Population specific template construction and brain structure segmentation using deep learning methods

Thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science by Research
in
Electronics and Communication Engineering

by

Raghav Mehta
201432625
raghav.mehta@research.iiit.ac.in

International Institute of Information Technology
Hyderabad - 500032, INDIA
July 2017

International Institute of Information Technology
Hyderabad, India

CERTIFICATE OF AUTHORSHIP

I, Raghav Mehta, declare that the thesis, titled 'Population specific template construction and brain structure segmentation using deep learning methods.', and the work presented herein are my own. I confirm that this work was done wholly or mainly while in candidature for a research degree at IIIT-Hyderabad.

25/07/2017

Date

Signature of the Candidate

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled 'Population specific template construction and brain structure segmentation using deep learning methods' by Raghav Mehta, has been carried out under my supervision and is not submitted elsewhere for a degree.

25/7/2017
Date

Adviser: Prof. Jayanthi Sivaswamy

To

Pabhai

Acknowledgements

First and foremost, I would like to thank my supervisor, Prof. Jayanthi Sivaswamy for her omnipresent guidance and support. I will always be grateful to her for introducing me to the area of Neuroimage analysis. She showed me how a small leap of faith and a little hard work can help us in achieving goals which are far beyond our dreams. Under her watchful eyes, I have seen myself transforming from a shy-bumbling person into a assured-methodical researcher.

Next, I would like to thank Prof. Bapi Raju, Dr. L.T. Kishore, Dr. Nagendra, Dr. Vishnu, Dr. Keshavdas, Dr. Kiranmayi, Mr. Ansari, Pawan, Ramesh, Mahesh and all the others who have helped me in the data collection task for the project of "Construction of Indian human brain atlas of young population", which was an integral part of my research work. I thank all the enthusiastic participants who volunteered for the MRI scans. This thesis would not have been possible without them.

I extend my gratitude to all the members of CVIT, especially of MIP group, for making this learning experience memorable. A special mention should go to R.S. Satyanarayan, for helping me out in administrative and financial processes during the Atlas project, and Mutiyalu, for greeting me with his evergreen smile every morning and for his excellent duty of making our workplace dust free.

Research is a long and arduous process, you require friends to keep you sane during that. I was lucky enough to have friends who kept me stable and made this journey memorable. Thank you Aditya and Soumyajit, for all the talks we had on various topics ranging from GoT to deep learning; Snehil and Shruti, for keeping me motivated during tough times; Karthik, for all the discussions we had on almost all the topics in the world; Pramod and Rahul, for the fun FIFA nights; Nausheen and Govinda, for showing me the importance of maintaining childlike innocence when you grow up; Maulik, for keeping me close to my native state Gujarat; and lastly Aabhas, for keeping me sane and calm during the paper deadlines.

I am fortunate enough to have a family who has supported me during all the highs and lows in my life. I thank my elder brother Parth, who introduced me to the world of research and helped me out in each and every step of my life, and twin sister Toral, who was always there for me when I needed her the most and who made me realize that every step of this journey is worthwhile. At last, I would like to wholeheartedly thank my parents, for their constant love and support and keeping faith in all the decisions I have taken.

Abstract

A brain template, such as MNI152 is a digital (magentic resonance image or MRI) representation of the brain in a reference coordinate system for the neuroscience research. Structural atlases, such as AAL and DKA, delineate the brain into cortical and subcortical structures which are used in Voxel Based Morphometry (VBM) and fMRI analysis.

Many population specific templates, i.e. Chinese, Korean, etc., have been constructed recently. It was observed that there are morphological differences between the average brain of the eastern and the western population. In this thesis, we report on the development of a population specific brain template for the young Indian population. This is derived from a multi-centeric MRI dataset of 100 Indian adults (21 - 30 years old). Measurements made with this template indicated that the Indian brain, on average, is smaller in height and width compared to the Caucasian and the Chinese brain.

A second problem this thesis examines is automated segmentation of cortical and non-cortical human brain structures, using multiple structural atlases. This has been hitherto approached using computationally expensive non-rigid registration followed by label fusion. We propose an alternative approach for this using a Convolutional Neural Network *(CNN)* which classifies a voxel into one of many structures. Evaluation of the proposed method on various datasets showed that the mean Dice coefficient varied from 0.844 ± 0.031 to 0.743 ± 0.019 for datasets with the least (32) and the most (134) number of labels, respectively. These figures are marginally better or on par with those obtained with the current state of the art methods on nearly all datasets, at a reduced computational time. We also propose an end-to-end trainable Fully Convolutional Neural Network (FCNN) architecture called the *M-net*, for segmenting deep (human) brain structures. A novel scheme is used to learn to combine and represent 3D context information of a given slice in a 2D slice. Consequently, the *M-net* utilizes only 2D convolution though it operates on 3D data. Experiment results show that the *M-net* outperforms other state-of-the-art model-based segmentation methods in terms of dice coefficient and is at least 3 times faster than them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Medical Imaging

Medical Imaging is a routine and essential part of medicine acquired by non-invasive, in-vivo techniques. These various imaging techniques acquire information about a particular body part of the subject. Typical examples of various medical imaging techniques are X-ray (teeth and bones), Ultrasound (Abdomen) , Computed Tomography (liver), Magnetic Resonance Image (heart), Optical images (Retina and skin lesion) etc. as shown in Fig:1.1. Major advantage of medical imaging is that it allows the health problems to be directly observed rather than derived from symptoms. The acquired images are used by clinicians and radiologist for diagnostic purposes, surgical planning, monitoring disease progression and to track the effect of a treatment. Availability of the abundance of medical imaging data, typical high resolution natures of images and critical nature of its application has made manual analysis of these images by human experts arduous. Automatic analysis systems use various image processing and machine learning techniques to help doctors in heterogeneous tasks such as abnormality detection and localization, disease classification, structure segmentation, quantitative analysis of structure of interest, multi-modality fusion and so on [9].

In this thesis, we will focus on structural neuroimaging which captures the anatomy of human brain.

1.2 Human Brain

The human brain is a crucial organ as it is the center of the nervous system. From a computer science point of view, brain acts like the processor of a human body, which receives information

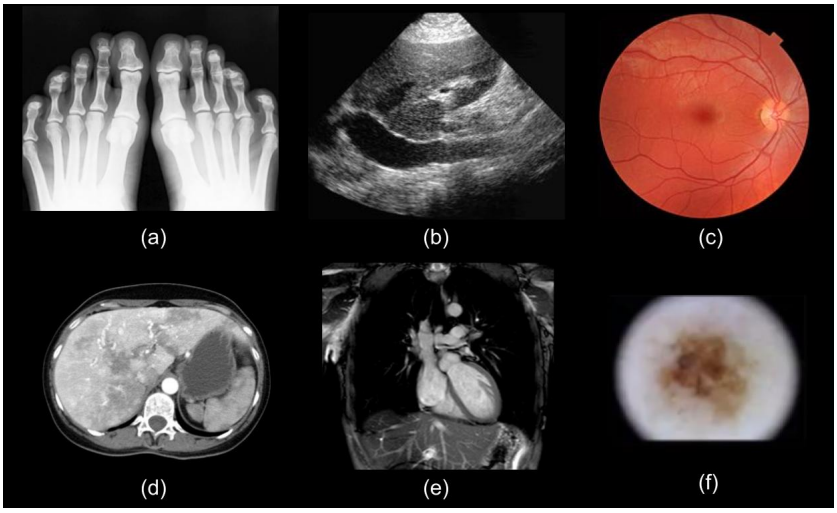

Figure 1.1: Sample medical images. (a) X-ray of feet (b) Ultrasound of abdomen (c) Optical image of retina (d) CT of liver (e) MRI of heart (f) Optical image of skin lesion

from various sensory systems (organs) like visual system (eyes), auditory system (ears), samotosenory system (skin) etc., processes it, interprets it, analyze it, makes decision based on that. These are then transmitted back to the various organs.

From a cellular science point of view it consists of millions and millions of neurons. This divides human brain in mainly two parts: White matter consisting of bundles of myelinated axons which connect various part of brain and Grey matter which is the processing part of neurons. Third and another important part of the brain is the cerebrospinal fluid (CSF) which acts as a cushion or buffer for the brain, providing basic mechanical and immunological protection to the brain inside the skull. All three parts are clearly distinguishable in terms of their color (Fig:1.3).

Figure 1.2: Pictorial view of Human brain. Image courtesy: National Geographic

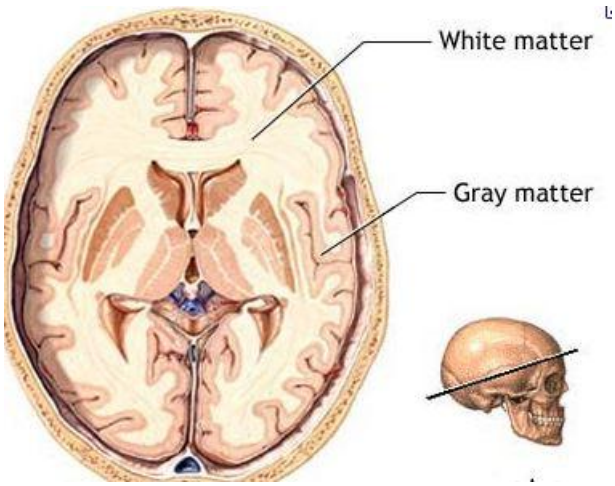

Figure 1.3: White Matter and Gray Matter in human brain (Image courtesy: Wikipedia)

From an anatomical point of view, the human brain can be broadly divided into three parts: cerebrum, brainstem and cerebellum as shown in Fig:1.4. The cerebrum is the largest part of human brain, which contains cerebral cortex as well as many subcortical structures like hippocampus and basal ganglia (Fig:1.5 (A)). Cerebral cortex, which consists of two cerebral hemisphere, can be broadly divided into four functional parts: Frontal lobe, parietal lobe, occipital lobe and temporal lobe (Fig:1.5 (B)). Cerebrum, as a whole structure, plays a major part in many human activities such as controlling emotions, hearing, vision, personality, movement, memory, language learning, olfaction etc. Each lobe can further be divided into smaller areas according to the specific type of activity it processes and controls (Fig:1.5 (C)). For example, the visual cortex which is a part of occipital lobe, plays an important role in processing the visual information.

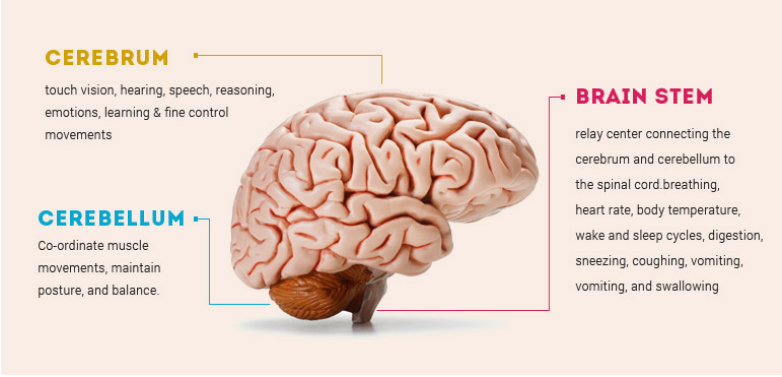

Figure 1.4: Human brain anatomy: Cerebrum, Brainstem and Cerebellum (Image courtesy: Wikipedia)

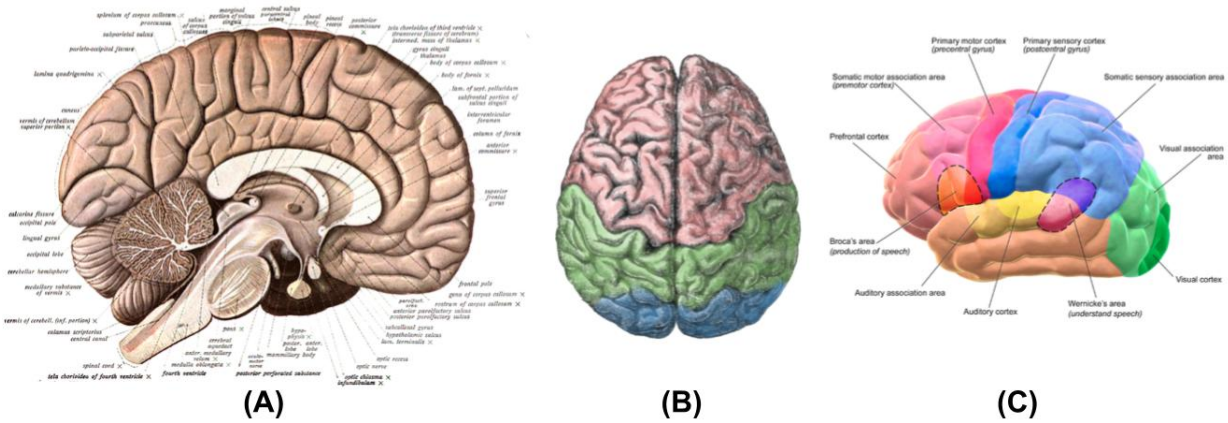

Figure 1.5: Structural (A) and Functional (B and C) areas of the human brain. (A) shows various cortical and subcortical structures of human brain like, sulcus, gyrus and basel ganglia. (B) depicts the frontal lobe (pink), parietal lobe (green) and occipital lobe (blue) of human brain. (C) represents various motor and sensory regions of cerebral cortex, like auditory cortex, visual cortex etc. (Image courtesy: Wikipedia)

The cerebellum or "little brain" is a major part of of human brain. It plays an important role motor control as well as many cognitive functions such as attention and language. It also regulates fear and pleasure responses. Damage to cerebellum can produce disorders in fine movement, equilibrium, posture and motor learning in humans.

The brainstem is a posterior part of brain which serves as a connection between spinal cord and cerebrum. It plays a role in relaying all the information from the various body part to cerebrum and vice versa. It is pivotal in maintaining consciousness and regulating sleep cycle.

1.3 Neuroimaging

Neuroimaging or brain imaging techniques is one of the most widely used imaging techniques, which directly or indirectly maps structural or functional information of human brain.

Neuroimaging can be broadly classified into following two categories: (i) Structural Neuroimaging and (ii) Functional Neuroimaging

1.3.1 Structural NeuroImaging

As the name suggests, this deals with the structure of the brain and aids diagnosis of large scale intracranial diseases and injury. Two major modalities for structural imaging are Computed

Tomography (CT) and Magnetic Resonance Imaging(MRI).

In this thesis, we will focus on MRI which uses a powerful magnetic field and radio-waves to produce high quality images of various structures of brain. Unlike, CT it doesn't use harmful x-ray for imaging purpose. MRI produces high quality images with soft tissue contrast that is much better than CT. Moreover, MRI can distinguish tissues that have similar intensities and are hard to distinguish using CT. Different modes of structural MRI like T1-weighted, T2-weighted, Proton Density (PD), Fluid Attenuated Inversion Recovery (FLAIR), Diffusion Tensor Imaging (DTI) and Diffusion Weighted Imaging (DWI) can be useful for various purposes like visualizing anatomy, visualizing pathology, identifying dementia, mapping white matter tractography etc. In this thesis we will focus on T1-weighted MRI images.

1.3.2 Functional NeuroImaging:

Measurement of brain function is of interest to understand the relationship between activity in certain brain areas and specific mental functions like motor, vision, sensory, memory etc. Three main neuroimaging modalities used for this purpose, are as follow: Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT) and Functional MRI (FMRI). While, every modality has its own pros and cons, use of FMRI for human functional analysis purpose has drastically increased in the last decade. It measures brain activity by detecting changes associated with blood flow. It acquires blood-oxygen-level dependent (BOLD) signals. Through a process called the hemodynamic response, blood releases oxygen to active neurons at a greater rate than to inactive neurons in the brain. This causes a change of the relative levels of oxyhemoglobin and deoxyhemoglobin (oxygenated or deoxygenated blood) that can be detected on the basis of their differential magnetic susceptibility by MRI scanner. This gives the images of brain regions which are highly active during a particular task.

1.4 Brain Template

A world atlas provides a standard co-ordinate system in terms of longitude and latitude, such that each area or region of earth has a specific values for it. This co-ordinate system can be useful to observe and analyze many different information such as geography, population, weather, terrain etc., by projecting them onto this standard co-ordinate system.

Similar to the world atlas, the human brain atlas represents a distinct anatomical portrayal of the brain depicting finer anatomical details. [10] This atlas provides a standard framework

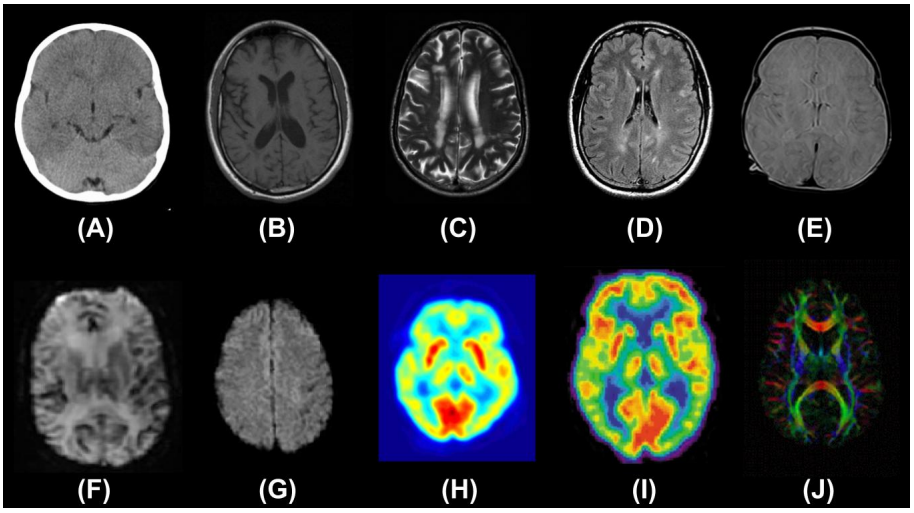

Figure 1.6: Example of various NeuroImaging Modalities. (A) CT (B) T1 MRI (C) T2 MRI (D) FLAIR MRI (E) PD MRI (F) PWI MRI (G) DWI MRI (H) SPECT (I) PET (J) DTI

in which population based assessment of brain function and anatomy is possible. This is done through spatial normalization of individual subjects onto standard atlas using registration methods.

Brain templates/atlases can be broadly divided into two categories: (i) Population specific template (ii) Structural atlas

1.4.1 Population specific Template

Although, in a broader sense, human brain anatomy is more or less consistent throughout the human population, there is an emerging evidence that there are morphological differences (e.g. brain shape and size) in the brain of different origin (e.g. age, gender, race) [11] [12] [13]. Hence, it is necessary to construct population specific brain atlas. These kind of brain atlases are constructed using structural MRI images, in particular T1 weighted MRI images. They serve as a standard template, for a population, against which neuroanatomy of different subjects can be compared. Montreal Neurological Institute (MNI) template (Fig:1.7) is one of the most widely used brain template [14]. It is used for various population based studies where there is a significant change in brain size and shape of subjects due to various factors like aging, disease or brain injuries.

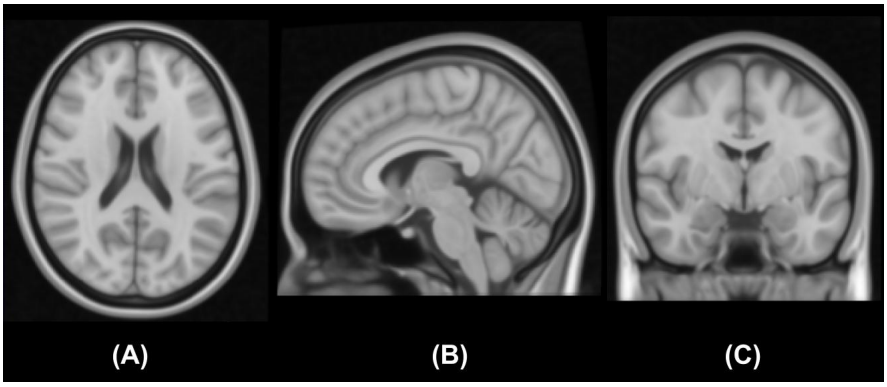

Figure 1.7: Axial (A), Sagittal (B) and Coronal (C) slices of MNI template [1]

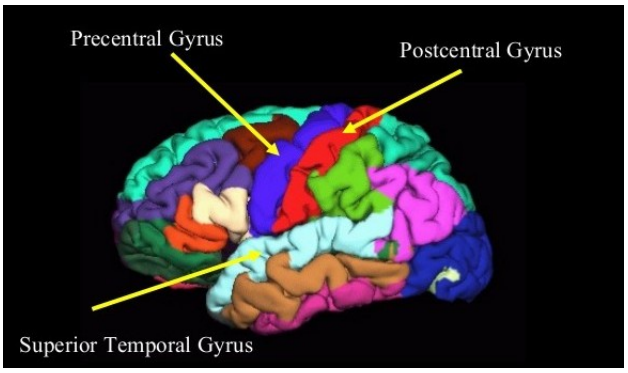

Figure 1.8: Automated surface parcellation of Desikan-Killiany atlas [2]

1.4.2 Structural Brain Atlas

Structural atlases divide the brain into number of brain regions. Here, each voxel in a brain is labeled as one of the many sub-regions of the brain. Process of dividing the brain into number of areas is known as brain parcellation or brain segmentation process. Various brain parcellation techniques either use mono-modal data like T1-weighted MRI [15] or use multi-modal data like T1 and fMRI images [16]. Parcellation delineates the brain into either coarse level regions (GM, WM and CSF) or fine-grained regions (upto 180 per hemisphere) [16]. Coarse level segmentation of brain is useful for various studies like aging where Voxel Based Morphometric (VBM) analysis of white matter plays an important role. Similarly, studies such as analysis of functional connectivity and structural connectivity of human brain requires fine-grained delineation of brain. Desikan-Killiany Atlas (Fig:1.8) is one of the most popular structural atlas [2]. This atlas delineates the brain into 68 Regions of interest (ROI) .

1.5 Thesis Focus

In this thesis, we focuses on two main problems, namely, population specific template construction for young Indian population and development of automatic brain segmentation techniques using deep learning, a task that is of interest in many atlas-based application.

Population specific brain MRI templates plays an important role in many neuroimage analysis techniques which require spatial normalization of a brain MRI to a standard co-ordinate system. This spatial normalization enables a straightforward qualitative and quantitative comparison of brain MRI of a wide population. MNI152 [1] is a widely used template which is constructed using T1 brain MRI scan of 152 Caucasian subjects. It is used as a standard template in many neuroimaging studies where comparison of different morphological features of brain like height, width, volume etc. plays an important role. Use of the template as a standard brain necessitate that this template should be as near to the average brain of the population as possible. Recent studies have shown that there is a morphological difference between Caucasian and Asian brain [12], especially in terms of global brain features like height, width and length. Similarly, difference between inter-continental population is also shown in a study by a Korean research group [17]. These studies motivate the construction and comparison on brain template of Indian population. In the first part of this thesis, we construct and validate the Indian brain MRI template of young Indian population. We show that an average Indian brain is smaller compare to other brain templates, especially in terms of height.

Functional MRI or FMRI provides the activation information of various brain area for a specific task in a subject. Analysis of fMRI data requires the segmentation of brain into known cortical and subcortical regions. This segmentation allows the measurement and comparison of functional activation of various brain regions during different tasks. Volumetric and shape analysis of sub-cortical structure like hippocampus and basal ganglia plays an important role in identifying diseases such as Alzheimer's and Huntington's [18]. This kind of analysis also needs segmentation of subcortical brain regions. Manual segmentation or parcellation of brain MRI is an arduous, time consuming task and marker dependent task. Hence, development of automatic segmentation methods is imperative. Time complexity and accuracy plays a critical role in determining the success of automatic segmentation methods. In the second half of this thesis, we propose two different Convolutional Neural Network (CNN) based solutions for the cortical and sub-cortical segmentation of brain MR images.

In neuroimaging studies like population specific atlas construction, acquired structural MRI scans can be from different scanning protocols. As MR image intensities do not possess a tissue

specific numeric meaning even in images acquired for the same subject, on the same scanner, for the same body region, by using the same pulse sequence, the intensity range of different scans vary significantly. Similarly, there is also intra-subject and inter-subject intensity variation in MRI scans which can hamper studies like longitudinal analysis, automatic segmentation and image registration methods [19]. The set of algorithms used to handle intensity variation in MRI scans are known as Intensity Standardization (IS) methods [20]. In this thesis, we propose a new variant for intensity standardization.

1.6 Contribution

Major contributions of this thesis are as follow:

1. A hybrid approach for landmark based intensity standardization of brain MRI images
2. Construction and validation of Indian brain MRI template for young (21-30 years) Indian population.
3. Proposal of two CNN based deep learning architectures for brain MRI segmentation

1.7 Outline of the thesis

The organization of the thesis is as follows: In chapter-2, we discuss various intensity standardization methods and propose a new variation for it. In chapter-3, we construct the population specific brain template for young Indian community and validate the constructed template against widelty used MNI-152 [1] and Chinese2020 [21] templates. In chapter-4 and chapter-5 we propose two different deep learning architectures for anatomical brain segmentation using multiple structural atlases. While the first architecture uses a traditional convolutional neural network which segments the brain voxel-by-voxel, second architecture is based on fully convolutional neural network (FCNN) which segments the brain slice by slice, keeping interslice consistency. In the last chapter, we give brief summary of the presented work and suggest some possible future work.

Chapter 2

Intensity Standardization

Magnetic Resonance Images (MRI) of the brain is widely used for the diagnosis of many neuro disorder starting from brain aneurysms and stroke, to Dementia, MS lesions and cancer. Atlas construction is another application widely used to get population specific normative parameters of the brain. This requires Registration of MRI of different patients from different scanner manufacturers. Disease diagnosis is based on quantitative information derived from segmentation and registration (to atlas) of MR volumes.

A major problem in MRI is that it lacks standardization of voxel intensity values. There is inter-scanner, intra-scanner, inter-patient, intra-patient variations of intensity profile for same tissue type. It is well known that this kind of variation affects processing steps like segmentation, registration [22] etc., which in turn affects above mentioned applications.

2.1 Related work

Various solutions have been proposed for the Intensity Standardization (IS) task [23, 24]. These approaches can be categorized as A: independent of or B: dependent on prior knowledge of tissue labels. Earlier approaches to IS belong to category A. Here, IS is performed in the histogram space using mappings. These include one-to-one, one-to-many and many-to-one mapping between non-standard to a standard-volume [25]. The importance of preserving tissue distributions were recognized and a Gaussian Mixture Modelling was used to estimate different tissue types and their variation across volumes was minimized [26]. One of the most popular method for IS uses landmark-based matching and mapping [27]. Landmarks are based on modes of histogram [20] or percentiles of the foreground [27][28]. IS has also been cast as a non-rigid registration

problem [29].

More recent methods are in category B. For example, [30] uses joint image histograms to determine intensity correspondence between the input and standard volumes in each tissue class, and uses it as landmark points. In contrast, [31] identifies the landmark points for each tissue class individually and matches them to standard scale. This method is reported to outperform [27].

Tissue distribution based IS (category B) is said to improve the performance in terms of tissue separability and tissue Gaussianity [31] [30]. However, the preprocessing steps of brain extraction and tissue segmentation are time consuming and hence these methods are slower than the methods in category A.

2.2 Proposed Method

We propose a hybrid approach to IS which exploits the strengths of both categories. This is motivated by the fact that i) In a practical scenario, it is not possible to have tissue segmentation for every volume, ii) IS is a preprocessing step for further image analysis and hence needs to be computationally light. We propose to restrict the use of tissue labelled volumes to the training phase of IS. Given a new volume with unlabelled voxels, we propose a novel way to identify the landmarks, by propagating the landmark points from similar volumes, identified using KL divergence, in the training dataset.

We use the method proposed in [31] to illustrate our idea. Nevertheless, our approach of finding similar image from training dataset and use its information for testing data is equally applicable to [30] and other approaches based on prior knowledge of tissue distributions.

The proposed method consists of two stages: offline training and online testing. During training we have a set of non-standardized MRI volumes I_n; and corresponding Tissue Masks M_n^j where $n = 1, ..N$ and j = Grey Matter (GM), White Matter (WM) and Cerebrospinal Fluid (CSF). During testing, we have a given MRI volume I_g with no tissue mask. All volumes (training and testing) are preprocessed to identify the foreground by thresholding the voxel intensity at a threshold value identified using Otsu's method [32].

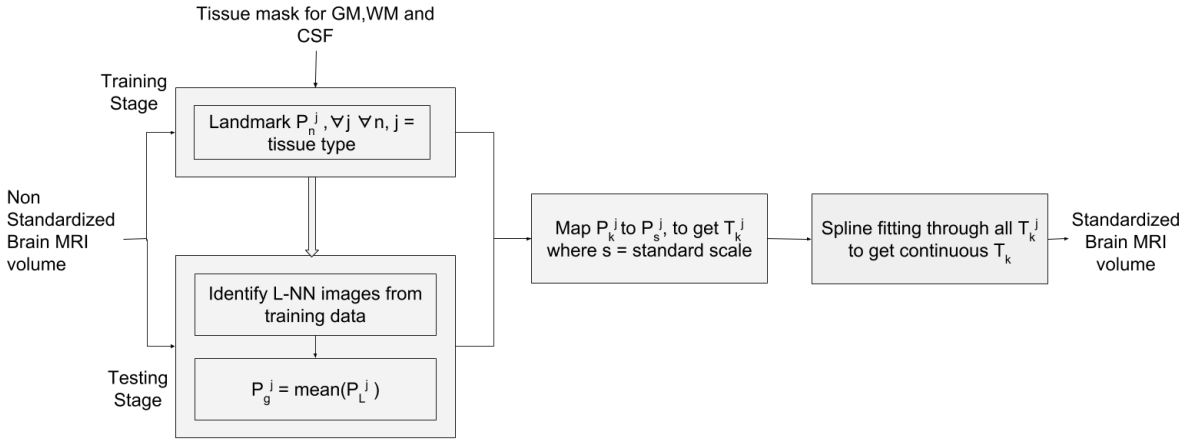

Figure 2.1: Overview of the proposed hybrid approach for intensity standardization of brain MR images

2.2.1 Training Stage

Given I_n and M_n^j the main steps in the training phase are:

1) Calculation of grey level histograms and corresponding percentiles for each tissue type P_n^j.

2) Determination of transformations T_n^j, $\forall j$ by matching the percentile landmarks P_n^j of I_n with the corresponding standard-scale percentile landmarks P_s^j.

3) Derivation of the continuous mapping T_n for the whole volume via spline-fitting through T_n^j as in [31].

2.2.2 Testing Stage

Our main contribution is in this stage. Given a test volume I_g our aim is to derive a transformation T. The key information needed is the landmarks on the tissue histograms. Since, no tissue-level information is available for I_g, we proposed to derive it by finding the nearest neighbours of I_g in the training set I_n.

2.2.2.1 Finding the Nearest Neighbours

First, the intensity range of the training volume is remapped to that of testing volume intensity range by a linear transformation. We choose the KL divergence (KLD) [33] metric to identify volumes from the training set I_n which are similar to I_g. Given two Probability Distribution

Functions(PDF) P and Q the KLD is given as

$$KLD(P,Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

In our case, KLD is computed between the histogram of the given volume and all the volumes in the training set to obtain $KLD(n) = KLD(h_n, h_g)$, where h_x denotes the histogram of an volume I_x.

Low KLD values indicate a high degree of similarity between volumes. Hence, $KLD(n)$ is thresholded to extract the required set of L training volumes, which are the highly similar to I_g. The threshold value controls the size of the set L and the diversity within the set.

2.2.2.2 Projecting the landmarks from Training volume

The L nearest neighbours of I_g, are used to determine the histogram landmarks for I_g as follows. The landmarks P_i^j, $i \in [1, L]$ from the L volumes are projected onto the intensity range of I_g and then averaged to get the required landmarks for the testing volume. The final landmarks are found as

$$P_g^j = \sum_{i=1}^{L} W_i P_i^j, \forall j$$

$$where, W_i = \frac{w_i}{\sum_{i=1}^{L} w_i}, w_i = \frac{1}{KLD(i)}, \forall i$$

A *weighted* averaging, with weights being inversely proportional to KLD values, is chosen to accommodate the possibility that the training and testing volumes may be quite diverse. In this case, the threshold may have to be lowered to permit L of sufficient size to be identified for accurate landmark identification.

2.2.2.3 Matching to standard scale percentiles

The percentile landmarks P_g^j are matched to the corresponding standard-scale landmarks P_s^j for each tissue type individually to get corresponding transformations T_g^j and a spline fitting is performed to obtain a continuous transformation T_g.

The effectiveness of this approach is illustrated in Fig:2.2 with a sample volume for which tissue masks were available. The transformation curves for mapping a given volume to the standard scale were derived with the proposed method by ignoring the tissue information and with tissue

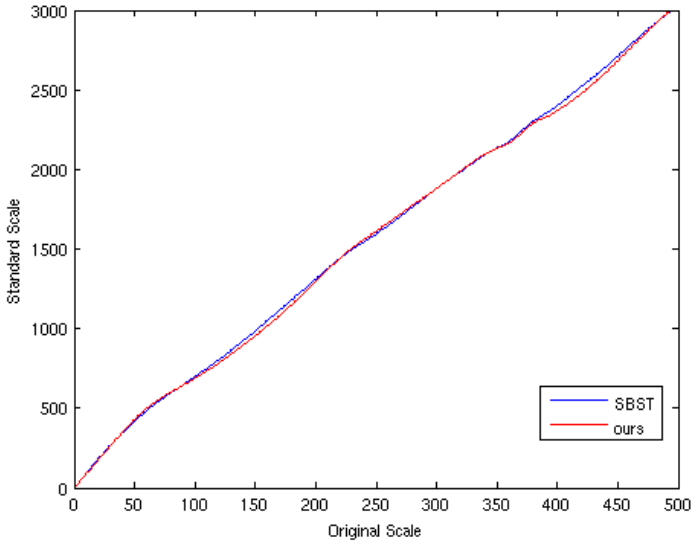

Figure 2.2: Mapping function comparison

information using [31]. These are shown in red and blue, respectively, in Fig:2.2. The two maps are highly overlapping emphasizing the effectiveness of our method.

2.3 Experiments And Results

Dataset: T1 weighted MRI volumes from 1.5 T (GE, Siemens and Phillips) scanners were used to construct a set of $8 \times 3 = 24$ volumes. The volumes from GE and Siemens scanners were sourced from local hospitals while the ones from the Phillips scanner were sourced from [34]. Details of the acquisition parameters for all scanners is given in Table:2.1.

Validation: Preprocessing steps like denoising and intensity in-homogeneity correction were done using [35] and [36] respectively, on all the MRI volumes. Tissue classification was done with the FAST tool of FSL Library [37]. Prior to tissue classification, skull stripping was done

	TE(ms)	**TR**(ms)	**TI**(ms)	**FA**(°)
GE	4.2	10.2	450	15
Siemens	2.9	2370	1000	7
Phillips	4.6	9.83	NA	8

Table 2.1: Scanning parameters.

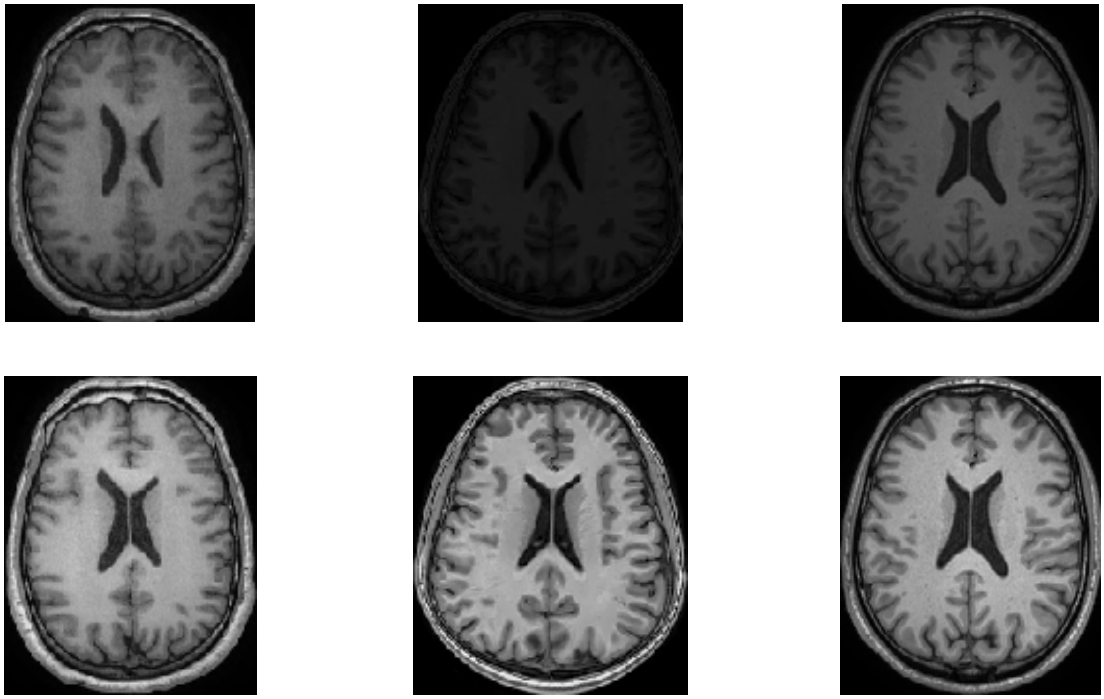

Figure 2.3: Sample images from GE, Siemens and Phillips scanners (left to right) before (top row) and after (bottom row) IS.

using BET [38].

Validation of the proposed method was done using a leave one out approach. All the computations were done on a PC with Intel $i5$ processor, with 4GB of RAM.

Landmarks for the standard scale were learnt from 30 MRI volumes drawn from different scanners, by projecting their landmarks to the voxel value range of 0:4096. The KLD values were thresholded to obtain (on average) L=3 similar volumes for a given test volume.

Sample images from 3 different scanners and results of IS with our method are shown in Fig:2.3. All images were shown at a fixed intensity window range of 0:4096. After IS, the images appear more uniform in terms of intensity.

A major issue in MRI is the intensity variations at the inter-scanner and intra-scanner levels. Our IS method reduces these variations as shown in Fig:2.5 and Fig:2.4 respectively.

Quantitative analysis of IS was done using the Jeffery Divergence (JD), which is defined for two PDFs P and Q as follow:

$$JD(P,Q) = 0.5[KLD(P,Q) + KLD(Q,P)]$$

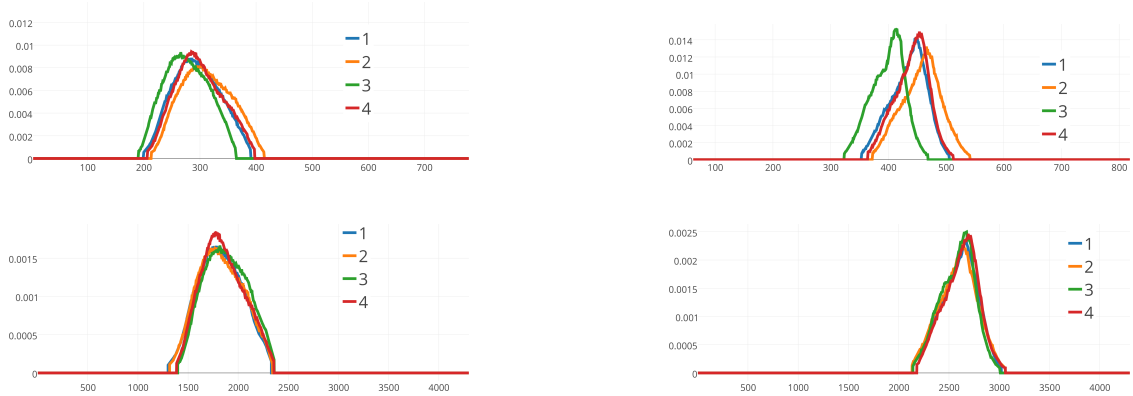

Figure 2.4: Intra-Scanner variation in tissue PDF before (top row) and after (bottom row) IS for the Siemens scanner.

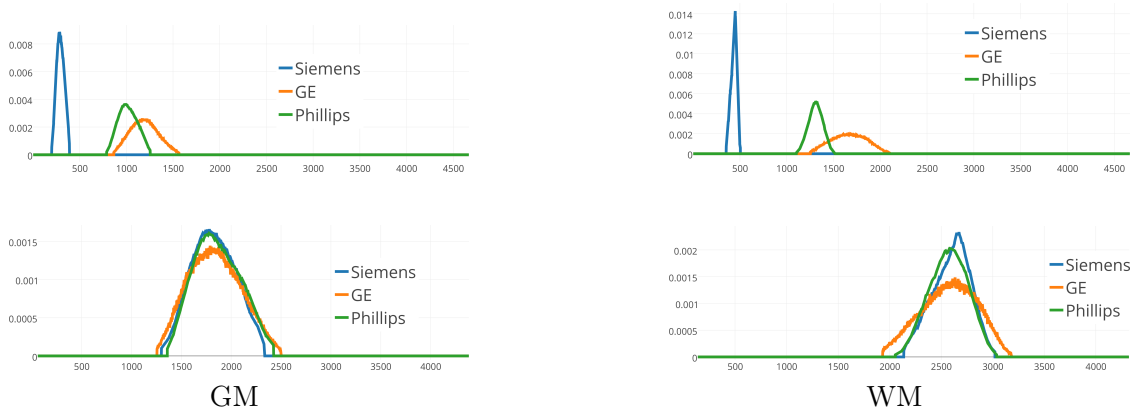

Figure 2.5: Inter-Scanner variation in tissue PDF before (top row) and after (bottom row) IS.

The JD was computed for each tissue class, both across volumes within a scanner and across scanners.

Let $I_n^j(s)$ denote the n^{th} volume from scanner s with tissue mask j; $n = 1, .., 8$, both j and $s = 1, 2, 3$. The intra-scanner ($JD_{intra}(s, j)$) and inter-scanner ($JD_{inter}(s, j)$) divergence were computed as follow:

$$JD_{intra}(s,j) = \frac{1}{mn} \sum_n \sum_m JD(I_n^j(s), I_m^j(s)),$$

		Intra-scanner JD (in x 10^{-2})			**Inter-scanner JD**			**NMI statistics** (across all volumes)		
		G	**S**	**P**	**G vs S**	**G vs P**	**S vs P**	σ_{NMI}	μ_{NMI}	**% CV**
CSF	Before	7.99	6.06	2.04	1.25	0.28	1.10	0.0240	0.1444	16.621
	L4	3.87	2.74	0.97	0.08	0.03	0.05	0.0127	0.2228	5.7001
	ours	3.53	2.40	0.87	0.04	0.03	0.04	0.0055	0.2506	2.1942
	SBST	3.50	2.35	0.84	0.03	0.02	0.03	0.0049	0.2412	2.0315
GM	Before	16.32	9.28	3.34	1.38	0.48	1.23	0.0305	0.2676	11.3999
	L4	7.70	3.54	2.28	0.15	0.08	0.05	0.0128	0.4129	3.1001
	ours	5.78	2.32	1.35	0.04	0.05	0.03	0.0094	0.4444	2.1152
	SBST	5.67	2.28	1.32	0.03	0.03	0.02	0.0084	0.4427	1.8974
WM	Before	19.53	8.71	3.46	1.38	0.88	1.25	0.0285	0.4049	7.0391
	L4	7.56	5.25	2.95	0.44	0.25	0.14	0.0196	0.5792	3.3836
	ours	5.32	2.47	2.27	0.09	0.07	0.05	0.0119	0.6200	1.9193
	SBST	5.19	2.40	2.19	0.07	0.05	0.04	0.0106	0.6205	1.7082

Table 2.2: Quantitative Analysis for all tissue types, across 3 scanners [GE(G), Siemens(S), Phillips(P)]

where, $m \neq n, \forall n, m \in s$

$$JD_{inter}(s_a, s_b, j) = \frac{1}{mn} \sum_{n} \sum_{m} JD(I_n^j(s_a), I_m^j(s_b)),$$

where,$\forall n \in s_a, \forall m \in s_b$

These are listed in Table:2.2. Low values signify similarity among volumes which is desirable especially after IS. The effect of IS with 3 different methods are presented to aid comparison. L4 [27], provides a baseline method as it does not use tissue information at any stage. IS results are comparable for all 3 methods for CSF whereas for GM/WM classes, SBST [31] and our method are (comparable and) better than L4. Similar trend is also seen in the inter-scanner divergence.

We also computed the Standard Deviation (σ_{NMI}),% Coefficient of Variance (CV) and mean (μ_{NMI}) for Normalized Mean Intensity (NMI) before and after IS, for all tissue types for the entire dataset. The NMI of a tissue class is with respect to the maximum intensity in a volume. These are tabulated in Table:2.2. A similar trend as in the case of JD, is observed for σ_{NMI} and CV. However, the μ_{NMI} is comparable for all 3 IS methods across tissues. This is to be expected as the effect of IS is to minimize variation rather than shift the mean intensity of a volume.

2.4 Conclusion

We proposed a hybrid method for IS of MRI volumes which relies on tissue information only during training. A technique for handling the lack of this information in test data was proposed using a KL divergence. Our method outperforms [27] which does not use any tissue information but is at par with [31], which uses prior tissue segmentation for test volume; it is robust to change of scanner. IS of a new volume with our method is significantly fast (as skull stripping and tissue segmentation are not required) and it can be easily adapted to any existing method which requires tissue information.

Chapter 3

Population specific brain template construction for young Indian population

3.1 Introduction

The significant difference in the shape and size of human brains across different races pose a great challenge for functional and structural comparison analysis in neuroscience research [12]. Spatial normalization is a critical pre-processing step in automated neuroimage analysis techniques like Voxel based morphometry (VBM). Spatial normalization involves matching the brain of different individual subjects to a standard co-ordinate framework using various registration methods [39]. Such a representation of a brain using standard co-ordinate framework is known as brain atlas. With the help of the brain atlas, it is possible for researchers to compare and/or combine the findings from different brain imaging modalities (functional or structural, e.g., fMRI, DTI, T1, T2, FLAIR etc.), different brain states (healthy or diseased), and different subjects [10].

The Talairach and Tournoux atlas is the oldest and the most widely used brain atlas [40]. This atlas was created manually from the post-mortem brain sections of 60 year old french women. The slice thickness of the brain varied from 3mm to 4 mm.

With the advances in MRI imaging techniques and availability of abundance of MRI volumes of different individual subjects, it is possible to create a digital atlas which represents the average brain of a population. The digital atlases are created using groupwise registration methods. These methods try to find a common co-ordinate system for a group of images such that all images are at a minimum equal deformation distance from the co-ordinate system. Template is created by warping these image into the co-ordinate system and taking their average. Groupwise registration can either be based on rigid registration or non-rigid registration. While rigid reg-

istration is capable of globally (i.e., translation, rotation, scaling, affine etc.) transforming the target image to align with the reference image, non-rigid registration locally warps the target image to align with the reference image. Both the methods have its pros and cons: rigid registration is faster compare to non-rigid registration (5 mins *vs* 2 hours), but non-rigid registration gives better alignment compared to rigid registration.

The Montreal Neurological Institute (MNI) and the International Consortium for Brain Mapping (ICBM) created the first digital human brain atlas using the brain MRI volume of 305 young normal caucasian subjects (M/F=239/66, 23.4 ± 4.1 years) in the year 1993 [14]. This template was created using 12 parameter linear groupwise registration method. In recent years, MNI and ICBM have released two more brain atlases, namely, MNI-152 [1] and ICBM-452 [41]. These templates are adopted by many groups and incorporated in various MR image analysis tools, such as Statistical Parametric Mapping (SPM) [1], Freesurfer [2], MriStudio [3], and FMRIB Software Library (FSL) [4]. MNI-152 is widely used as standard co-ordinate system in many neuroscience studies. Some of the researchers have also constructed atlases from the multiple scan of an individual subjects, for example Colin27 [42] and French Template [43].

However, these brain templates do not account for differences across phenotypic groups (e.g., age, gender, race, or disease conditions) and they are created using brain MRI of Caucasian population. There have been many studies which show that there is a morphological difference between the brains of Caucasian and Eastern populations [12] [13] as well as between Caucasian and African-Americans [11]. These studies lead to the creation of population specific brain templates, as spatial normalization of non-caucasian brain to Caucasian template can result into mismeasurements [44].

In recent years, several efforts have been made to construct brain templates for eastern population. For example, Chinese template was created using 56 right-handed Chinese young male subjects (24.46 ± 1.81 years) [12]. Similarly, a Korean brain template was also constructed using MRI and Positron-Emission Tomography (PET) images of 78 normal right-handed Koreans aged between 18 to 77 years (M/F=49/29, 44.6 ± 19.4 years) [45].

In the last two years, large scale studies have been undertaken to create Chinese template and Korean Template. In 2015, the Chinese adult brain template based on a multi-center, large

[1] `http://www.fil.ion.ucl.ac.uk/spm/`
[2] `http://freesurfer.net/`
[3] `https://www.nitrc.org/projects/mri_studio/`
[4] `https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/`

scale dataset (over 2000 subjects) was created [5] [21]. Similarly, in 2016, 96 Korean subjects (M/F=48/48) aged 60 years or older (M=69.5Âś6.2 years, F = 70.1Âś7.0 years) were used to construct a Korean normal elderly template (i.e., KNE96 template) [6] [17].

A recent study which compares the neonatal brain images of Chinese, Malay and Indian population found global and regional anatomical variation between these populations which emphasis the need of population specific brain atlas for Indian population [46]. In our study, we have created a population specific brain template for young Indian community (IBA100). We have compared and validated the created template, using global brain features, against the widely used MNI-152 template and other population specific templates like KNE96, Chinese2020 etc.

A good survey of various population specific atlases and description of the automated workflow-based protocol for designing these atlases from structural MRI data can be found in [47] and [48].

3.2 Materials and Methods

Data collection for constructing brain MRI template of young population was an arduous task and it required many organizations to work in coherence. Interested readers are requested to read Appendix(6.1) for the detail of data collection process.

3.2.1 Subjects

100 young healthy adults (M/F=50/50), aged 21-30 years, were recruited for MRI scan at three different hospitals in Hyderabad (CARE hospital, Nampally; Continental hospital and CARE hospital, Hi-tech). All volunteers belong to different states of India. Age distribution and native state distribution of all the volunteers is shown in Fig:3.1 and Fig:3.2 respectively. Medical examination was conducted to exclude subjects with psychological or neurological disorder. All images were examined by an experienced neuroradiologist and reported to have no structural abnormalities. Written consent of all the volunteers was taken to use their MRI scan for atlas construction.

[5] http://www.chinese-brain-atlases.org
[6] http://recode.webnode.kr/news/kne-korean-normal-elderly-group-template/

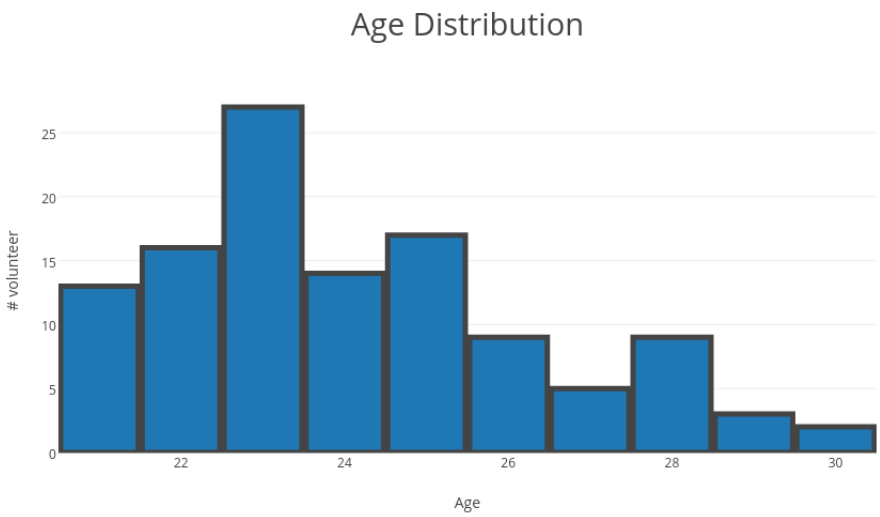

Figure 3.1: Age histogram of 100 volunteers

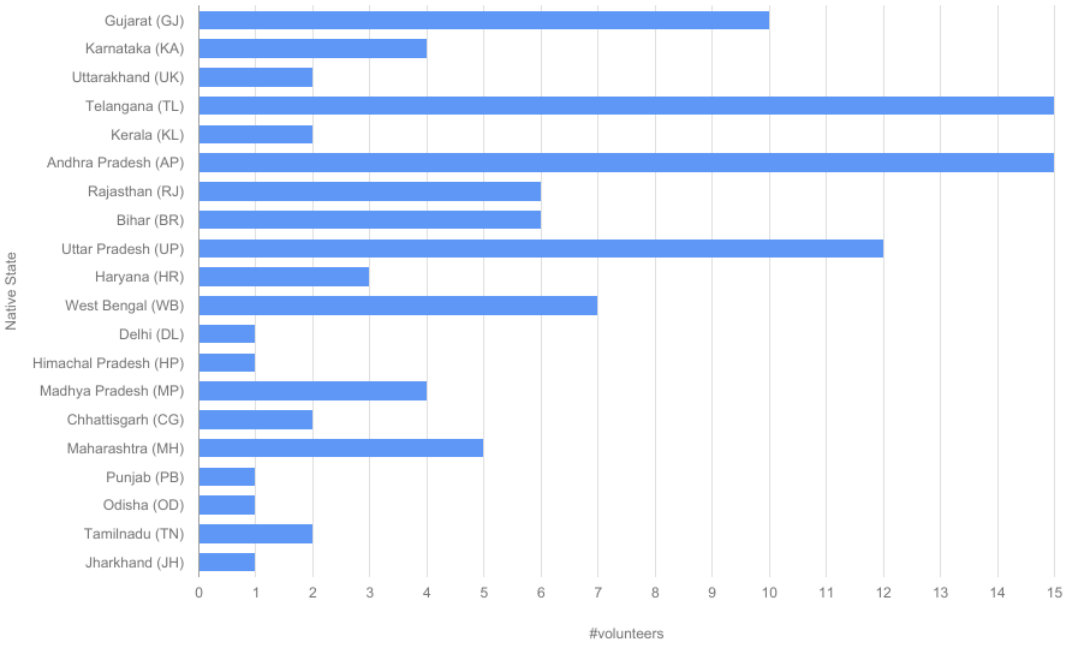

Figure 3.2: Histogram depicting the number of volunteers according to their native state

3.2.2 Image Acquisition

100 subjects were divided in three groups with equal male and female subjects. One group was scanned using Siemens 1.5T MRI scanner for T1 MPRAGE sequence, second group was scanned using GE 1.5T MRI scanner for T1 BRAVO sequence, while the last group was scanned using Phillips 1.5T MRI scanner for T1 3D TFE sequence. MRI volumes were acquired using Sagittal cuts for all the scanners. Quality of the MRI volumes was assessed by an experienced radiologist and was assured to be in good condition.

Scanning Parameters can be found in Table:3.1.

Table 3.1: Scanning Parameters for three different scanners

	Sequence	**TE (ms)**	**TR (ms)**	**TI (ms)**	**Flip Angel**	**Slice Thick. (mm)**	**Acqu. Matrix**	**No. of Slices**	**Voxel Size (mm^3)**
GE	BRAVO	4.2	10.2	450	15°	1	256 x 256	192	1 x 1 x 1
Siemens	MPRAGE	2.9	2370	1000	7°	1	256 x 256	192	1 x 1 x 1
Phillips	3D TFE sense	3.8	8.2	-	7°	1	256 x 256	192	1 x 1 x 1

3.2.3 Data Pre-processing

All 100 brain MRI volumes were pre-processed using standard pipeline, which consisted of N4-Biasfield correction [49] and denoising using Non-local Mean filtering [50]. As we have collected the data from three different scanner manufacturers with different acquisition parameters, intensity range of all the volumes varies significantly. Hence, we need to standardize this intensity range. This was done using histogram matching based intensity standardization method used in [51]. We chose this IS methods as it was an internal part of the ANTs package used for Atlas construction. All the volumes were automatically aligned with respect to their Anterior Commissure (AC) - Posterior Commissure (PC) line using an algorithm proposed in [52], available with ART package [7]. This was further corrected by a senior radiologist using MIPAV tool [8].

3.2.4 Template Construction

We constructed a template using all the 100 T1 MRI volumes using ANTs toolbox [9]. We employed Non-rigid groupwise registration method proposed in [53]. It derives optimal template that is unbiased with respect to both shape and appearance, in the diffeomorphic space, and with fully symmetric methods. The resultant template will henceforth be referred to as IBA100

[7] `https://www.nitrc.org/projects/art/`
[8] `http://mipav.cit.nih.gov/`
[9] `http://stnava.github.io/ANTs/`

template.

In addition, to this we also constructed two gender specific templates (IBA50M and IBA50F) dividing subjects into two subgroups according to their gender and repeating template construction procedure for 50 male and 50 female subjects respectively.

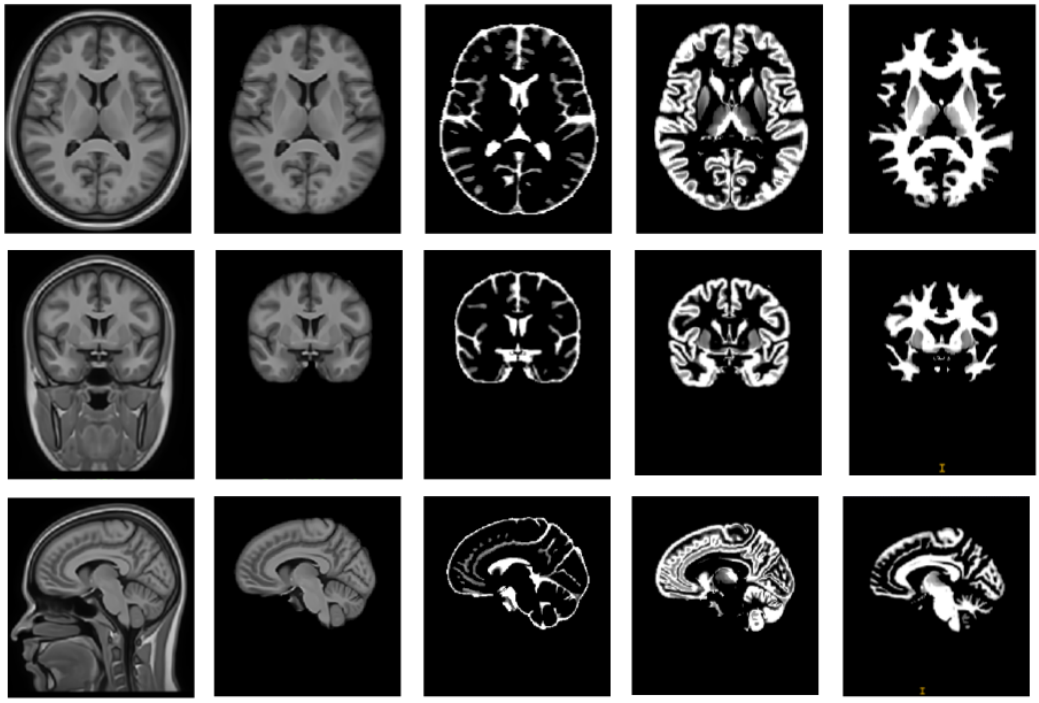

Figure 3.3: Indian brain template (i.e., IBA100) of the young population together with its tissue probability maps. Top to bottom: Axial, Coronal and Sagittal slices. Left to right: MRI, skull stripped brain, Cerebrospinal Fluid (CSF) probability map, Grey Matter (GM) probability map and White Matter (WM) probability map

3.3 Validation

3.3.1 Global brain feature comparison with other Templates

To validate the constructed template against templates available for other population [40] [14] [42] [41] [1] [12] [21] [17], we measured global brain features like length, width, height and volume of whole brain for each template and compared their values. Length, width and height were measured as distance between most anterior (A) and posterior (P), right (R) and left (L), superior (S) and inferior (I) points of cerebrum, respectively. These measurements were per-

Table 3.2: Comparison of Brain Size and shape of Indian Templates and other templates.

	AC-PC (mm)	**Length (mm)**	**Width (mm)**	**Heigth (mm)**	**W/L**	**H/L**	**H/W**	**Volume** (dm^3)
IBA100	25	160	130	88	0.81	0.55	0.68	1.39
IBA50M	25	162	131	91	0.81	0.56	0.69	1.45
IBA50F	24	157	128	86	0.82	0.55	0.67	1.32
Talairach [40]	-	180	146	115	0.81	0.64	0.79	-
MNI305 [14]	28	181	142	111	0.78	0.61	0.77	2.39
MNI152 [1]	28	179	142	110	0.79	0.61	0.77	2.06
ICBM452 [41]	28	176	144	109	0.81	0.57	0.70	1.56
Chinese56 [12]	26	175	145	100	0.83	0.57	0.69	1.89
Chinese2020 [21]	26	162	137	94	0.85	0.58	0.69	1.51
Korean96 [17]	26	160	136	92	0.85	0.58	0.68	1.63
Colin27 [42]	27	176	141	114	0.80	0.65	0.81	2.07

formed by a senior radiologist. Following this, width to length (W/L), height to length (H/L) and height to width (H/W) ratio were also computed. AC-PC length was also computed, which is the distance between Anterior Commissure (AC) and Posterior Commissure (PC) of brain. We automatically segmented brain MRI volume in tissue probability class of grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF) using FAST [37] algorithm of FSL toolbox [10]. Total brain volume was measured as sum of all the voxels for which sum of the three tissue probabilities was more than 0.5. Comparison of these brain features is given in Table:3.2

From this table, we can observe that brain size and shape of western population (MNI305, MNI152 and ICBM452) is significantly higher compare to eastern population (Chinese2020 and Korean96). It should be noted that Chinese56 template was created using MRI scans of Chinese population, which were residing in USA. Hence, shape and size of Chinese56 template is more near to MNI152 template compare to Chinese2020 template. We can also observe that brain of eastern population is highly variable depending on the ethnicity. This underscores the need for the construction of brain template of Indian population. Based on the tabulated results for IBA100, we can conclude that the Indian brain, on an average, is smaller in height, width and volume compare to MNI152 [1], Chinese2020 [21] and Korean96 [17] templates.

3.3.2 Gender Specific Template Validation

We hypothesize that we need separate template for male and female population, as essentially their brain size and shape are different. To validate this we measured GM, WM and CSF vol-

[10] `http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/`

Table 3.3: Comparison of global brain features for Indian young adult subjects in the construction group

	Young Adults	**Male**	**Female**	**p-value**
AC-PC (mm)	25.27 ± 1.29	25.57 ± 1.30	24.99 ± 1.23	0.017
Length (mm)	159.47 ± 7.55	162.39 ± 7.73	156.67 ± 6.24	< 0.001
Width (mm)	130.63 ± 6.10	132.92 ± 6.23	128.43 ± 5.12	< 0.001
Height (mm)	88.75 ± 4.35	90.29 ± 4.02	87.27 ± 4.18	0.001
W/L	0.82 ± 0.05	0.82 ± 0.05	0.82 ± 0.05	0.625
H/L	0.56 ± 0.03	0.56 ± 0.02	0.56 ± 0.03	0.421
H/W	0.68 ± 0.04	0.68 ± 0.03	0.68 ± 0.04	1.000
CSF volume (dm^3)	0.278 ± 0.037	0.293 ± 0.038	0.265 ± 0.031	< 0.001
GM volume (dm^3)	0.646 ± 0.067	0.681 ± 0.069	0.613 ± 0.043	< 0.001
WM Volume (dm^3)	0.516 ± 0.067	0.551 ± 0.065	0.482 ± 0.051	< 0.001
Brain Volume (dm^3)	1.442 ± 0.143	1.526 ± 0.135	1.361 ± 0.096	< 0.001
CSF/Brain	0.194 ± 0.020	0.192 ± 0.019	0.195 ± 0.021	0.477
GM/Brain	0.449 ± 0.021	0.446 ± 0.023	0.451 ± 0.018	0.315
WM/Brain	0.357 ± 0.023	0.361 ± 0.026	0.354 ± 0.020	0.067

umes, in addition to the above mentioned global brain features for all 100 subjects. We divided them according to their gender into two groups of male and female and examined the global brain size difference them using independent t-test (2-tailed). Results are listed in Table:3.3.

From these results we can say that the average male brain is longer, wider and taller in comparison to the average female brain, while W/L, H/L and H/W ratios are comparable for both gender. It can also be observed that the male brain has higher total volume as well as higher CSF, GM and WM volumes compare to the female brain. This was expected as brain volume is directly correlated with size of the brain. This finding that the average female brain is smaller in size compare to the average male brain is consistent with the result for other populations [54] [17].

3.3.3 Template Validation using Linear Registration

We also collected MRI of 15 additional volunteers (8M/7F) for validation purposes. They were divided into three groups of 5 volunteers. Each group was scanned using either GE, Siemens or Phillips scanner using parameters mentioned in Table:3.1. These scans were used to validate the construction of IBA100 template. Validation was done using linear registration as it will compensate any the global difference, like rotation, scaling, translation, affine etc., between the original validation brain and the template brain. This will allow us to compare global brain

Table 3.4: Brain shape and size differences across population templates. Validation data consists of 15(8M/7F) MRI scan of Indian subjects. They were registered onto the MNI152, the Chinese2020 and the IBA100 template spaces separately. Note: P1 was the statistical significance for the measured values of original brains and the brains registering to the MNI152 atlas. P2 was the statistical significance for the measured values of original brains and the brains registering to the Chinese2020 atlas. P3 was the statistical significance for the measured values of original brains and the brains registering to the IBAP50 atlas. *P<0.0001.

Measurement (mm)	**Original Brains**	**Registered to MNI152**	**Registered to Chinese2020**	**Registered to IBA100**	**P - values**		
					P	**P2**	**P3**
AC-PC	25.09 ± 1.39	28.39 ± 1.19	26.52 ± 0.71	25.83 ± 0.93	$< 0.0001^*$	0.1349	0.2347
Length (L)	160.93 ± 4.99	179.80 ± 1.08	168.17 ± 1.45	161.80 ± 1.26	$< 0.0001^*$	0.0028	0.4618
Width (W)	129.93 ± 5.43	142.73 ± 1.28	134.90 ± 2.12	129.93 ± 0.88	$< 0.0001^*$	$< 0.0001^*$	0.8301
Height (H)	88.80 ± 3.41	109.07 ± 1.53	96.77 ± 2.00	89.00 ± 1.73	$< 0.0001^*$	$< 0.0001^*$	0.8177
W / L	0.81 ± 0.04	0.79 ± 0.01	0.80 ± 0.01	0.80 ± 0.01	0.0815	0.0058	0.7336
H / L	0.55 ± 0.02	0.61 ± 0.01	0.57 ± 0.01	0.55 ± 0.01	$< 0.0001^*$	$< 0.0001^*$	0.7014
H / W	0.68 ± 0.04	0.76 ± 0.01	0.72 ± 0.01	0.69 ± 0.02	$< 0.0001^*$	0.6096	0.4920

features, like height, width and length, of the validation set before and after registration to a template space.

Brain MRI volumes of these 15 volunteers (5 male, 24.50 ± 2.67) were aligned to the Indian brain template (IBP100), the Chinese2020 template and the MNI152 template respectively using a 12-parameter transformation as implemented in ANTs. The global brain features were then compared between the original brain (i.e., in the native space) and the brains registered to MNI152 (P1), Chinese2020 (P2), and MNI152 (P3), by using a paired t-test. With the help of this the global deformations required to register a brain to the Indian Template, Chinese2020 template and MNI152 template could be quantitatively evaluated. Similar validation method was recently used in [21] for validating Chinese2020 template and it was observed that the global deformation required to register brains of Chinese subjects to the Chinese2020 template was smaller compare to the global deformation required to register them to MNI152 template. We can expect similar results for Indian population, namely, smaller global deformation should be required to register them to IBA100 template compared to MNI152 or Chinese2020 templates.

From Table:3.4, we can observe that more global deformation is required to register Indian brain to MNI152 and Chinese2020 template as P1 and P2 have lower values for all the global brain features. The lower values of P1 and P2 indicates that the two distributions (orginal brain *vs* registered to MNI152/Chinese2020) are significantly different then each other, which implies the higher global deformation. While this is not true for IBA100 as P3 has higher values for all the brain features, indicating marginal difference between global features of original brain *vs* brain registered to IBA100 templates. This shows that IBA100 template is near to the average Indian brain then MNI152 or Chinese2020 template.

3.4 Conclusion

In this study, we constructed the Indian brain template for young population and validated it against MNI152, Chinese2020 and Korean96 templates. The result of this study indicates that there is a significant difference between global size of Indian (IBA100) and Caucasian brain (MNI152). This difference was noteworthy for all three global measures, i.e. length, width and height. Comparison of Indian template (IBA100) with Chinese (Chinese2020) and Korean (KNE96) shows that Indian brain is comparable in terms of length with both the templates, while it was significantly smaller in terms of height and width.

Findings of this study have many implications for various neuroimage analysis techniques, especially studies involving Indian population. VBM and fMRI analysis which require spatial normalization should benefit from the Indian Template, as there are studies which reports spatial mismatch and mislocalization between brain structures when MNI-152 template was used instead of population specific template for spatial normalization [44].

Future work should focus on further validation of these templates using VBM analysis of new MRI data, by doing registration of that data to the popular MNI152 template and the constructed Indian Template. While we used global features to validate the constructed IBA100 template, local level validation is also possible by comparing the local deformation field required to register Indian brain to IBA100 and MNI152/Chines2020 templates using non-rigid registration. Template for different age group can also be constructed for Indian population to check the effect of age related factors on the average brain size.

Chapter 4

Convolutional Neural Network (CNN) for whole brain segmentation

Quantitative analysis of the neuroimaging data requires cortical and non-cortical structural segmentation. Such analysis is critical in many tasks like the assessment of several neurodegenerative disorders, development of neonatal brain, fMRI studies, connectivity analysis etc. Manual labeling of these structures is unsuitable for studies involving large datasets since it is a slow process and prone to human errors. Automatic segmentation addresses these problems. Though there are different methods for automatic segmentation, a popular one is based on the use of multiple atlases, known as Multi-Atlas Segmentation (MAS).

4.1 Related Work

The most widely used approach for *MAS* is based on non-rigid registration and label fusion. A typical non-rigid registration based method for *MAS* as shown in Fig:4.1, follows these steps: selection of the relevant source atlases (with labeled voxels) from a training set [55][56] and their non-rigid registration to the target volume; label propagation from the source to the target space [15] and finally label fusion [57][58] to combine the propagated labels into a segmentation estimate for the target volume. The non-rigid registration and the label fusion technique determine the accuracy of these methods. This approach for *MAS* is know as Multi-Atlas Label Propagation (*MALP*). A good survey of different methods for *MAS* task can be found in [59].

A key drawback of the *MALP* methods is the computational cost. The non-rigid registration step can require as long as 2-20 hours [60], while the label fusion [58] and atlas selection [56] typically require 2-3 hours. Further, since the total computational time is linearly proportional to the number of training atlases, it linearly increases with increase in the number of training

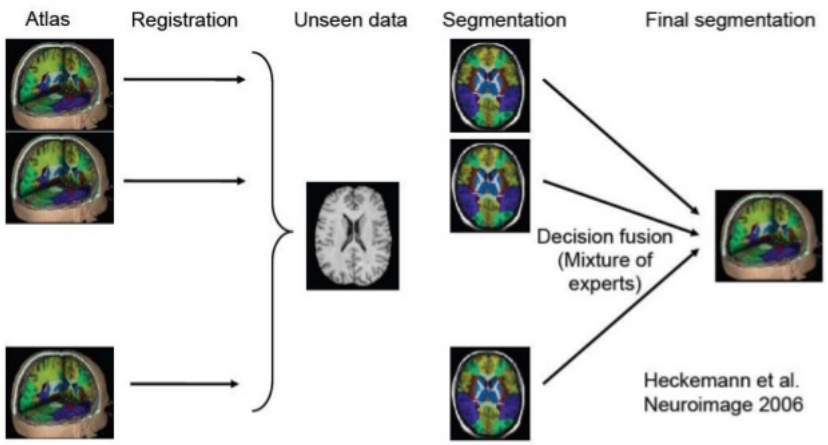

Figure 4.1: Illustration of typical pipeline for Multi Atlas Segmentation.(Image Courtesy: qure.ai)

atlases.

Alternatively, patch based techniques [61, 62, 63] have also been investigated as a solution for *MAS*. Here, the labeling of each voxel is done by comparing its surrounding patch with patches in training data in which the labels of the central voxels are known. These patch based techniques also typically require 2-3 hours for labeling of new volume [6].

A reduction in the computational cost is possible with offline learning [4, 5, 64, 7, 65, 8], where structure segmentation is proposed as a voxel classification problem. Here, a model is learnt on the training data using different machine learning algorithms and used to classify the voxels in an unseen volume in a short time. An example of this is the use of a set of Atlas Forests (AFs) [4]. Here, each AF encodes a single atlas and a probabilistic atlas is constructed by iterative non-rigid groupwise registration of training samples to their mean. During testing, a new volume is first affine registered to the probabilistic atlas followed by a coarse, non-rigid registration, thus leading to computational efficiency.

Spurred by the success of deep learning in computer vision [66, 67], Convolutional Neural Network (CNN) based techniques have been explored for the segmentation of various anatomical structures [68, 69], from the pancreas in CT images[70] to neuronal structures in electron microscopic stacks [71].

In neuroimages, segmentation of images into 3 basic tissue types such as Grey Matter (GM), White Matter (WM) and CerebroSpinal Fluide (CSF) in different age groups has been attempted using CNN [72, 73], while segmenting 8 different tissue types have been reported in [74].

The task of segmenting cortical and sub-cortical structures is similar to the tissue labeling task, albeit, more challenging as the number of structures of interest is generally large, typically 32-134, with the intensity information being inadequate as many structures belong to both GM and WM tissues. Hence, the contextual information for this task is essential to distinguish between structures *and* correctly identify their location in left and right hemispheres of the brain as shown in [64]. Non-rigid registration (of a labeled atlas) has been a natural choice for solving this task as it has embedded contextual information. This contextual information is obtained with the help of propagated atlas priors [15, 75] or alternately the spatial information is provided explicitly as prior probability of structure in patch based approaches [61, 6, 62].

Some recent works take the *CNN* based approach for structure segmentation [64, 65, 7, 8]. Every voxel is represented in [64] by a set of features which includes local appearance information and position relative to the centroids of different sub-structures/segments of the brain as contextual information. Segmenting a new volume involves an iterative 2-step solution: one to generate a rough labeling (for centroid information) and the other, to refine it. These two steps are repeated until convergence.

A Fully Convolutional Network (*FCN*) based CNN approach has also been proposed [7] to segment deep brain structures (typically 10-12 structures). Since contextual or 3D information are not part of the input in this method, it is necessary to use a separate Markov Random Field (MRF) as post-processing for label consistency. Alternately, Hough voting has also been used as the post-processing step [65]. Training a separate *CNN* for each structure of interest has also been proposed for deep brain structure segmentation in [8]. Dynamic random walker with decayed region of interest is then used to enforce label consistency. The use of separate *CNNs* for each structure makes it impractical for scenario where it is necessary to segment up to 134 structures. Thus, existing CNN-based solutions for structure segmentation have the following drawbacks: they are iterative in nature, require postprocessing, do not permit end-to-end training and often require a registration step. Further, most of them have addressed largely labeling of the sub-cortical structures.

4.2 Key Contributions

We propose a non-iterative as well as end-to-end trainable *CNN*-based solution for this task. The salient features of our solution are:

- Novel patches: The *CNN* is trained on 2D and 3D patches to capture 4 types of voxel appearance information which encode local intensity profile as well as global context.
- No requirement for registration: Unlike the existing methods, no registration is required to segment a new volume.
- Comprehensive labeling: Structure segmentation task is been demonstrated for segmentation of both cortical and sub-cortical structures.
- Good performance: A reduction in computational time by a factor of 70 is achieved relative to non-rigid registration based approaches; this is independent of the training set size. The proposed solution also gives comparable or marginally better performance than the current state of the art techniques on 4 out of 5 public datasets in terms of accuracy.

4.3 Methodology

CNN is a deep learning architecture inspired by biological networks akin to the multilayer perceptron. Basic blocks of a *CNN* are (i) A convolutional layer (2D/3D) to detect local features at different positions in an image through a set of learnable filters (or kernels); (ii) A maxpooling layer (2D/3D) which downsamples the output of a layer thus reducing the number of parameters and computation progressively; (iii) A fully connected layer which is an extension of the original multilayer perceptron; (iv) A dropout layer [76] to effectively regularize the network and reduce overfitting to the training data; and (v) Activation functions such as Rectified Linear Unit (ReLU), tanh and softmax [66]. The architecture we propose is shown in Fig:4.2.

In the structure segmentation task, mainly three types of information is useful for segmentation. (i) Local intensity profile to distinguish between structures belonging to different tissues (ii) Context, necessary to encode the spatial configuration of structures and (iii) 3D information for attaining label consistency across slices.

The proposed *CNN* architecture, henceforth referred to as *BrainSegNet*, has these information provided via 2D/3D patches of various sizes. Around every voxel, the following patches are extracted: i) three, 2-D orthogonal patches of size 31x31, extracted from the sagittal, coronal and axial (*sca*) planes to provide local 2D intensity profile (31x31x3), ii) a 3-D patch of size

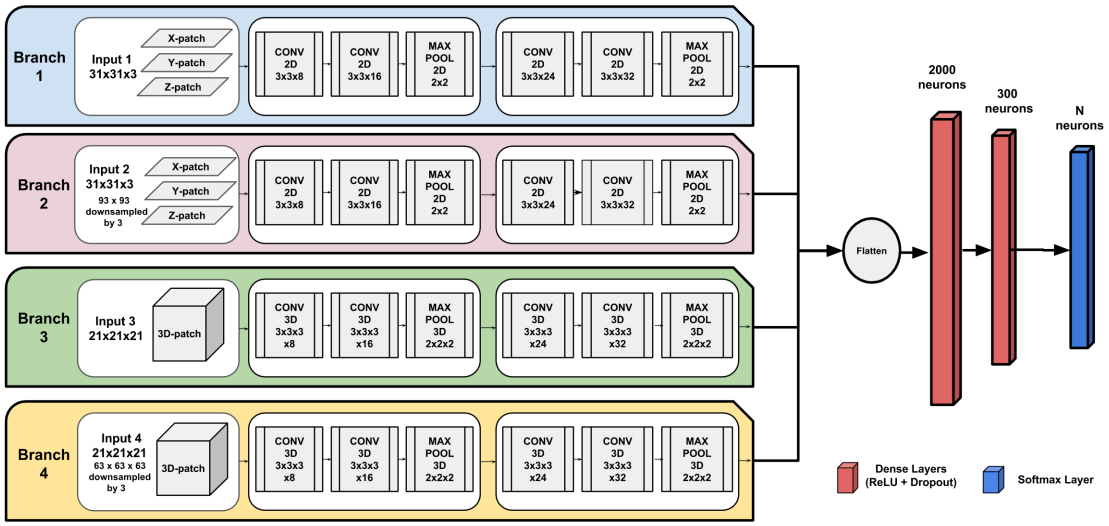

Figure 4.2: Schematic overview of the proposed *CNN* Architecture. The number of neurons N is same as the number of manually marked structures in a dataset (including background).

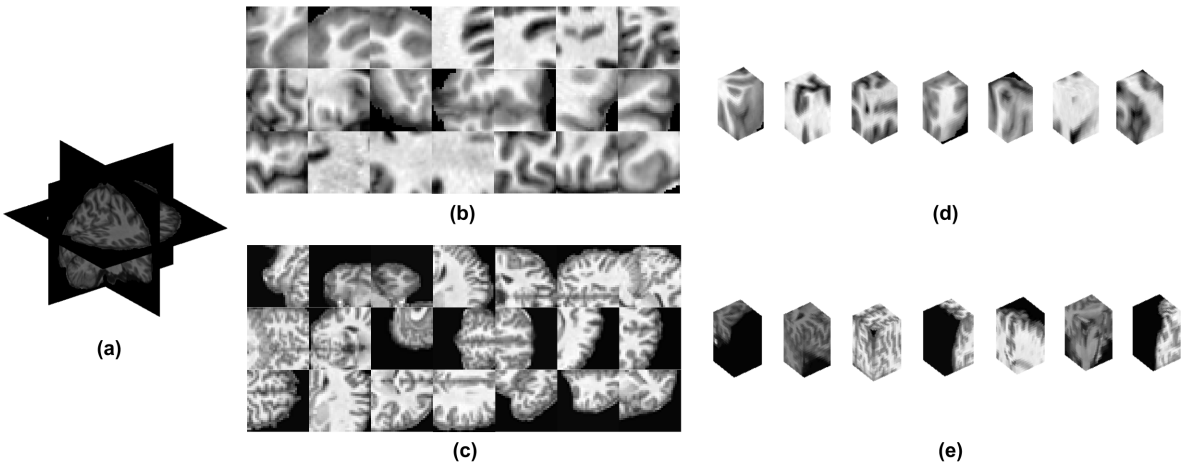

Figure 4.3: Sample input patches. (a) 2.5D representation of the brain MRI volume. For seven different voxels, the Branch 1 (31x31x3) (b), Branch 2 (93/3 x 93/3 x 3) (c), Branch 3 (21x21x21) (d) and Branch 4 (63/3 x 63/3 x 63/3) (e) patches/cubes are also shown. The ordering for (b) and (c) are: coronal (top row), sagittal (middle row) and axial (bottom row) slices.

21x21x21, to provide local 3D intensity profile (21x21x21x1), iii) three 2-D orthogonal patches of size 93x93, down-sampled by a factor of 3, to provide global information regarding input voxel, from the *sca* planes, (31x31x3), iv) a 3-D patch of size 63x63x63, downsampled by a factor of 3, to provide global 3D information (21x21x21x1). Sample patches for 7 different voxels are shown in Fig:4.3.

Inspired by the VGGnet [77], small-sized kernels were chosen for the convolutional layers in our architecture, as shown in Fig:4.2. Each of the four type of patches mentioned above is considered as an input in our network and has a separate processing pipeline for them. Each branch has a cascade of two convolutional layers followed by one maxpooling layer. Each convolutional layer has 3x3(2D) / 3x3x3(3D) kernel size, while the number of filters are 8, 16, 24 and 32 respectively. The maxpooling uses 2x2(2D) / 2x2x2(3D) filters with the stride equal to 2, thus reducing the patch size by half.

The output of all the four branches are flattened and concatenated to form a single 1-D array. This is passed through two Fully Connected (FC) layers with 2000 and 300 neurons respectively, and softmax layers (N neurons, N = total number of desired labels including background) in a sequential manner. A dropout layer (with probability 0.5) is applied between the two FC layers as well as between the last FC layer and the softmax layer (in blue).

A weighted categorical cross entropy function (with respect to classes) is used to handle the class imbalance problem. This loss function and the weights are defined such that the weight increases whenever there are fewer voxels in a particular class.

$$L^i = -\sum_n \sum_l w_l * t^i_{n,l} \log p^i_{n,l} \qquad \text{where, } w_l = \frac{\sum_{k=0}^{k=N} m_k}{m_l}$$

where L^i is the loss for volume i; $t^i_{n,l}$ is 1 if the true label of voxel n of volume i is l otherwise it is 0; $p^i_{n,l}$ is the probability that the *CNN* will predict label l for voxel n of volume i; w_l denotes the weight for the l^{th} class and m_l is the number of voxels of l^{th} class in the training dataset.

4.3.1 Different variants of the proposed *CNN*

As mentioned above, the BrainSegNet has four types of inputs (2D/3D local/global), which contribute to the CNN architecture. Analysis of their relative contribution is possible by considering various architectures.

In general, it can be observed that the variation in inter-structure intensity of non-cortical regions is greater than their spatial positions. This is contrary to the cortical structures, where spatial position (contextual information) varies more than their local intensity information. Thus, the segmentation accuracy should vary with the type of input that is available in the CNN. These observations were verified by considering CNN variants with only one type of input. CNN1: a simple architecture which has only 2D local information as input. Specifically, an input of size 31x31x3 is derived from a 31x31 patch from the three *sca* planes. CNN2a: a network with only 2D *global* information as input. The input is now a patch of size 21x21x3 derived from 63x63 patches from 3 *sca* planes after downsampling by 3. CNN2b: Here, the size of global context considered is wider. In place of the 63x63 patch, a 93x93 patch is used to derive the 2D global information.

One can say that a combination of both local and global information is beneficial for the accurate segmentation of both cortical and non-cortical structures. To observe the performance of this combination of inputs, we considered a variant of the CNN architecture (CNN3): Here, the inputs are both 2D and 3D. Specifically, the input consists of 2D local information from sca planes (31x31x3), 2D global information from sca planes (63x63 downsampled by factor 3, 21x21x3) and 3D local (15x15x15) information. The 3D input patches can provide segmentation consistency between successive slices and remove the requirement of post-processing. The proposed *BrainSegNet* has all the necessary 2D/3D local/global information as input.

An alternative to using both 2D and 3D patches is to use only 3D inputs as in [78]. However, processing 3D patches requires larger GPU memory and increases the computational time (for the same size of input) as 3D convolution is computationally more intensive than 2D. Hence, this variant was not considered. All the 4 variants and the *BrainSegNet* were tested on a public dataset and the results are presented in Section:4.5.1.

4.4 Datasets

BrainSegNet was evaluated on 5 different publicly available datasets, with varying number of structures per volume, as described in Table:4.1. As indicated in the table, except for the MICCAI-2012, every dataset is split randomly into two equal-sized training and testing sets.

Table 4.1: Dataset Description

Dataset	No. of structures	No. atlases	Train Volumes	Test Volumes
MICCAI-2012	134	35	15	20
IBSR	32	18	9	9
LPBA40	54	40	20	20
Hammers67n20	67	20	10	10
Hammers83n30	83	30	15	15

4.4.1 MICCAI-2012 Dataset:

This dataset was released as a part of a workshop on Multi-Atlas Labeling in MICCAI-2012[1]. It consists of 15 training images and 20 testing images from the OASIS [79] project. A detailed description of the acquisition parameters can be found on the OASIS website[2]. All images have 134 manually segmented structures provided by Neuromorphometrics, Inc[3]. The large number of marked structures makes it a challenging dataset for Structure segmentation task.

4.4.2 IBSR Dataset:

The International Brain Segmentation Repository (IBSR) dataset[4] has 18, 3D T1 weighted MR images of 1.5 mm thick cortical slices. Manual Segmentation of 32 structures (primarily non-cortical) is provided by the Center for Morphometric Analysis at Massachusetts General Hospital[5].

4.4.3 LONI-LPBA40 Dataset:

The LONI-LPBA40 dataset consists of T1-weighted MRI scans of 40 healthy volunteers. A total of 50 cortical and 4 subcortical structures along with the brainstem and the cerebellum are delineated by trained raters for each volume. Details of acquisition parameters can be found in [80]. In our work, the brainstem and cerebellum were excluded from assessment, as they were removed by the skull-stripping step[6].

[1] `https://masi.vuse.vanderbilt.edu/workshop2012`
[2] `http://www.oasis-brains.org/`
[3] `http://neuromorphometrics.com/`
[4] `http://www.nitrc.org/projects/ibsr`
[5] `http://www.cma.mgh.harvard.edu/`
[6] `http://loni.usc.edu/atlases/Atlas_Methods.php?atlas_id=12`

4.4.4 IXI Datasets:

Hammers67n20 and Hammers83n30 are 2 sets, consisting of 20 and 30 T1-weighted MR image data, provided as a part of the IXI database[7]. Details of the acquisition parameters can be found in [15]. Hammers83n30 has much more detailed segmentation of gyrus in the frontal and temporal lobs compared to Hammers67n20. This is clearly visible in the column (d) and (e) of Fig:4.4.

4.5 Experiments and Results

All the volumes were pre-processed as follows: Intensity in-homogeneity correction was done using N4-bias correction algorithm [49] followed by skull stripping [38] (BET) [8] and intensity normalization by subtracting mean intensity of a volume and dividing by the standard deviation of a volume.

BrainSegNet was trained on NVIDIA K40 GPU, with 12GB of RAM for 30 epochs using Stochastic Gradient Descent with momentum 0.75 and learning rate 0.05. Learning rate was reduced by half at every 10 epochs. 2 Million patches were extracted from each training volume. The training time was roughly 2 days. The code was written in Python using Keras library. A new test volume can be segmented in 15-20 minutes, which is a *60-70* fold reduction compared to standard non-rigid registration based methods, which requires 20-25 hours for segmentation.

The segmentation performance was quantitatively assessed using the mean Dice Coefficient (DC) which is defined as follows. Let A and B denote the binary segmentation labels generated manually and computationally, respectively. The DC is defined as:

$$DC(A, B) = \frac{2|AB|}{|A| + |B|}$$

where $|A|$ denotes the number of positive elements in the binary segmentation A, and $|AB|$ is the number of shared positive elements by A and B. DC$\in[0, 1]$. A higher DC value indicates better segmentation performance.

Sample axial slices drawn from 5 datasets, their manual segmentation and output of the *BrainSegNet* are shown in Fig:4.4. It should be noted that a *smooth* segmentation of cortical and sub-cortical structures is obtained without any post-processing. This is due to the fact that the

[7] `http://brain-development.org/brain-atlases/`
[8] `http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BET`

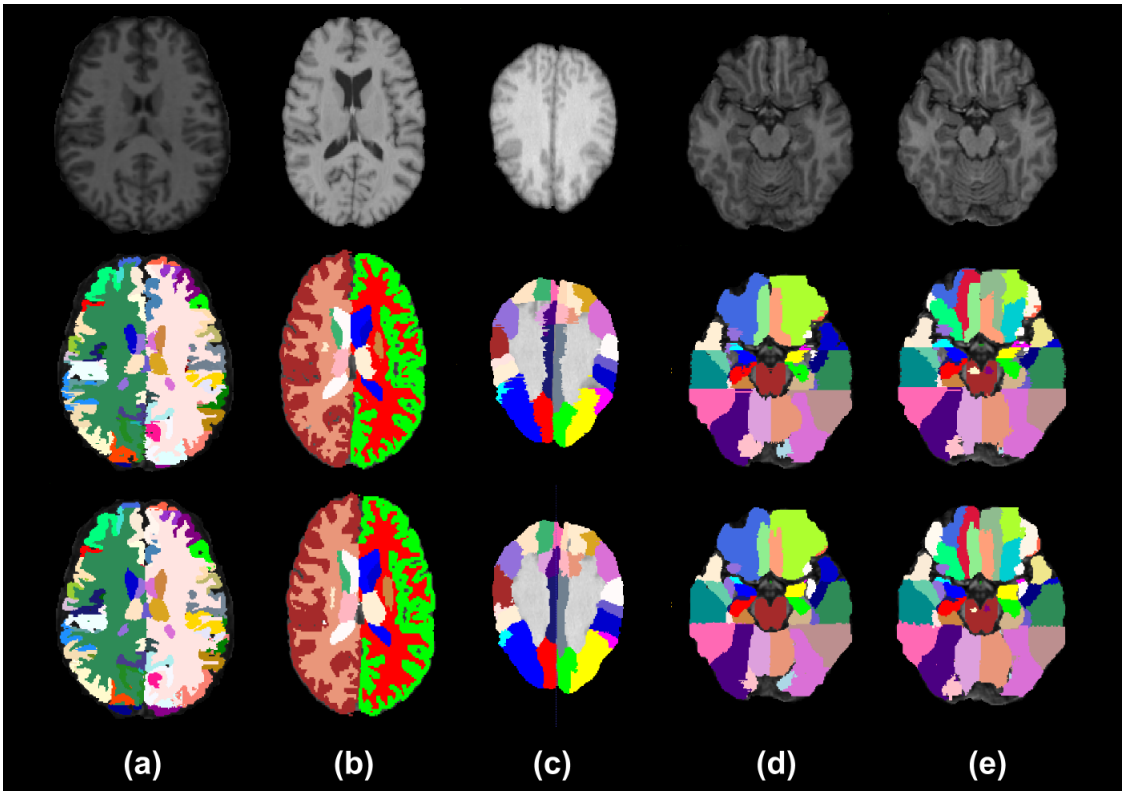

Figure 4.4: Sample images from the different datasets (top row); their manual segmentation (middle row) and output of the *BrainSegNet* (bottom rom). (a) MICCAI-2012 (b) IBSR (c) LONI-LPBA40 (d) Hammers67n20 (e) Hammers83n30.

Table 4.2: Performance (Mean DC) comparison of the *BrainSegNet* with the various methods (MALP based, patch based and classification based) for different datasets. Here, $*$ indicates statistically significant difference ($p < 0.01$) between the proposed and Random Forest based method [4], similarly † indicates statistical significant difference between the proposed and the state-of-the-art method (underlined) for respective datasets. (Bold values indicates the best performing methods)

<table>
<tr><th></th><th colspan="5">Various Segmentation methods</th><th>BrainSegNet</th></tr>
<tr><td rowspan="2">MICCAI 2012</td><td colspan="4">MICCAI challenge [60]</td><td>[4]</td><td rowspan="2">0.7432*</td></tr>
<tr><td colspan="3">Max: <u>0.764</u></td><td>Min: 0.7107</td><td>0.7275</td></tr>
<tr><td>IBSR</td><td colspan="3">[6]
<u>0.835</u></td><td colspan="2">[4]
0.835</td><td>0.844</td></tr>
<tr><td>LONI LPBA40</td><td>[61]
0.783</td><td>[5]
0.784</td><td>[62]
0.799</td><td>[4]
0.801</td><td>[75]
<u>0.814</u></td><td>0.824*†</td></tr>
<tr><td rowspan="2">Hammers 67n20</td><td colspan="5">[15]</td><td rowspan="2">0.840</td></tr>
<tr><td colspan="3">Non Rigid: <u>0.836</u></td><td colspan="2">Affine: 0.754</td></tr>
<tr><td rowspan="2">Hammers 83n30</td><td colspan="5">Feature Based Reg. [81]</td><td rowspan="2">0.808†</td></tr>
<tr><td colspan="3">Max: <u>0.801</u></td><td colspan="2">Min: 0.752</td></tr>
</table>

input patches are rich in information, both local intensity as well as global context is provided.

A comparison of the performance of the *BrainSegNet* with the state-of-the-art standard methods for *MALP* is given in Table:4.2. This table shows that the performance of the *BrainSegNet* is comparable or marginally better than other methods, on all 5 datasets except one.

Since the methods taken up for comparison are computationally expensive, the tabulated results for other methods are drawn from the respective papers without re-implementing them. Hence, to establish the statistical significance (p-value) of the results, the one sample z-test was used.

A detailed evaluation of all the variants listed in Section:4.3.1 was done on MICCAI-2012 dataset and results are provided in Section:4.5.1. A detailed evaluation of the proposed method on all the other datasets is provided in Section:4.5.2.

4.5.1 Performance of different variants of the proposed CNN on MICCAI-2012

The CNN variants were evaluated on the MICCAI-2012 dataset as it has the most number of labeled structures (134). It should be noted that all these architectures were trained using the same number of patches and optimization parameters as mentioned above in Section:4.5.

The mean DC for the cortical and non-cortical structures for different variants are listed in Table:4.3. The box plots for different non-cortical and cortical structures, for these variants are shown in Fig:4.5 and Fig:4.6 respectively.

Table 4.3: mean DC values for different variants of the proposed CNN architecture on MICCAI-2012 dataset. (Bold values indicates the best performing methods)

	Cortical Structures	Non-cortical Structures	Overall
CNN1:	0.6306 ± 0.023	0.7701 ± 0.024	0.6685 ± 0.021
CNN2a:	0.6370 ± 0.011	0.7535 ± 0.024	0.6683 ± 0.010
CNN2b:	0.6576 ± 0.011	0.7604 ± 0.028	0.6852 ± 0.012
CNN3:	0.6758 ± 0.013	0.7793 ± 0.022	0.7036 ± 0.011
BrainSegNet:	**0.7204 ± 0.012**	**0.8053 ± 0.028**	**0.7432 ± 0.019**

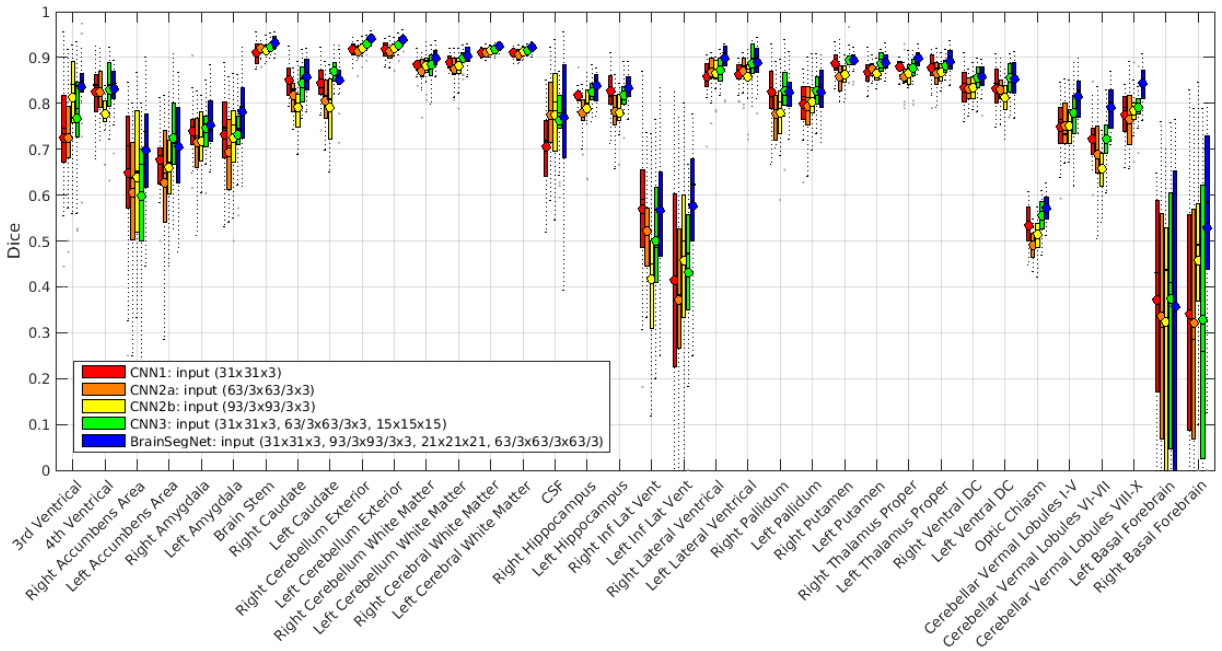

Figure 4.5: Results of labeling of non-cortical structures on test volumes of MICCAI-2012 dataset.

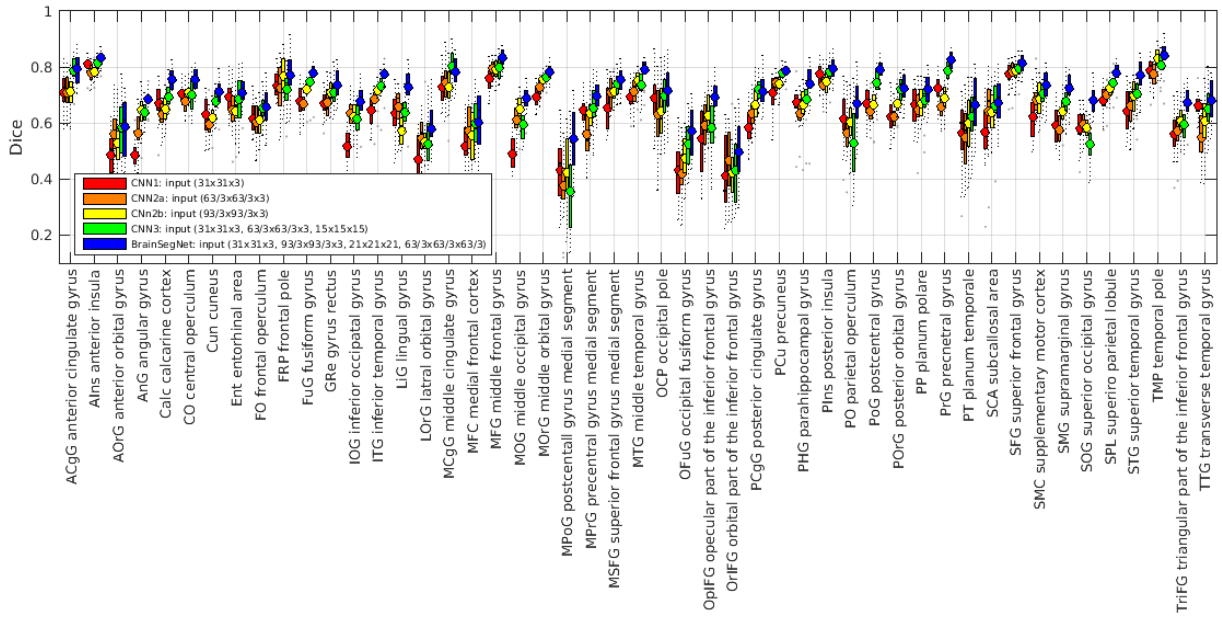

Figure 4.6: Results of labeling of cortical structures on test volumes of MICCAI-2012 dataset (left and right labels shown jointly).

CNN1 is expected to give a superior performance for non-cortical structures rather than cortical structures and vice verse for CNN2a, as CNN1 encodes local information necessary for non-cortical structures while CNN2a encodes global context necessary for cortical structures. This is borne out to be true as CNN1 has better DC (cortical: 0.6306 ± 0.023, non-cortical: 0.7701 ± 0.024) than CNN2a (cortical: 0.6370 ± 0.011, non-cortical: 0.7535 ± 0.024) for non-cortical structures and vice versa for cortical structures.

Similarly, it can also be expected that an increase in the global context should increase the DC. This is also true as CNN2b (0.6852 ± 0.012) has better overall DC than CNN2a (0.6683 ± 0.010), underscoring the beneficial effect of context in cortical structure labeling. However, the improvement in DC for cortical (CNN2a: 0.6370 ± 0.011, CNN2b: 0.6576 ± 0.011) versus non-cortical (CNN2a: 0.7535 ± 0.024, CNN2b: 0.7604 ± 0.028) structures differs, with only a marginal increase for the latter class, which is to be expected.

Since CNN3 has a combination of 2D local/global and local 3D information as input, its performance should be better than the variants with only one type of input. The results in Table:4.3 affirm this, as an improved labeling performance is evident for both cortical (0.6758 ± 0.013) and non-cortical (0.7793 ± 0.022) structures relative to CNN1 and CNN2.

Likewise, the performance of the proposed *BrainSegNet* should be superior to that of all these variants as it has all four global/local 2D/3D information as input. This also holds, as the *BrainSegNet* has the best performance among all variants with a mean DC of 0.7432 ± 0.019 for all structures in the MICCAI-2012 dataset. This is also better than the Atlas Forest based method [4], which reports a mean DC of 0.7275 ± 0.070 ($p < 0.01$).

The ranked results of 25 methods, evaluated during the MICCAI-2012 challenge, are reported in [60]. The obtained performance of *BrainSegNet* places it at rank 5, with the difference in DC value between the best (0.764) and *BrainSegNet* (0.7432 ± 0.019) being 0.02. This is noteworthy as all the methods in [60] are based on (linear/non-linear) registration and none use a *CNN* based solution. It should be noted that this ranking is from 2012 and continues to be valid as of date, since most other methods for brain structure segmentation (including CNN-based) in literature have not reported results on this dataset. This perhaps is due to the large number (134) of labeled structures in this dataset.

4.5.2 Evaluation and comparison of BrainSegNet on multiple datasets

In addition to MICCAI-2012, *BrainSegNet* is also evaluated on 4 other public datasets. We present the performance figures for the whole brain as well as for the individual structures on each of these datasets. These individual structures are chosen according to the reports in the literature on each datasets. It should be noted that the mean DC values, reported from respective papers, are based on leave-one-out assessment whereas ours is based on half-split assessment.

4.5.2.1 IBSR Dataset

This dataset has the fewest labeled structures. The mean DC obtained for this dataset with *BrainSegNet* is 0.844 ± 0.031, which is marginally higher than those reported for the current state-of-the-art methods such as [6] (0.835) and [4] ($0.835 \pm 0.042, p = 0.1467$). Table:4.4, lists the mean DC values for non-cortical structures, for three methods: proposed, a patch based [6] and a random forest based method [5]. It can be observed that proposed method gives equal or marginally better performance than patch based method [6] on 13 out of 18 structures, while it is consistently better than Random Forest based method [5] for all structures.

Fully Convolutional Network (FCN) based CNN method in [7] and [8] also report on this dataset for sub-cortical structure segmentation. A comparison of their reported DC values and ours is shown in Table:4.5. *BrainSegNet* appears to perform superior compared to both [7] and [8]. It should be noted that these methods utilize post-processing steps like MRF and random walker. A reason for the increased performance of our method, which has no post-processing, can be attributed to the better use of contextual information in the proposed architecture.

4.5.2.2 LONI-LPBA40 Dataset

In this dataset [80], largely *cortical* structures are labeled. *BrainSegNet* has a mean DC value of 0.824 ± 0.040 for this set, which is higher than 0.814 ± 0.023 ($p < 0.01$) reported by the current state-of-the-art patch based technique [75]. The random forest based method [5] also uses a classification approach like the proposed method. The DC values for 54 structures in LPBA40 dataset are listed in Table:4.6 for comparison. Based on these values, *BrainSegNet* appears to perform consistently better.

Table 4.4: Quantitative Comparison on IBSR dataset. mean DC values are listed for non-cortical structures. Here, $*/\dagger$ indicates the statistically significance difference ($p < 0.05$) between *BrainSegNet* and [5]/[6]. (Bold values indicates the best performing methods)

	[5]	[6]	BrainSegNet
L. Lateral Ventrical	0.85	**0.93**	**0.93 ± 0.039***
L. Thalamus	0.88	**0.89**	0.88 ± 0.050
L. Caudate	0.83	**0.88**	0.86 ± 0.047
L. Putamen	0.84	0.89	**0.91 ± 0.022***†
L. Pallidum	0.74	0.79	**0.81 ± 0.089***
L. Hippocampus	0.74	**0.83**	0.81 ± 0.065*
L. Amygdala	0.68	0.75	**0.76 ± 0.087***
L. Ventral DC	0.81	**0.82**	**0.82 ± 0.033**
3rd Ventrical	0.74	0.80	**0.81 ± 0.079***
4th Ventrical	0.76	0.84	**0.85 ± 0.094***
R. Lateral Ventrical	0.85	**0.92**	**0.92 ± 0.026***
R. Thalamus	0.87	0.89	**0.90 ± 0.029***
R. Caudate	0.81	**0.89**	0.88 ± 0.048*
R. Putamen	0.84	0.89	**0.91 ± 0.023***†
R. Pallidum	0.75	0.79	**0.83 ± 0.086***†
R. Hippocampus	0.76	**0.83**	**0.83 ± 0.071***
R. Amygdala	0.66	**0.75**	0.71 ± 0.087
R. Ventral DC	0.81	**0.82**	**0.82 ± 0.051**

Table 4.5: Quantitative comparison on IBSR dataset for *CNN* based methods. DC values are shown for sub-cortical structures used by [7] and [8] for evaluation. Here, $*/\dagger$ indicates the statistically significance difference ($p < 0.05$) between *BrainSegNet* and [7]/[8].

	[7]	[8]	BrainSegNet
Thalamus	0.87	0.89	**0.89 ± 0.041**
Putamen	0.83	0.88	**0.91 ± 0.022***†
Caudate	0.78	**0.87**	**0.87 ± 0.047***
Pallidum	0.75	0.79	**0.82 ± 0.083***†
Hippocampus	0.77	0.81	**0.82 ± 0.066***
Amygdala	0.65	0.67	**0.74 ± 0.089***,†

4.5.2.3 IXI Datasets

IXI database consists of two datasets with 67 and 83 labeled structures as mentioned in Section:4.4.

Hammers67n20 Dataset[15]: *BrainSegNet* obtains a mean DC of 0.840 ± 0.013 on this dataset, which is comparable to 0.836 ± 0.009 ($p = 0.35$) reported in [15] with non-rigid registration. [15] also reports a lower figure of 0.754 ± 0.016 ($p < 0.01$) with affine registration. It should be noted that though [15] reports on 30 MRI scans, only 20 scans are available for public access.

Hammers83n30 Dataset[15]: *BrainSegNet* has a mean DC of 0.808 ± 0.014 on this dataset, which is better than that obtained by different non-rigid registration based methods reported in [81] such as demons ($0.785, p < 0.01$), PCA ($0.754, p < 0.01$), HAMMER ($0.789, p < 0.01$), ISA ($0.801, p < 0.01$).

Table:4.7 lists the DC values obtained with the proposed method (*BrainSegNet*) and those reported by [75] for sub-cortical structures common to both the above 2 datasets. It is desirable for a method to perform equally well on same structures from both the datasets. This is true for our method. It can also be observed that *BrainSegNet* has equal or marginally higher DC than [75] for 4 out of 7 structures while it is marginally less for the rest.

Finally, the generalizability of the obtained results with *BrainSegNet* was also assessed. Training was done with the MICCAI-2012 dataset and testing was done on the IXI datasets (Hammers67n20 and Hammers83n30). As subcortical structures are common between these three datasets, the performance is compared for only these structures in Table:4.8. The tabulated results indicate only marginal degradation with cross training which implies that the proposed method is robust to change in dataset. This eliminates the need for re-training for a new dataset.

4.6 Discussion and Conclusion

Traditional approaches like MALP for the brain structure segmentation rely on non-rigid registration to label a new test volume. Various machine learning based algorithms like random forest [4] and *CNN* [64, 7, 65, 8] treat structure segmentation as a classification task. These existing *CNN* based methods however have the following drawbacks: multiple passes over test image to obtain final labels and dependency on initial model [64]; need for post-processing (such as MRF [7], Hough voting [65] or random walker [8]), which prohibits an end-to-end trainable

CNN-based solution.

We presented an end-to-end trainable *CNN* based solution (*BrainSegNet*) for brain structure segmentation by employing 2D/3D patches of varying size as input. The experimental results of the variants of the proposed method (see Table:4.3) demonstrated that both context (branches 2 and 4) and appearance (branches 1 and 3) are important for labeling, with wider context boosting the DC values of cortical more than non-cortical structures.

The results on 5 different publicly-available datasets and especially the IXI datasets (with 67 and 84 labeled structures), indicate that DC tends to degrade with an increase in the number of labeled structures in a dataset. This suggests that the source of degradation could potentially be either a class-imbalance problem, which arises in *MAS* problem with large number of classes, or the chosen evaluation metric. Hence, we computed the *Accuracy* of segmentation for the 2 IXI datasets with 67 and 83 structures. It was found that while the mean *Accuracy* was 0.868 for both datasets, the mean DC values were different. This can be attributed to the fact that the loss function optimises *Accuracy* and not DC. Thus, the degradation appears to be due to the evaluation metric (which was chosen based on its popularity in literature) and not the class imbalance. In the future, it would be of interest to explore a loss function which optimizes DC to confirm the above observation.

It was also observed that labeling error was higher in the brain-skull boundary region. Hence, a brain mask was generated using the 'GroundTruth' mask. With this modification, the mean DC was found to increase by $\sim 1\%$ for all the datasets (new values for MICCAI: 0.7542, IBSR: 0.853, LONI-LPBA40: 0.834, Hammers67n20: 0.849 and Hammers83n30: 0.815). Hence, skull stripping does play a role in performance.

The proposed *BrainSegNet* adopts a classification-based approach for brain structure segmentation. A comparative analysis done against other classification-based approaches, patch-based and registration-based methods on 4 datasets showed that *BrainSegNet* has comparable or marginally better performance and at a reduced computational time. (see Table:4.2). Experimental results in Section:4.5.2.3, showed it to be robust to changes in datasets obviating the need to retrain when the structures of interest are common across the datasets. This coupled with the fact that a non-rigid registration step is not required should make the proposed solution attractive for many applications where computational time plays a critical role.

Table 4.6: Quantitative Comparison on LPBA40 dataset. [5] is based on random forest. DC values are listed for 54 structures. Here, $*$ indicates the statistically significance difference ($p < 0.05$) between *BrainSegNet* and [5]. (Bold values indicates the best performing methods)

	Left Brain		Right Brain	
	[5]	BrainSegNet	[5]	BrainSegNet
Superior Frontal Gyrus	0.86 ± 0.024	**0.89 ± 0.023***	0.86 ± 0.020	**0.89 ± 0.024***
Middle Frontal Gyrus	0.85 ± 0.029	**0.89 ± 0.031***	0.85 ± 0.031	**0.88 ± 0.026***
Inferior Frontal Gyrus	0.79 ± 0.046	**0.85 ± 0.025***	0.80 ± 0.034	**0.85 ± 0.036***
Precentral Gyrus	0.81 ± 0.042	**0.86 ± 0.040***	0.82 ± 0.039	**0.84 ± 0.047***
Middle Orbitofrontal Gyrus	0.75 ± 0.069	**0.81 ± 0.067***	0.75 ± 0.068	**0.80 ± 0.056***
Lateral Orbitofrontal Gyrus	0.69 ± 0.096	**0.76 ± 0.061***	0.70 ± 0.073	**0.72 ± 0.082**
Gyrus Rectus	0.76 ± 0.051	**0.79 ± 0.068***	0.75 ± 0.051	**0.79 ± 0.082**
Postcentral Gyrus	0.77 ± 0.052	**0.83 ± 0.051***	0.78 ± 0.071	**0.84 ± 0.041***
Superior Parietal Gyrus	0.80 ± 0.040	**0.85 ± 0.031***	0.81 ± 0.029	**0.84 ± 0.029***
Supramarginal Gyrus	0.74 ± 0.066	**0.79 ± 0.054***	0.75 ± 0.073	**0.78 ± 0.070***
Angular Gyrus	0.75 ± 0.041	**0.76 ± 0.070**	0.74 ± 0.070	**0.79 ± 0.045***
Precuneus	0.77 ± 0.043	**0.79 ± 0.064**	0.77 ± 0.039	**0.82 ± 0.053***
Superior Occiital Gyrus	0.69 ± 0.075	**0.73 ± 0.090***	0.71 ± 0.073	**0.72 ± 0.081**
Middle Occiital Gyrus	0.77 ± 0.048	**0.79 ± 0.064**	0.78 ± 0.050	**0.80 ± 0.063**
Inferior Occiital Gyrus	0.75 ± 0.056	**0.81 ± 0.056***	0.76 ± 0.057	**0.82 ± 0.054***
Cuneus	0.74 ± 0.072	**0.83 ± 0.067***	0.75 ± 0.067	**0.78 ± 0.082**
Superior Temporal Gyrus	0.84 ± 0.027	**0.87 ± 0.031***	0.84 ± 0.038	**0.88 ± 0.033***
Middle Temporal Gyrus	0.78 ± 0.040	**0.81 ± 0.045***	0.76 ± 0.046	**0.83 ± 0.037***
Inferior Temporal Gyrus	0.78 ± 0.051	**0.83 ± 0.047***	0.76 ± 0.048	**0.83 ± 0.039***
Parahippocampal Gyrus	0.79 ± 0.039	**0.82 ± 0.053***	0.79 ± 0.036	**0.83 ± 0.044***
Lingual Gyrus	0.80 ± 0.054	**0.86 ± 0.028***	0.79 ± 0.057	**0.85 ± 0.041***
Fusiform Gyrus	0.80 ± 0.051	**0.85 ± 0.055***	0.81 ± 0.044	**0.86 ± 0.044***
Insular Cortex	0.84 ± 0.027	**0.88 ± 0.033***	0.86 ± 0.020	**0.87 ± 0.038**
Cingulate Gyrus	0.77 ± 0.065	**0.79 ± 0.042**	0.79 ± 0.044	**0.80 ± 0.051**
Caudate	0.81 ± 0.046	**0.84 ± 0.060***	0.81 ± 0.064	**0.82 ± 0.076**
Putamen	0.82 ± 0.028	**0.83 ± 0.039**	0.81 ± 0.028	**0.85 ± 0.043***
Hippocampus	0.81 ± 0.026	**0.83 ± 0.036***	0.81 ± 0.048	**0.83 ± 0.038***
Overall	0.78 ± 0.048	**0.82 ± 0.039***	0.79 ± 0.048	**0.82 ± 0.041***

Table 4.7: Quantitative comparison on two variants (Hammers67n20 and Hammers83n30) of IXI database. DC values are given for sub-cortical structures that are common for both the datasets. (Bold values indicates the best performing methods)

	[75]	BrainSegNet	
		Hammers67n20	Hammers83n30
Hippocampus	**0.85**	0.84 ± 0.038	0.83 ± 0.035
Amygdala	**0.82**	0.81 ± 0.089	0.81 ± 0.070
Caudate	**0.90**	**0.90 ± 0.029**	0.88 ± 0.026
Nuc. Accumbens	**0.71**	0.70 ± 0.112	0.68 ± 0.105
Putamen	**0.89**	**0.89 ± 0.031**	**0.89 ± 0.039**
Thalamus	0.90	**0.91 ± 0.015**	0.90 ± 0.021
Pallidum	0.80	**0.81 ± 0.076**	0.80 ± 0.066

Table 4.8: Quantitative comparison of two variants (Self-training and Cross-training) on IXI databases. DC values are given for sub-cortical structures that are common for both the datasets.

	Self-training (same dataset)		Cross-training (MICCAI-2012 dataset)	
	Hammers67n20	Hammers83n30	Hammers67n20	Hammers83n30
Hippocampus	0.84 ± 0.038	0.83 ± 0.035	0.82 ± 0.042	0.82 ± 0.039
Amygdala	0.81 ± 0.089	0.81 ± 0.070	0.80 ± 0.096	0.80 ± 0.078
Caudate	0.90 ± 0.029	0.88 ± 0.026	0.89 ± 0.022	0.88 ± 0.029
Nuc. Accumbens	0.70 ± 0.112	0.68 ± 0.105	0.69 ± 0.119	0.69 ± 0.101
Putamen	0.89 ± 0.031	0.89 ± 0.039	0.88 ± 0.043	0.89 ± 0.043
Thalamus	0.91 ± 0.015	0.90 ± 0.021	0.90 ± 0.017	0.89 ± 0.023
Pallidum	0.81 ± 0.076	0.80 ± 0.066	0.80 ± 0.081	0.80 ± 0.073

Chapter 5

Fully Convolutional Neural Network (FCNN) for sub-cortical structure segmentation

Diagnosis of neuro-degenerative diseases, analysis of development of neonatal brain etc., rely heavily on quantitative and qualitative measurements of different human brain structures, especially deep brain structures [82]. For instance, morphometry and volumetry of hippocampus plays an important role in Alzheimer's disease assessment [18] while that of putamen is relevant for obstructive sleep apnea patients [83]. Other deep brain structures also plays a useful role in diagnosis of different disorders like schizophrenia, attention deficit, and subtypes of epilepsy [82]. Thus, segmentation of these deep brain structures, *in less time*, is critical. A pictorial view of these deep brain structures is givein in Fig:5.1

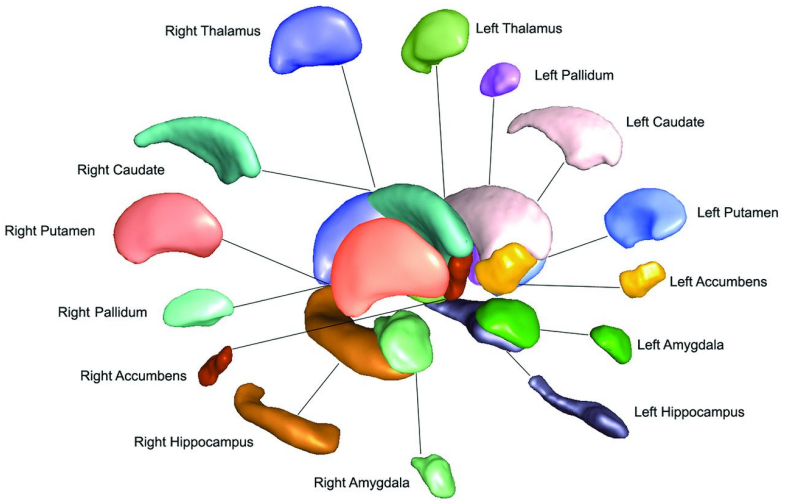

Figure 5.1: Pictorial view of seven subcortical structures (Image Courtesy: [3])

5.1 Related Work

This problem of automatic segmentation is generally solved either using Non-rigid Registration [15] or Model based techniques [84], with both relying on training atlases (with manual segmentation) to segment new volumes, albeit in different ways. The former class of techniques label a new volume by registering (non-rigid) training atlases to it and then fusing the propagated labels whereas, the latter techniques predict the labels from a mathematical model learnt from the training atlases. Model based techniques typically require 15-20 *minutes* to label new volumes as against 20-25 *hours* required by registration based methods [59].

Here, we focus on model based segmentation techniques given their computational efficiency and hence potential for clinical use. FSL-FIRST is a popular technique based on statistical shape modeling [84]. Here, the relationship between shape and intensity are modeled via the conditional distribution of intensity given shape aiding in producing smooth shapes. The smooth shapes is useful for shape analysis of structures like hippocampus which plays an important role in tracking progression of Alzheimer diseases.

Voxel-level classification models have also been used for deep brain structure segmentation. A set of random forests are learnt in [85] for each training atlases separately using intensity and contextual features, majority voting of the predicted output of these set of random forests is used to label voxels in a new volume. Deep learning techniques, especially Convolution Neural Networks (CNN), have also been explored for many medical image analysis tasks (ex. [72]) including deep brain segmentation, given the network's ability to learn features for the given task from training data. A Multi-Scale CNN (MS-CNN) architecture is proposed in [8] using intensity as well as contextual information as inputs. Final voxel labeling is achieved by postprocessing the CNN output using Random Walker (RW) based graphical model. A Fully Convolutional Network (FCN) based on 2D-CNN architecture is proposed in [7] to label voxels with a single slice of MRI volume as input, here label consistency across slices is achieved using Markov Random Field (MRF). Both [8] and [7] perform well in terms of labeling accuracy but need separate graphical models-based post-processing for final segmentation which precludes end-to-end CNN training in addition to increasing the run time.

5.2 Proposed method

We present a novel CNN architecture, which eliminates the need of any post-processing step making it end-to end trainable. It is designed to leverage 3D information around a slice at the

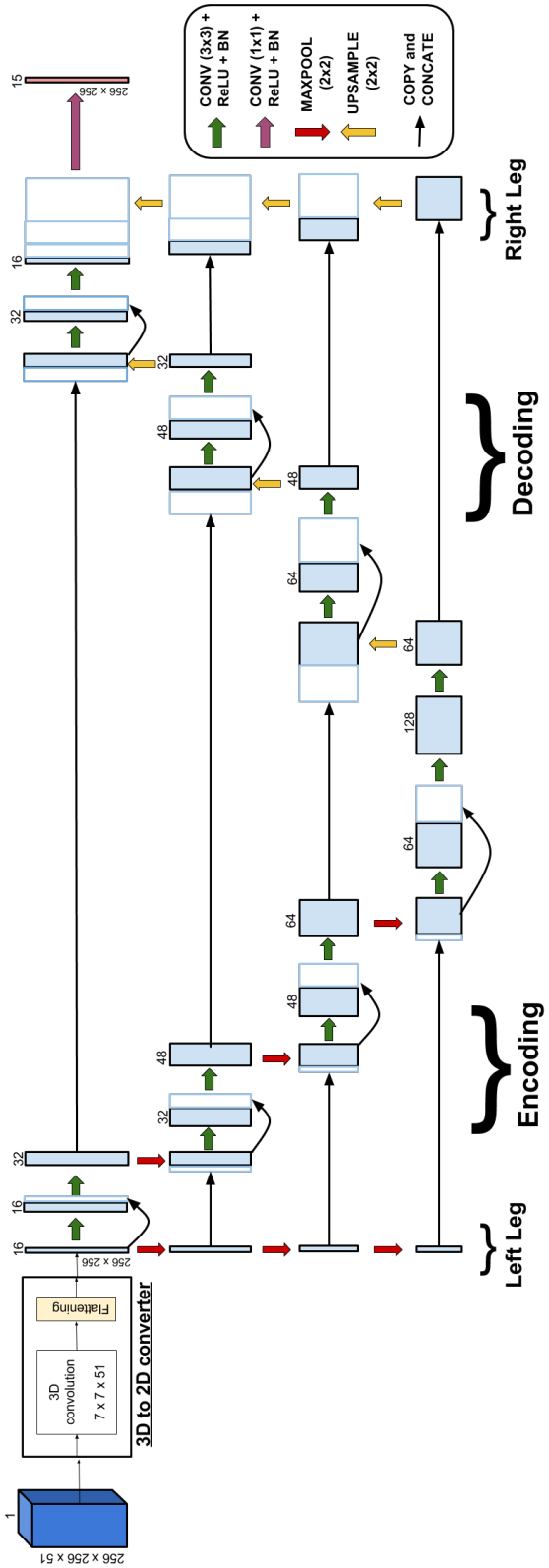

Figure 5.2: Schematic representation of the *M-net* CNN architecture. Solid blue boxes represent multi-channel feature maps. Blue framed boxes represent copied feature maps. Number of channels is denoted on the top of the box.

input level and yet operate only on 2D information beyond the first stage to produce a labeled slice as output. This ensures labels are consistent and accurate across slices without using any post-processing steps, which in-term reduces run-time and memory requirement.

CNN is a deep learning architecture inspired by the biological networks akin to the multilayer perceptron. It has been widely used for the segmentation and recognition tasks. Basic blocks of a CNN are Convolutional Layer (2D/3D), Maxpooling layer (2D/3D), Fully Connected Layers, Dropout layer [76], Activation Functions like Rectified Linear Unit (ReLU), tanh, Softmax functions, and Batch Normalization Layer [86]. For a detailed description of these blocks readers are refered to [66]. The proposed architecture is shown in Fig:5.2 .

Our architecture is inspired by the U-net proposed for 2D and 3D microscopy segmentation in [71] and [87], respectively. In [71], a 2D CNN architecture is used for segmentation of neuronal structures in electron microscopic stacks while in [87], the same architecture with 3D filters is used for segmentation of the Xenopus kidney from 3D volumes of confocal microscopy. Although, both give good performance for their respective tasks, they are not appropriate for segmentation of MRI volume of size 256x256x256, as 2D U-net does not utilize any 3D information, while 3D U-net does it but at the cost of a high memory ($\sim$10 GB) requirement for a small input of 200x200x50. The latter is a bottleneck as maximum memory available in most advanced GPU is only 12 GB.

We address this issue of memory constraint, while still utilizing necessary 3D information for MRI segmentation, in a novel way in our proposed CNN architecture, henceforth known as *M-net*. As shown in Fig:5.2, a slice s and its neighbours s-n:s+n are used to form a stack which serves as an input. The value of n is determined empirically. This allows us to utilize 3D information. The stack of slices is passed through a *3D-to-2D* converter block, which learns a 3D convolution filter of size 7x7x(2n+1), to combine the stack of 2D slices into one single 2D slice $\bar{s}$. This $\bar{s}$ is then processed through the *M-net* architecture to obtain the desired segmentation, as shown in Fig:5.2. Thus, segmentation of a whole volume is done slice by slice.

M-net has mainly 4 pathways of 2D filters, through which $\bar{s}$ is passed, two main encoding and decoding paths, as well as two side paths which gives our architecture functionality of deep-supervision [88]. Each pathway has 4 steps. In the encoding path, each step has a cascade of 2D convolution filters of size 3x3 and maxpooling by 2x2, which reduces the size of input by half and allows network to learn contextual information. In the cascade of convolution filters, skip connection is introduced to enable the network to learn better features [89]. The decoding layer is identical to encoding layers with one exception: maxpooling is replaced by upsampling

layer by 2x2 to double the size of input and allow recovery of an output image of original size. Similarly, skip connections are also implemented between corresponding encoding and decoding layers, as shown in Fig:5.2. This type of design ensures the network has sufficient information to derive fine grain labeling of an image while eliminating the need for any post-processing. The left leg operates on $\bar{s}$ with 4 maxpooling layers of size 2x2 and the outputs are given as inputs to the corresponding encoding layers. The right leg upsamples the output of each of the decoding layers to the original size of $\bar{s}$. Finally, the output of the decoding layer and the right leg is processed by a 1x1 convolution layer with L channels, where L is the number of structures of interest including background. Dropout (with probability 0.3) [76] and batch normalization (BN) [86] are applied after each step and each convolution layers respectively, to reduce overfitting. For all the layers except the last, a ReLU activation is applied after every convolution layer. For the last layer, a softmax activation is applied, which gives the probability of each voxel belonging to any particular structure. The final label for any voxel is the structure with maximum probability.

A weighted Categorical Cross Entropy function (with respect to classes) was used to tackle the class imbalance problem. This loss function and the weights are defined such that the weight increases whenever there are fewer voxels in a particular class.

$$Ls^i = -\sum_v \sum_l w_l * t^i_{v,l} \log p^i_{v,l} \qquad \text{where, } w_l = \frac{\sum_{k=0}^{k=N} m_k}{m_l}$$

Ls^i is the loss for volume i; $t^i_{v,l}$ is 1 if the true label of voxel v of volume i is l else it is 0; $p^i_{v,l}$ is the probability that the *CNN* will predict label l for voxel v of volume i; w_l denotes the weight for the l^{th} class and m_l is the number of voxels of l^{th} class in the training dataset.

The advantage of the *M-net* is that barring one 3D convolution filter, all other filters are 2D filters which allows end-to-end training of the network with considerably low memory requirement ($\sim$5GB).

5.3 Experiments and Results

The proposed architecture is demonstrated on the task of segmenting deep brain structures like Thalamus, Putamen, Pallidum, Hippocampus, Amygdala, Caudate and Accumbens area. Data of varying size is drawn from two publicly available datasets.

First dataset considered is the International Brain Segmentation Repository (IBSR) dataset[1] [90] which has 18, 3D T1 MR images of 1.5 mm thick cortical slices. The size of the MRI volume is 256x256x128. Manual Segmentation of 32 structures is available, however we restrict our attention to only the deep brain structures for the left and right hemispheres.

Second dataset considered is a Diencephalon dataset released as a part of the MICCAI 2013 Challenge workshop on Segmentation: Algorithms, Theory and Applications (SATA) [91]. The data consists of 35 training and 12 testing T1 MR images from the OASIS [79] [2] project with corresponding 14 sub-cortical label maps as provided by Neuromorphometrics Inc.[3] Input volume size is 256x256x300. This dataset is more challenging than the IBSR dataset, as the test volumes are of patients with different abnormalities like white matter degeneration, whereas the training volumes are all of normal patients. Manual segmentation is only available for the training set and evaluation requires submitting the segmentation results to the challenge organisers. [4].

5.3.1 Implementation Details

The proposed CNN was trained on a NVIDIA K40 GPU, with 12GB of RAM for 30 epochs. Approximate training time was 3 days. The CNN was trained using Adam Optimizer [92] with following hyper parameters: learning rate =0.001, beta1=0.9, beta2=0.999, epsilon=10^{-08} and decay=0.0. Learning rate was reduced by a factor of 10 after 20 epochs. Code was written in Keras Library using Python. A hyperparameter of *M-net* is n, which denotes the number of neighbor slices given as additional input to CNN. This was empirically finalized to be 25. Thus, segmentation of any given slice s is done by taking that as a central slice and (25+25=) 50 of its neighbouring slices as input.

The IBSR dataset was randomly divided into two equal sets with 9 volumes each. CNN was trained 2 times on the two sets separately, considering the other set as a testing set. For diencephalon dataset of SATA challenge, CNN was trained on all the 35 training volumes. Performance on testing set was evaluated by uploading the segmented volume on the challenge server.

[1] http://www.nitrc.org/projects/ibsr
[2] http://www.oasis-brains.org/
[3] http://neuromorphometrics.com/
[4] http://tinyurl.com/SATAchallenge

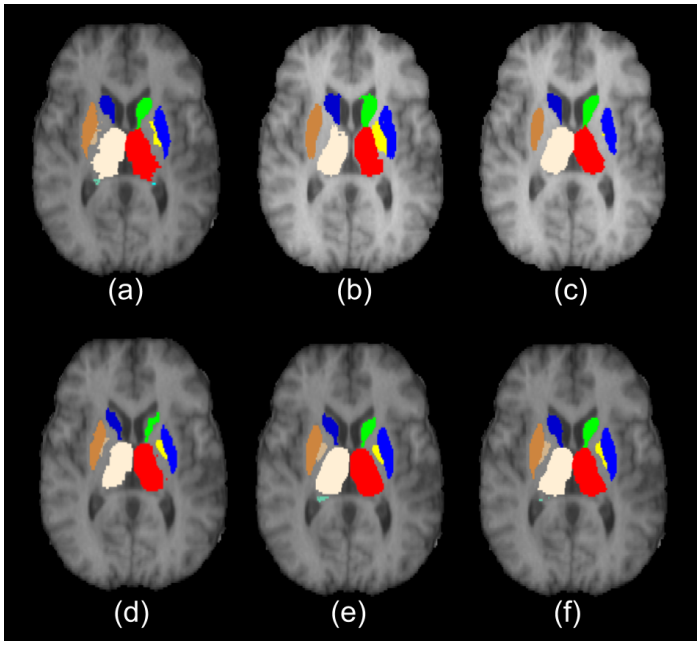

Figure 5.3: Qualitative comparison of 2D slice segmentation for IBSR dataset. (a) GroundTruth (b) RF+MRF (c) FCN+MRF (d) Freesurfer (e) FSL-FIRST (f) *M-net*

5.3.2 Results and Comparison with other methods

The segmentation performance is quantitatively evaluated using the mean Dice Coefficient (DC) across a dataset. Let A and B denote the binary segmentation labels generated manually and computationally, respectively. The DC is defined as:

$$DC(A,B) = \frac{2|AB|}{|A| + |B|}$$

where $|A|$ denotes the number of positive elements in the binary segmentation A, and $|AB|$ is the number of shared positive elements by A and B. DC$\in[0,1]$. A higher DC value indicates better segmentation performance.

On IBSR dataset, we compare the output of *M-net* with 6 different model based methods: FSL-FIRST [84] [5], MS-CNN+RW [8], FCN+MRF [7], U-net [71] + 3D-to-2D converter, RF+MRF [85] and Freesurfer [93] [6]. The approximate CPU run time, for segmenting a new volume, for all of these methods are: FSL-FIRST (∼15 min), Freesurfer (∼90 min), MS-CNN+RW (∼20 min), FCN+MRF (∼15 min),U-net + 3D-to-2D converter (∼5 min) and *M-net* (∼5 min).

[5] `http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FIRST/UserGuide`
[6] `http://freesurfer.net/fswiki/SubcorticalSegmentation`

Table 5.1: Quatitative comparison of performance on the IBSR dataset. Reported Dice coefficient values for a structure are averaged over the values for the 2 hemispheres.

	Freesurfer	FSL-FIRST	RF + MRF	FCN + MRF	MS-CNN + RW	U-net+3D-to-2D	*M-net*
Amygdala	0.69	0.70	0.62	0.64	0.67	0.70	**0.73**
Caudate	0.82	0.83	0.78	0.78	**0.87**	0.85	**0.87**
Hippocampus	0.77	0.81	0.59	0.71	**0.82**	0.81	**0.82**
Pallidum	0.71	0.76	0.62	0.75	0.80	0.80	**0.82**
Putamen	0.81	0.84	0.77	0.83	0.88	0.89	**0.90**
Thalamus	0.86	0.88	0.80	0.87	**0.90**	0.89	**0.90**
Accumbens Area	0.69	0.73	0.60	0.63	0.69	0.71	**0.75**
Overall	0.76	0.79	0.69	0.75	0.80	0.81	**0.83**

The mean DC values for each structure as well as for the whole volume is listed in Table:5.1. Segmentation result for a sample slice is shown in Fig:5.3. Based on these results, we can observe that *M-net* is able to outperform all the other segmentation methods on all the deep brain structures. It should be noted that there is major boost in DC (0.06) for small structures like amygdala and Accumbens area. Results in Fig:5.3 indicate tendency to undersegment by RF+MRF, FCN+MRF and Freesurfer methods, which is not the case with *M-net*. It is notable that unlike *M-net*, the other methods do employ post-processing.

Next, we experimented with Diencephalon dataset. As per the evaluation results displayed in the website hosted by SATA challenge organizers, the mean dice coefficient for *M-net* (IIIT_M_net_try_6) was *0.85780*. This is considerably better than the values for other model based methods: FSL_FIRST - 0.82437, Atlas Forrest based method (MSRC_AtlasForest_Stage2) - 0.82819), Freesurfer- 0.75761. [7].

A bar plot of DC values for all the deep brain structures across the dataset is also made available after evaluation by the SATA challenge organisers. This is shown in Fig:5.4. We can see that, as is the case with all the other methods, the performance of *M-net* is marginally lower for smaller structures compared to bigger structures. Qualitative comparison of segmentation using 3D rendering for *M-net*, FSL-FIRST and Freesurfer is shown in Fig:5.5. The results for FSL-FIRST and *M-net* are smooth for all the structures, compared to Freesurfer.

5.4 Conclusion

We proposed a novel CNN architecture which utilizes 3D information with the help of comparatively less expensive 2D convolution filter, this is achieved by using a single 3D convolution filter to a slice and its neighbouring slices into one slice as a first step. We also introduced skip

[7] `http://tinyurl.com/SATAleaderboard`

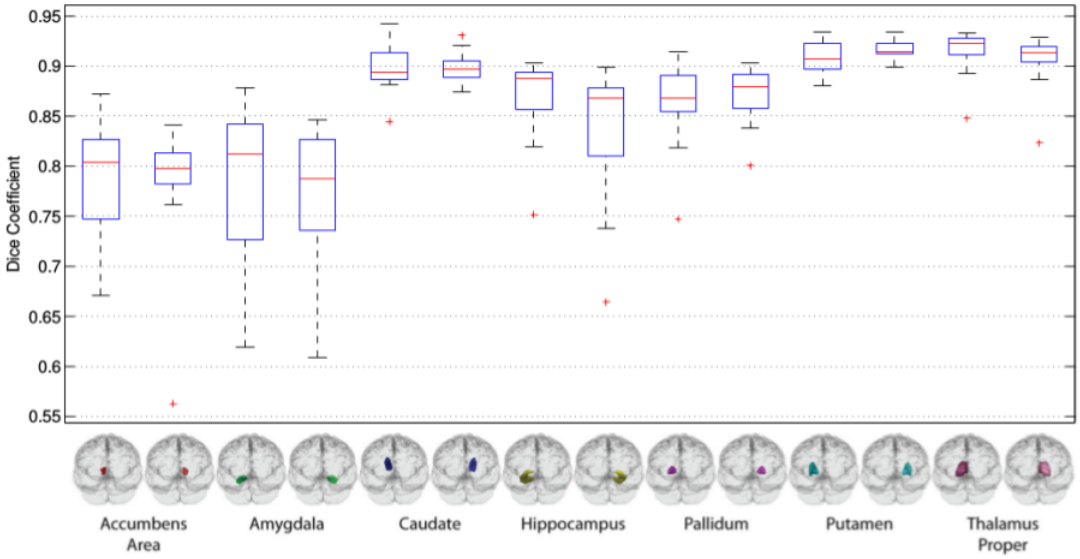

Figure 5.4: DC values for different structures obtained with *M-net* on the Diencephalon dataset

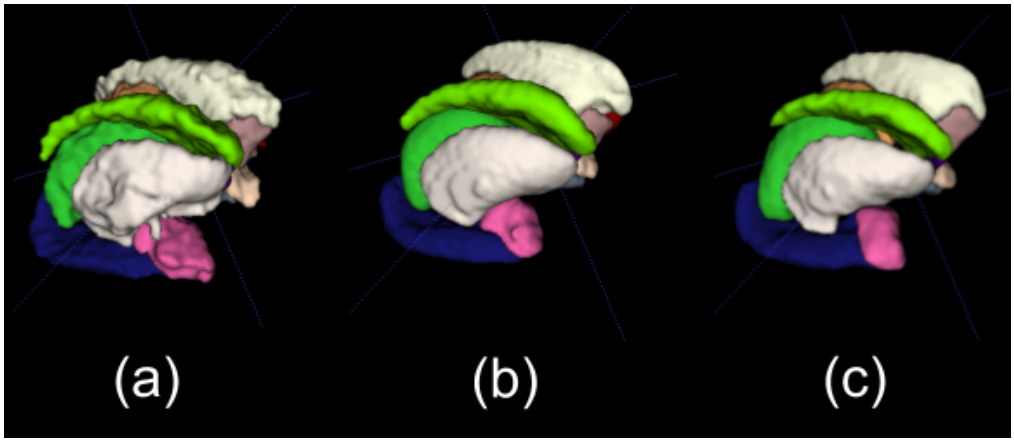

Figure 5.5: Qualitative comparison of segmentation results, using 3D rendering, for the Diencephalon dataset. From left to right: Freesurfer,FSL-FIRST and *M-net*

connections between convolution filters and deep supervision functionality in our network which allows it to learn better features. Experimental results on two publicly available datasets, with different volume dimensions, shows that proposed network outperforms the current state of the art model based segmentation techniques and at considerably ($\sim\frac{1}{3^{rd}}$) less processing time. These aspects paves way for its use in clinical scenario. The *M-net* segments a volume slice by slice and hence it can potentially be used to segment any 3D dataset which is to be explored in the future.

For example, we used M-net to segment the whole brain into 134 structures using MICCAI-2012 dataset (for dataset detail refer chapter-3). M-net achieved the mean DC of 0.7278. This was better then 2D U-net and U-net + 3D-to-2D converter which had reached mean DC of 0.6624 and 0.6971, respectively. This experiment also verifies the contributions of both 3D-to-2D converter and introduction of skip connections. There is an increase of 0.0347 in mean DC due to 3D-to-2D converter (2D U-net *vs* U-net + 3D-to-2D converter) and an increase of 0.0307 with the introduction of skip connections (U-net + 3D-to-2D converter *vs* M-net). While this result shows the superiority of M-net over U-net, it also shows that it is not able to outperform Multi-Atlas segmentation (MAS) method [60] which achieves mean DC of 0.764. It should be noted that while on a standard CPU M-net takes only 5 mins for whole brain segmentation, MAS method takes upto 20 hours for it.

Chapter 6

Summary and Future Work

In this thesis we have focused on the construction of population specific human brain atlas for young Indian population and the use of convolutional neural network based architectures for segmenting cortical and sub-cortical structures in human brain MRI using multiple structural atlases.

We began the thesis by looking into the problem of Intensity standardization (IS), which arises due to the intensity variations in MR images. In chapter-2, we proposed a method for IS in which tissue information (via segmentation) is needed during training but not during testing by using landmark propagation. Evaluation and comparison of the proposed method on 24 MRI volumes from 3 different scanners was done against two state of the art IS methods. We observed that the use of KL divergence based information theoretical approach, to find volumes which have similar tissue intensity distribution, can perform better IS as it reduces Intra-scanner and Inter-scanner Jeffery divergence between histogram of MRI volumes, without any overhead of tissue segmentation during testing time.

Next in chapter-3, we looked into the problem of population specific brain MRI template construction for young Indian Population using the collected T1 weighted brain MRI of 100 young volunteers. We validated and compared the constructed template (IBA100), using another 15 subjects, against the popular MNI template as well as recent template of the Chinese population. We found that the Indian brain, on average, is smaller in size compare to the Caucasian and the Chinese brain templates. In addition to this, we also found that the female brain, on average, is smaller in size compare to the male brain, which necessitates the need of construction of two separate template for both gender. We validated the constructed template by comparing the global deformation field required to register Indian population brain MRI to IBA100 and MNI152/Chinese2020 template. We concluded that smaller global deformation is required for IBA100 template compared to MNI152 and Chinese2020 templates. This paves the way for its

use in many neuroscience research methodologies like fMRI and VBM analysis which relies on spatial normalization.

With the recent advances in deep learning approaches for many computer vision tasks such as classification, segmentation etc. [66], it was of interest of to check its effectiveness for MRI structural segmentation which is hitherto approached using computationally expensive Multi-Atlas segmentation (MAS) techniques based on non-rigid registration. In chapter-4 and chapter-5, we proposed two variation Convolutional Neural Network *(CNN)* for structural segmentation in brain MRIs.

In chapter-4, we proposed an alternative approach (BrainSegNet) for automatic segmentation of brain MRI using a CNN which classifies a voxel into one of many structures. Four different kinds of 2D and 3D intensity patches were extracted for each voxel, providing local and global (context) information to the *CNN*. With the help of experiment on MICCAI-2012 dataset, it was observed that while local 2D patches are helpful for segmenting sub-cortical structures, global 2D patches help in achieving better segmentation for cortical structures. Use of 3D patches also boost the segmentation performance, as it gives inter-slice consistency for MRI without the need of post processing steps like MRF, random walker etc. Comparison of the proposed approach on 5 different publicly available datasets against many different segmentation methods showed that its performace is marginally better or on par than the current state of the art methods on nearly all datasets, in terms of mean Dice coefficient at a reduced computational time.

In chapter-5, we proposed an end-to-end trainable Fully Convolutional Neural Network (FCNN) architecture called the *M-net*, for segmenting deep (human) brain structures from MRI. The M-net was a modified version of previously proposed U-net. Two main modification were (i) introduction of 3D-to-2D converter, which gives the inter-slice consistency without using 3D convolutional filter beyond first stage and (ii) use of skip connection and deep supervision which imporves the segmentation performce. The M-net was evaluated on 2 publicly available datasets and was compared against publicly available model based segmentation algorithms (FSL, Freesurfer) as well as other classification based algorithms such as Random Forrest and 2D CNN based approaches. From the qualitative and quantitative comparison we noticed that the M-net outperformed the state-of-the-art methods with at least 3 times speed up in segmentation time. We also showed that the M-net can be used for any 3D dataset by using it for whole brain structure segmentation on MICCAI-2012 dataset.

Results in chapter-4 and chapter-5 indicate the potential use of Deep learning based models for structural segmentation in brain MRI.

6.1 Future Work

In the first half of the thesis we focused on construction and validation of Indian brain template for young population. We found that brain of Indian population is smaller compare to the Caucasian and the Chinese brain in terms of height and width. Further validation of the constructed template is needed. One possible method for validating the template can be via atlas based segmentation of different anatomical structures using Indian template and comparing the results against the popular structural atlases like AAL [94] and DKS which is constructed based on Caucasian brain data. Quantitative comparison of the local deformation field (using non-rigid registration) required to warp Indian brain to the IBA100 template and MNI-152/Chinese2020 template can also be useful for validation purpose. It should we noted that the constructed Indian brain template is static as it was constructed using MRI of young Indian population (21-30). This did not capture the brain template as a function of age. Studies like [95] [54] show that structure of brain is highly variable and it changes with age even for healthy humans. This emphasizes the need for a spatio-temporal brain atlas for the Indian population, which can be used as a standard template for all the age groups of the society.

In the second half of this thesis we proposed two CNN architecture for structure brain segmentation. BrainSegNet proposed in chapter-4 was able to marginally outperform the state-of-the-art multi atlas segmentation (MAS) method in terms of dice coefficient. Similarly, M-net proposed in chapter-5 was able to segment whole brain in significantly lower computational time than MAS methods (M-net: 5 mins, MAS: 20 hours) but with a reduction in accuracy. In the future, a better architecture which can achieve mean DC comparable or better to BrainSegNet at a computational time of M-net should be designed. While the proposed CNN based methods (BrainSegNet and M-net) gave satisfactory performance for structural segmentation in healthy brain MRI, it would of interest to analyze the same for diseased (Alzheimer's and Huntington's) [96] and pathological brain MRIs (MS lesions and Tumor) [97] .

Appendix

The project of "Digital human brain atlas construction for Indian population" was started in June 2015. This project required the collection of 100 brain MRI scans of Indian population.

First step in the data collection process required an approval by the ethics committee of the college, as the participants/volunteers had to undergo MRI scan in hospitals and the MRI scanners contain 1.5T magnet which may be of concern for many participants.

Most of the population specific templates, i.e. Chinese2020 [21], Korean96 [17], are based on the MRI scans acquired using scanners from one manufacturer with same scanning protocols. This allowed them to have MRI scans which have the same intensity range, this made the process of template construction easier as they didn't have to handle inter-scanner intensity variation. It is of interest to neuroscience community to construct the template using MRI scans which are acquired on scanner by different manufacturers using different scanning protocols, because in a practical scenario where constructed template will be used, we cannot guarantee that the acquired MRI will have same scanning protocols. Keeping this in mind, for this project collaboration with three different hospitals, i.e. Continental, CARE Nampally and CARE Hitech, was done so that the collected data can have diversity in terms of scanner manufactures, i.e. GE, Siemens and Phillips, and scanning protocols.

The aim of this project was to construct the template for Indian population such that the constructed template can serve as a reference brain MRI, and we can identify the brain MRI scans which are affected by different disease like Alzheimer's. Similarly, the constructed template can also be used to detect the effect of aging. In many studies [98, 99, 100, 101], it was shown that human brain matures at the age of 21 and the effect of aging starts as early as 30 year age. Keeping this in mind, we recruited the MRI volunteers aged between 21 to 30. As IIIT-H is an research based institute which has many post-graduation program, like M-tech, MS and PhD, with high intake and students in the age group of 21-30, we recruited all the participants/volunteers from the institute. It was of critical importance to make sure that the volunteers are "Normal and

Healthy" and were not affected by any psychological or neurological disorders. Psychological examination of all the volunteers wae conducted by an on campus senior psychologist using self reporting questionnaire designed by World Health Organization (W.H.O). Senior radiologist at each hospital inspected the acquired MRI scans to discard scans which has neurological disorder.

As the collected data was from three different hospitals which has MRI scanners from three different scanner manufacture, it was necessary to fix scanning protocol such that the images are less noisy and they are as near to each other as possible in terms of intensity range. The scanning protocol was decided by three senior radiologist from all the collaborating hospitals in a series of meetings discussing pros and cons, like total scanning time and image quality, for all the possible scanning protocols. A volunteer, yours truly, was recruited as a human phantom to inspect the quality of acquired images using different scanning protocols at all the hospitals. After almost 1 month and 10 MRI scans of the volunteer, scanning protocols were fixed. The final protocols are tabulated in Table:3.1.

The task of data collection is challenging. First and somewhat most difficult task was to recruit the volunteers the MRI scans. This was done using an online volunteering form. Although, many volunteers eagerly filled up the form for participation, some of them backed out after the initial process due to the false assumptions prevailing in Indian society like MRI scan is harmful for the health as it uses radiation! The aim of this project was to collect MRI scan data which was not biased by gender, i.e. 50 male and 50 female MRI scans. One difficulty, which we faced, was to get 50 female volunteers as there were less female volunteers who were willing to participate in this study. The reason for this varied from smaller number of female students in IIIT-H to non-consent by student parents.

All the three collaborating hospitals in this study are commercial hospitals and gave priority the scanning patients over research volunteers. Hnece, all the MRI scans for this study were done either during early morning (6AM-10AM) or at late night (9PM-12PM). This caused difficulty for many volunteers, especially the ones who had to go to CARE Nampally for scan as the hospital is 20 Km away from IIIT-H. In the future, it would be better if data collection is done with research hospitals so that this kind of issue is avoided.

Validation of the constructed atlas was done measuring global feature of the brain like height, width and length. In addition to this, required global linear deformation for registering Indian brain to the constructed IBA100 and MNI152/Chinese2020 was compared for a validation purpose. While this kind of analysis legitimizes the template construction, a better way of validating is through comparing segmentation performance for various brain structures using

multi atlas segmentation by IBA100 and MNI152 template. This kind of validation requires manual groundtruth marking for the structure of interest by an expert neuroradiologist. Manual marking is a strenuous task which requires concentration for long hours and precision in terms of marking. It takes 2-3 hours for manual marking of two structures, namely, hippocampus and putamen on a single volume as each volume has atleast 20-30 slices which contain these structures and 3D consistant between slices is also necessary. In this ongoing project, we have collaborated with neuroradiologist from SCMIST, Kerala and CARE hospital Nampally for manual marking. As both of these hospitals are commercial hospitals, the time taken by the neuroradiologist for manual marking was long as they have to attend various hospital related chores. In future, recruitment of an expert neuroradiologist, who can primarily work for manual structure marking, as a team member is of necessity.

Data collection of total 115 MRI scans, 100 for template construction and 15 for template validation, took almost eighteen months. The first scan was done on 5th August 2015 while the last scan was done on 27th January 2017. Reasons for this long process varied from non-availability of volunteers to other organizational constraints of participating hospitals.

As mentioned above, for the template validation purpose we are collecting groundtruth marking of hippocampus and putamen by an expert neuroradiologist. At this stage of the project, we have collected manual segmentation for 10 volumes which took approximately 15 days with daily 3-4 hours workload for a neuroradiologist. Same process for other validation volumes is a work in progress.

Related Publications

- Mehta, R. and Sivaswamy, J., 2016, April. A hybrid approach to tissue-based intensity standardization of brain MRI images. In Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on (pp. 95-98). IEEE.
- Mehta, R. and Sivaswamy, J., 2017, April. M-net: A convolutional Neural Network for Deep Brain Segmentation. In Biomedical Imaging (ISBI), 2017 IEEE 14th International Symposium on. IEEE.
- Mehta, R., Majumdar, A. and Sivaswamy, J., 2017. BrainSegNet: a convolutional neural network architecture for automated segmentation of human brain structures. Journal of Medical Imaging (JMI), 4(2), pp.024003-024003.

Bibliography

[1] J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, T. Paus, G. Simpson, B. Pike, *et al.*, "A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (icbm)," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 356, no. 1412, pp. 1293–1322, 2001.

[2] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, *et al.*, "An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest," *Neuroimage*, vol. 31, no. 3, pp. 968–980, 2006.

[3] M. Walterfang, B. Patenaude, L. Abel, H. Kluenemann, E. Bowman, M. Fahey, P. Desmond, W. Kelso, and D. Velakoulis, "Subcortical volumetric reductions in adult niemann-pick disease type c: a cross-sectional study," *American Journal of Neuroradiology*, vol. 34, no. 7, pp. 1334–1340, 2013.

[4] D. Zikic, B. Glocker, and A. Criminisi, "Atlas encoding by randomized forests for efficient label propagation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 66–73, Springer, 2013.

[5] L. Zhang, Q. Wang, Y. Gao, G. Wu, and D. Shen, "Automatic labeling of mr brain images by hierarchical learning of atlas forests," *Medical physics*, vol. 43, no. 3, pp. 1175–1186, 2016.

[6] F. Rousseau, P. A. Habas, and C. Studholme, "A supervised patch-based approach for human brain labeling," *IEEE transactions on medical imaging*, vol. 30, no. 10, pp. 1852–1862, 2011.

[7] M. Shakeri, S. Tsogkas, E. Ferrante, S. Lippe, S. Kadoury, N. Paragios, and I. Kokkinos, "Sub-cortical brain structure segmentation using f-cnn's," *arXiv preprint arXiv:1602.02130*, 2016.

[8] S. Bao and A. C. Chung, "Multi-scale structured cnn with label consistency for brain mr image segmentation," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pp. 1–5, 2016.

[9] M. N. Wernick, Y. Yang, J. G. Brankov, G. Yourganov, and S. C. Strother, "Machine learning in medical imaging," *IEEE signal processing magazine*, vol. 27, no. 4, pp. 25–38, 2010.

[10] A. C. Evans, A. L. Janke, D. L. Collins, and S. Baillet, "Brain templates and atlases," *Neuroimage*, vol. 62, no. 2, pp. 911–922, 2012.

[11] N. Isamah, W. Faison, M. E. Payne, J. MacFall, D. C. Steffens, J. L. Beyer, K. R. Krishnan, and W. D. Taylor, "Variability in frontotemporal brain structure: the importance of recruitment of african americans in neuroscience research," *PLoS One*, vol. 5, no. 10, p. e13642, 2010.

[12] Y. Tang, C. Hojatkashani, I. D. Dinov, B. Sun, L. Fan, X. Lin, H. Qi, X. Hua, S. Liu, and A. W. Toga, "The construction of a chinese mri brain atlas: A morphometric comparison study between chinese and caucasian cohorts," *Neuroimage*, vol. 51, no. 1, pp. 33–41, 2010.

[13] H. T. Uchiyama, A. Seki, D. Tanaka, T. Koeda, *et al.*, "A study of the standard brain in japanese children: Morphological comparison with the mni template," *Brain and Development*, vol. 35, no. 3, pp. 228–235, 2013.

[14] A. C. Evans, D. L. Collins, S. Mills, E. Brown, R. Kelly, and T. M. Peters, "3d statistical neuroanatomical models from 305 mri volumes," in *Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record.*, pp. 1813–1817, IEEE, 1993.

[15] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers, "Automatic anatomical brain mri segmentation combining label propagation and decision fusion," *NeuroImage*, vol. 33, no. 1, pp. 115–126, 2006.

[16] M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, *et al.*, "A multi-modal parcellation of human cerebral cortex," *Nature*, 2016.

[17] H. Lee, B. I. Yoo, J. W. Han, J. J. Lee, S. Y. W. Oh, E. Y. Lee, J. H. Kim, and K. W. Kim, "Construction and validation of brain mri templates from a korean normal elderly population," *Psychiatry investigation*, vol. 13, no. 1, pp. 135–145, 2016.

[18] L. G. Apostolova, L. Mosconi, P. M. Thompson, A. E. Green, K. S. Hwang, A. Ramirez, R. Mistur, W. H. Tsui, and M. J. de Leon, "Subregional hippocampal atrophy predicts

alzheimer's dementia in the cognitively normal," *Neurobiology of aging*, vol. 31, no. 7, pp. 1077–1088, 2010.

[19] B. M. Dawant, A. P. Zijdenbos, and R. A. Margolin, "Correction of intensity variations in mr images for computer-aided tissue classification," *IEEE transactions on medical imaging*, vol. 12, no. 4, pp. 770–781, 1993.

[20] L. G. Nyu and J. K. Udupa, "On standardizing the mr image intensity scale," *image*, vol. 1081, 1999.

[21] P. Liang, L. Shi, N. Chen, Y. Luo, X. Wang, K. Liu, V. C. Mok, W. C. Chu, D. Wang, and K. Li, "Construction of brain atlases based on a multi-center mri dataset of 2020 chinese adults," *Scientific reports*, vol. 5, 2015.

[22] U. Bağcı, J. K. Udupa, and L. Bai, "The role of intensity standardization in medical image registration," *Pattern Recognition Letters*, vol. 31, no. 4, pp. 315–323, 2010.

[23] S. Roy, A. Carass, and J. L. Prince, "Patch based intensity normalization of brain mr images," in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, pp. 342–345, IEEE, 2013.

[24] J. D. Christensen, "Normalization of brain magnetic resonance images using histogram even-order derivative analysis," vol. 21, pp. 817–820, Elsevier, 2003.

[25] I. J. Cox, S. Roy, and S. L. Hingorani, "Dynamic histogram warping of image pairs for constant image brightness," in *Image Processing, 1995. Proceedings., International Conference on*, vol. 2, pp. 366–369, IEEE, 1995.

[26] P. Hellier, "Consistent intensity correction of mr images," in *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, vol. 1, pp. I–1109, IEEE, 2003.

[27] L. G. Nyúl, J. K. Udupa, and X. Zhang, "New variants of a method of mri scale standardization," *Medical Imaging, IEEE Transactions on*, vol. 19, no. 2, pp. 143–150, 2000.

[28] M. Shah, Y. Xiao, N. Subbanna, S. Francis, D. L. Arnold, D. L. Collins, and T. Arbel, "Evaluating intensity normalization on mris of human brain with multiple sclerosis," *Medical image analysis*, vol. 15, no. 2, pp. 267–282, 2011.

[29] F. Jäger, Y. Deuerling-Zheng, B. Frericks, F. Wacker, and J. Hornegger, "A new method for mri intensity standardization with application to lesion detection in the brain," *Procs*, vol. 1010, pp. 269–276, 2006.

[30] N. Robitaille, A. Mouiha, B. Crépeault, F. Valdivia, and S. Duchesne, "Tissue-based mri intensity standardization: application to multicentric datasets," *Journal of Biomedical Imaging*, vol. 2012, p. 4, 2012.

[31] G. De Nunzio, R. Cataldo, and A. Carlà, "Robust intensity standardization in brain magnetic resonance images," *Journal of digital imaging*, pp. 1–11, 2015.

[32] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.

[33] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, pp. 79–86, 1951.

[34] L. Imperial College, "Brain development dataset,." `http://biomedic.doc.ic.ac.uk/brain-development/index.php?n=Main.Datasets/`.

[35] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[36] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in mri data," *Medical Imaging, IEEE Transactions on*, vol. 17, no. 1, pp. 87–97, 1998.

[37] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm," *IEEE transactions on medical imaging*, vol. 20, no. 1, pp. 45–57, 2001.

[38] S. M. Smith, "Fast robust automated brain extraction," *Human brain mapping*, vol. 17, no. 3, pp. 143–155, 2002.

[39] J. Ashburner and K. J. Friston, "Voxel-based morphometry: the methods," *Neuroimage*, vol. 11, no. 6, pp. 805–821, 2000.

[40] J. Talairach and P. Tournoux, "Co-planar stereotaxic atlas of the human brain. 3-dimensional proportional system: an approach to cerebral imaging," 1988.

[41] J. C. Mazziotta, A. W. Toga, A. Evans, P. Fox, and J. Lancaster, "A probabilistic atlas of the human brain: Theory and rationale for its development: The international consortium for brain mapping (icbm)," *Neuroimage*, vol. 2, no. 2, pp. 89–101, 1995.

[42] D. L. Collins, C. J. Holmes, T. M. Peters, and A. C. Evans, "Automatic 3-d model-based neuroanatomical segmentation," *Human brain mapping*, vol. 3, no. 3, pp. 190–208, 1995.

[43] F. Lalys, C. Haegelen, J.-C. Ferre, O. El-Ganaoui, and P. Jannin, "Construction and assessment of a 3-t mri brain template," *Neuroimage*, vol. 49, no. 1, pp. 345–354, 2010.

[44] T. Jao, C.-Y. Chang, C.-W. Li, D.-Y. Chen, E. Wu, C.-W. Wu, C.-H. Tsou, C.-C. Ho, and J.-H. Chen, "Development of ntu standard chinese brain template: morphologic and functional comparison with mni template using magnetic resonance imaging," in *Engineering*

in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE, pp. 4779–4782, IEEE, 2009.

[45] J. S. Lee, D. S. Lee, J. Kim, Y. K. Kim, E. Kang, H. Kang, K. W. Kang, J. M. Lee, J.-J. Kim, H.-J. Park, *et al.*, "Development of korean standard brain templates," *Journal of Korean medical science*, vol. 20, no. 3, pp. 483–488, 2005.

[46] J. Bai, M. F. Abdul-Rahman, A. Rifkin-Graboi, Y.-S. Chong, K. Kwek, S.-M. Saw, K. M. Godfrey, P. D. Gluckman, M. V. Fortier, M. J. Meaney, *et al.*, "Population differences in brain morphology and microstructure among chinese, malay, and indian neonates," *PLoS One*, vol. 7, no. 10, p. e47816, 2012.

[47] P. K. Mandal, R. Mahajan, and I. D. Dinov, "Structural brain atlases: design, rationale, and applications in normal and pathological cohorts," *Journal of Alzheimer's Disease*, vol. 31, no. s3, pp. S169–S188, 2012.

[48] D. A. Dickie, S. D. Shenkin, D. Anblagan, J. Lee, M. B. Cabez, D. Rodriguez, J. P. Boardman, A. Waldman, D. E. Job, and J. M. Wardlaw, "Whole brain magnetic resonance image atlases: a systematic review of existing atlases and caveats for use in population imaging," *Frontiers in neuroinformatics*, vol. 11, 2017.

[49] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, "N4itk: improved n3 bias correction," *IEEE transactions on medical imaging*, vol. 29, no. 6, pp. 1310–1320, 2010.

[50] J. V. Manjón, P. Coupé, L. Martí-Bonmatí, D. L. Collins, and M. Robles, "Adaptive non-local means denoising of mr images with spatially varying noise levels," *Journal of Magnetic Resonance Imaging*, vol. 31, no. 1, pp. 192–203, 2010.

[51] B. Avants, T. Sundaram, J. Duda, J. Gee, and L. Ng, "Insight into images. ch. nonrigid image registration," pp. 307–348, 2004.

[52] B. A. Ardekani and A. H. Bachman, "Model-based automatic detection of the anterior and posterior commissures on mri scans," *Neuroimage*, vol. 46, no. 3, pp. 677–682, 2009.

[53] B. B. Avants, P. Yushkevich, J. Pluta, D. Minkoff, M. Korczykowski, J. Detre, and J. C. Gee, "The optimal template effect in hippocampus studies of diseased populations," *Neuroimage*, vol. 49, no. 3, pp. 2457–2466, 2010.

[54] W. Xing, C. Nan, Z. ZhenTao, X. Rong, J. Luo, Y. Zhuo, S. DingGang, and L. KunCheng, "Probabilistic mri brain anatomical atlases based on 1,000 chinese subjects," *PloS one*, vol. 8, no. 1, p. e50939, 2013.

[55] T. R. Langerak, F. F. Berendsen, U. A. Van der Heide, A. N. Kotte, and J. P. Pluim, "Multiatlas-based segmentation with preregistration atlas selection," *Medical physics*, vol. 40, no. 9, p. 091701, 2013.

[56] A. K. H. Duc, M. Modat, K. K. Leung, M. J. Cardoso, J. Barnes, T. Kadir, S. Ourselin, A. D. N. Initiative, *et al.*, "Using manifold learning for atlas selection in multi-atlas segmentation," *PloS one*, vol. 8, no. 8, p. e70059, 2013.

[57] H. Wang, B. Avants, and P. Yushkevich, "A combined joint label fusion and corrective learning approach," in *MICCAI Workshop on Multi-Atlas Labeling*, 2012.

[58] A. Asman and B. Landman, "Multi-atlas segmentation using non-local staple," in *MICCAI Workshop on Multi-Atlas Labeling*, 2012.

[59] J. E. Iglesias and M. R. Sabuncu, "Multi-atlas segmentation of biomedical images: a survey," *Medical image analysis*, vol. 24, no. 1, pp. 205–219, 2015.

[60] B. Landman and S. Warfield, "Miccai 2012 workshop on multi-atlas labeling," in *Medical Image Computing and Computer Assisted Intervention Conference 2012: MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling Challenge Results*, 2012.

[61] P. Coupé, J. V. Manjón, V. Fonov, J. Pruessner, M. Robles, and D. L. Collins, "Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation," *NeuroImage*, vol. 54, no. 2, pp. 940–954, 2011.

[62] D. Zhang, Q. Guo, G. Wu, and D. Shen, "Sparse patch-based label fusion for multi-atlas segmentation," in *International Workshop on Multimodal Brain Image Analysis*, pp. 94–102, Springer, 2012.

[63] G. Wu, Q. Wang, D. Zhang, and D. Shen, "Robust patch-based multi-atlas labeling by joint sparsity regularization," in *MICCAI Workshop STMI*, 2012.

[64] A. de Brebisson and G. Montana, "Deep neural networks for anatomical brain segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 20–28, 2015.

[65] F. Milletari, S.-A. Ahmadi, C. Kroll, A. Plate, V. Rozanski, J. Maiostre, J. Levin, O. Dietrich, B. Ertl-Wagner, K. Bötzel, *et al.*, "Hough-cnn: Deep learning for segmentation of deep brain regions in mri and ultrasound," *arXiv preprint arXiv:1601.07014*, 2016.

[66] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[67] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, 2015.

[68] H. R. Roth, L. Lu, A. Seff, K. M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, and R. M. Summers, "A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 520–527, Springer, 2014.

[69] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 246–253, Springer, 2013.

[70] H. R. Roth, A. Farag, L. Lu, E. B. Turkbey, and R. M. Summers, "Deep convolutional networks for pancreas segmentation in ct imaging," in *SPIE Medical Imaging*, pp. 94131G–94131G, International Society for Optics and Photonics, 2015.

[71] 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*, pp. 234–241, Springer, 2015.

[72] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, 2015.

[73] M. F. Stollenga, W. Byeon, M. Liwicki, and J. Schmidhuber, "Parallel multi-dimensional lstm, with application to fast biomedical volumetric image segmentation," in *Advances in Neural Information Processing Systems*, pp. 2998–3006, 2015.

[74] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. Benders, and I. Išgum, "Automatic segmentation of mr brain images with a convolutional neural network," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1252–1261, 2016.

[75] G. Wu, M. Kim, G. Sanroma, Q. Wang, B. C. Munsell, D. Shen, A. D. N. Initiative, *et al.*, "Hierarchical multi-atlas label fusion with multi-scale feature representation and label-specific patch partition," *NeuroImage*, vol. 106, pp. 34–46, 2015.

[76] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting.," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[77] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[78] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," *arXiv preprint arXiv:1603.05959*, 2016.

[79] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults," *Journal of cognitive neuroscience*, vol. 19, no. 9, pp. 1498–1507, 2007.

[80] D. W. Shattuck, M. Mirza, V. Adisetiyo, C. Hojatkashani, G. Salamon, K. L. Narr, R. A. Poldrack, R. M. Bilder, and A. W. Toga, "Construction of a 3d probabilistic atlas of human cortical structures," *Neuroimage*, vol. 39, no. 3, pp. 1064–1080, 2008.

[81] G. Wu, M. Kim, Q. Wang, Y. Gao, S. Liao, and D. Shen, "Unsupervised deep feature learning for deformable registration of mr brain images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 649–656, Springer, 2013.

[82] Y. Chudasama and T. Robbins, "Functions of frontostriatal systems in cognition: comparative neuropsychopharmacological studies in rats, monkeys and humans," *Biological psychology*, vol. 73, no. 1, pp. 19–38, 2006.

[83] R. Kumar, S. Farahvar, J. A. Ogren, P. M. Macey, P. M. Thompson, M. A. Woo, F. L. Yan-Go, and R. M. Harper, "Brain putamen volume changes in newly-diagnosed patients with obstructive sleep apnea," *NeuroImage: Clinical*, vol. 4, pp. 383–391, 2014.

[84] B. Patenaude, S. M. Smith, D. N. Kennedy, and M. Jenkinson, "A bayesian model of shape and appearance for subcortical brain segmentation," *Neuroimage*, vol. 56, no. 3, pp. 907–922, 2011.

[85] S. Alchatzidis, A. Sotiras, and N. Paragios, "Discrete multi atlas segmentation using agreement constraints," in *British Machine Vision Conference*, 2014.

[86] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[87] O. Cicek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 424–432, Springer, 2016.

[88] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets.," in *AISTATS*, vol. 2, p. 6, 2015.

[89] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Highway networks," *arXiv preprint arXiv:1505.00387*, 2015.

[90] T. Rohlfing, "Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable," *IEEE transactions on medical imaging*, vol. 31, no. 2, pp. 153–163, 2012.

[91] A. Asman, A. Akhondi-Asl, H. Wang, N. Tustison, B. Avants, S. K. Warfield, and B. Landman, "Miccai 2013 segmentation algorithms, theory and applications (sata) challenge results summary," in *MICCAI Challenge Workshop on Segmentation: Algorithms, Theory and Applications (SATA)*, 2013.

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

[93] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. Van Der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, *et al.*, "Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 33, no. 3, pp. 341–355, 2002.

[94] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain," *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002.

[95] R. Takahashi, K. Ishii, T. Kakigi, and K. Yokoyama, "Gender and age differences in normal adult human brain: Voxel-based morphometric study," *Human brain mapping*, vol. 32, no. 7, pp. 1050–1058, 2011.

[96] O. T. Carmichael, H. A. Aizenstein, S. W. Davis, J. T. Becker, P. M. Thompson, C. C. Meltzer, and Y. Liu, "Atlas-based hippocampus segmentation in alzheimer's disease and mild cognitive impairment," *Neuroimage*, vol. 27, no. 4, pp. 979–990, 2005.

[97] B. Wang, M. Prastawa, A. Irimia, A. Saha, W. Liu, S. M. Goh, P. M. Vespa, J. D. Van Horn, and G. Gerig, "Modeling 4d pathological changes by leveraging normative models," *Computer Vision and Image Understanding*, vol. 151, pp. 3–13, 2016.

[98] L. Svennerholm, K. Boström, and B. Jungbjer, "Changes in weight and compositions of major membrane components of human brain during the span of adult human life of swedes," *Acta neuropathologica*, vol. 94, no. 4, pp. 345–352, 1997.

[99] R. Peters, "Ageing and the brain," *Postgraduate medical journal*, vol. 82, no. 964, pp. 84–88, 2006.

[100] J. N. Giedd, J. Blumenthal, N. O. Jeffries, F. X. Castellanos, H. Liu, A. Zijdenbos, T. Paus, A. C. Evans, and J. L. Rapoport, "Brain development during childhood and adolescence: a longitudinal mri study," *Nature neuroscience*, vol. 2, no. 10, pp. 861–863, 1999.

[101] A. W. Toga, P. M. Thompson, and E. R. Sowell, "Mapping brain maturation," *Focus*, 2006.