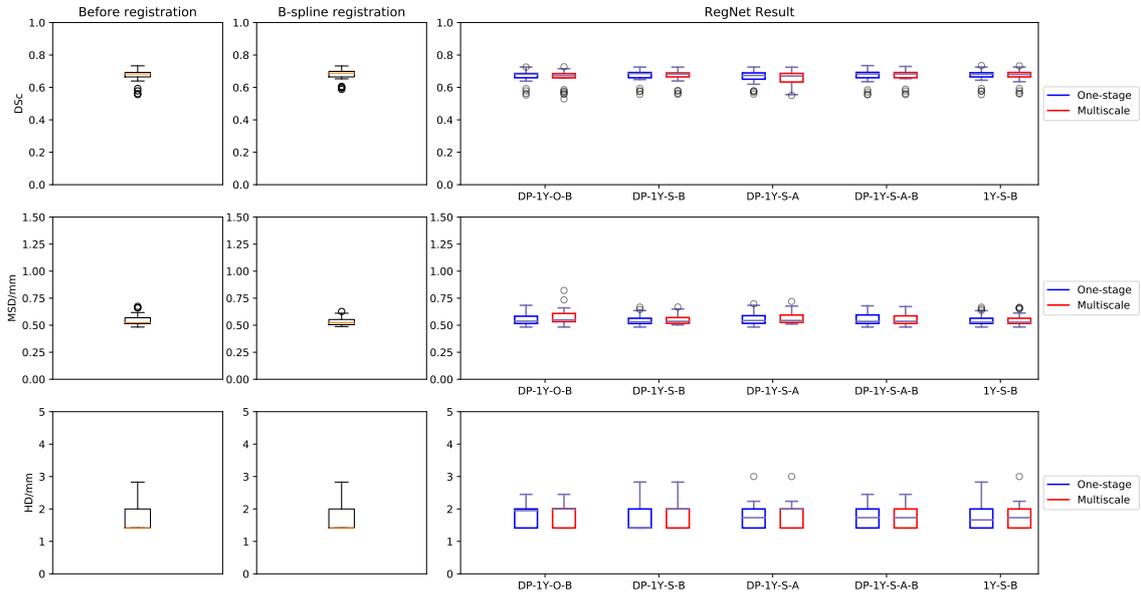
Figure 4.3: Box-plot result of test using 1-Year test images and hippocampus for evaluation



(a) DP-1Y-O-B-One deformed image

(b) DP-1Y-S-B-One deformed image

(c) DP-S-B-One deformed image

(d) DP-1Y-S-A-One deformed image

(e) DP-1Y-S-A-B-One deformed image

(f) DP-1Y-O-B-One mismatch

(g) DP-1Y-S-B-One mismatch

(h) DP-S-B-One mismatch

(i) DP-1Y-S-A-One mismatch

(j) DP-1Y-S-A-B-One mismatch
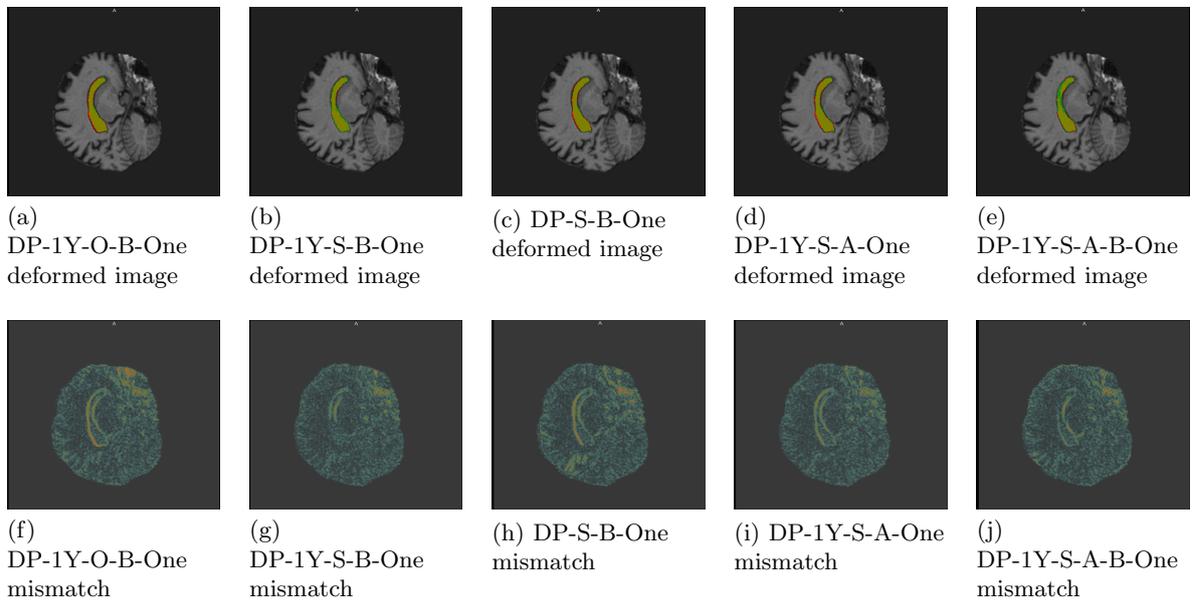
Figure 4.4: One-stage RegNet deformed moving image (Ventricle map in fixed image in green, deformed ventricle map in red) in DP test images and intensity difference with the corresponding fixed image (Minor mismatch in blue, larger mismatch in red)

(a) DP-1Y-O-B-Mul deformed image  (b) DP-1Y-S-B-Mul deformed image  (c) DP-S-B-Mul deformed image  (d) DP-1Y-S-A-Mul deformed image  (e) DP-1Y-S-A-B-Mul deformed image

(f) DP-1Y-O-B-Mul mismatch  (g) DP-1Y-S-B-Mul mismatch  (h) DP-S-B-Mul mismatch  (i) DP-1Y-S-A-Mul mismatch  (j) DP-1Y-S-A-B-Mul mismatch
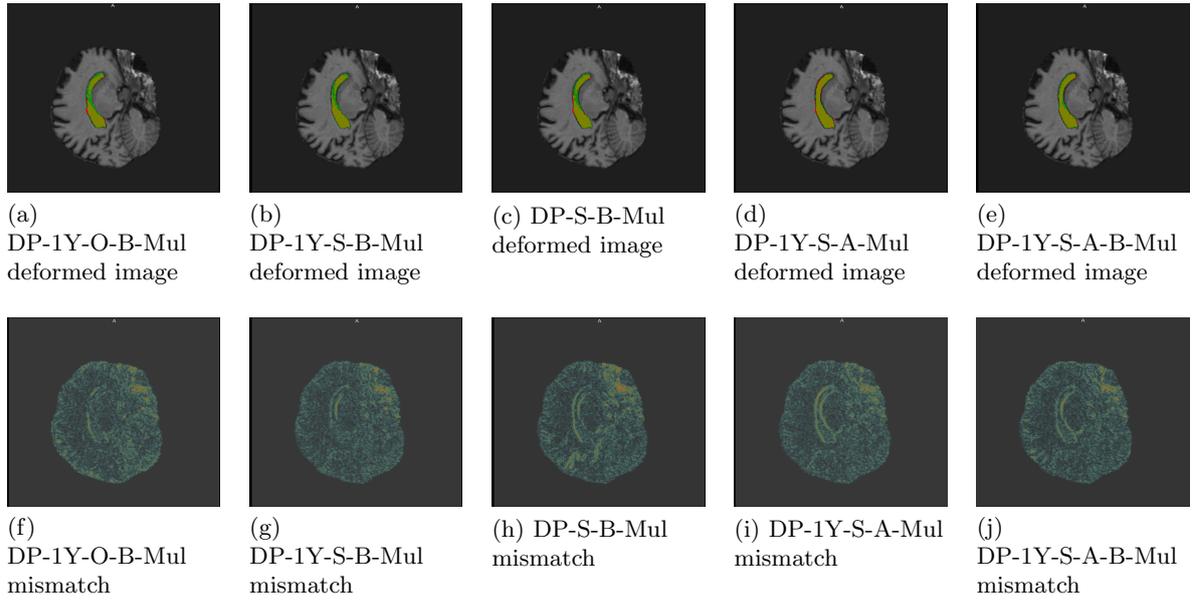
Figure 4.5: Multistage RegNet deformed moving image (Ventricle map in fixed image in green, deformed ventricle map in red) in DP test images and mismatch with the corresponding fixed image (Minor mismatch in blue, larger mismatch in red)



(a) DP-1Y-O-B-One deformed image  (b) DP-1Y-S-B-One deformed image  (c) 1Y-S-B-One deformed image  (d) DP-1Y-S-A-One deformed Image  (e) DP-1Y-S-A-B-One deformed image

(f) DP-1Y-O-B-One mismatch  (g) DP-1Y-S-B-One mismatch  (h) 1Y-S-B-One mismatch  (i) DP-1Y-S-A-One mismatch  (j) DP-1Y-S-A-B-One mismatch
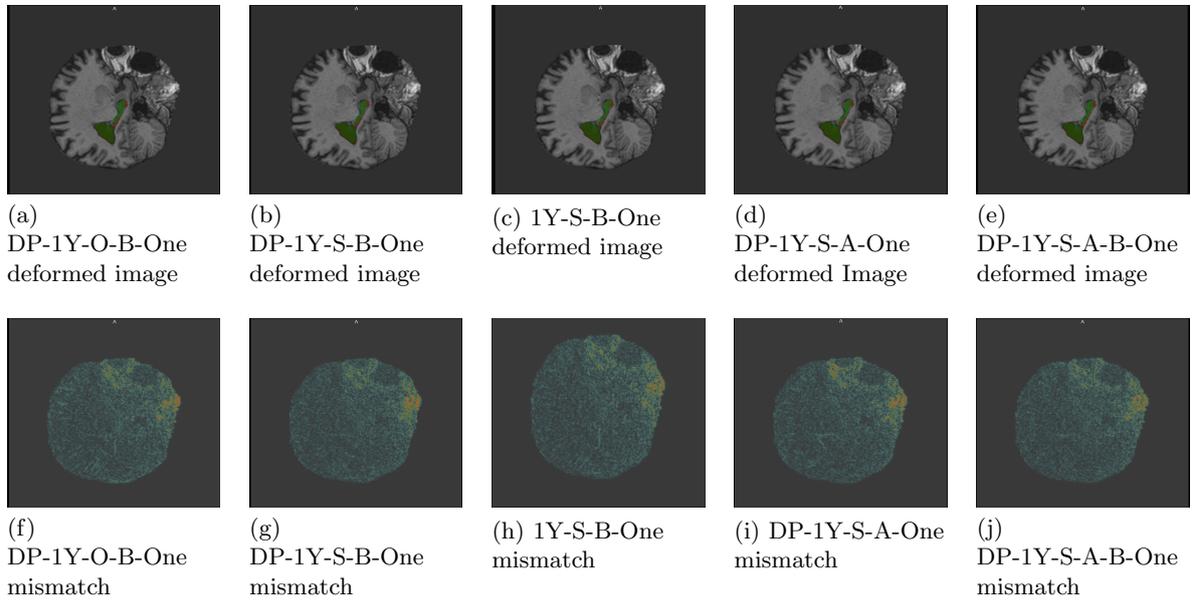
Figure 4.6: One-stage RegNet deformed moving image (Ventricle and hippocampus map from fixed image in green, deformed ventricle and hippocampus map in red) in DP test images and mismatch with the corresponding fixed image (Minor mismatch in blue, larger mismatch in red)

(a)
DP-1Y-O-B-Mul
deformed image

(b)
DP-1Y-S-B-Mul
deformed image

(c) 1Y-S-B-Mul
deformed image

(d)
DP-1Y-S-A-Mul
deformed image

(e)
DP-1Y-S-A-B-Mul
deformed image

(f)
DP-1Y-O-B-Mul
mismatch

(g) DP-1Y-S-B-Mul
mismatch

(h) 1Y-S-B-Mul
mismatch

(i) DP-1Y-S-A-Mul
mismatch

(j)
DP-1Y-S-A-B-Mul
mismatch
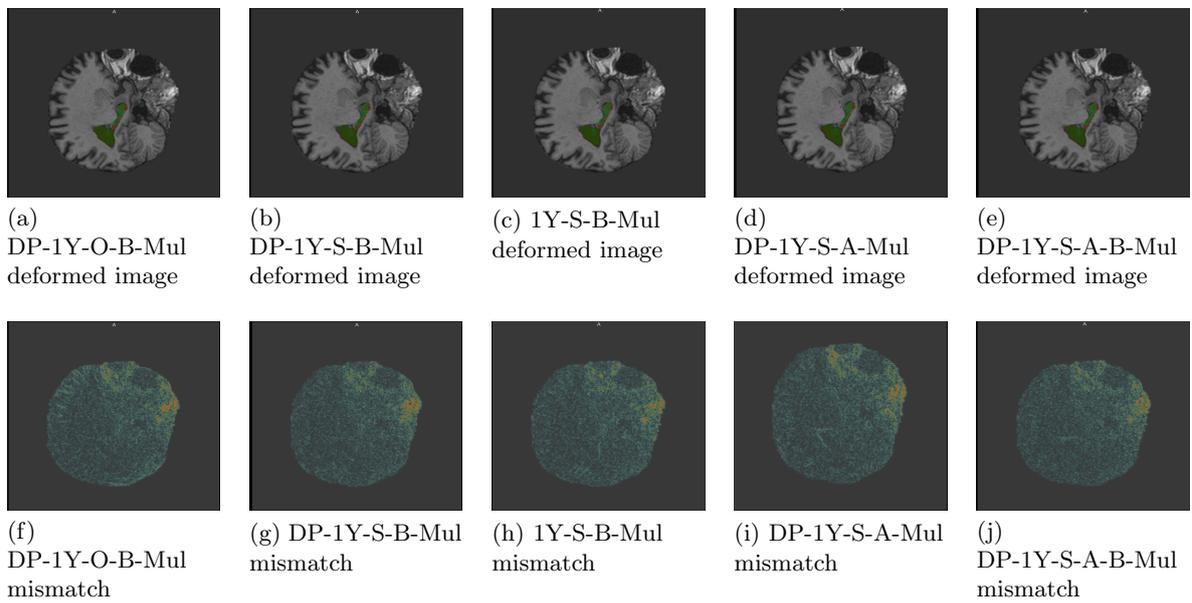
Figure 4.7: Multistage RegNet deformed moving image (Ventricle and hippocampus map from fixed image in green, deformed ventricle and hippocampus map in red) in DP test images and mismatch with the corresponding fixed image (Minor mismatch in blue, larger mismatch in red)

# Discussion and Future Works

<div style="text-align: right">**5**</div>

We have presented a supervised training approach of RegNet for 3D image registration problem, and evaluated the result on an intrapatient brain registration problem. B-spline estimated DVF and artificial simulated DVF were used as ground truth during the training. Applying RegNet in evaluation based on images from both the DP dataset and the 1-Year dataset had achieved promising result in short inference time.

One of the reasons behind the good performance of RegNet is to use skull stripped images in training. As shown in table 4.2 and table 4.3, the DP-1Y-S-B-One model using the skull stripped images is outperforming the DP-1Y-O-B-One model trained with the original images. When the skull stripping technique is applied, only the region within the intracranial region would be processed during the training, and the large background area which is out of interest in medical application would be discarded. Thus, the network could focus more on the region of interest.

We used two types of ground truth DVF, the B-spline estimated DVF and the artificial simulated DVF. The B-spline estimated DVF, as shown in fig. 3.8, contains a lot local variations. In contrast, the DVF simulated artificially, as shown in fig. 3.11, is more smooth. During test time, the RegNet model trained with the B-spline estimated DVF performs better than the RegNet model trained with the artificial simulated DVF. There is also a model DP-1Y-S-A-B-Mul that trained with a combination of B-spline estimated DVF and artificial simulated DVF. Though trained with a larger amount of training data, there is no significant performance improvement for DP-1Y-S-A-B-Mul in contrast with DP-1Y-S-B-Mul. It is not hard to figure out that, artificial simulated DVF generated in this method does not bring additional insights when mixed up with the B-spline estimated DVF during the training. This directly reveals that, realistic ground truth similar to the clinical data is the key of good supervised registration method.

In this project, the performance of the multistage framework is also analyzed. Multistage framework would perform better when there is large deformation between the moving image and the fixed image. As shown in fig. 4.1, the result of RegNet models with multistage framework can outperform the model with one-stage framework on DP test images. Because with the same CNN, using a multistage framework means that the input patches will be first processed by the stage-2-model. The introduction of stage-2-model increases the capture range. The first deformation predicted by the stage-2-model could provide a good basis for the subsequent RegNet to further refine the deformation. What should be pointed out is that using the multistage framework, the RegNet is not showing a performance improvement in the test using 1-Year test images, especially when the hippocampus was used for evaluation. The reason is that the resolution of stage-2-RegNet model is 2 $mm$, in contrast, the 99% HD for the ventricle and hippocampus in the 1-Year test set is 2.0 $mm$ and 1.7 $mm$ respectively, as shown in table 4.4. This type of deformation is too minor for multistage framework to be

effective. Above all, the multistage framework is not suitable to solve the registration problem with only small deformation.

Moreover, when hippocampus maps in the 1-Year dataset were used for evaluation, using RegNet could only reach a registration performance almost equal as the result using the affine registration. This is because hippocampus is relatively a small tissue. For patients with around one year interval between the baseline scan and the follow-up scan, there are no large local deformation, and the affine registration could already reach good alignment.

To further evaluate the performance of the RegNet and the multistage framework, it is suggested to test the method on interpatient problem. It will also be interesting to see, using other methods for ground truth generation, for instance, applying segmentation for support, what kind of influence will be made on the registration performance. We can also try to mask out the region outside the skull for the artificial DVF generation to further check the registration performance.

# Conclusion

# 6

RegNet is a generic medical image registration model, which was originally designed for lung CT image registration problem[41]. In this thesis, trained with two types of ground truth, RegNet is also applied for the intrasubject brain registration problem, obtaining result almost equal with the conventional registration method within a shorter time frame. We suggest applying other ground truth generation method to obtain a better performance. It is also suggested to apply the same method and framework into intersubject MR image registration problem to further validate the performance of this method.

# Bibliography

[1] Zeynettin Akkus, Alfiia Galimzianova, Assaf Hoogi, Daniel L Rubin, and Bradley J Erickson. Deep learning for brain mri segmentation: state of the art and future directions. *Journal of digital imaging*, 30(4):449–459, 2017.

[2] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 2019.

[3] DE Barnaal and IJ Lowe. Proton spin–lattice relaxation in hexagonal ice. *The Journal of Chemical Physics*, 48(10):4614–4618, 1968.

[4] Jan-Philip Bergeest and Florian Jäger. A comparison of five methods for signal intensity standardization in mri. In *Bildverarbeitung für die Medizin 2008*, pages 36–40. Springer, 2008.

[5] Felix Bloch. Nuclear induction. *Physical review*, 70(7-8):460, 1946.

[6] Paul A Bottomley, Thomas H Foster, Raymond E Argersinger, and Leah M Pfeifer. A review of normal tissue hydrogen nmr relaxation times and relaxation mechanisms from 1–100 mhz: dependence on tissue type, nmr frequency, temperature, species, excision, and age. *Medical physics*, 11(4):425–448, 1984.

[7] Richard G Boyes, Jeff L Gunter, Chris Frost, Andrew L Janke, Thomas Yeatman, Derek LG Hill, Matt A Bernstein, Paul M Thompson, Michael W Weiner, Norbert Schuff, et al. Intensity non-uniformity correction using n3 on 3-t scanners with multi-channel phased array coils. *Neuroimage*, 39(4):1752–1762, 2008.

[8] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[9] Xiaohuan Cao, Jianhua Yang, Jun Zhang, Dong Nie, Minjeong Kim, Qian Wang, and Dinggang Shen. Deformable image registration based on similarity-steered cnn regression. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 300–308. Springer, 2017.

[10] Marie Chupin, Emilie Gérardin, Rémi Cuingnet, Claire Boutet, Louis Lemieux, Stéphane Lehéricy, Habib Benali, Line Garnero, and Olivier Colliot. Fully automatic hippocampus segmentation and classification in alzheimer's disease and mild cognitive impairment applied on data from adni. *Hippocampus*, 19(6):579–587, 2009.

[11] Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.

[12] William R Crum, Thomas Hartkens, and DLG Hill. Non-rigid image registration: theory and practice. *The British journal of radiology*, 77(suppl_2):S140–S153, 2004.

[13] Bob D de Vos, Floris F Berendsen, Max A Viergever, Hessam Sokooti, Marius Staring, and Ivana Išgum. A deep learning framework for unsupervised affine and deformable image registration. *Medical image analysis*, 52:128–143, 2019.

[14] Bob D de Vos, Floris F Berendsen, Max A Viergever, Marius Staring, and Ivana Išgum. End-to-end unsupervised deformable image registration with a convolutional neural network. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 204–212. Springer, 2017.

[15] E James Delikatny, Sanjeev Chawla, Daniel-Joseph Leung, and Harish Poptani. Mr-visible lipids and the tumor microenvironment. *NMR in biomedicine*, 24(6):592–611, 2011.

[16] Maurice Goldman. Formal theory of spin–lattice relaxation. *Journal of Magnetic Resonance*, 149(2):160–187, 2001.

[17] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[18] Grant Haskins, Uwe Kruger, and Pingkun Yan. Deep learning in medical image registration: A survey. *arXiv preprint arXiv:1903.02026*, 2019.

[19] Christopher P Hess and D Purcell. Exploring the brain: is ct or mri better for brain imaging. *UCSF Department of Radiology & Biomedical Imaging*, 11, 2012.

[20] Peter J Huber. Robust estimation of a location parameter. In *Breakthroughs in statistics*, pages 492–518. Springer, 1992.

[21] Juan Eugenio Iglesias, Cheng-Yi Liu, Paul M Thompson, and Zhuowen Tu. Robust brain extraction across datasets and comparison with publicly available methods. *IEEE transactions on medical imaging*, 30(9):1617–1634, 2011.

[22] Clifford R Jack Jr, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L. Whitwell, Chadwick Ward, et al. The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4):685–691, 2008.

[23] Stefan Klein and Marius Staring. elastix: the manual. [http://elastix.isi.uu.nl/download/elastix_manual_v4.3.pdf](http://elastix.isi.uu.nl/download/elastix_manual_v4.3.pdf), Jan 2010.

[24] Stefan Klein, Marius Staring, Keelin Murphy, Max A Viergever, and Josien PW Pluim. Elastix: a toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging*, 29(1):196–205, 2010.

[25] Stefan Klein, Marius Staring, and Josien PW Pluim. Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines. *IEEE transactions on image processing*, 16(12):2879–2890, 2007.

[26] Seymour H Koenig. Classes of hydration sites at protein-water interfaces: the source of contrast in magnetic resonance imaging. *Biophysical journal*, 69(2):593–603, 1995.

[27] Julian Krebs, Tommaso Mansi, Hervé Delingette, Li Zhang, Florin C Ghesu, Shun Miao, Andreas K Maier, Nicholas Ayache, Rui Liao, and Ali Kamen. Robust non-rigid registration through agent-based action learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 344–352. Springer, 2017.

[28] Jorge Larrey-Ruiz and Juan Morales-Sánchez. Optimal parameters selection for nonparametric image registration methods. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 564–575. Springer, 2006.

[29] Dwarikanath Mahapatra, Suman Sedai, and Rahil Garnavi. Elastic registration of medical images with gans. *arXiv preprint arXiv:1805.02369*, 2018.

[30] Andriy Myronenko, Xubo Song, and Miguel Carreira-Perpinan. Non-parametric image registration using generalized elastic nets. 2006.

[31] National Institutes of Health. Ventricles of the brain, December 2011.

[32] Gerald M. Pohost, Gabriel A. Elgavish, and William T. Evanochko. Nuclear magnetic resonance imaging: With or without nuclear? *Journal of the American College of Cardiology*, 7(3):709 – 710, 1986.

[33] Ivo Rausch, Harald H Quick, Jacobo Cal-Gonzalez, Bernhard Sattler, Ronald Boellaard, and Thomas Beyer. Technical and instrumentational foundations of pet/mri. *European Journal of Radiology*, 94:A3–A13, 2017.

[34] Jimin Ren, Ivan Dimitrov, A Dean Sherry, and Craig R Malloy. Composition of adipose tissue and marrow fat in humans by 1h nmr at 7 tesla. *Journal of lipid research*, 49(9):2055–2062, 2008.

[35] John S Rigden. Quantum states and precession: The two discoveries of nmr. *Reviews of modern physics*, 58(2):433, 1986.

[36] Marc-Michel Rohé, Manasi Datar, Tobias Heimann, Maxime Sermesant, and Xavier Pennec. Svf-net: Learning deformable image registration using shape matching. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 266–274. Springer, 2017.

[37] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[38] Daniel Rueckert, Luke I Sonoda, Carmel Hayes, Derek LG Hill, Martin O Leach, and David J Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE transactions on medical imaging*, 18(8):712–721, 1999.

[39] Julia A Schnabel, Daniel Rueckert, Marcel Quist, Jane M Blackall, Andy D Castellano-Smith, Thomas Hartkens, Graeme P Penney, Walter A Hall, Haiying Liu, Charles L Truwit, et al. A generic framework for non-rigid registration based on non-uniform multilevel free-form deformations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 573–581. Springer, 2001.

[40] John G Sled, Alex P Zijdenbos, and Alan C Evans. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE transactions on medical imaging*, 17(1):87–97, 1998.

[41] Hessam Sokooti, Bob de Vos, Floris Berendsen, Boudewijn PF Lelieveldt, Ivana Išgum, and Marius Staring. Nonrigid image registration using multi-scale 3d convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 232–239. Springer, 2017.

[42] Arthur W Toga and Paul M Thompson. Mapping brain asymmetry. *Nature Reviews Neuroscience*, 4(1):37, 2003.

[43] Michael Unser. Splines: A perfect fit for signal and image processing. *IEEE Signal processing magazine*, 16(ARTICLE):22–38, 1999.

[44] Hristina Uzunova, Matthias Wilms, Heinz Handels, and Jan Ehrhardt. Training cnns for image registration from few samples with model-based data augmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 223–231. Springer, 2017.

[45] Xiao Yang, Roland Kwitt, Martin Styner, and Marc Niethammer. Quicksilver: Fast predictive image registration–a deep learning approach. *NeuroImage*, 158:378–396, 2017.

[46] Paul A Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C Gee, and Guido Gerig. User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*, 31(3):1116–1128, 2006.

[47] Song Chun Zhu and Alan Yuille. Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (9):884–900, 1996.

# Appendix

# A
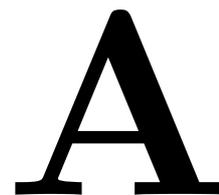
## A.1 ADNI index of DP training set

'002_S_0619',   '002_S_0938',   '002_S_1018',   '002_S_1070',   '003_S_1059',   '005_S_0221',
'005_S_0222',   '005_S_0223',   '005_S_0572',   '005_S_0814',   '005_S_1341',   '006_S_0547',
'006_S_0675',   '006_S_1130',   '007_S_0249',   '007_S_0293',   '007_S_0316',   '007_S_0344',
'007_S_1339',   '011_S_0003',   '011_S_0053',   '012_S_0689',   '012_S_0803',   '013_S_0996',
'013_S_1205',   '014_S_0328',   '014_S_0548',   '014_S_0558',   '014_S_0563',   '014_S_0658',
'016_S_0769',   '016_S_0991',   '018_S_0057',   '018_S_0286',   '018_S_0335',   '018_S_0633',
'021_S_0141',   '021_S_0231',   '021_S_0343',   '021_S_0424',   '021_S_0753',   '022_S_0007',
'022_S_0129',   '022_S_0543',   '022_S_0750',   '023_S_0030',   '023_S_0083',   '023_S_0084',
'023_S_0093',   '023_S_0139',   '023_S_0331',   '023_S_0388',   '023_S_0604',   '023_S_0625',
'023_S_1126',   '023_S_1247',   '023_S_1262',   '024_S_0985',   '024_S_1171',   '024_S_1307',
'024_S_1393',   '027_S_0179',   '027_S_0404',   '027_S_0408',   '027_S_0835',   '027_S_0850',
'027_S_1081',   '027_S_1082',   '027_S_1213',   '027_S_1254',   '027_S_1385',   '029_S_0843',
'029_S_0878',   '029_S_1056',   '031_S_0294',   '031_S_0554',   '031_S_0568',   '031_S_1209',
'032_S_0147',   '032_S_0214',   '032_S_0400',   '032_S_0978',   '033_S_0514',   '033_S_0567',
'033_S_0723',   '033_S_0724',   '033_S_0725',   '033_S_0733',   '033_S_0889',   '033_S_0906',
'033_S_0922',   '033_S_1281',   '033_S_1283',   '033_S_1285',   '033_S_1308',   '035_S_0204',
'035_S_0341',   '035_S_0555',   '036_S_0577',   '036_S_0759',   '036_S_0760',   '036_S_0869',
'037_S_0539',   '037_S_0566',   '041_S_0314',   '041_S_1368',   '051_S_1123',   '051_S_1296',
'052_S_1054',   '057_S_0779',   '057_S_0839',   '057_S_0941',   '057_S_1371',   '057_S_1373',
'057_S_1379',   '062_S_0535',   '062_S_0690',   '062_S_0730',   '062_S_0793',   '062_S_1299',
'067_S_0029', '067_S_0076', '067_S_0077', '067_S_0336', '067_S_0812'

## A.2 ADNI index of DP test set

'073_S_0518',   '073_S_0565',   '094_S_0434',   '098_S_0269',   '114_S_0166',   '114_S_0374',
'114_S_0979',   '114_S_1106',   '116_S_0392',   '116_S_0487',   '116_S_0649',   '126_S_0891',
'126_S_1077',   '126_S_1221',   '127_S_0394',   '127_S_0754',   '127_S_0844',   '127_S_1382',
'131_S_0457',   '131_S_0497',   '131_S_1389',   '133_S_0727',   '137_S_0366',   '137_S_0459',
'137_S_0631', '941_S_1311'

## A.3 ADNI index of 1Y training set

'021_S_0159',   '013_S_1035',   '013_S_0502',   '007_S_1339',   '021_S_0276',   '013_S_1120',
'014_S_0558',   '016_S_1028',   '005_S_0546',   '007_S_0068',   '014_S_0328',   '007_S_0414',
'002_S_0782',   '003_S_0907',   '002_S_0295',   '006_S_0681',   '009_S_1030',   '016_S_0702',
'002_S_0938',   '007_S_0041',   '005_S_0610',   '020_S_0883',   '007_S_0698',   '003_S_0981',
'014_S_0658',   '013_S_1186',   '002_S_0954',   '006_S_0498',   '011_S_0003',   '006_S_0731',
'006_S_1130',   '013_S_1205',   '002_S_1155',   '016_S_0359',   '005_S_0448',   '021_S_0337',
'002_S_0413',   '007_S_0101',   '014_S_0519',   '005_S_0814',   '021_S_0984',   '005_S_1341',
'014_S_0563',   '016_S_1117',   '014_S_0169',   '009_S_0751',   '002_S_0729',   '002_S_0685',
'002_S_1018',   '011_S_0022',   '016_S_0538',   '021_S_0231',   '014_S_0548',   '014_S_0557',
'022_S_0004',   '023_S_0217',   '023_S_0388',   '033_S_0724',   '029_S_0824',   '027_S_0118',
'036_S_0869',   '033_S_0725',   '023_S_0855',   '036_S_0945',   '033_S_0514',   '035_S_0048',
'033_S_0923',   '022_S_0066',   '023_S_0078',   '027_S_0120',   '027_S_0179',   '027_S_0074',
'029_S_1215',   '027_S_1277',   '035_S_0156',   '029_S_0843',   '029_S_1073',   '023_S_0042',
'033_S_0906',   '035_S_0555',   '023_S_0061',   '033_S_0567',   '027_S_0461',   '035_S_0341',
'022_S_0096',   '036_S_0577',   '036_S_0759',   '035_S_0997',   '033_S_0741',   '035_S_0292',
'033_S_1098',   '033_S_0922',   '033_S_1016',   '029_S_0878',   '033_S_1285',   '027_S_0404',
'036_S_0656',   '022_S_0130',   '022_S_0014',   '033_S_1281',   '024_S_1171',   '036_S_0673',
'033_S_0723',   '023_S_0084',   '027_S_0644',   '033_S_0511',   '036_S_1023',   '033_S_0920',
'033_S_0733',   '023_S_0081',   '031_S_0867',   '029_S_0866',   '036_S_0672',   '023_S_0963',
'033_S_0734',   '027_S_0307',   '033_S_0739',   '033_S_1279',   '036_S_1240',   '022_S_1351',
'022_S_0750',   '027_S_0256',   '033_S_0516',   '027_S_1045',   '036_S_0976',   '022_S_0961',
'029_S_0845', '024_S_0985', '036_S_0760', '036_S_0748', '036_S_0813'

## A.4 ADNI index of 1Y test set

'005_S_0222',   '006_S_0675',   '006_S_1130',   '016_S_0359',   '020_S_0213',   '020_S_0883',
'021_S_0273',   '022_S_0014',   '029_S_0999',   '035_S_0048',   '035_S_0341',   '133_S_0525',
'035_S_0997',   '036_S_1023',   '041_S_0549',   '041_S_0598',   '062_S_0768',   '067_S_0076',
'067_S_0077',   '067_S_0257',   '073_S_0518',   '073_S_0565',   '073_S_0746',   '116_S_0834',
'131_S_0384', '131_S_0457'