FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Brain Pre-surgical white matter Mapping

João Salgueiro Rocha

DISSERTATION

MESTRADO EM ENGENHARIA BIOMÉDICA

Supervisor: João Manuel R. S. Tavares

18th September

Brain Pre-surgical white matter Mapping

João Salgueiro Rocha

MESTRADO EM ENGENHARIA BIOMÉDICA

18th September

Resumo

A visualização dos caminhos da matéria branca do cérebro através de ressonância magnética de difusão está cada vez mais no centro das atenções para o planeamento neurocirúrgico, pois estes métodos permitem a visalização pré-operatória, tridimensional, não invasiva e in vivo da localização e trajetória dos caminhos que a matéria branca percorre no cérebro. O rastreamento de fibras axonais de matéria branca no cérebro fornece uma maneira de analisar espacialmente a organização de estruturas cerebrais e massas benignas ou malignas. Esta capacidade é usada por médicos clínicos para orientar o planeamento pré-cirúrgico em vários cenários, não apenas quando pistas importantes estão no caminho de uma intervenção cirurgica seja cavidade criada por um tumor ou à volta de uma lesão superficial, mas também indiretamente quando o alvo é uma lesão localizada profundamente ou uma estrutura funcional. Apesar disso, os métodos atuais de tractografia sofrem problemas de legibilidade e validade, dependendo fortemente de segmentações manuais ou usando regiões de interesse na imagem para semear algoritmos probabilísticos que seguem os caminhos. Como tal, sistemas baseados em técnicas de deep learning estão a ser cada vez mais propostos para reduzir o tempo gasto na geração desses mapas. Esta dissertação começa com uma visão geral do sistema nervoso central, resumindo estruturas anatómicas importantes e a composição dos tecidos culminando com a descrição de 2 grandes caminhos de matéria branca com maior importância clínica nas cirurgias para remoção de tumores: Radiações Ópticas e Trato Corticoespinhal. Da mesma forma é importante discutir os aspectos físicos relacionados com a RM e a construção de imagens estruturais e de difusão. Em seguida, é apresentado um resumo simples dos conceitos relacionados com Deep learning com ênfase nas CNNs e seus componentes mais comuns. Uma revisão detalhada do estado da arte em segmentação de imagens cerebrais é subsequentemente apresentada para estabelecer uma fundação a partir do qual o trabalho experimental foi construído. Como a arquitetura base mais utilizada para segmentação de imagens médicas é actualmente V-net ou U-net, o trabalho experimental concentra-se em investigar o uso de análise multimodal de imagens médicas através da rede AgYnet. Este modelo incorpora mecanismos de atenção para fornecer informações sobre áreas mais salientes da imagem orientando a previsão dessa maneira. A investigação concentrou-se em estabelecer uma linha de base para o AgYnet, depois experimentar uma arquitetura mais densa derivada do modelo Unet++ e, finalmente, ampliar a quantidade e a qualidade dos dados de um conjunto de dados maior de modo a melhorar as previsões. Neste trabalho foi possivel obter uma segmentação dos caminhos das radições ópticas com um DICE médio de 65,79% (direito) e 64,87% (esquerdo) e para o caminho corticoespinal de 63,72% (direito) e 61,63% (esquerdo). As contribuições deste trabalho incluem a implementação em tensorflow e keras de AgYnet e nestedAgynet; e a constatação de que com mais dados o AgYnet pode produzir resultados muito promissores na previsão correta dos tratos da matéria branca no cérebro. Por outro lado, foi possível concluir que a arquitetura aninhada oferece pouco em termos de benefícios para a qualidade da previsão, ao mesmo tempo em que amplia imensamente o tempo necessário para obtê-los.

Abstract

The visualization of the brain's white matter pathways via diffusion magnetic resonance imaging is receiving increasing attention for neurosurgical planning, as these methods allow preoperative, three-dimensional, non-invasive, in vivo demonstration of the location and trajectory of white matter tracts. Tracking white matter fibers through the brain provides a way to spatially analyze the organization of brain structures and benign or malign masses. This capability is used by clinicians to guide pre surgical planning in various settings not only when important tracks are in the way of an intervention into a resection cavity of a superficial lesion, but also indirectly when targeting a deeply located lesion or functional structure through probable eloquent white matter. Despite this, current tractography methods suffer from readability and validity issues, heavily relying on manual segmentations or using regions of interest in the image for seeding probabilistic algorithms that follow the pathways. As such systems based on deep learning techniques are being proposed to reduce the time spent on the generation of these maps. This dissertation starts with a overview of the central nervous system, summarizing important anatomical structures and tissue composition culminating with the description of 2 major white matter tracts with the most clinical significance when discussing tumor resection: Optical Radiations and Corticospinal tract. It is also important to discuss the physical aspects that relate to MRI and the assembly of both structural and diffusion weighted images.Then simple summary of concepts related to Deep learning is then presented with an emphasis on CNNs and their most common components. A detailed review of the state of the art in brain image segmentation is then presented in order to establish a background from which the experimental work was built. As the most used base architecture for medical image segmentation was found to V-net or U-net variations, the experimental work focuses on investigating the use of multi-modal analysis of medical images through the attention gated y network. This model incorporates attention mechanisms in order to provide information on more salient areas of the image guiding the prediction in this way. The investigation focused on establishing a baseline for the AgYnet, then experimenting with a nested architecture derived from the Unet++ model and finally broadening the amount and quality of data from a larger dataset. In this work it was possible to obtain a segmentation of the optical radiations tract with a average DICE score of 65,79% (right) and 64,87% (left) and for the corticospinal tract of 63,72% (right) 61,63% (left). The contribuitions of this work include the tensorflow and keras implementation of AgYnet and nestedAgynet; and the verification of the hypothesis that with more data the AgYnet can produce very promissing results in the correct prediction of white matter tracts in the brain. On the other hand it was possible to conclude that the nested architecture provides little in the way of benifits to the quality of prediction while expanding immensely the ammount of time requiered to obtain them.

Acknowledgments

Em primeiro lugar quero de agradecer aos meus pais por me terem dado a oportunidade de frequentar um curso superior. O meu amor por eles e pelo seu sacrifício é única coisa que me motiva a acabar este precurso. Obrigado Pai e Mãe. Em seguida quero estender o agradecimento à minha irmã Mariana pelo apoio nos momentos mais nervosos e irritadiços dos meus dias. Quero agradecer tambem ao João, Gonçalo, a ambos os Franciscos, ao Frederico e à Lia, Pi, Clara,Maria,Isa, X, Estru e à Daniela e a todas as pessoas que me impediram de ficar no mesmo sítio. Agradeço ao Filipe e ao André pelo amor ao papel que partilhamos. Agradeço finalmente ao meu orientador o Professor João Tavares pela paciencia que teve e pela motivação que me deu. I wish to also thank Ilya for answering the basic questions of a student a whole continent away.

Data was provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

Jag har aldrig lärt mig att leva. Jag tränar varje dag.

Eva, read by Victor

Contents

List of Figures

Abbreviations

ANN	Artificial Neural Networks
AWS	Amazon Web Services
BET	Brain Extraction Tool
BKS	Behaviour-Knowledge Space method
CNN	Convolutional Neural Networks
CSF	Cerebrospinal fluid
DECMAP	Directionally encoded color map
DTI	Diffusion Tensor Imaging
DW	Diffusion weigted
DWI	Diffusion weighted imaging
FA	Fractional Anisotropy
fMRI	Functional Magnetic Resonance Imaging
GM	Grey matter
HCP	Human Connectome Project
IoU	Intersection over Union
MEG	Magnetoencephalography
MLP	Multi Layer Perceptron
MNI	Montreal Neurological Institute
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
PReLU	Parametric Rectified Linear Unit
ReLU	Rectified Linear Unit
WM	White matter

Chapter 1

Introduction

1.1 Contextualization

Medical imaging is fundamental in current clinical practice it comprises several techniques that allow clinicians to visualize structure and function of internal organs, bone and malign or benign growths. Magnetic Ressonance Imaging (MRI) is a technology that is mostly non-invasive if adding contrast is considered invasive, free of ionizing radiation unlike other imaging techniques as computed tomography, employed mainly to understand tissue and organ function as opposed to structure. Despite this, many soft tissue structures are not clearly visible in other image modalities such as the above-mentioned computed tomography. These characteristics make it ideal for brain imaging, an organ comprised of mostly soft tissues and encased in bone and dura matter. Not only is MRI applied to research connections and information pathways in the myriad of neurons that comprise brain but also is fundamental in providing insight for clinical diagnosis and treatment. Another application is not to diagnose but to plan interventions like tumour resections. The ability to visualize the position of malign masses, bleedings or other lesions makes MRI a very powerful tool.

These advantages are supplanted by the necessity of expert eyes to analyse the images obtained. Often surgeon spend several hours per patient trying to segment important white matter tracts that can be in the way of a tumour resection or trying to assess the size and shape of a mass. This time could be better spent accompanying the patient's treatment and wellbeing.

1.2 Motivation

To overcome the time consumed on expert segmentation of brain structures for either research or surgical planning, automated image analysis must be employed. As such motivation for this work comes from the need to leverage algorithms that segment and maybe provide contextual, readable and valid information.

1.3 Objectives

The objective of this dissertation is to produce a body of work that summarizes the current state of the art methods for brain image data segmentation using deep learning; provides insights into popular semantic segmentation deep learning architectures and how they relate to this specific problem and to provide some insights into solving the problem of automatic white matter segmentation in the clinical setting. In a more concrete way the objective is to improve the predictions of the AgYmodel by testing its variations and training it with more and better data.

1.4 Contributions

In this dissertation the main contribution to the current state of the art in brain imaging is the exploration of attention mechanisms in CNNs for the context of white matter segmentation. More over a up to date review of the state of the art is presented as are the results of experiments with the AgYnet and its variants.

1.5 Document Structure

This document begins with a contextualization of the biological and anatomical aspects regarding the central nervous system and its importance in 2.In 3 an overview of the MRI techniques is presented along with a brief presentation of the process and common mechanisms of deep learning techniques. In 4 the data organization and preprocessing required for this work is laid out and defined in order to provide a better understanding of the images and their specificities. The following chapter, 5, presents the methodologies and experiments made with deep learning techniques in order to improve the current state of the art, while presenting the DICE and IoU scores for each experiment, as well as a discussion at the end of the description of each.Finally in 6 the conclusions of this work are presented along with discussion on future work.

Chapter 2

Mapping the brain

This Chapter is intended to be a theoretical contextualization of the brain mapping problem starting with a description of the anatomy of the nervous system, then proceeding into MRI technology, describing the modalities of images and data touching on how they are used for tractography and finishing with a high level review of deep learning techniques with an emphasis on CNNs due to their importance in the field. The neuroimage visualization and processing software FSL is also briefly touched on due to its relevance to the field and this work.

2.1 The nervous system

The human nervous system, can be divided into two different complexes which are considered, for nomenclature purposes, two individual systems (Figure 2.1) [9]:

Central nervous system (CNS):
Comprising the brain and the spinal cord, encased in bone structures, is responsible for the processing, integration, storage and also the for the response to the information inputted by the the peripheral nervous system [9].

Peripheral nervous system (PNS):
Being external to the CNS, it includes the sensory receptors, nerves, ganglia and plexuses. Further, it can be split into two different divisions: the sensory/afferent division, responsible for the delivery of electric action potentials from sensory receptors to the CNS, and motor/efferent division that, in turn, controls the transmission of the electric signals from the CBS to the effector organs [9].

As one, the nervous system, is responsible for most of the body functions. Some of the functions with major importance are: sensory input, integration, control of muscles and glands, homeostasis regulation and mental activity [9].

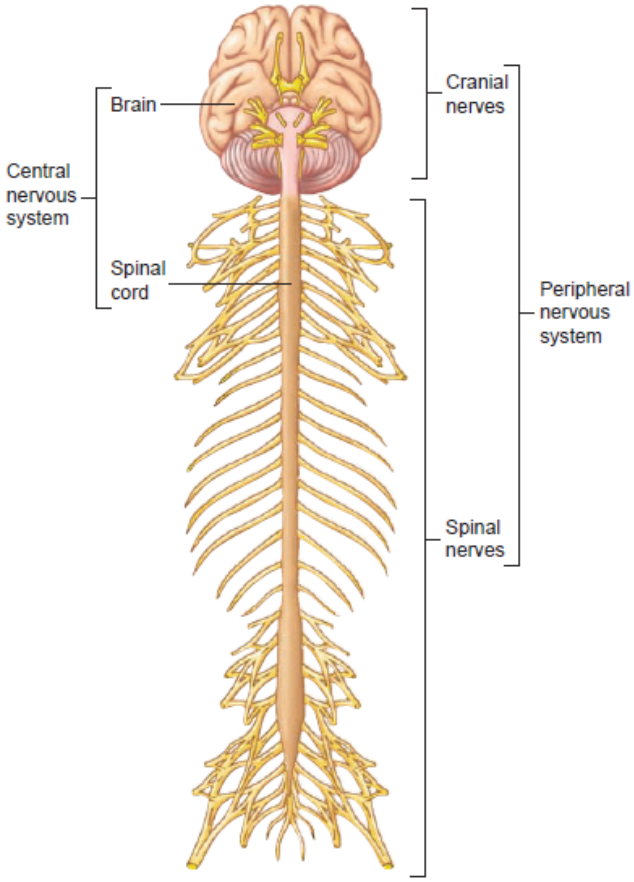


Figure 2.1: Functional organization of the nervous system. From [9].

2.1.1 The nervous tissue

2.1.1.1 Cells of the nervous system

Two distinct types of cells can be found when studying the nervous system: neurons, responsible for the reception and transmission of action potentials, and glial cells that give the neurons the needed support and protection, among other functions [9].

Neurons

The neurons are the basic unit of the entire nervous system and organize themselves in sophisticated networks for the constant transmission of action potentials [9, 10].

Each neuron is composed by three distinct structural regions: the neuron cell body, the dendrites and the axons (See Figure 2.2). Regarding the cell body, also know as soma, a single, rather large nucleus, centrally located and delimited by the nuclear envelop can be found, among regular cellular organelles [9, 10]. From the body, two types of processes arise. For most neurons, a single axon emerges from a cone-shaped section of the neuron body - the axon hillock. With a

constant diameter but variable length, the axon branches in order to form the presynaptic terminals, responsible for accommodating the neurotransmitters that trigger or hinder the postsynaptic cell. In turn, the dendrites are shorter cytoplasmic extensions usually extremely branched that act as the neuron input region, being able to produce electric currents that are conducted throughout the entire neuron [9].

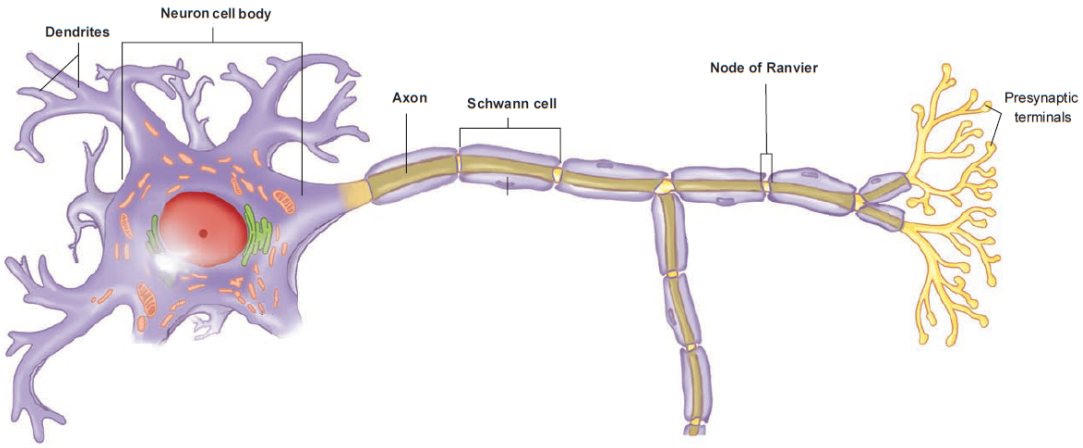
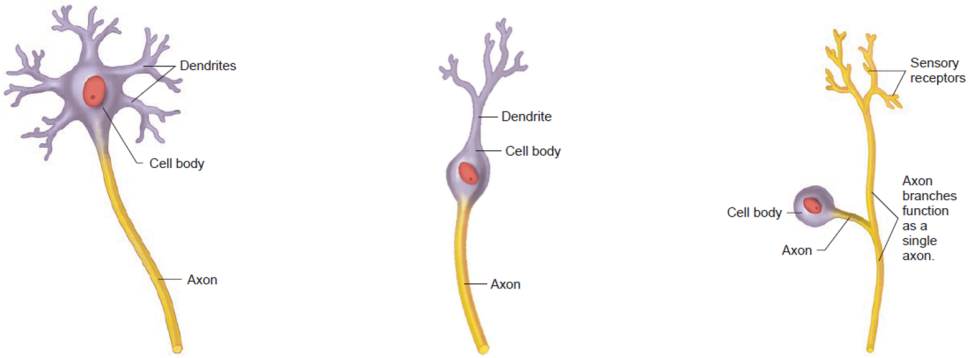


Figure 2.2: The structural features of a neuron. From [9].

It is possible to classify the neurons according to their function or to their structure [9]. Functionally speaking, a neuron can be sensory/afferent, motor/efferent or interneurons. While the first class conducts the signals from the CNS to the muscles and glands, the second one is responsible for the communication in the opposite direction. Interneurons, or association neurons, are the ones that control the conduction of the action potential within the CNS itself [9].



(a) Multipolar neuron (b) Bipolar neuron (c) Unipolar neuron

Figure 2.3: Classification for structurally different neurons. From [9].

Due to the large diversity of form and size that the neurons can present, also a structural classification is needed. The three major structural classes of neurons are presented in Figure 2.3. This classification approach takes into consideration the number of processes that arise from the

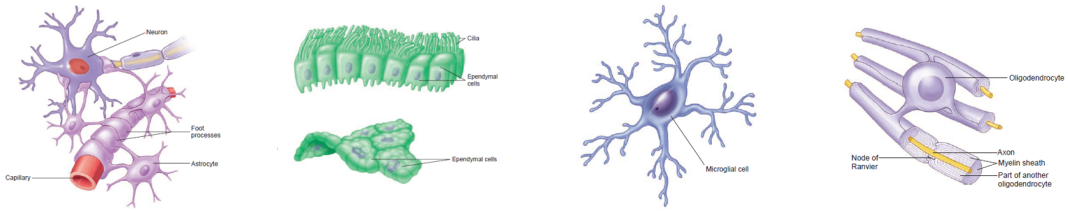
neuron cell body. Multipolar neurons (Figure 2.3a) display a single axon and multiple dendrites that differ in their branching extent. Bipolar neurons (Figure 2.3b) present only two processes, having a single axon and also a single dendrite. In contrast, from unipolar neurons (Figure 2.3c), arises a single process which, at a short distance form the cell body, splits into two branches; one leading to the CNS and another one, with dendrite-like sensory receptors, that extends to the periphery. Both branches act as a single cohesive axon [9].

Glial cells

Glial cells, or neuroglia, are long-lived cells [11] responsible for the establishment of a permeability barrier connecting neurons and blood. They act as a major defense source of the CNS cells, being able to produce myelin sheaths around axons and capable of execute some immune functions autonomously [12].

2.1.1.2 Glial cells of the CNS

Astrocytes (Figure 2.4a) are star-shaped cells that account for between 20 to 40% of all glial cells in the CNS [9, 13]. Having cytoplasmic processes that cover the surface of neurons, blood vessels and the pia mater, they provide structural and nutritional support to the neurons, modulate the neuronal activation's blood flow and are crucial regulators of diseases of the CNS [9, 13].



(a) Astrocyte (b) Ependymal cells (c) Microglia (d) Oligodendrocytes

Figure 2.4: Glial cells of the central nervous system. From [9].

In turn, ependymal cells (Figure 2.4b), which line the ventricles of the brain and the central canal of the spinal cord, when in combination with blood vessels, form choroid plexuses that secrete cerebrospinal fluid [9].

Microglia (Figure 2.4c) are cells that, in response to an inflammatory input are capable to migrate to the affected areas and become phagocytic to remove necrotic tissue, microorganisms or external substances that invade the CNS [9].

The final type of CNS glial cells, oligodendrocytes (Figure 2.4d) have cytoplasmic extensions that, when wrapped multiple times around the axons, create myelin sheaths [9].

Glial cells of the PNS

Regarding the PNS, two different classes of glial cells can be found 2.5.Firstly, Schwann cells, also known as neurolemmocytes, are cells which, likewise oligodendrocytes form myelin sheaths by wrapping around axon. However, contrarily to their homologous cell in the CNS that can cover several axons, each Schwann cell only wraps around a single axon.

Satellite cells can be found in sensory ganglia, surrounding the cell bodies of the neurons and are responsible for providing support, nutrition and protection against heavy metal poisoning to the neurons' cell body [9].

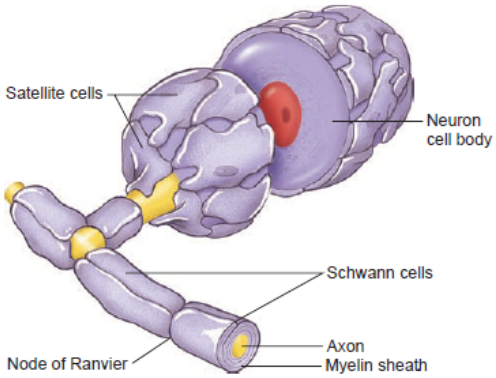


Figure 2.5: Glial cells of the peripheral nervous system, Schwann cells and satellite cells, associated to the neuron cell body. From [9].

2.1.1.3 Organization

When talking about major organization levels of the overall neuronal tissue, it is important to understand the myelin's importance.

The myelin sheaths produced by the oligodendrocytes in the CNS and by the Schwann cells in the PNS cover the neurons' axons forming either myelinated or unmyelinated axons, acording to the casing extent. Myelin, rich in phospholipids, is responsible for the protection and electric insulation of the axons. In myelinated axons, due to the discontinous nature of the myelin sheaths, there are locations every 0.3 to 1.5 mm were slight constritions exist, leaving 2 to 3 μm of the axon's lenght not covered. These interruptions, named Node of Ranvier, are also graphically presented in Figure 2.5 [9].

In this way, two significant classes of tissue arise. The first, white matter, consists in bundles of mostly myelinated axons and oligodendrocytes, presenting a whitish shade [9, 10]. In turn, the gray matter is a more vascularized tissue and comprises the majority of neuronal cell bodies, dendrites, glial cells and unmyelinated axons, imparting a gray colour [10].

It is possible, for example histologically, to distinguish the two different tissues. Figure 2.6 presents a microscopic image of a spinal cord's cross-section. Due to the staining process which

colours myelin in blue-black, it is possible to unveil the location of the white matter, richer in myelin, that, through these method presents itself darker than the gray matter tissue [14].

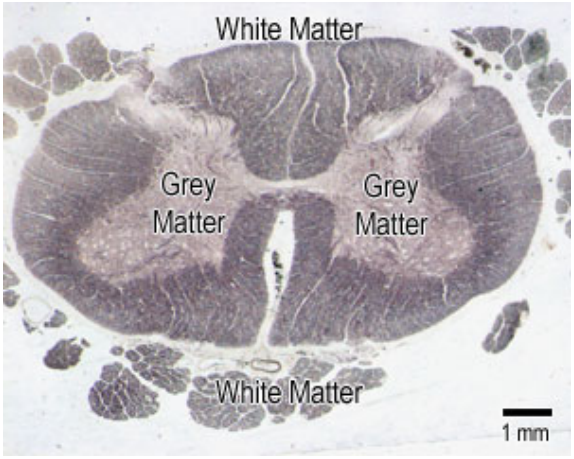


Figure 2.6: Spinal cord's cross-section microscopic image for tissue identification. From [14].

2.1.2 The command center: the brain

Without a doubt, the brain out-tops all the organs with its complex functionalities, being responsible for the majority of the body functions [9]. It is a gelatinous mass that weighs, on average, 1500 g and is protected by three layers of connective tissue, called meninges: dura mater, arachnoid mater and pia mater. Circulating along all the cerebral perimeter, passing between two of the meninges' layers, it is possible to find cerebrospinal fluid which is responsible for acting as a damper for the brain in rapid head movements [9, 10].

When analyzing the brain gross anatomy, one can consider four significant structures (See Figure 2.7).

The brainstem is responsible for the connection of the spinal cord to the cerebrum and is related to the to actions of the cranial nerves [9, 10]. The medulla oblongata, pons and midbrain are its primary constituints and, as many of the essential reflexes for basic survival are coordinated by this structure, small damage in the brainstem are often cause of death [9].

In turn, the cerebellum controls the muscle movement and balance. Further, it is responsible for the regulation of intentional movements and is also involved in the process of learning new motor skills [9].

As to the diencephalon, containing the thalamus, subthalamus, epithalamus, and hypothalamus, it can be found between the brainstem and the cerebrum. Altogether, it is responsible for acting as a control center for sensory relay and homeostasis and endocrine function regulation [9].

Finally, the cerebrum accounts for the biggest portion of the total brain weight; approximately 1200 g and 1400 g for females and males, respectively. Being able to override most other systems, it controls conscious perception, thought and conscious motor activity [9]. It is divided by a longitudinal fissure, giving rise to two distinct hemispheres where a great number of folds, that

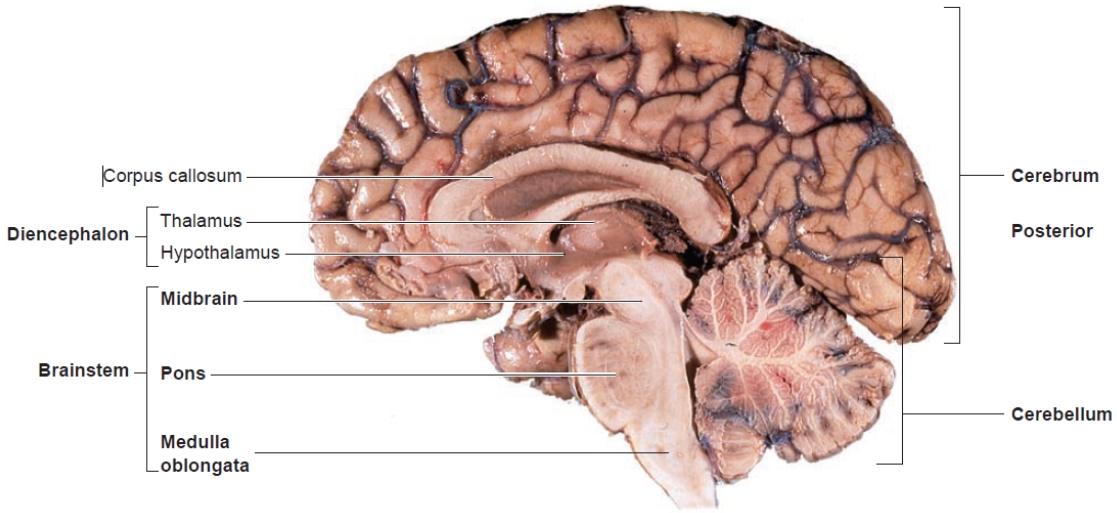


Figure 2.7: Regions of the right half of the brain - a medial view. From [9].

highly increase the surface, are the greatest features. Each hemisphere is divided into lobes that take the name of the overlying bone. Taking into consideration the local tissue identity, it is also possible to label some areas of interest. The cortex corresponds to the gray matter on the outer surface while deeper clusters inside the brain are nuclei. Between the cortex and nuclei, one can find cerebral medulla, rich in white matter [9].

2.1.3 Major white mater tracts

2.1.3.1 Corticospinal Tracts

The corticospinal tracts are probably the best known of all white matter tracts. They are projection rather than association fibers. They originate from the motor cortex of the pre-central gyrus, which gives off tracts in a pattern corresponding to the well-known “motor homunculus.” This fan-like projection of fibers converges and travels ventrally (downwards) via posterior limb of the internal capsule. The fibers originating at the more lateral regions of the pre-frontal motor cortex “twist” around to travel on the medial aspect of the descending corticospinal tract, giving it a characteristic appearance that can be replicated by tractography[15]. From the internal capsule, the corticospinal fibers continue ventrally into the brainstem via cerebral peduncle. They pass downwards through the pons and at the anterior aspect of the medulla they decussate (cross over the midline) in structures known as the pyramids. The result of decussation is that each motor cortex controls the limbs on the opposite side of the body. Corticospinal fibers originating from the motor cortex travelling latero-medially meet the perpendicularly travelling fibers of the arcuate and superior longitudinal fasciculi as they travel antero-posteriorly[15]. The area where these meetings occur is known as the centrum semiovale. Older tractography algorithms using diffusion tensor imaging (DTI) may have trouble tracking crossing fibers due to this interaction.

2.1.3.2 Optic Radiations

The optic radiations are the most posterior component of the visual pathways. They originate from the lateral geniculate nucleus (LGN), which lies on the undersurface of the thalamus. These fibers fan out from the LGN, first travelling medially and anteriorly within the medial temporal lobe, medial to the anterior horn of the lateral ventricle. They then turn obliquely, almost 180° as they pass over the ceiling of the lateral ventricle, whilst fanning out[15]. This turning is known as Meyer's loop. The fibers then continue posteriorly after turning, now lateral to the lateral wall of the third ventricle. They are separated from the actual lateral ventricular wall by another thin sheet of white matter known as the tapetum. The optic radiations with the sagittal stratum, deep to the inferior fronto-occipital fasciculus, to terminations above and below the calcarine sulcus (primary visual cortices)[15]. If a portion of the optic radiations is damaged due to a stroke or other insult, it usually results in partial loss of visual field corresponding to the damaged area. Meyer's loop is a very important surgical landmark, especially for temporal lobe surgery. Surgeons must be aware of the anterior extent of Meyer's loop and its distance from the temporal pole to avoid unintentional damage and subsequent post-surgical visual deficits[15].

2.2 Magnetic Resonance Imaging and Diffusion Tensor Imaging

Magnetic resonance imaging (MRI) is a medical imaging technique used in radiology to form pictures of the anatomy and the physiological processes of the body. MRI scanners use strong magnetic fields, magnetic field gradients, and radio waves to generate images of the organs in the body. In contrast to CT and PET which use ionizing radiation, MRI doesn't which makes it safer that these techniques[16].

Compared to CT, MRI provides better contrast in images of soft tissues, in the brain or abdomen. As such it is manly used to observe organ function, not structure as X-ray based techniques. Having said this, MRI can produce anatomical images but mainly for reference. However, it may be perceived as less comfortable by patients, due to the usually longer and louder measurements with the subject in a long, confining tube. Additionally, implants and other non-removable metal in the body can pose a risk and may exclude some patients from undergoing an MRI examination safely.

MRI was originally called NMRI (nuclear magnetic resonance imaging), but "nuclear" was dropped to avoid negative associations[17]. Certain atomic nuclei, like hydrogen, are able to absorb radio frequency energy when affected by an external magnetic field; the resultant evolving spin polarization can induce a signal in a radio frequency coil and thereby be detected[17].

In clinical and research MRI, hydrogen atoms are most often used to generate a macroscopic polarization that is detected by antennae close to the subject being examined.[18]. Hydrogen atoms are naturally abundant in humans and other biological organisms, particularly in water and fat. For this reason, most MRI scans essentially map the location of water and fat in the body. Pulses of radio waves excite the nuclear spin energy transition, and magnetic field gradients localize the

polarization in space. By varying the parameters of the pulse sequence, different contrasts may be generated between tissues based on the relaxation properties of the hydrogen atoms therein.

2.2.1 Types of MRI

2.2.1.1 T1 and T2

Each tissue returns to its equilibrium state after excitation by the independent relaxation processes of T1 (spin-lattice; magnetization in the same direction as the static magnetic field) and T2 (spin-spin; transverse to the static magnetic field). To create a T1-weighted image, magnetization is allowed to recover before measuring the MR signal by changing the repetition time. This image weighting is useful for assessing the cerebral cortex, identifying fatty tissue, characterizing focal liver lesions, and in general, obtaining morphological information, as well as for post-contrast imaging. To create a T2-weighted image, magnetization is allowed to decay before measuring the MR signal by changing the echo time . This image weighting is useful for detecting edema and inflammation, revealing white matter lesions, and assessing zonal anatomy in the prostate and uterus[19].

From a clinical perspective, T1 images show Lower signal for more water content as in edema, tumor, infarction, inflammation, infection, hyperacute or chronic hemorrhage[17].In contrast, they display higher signal for fat[17], paramagnetic substances, such as MRI contrast agents.T2 images show higher signal for more water content and for fat while showing low signal for paramagnetic substances and bone[17].

Together these comprise the standard foundation and comparison for other sequences.

2.2.2 Diffusion weighted imaging and Diffusion tensor imaging

2.2.2.1 Diffusion weighted imaging

In diffusion weighted imaging (DWI), the intensity of each image element (voxel) reflects the best estimate of the rate of water diffusion at that location. Because the mobility of water is driven by thermal agitation and highly dependent on its cellular environment, the hypothesis behind DWI is that findings may indicate pathological change. For instance, DWI is more sensitive to early changes after a stroke than more traditional MRI measurements such as T1 or T2 relaxation rates[20].

Diffusion-weighted images are very useful to diagnose vascular strokes in the brain. It is also used more and more in the staging of non-small-cell lung cancer, where it is a serious candidate to replace positron emission tomography as the 'gold standard' for this type of disease. Diffusion tensor imaging is being developed for studying the diseases of the white matter of the brain as well as for studies of other body tissues. DWI is most applicable when the tissue of interest is dominated by isotropic water movement e.g. grey matter in the cerebral cortex and major brain nuclei, or in the body—where the diffusion rate appears to be the same when measured along any

axis. However, DWI also remains sensitive to T1 and T2relaxation[20]. A special application of this technique is described below.

2.2.2.2 Diffusion tensor imaging

Diffusion tensor imaging is important when a tissue—such as the neural axons of white matter in the brain or muscle fibers in the heart—has an internal fibrous structure analogous to the anisotropy of some crystals. Water will then diffuse more rapidly in the direction aligned with the internal structure (axial diffusion), and more slowly as it moves perpendicular to the preferred direction (radial diffusion). This also means that the measured rate of diffusion will differ depending on the direction from which an observer is looking[20].

In order to discuss this technique in more detail one must clarify the concept of Brownian motion.

2.2.2.3 Brownian motion

The first formal description of diffusion was made in 1827 by Robert Brown, a Scottish botanist. While looking at microscopic particles through a microscope, he noticed that the particles seemed to move randomly - similar to how motes of dust, when observed moving through a shaft of light, appear to move in random patterns. Browning concluded that the movement was caused by smaller particles impinging upon the larger molecules seen through the microscope[21].

The properties of the medium determine the speed of Brownian Movement,increasing the temperature of the surrounding water will cause the water molecules to move faster, and consequently make the particles move faster as well. If the viscosity of the medium is higher the molecules move slower.Decreasing the temperature and decreasing the viscosity, on the other hand, will have the opposite effects.

This Brownian Movement of both the particles and the molecules is determined by the size and shape of the container. Food dye dropped into a spherical bowl of water will diffuse randomly in all directions. Food dye dropped into a cylindrical beaker, on the other hand, will quickly diffuse along the length of the beaker; the particles will soon run into the walls of the container, and be forced to move either up or down. We call this type of container anisotropic, meaning that the dimensions of the container cause the particles to diffuse along a predominant axis[21]. All of these factors -temperature, particle size, and viscosity- were combined by Albert Einstein into a single equation known as the Stokes-Einstein Equation 2.1.

$$D = (kT)/6\pi\eta r \tag{2.1}$$

The diffusion coefficient, D, increases as the temperature (T) increases, and deceases with higher viscosity (symbolized by eta) and a higher particle radius (r). k stands for Boltzmann's constant. This diffusion coefficient will play a role in how we acquire diffusion-weighted images, which we turn to next.

T1-weighted images are often used for anatomical scans, and that T2-weighted images are usually for functional scans. For T2-weighted images, the presence or absence of nearby oxygenated hemoglobin leads to an increase or decrease in the signal emitted by hydrogen protons of the water molecules in the brain. In this scenario a radiofrequency pulse is turned on in order to tilt the spin of the hydrogen atoms and is quickly shut off; the signal is then emitted by the hydrogen protons and detected by a sensor inside the magnet, and the process repeats until an entire time-series of functional data is generated[22].But if a magnetic gradient was applied while scanning the brain,it could make the magnetic field stronger along one direction and weaker along the opposite direction. Since the frequency of the spins of the hydrogen atoms is proportional to the strength of the magnetic field, it would be expected or the spins on the left side of the brain to be slightly slower than those on the right if , for example, the magnetic gradient applied was slightly weaker on the left side of the brain[22]. At this point, the spins would be out of phase with respect to each other, as the protons are out of phase we call the gradient we just applied the Dephasing Gradient[20]. If an equal and opposite Rephasing Gradient was applied the spins of the atoms would then be realigned with each other[22] .

These gradients are defined by the parameters on equation 2.2 where b is not surprisingly called b-value.

$$b = \gamma^2 G^2 \delta^2 (\Delta - \frac{\delta}{3}) \tag{2.2}$$

Note that the b-value is proportional to the magnitude of the gradient, duration of the gradient, and time between the gradients; if any of these parameters increase, the b-value increases as well[22].

In the above example, it is assumed that a rephasing gradient would put the hydrogen atoms back into alignment. This assumption is true, but only if the hydrogen atoms don't move in between the turning on and off of gradients. If, on the other hand, they do move - if they diffuse, according to the principles of Brownian movement that we discussed earlier - then the rephasing gradient will not lead to a realignment of the hydrogen atoms. Rather, they will be out of alignment in proportion to how much they have diffused in the time between the gradients[22]. This is the key insight through which it is possible to measure how much diffusion occurs in a direction between gradient changes and in which direction, in favor of the gradient, or against it. What is truly interesting is if there is a preferred direction of diffusion.

The result is a contrast image that looks similar to the T2-weighted images except in this case, greater signal loss implies more diffusion, with the amount of loss being relative to a scan that was acquired without any diffusion gradients being applied - in other words, relative to a scan that had a b-value of zero. A diffusion-weighted image with a b-value of zero is virtually identical to a typical T2-weighted image - CSF is bright and grey matter is dark. As we increase the b-values, we see that there is greater signal loss in specific parts of the brain, primarily within the white matter. This is because the water within those white matter tracts is diffusing primarily along the direction of the tract, and the image that is generated shows correspondingly lower signal[22]. In order to get the complete information from a DWI scan, it is necessary to know the direction of

the gradients that were applied. These directions are known as b-vectors.

2.2.2.4 Modeling the tensor

In order to build the tensor representation for each voxel the eigenvalues and eigenvectors are calculated. After getting combination of values that best represents the signal observed in the current voxel, we can use a number of different equations to calculate different properties of the diffusion at that voxel with the most popular being fractional anisotropy[23]. This measure allows the construction of color maps condensing the information from a multi sequence DW image into an RGB 3D image where each color channel represents one of the three principal directions and its intensity the amount of diffusion in that direction [23].

The diffusion tensor model describes the magnitude and direction of diffusion within a voxel as described in[24]. The diffusion tensor model the diffusion signal as:

$$\frac{S(g,b)}{S_0} = e^{-bg^T Dg} \tag{2.3}$$

Where g is a vector in 3D space that indicates the direction of measurement and b represents the parameters of that measurement such as the duration and magnitude of the diffusion weighting gradient; $S(g,b)$ represents the diffusion weighted signal measured in the g direction with b parameters and S_0 is the signal acquired with no diffusion weighting B_0.D is a positive-definite quadratic form which contains six parameters to fit as represented in 2.4.

$$D = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix} \tag{2.4}$$

This is a covariance matrix of the diffusivity along each spatial dimension. As noted in [24] has symmetry relative to the diagonal so for example $D_{xy} = D_{yx}$. This is why there are only six elements to estimate as opposed to nine. Bellow an example of a DTI image is presented 2.8.

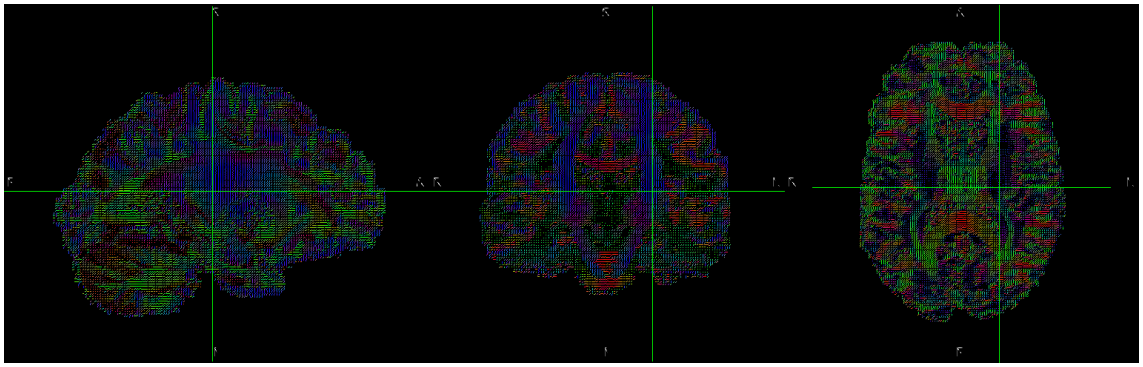


Figure 2.8: Example DTI image using lines to represent Fractional anisotropy

2.2.2.5 Crossing Fibers

As a final remark, this technique falls short when confronted with crossing white matter fibers since this representation has trouble representing more than one direction at the same time. When fibers cross what happens is the reduction of the overall signal and failure for tractography methods that rely on following the principal direction of diffusion. More recent algorithms that use a projection of b-values and b-vectors into spherical space exist[25] but are very time consuming and require acquisitions for tens of b-values and 32 to 64 orientations.

2.3 Deep Computer Vision Using Convolutional Neural Networks

This section is motivated by the extensive use of this technology in the current state of the art in medical imaging. A overview of the more repeated implemented layers is presented for completeness and better readability.

2.3.1 Convolutional Neural Networks

CNNs emerged form the study of the brains visual cortex and have been used in image recognition since the 1980s. Thanks to the contemporary increase in computational power, widespread availability of data and new strategies developed for the training of ever deeper networks these networks have been able to achieve superhuman performance in complex visual tasks. Examples of their application include self driving cars, image search services, video classification systems, computer assisted diagnosis and more.

In the late 1950's David H.Huble and Torsten Wiesel per formed several experiments on animal subjects [26] that earned the 1981 Nobel Prize in Physiology or Medicine. Specifically they showed that many neurons in the visual cortex have a small local receptive field, as such they only react to stimuli placed inside a specific region smaller than their overall range. As receptive fields overlap they form the whole visual field. Additionally they showed that some neurons react only to horizontal lines as other react only to lines with different orientations as they share the same receptive field but detect different shapes separately. Other neurons, with larger receptive fields, reacted only to combinations of patterns that lower level neurons already detect in isolation. These observations led to[27] and the 1998 paper [28] introduced the LeNet-5 architecture widely used to recognize hand written numbers.

2.3.1.1 Convolutional Layers

The most important building block of a CNN is, of course, the convolutions layer. Unlike fully connected layers, the first layer of a CNN is not connected to all pixels of an input image but only to those in its receptive field. Similarly the second convolutional layer is only connected to the neurons of a small area inside the first layer[29]. This allows the algorithm to focus on low-level features in the first layers and assemble them into complex combinations in later layers for a high

level representation of the input image. This hierarchical representation is common in images and that is one of the reasons CNNs work so well in image recognition tasks.

A neuron located in row i, column j of a given layer is connected to the outputs of the neurons located in rows i to $i+f_h-1$, columns j to $j+f_w-1$,where f_h and f_w are the height and width of the receptive field. In order for a layer to have the same size of the previous layer it is common to use zero padding. It is also possible to connect layers with very different sizes by spacing out the receptive fields, reducing computational load. The shift from one receptive field to the next is called the stride. The stride is usually the same in all directions but it is not mandatory. One can reduce the size of feature maps from one layer to the next by using a stride of 2 for example, undersampling the layer above[29].

In the case of a 2D image, which is useful for illustration, a neurons weights can be expressed as an image the size of the receptive field. A set of weights in a convolutional layer is commonly referred to as a filter or convolution kernel. If a layer is full of neurons using the same filter it outputs a feature map highlighting the regions that activate the filter the most[29]. These are not defined manually as the layer will learn the most useful filters for its task automatically.

In reality a convolutional layer uses as many filters as the designer wants, and outputs a feature map per filter. Consequently the output of a convolutional layer is best represented by a stack of feature maps. Having one neuron per pixel or voxel in each feature map, all neurons in the same feature map share the same weights and biases(parameters)[29]. In summary a convolutional layer applies multiple trainable filter to its inputs making it able to detect multiple features throughout the inputs. In 2.5 the output of a convolutional layer is defined for 2D.

$$z_{i,j,k} = b_k + \sum_{u=0}^{f_h-1} \sum_{v=0}^{f_w-1} \sum_{k'=0}^{f_{n'}-1} x_{i',j',k'} \times w_{u,v,k',k} \tag{2.5}$$

With $i' = i \times s_h + u$ and $j' = j \times s_w + v$ where $z_{i,j,k}$ is the neuron located in row i column j in feature map k of the layer l; s_h and s_w are the horizontal and vertical stride; f_h and f_w are the height and width of the receptive field, and $f_{n'}$ is the number of feature maps in the previous layer $l-1$;$x_{i',j',k'}$ is the output of the neuron located in layer $l-1$, row i' column j' and feature map k';b_k is the bias term for feature map k; $w_{u,v,k',k}$ is the connection between any neuron in a feature map k of the layer l with its input located at row u , column v relative to the receptive field and feature map k'[29].

2.3.1.2 Pooling layer

The role of a max pooling layer is to subsample its inputs, it is exactly the same as a convolutional layer but it has no weights, it only agregates the maximum responses in its receptive field. This reduces computational load, trims smaller responses to the filters of a convolutional layer and introduces some invariance to small translations[29]. But it is also very destructive since it is not obvious that a neuron ignored by a max pooling operation in a early layer wouldn't be useful in some deeper part of the network.

2.3.1.3 Activation

The feature maps from a convolutional layer are fed through nonlinear activation functions. This makes it possible for the entire neural network to approximate almost any nonlinear function.The activation functions are generally the very simple rectified linear units, or ReLUs, defined as $ReLU(z) = max(0,z)$, or variants like leaky ReLUs or parametric ReLUs. Feeding the feature maps through an activation function produces new feature maps[30][29].

2.3.1.4 Batch normalization

These layers are typically placed after activation layers, producing normalized feature maps by subtracting the mean and dividing by the standard deviation for each training batch. Including batch normalization layers forces the network to periodically change the neuron outputs to zero mean and unit standard deviation as the training batch hits these layers, which works as a regularizer for the network, speeds up training, and makes it less dependent on careful parameter initialization[31][29].

2.3.2 Transfer Learning

One of the most difficult parameters to tune when using deep learning models is weight initialization. There is no correct way to do it for any specific task so one answer could be to initialize weights randomly. An assumption of traditional machine learning methodologies is the training data and testing data are taken from the same domain, such that the input feature space and data distribution characteristics are the same. However, in some real-world machine learning scenarios, this assumption does not hold. There are cases where training data is expensive or difficult to collect. Therefore, there is a need to create high-performance learners trained with more easily obtained data from different domains. This methodology is referred to as transfer learning[32].This methodology will be used in this work in an attempt to improve the results that some model gives and as a way of enlarging the overal amount of data available for training a convolutional neural network.

2.4 FSL as a preprocessing and visualization tool

FMRIB[33] abreviated as FSL is a software library developed by the Analysis group at Oxford University that comprises a series of tools for functional, structural and diffusion MRI data. It is available for both MacOS and Linux operating systems with its first release being made in the year 2000 and more or less yearly updates. This software has many useful packages for neuroimaging but for this work the most relevant tools are BET[34] and FLIRT [35][36][37]. The brain extraction tool algorithm shipped with FSL is the golden standard for skullstripping MR images and is ubiquitous throughout literature in the preprocessing section and was used in this work for that purpose. FMRIB's Linear Image Registration Tool FLIRT is an algorithm for linear registration of brain images allows a user to register a low resolution MR image like a diffusion weighted image

to one with a higher resolution such as a T1 weighted image or some MR image to the MNI 152 template. This allows researchers to interpret results in some area of the brain against a brain atlas giving powerful insights for virtually any work in the neuroimaging field. This tool was used in this work to register diffusion weighted images to their higher resolution T1 weighted counterparts ensuring the cohesiveness of a dataset between modalities. The masking of the diffusion images and their subsequent reslicing was also done with recourse to this tool as described in 4.2. Lastly the FSLeyes module, presented in Figure 2.9, is an interactive display for 3D and 4D data image data is a fundamental tool for the visualization and qualitative assessment of brain images. It allows a variety of different perspectives and modalities to be visualized, masked and overlaid providing a reliable way to inspect brain images.

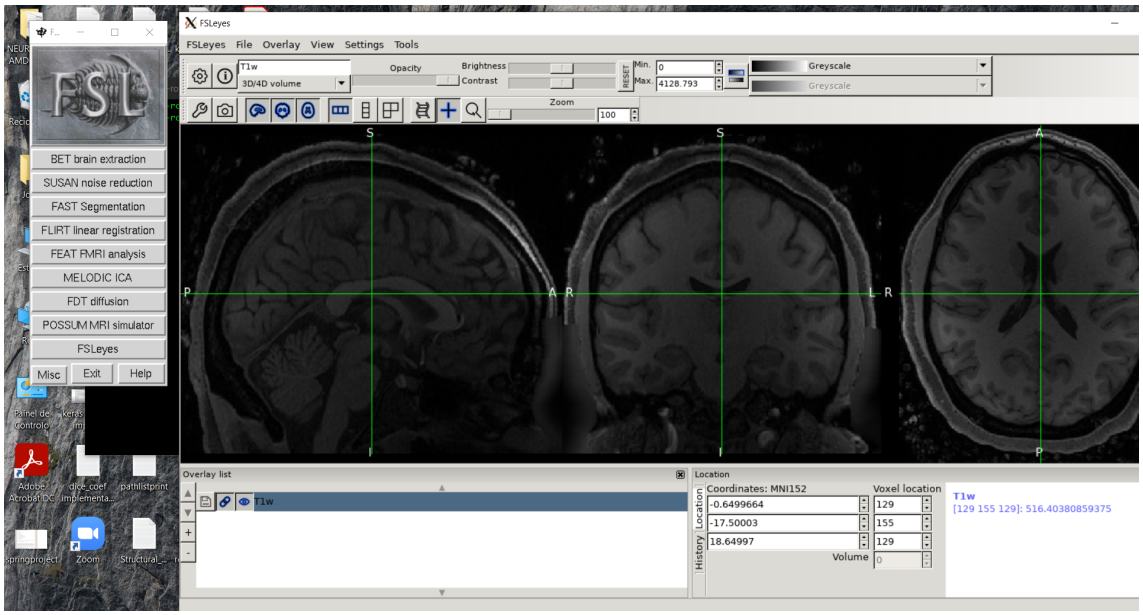


Figure 2.9: FSL and FSLeyes GUI example

2.5 Conclusion

In this chapter an overview of each element present in this work is presented. The intent is to establish a foundation that merges knowledge from the medical and engineering fields resulting in the start of a multidisciplinary work. The main take home messages of this chapter are the great diversity and complexity of cerebral tissue which in turn makes in intrinsically difficult to visualize; the fundamental principles behind the MRI acquisition process and the difference between structural and diffusion images; the basic functional parts of a deep learning algorithm and the popular neuroscience tool that is FSL for both visualization and image processing.

Chapter 3

State of the art in brain MRI image segmentation

3.1 Introduction

In this section a review of the current state of the art in MRI image analysis is made with an emphasis on segmentation. Currently most of the literature shows that CNNs are the standard for medical image segmentation. As such most of the papers reviewed use this technique to solve various challenges regarding brain MRI image segmentation.

3.2 MRI brain image segmentation state of the art

In this section, a review of deep learning methods for segmentation of brain or otherwise anatomical structures is presented. The reason for this is hat firstly the segmentation of white matter using learning algorithms is fairly recent and as such it is useful to also review segmentation and classification problems that use 3D images as input or 2D slices of 3D image modalities, preferably MRI, and aim to segment or localise biological structures for completeness.

The first paper to be reviewed is [38]. Here the authors intend to classify white matter lesions caused by Multiple Sclerosis in the brain. To do so the authors use a database of 150 brain MRI which they aim to classify into healthy and MS patients. This dataset is composed of 72 non MS brain images and 81 suggestive MS images and was made available in three orthogonal views – axial, sagittal, and coronal – although the majority of algorithms use only the axial view with a size of 508x508[38].

The pipeline proposed for this classification begins with the skull striping and noise removal of the MRI images as well as the registration reducing the image slice size to 180 x 240. This is a standard procedure throughout the following literature and some authors to be discussed further stop referring to it as this procedure became standard at the time of dataset construction. The authors do not discuss the method used for any of these pre-processing steps.

Sequentially these slices are fed to a CNN which the authors describe as being three layers deep composed of an input layer with as many channels as the input image, a convolutional layer and an output layer. In fact, to characterize the network , it is more useful describe how the network is structured. The convolutional part of this network starts with a convolutional layer with 6 channels and 7x7 kernels followed by a max pooling layer with 2x2 kernels. This architecture is repeated in sequence three more times with 16,30 and 60 convolutional kernels of size of 8x8,7x8,8x7, each of them followed by the max pooling layer mentioned above. The authors chose to use a fully connected (dense) layer at the end of the network to generate a vector with 120 features for each image which are then fed to a MLP classifier.

The MLP classifier decides between normal or suggestive MS and is designed with 120 neurons in the input layer, equal to the number of features as expected, 52 neurons in the hidden layer and one in the output layer which produces the output of the pipeline.

In practice and for comparison purposes the MLP was implemented and applied to the dataset with 2 activation functions tanh and tansig and compared to a pipline using GLCM followed by PCA and finaly the MLP. The evaluation was done using sensitivity, specificity and accuracy as metrics. The proposed algorithm of CNN plus MLP outperformed the more classical methods with both activation functions achieving the best result of 96.1% ,97.5% and 92.9% respectively. This work shows the early adoption and good results of CNN as a feature extractor and selector and is developed further by other authors.

Applied to the problem of brain tumour tissue segmentation, [39] the authors of this work apply CNN without a classifier as the final decision step as described above but with a final softmax layer within the network's architecture. The data used in this work consists of 4 different 3D MR contrast images: contrast enhanced T1 (T1c), T1, T2 and FLAIR and was provided by the BraTS challenge organizers. While T1c usually has an isotropic resolution, the other channels originally havea slice distance which is larger than the in-slice element spacing [39], as such all images provided were rescaled to the T1c resolution.

For computational efficiency and because of the lower resolution present in the axial direction the authors chose to extract 2D multi modal patches from the images for each point to be classified and stacked them in the axial direction. As such the problem is reduced to the 2D space enabling the authors to use implementations of 2D CNNs available online.

For pre-processing the authors use N3 bias field correction[40], normalization of the data and downsampling by two using nearest neighbour interpolation [39]. The algorithm is also evaluated on downsampled data. The reason for this reduction in resolution is not adressed in the paper but one can assume it is done for computational efficiency or to capture a larger area on the CNNs receptive field.

The network is comprised of six layers with an input layer of 19x19 with 4 channels followed by a convolutional layer with 5x5 kernels and 64 channels. Max pooling is applied next with 3x3 kernels and a stride of 3 followed by another convolutional layer with 3x3 kernels and 64 channels, then a fully connected layer with 521 neurons and a final layer that applies the softmax operation to the nodes outputting a probability of tumour or non-tumour to each point in the input patch. The

authors specify that all inner nodes use ReLU as non-linearity term [39]. For training logarithmic loss was used as the loss function and stochastic gradient descent with momentum as the learning algorithm.

The proposed method is evaluated against a previous method proposed by the authors [41] based on random forests. The results for CNN are obtained by a2-fold data split for training and testing. The results for RF are obtained with a leave-one-out experiment [39]. This preliminary evaluation heavily favors the CNN method.

Another patch-based approach is proposed by [1]. What sets this paper apart is the use of a structured approach to the segmentation problem and the separation of a four class segmentation problem into three binary segmentation problems. The work was developed with data made available by the BraTS 2014 challenge which asks for segmentation of tumour tissue in the brain with recourse to the four image modalities already discussed above. Alongside the implementation of a CNN the authors first attempt to extract structure form the available labelled data. The authors claim that 3D implementations of CNNs where out of scope for contemporary computers based on [42]. As such they take a patch based 2D aproach using only the plane with the highest resolution of the dataset claimed by the authors to be the axial plane.

For the structured part of the approach the authors choose a smaller patch centred around the centre pixel of the label image and build clusters using k-means and thus creating a cluster dictionary. From this, a cluster template is generated to be used in the final decision step. The authors then change binary segmentation problem into a K class prediction problem [1].

The authors identify each training image patch x with the group k that the corresponding label patch has been assigned to during the label patch dictionary generation. In prediction, the label template t of the predicted group k is assigned to the location of each image patch and all overlapping predictions of a neighbourhood are averaged. According to the experiments a discrete threshold 0.5 was chosen for the final label prediction[1].

The choice of CNN was made as it has the advantage of preserving the spatial structure of the input[1] and is organized as follows: 4 convolutional layers each with 24 channels and 20x20,10x10,6x6,3x3 kernel size filters respectively. After each convolutional layer sub sampling is made with recourse to a max pooling operation with 2x2 kernels. The network ends with a fully connected layer of K neurons referring to the number of clusters generated above. For inference on each slice the authors note that due to the pooling operations inside the network a prediction is made for each 4x4 region of the input image. To overcome this, two methods are tested: feeding the network with several versions of input image shifted on X and Y axis and merged the outputs properly versus a more common and faster approach as to upscale the label map to the size of the input image [1]. A schematic representation is presented in 3.1

As for pre-processing only N4 bias field correction was made. The evaluation of the algorithm's performance using optimal parameters for this dataset is made using the Dice score coefficient. For the test set, average Dice scores 83% (whole tumor), 75% (tumor core), and 77%(active tumor). The resulting Dice scores are comparable to intra-rater similarity that had been reported for the three annotation tasks in the BRATS data set [11]with Dice scores 85%

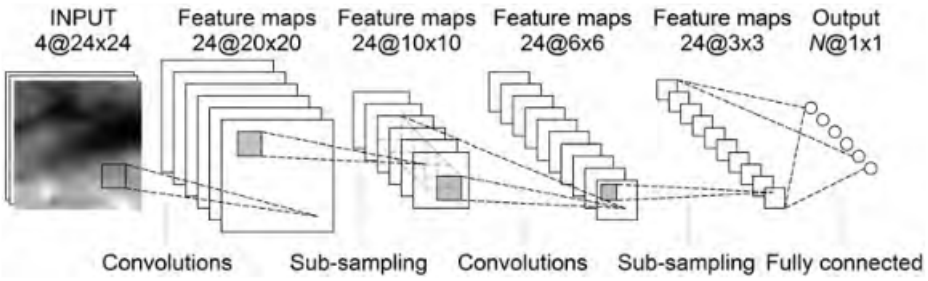


Figure 3.1: Schematic representation of the network used in [1]

(whole tumor), 75% (tumor core) and 74% (active tumor)and to the best results of automated segmentation algorithms with the Dice score of the top three in between 79%–82% (here: 83%) for the whole tumors segmentation task, 65%–70% (here: 75%) for the segmentation of the tumor core, and 58%–61% (here: 77%) for the segmentation of the active tumor region [1] paper. The authors noted that the algorithm improves upon the contemporary state of the art and presume that due to the fast inference time the approach could be improved with online updates for other datasets.

This approach using CNNs has the objective of identifying cerebral micro bleeds. Cerebral microbleeds (CMBs) are small haemorrhages of blood vessels in the brain, which are prevalent in the elderly population[43]. Images used in this paper were obtained using susceptibility weighted imaging, a subset of gradient echo MRI clinically employed specifically to detect internal haemorrhaging.

For this application the authors decided to start by identifying candidate regions of interest on non-normalized images by setting a statistical threshold. Then the center of mass of each 3D connected component is calculated and unusually large or small components are removed. The parameters for this operation were obtained empirically. Components that passed this observation are used for the rest of the method.

However, direct adoption of CNN with 2Dconvolutional filters in 3D medical imaging modalities could face the risk of over-fitting. Because more input channels of the third dimension could introduce more weights in the neural network compared to less input channels as would an RGB image have, for example. Moreover, limited available data can degrade the performance, especially in medical domain[43]. To address this consideration only 2D patches from the axial plane are extracted and fed to the network.

In this work a CNN was implemented only to play the role of a feature extractor whose predictions would be fed to a SVM classifier to further refine the segmentation. Training of the following CNN model was performed using negative log likelihood as a loss function, data augmentation and normalization were implemented, and dropout in the final layer. In order to preserve some tridimensionality in the decision process, features extracted using the CNN were concatenated along the axis from which they were retrieved. As such the output of this network was effectively a stack of 2D feature patches extracted from the same image, per image.

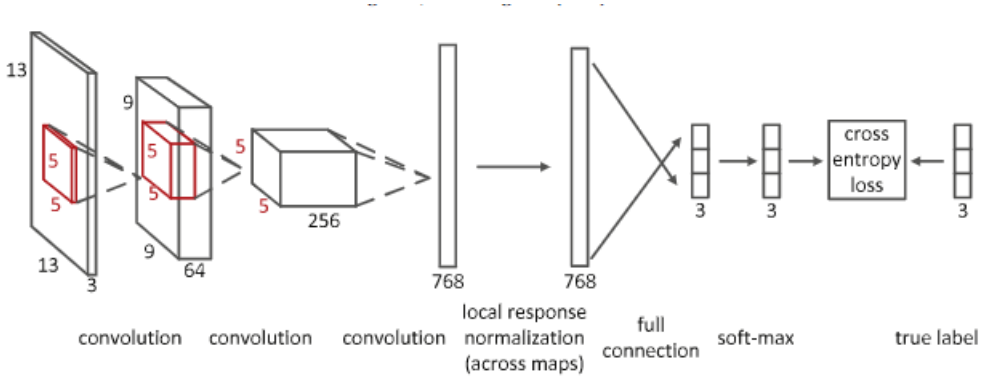


Figure 3.2: Ilustration of the network used in[2]

In practice, the data was normalized to have zero mean and unit standard deviation after trimming the top 1% of the intensities, augmentation was also performed by rotating the input patch by 90,180 and 270 degrees and a translation of the image between 1 and 4 voxels around the ground truth[43]. The authors noted that the thresholding step produced a large quantity of false positives per subject. To balance the positive and negative classes, the latter was sampled randomly from each image in the same quantity as the positive samples, labelled by experts. The authors state that the proposed method was compared to a CNN and RF method, but it is unclear in the paper whether they are referring to the feature extraction part or to the algorithm as a whole. In spite of this the paper claims to outperform other methods specifically by reducing the number of false positives, presenting 89.1% sensitivity and an average of 6.4 False positives per image volume with the closest method having 86.9% and 12.4 sensitivity and false positives per subject.

The following paper presented uses multi-modality MR brain images to segment white matter, grey matter and cerebrospinal fluid using CNNs. The modalities used are T1, T2 and fractional anisotropy images (FA). As such the authors attempt to solve a three-class segmentation problem. From a total of 10 subjects only 8 were used due to difficulties in obtaining T2 and FA modalities. The data was pre-processed by intensity inhomogeneity correction on both T1 and aligned T2 images. After that, skull stripping and removal of cerebellum and brain stem on the T1 image by using in-house tools was applied[2]. After registration, these structures were removed form the other image modalities.

In this method the authors start by dividing the image volume into 2D patches centred around each voxel for each modality totalling 10 000 patches. These were then fed to a CNN3.2 illustrated by the authors using input patches of size 13x13 into a convolutional layer with kernel size of 5x5 with 64 channels repeated twice more using ReLU as a non-linearity term after all convolutional layers. Then the feature maps produced were passed through a local response normalization layer to enforce competitions between features at the same spatial location across different feature maps (feature selection). Finally, the output was generated by a fully connected layer, with dropout, with three outputs corresponding to the classes to segment. A probability was given to each of the outputs via a softmax operation. For this multi-class problem, the chosen loss function was cross-entropy loss.

The authors experimented with other patch sizes and state that for the maximum patch size used, 22x22, a max pooling layer, with a 2x2 kernel and stride 2, was added after each convolutional layer.

Using the dice coefficient and modified hausdorf distance[44] the method was compared to a SVM and RF algorithms. The authors present extensive results and comparisons between the variations of initializing parameters and patch sizes for all algorithms and also access the importance of the multi modal nature of the data. While presenting all results is somewhat out of scope of this monograph the authors claim that results showed that the proposed model significantly outperformed prior methods on infant brain tissue segmentation[2].

DeepNAT is an algorithm conceived to segment neural anatomy using deep learning algorithms specifically CNN using three dimensional patches. The main innovation of this work is the use of two CNN stacked sequentially where the first learns to differentiate between background and brain and the second learns to identify each of the 25 anatomical structures the authors propose to segment. Unlike the former paper discussed, only T1 images were used here. Spectral features of the images are also used in this work, which was never specifically done before. The algorithm also performed predictions in an interesting way: predictions are made not only for a central voxel but also for its immediate, small neighbourhood. Then the final decision is made based on the average probability of the neighbourhood. The authors refer to this as multi-task learning as the network learns to make predictions on several voxels at the same time. The authors refer to tasks as the number of neighbours that participate in this process. Neighbourhoods of the 6 immediately connected voxels and the full cubed region were considered. In order to produce and refine the final prediction the authors also propose the use of a conditional random field after the networks output using the highly efficient approximate inference algorithm proposed by[45]to infer a fully connected CRF model on the entire 3Dbrain[46]. The authors propose spectral brain coordinates as an alternative parameterization of the brain volume, which can be obtained by computing eigenfunctions of the Laplace-Beltrami operator inside the 3D brain mask. Eigenfunctions of the cortex surface have previously been used for brain matching [47] and eigenvalues as shape descriptors [48]. In contrast, spectral coordinates are computed on the solid (volume) and use it as an intrinsic coordinate system for learning[46].

Both of the stacked networks are identical except for the last layer which has in the first case two output neurons (brain and background) and in the second 25 output neurons times the number of tasks. They are structured as follows: three convolutional layers with kernels 7x7x7 with 32 channels, 5x5x5 with 64 channels, 3x3x3 with64 respectively, each of them followed by a ReLU. Max pooling is done after the first layer with a 2x2x2 kernel with a stride of 2. Batch regularization is applied to the last two convolutional layers after ReLU. After this last convolutional layer a fully connected layer with ReLU and dropout is applied with 1024 neurons. After this the author remark that the coordinates for each patch are given to the network in order to reduce location ambiguity by concatenating the image content after the first fully connected layer with the location information (coordinates mentioned above) [46]. Finally, another fully connected layer with 512 neurons with ReLU and dropout finish the network with the last fully connected layer followed

by a softmax operation. The network is trained using multinomial logistic loss, an extension of multinomial logistic regression model that involves changing the loss function to cross-entropy loss and predict probability distribution to a multinomial probability distribution in order to apply it to multi class problems.

The algorithm is evaluated using the Dice score and across all structures, DeepNAT with the conditional random fields yields significantly higher Dice scores in comparison to DeepNAT , FreeSurfer [49],and STAPLE [50][46].

Deepmedic[3] is a 11-layers deep, three-dimensional Convolutional Neural Network for the challenging task of brain lesion segmentation. Four main ideas are prevalent in this work: firstly, what the authors call dense training[51]; secondly the use of segments and batches; thirdly smaller kernel sizes enable deeper networks with very similar receptive fields with mandatory batch normalization as a consequence; and finally multiscale analysis through parallel networks [3]. The work is very extensive as the authors test and present results in several BraTS and ISLES datasets as well as clinical data. A summarized review of the system is presented.

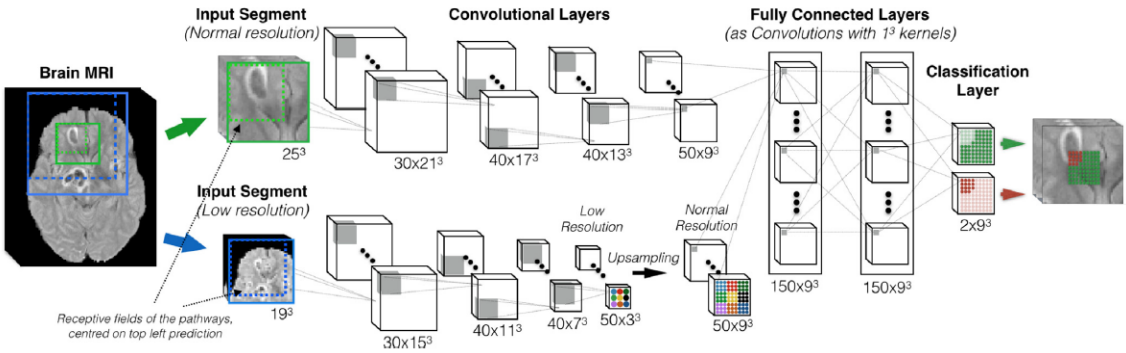


Figure 3.3: Representation of the Deepmedic network from [3]

Inspired by VGG16[52] Deepmedic is a CNN organized into two main paths: one path takes several 3D segment from an image and extracts 3D patches from it with the same resolution as the image; the other path takes the same segment but with lower resolution also extracting 3D patches. This allows the system to on one hand have a cascaded CNN extracting feature maps from high resolution small volumes with a lot of detail from which low level characteristics can be taken into account at the time of decision; on the other hand, it extracts high level features from the lower resolution segments in order to infer contextual relationships within the same space. This pipeline for segmentation also leverages high computational efficiency by using small 3x3x3 kernels for the convolution operations allowing more layers which turns out to be very effective. The conditional random field implemented at the end of the network helps to refine the segmentation and remove false positives[3]. The implementation of the network is summarized in3.3

Networks that are implemented as fully-convolutional are capable of dense-inference, which is performed when input of size greater than the CNN receptive field is provided[53]. This strategy significantly reduces the computational costs and memory loads since the otherwise repeated computations of convolutions on the same voxels in overlapping patches are avoided[3]. This approach

leverages the amount of voxels times the number of image segments per training batch. As the network makes predictions for each voxel this strategy for training as opposed to a prediction for a batch thus, according to the authors, reducing training time and speeding up convergence. As a consequence of this strategy, a way to reduce class imbalance becomes apparent: the authors simply sample image segments with a 50% probability of being centred on a foreground voxel, alleviating class imbalance.

Finally, a CRF is employed[45] and extended into three dimensional image space in order to smooth out outliers in the soft prediction maps outputted by the CNN.

The authors analyse the advantages of training in 3D space, the added scope and contextual information from using parallel pathways with different resolutions and state that Deepmedic surpasses contemporary state-of the art techniques achieving top results on both public benchmarks of brain tumours (BRATS 2015) and stroke lesions (SISS ISLES 2015). One of the most interesting findings is that the network learns to identify the ventricles, CSF, white and grey matter, suggesting that tissue differentiation plays a big role in lesion segmentation [3]. The authors also note a drop in performance when inferring on datasets retrieved from clinical data form other MRI scanners and imaging centres. One possible way of making the CNN invariant to the data heterogeneity is to learn a generative model for the data acquisition process and use this model in the data augmentation step[3].

This next article[54] deals with white matter lesion segmentation. The authors tackle this problem using three MRI image modalities T1,T2 and FLAIR and using two CNN in sequence where the first predicts if a voxel is lesion or non-lesion. The second one learns to reduce the number of false positives. Before training the author apply normalization as a pre-processing step and decide to use axial 3D patches centred on the voxel being analysed. Authors use the FLAIR modality as a reference since most lesions are placed in the boundaries between white matter and grey matter, as such voxels with an intensity less than 0.5 in the FLAIR channel are removed from all modalities. The authors justify this step in order to generate more challenging negatives. To deal with class imbalance authors also randomly undersample the non-lesion class to the same number of the positive class. This step is applied again to the predictions of the first CNN so the input of the second CNN is comprised a number of correctly classified for the positive(lesion)class and the same number of incorrectly classified voxels per patch. Finally the second CNN is trained with this new balanced set of input data.

Each network is composed by two stacks of convolution and maxpooling layers with 32 and 64 filters, respectively. Convolutional layers are followed by a fully connected layer of size 256 and a soft-max FC layer of size 2 that returns the probability of each voxel to belong to the positive and negative class[54]. ReLU is applied to all layers[55] to speed up training. Batch normalization[31] is applied after each convolutional layer and dropout[56] before the last fully connected layer.

The algorithm is tested on the MICCAI2008 dataset using metrics defined by the challenge alongside two private clinical datasets using dice score and precision rate. The paper surpassed other participants in all metrics.

Compared to other available methods, the performance of the proposed approach shows a

significant increase in the sensitivity while maintaining a reasonable low number of false positives. As a result, the method exhibits a lower deviation in the expected lesion volume in manual lesion annotations with different input image modalities and image datasets[54].

The most popular, contemporary design of CNN for medical images is the V-net[4]. This CNN architecture is an extension to 3D inputs of the 2D proposal U-net[57]. The network is fully convolutional and does not use max pooling layers, instead in employs further convolutions with larger stride in order to downsample the input and feature maps in an encoder-decoder setup. It also makes use of residual and skip connections in and additive fashion in order to preserve contextual and spatial information from higher levels of the network into the lower levels. These connections are also implemented to connect same level layers in the encoder and decoder. Encoder-decoder networks are a class of models that attempt to rebuild its inputs. As such V-net applied to a segmentation problem outputs a number of images equal to the number of segmentation classes. In the encoder or analysis path the network executes convolution steps increasing the number of channels while descending into deeper layers and smaller feature maps with residual connections surrounding every convolution block. In the subsequent decoder or synthesis path, the network upsamples the small feature maps into the output volume size while implementing the same skip connections around convolution blocks, in this case to preserve detailed information contained in in previous channels while also incorporating information from features coming from the analysis pathway.The authors also developed a custom loss function based on the derivative of the dice score. A schematic representation is provided in 3.4.

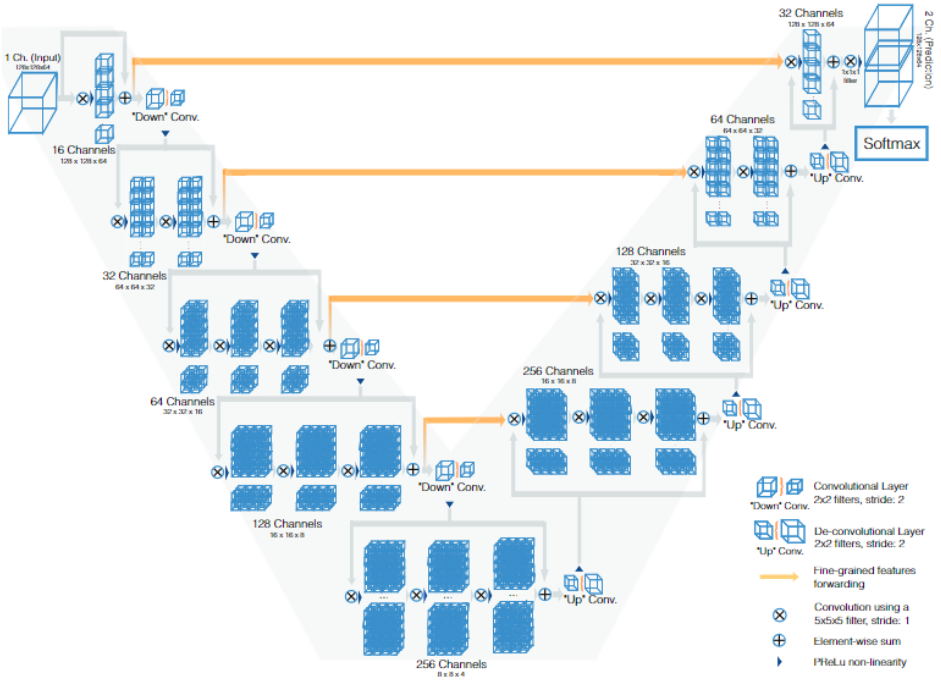


Figure 3.4: Illustration of the 3D V-net in[4]

Using this formulation, the authors did not need to establish the right balance between foreground and background voxels[4]. This is because the dice score compares the ground truth segmentation the one obtained through the CNN. As a consequence, the optimization of the similarity between ground truth and predicted segmentation is embedded in the learning process. In fact, the authors state that they obtain results that are much better than the ones computed through the same network trained optimising a multinomial logistic loss with sample re-weighting[4].

The current state of the art method for white matter tract segmentation is presented in the Tractseg paper[5] and implements this very architecture albeit in a 2D setting. This paper combines the usage of tractography algorithms with the architecture in [4]. It also leverages heavy data augmentation in order to segment 72 WM structures in the brain. Besides the state-of-the art WM mapping algorithm the authors discuss at length the methods employed to create reference segmentations for testing the algorithm and make this dataset available for public use[5]. The design of the algorithm starts with extraction of peaks of the result of a CSD of a set of MR images. This provides the principal direction of diffusion of water molecules in the brain. Each voxel then contributes with three principal directions comprising nine channels. These input images are then separated into slices according to the sagittal, coronal and axial orientations. The slices are fed to a V-net like 2D CNN[57] which is trained once for each of the orientations. This first CNN already produces a segmentation, but the authors feed its outputs to a second CNN also run once for each orientation and once for each target structure. This last network is implemented in order to imbue the algorithm with the ability to learn the best way to combine the segmentations on each orientation outputted by the first network. The second network is also rerun three times per voxel per tract outputting then 3 predictions per voxel per tract. Since the prediction is a probability resulting from a sigmoid operation, these are averaged, and the images are re-stacked, and thresholding is used to obtain the final 3D image segmentation. As the system only predicts one class at a time, binary cross entropy loss is used. A scheme of the system is presented in3.5.

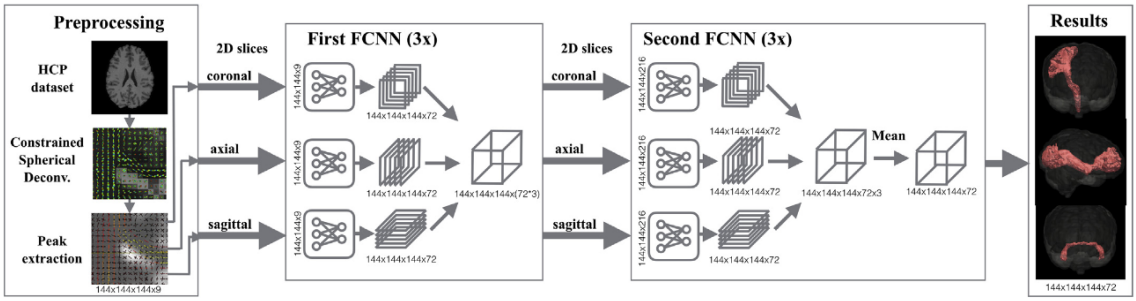


Figure 3.5: Scheme representing the network employed in [5]

Pushing beyond the current state of the art, two interesting works are presented, both based on the concept of attention mostly used in natural language processing now adapted to computer vision. In deep learning attention is a technique that allows models to focus more resources, giving more weight, to specific parts

of the input data. In a nutshell, attention in deep learning can be broadly interpreted as a vector of importance weights: in order to predict or infer one element, such as a pixel in an image or a word in a sentence, we estimate using the attention vector how strongly it is correlated with other elements and take the sum of their values weighted by the attention vector as the approximation of the target.

In [6] a V-net[4] is used in combination with attention mechanisms developed in [7]. Here the network is provided two sets of inputs, the T1 and the colour map representing the principal direction of diffusion[23]. These run in parallel network paths through an encoder network almost identical to [4] except the skip connections to the decoder network which are fused by an attention model developed in [7] that provides more weight to features that already exhibit high responses in previous convolution operations. The result of this attention based fusion is then concatenated with the information of the same hierarchical level in the decoder network. For comparability the authors use the dataset made available by the authors of the TractSeg algorithm[5] but do not present results for all 72 tracts. This paper is set as the baseline for the challenge that motivated this monograph and dissertation.

Recently attention was taken a step further, with the paper [58]. This paper proposes an architecture for classification with no convolutions, just attention mechanisms. Inspired by this the authors of [59] apply the transformer architecture to the encoder part of the V-net design. In [59] the first study which explores the potential of transformers in the context of medical image segmentation is presented. However, interestingly, it was found that a naive usage (i.e., use a transformer for encoding the tokenized image patches, and then directly upsamples the hidden feature representations into a dense output of full resolution) cannot produce a satisfactory result. This is due to the fact that transformers, unlike CNNs, do not use a hierarchical representation of features but instead try to capture global context in all stages. In this paper the authors propose TransU-net, establishing mechanisms self-attention mechanisms from the perspective of sequence-to-sequence prediction[59]. Self-attention refers to is an attention mechanism relating different positions of a single sequence in order to compute a representation of the same sequence.

The presented system leverages a hybrid Transformer-CNN architecture in order to compensate for the shortcomings of Transformers referenced above. Hence, the system first employs a 2D CNN in order to build image patches which are then flattened and used as input to the transformer. They are also saved to use as skip connections in the decoder part of the system similar to [4]. The flattened feature maps are fed to the transformer who produces patch embeddings, in other words, the transformer learns which features are important and which aren't. These embeddings are them upsampled before entering the decoder once again employing the same system as [4]. One problem with the use of transformers, which illustrates why they are so prevalent in natural language processing, is the amount of data necessary for training which is widely available in other fields but not in medical imaging. As such the transformer implementations presented in this study were pretrained in ImageNet[60]. This work was developed using CT scans of the chest aiming to segment organs. The authors conclude the work by stating that the proposed method is less likely to over or under segment(less false positives) and is favourably compared to other

methods which lack the contextual spatial information and high detail obtain using their design.

3.3 Conclusion

In this chapter the state of the art in brain image segmentation was detailed. We can conclude that the V-net CNN architecture and its variants are the most popular algorithms for brain image segmentation, and for good reason. Currently variations of this architecture are competing to be the best segmentation model in a variety of contexts unimodal or multi-modal. Finnaly we can notice a shift toward attention mechanisms and even attention based architectures which not only seem to be more effective but also can provide an avenue for explainability.

Chapter 4

Datasets and preprocessing

This chapter is intended to document the tasks that need to be done before even thinking about doing any deep learning. The standardization and preprocessing requiered to achieve data quality to feed to a deep learning algorithm are discussed here.

4.1 Datasets and data format

In this section a detailed description of the datasets used in this work is provided along with their advantages,disadvantages and the motivation behind their assembly.

4.1.1 NIFTI file format

The Neuroimaging Informatics Technology Initiative (nifti)[61] file format was developed as an update for the widespread in neuroimaging but problematic analyze 7.5 format. One of its main flaws was the lack of adequate spatial orientation information such that the stored data could not be unambiguously interpreted. The new format was defined in two meetings of the so called Data Format Working Group (dfwg) at the National Institutes of Health (nih) in which the are Representatives of some of the most popular neuroimaging software agreed upon a format that would include new information, and upon using the new format, either natively, or have it as an option to import and export. Perhaps the most visible consequence of the lack of orientation information was the then reigning confusion between left and right sides of brain images during the years in which the analyze format was dominant. At this time researchers became used to describing an image as being adherent to "neurological" or "radiological" orientation. These terms were but a reflection of the physician's clinical activity since radiological exams show the patient usually from the front as a practitioner would observe a patient, whereas neurological practice more commonly displays images from the back of the patient. This was troublesome since some software would display the images one way or the other without a set standard. The nifti format obviated all these issues, rendering these terms obsolete. Software can now mark the left or the right side correctly, sometimes giving the option of showing it flipped to better adapt to the user personal orientation preference. Moreover, the nifti format now included the medical image data

and corresponding header file detailing the way the image was acquired, the affine transformation between real space and the image space, and other information which was until then left separate leading loss of corresponding files as well as other errors. In practice it is also common to have binary masks, regions of interest, and images with large amounts of background voxels which took up a lot of space on disk but do not provide much actual information. As such nifti files were design to be compressed by the deflate algorithm used by gzip[62]. As such, it is common to see and use the .nii.gz file extention. In the nifti format, the first three dimensions are reserved to define the three spatial dimensions — x, y and z —, while the fourth dimension is reserved to define the time points — t. The rest of the file is comprised of metadata relating to image acquisition.

All images provided in the following datasets were originaly in the nifti file format and compressed.

4.1.2 BrainPTM 2021 dataset

The dataset was provided by the organizers of the BrainPTM2021 challenge and made publicly available at https://brainptm-2021.grand-challenge.org/. The challenge organizers are CILAB https://www.cilab.org.il/ from Sheba medical center, Israel. It is comprised of 75 clinical cases of patients indicated for brain tumor resection with T1w Structural and Diffusion Weighted modalities. Patient diagnoses include oligodendrogliomas, astrocytomas, glioblastomas and cavernomas, on first occurrence or in a post-surgical recurrence.The patients underwent pre-surgical DTI mapping of the motor, language, visual or a combination of those tracts, depending on their proximity to the lesion. According to the neuroradiologist's estimation, the tumor volumes ranged from 4 (cavernoma) to 60cm^3 (glioblastoma multiforme). Also, different levels of edema are present around the dataset tumors from inexistent to very significant[63]. Image acquisition was made on a 3T Signa machine (GE healthcare, Milwaukee) with a diffusion weighted protocol of 64 gradient directions at B0 = 1000 and one scan at B0 = 0. T1w with no contrast injection was acquired at 1x1x1 mm^3 resolution and the diffusion scan with 1x1x2.6 mm^3. The Diffusion scan was processed with MrDiffusion toolbox [64] for Matlab and registered to the higher resolution T1w. Finally T1w scans were skull stripped with BET [34] and resliced to 1.25x1.25x1.25 mm^3, diffusion scans (registered to T1w) were resliced to 1.25x1.25x1.25 mm^3 as well to ensure dimensional equivalence. Each T1w and diffusion weighted scan is stored in a separate nifti file with 128x144x128 spatial dimension. Ground truth images were acquired semi-manual process for each scan optic radiation and Corticospinal tracts were manually segmented in 3-D for both hemispheres of each brain by an expert in neuroanatomy and verified by a senior neuroradiologist using the MrDiffusion GUI [64]. Optic Radiation (OR) and Corticospinal tracts (CST) (left and right) were chosen as they are of significant interest during neuro-surgical planning. Example images are shown in 4.1.

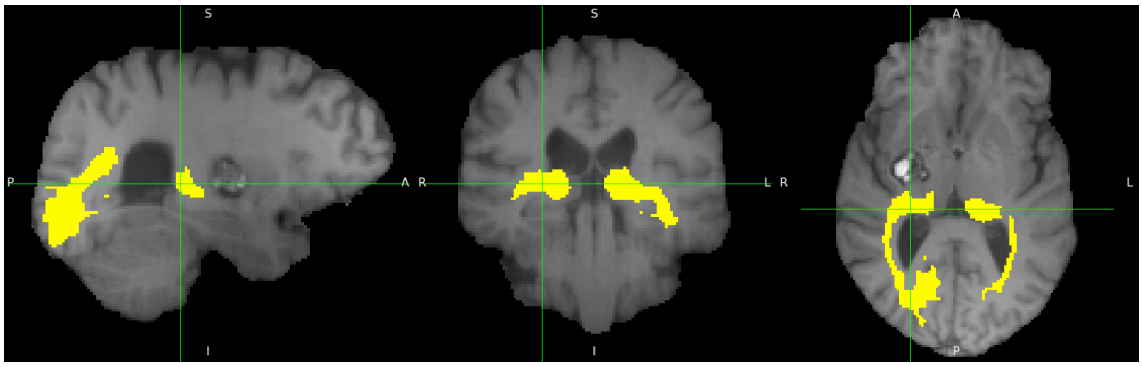


Figure 4.1: T1 weighted image of case_1 from the BrainPTM2021 dataset overlaid with the ground truth for the optical radiations right and left tract (yellow)

Tracts annotations a were provided only for 60 cases (train set) while the rest 15 cases are used for participants algorithms evaluation, so only the input cases scans were provided (T1w and DW). For four of the 60 training set cases 4 were excluded due to lack if CST annotations. Challenge organizers claim that this was due to difficulties in tractography for these cases.

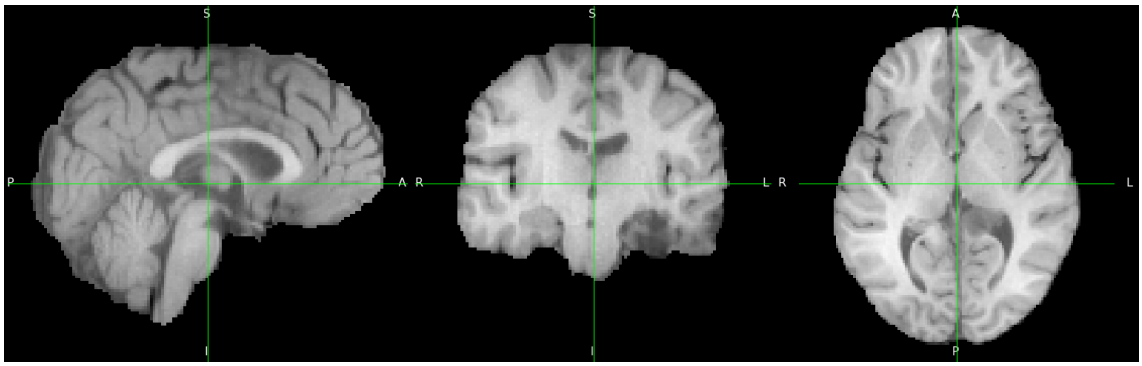


Figure 4.2: T1 weighted image of case_3 from the BrainPTM2021 dataset

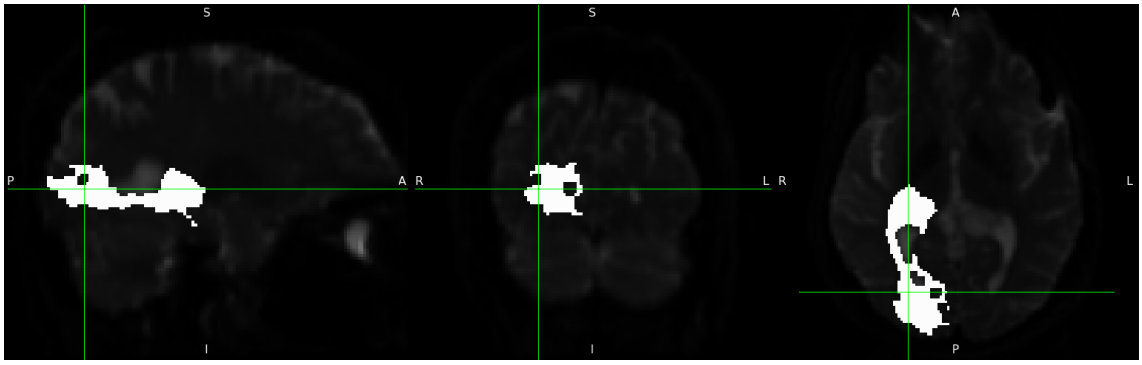


Figure 4.3: Diffusion weighted image of case_3 from the BrainPTM2021 dataset overlaid with the Optical radiations right tract

4.1.3 HCP 1200 young adult dataset

The 1200 Subjects Release (S1200) includes behavioral and 3T MR imaging data from 1206 healthy young adult participants (1113 with structural MR scans) collected in 2012-2015. In addition to 3T MR scans, 184 subjects have multimodal 7T MRI scan data and 95 subjects also have some resting-state MEG (rMEG) and/or task MEG (tMEG) data available. For the purpose of this work only T1 weighted and diffusion weighted modalities were used. Further more the subset of cases used for training and evaluation on this dataset were the 105 cases whose ground truth tract segmentations were made available in [5]. For these 105 cases the data was downloaded into the remote machine used for computation in this study via the AWS S3 open access data provided. Relevant data was identified in the database using S3 Browser and a python script was created using the AWS SDK for Python downloading the 150 Gb of data. A similar approach was made for the collection of ground truth images relating to these HCP cases provided publicly by [5] and located at [65]. This data was also downloaded directly into the remote machine via python script. Here the data provided was in the form of tractograms for each of the tracts used in [5] of which only four were used in this work, namely: OR_right, OR_left, CST_right and CST_left. Originally as extracted from the open access data base the T1w data4.4 was 260x311x260 with a slice thickness of 0.7x0.7x0.7 mm^3 and the diffusion weighted data4.5 was 145x174x145 with 270 gradient directions with 3 b-values(1000, 2000, 3000 $s/mm2$) and 18 B_0 images with a slice thickness of 1.25x1.25x1.25 mm^3. In this case the images are from young and healthy adults, as such no pathological structures were expected to appear in the images and none did. The images used for the deep learning pipelines that had already been processed by the minimal preprocessing pipeline (e.g. distortion correction, motion correction, registration to MNI space and brain extraction had already been completed)[66].

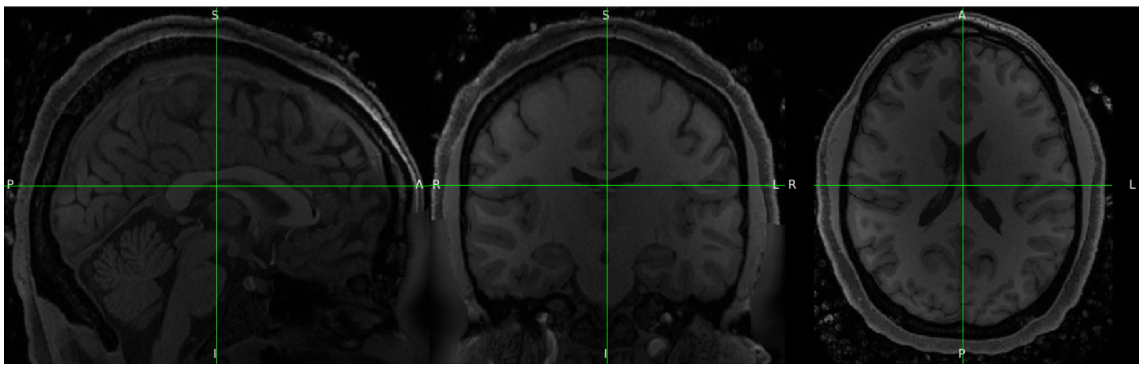


Figure 4.4: T1 weighted image from case 599469 of the open access HCP 1200 young adult dataset

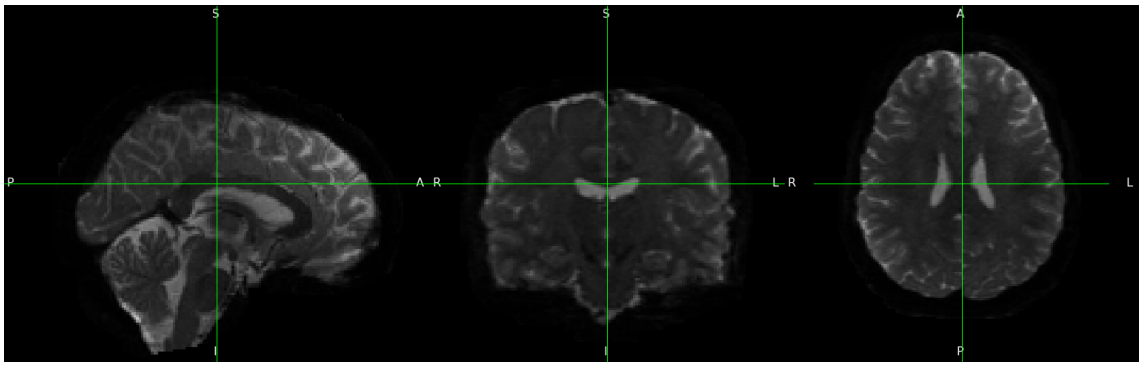


Figure 4.5: Diffusion weighted image from case 599469 of the open access HCP 1200 young adult dataset

From the tractograms provided the derived ground truth images were created to fit the diffusion weighted data. Details on this process and changes made for the T1 data and diffusion data are discussed in 4.2.3.

4.2 Preprocessing

In this section all image preprocessing that was done on the data provided and data organization for this work is discussed.

4.2.1 Dataset organization

In order to efficiently use a dataset for any machine learning or deep learning application, there needs to be a structure of folders and subfolders that hold training images for each modality and their respective labels. For each dataset set used, images were saved in a training, validation and test folder respectively. Every one of these folders contains a subfolder holding T1 weighted images; a subfolder for the colormaps that represent the DTI;a labels folder that holds four other folders with the label names that contain the respective binary masks. All images were transformed to the .npy format which is easily readable by the image data generator and subsequent deep learning models.

4.2.2 BrainPTM2021

For this dataset most of the preprocessing was done by its providers. Only the transformation to DECMAP and finally to the .npy format was done locally as all images had already been re-sliced to match in dimensions and resolution. The subjects that had no CST annotations were ignored do the image count was lowered to 61. As for the training, validation and test splits the data was divided into:46 subjects for training, 10 subjects for validation and 5 for testing.

4.2.3 Human Connectome Project 1200 young adult

Processing for this dataset started with skullstripping the T1 weighted images. This is standard procedure in neuroimage as the bone and dura mater are of no concearn for the problem at hand. The method used was FSL's Brain Extraction Tool [34] a well known and trusted method. Also using FSL, the higher resolution T1 images were resliced to the same resolution as the diffusion images using FLIRT, FSl's linear registration tool[35][36][37]. As discussd above, image labels were not provided directly, what was provided were the tractograms for the 105 subjects used in Tractseg[5]. These tractograms were transformed into binary masks using the Dipy module to open and process the tractograms. The specific method used was the same as in [5] not only for comparison fairness but also because its authors provided the tractograms and made the script for the transformation to mask publicly available. The only changes made to the script were due to function deprecation issues in the Dipy module. As above the diffusion images were processed to their DECMAP[23] counterparts using the Dipy module as described in 4.2.4 and all images were saved in the .npy format.

As recommended in the dataset in [65] the data was divided into the 5 folds that the authors used to perform 5-fold cross validation. As in this work cross validation was not performed due to time constraints the first four folds were used for training and the fifth for validation. As presented before the dataset is comprised of 105 cases but no T1 weighted image was found for one case. This was of no consequence for the authors of [5] since they only used diffusion tensor images but for the purpose of a multi modal approach it does matter so the case was discarded. As such the data was organized into 86 cases for training and 18 for validation.

4.2.4 Diffusion Images Diffusion Tensor Images

In order to translate diffusion images to DTI in the form of the directionally encoded colormaps proposed in [23] the Dipy [67]module was used. Firstly, images were loaded and their affine transformation was extracted for posterior translation into real space. Then the image was masked using a binary image were only the brain region was true. This step was done in order to avoid tensor calculation in the image background. Then the b-values and b-vectors data was loaded and matched to the corresponding channel in the diffusion image. As described in 2.2.2.4 and proposed in [24] a tensor model was created and fitted to the image producing the FA value for each voxel. This FA value was first scaled to fit between 0 and 1 then used to compute the DECMAP atribuiting a color to each direction and using the FA value for magnitude,thus generating a 3 channel image with the same dimentions as the diffusion weighted image. This process was done once for each image. An example is presented bellow4.6 for the HCP dataset and 4.7.

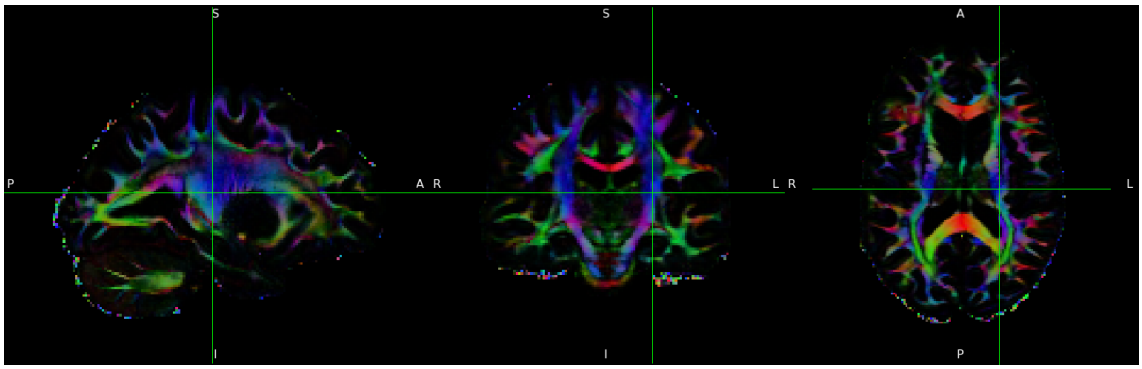


Figure 4.6: DTI from case 599469 of the open access HCP 1200 young adult dataset

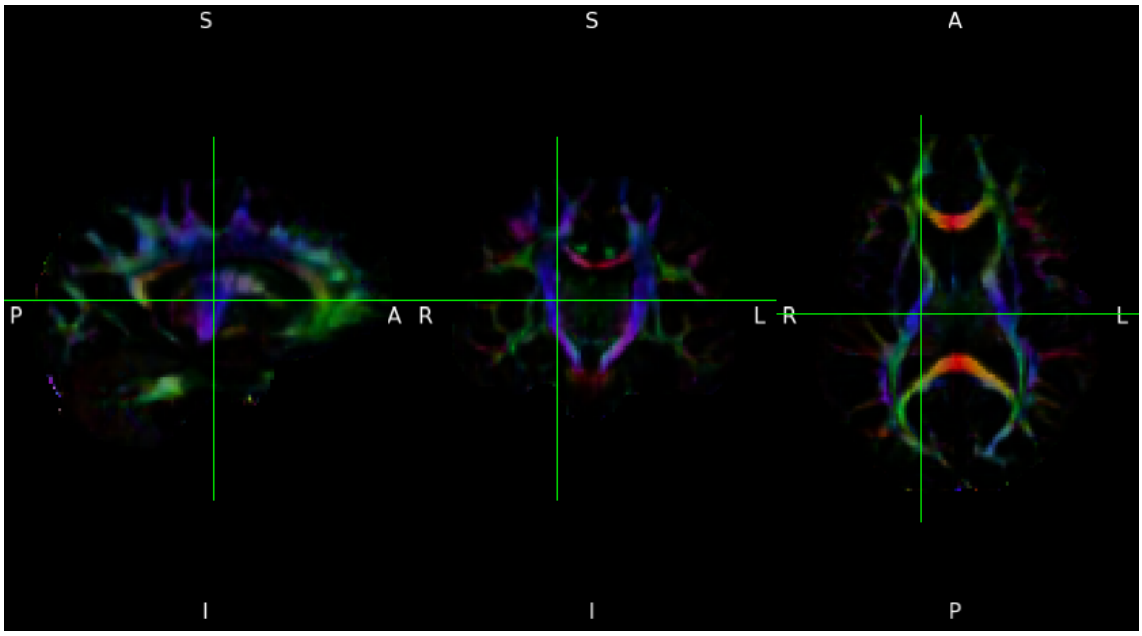


Figure 4.7: DTI from case_3 of the open access BrainPTM2021 dataset

Differences in quality of the tensor image are due to the large difference in the number and direction of acquisitions made during diffusion weighted image scanning. As mentioned before the DWI from the BrainPTM2021 dataset were acquired with 64 gradient direction and 1 B_0 image while those from the HCP dataset were scanned in 270 directions with 18 B_0 images which provides much more detailed information at the time of tensor fitting.

4.2.5 Image data generator and data augmentation

Deep learning pipelines benefit greatly from large and diverse datasets but in a clinical setting, data is not widely available and it takes along time to perform the scans and processing required to get it. Moreover datasets of brain images in structural segmentation process suffer from a inherent

class imbalance problem. This is easy to understand since most of the brain scan is background when the objective is to segment a white matter tract. To tackle this problem data augmentation is performed. For each image a random combination of transformations is performed just before the image is fed to the network during training. The value of the scalars used in each transformation is also determined randomly within the range present. The set of transformations used is listed in 4.2.5.

- Gaussian noise with mean 0 and std 0.05
- Displacement by $[-10, 10]$ voxels
- Rotation by $\theta = [-45, 45]$
- Brightness by a factor of $[0.7, 1.3]$
- Contrast by a factor of $[0.7, 1.3]$

When performing data augmentation, one option is to apply a set of transformations to all images in a dataset and save them, repeating this process for each desired transformation. This course of action effectively takes up a lot of disk space and might slow down learning. A better way is to create a generator, a small script that gets an image from disk, applies the desired data augmentation transformation and feeds it to the network during training. The generator implemented for this task has two modes: training and testing. During training, the generator fetches an image pair and the corresponding label, then it randomly selects a transformation from the list presented above 4.2.5 and applies it. Thus the data augmentation is performed for each image in each training step per epoch without occupying space in disk. During testing the generator does not apply any augmentation and feeds the data to the network as is.

In order to implement the above mentioned transformations, functions from the ndimage package of the SciPy[68] library were used for rotation and displacement and NumPy[69] for the other transformations as well as the random number generation. Rotation was applied to each channel of the image, in the case of the DECMAPS, individually; spline interpolation of order 1 was used as higher order interpolations cannot avoid overshooting.

4.3 Conclusion

In this chapter a detailed description of the data used in this work is presented alongside a review of the preprocessing steps taken to ensure data quality and cohesion across experiments. A summary of the methods used to correctly generate different image modalities presented in order to emphasize that a Deep learning model does not solely depend on the architecture or the initialization hyperparameters but also, and very significantly so, on the quality and quantity of data that is used to train it. The image data augmentation techniques used in this work are also explained since they are fundamental for the success of the learning process translating into better results

and performance enhancement for both training and testing. It is also important to say that all images of all datasets used in this work were first standardized to have zero mean and unit standard deviation as is common practice in all image analysis applications.

Chapter 5

Methodology, Experiments and Results

5.1 Network Architectures

In this work different network architectures and training schemes were tested and compared in order to achieve a deeper understanding and possibly better results for the white matter tract segmentation task. The network architectures and implementations are discussed below. All networks were developed and tested in a remote workstation NVIDIA DGX with 4 Teslas V100 (32GB) with Intel Xeon (40 cores) and 256GB of memory with Ubuntu operating system. Access was made via SSH client and was provided by this work's supervisor Professor João Tavares. Also all final segmentation results were obtained by thresholding the output of the network by 0.5. No cross validation was made due to timing restrictions since these are models that take each over 48h to train per network.

5.1.1 Baseline AgYnet

In [6] a V-net[4] is used in combination with attention mechanisms developed in [7]. Here the network is provided two sets of inputs, the T1 and the colour map representing the principal direction of diffusion[23]. These run in parallel network paths through an encoder network almost identical to [4] except the skip connections to the decoder network which are fused by an attention model developed in [7] that provides more weight to features that already exhibit high responses in previous convolution operations. The scheme of the network is presented in 5.1.

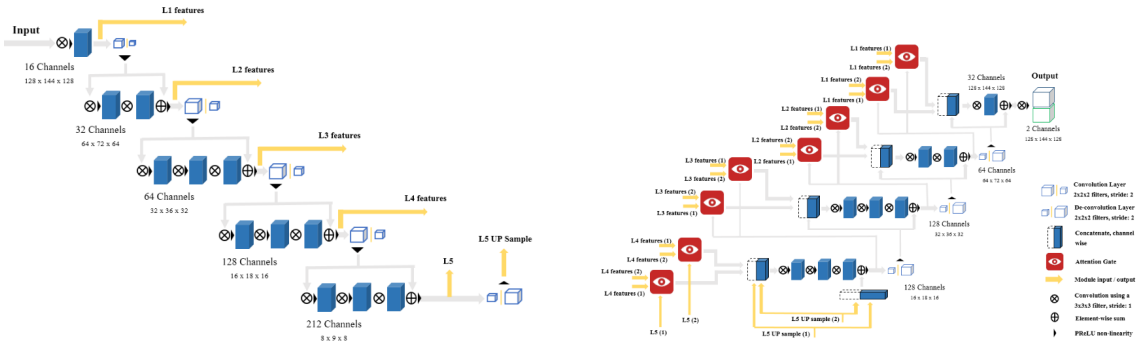


Figure 5.1: A schematic representation of the AgYnet from [6]

On the left side of 5.1 the scheme of the encoder network is presented. The convolution blocks were implemented using the layers API of Keras [70] using a kernel size of 3x3x3, a stride of 1 in each direction and zero padding when necessary. The number of filters used in each convolutional layer were [16,32,64,128,256] from top to bottom, values which were mirrored and inverted on the encoder path. Notice that there are two encoder pathways, one for the T1 modality and another for the DECMAPS, that are fused at the attention block which is discussed in5.1.1.1. The activation function chosen for this architecture was PReLU also implemented as a layer and was initialized at 0.25[71]. The model outputs a tensor with the same spacial dimentions as the input and only one channel where the value for each voxel is the result of the sigmoid activation function, a value between 0 and 1 which indicates the probability of that voxel to belong to the positive class.

The network introduced here was implemented in the tensorflow and keras frameworks and trained with the dataset provided in the challenge in order to establish a baseline for subsequent comparison to other network architectures.

5.1.1.1 Attention gate for two inputs

The idea for this attention mechanism first came forth in [7]. As discussed before in this work namely in **??** the attention gate is a mechanism that allows a neural network to scale the input image by the the highest response of the filters. This has two benefits: on one hand it allows the network to give higher values to voxels that have a high response to the filter kernels applied to it making important regions more salient; on the other it dampens regions that show little response to the filters in convolutional blocks. This translates to an output image that carries that is now easier to interpret further in the network since clear differences in regions of interest have been exacerbated. The scheme for the attention mechanism prosed in [7] is presented bellow in 5.2.

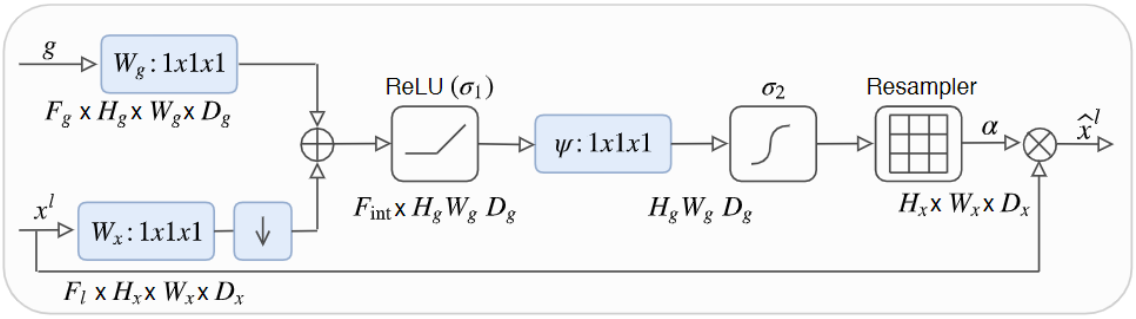


Figure 5.2: Scheme of the additive attention gate proposed in [7].

Input image (xl) is scaled with attention coefficients (α) computed in the gate. Spatial regions are selected by analyzing both the activations and contextual information provided by the gating signal (g) which is collected from a coarser scale. Image resampling before the output is done by trilinear interpolation[7]

This idea was modified to work as both a mechanism for making important regions more salient as well as way to fuse different modalities in [6]. As such a dual input attention gate was formulated as in 5.3.

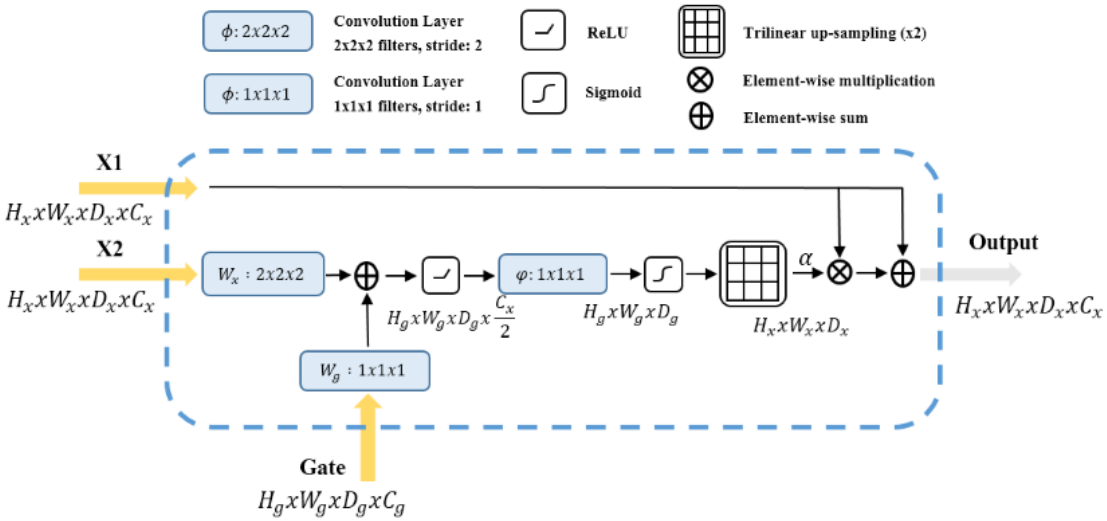


Figure 5.3: Attention gate scheme used in [6]

These redesigned attention gates are used in pairs in order to use one modality to gate the other and vice versa as represented in5.1.

5.1.1.2 Training scheme

For this experiment the network was trained once for each tract with the 46 cases from the BrainPTM2021 dataset and 10 cases for validation which was done once at the end of each epoch.Training was started with an initial learning rate of 1e-6 and ran for 120. Callbacks were

used to get only the best model and a monitor for the validation accuracy and mean Intersection over Union were implemented with a patience of 10 epochs. The batch size for this experiment was 1 since the volumes are relatively large: [128,144,128,1] for the T1 modality and [128,144,128,3] for the DECMAPS. As a loss function, the DICE loss function first presented in [4] was used as it theoretically seems more suited for a problem with so many voxels in background as opposed to binary cross-entropy loss.

5.1.1.3 Results

In this section the results for the training and validation of the AgYnet are presented. The first thing to note is that training took 15 seconds per step making the total training time per tract to be around 30 hours even with the callback system having stopped training at around 100 epochs for each tract due to the lack of evolution in the mean IoU score. Below, plots representing the DICE loss, accuracy and mean intersection over union are presented for the optic radiations right tract during training and validation. Only one is presented since they are all similar.

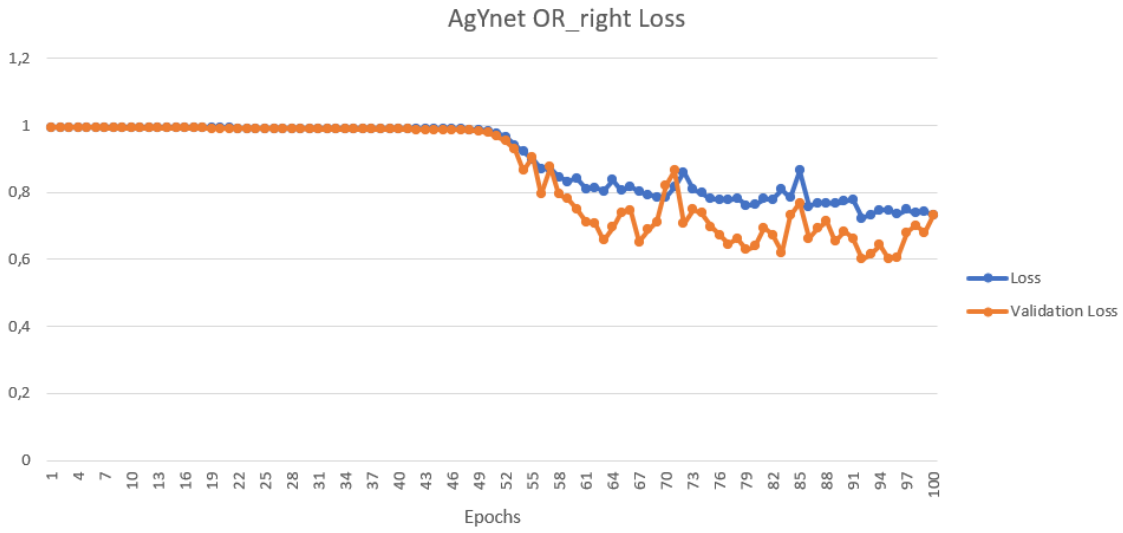


Figure 5.4: Plot of training and validation loss for the AgYnet model per epoch for the right Optical Radiations tract

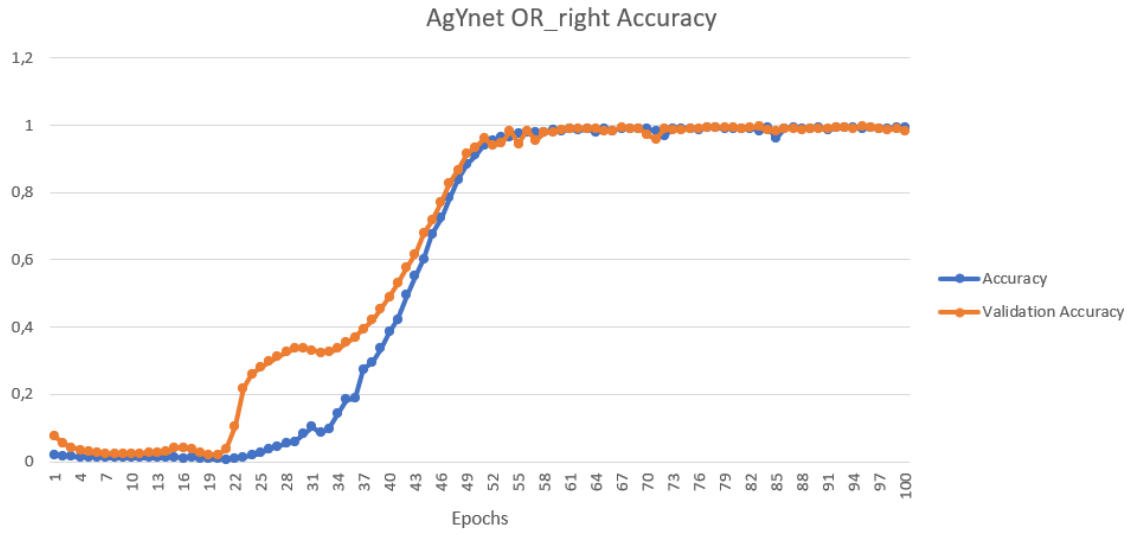


Figure 5.5: Plot of training and validation accuracy for the AgYnet model per epoch for the right Optical Radiations tract

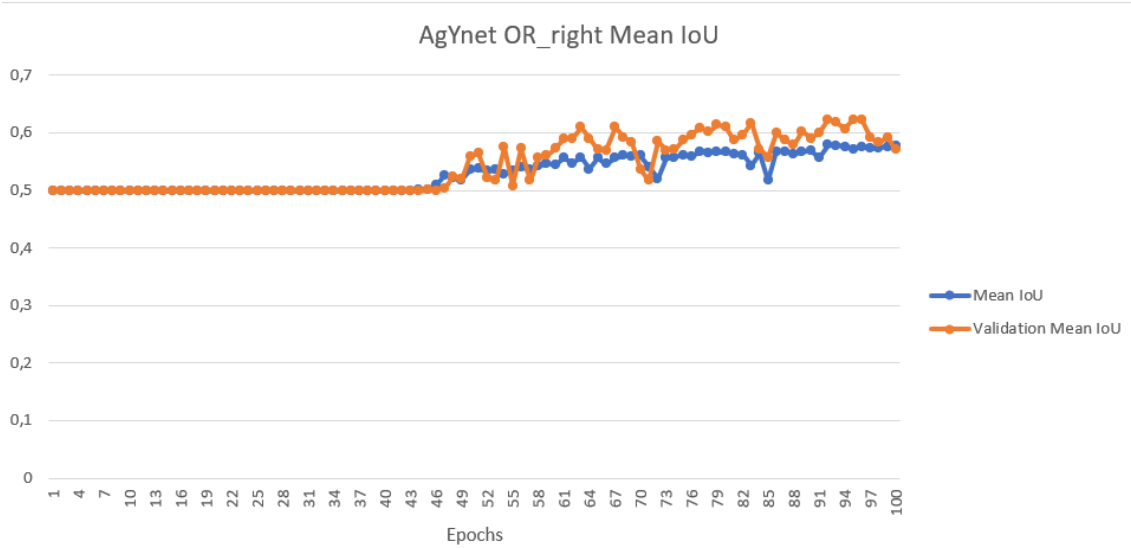


Figure 5.6: Plot of training and validation mean IoU for the AgYnet model per epoch for the right Optical Radiations tract

Regarding these plots it is possible to see a few patterns: accuracy initially falls and then goes to very high values very fast across a few epochs and then stabilizes. Loss stays very high throughout training and validation and decreases very slowly and never bellow 50%. Mean IoU stays almost flat throughout almost half of the training and then goes up just a little above the 60% mark but in a very unstable way.

Bellow a table indicating the DICE score and IoU of each of the 5 testing cases is presented for the OR right tract since the results for the other tracts are similar.

case	DICE	IOU
56	0.5723	0.5676
57	0.5864	0.5584
58	0.5344	0.5851
59	0.5435	0.5684
60	0.5372	0.5520

Table 5.1: Results for the AgYnet BrainPTM data per test case

5.1.1.4 Discussion

These results are unsatisfactory in comparison to the work made by the designers of AgYnet in [6]. The authors of that paper claim to have train the network on a larger dataset but the results reported above are underwhelming. This might be due to the fact that no cross validation was performed but the more probable aspect is weight initialization. As this is a random parameter in almost every machine learning study involving CNNs authors end up reporting the best results with very low reproducibility. As the authors of [6] did not make their implementation of the network public there is little way of directly comparing results. Despite this, the high instability of the network during training may indicate that this architecture would benefit from some batch normalization layers before activation functions. These not only help to stabilize the learning of the network by regularizing the range of values during optimizer computations but also help to achieve reproducible results.

5.1.2 Nested AGYnet

In this section the implementation and training and results of a variation of AGYnet is discussed. UNET++ [8] is a variation of the original Unet design with the main feature being the redesign of skip pathways and the possibility of deep supervision. In the original Unet architecture feature maps outputted by the descending part of the network (encoder) are directly received on the ascending path(decoder); in contrast, with this architecture, they first must go through a dense convolution block whose number of convolutional layers depends on the level of the layer which provides them. For instance, the skip pathway of the first layer starts is concatenated with the upsampled output of the second layer and convolved. Then it is concatenated with the output of the first convolutional nested block of the second layer which is the result of the same process between the second- and third-layer's convolutional block. This happens again and again until all convolutional blocks in the encoder are able to contribute to the last convolutional block in the first skip connection as described in 5.7.

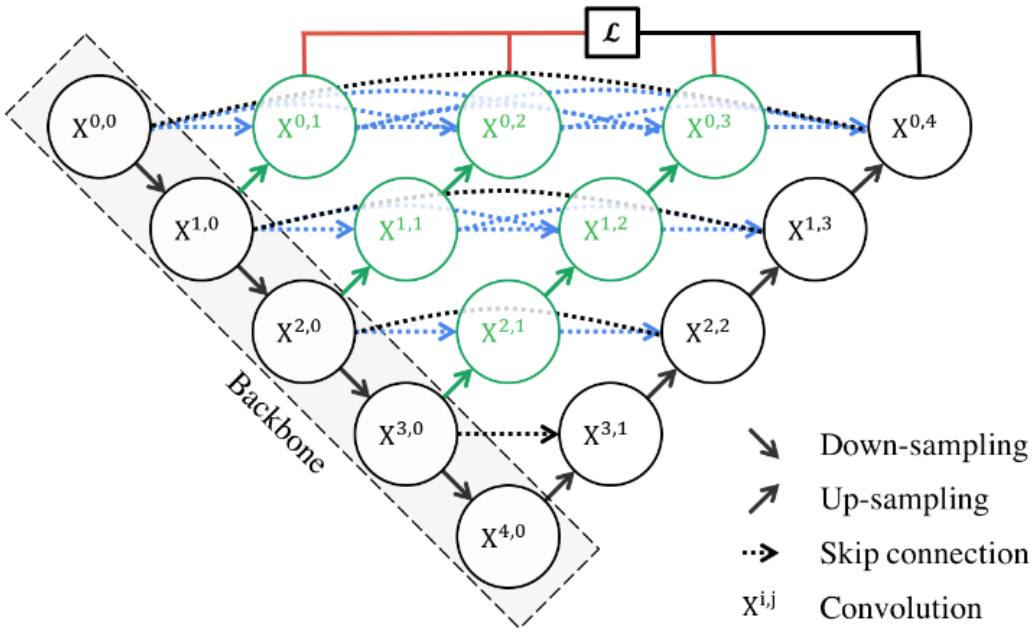


Figure 5.7: A scheme of the concept of the nested [8] network

In essence, the dense convolution block is intended to bring the semantic level of the encoder feature maps (filter responses). Essentially, the dense convolution block brings the semantic level of the encoder feature maps closer to that of the feature maps awaiting in the decoder. The hypothesis is that the optimizer would face an easier optimization problem when the received encoder feature maps and the corresponding decoder feature maps are semantically similar[8]. The idea Is that in this way the optimizer would face and easier problem, computation wise, if the outputs of the encoder are in a similar semantic level as the upsampled inputs of a decoder layer. Adding to this, the attention gate mechanism is still employed as the last part of this dense convolutional block and as a consequence still preforms the fusion between the modalities. The Unet++ architecture also proposes a way to perform deep supervision by outputting images at the end of each level (layer) of dense convolutional block either averaging the results for the final prediction or selecting one of these outputs for prediction. Here the idea is to iteratively prune the network in order to achieve better performance. It is important to understand that this model was designed for 2D image segmentation so it had to be modified to the 3D implementation that this work is focused on. Other modifications were the abandoning of max-pooling layers using instead convolutional layers with stride 2 as was the case in the previous network discussed. Activation functions were also changed to be PReLU.

5.1.2.1 Training scheme

For this model the training scheme is analogous to the on 5.1.1.2. The concept of nesting was applied to the AgYnet, as such this was not a direct application of [8] but more of an expansion of the AgYnet architecture. Unet++ was implented on 2D images, the model here was implemented

for 3D. Another difference is that there are no max pooling operations but just convolutional layers with stride 2.

5.1.2.2 Results

The plots for the training, validation and accuracy curves are shown bellow .

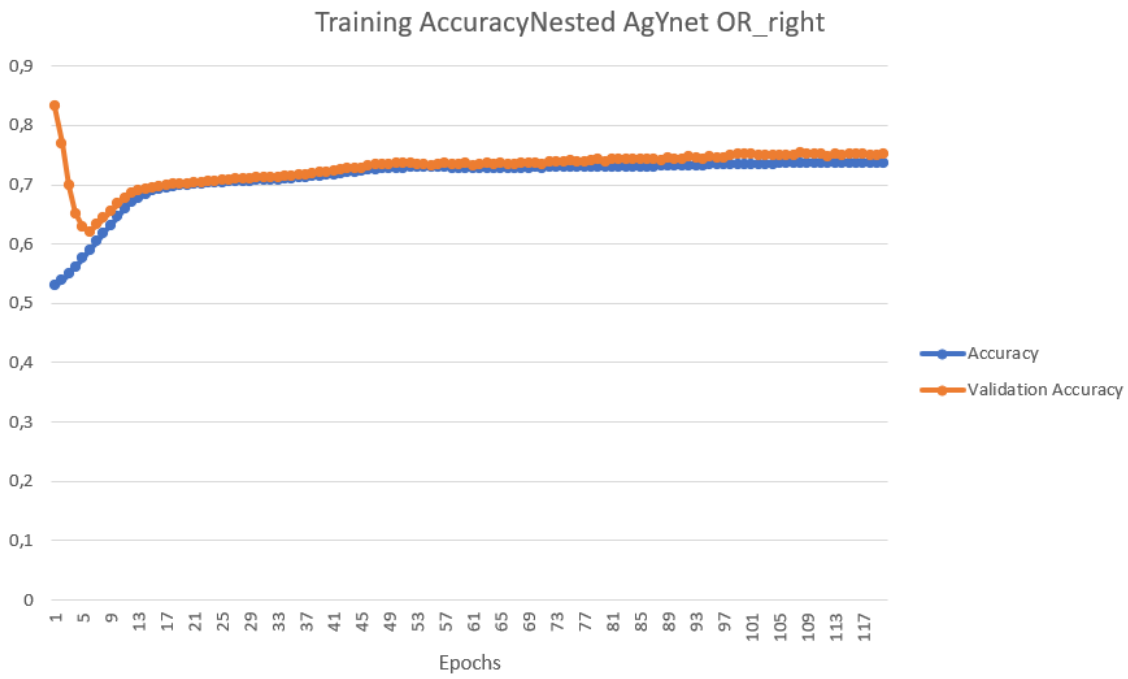


Figure 5.8: Plot of training and validation loss for the nested AgYnet model per epoch for the right Optical Radiations tract

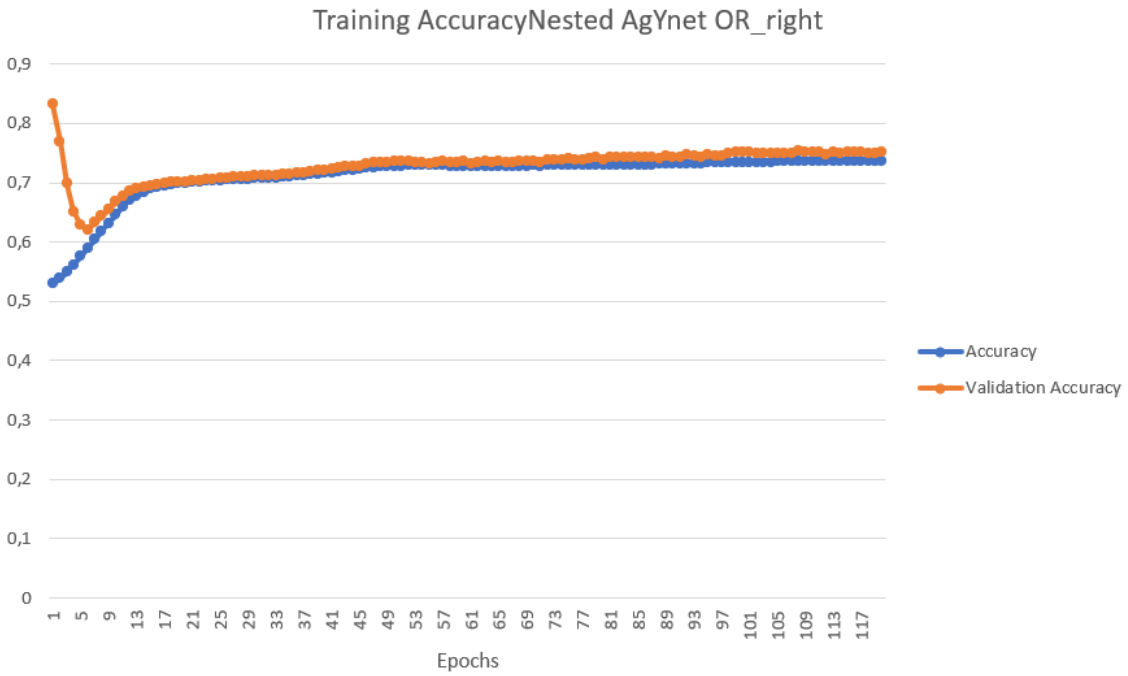


Figure 5.9: Plot of training and validation accuracy for the nested AgYnet model per epoch for the right Optical Radiations tract

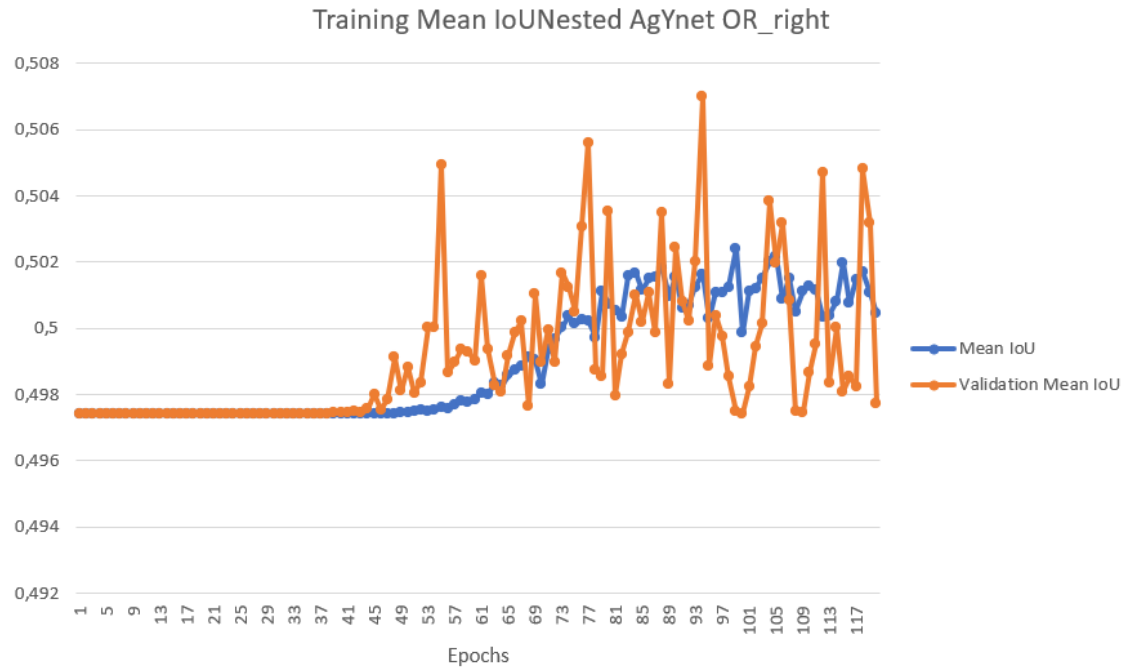


Figure 5.10: Plot of training and validation mean IoU for the nested AgYnet model per epoch for the right Optical Radiations tract

Bellow a table indicating the DICE score and IoU of each of the 5 testing cases is presented.

case	DICE	IOU
56	0.5379	0.5375
57	0.5154	0.5244
58	0.5054	0.5143
59	0.5263	0.5471
60	0.5185	0.5273

Table 5.2: Results for the nested AgYnet data per test case

5.1.2.3 Discussion

A few things are of note in this experiment. Firstly the use of the nested architecture dramatically increased the training time for each epoch from 15 seconds in the original AgYnet model to 50 seconds here. This is due to the fact that in the work where UNET++ [8] is presented, the authors designed it for an unimodal application. In this case however the inputs are multi modal as are the branches of the encoder block. A reason for the downgrade in both DICE score and IoU is unclear as further experimentation would be needed but with this huge amount of time requiered for training the architecture was abandoned.

5.1.3 Transfer Learning - HCP to BrainPTM

In order to improve upon the results given by the AgYnet architecture one option considered was the use of more data in training. As such the HCP dataset was used in order to pre-train the AgYnet in order to investigate how this would affect the output of the network. The model used for this experiment was the version of AgYnet put forth in 5.1.1 and shown in 5.1. The expectation was that with pretraining being made with more data and mainly with better quality data as described in 4.1.3, the same architecture would display better results in segmentation of the white matter tracts.

5.1.3.1 Training scheme

Firstly the AgYnet was trained on the HCP dataset for 120 epochs with 86 steps per epoch, a learning rate of 1×10^{-6}, a batch size of 1 and the DICE loss function. Callbacks were implemented in the same way as described in the previous experiments. The dataset was comprised of 86 training cases, and 18 cases for validation and then retrained for the BrainPTM dataset using 46 cases for training, 10 for validation and 5 for testing for another 50 epochs to avoid overfitting and the same hyperparameters. Results are reported in the 5 cases held out for testing from the BrainPTM2021 dataset.

5.1.3.2 Results

Bellow the plots for the training loss, accuracy and mean IoU are presented.

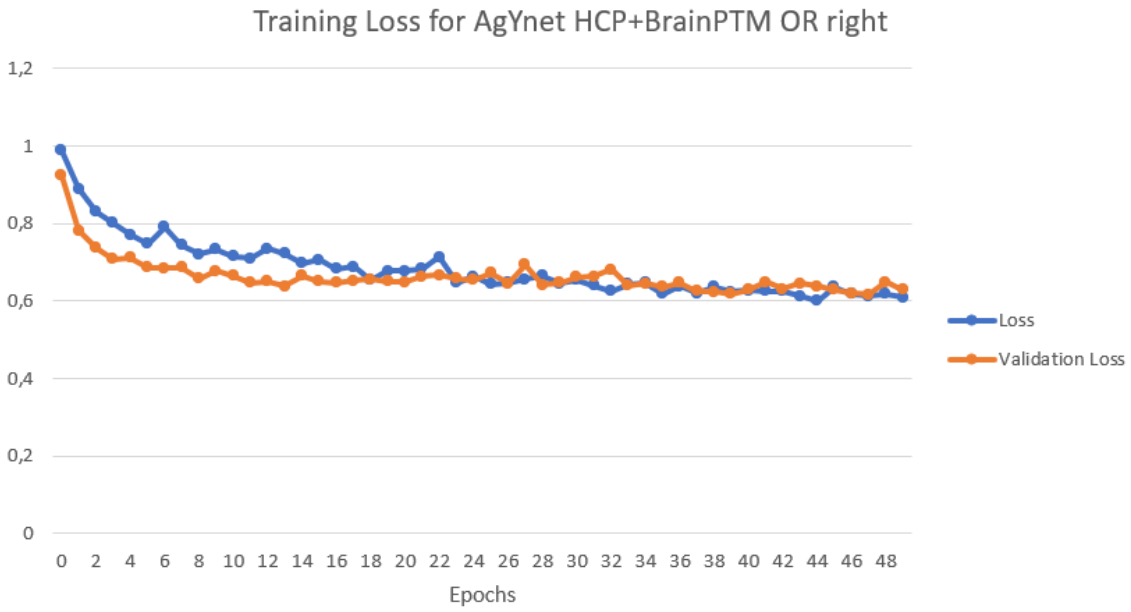


Figure 5.11: Plot of training and validation loss for the AgYnet model pretrained with HCPdata per epoch for the right Optical Radiations tract

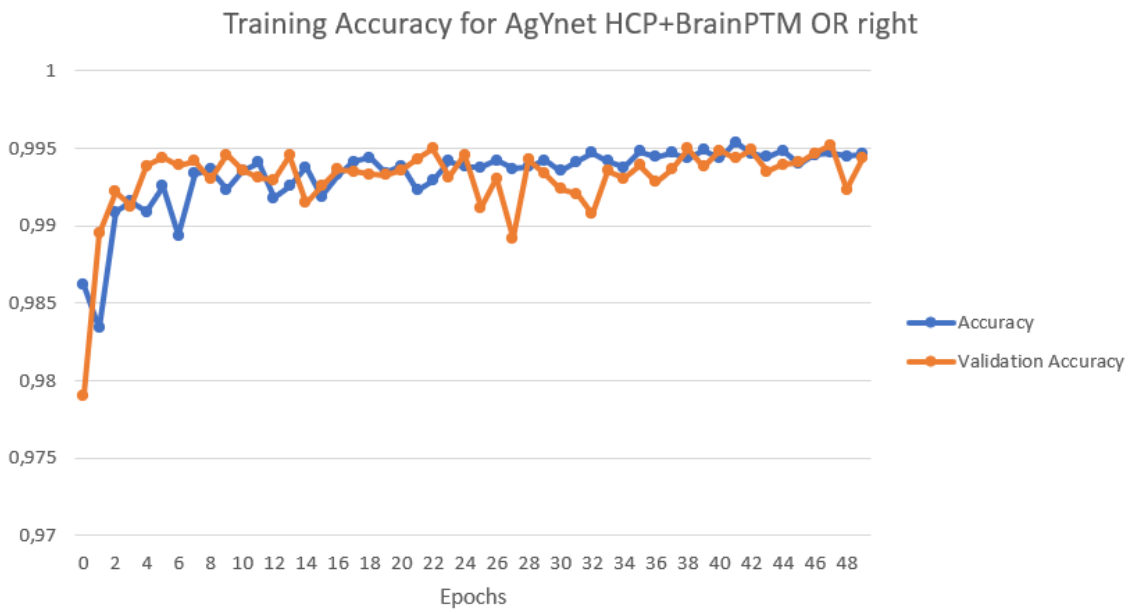


Figure 5.12: Plot of training and validation accuracy for the AgYnet model pretrained with HCP-data per epoch for the right Optical Radiations tract

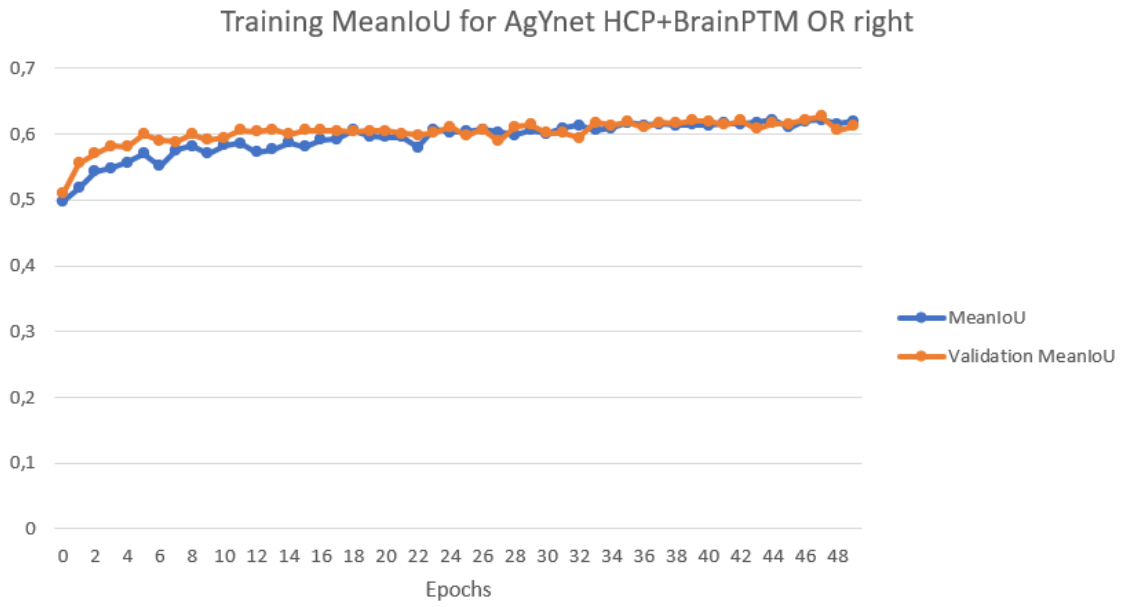


Figure 5.13: Plot of training and validation mean IoU for the AgYnet model pretrained with HCPdata per epoch for the right Optical Radiations tract

Bellow a table indicating the DICE score and IoU of each of the 5 testing cases is presented for the final training scheme.

case	DICE	IOU
56	0.6543	0.6423
57	0.6744	0.6532
58	0.6566	0.6325
59	0.6622	0.6547
60	0.6421	0.6366

Table 5.3: Results for the final training scheme AgYnet+ HCP+BrainPTM data per test case

Finally the mean DICE score and mean IoU score are presented below As well as the final segmentation for all four tracts.

Tract	Mean DICE	Mean IOU
OR_right	0,6579	0.6438
OR_left	0.6487	0.6512
CST_right	0.6372	0.6233
CST_left	0.6163	0.6243

Table 5.4: Results for the final training scheme AgYnet+ HCP+BrainPTM data per tract

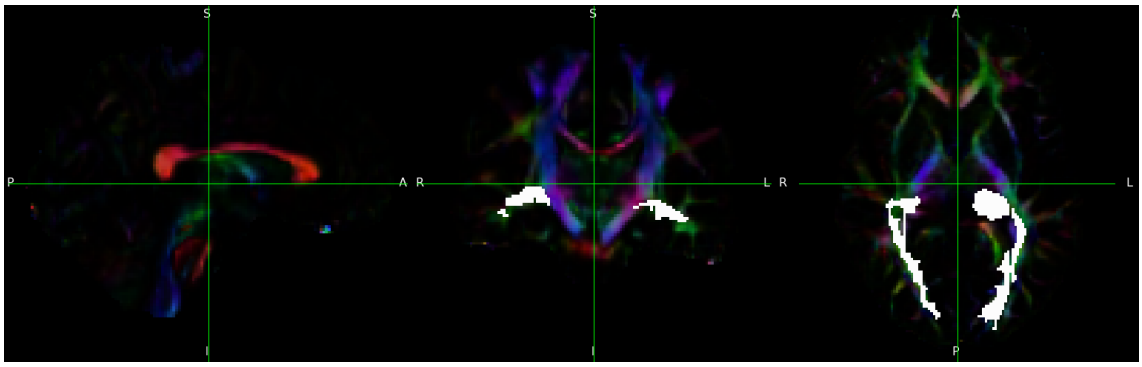


Figure 5.14: Best segmentation obtained for the Optical Radiations tract right and left over the corresponding dti image (case 60)

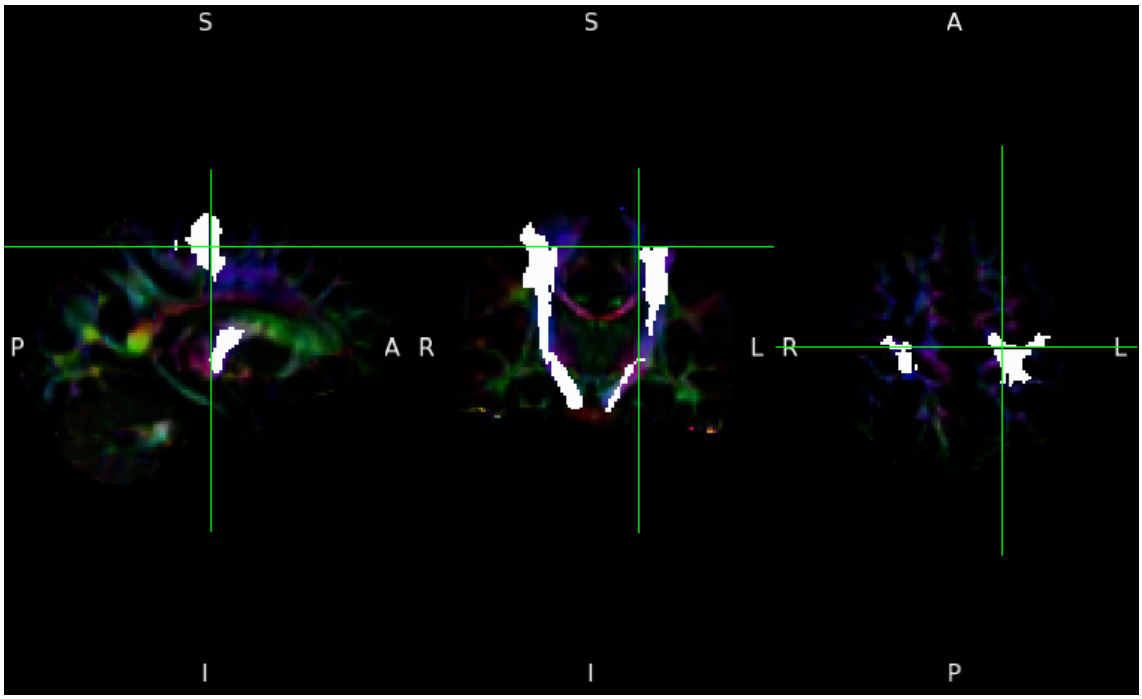


Figure 5.15: Best segmentation obtained for the Corticospinal tract right and left over the corresponding dti image (case 60)

5.1.3.3 Discussion

As expected the pretraining improved the overall performance of the model, however, the results are still very underwhelming when compared to the ones reported on [6]. Instability in learning is still very present throughout the network which could have countered by the use of batch normalization layers. Such was not done due to the increase in training time of 4 seconds. Another main difference between this work and [6] is that the results were not reported on HCP images but on

the lower quality BrainPTM images. This was motivated by the will to see how the architecture performs in the clinical setting as most of the works done in this field utilize images that are of very high quality with very computationally intensive DTI algorithms like CSD in [5].

Chapter 6

Conclusion

6.1 Conclusion & Future work

This dissertation attempts to investigate the difficult topic of white matter segmentation using deep learning techniques in the clinical setting. In this manner there is an attempt to incentivise the testing and procurement of new models and architectures that present incremental improvements to the current clinical practice and, of course, the state of the art in brain image segmentation. While it is true that currently it is possible to visualize whole brain tractograms it is important to make the reminder that despite these images being highly detailed they do not correspond to the truth. These tractograms are probabilistic algorithms that need to be seeded, as such, they are not a one to one correspondence with what is actually happening in the human brain. To attain the real picture, the best bet currently are deep learning techniques that are no longer in their infancy but are still extremely varied and sometimes make up an uncohesive tissue. Nowadays it seems, this technology is moving ever more heavily in the direction of multi-modal approaches reflecting the actual clinical practice: no one exam tells the whole story. In this vein these approaches are considered innovative and are very challenging to implement and make work since they require a holistic view of the diagnostical process from the selection of imaging modalities, to their pre-processing, to the classification or in this case segmentation. This subject is very current and necessary for the clinical practice as time being saved in the decision process is time gained by the patient in life expectancy. As such an attempt was made to implement, train and discuss the aspects required for a truly automatic mapping of the human brain with a focus on white matter tracts that are of special importance for presurgical evaluation: Optic radiations and Corticospinal tract. Results were underwhelming but confirmed the expectation that better data yields better performance. A huge number of architectures remain to be evaluated and tested, if deadlines did not exist, all the architectures mentioned should have been subjected to cross-validation in order to access the capacity for generalization of the model. With better generalization come better results when it comes to dealing with different datasets. The monograph preceding this work had plans for the utilization of Tranformers for use as encoders V-net like architectures but such was not possible due to time restricitons.Another possible avenue of exploration could be the use of

atrous and dilated convolutions where a network like AgYnet could have a high resolution and low resolution pathway. Despite this work comprises a wholesome view of the tools and methods used in the neuroimaging field specifically centered around diffusion tensor imaging at its application as a tool to generate segmentation maps of white matter tracts in the brain. As such it could be used as a starting point for research in the future. Finalizing this set of concluding remarks, it should be highlighted that the efforts of this thesis aim to make a valuable contribution to the field of neuroimaging and more specifically to its applicability to healthcare services. This involvement is placed within the scope of the defense of deep learning technology and its use as a second opinion or a start up point for healthcare diagnosis that could be available readily and speedly while clinical practitioners practice their vocation to take are of others.

The dissertation proposes to investigate and prospect multimodal solutions (textual and visual) for generating explanations in a medical or clinical setting. In this manner, there is an attempt to promote the complementarity of explanations and increase diversity and, consequently, the likelihood of acceptance by stakeholders. To attain such a proposal, two aspects are required: guarantee the feasibility of unimodal tasks, where both visual and textual explainability techniques can minimally fulfill what is expected of them, being coherent and clearly interpretable; ensure forms of multimodal interaction in which the sharing and/or transfer of knowledge between multimodal approaches is maximized with a perspective of achieving quality compliance for the explanations. Considering the diagnostic domains or assignments where visual evidence is normally relied upon, the relevance of establishing methodologies for merging visual saliency maps with semantic concepts incorporated in textual explanations is substantial. In this vein, the fundamentally innovative aspect is present in the multiple creations of explanations. That is, granted the example of a chest X-ray, it should be possible to obtain a decision supported by the demonstration of the pixels considered in it, and by a free-text radiology report style description that complements the previous explanation.

References

[1] P. Dvorak and B. Menze, "Structured prediction with convolutional neural networks for multimodal brain tumor segmentation," pp. 13–24, 10 2015. Cited on pages ix, 21, and 22.

[2] 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. Cited on pages ix, 23, and 24.

[3] 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," *Medical image analysis*, vol. 36, pp. 61–78, 2017. Cited on pages ix, 25, and 26.

[4] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*, pp. 565–571, IEEE, 2016. Cited on pages ix, 27, 28, 29, 41, and 44.

[5] J. Wasserthal, P. Neher, and K. H. Maier-Hein, "Tractseg-fast and accurate white matter tract segmentation," *NeuroImage*, vol. 183, pp. 239–253, 2018. Cited on pages ix, 28, 29, 34, 36, and 54.

[6] I. Nelkenbaum, G. Tsarfaty, N. Kiryati, E. Konen, and A. Mayer, "Automatic segmentation of white matter tracts using multiple brain mri sequences," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pp. 368–371, IEEE, 2020. Cited on pages ix, 29, 41, 42, 43, 46, and 53.

[7] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: Learning to leverage salient regions in medical images," *Medical image analysis*, vol. 53, pp. 197–207, 2019. Cited on pages ix, 29, 41, 42, and 43.

[8] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pp. 3–11, Springer, 2018. Cited on pages ix, 46, 47, and 50.

[9] R. R. Seeley, T. D. Stephens, P. Tate, C. M. Eckel, and J. L. Regan, "Anatomy Physiology Eighth Edition," *McGraw-Hill Companies Inc.*, p. 1266, 2008. Cited on pages 3, 4, 5, 6, 7, 8, and 9.

[10] C. R. Noback, N. L. Strominger, R. J. Demarest, and D. A. Ruggiero, *The human nervous system: Structure and function: Sixth edition*. 2005. Cited on pages 4, 7, and 8.

[11] V. H. Perry and J. Teeling, "Microglia and macrophages of the central nervous system: The contribution of microglia priming and systemic inflammation to chronic neurodegeneration," *Seminars in Immunopathology*, vol. 35, no. 5, pp. 601–612, 2013. Cited on page 6.

[12] W. J. Streit, M. B. Graeber, and G. W. Kreutzberg, "Functional plasticity of microglia: A review," *Glia*, vol. 1, no. 5, pp. 301–307, 1988. Cited on page 6.

[13] R. R. Ji, C. R. Donnelly, and M. Nedergaard, "Astrocytes in chronic pain and itch," *Nature Reviews Neuroscience*, vol. 20, no. 11, pp. 667–685, 2019. Cited on page 6.

[14] M. Peckham, A. Knibbs, and S. Paxton, "The Histology Guide," 2003. Cited on page 8.

[15] B. J. Jellison, A. S. Field, J. Medow, M. Lazar, M. S. Salamat, and A. L. Alexander, "Diffusion tensor imaging of cerebral white matter: a pictorial review of physics, fiber tract anatomy, and tumor imaging patterns," *American Journal of Neuroradiology*, vol. 25, no. 3, pp. 356–369, 2004. Cited on pages 9 and 10.

[16] O. Grad, M. A. Moskowitz, *et al.*, "Magnetic resonance imaging: overview of the technology and medical applications," *International Journal of Technology Assessment in Health Care*, vol. 1, no. 3, pp. 481–498, 1985. Cited on page 10.

[17] D. W. McRobbie, E. A. Moore, M. J. Graves, and M. R. Prince, *MRI from Picture to Proton*. Cambridge university press, 2017. Cited on pages 10 and 11.

[18] D. I. Hoult and B. Bhakar, "Nmr signal reception: Virtual photons and coherent spontaneous emission," *Concepts in Magnetic Resonance: An Educational Journal*, vol. 9, no. 5, pp. 277–297, 1997. Cited on page 10.

[19] M. T. Vlaardingerbroek and J. A. Boer, *Magnetic resonance imaging: theory and practice*. Springer Science & Business Media, 2013. Cited on page 11.

[20] X. Liu, M. Kinoshita, H. Shinohara, O. Hori, N. Ozaki, and M. Nakada, "A fiber dissection study of the anterior commissure: Correlations with diffusion spectrum imaging tractography and clinical relevance in gliomas," *Brain topography*, pp. 1–9, 2021. Cited on pages 11, 12, and 13.

[21] P. Mörters and Y. Peres, *Brownian motion*, vol. 30. Cambridge University Press, 2010. Cited on page 12.

[22] S. Mori and J. Zhang, "Principles of diffusion tensor imaging and its applications to basic neuroscience research," *Neuron*, vol. 51, no. 5, pp. 527–539, 2006. Cited on page 13.

[23] S. Pajevic and C. Pierpaoli, "Color schemes to represent the orientation of anisotropic tissues from diffusion tensor data: application to white matter fiber tract mapping in the human brain," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 42, no. 3, pp. 526–540, 1999. Cited on pages 14, 29, 36, and 41.

[24] P. J. Basser, J. Mattiello, and D. LeBihan, "Mr diffusion tensor spectroscopy and imaging," *Biophysical journal*, vol. 66, no. 1, pp. 259–267, 1994. Cited on pages 14 and 36.

[25] B. Jeurissen, J.-D. Tournier, T. Dhollander, A. Connelly, and J. Sijbers, "Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion mri data," *NeuroImage*, vol. 103, pp. 411–426, 2014. Cited on page 15.

[26] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *The Journal of physiology*, vol. 148, no. 3, p. 574, 1959. Cited on page 15.

[27] K. Fukushima and S. Miyake, "Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition," in *Competition and cooperation in neural nets*, pp. 267–285, Springer, 1982. Cited on page 15.

[28] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. Cited on page 15.

[29] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, pp. 1–6, 2017. Cited on pages 15, 16, and 17.

[30] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on mri," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 102–127, 2019. Cited on page 17.

[31] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*, pp. 448–456, PMLR, 2015. Cited on pages 17 and 26.

[32] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big data*, vol. 3, no. 1, pp. 1–40, 2016. Cited on page 17.

[33] M. W. Woolrich, S. Jbabdi, B. Patenaude, M. Chappell, S. Makni, T. Behrens, C. Beckmann, M. Jenkinson, and S. M. Smith, "Bayesian analysis of neuroimaging data in fsl," *Neuroimage*, vol. 45, no. 1, pp. S173–S186, 2009. Cited on page 17.

[34] M. Jenkinson, M. Pechaud, S. Smith, *et al.*, "Bet2: Mr-based estimation of brain, skull and scalp surfaces," in *Eleventh annual meeting of the organization for human brain mapping*, vol. 17, p. 167, Toronto., 2005. Cited on pages 17, 32, and 36.

[35] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, "Improved optimization for the robust and accurate linear registration and motion correction of brain images," *Neuroimage*, vol. 17, no. 2, pp. 825–841, 2002. Cited on pages 17 and 36.

[36] M. Jenkinson and S. Smith, "A global optimisation method for robust affine registration of brain images," *Medical image analysis*, vol. 5, no. 2, pp. 143–156, 2001. Cited on pages 17 and 36.

[37] D. N. Greve and B. Fischl, "Accurate and robust brain image alignment using boundary-based registration," *Neuroimage*, vol. 48, no. 1, pp. 63–72, 2009. Cited on pages 17 and 36.

[38] M. Maleki, M. Teshnehlab, and M. Nabavi, "Diagnosis of multiple sclerosis (ms) using convolutional neural network (cnn) from mris," *Global Journal of Medicinal Plant Research*, vol. 1, no. 1, pp. 50–54, 2012. Cited on page 19.

[39] D. Zikic, Y. Ioannou, M. Brown, and A. Criminisi, "Segmentation of brain tumor tissues with convolutional neural networks," *Proceedings MICCAI-BRATS*, vol. 36, no. 2014, pp. 36–39, 2014. Cited on pages 20 and 21.

[40] 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. Cited on page 20.

[41] D. Zikic, B. Glocker, E. Konukoglu, J. Shotton, A. Criminisi, D. Ye, C. Demiralp, O. Thomas, T. Das, R. Jena, *et al.*, "Context-sensitive classification forests for segmentation of brain tumor tissues," *Proc. MICCAI-BRATS*, pp. 22–30, 2012. Cited on page 21.

[42] 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. Cited on page 21.

[43] H. Chen, L. Yu, Q. Dou, L. Shi, V. C. Mok, and P. A. Heng, "Automatic detection of cerebral microbleeds via deep learning based 3d feature representation," in *2015 IEEE 12th international symposium on biomedical imaging (ISBI)*, pp. 764–767, IEEE, 2015. Cited on pages 22 and 23.

[44] M.-P. Dubuisson and A. K. Jain, "A modified hausdorff distance for object matching," in *Proceedings of 12th international conference on pattern recognition*, vol. 1, pp. 566–568, IEEE, 1994. Cited on page 24.

[45] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," *Advances in neural information processing systems*, vol. 24, 2011. Cited on pages 24 and 26.

[46] C. Wachinger, M. Reuter, and T. Klein, "Deepnat: Deep convolutional neural network for segmenting neuroanatomy," *NeuroImage*, vol. 170, pp. 434–445, 2018. Cited on pages 24 and 25.

[47] H. Lombaert, J. Sporring, and K. Siddiqi, "Diffeomorphic spectral matching of cortical surfaces," in *International Conference on Information Processing in Medical Imaging*, pp. 376–389, Springer, 2013. Cited on page 24.

[48] C. Wachinger, P. Golland, W. Kremen, B. Fischl, M. Reuter, A. D. N. Initiative, *et al.*, "Brainprint: A discriminative characterization of brain morphology," *NeuroImage*, vol. 109, pp. 232–248, 2015. Cited on page 24.

[49] B. Fischl, "Freesurfer," *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012. Cited on page 25.

[50] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation," *IEEE transactions on medical imaging*, vol. 23, no. 7, pp. 903–921, 2004. Cited on page 25.

[51] 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. Cited on page 25.

[52] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014. Cited on page 25.

[53] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *arXiv preprint arXiv:1312.6229*, 2013. Cited on page 25.

[54] S. Valverde, M. Cabezas, E. Roura, S. González-Villà, D. Pareto, J. C. Vilanova, L. Ramió-Torrentà, À. Rovira, A. Oliver, and X. Lladó, "Improving automated multiple sclerosis lesion segmentation with a cascaded 3d convolutional neural network approach," *NeuroImage*, vol. 155, pp. 159–168, 2017. Cited on pages 26 and 27.

[55] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016. Cited on page 26.

[56] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014. Cited on page 26.

[57] 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. Cited on pages 27 and 28.

[58] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017. Cited on page 29.

[59] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021. Cited on page 29.

[60] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009. Cited on page 29.

[61] "Neuroimaging informatics technology initiative." https://nifti.nimh.nih.gov/. Accessed: 2022-06-30. Cited on page 31.

[62] P. Deutsch, "Deflate compressed data format specification version 1.3," tech. rep., 1996. Cited on page 32.

[63] I. Avital, I. Nelkenbaum, G. Tsarfaty, E. Konen, N. Kiryati, and A. Mayer, "Neural segmentation of seeding rois (srois) for pre-surgical brain tractography," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1655–1667, 2019. Cited on page 32.

[64] Cited on page 32.

[65] W. Jakob, N. Peter, and M.-H. Klaus, "High quality white matter reference tracts," Nov 2018. Cited on pages 34 and 36.

[66] M. F. Glasser, S. N. Sotiropoulos, J. A. Wilson, T. S. Coalson, B. Fischl, J. L. Andersson, J. Xu, S. Jbabdi, M. Webster, J. R. Polimeni, *et al.*, "The minimal preprocessing pipelines for the human connectome project," *Neuroimage*, vol. 80, pp. 105–124, 2013. Cited on page 34.

[67] E. Garyfallidis, M. Brett, B. Amirbekian, A. Rokem, S. Van Der Walt, M. Descoteaux, I. Nimmo-Smith, and D. Contributors, "Dipy, a library for the analysis of diffusion mri data," *Frontiers in neuroinformatics*, vol. 8, p. 8, 2014. Cited on page 36.

[68] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020. Cited on page 38.

[69] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, pp. 357–362, Sept. 2020. Cited on page 38.

[70] F. Chollet *et al.*, "Keras," 2015. Cited on page 42.

[71] L. Datta, "A survey on activation functions and their relation with xavier and he normal initialization," *arXiv preprint arXiv:2004.06632*, 2020. Cited on page 42.