جامعة مولاي إسماعيل
UNIVERSITÉ MOULAY ISMAÏL

كلية العلوم
FACULTÉ DES SCIENCES

End-Of-Studies Project

Speciality : Artificial Intelligence

Presented By

OUBRAYME MUSTAFA

Framed by

Mr. El Mehdi ISMAILI ALAOUI

On the theme:

DEEP NEURAL NETWORK BASED APPROACH FOR 3D BRAIN TUMOR REGISTRATION AND SEGMENTATION

Defended on September 30, 2023 in front of the jury:

Mr. Nom et PrÃ©nom	FacultÃ© des Sciences, MeknÃ¨s
Mr. Nom et PrÃ©nom	FacultÃ© des Sciences, MeknÃ¨s
Mr. Nom et PrÃ©nom	FacultÃ© des Sciences, MeknÃ¨s

Faculty of Sciences, B.P. 11201 Zitoune Meknes, Morocco.

Tel: +212 5 35 53 73 21, Fax: +212 5 35 53 68 08, http://www.fs-umi.ac.ma.

Acknowledgments

I would like to express my deepest gratitude to my supervisor, Mr. El Mehdi Ismaili Alaoui, for his invaluable guidance, unwavering support, and mentorship throughout this research endeavor. His expertise and dedication have been instrumental in shaping the direction of this work.

I am also thankful to my esteemed professors, Mr. Moulay Ali Bekri and Mr. Ali Oubelckacem, for their insightful feedback, encouragement, and scholarly advice, which have enriched my academic journey. Their commitment to fostering intellectual growth has been a constant source of inspiration.

Last but not least, I would like to acknowledge the support and understanding of my family and friends, whose unwavering encouragement and patience have been my pillars of strength.

This work would not have been possible without the collective effort and encouragement of all those mentioned above. Thank you for your invaluable contributions to this endeavor.

Table of Contents

List of Figures

CHAPTER 1

INTRODUCTION

1. 1 Project Overview

In the realm of medical imaging and diagnosis, the project at hand holds significant promise in advancing our understanding and treatment of brain tumors, particularly gliomas. Brain tumors, including the aggressive forms, remain among the most challenging and perilous forms of cancer worldwide. Accurate and timely diagnosis is paramount, and this project emerges at the intersection of cutting-edge technology and clinical need. Leveraging the power of deep learning and artificial intelligence, the project seeks to revolutionize the way we analyze and interpret brain tumor images, with a particular focus on Magnetic Resonance Imaging (MRI). Through the automated processes of image registration, which aligns images from different sources or time points, and segmentation, which delineates tumor boundaries, the project aims to enhance the precision and efficiency of diagnosis and treatment planning. This approach offers a unique opportunity to improve patient outcomes by providing clinicians with a more accurate and comprehensive understanding of brain tumor characteristics and evolution. The synergy between state-of-the-art AI techniques, advanced medical imaging, and the vital role of registration forms the foundation of this transformative research endeavor.

1. 2 The Problem Statement

1. 2. 1 MRI-Based Image Registration

Image registration, specifically within the context of MRI scans, is the process of aligning multiple MRI images taken at different times or from different imaging modalities. It involves ensuring that these images are not only aligned spatially but also coherently. This alignment is essential for tracking the evolution of brain tumors over time, planning treatments, and assessing their effectiveness.

The challenge in MRI-based image registration lies in the inherent variability in MRI scans, caused by

differences in patient positioning, imaging protocols, and scanner settings. Furthermore, brain tissue's deformability and the dynamic nature of tumors make the registration task particularly complex.

1. 2. 2 MRI-Based Image Segmentation

MRI-based image segmentation is the task of precisely delineating regions of interest within MRI scans, such as the boundaries of brain tumors. Accurate segmentation plays a pivotal role in treatment planning, as it allows for the quantification of tumor size, tracking changes, and making informed decisions regarding therapy. Segmenting brain tumors from MRI scans is a complex endeavor due to the variability in tumor shapes, intensities, and appearances in different MRI sequences. Additionally, MRI scans can contain noise and artifacts, further complicating the segmentation process.
As a consequence, the successful registration (aligning images from different sources or time points) and segmentation (delineating tumor boundaries) using advanced techniques can significantly enhance the accuracy and efficiency of diagnosis and treatment planning.

1. 3 Motivation

The motivation driving this project stems from the pressing need to address the complexities and challenges posed by brain tumors, particularly gliomas, in the field of medical imaging and diagnosis. Brain tumors, with their various malignancy grades and histological diversity, continue to exact a significant toll on patients' lives. The imperative to improve early detection and accurate characterization of these tumors is underscored by their formidable impact. With the emergence of deep learning and artificial intelligence, there exists an unprecedented opportunity to revolutionize the way we approach brain tumor analysis. The traditional methods of manual segmentation and interpretation of MRI images, often time-consuming and reliant on experienced neuroradiologists, call for innovative solutions. By harnessing the power of image registration and segmentation, this project aims to streamline and enhance the diagnostic process. Through automation and advanced algorithms, it seeks to empower medical professionals with precise, timely, and comprehensive insights into brain tumor morphology and evolution. The ultimate motivation lies in contributing to improved patient care and outcomes by providing clinicians with cutting-edge tools to navigate the challenges posed by brain tumors.

1. 4 Goals

The overarching goals of this project are multifaceted and driven by the imperative to advance our understanding and treatment of brain tumors. Firstly, the project aims to develop and implement state-of-the-art

deep learning techniques for automated image registration and segmentation of brain tumor images, specifically focusing on Magnetic Resonance Imaging (MRI). These techniques will facilitate precise alignment of images from various sources or time points, as well as accurate delineation of tumor boundaries. Secondly, the project seeks to enhance the efficiency of the diagnostic process by significantly reducing the time and manual effort required for image interpretation. By doing so, it aims to empower medical professionals with the ability to make timely and informed decisions regarding patient care. Thirdly, the project aspires to contribute valuable insights to the field of medical imaging, with the potential for broader applications beyond brain tumors. Ultimately, the central goal is to improve patient outcomes by revolutionizing the way we diagnose and treat brain tumors, thus paving the way for more effective and personalized treatment strategies.

1. 5 Report organisation

The report is thoughtfully structured to provide a cohesive exploration of deep learning techniques applied to the field of medical image processing. It commences with an introduction to deep learning, covering fundamental concepts such as machine learning, the history of deep neural networks, and activation functions. Following this, it delves into various classification approaches within deep learning and their diverse applications, particularly in medical image analysis. Transitioning into the core of the report, the chapter on image registration provides an exhaustive examination of both traditional and deep learning-based techniques, each offering unique insights into the state-of-the-art methods.

The subsequent chapter centers on the Voxelmorph framework, offering a comprehensive look at its architecture, loss functions, and experimental setup for medical image registration, along with the corresponding results. Following this, the report shifts its focus to image segmentation, discussing traditional and deep learning techniques, data preprocessing, model selection, and the corresponding experimental results. This structured arrangement ensures a logical progression of ideas, fostering a comprehensive understanding of the project's scope and contributions.

Finally, the report culminates with a concise conclusion summarizing key findings and the impact of the applied deep learning techniques in the domain of medical image processing. This organized structure guides the reader through the project's journey, providing clarity and context at every step.

CHAPTER 2

DEEP LEARNING

Preambule

This chapter presents a comprehensive exploration of deep learning, beginning with an overview of machine learning and diving into the historical development of deep neural networks. It also delves into activation functions and provides a classification of various deep learning approaches, with a special focus on their applications, particularly in the medical image domain

2. 1 Machine learning

Machine learning is a field of artificial intelligence, which consists in programming a machine to learn how to perform tasks by studying examples of these tasks. From a mathematical point of view, these examples are represented by data that the machine uses to develop a model. E.g a function of the type f(x) = a x + b. The goal of the machine learning game is to find the parameters a and b that give the best possible model, to fit our data. For this, we program an optimization algorithm that will test different values of a and b until we obtain the combination that minimizes the distance between the model and the points. And that's it; It's all about developing a model using an optimization algorithm to minimize the errors between the model and our data. And there are lots of models: like linear models, decision trees, or Support Vector Machines. Each coming with its own optimization algorithm: gradient descent for linear models, the CART algorithm for decision trees, or the maximum margin for Support Vector Machines. Now, what about deep learning? Well, the Deep Learning is a domain of Machine Learning in which, instead of developing one of the models we just mentioned, we develop what we call artificial neural network. So, the principle remains exactly the same: we provide the machine with data, and it uses an optimization algorithm to adjust the model to these data. The model to this data. But this time, our model is not a simple function of the type f(x) = ax + b, but rather a network of interconnected functions; A neural network. that is, deep learning, when we develop artificial neural networks. So, remember that actually Deep Learning is a field of machine

learning, which is based on the same foundations as machine learning, and that machine learning is itself a domain of artificial intelligence.

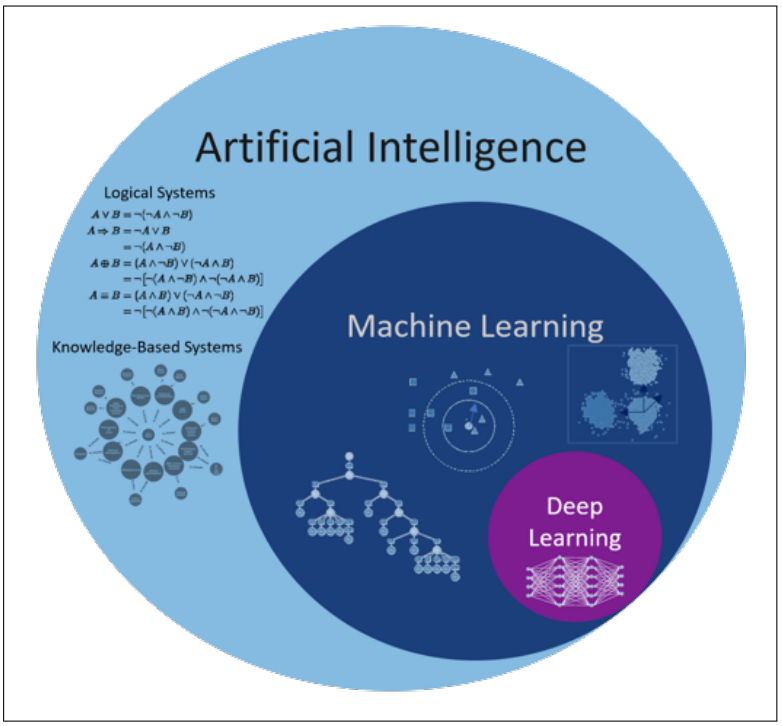

FIGURE 2.1 – Artificial-intelligence

2. 2 History of deep neural network

To understand how artificial neural networks work, I would like to go back to the origin of their history. The first neural networks were invented in 1943 by two mathematicians and neuroscientists named **Warren McCulloch** and **Walter Pitts** [**?**]. In their scientific paper entitled: "A Logical Calculus of the ideas immanent in nervous activity", they explain how they were able to program artificial neurons inspired by the functioning of biological neurons. Remember, in biology, neurons are excitable cells connected to each other, and their role is to transmit information in our nervous system. Each neuron is composed of several dendrites, a soma (cell body), and an axon. Dendrites are the gateways to a neuron.

It is at this point, at the synapse, that the neuron receives signals from the preceding neurons. These signals can be excitatory or inhibitory. When the sum of these signals exceeds a certain threshold, the neuron activates and produces an electrical signal. This signal travels along the axon to the endings and is sent to other neurons in our nervous system; neurons that will work exactly the same way. What Warren McCulloch and Walter Pitts have tried to do is to model this operation by considering that a neuron could be represented

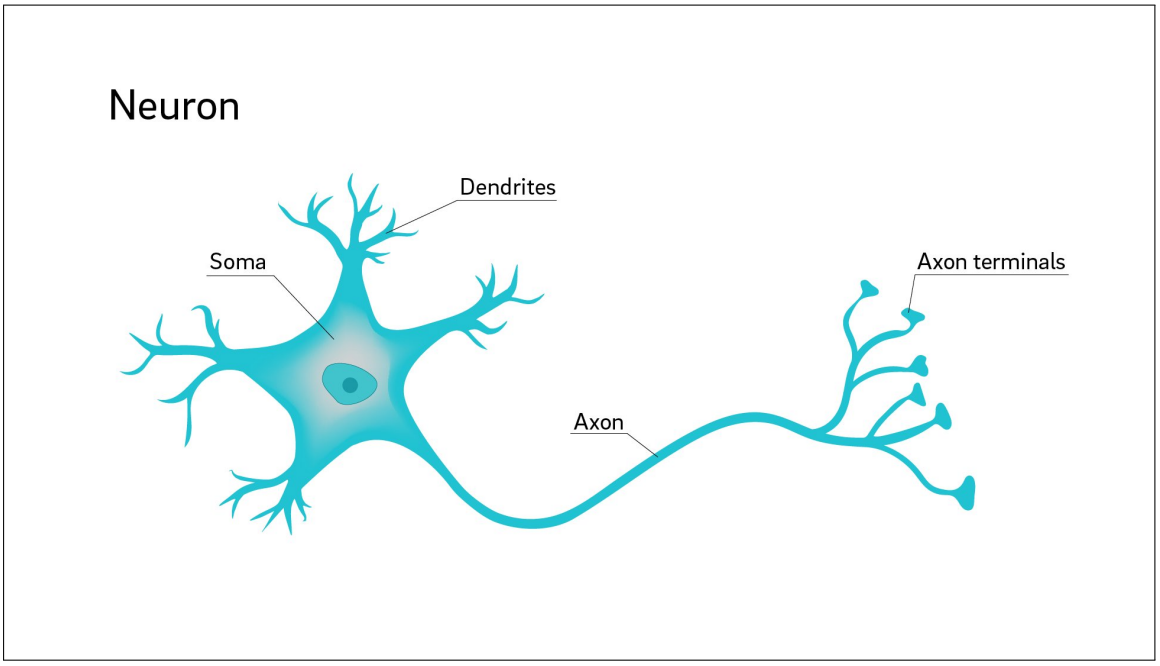

FIGURE 2.2 – Neuron biology

by a transfer function, which takes as input X signals and returns an output y. Inside this function, there are 2 main steps. The first one is an aggregation step. We make the sum of all the inputs of the neuron, by multiplying each input by a coefficient W. This coefficient represents the synaptic activity, i.e., whether the signal is excitatory, in which case w is +1, or inhibitory, in which case it is -1. In this aggregation phase, we obtain an expression of the form w1 x1 + w 2 x 2 + w3 x3 etc. Once this step has been completed, we move on to the activation phase. We look at the result of the calculation made previously, and if it exceeds a certain threshold, usually 0, then the neuron is activated and returns an output y = 1. Otherwise, it remains at 0. So, this is how Warren McCulloch and Walter Pitts managed to develop the first artificial neurons later renamed **"Threshold Logic Unit"**. This name comes from the fact that their model was originally designed only to process logic inputs that are either 0 or 1. They were able to demonstrate that with this model, it was possible to certain logic functions, such as the END gate and the OR gate. They also showed that by connecting several of these functions to each other, a little bit like the neurons in our brain, then it would be possible to solve any Boolean logic problem. Some people even thought that in a few years we would be able to develop artificial intelligence capable of completely replacing human beings! Of course, this was not the case... because even though this model lays the foundations of what Deep Learning is still today, it contains a number of flaws... notably the fact that it does not have a learning algorithm, and that we have to find ourselves the values of the W parameters if we want to use it for real world applications. Fortunately, about fifteen years later, in 1957, **Franck Rosenblatt** (an American psychologist) found a way to improve this model, by proposing the first learning algorithm in the history of Deep Learning.

Which is the Perceptron [**?**], it looks in fact very closely to the one we have just studied. This is an artificial neuron, which is activated when the weighted sum of its inputs exceeds a certain threshold, usually 0. But with this, the perceiver also has a learning algorithm allowing it to find the values of its parameters w in order to obtain the outputs y that we want. To develop this algorithm, Frank Rosenblatt was inspired by **Hebb's theory** . This theory suggests that when two biological neurons are jointly excited, then they strengthen their synaptic links that is, they strengthen the connections between them. In neuroscience, this is called synaptic plasticity, and this is what allows our brain to build memory, to learn new things, to make new associations. So, from this idea, Frank Rosenblatt developed a learning algorithm, which consists in training an artificial neuron on reference data (X, y) so that it reinforces its parameters w each time an input X is activated at the same time as the output y present in these data.

Following this invention, there was again an unbridled craze for artificial intelligence. It was thought that with Perceptron, it would be possible to build machines that could read, speak, walk, and even have a conscience; But all this craze collapsed a few years later, when we realized that these promises could not be kept because the perceptron is a **linear model**. The first winter of artificial intelligence was known, from 1974 to 1980, period during which there were almost no more investors to finance A.I. research Artificial intelligence was about to die. Fortunately, everything changed in the 1980s when **Geoffrey Hinton**, developed the multi-layer Perceptron, the first true artificial neural network. The only problem is that a lot of the phenomena in our universe are not linear. But remember the idea of McCulloch and Pitts: by connecting several neurons together, it is possible to solve more complex problems than with one

2. 3 Activation functions

The basic component of neural networks is the perceptron. It is a binary classification model capable of linearly separating two distinct classes of points. And to improve this model a good thing to do would be to accompany each prediction of a probability, the probability of each point that it belongs to its class, for that we could use an activation function, The activation functions are used to generate nonlinear relationships between the input and the output. This non-linearity, combined with many neural nodes and many layers, mimics the human brain like structure, which is why it is called a neural network [**?**]. There are lots of activation functions, some of them are shown in Fig.2.3

The activation function transforms and abstracts data into a more classifiable plane. In general, the data is extremely densely clustered; it is the task of the activation function to translate the data into a different plane,see Fig 2.4
offering the impacts of multiple dimensions in the given situation to be observed. The greatest and most

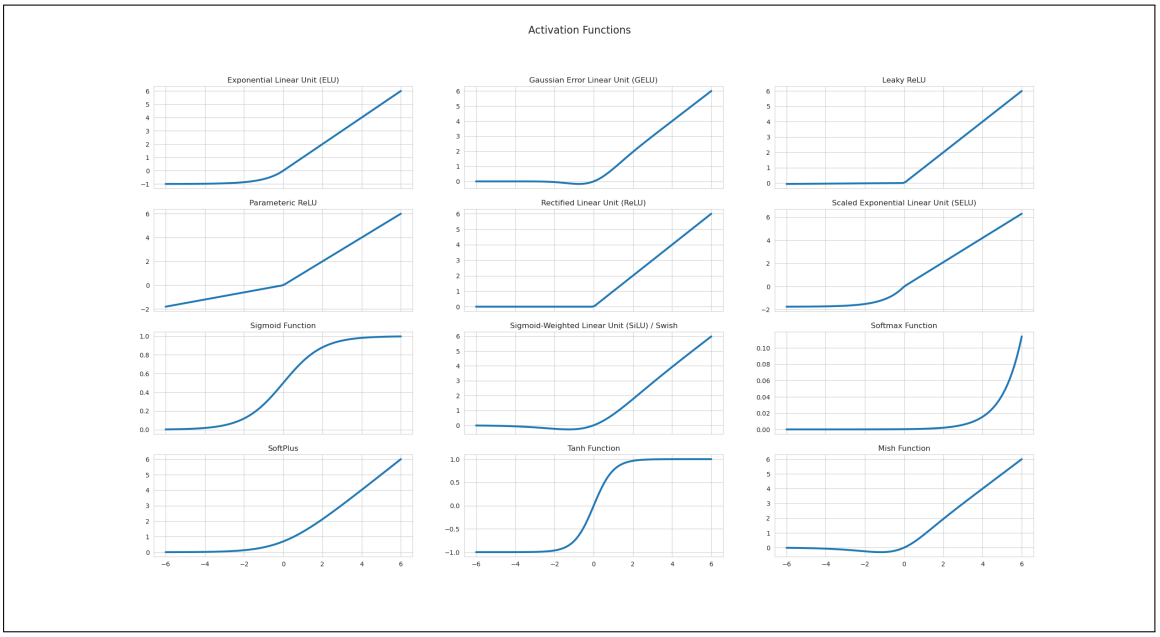

FIGURE 2.3 – Activation Functions

well-known example of an activation function is sigmoid activation, which is utilized in logistic regression. In fact, logistic regression may be thought of as a single brain unit. The sigmoid function's duty is to accept any input and produce an output between 0 and 1, which is used in binary classification problems. And there are also what we called SoftMax function. It is a generalization of logistics function to multiple dimensions, and used in multinomial logistic regression (Multiclass classification problem).

2. 4 Classification of DL approaches

Classification of DL approaches DL techniques are classified into three major categories: unsupervised, partially supervised (semi-supervised) and supervised. Furthermore, deep reinforcement learning (DRL), also known as RL, is another type of learning technique, which is mostly considered to fall into the category of partially supervised (and occasionally unsupervised) learning techniques [**?**].

2. 4. 1 Deep supervised learning

This technique deals with labeled data. When considering such a technique, the environs have a collection of inputs and resultant outputs $(x_t, y_t) \sim \rho$. For instance, the smart agent guesses $y_t^{'} = f(x_t)$ if the input is x_t and will obtain as a loss value. Next, the network parameters are repeatedly updated by the agent to obtain an improved estimate for the preferred outputs. Following a positive training outcome, the agent acquires the ability to obtain the right solutions to the queries from the environs. For DL, there are

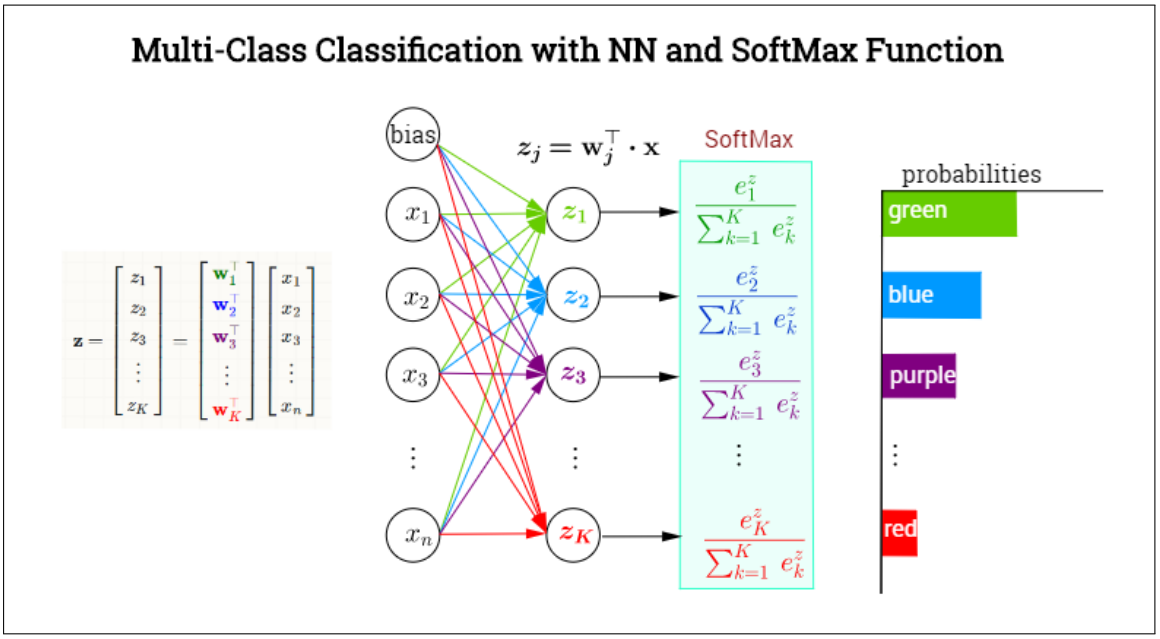

FIGURE 2.4 – Softmax Function

several supervised learning techniques, such as recurrent neural networks (RNNs), convolutional neural networks (CNNs), and deep neural networks (DNNs). In addition, the RNN category includes gated recurrent units (GRUs) and long short-term memory (LSTM) approaches. The main advantage of this technique is the ability to collect data or generate a data output from the prior knowledge. However, the disadvantage of this technique is that decision boundary might be overstrained when training set doesn't own samples that should be in a class. Overall, this technique is simpler than other techniques in the way of learning with high performance.

2. 4. 2 Deep semi supervised learning

In this technique, the learning process is based on semi-labeled datasets. Occasionally, generative adversarial networks (GANs) and DRL are employed in the same way as this technique. In addition, RNNs, which include GRUs and LSTMs, are also employed for partially supervised learning. One of the advantages of this technique is to minimize the amount of labeled data needed.

On other the hand, one of the disadvantages of this technique is irrelevant input feature present training data could furnish incorrect decisions. Text document classifier is one of the most popular examples of an application of semi-supervised learning. Due to difficulty of obtaining a large amount of labeled text documents, semi-supervised learning is ideal for text document classification task.

2. 4. 3 Deep unsupervised learning

This approach enables the execution of the learning process without relying on available labeled data; in other words, it does not necessitate labels. Within this framework, the agent acquires the essential features or internal representations necessary for uncovering latent structures or relationships within the input data. Unsupervised learning encompasses various techniques, including generative networks, dimensionality reduction, and clustering. Several deep learning models, such as restricted Boltzmann machines, auto-encoders, and more recently, GANs, have demonstrated proficiency in non-linear dimensionality reduction and clustering tasks. Additionally, recurrent neural networks (RNNs), including GRUs and LSTM approaches, have found application in diverse unsupervised learning scenarios. However, unsupervised learning has limitations, notably its inability to provide precise data categorization and its computational complexity. Among the most popular methods in unsupervised learning is clustering [**?**].

2. 4. 4 Deep reinforcement learning

Reinforcement Learning operates on interacting with the environment, while supervised learning operates on provided sample data. This technique was developed in 2013 with Google Deep Mind. Subsequently, many enhanced techniques dependent on reinforcement learning were constructed. This method is sometimes referred to as semi-supervised learning. In comparison with traditional supervised techniques, performing this learning is much more difficult, as no straightforward loss function is available in the reinforcement learning technique. In addition, there are two essential differences between supervised learning and reinforcement learning: first, there is no complete access to the function, which requires optimization, meaning that it should be queried via interaction; second, the state being interacted with is founded on an environment, where the input x_t is based on the preceding actions. For solving a task, the selection of the type of reinforcement learning that needs to be performed is based on the space or the scope of the problem. For example, DRL is the best way for problems involving many parameters to be optimized. By contrast, derivative-free reinforcement learning is a technique that performs well for problems with limited parameters. Some of the applications of reinforcement learning are business strategy planning and robotics for industrial automation. The main drawback of Reinforcement Learning is that parameters may influence the speed of learning. [**?**]

2. 5 Applications of deep learning

Presently, various DL applications are widespread around the world. These applications include healthcare, social network analysis, audio and speech processing (like recognition and enhancement), visual data processing methods (such as multimedia data analysis and computer vision), and NLP (translation and

sentence classification), among others (Fig.2.5) [**?**].

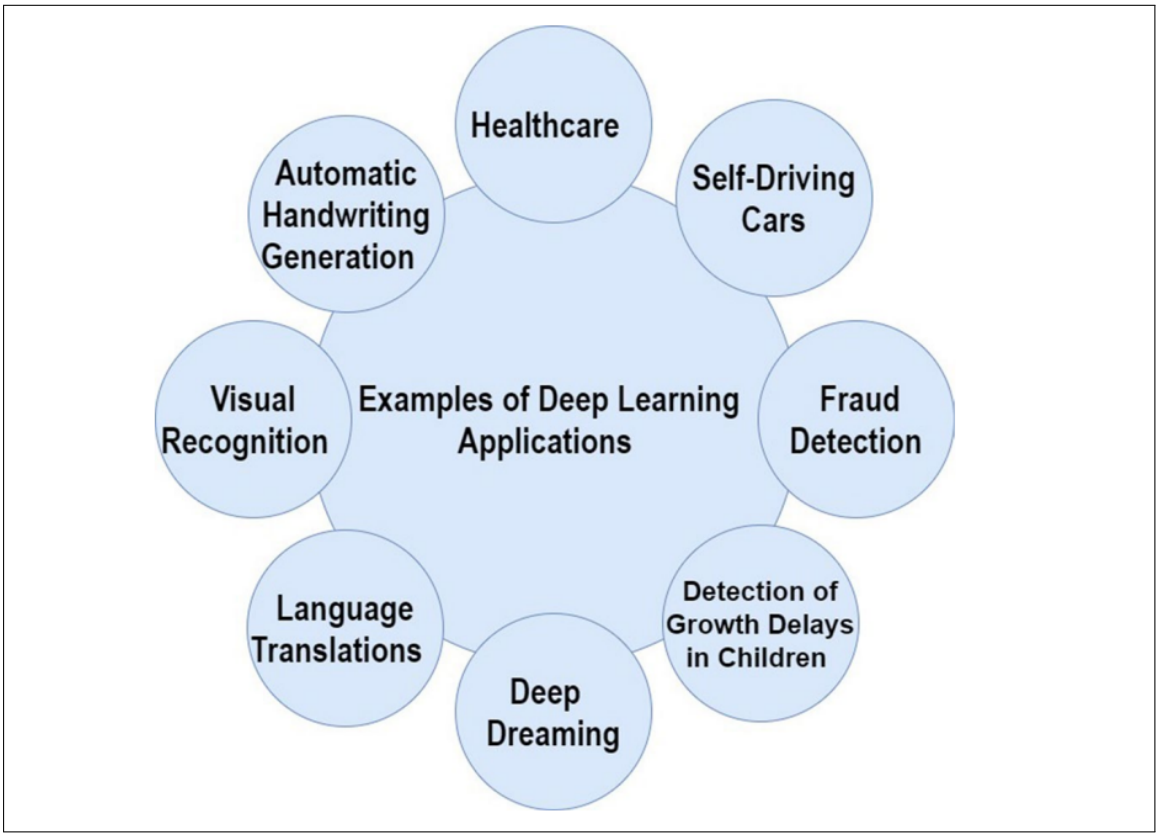

FIGURE 2.5 – Examples of DL applications

These applications have been classified into five categories: classification, localization, detection, segmentation, and registration. Although each of these tasks has its own target, there is fundamental overlap in the pipeline implementation of these applications as shown in Fig.2.6.

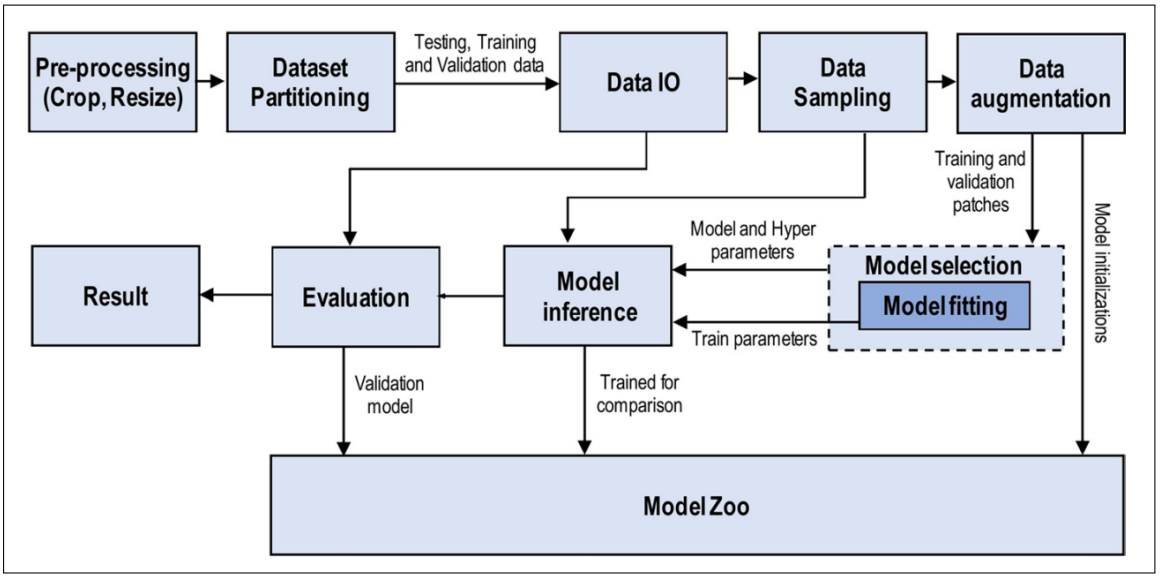

FIGURE 2.6 – Workflow of deep learning tasks

Classification is a concept that divides a collection of data into categories. Finding interesting things in an image requires detection, which takes the backdrop into account. Multiple items that may belong to different classes are encircled by bounding boxes during detection. Localization is utilized to identify the item, which has a single bounding box around it. In semantic segmentation, each pixel is assigned a label that represents the class of the object or region to which it belongs. moreover, fitting a single image (which could be 2D or 3D) onto another refers to registration. One of the most significant and varied Deep Learning applications is in the field of healthcare. Because of this, we take DL applications in the area of medical image analysis to explain Deep Learning applications. [**?**] [**?**]

2. 5. 1 The classification of medical image

Diabetic retinopathy detection In the field of deep learning, image classification and its application have made great progress in the year 2020. On the one hand, the academic circles have made great efforts to design a variety of efficient CNN models, which have achieved high accuracy and even exceeded the human recognition ability. On the other hand, the application of CNN model in medical image analysis has become one of the most attractive directions of deep learning. In particular, the retinal fundus image obtained from fundus camera has become one of the key research objects of deep learning in the field of image classification. Furthermore, the classification of breast nodules; The current deep learning technology has achieved research results in the field of ultrasound imaging such as breast cancer, cardiovascular and carotid arteries. [**?**]

2. 5. 2 Localization

Although applications in anatomy education could increase, the practicing clinician is more likely to be interested in the localization of normal anatomy. Radio-logical images are independently examined and described outside of human intervention, while localization could be applied in completely automatic end-to-end applications. Zhao & al [**?**] introduced a new deep learning-based approach to localize pancreatic tumor in projection X-ray images for image-guided radiation therapy without the need for fiducials. Roth & al. constructed and trained a CNN using five convolutional layers to classify around 4000 transverse-axial CT images. [**?**]

2. 5. 3 Regisration

Deep Learning for medical image registration has numerous applications, which were listed by some review papers. Yang & al. implemented stacked convolutional layers as an encoder-decoder approach to predict the morphing of the input pixel into its last formation using MRI brain scans from the OASIS dataset. They employed a registration model known as Large Deformation Diffeomorphic Metric Mapping (LDDMM) and attained remarkable enhancements in computation time. Miao & al. used synthetic X-ray images to train a five-layer CNN to register 3D models of a trans-esophageal probe, a hand implant, and a knee implant onto 2D X-ray images for pose estimation. They determined that their model achieved an execution time of 0.1 s, representing an important enhancement against the conventional registration techniques based on intensity; moreover, it achieved effective registrations 79–99% of the time.

2. 5. 4 Segmentation

Brain MRI analysis is mainly for the segmentation of different brain regions and the diagnosis of brain diseases, such as brain tumor segmentation, schizophrenia diagnosis, early diagnosis of Parkinson's syndrome and early diagnosis of AD. Among them, the broadest field of deep learning applications is the early diagnosis of AD. AD diagnosis based on deep learning is mainly based on segmentation of hippocampus, cortical thickness and brain volume in brain MRI images Furthermore, Left ventricular segmentation Cardiac MRI analysis diagnoses heart disease by dividing the left ventricle to measure left ventricular volume, ejection fraction, and wall thickness. Among them, deep learning is widely used in left ventricular segmentation. In recent years, deep learning algorithms for left ventricular segmentation on MRI image have emerged in an endless stream. [**?**]

CHAPTER 3

IMAGE REGISTRATION : THE STATE-OF-THE-ART

Préambule

This chapter presents an in-depth examination of image registration, starting with an introduction and defining the concept. It explores the various types of image registration, including deformable methods. Furthermore, it discusses both conventional and deep learning-based techniques, offering insights into their applications and significance in the realm of medical imaging.

3. 1 Introduction

For reasons of diagnosis, prognosis, therapy, and follow-up, there are a number of situations in which images must be taken during the majority of medical treatments. These images can have different temporal, spatial, dimensional, or modular characteristics. When it comes to online and real-time decision-making, image fusion that results in information synergy can significantly aid and support doctors. Lack of alignment is unavoidable for these images taken in different times and conditions; hence, can challenge the quality and accuracy of the subsequent analyses.

3. 2 What is Image Registration

Image registration in the medical domain refers to the process of aligning and matching two or more medical images of the same patient or anatomical region, taken at different times or using different imaging modalities. The goal of image registration is to find the spatial transformation that maps one image onto another, allowing for a meaningful comparison or fusion of the information contained in the images.

The aim is at finding an optimum spatial transformation that registers the structures-of-interest in the best way. This problem is important in numerous ways in the field of machine vision e.g., for remote sensing, object tracing, satellite imaging and so on (Goshtasby 2017) [**?**].

3. 3 Types of Image Registration

There are several types of image registration techniques, depending on the specific application and requirements:

3. 3. 1 Rigid Registration

In rigid registration, only translation, rotation, and scaling transformations are considered. It assumes that the images have no deformations, and the relationship between corresponding points is purely rigid. Rigid registration is commonly used when the objects in the images undergo minimal deformations, such as in aerial or satellite imagery [**?**].

FIGURE 3.1 – Satellite Imagery Registration

3. 3. 2 Affine Registration

Affine registration allows for more complex transformations than rigid registration by incorporating shearing and non-uniform scaling. It is useful when the images exhibit moderate deformations while maintaining some level of rigidity. Affine registration is commonly used in applications like 3D medical image registration or image stitching in photography.

3. 3. 3 Intensity-based Registration

Intensity-based registration methods rely on the pixel intensity or voxel values of the images to find the best alignment. Common similarity measures used in intensity-based registration include Normalized Mutual Information (NMI) or Cross-Correlation.

FIGURE 3.2 – image stitching

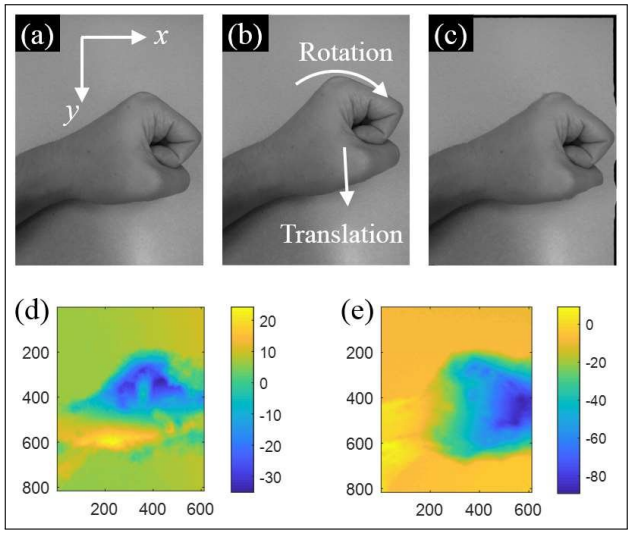

FIGURE 3.3 – Demonstration of intensity-based image registration: (a) input Image A; (b) input Image B; (c) newly-registered Image C; (d) displacement field in x direction; and (e) displacement field in y direction.

3. 3. 4 Feature-based Registration

Feature-based registration techniques extract and match distinctive features or key points in the images, like corners or edges, to determine the transformation. Feature-based methods are robust to variations in image intensities and are commonly used in computer vision tasks.

3. 3. 5 Hierarchical Registration

Hierarchical registration approaches use a multi-resolution or pyramid-based strategy. The registration process used to align two or more datasets in a hierarchical manner. Instead of performing registration in one single step, hierarchical registration divides the datasets into multiple levels or resolutions, starting from coarse to fine details

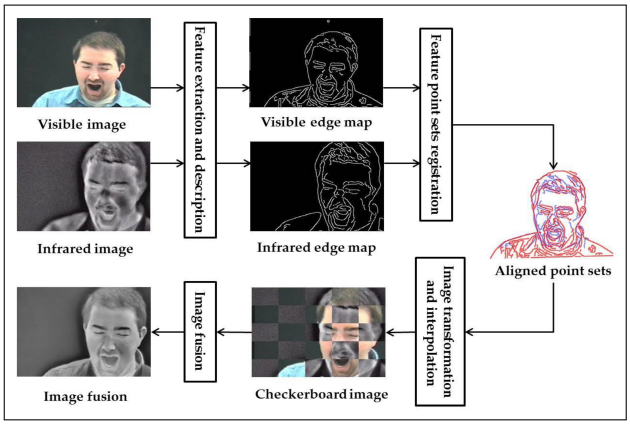

FIGURE 3.4 – Multi-Sensor Face Registration Based on Global and Local Structures

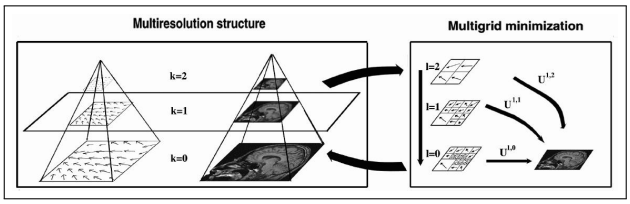

FIGURE 3.5 – Example of multiresolution/multigrid minimization.

3. 3. 6 Deformable Registration

Deformable registration [**?**], also known as Non-rigid registration, is a technique used to align two or more datasets that undergo significant shape-changing deformations. Unlike rigid registration, which assumes that the objects maintain their shape and only undergo translation and rotation, non-rigid registration allows for local deformations and changes in the shape of the objects being aligned.

The main characteristic of non-rigid registration is its ability to model complex and variable transformations, making it suitable for tasks involving organ motion in medical images, facial expression changes in computer vision, and animation in computer graphics. The process involves finding a smooth and continuous transformation that best matches the features or points in one dataset to their corresponding locations in the other dataset.

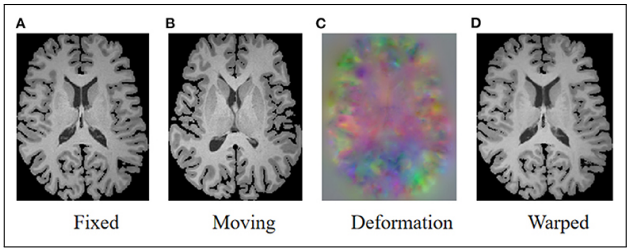

FIGURE 3.6 – Defromable registration

3. 4 Conventional image registration

Basically, conventional image registration is an iterative-based optimization process that requires extracting proper features, selecting a similarity measure (to evaluate the registration quality), choosing the model of transformation, and finally, a mechanism to investigate the search space [**?**]. As shown in Fig. 3.7, the system receives two pictures, one of which is entered as a fixed image and the other as a moving image. Iteratively sliding the moving picture over the fixed image will result in the best alignment. The considered similarity measure first determines how closely the submitted photos relate to one another. The parameters for the new transformation are calculated using an optimization algorithm applying an update mechanism.The moving image improves its correspondence with the fixed image with each iteration, and this process is repeated until no more registration can be accomplished

This approach has the following two key drawbacks:

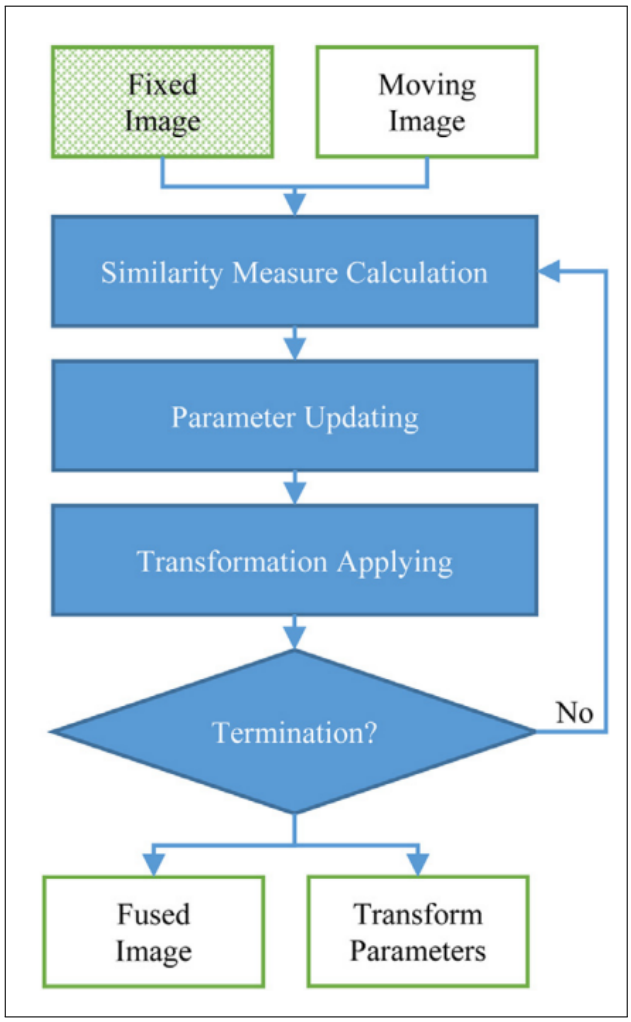

FIGURE 3.7 – The workflow of conventional optimization-based image registration techniques

— Even with an efficient implementation, this iterative method is exceedingly slow and runtimes for standard deformable image registration approaches average in the tens of minutes. Since real-time clinical operations are the norm, this protracted time waste is not acceptable.
— Most similarity measures have a large number of local optima surrounding the global one, especially when dealing with images from different modalities (known as multimodal image registration); as a result, they lose their effectiveness due to premature convergence or stagnation, two common confining dilemmas in the field of optimization.

As a result, to overcome these two constraints, learning-based registration approaches have grown in popularity in recent years; meanwhile, deep neural networks (DNNs), as one of the most powerful techniques ever seen by the machine intelligence community, have been applied to a variety of image processing applications. Medical image registration is no exception, and several Deep Learning (DL)-based approaches have been proposed in the literature. nevertheless, the number of works and techniques used are very limited, and there is a promising potential for further research.

3. 5 Deep Learning based techniques

Iterative optimization strategies are used in traditional image registration. Based on a preset similarity metric, improved alignment is intended to be attained in each cycle. Operations continue until either better registration cannot be made or a set of predetermined requirements is met. Long reaction times and the proposed similarity measures' flawed nature, particularly for multimodal registration, which results in becoming stuck in local minima, are the most difficult problems to solve for utilizing this paradigm. For tackling the aforementioned challenges, DL-based techniques have developed in prominence in recent years.

Five key generations of techniques may be identified in a taxonomy based on literature breakthroughs; Deep Similarity Metrics (DSM), Supervised End-to-End Registration (SE2ER), Deep Reinforcement Learning (or Agent-Based Registration) (DRL), Unsupervised End-to-End Registration (UE2ER), and Weakly/Semi-Supervised End-to-End Registration (WSE2ER) [**?**].

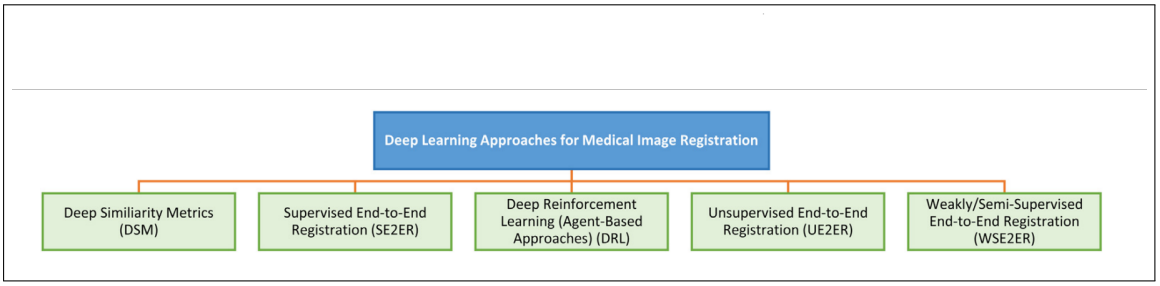

FIGURE 3.8 – The taxonomy on deep learning approaches for medical image registration

3. 5. 1 Deep Similarity Metrics

The first wave of studies relied on the use of various types of DNNs to learn visual similarity metrics from a huge number of annotated paired ground-truths (see fig 3.9). They are known as deep similarity measures/metrics. In order to obtain the final transformation parameters, DSM techniques frequently supply the metric for the common iterative deformable registration algorithms. It can definitely be a potent rival for the conventional multimodal similarity measures, e.g., Mutual Information (MI), if and only if an adequate number of clearly annotated ground-truths are available, which is a severe restricting factor to develop such approaches. Today, there have been many similar approaches conducted. Additionally, it has been established that the use of deep similarity measures for unimodal registration is not strongly justified if the similarity measure may be appropriately chosen based on the context and modality [**?**].

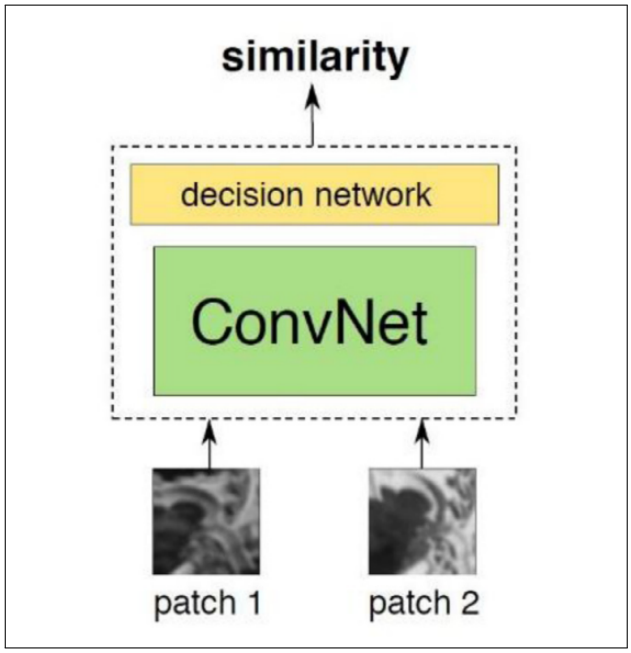

FIGURE 3.9 – A deep similarity metric based on the CNN.

3. 5. 2 Supervised End-to-End Registration

The second wave is part of the end-to-end supervised registration technique (see fig 3.10), in which various DNNs types were trained on ground truth to build regression models that produced the transformation parameters all at once. CNN and U-Net are the predominate methods for the affine and deformable transformation models. It involves training a neural network to directly learn the mapping or transformation between a pair of images: one acting as a reference and the other as the input image. This method requires a dataset comprising such pairs, with known ground truth registration parameters representing the transformation

needed to align the images properly. Through this supervised learning process, the model learns to extract meaningful features from the images and predict the accurate registration parameters. The beauty of this technique lies in its ability to handle complex image deformations and various types of transformations. However, obtaining a diverse and accurately annotated dataset can be challenging, especially for applications like medical imaging [**?**].

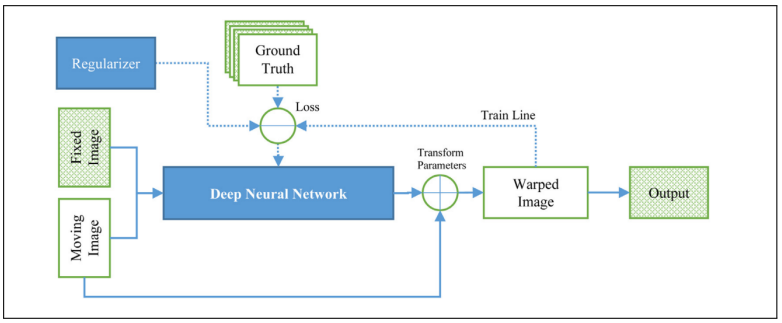

FIGURE 3.10 – The main framework for supervised end-to-end medical image registration.

3. 5. 3 Deep Reinforcement Learning

The third wave of image registration techniques falls under Deep Reinforcement Learning (DRL). Similar to Fig. 3.11, in this wave, a deep agent or multiple agents learn to iteratively produce the final transformation. The aim is to maximize positive feedback from the environment, which is often represented by a predefined similarity measure. Unlike the previous deep similarity measure (DSM) approach, these methods use conventional similarity measures such as Normalized Mutual Information (NMI) or Local Cross-Correlation (LCC). However, a significant limitation of this paradigm is that the agents struggle to interact effectively with the vast state-space introduced by deformable registration tasks. Consequently, they cannot capture the necessary deformations required for successfully registering elastic organs or dealing with longer registration times. This limitation hampers the effectiveness of this approach and limits its potential for successful implementation [**?**].

3. 5. 4 Unsupervised End-to-End Registration

In the past, the paradigms introduced relied on ground-truth data to build the models. However, in the field of medicine, particularly in image registration, annotated datasets are usually small in size and not well-suited for exhaustive deep learning. As a result, researchers shifted their focus towards the fourth generation of approaches known as unsupervised end-to-end registration. The main framework for unsupervised end-to-end medical image registration (see fig 3.13) involves training Deep Neural Networks (DNNs) without relying on ground-truth data. Unlike previous paradigms that heavily depended on annotated datasets for

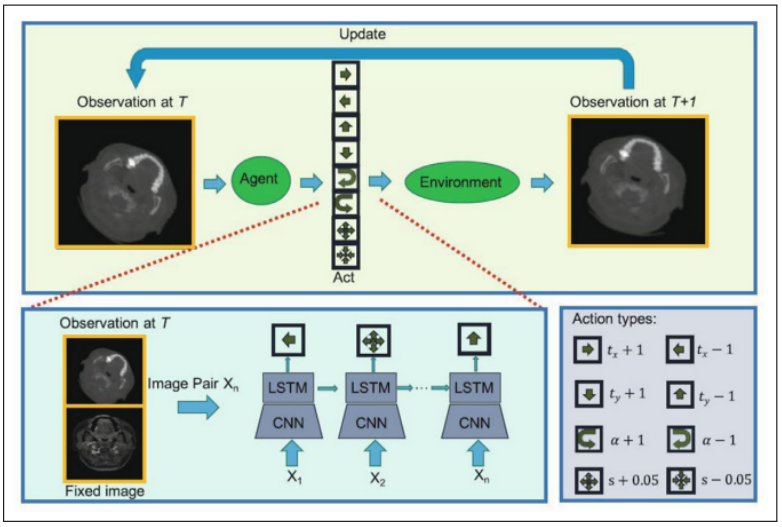

FIGURE 3.11 – Deep Reinforcement Learning (DRL) architecture applied to medical image registration

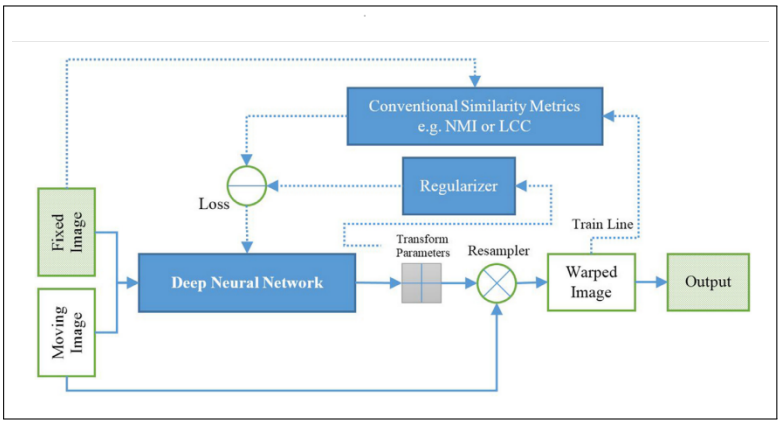

FIGURE 3.12 – The main framework for unsupervised end-to-end medical image registration

constructing registration models, this fourth-generation approach takes a more self-learning approach. The DNNs are designed to produce the transformation parameters directly from the input images, bypassing the need for explicit ground-truth annotations. Instead, the networks learn to find the optimal transformations by maximizing certain objective functions or similarity metrics between the images [**?**].

3. 5. 5 Weakly/semi-supervised End-to-End Registration

The fifth generation belongs to weakly/semi-supervised techniques (see fig 3.9), which can be divided into two key paradigms. In one approach, a few fully-annotated ground-truth data with multiple landmarks, such as contours, regions, corners, and lines, are used. Each landmark is assigned a distinct class label, and the network is trained on these labeled data. Besides its primary task of image registration, the network also

learns to detect landmarks in any input image pair. This landmark detection is crucial for constructing efficient models and improving system accuracy. Moreover, to train the network, the Target Registration Error (TRE), a valuable structural similarity measure, can be employed as a non-trivial loss function. Notably, the work by Hu et al. in 2018 stands as a comprehensive representation of this paradigm.

Another approach in this generation involves the utilization of Generative Adversarial Networks (GANs). Here, the generator takes fixed and moving input images and attempts to produce transformation parameters, ensuring that the transformed moving image (warped image) cannot be distinguished from the ground-truth by the discriminator, akin to an expert registration agent's expectation. By employing game theory-like survival competition between the generator and discriminator, the network can be trained on a small dataset so that the generated samples become indistinguishable, leading to the network achieving an equilibrium state [**?**].

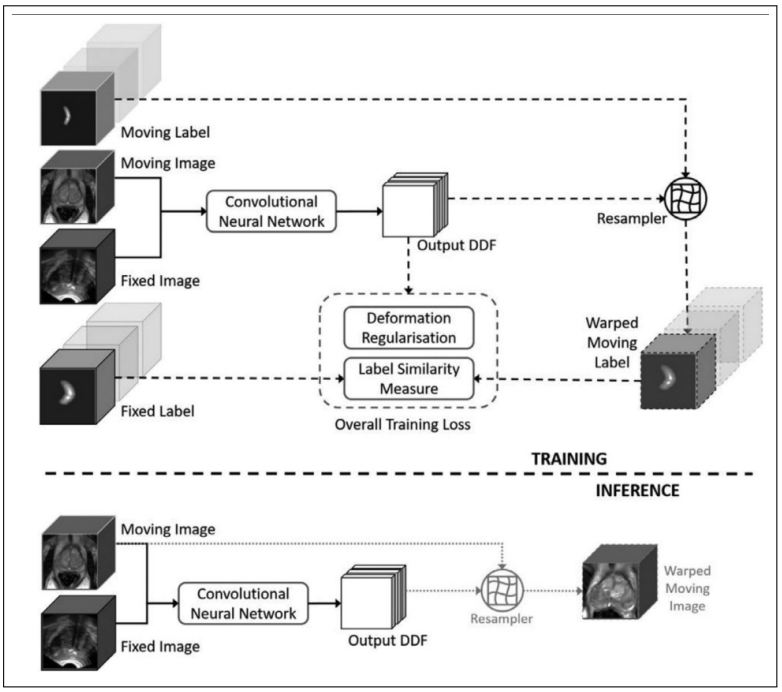

FIGURE 3.13 – The main framework for weakly-supervised label-driven medical image registration

CHAPTER 4

VOXELMORPH : A LEARNING FRAMEWORK FOR IMAGE MEDICAL REGISTRATION

Preambule

This chapter presents Voxelmorph, a sophisticated learning framework designed for medical image registration. It begins with an introduction to MRI, brain tumors, and tumor grades. It then outlines the problem statement, architecture, and VoxelMorph CNN architecture. The chapter delves into the intricacies of the loss function, including the image similarity term, smoothness term, and auxiliary data loss function. Experimental setups, evaluation metrics, and conclusions regarding Voxelmorph are also discussed.

4. 1 Introduction

Image registration is a sophisticated image processing technique that plays a pivotal role in the field of medical imaging. It involves aligning and spatially transforming multiple images of the same subject or region acquired at different times or using different imaging modalities. As mentioned in the previous chapter, image registration enhances the clinical utility of medical imaging by providing valuable insights into anatomical changes, disease progression, and treatment response.

In the context of brain tumor assessment, image registration holds immense significance as it enables the comparison and fusion of various Magnetic Resonance Imaging (MRI) scans. MRI, being a non-invasive and high-resolution imaging modality, is extensively used for visualizing brain tumors and their surrounding structures. The combination of image registration and MRI in brain tumor grading allows for a comprehensive and precise evaluation of tumor characteristics, aiding in treatment planning and monitoring.

4. 1. 1 What is MRI

MRI utilizes a powerful magnetic field and radiofrequency pulses to produce detailed images of the human body's internal structures, providing exceptional soft tissue contrast. Unlike X-rays or CT scans, MRI does not use ionizing radiation, making it a safer imaging option for patients. The technique is particularly valuable for visualizing soft tissues, such as the brain, spinal cord, muscles, and organs. Its multi-planar capabilities allow imaging from various angles, aiding in comprehensive anatomical assessment.

4. 1. 2 Brain tumors

A brain tumor is an intracranial mass produced by an uncontrolled growth of cells either normally found in the brain such as neurons, lymphatic tissue, glial cells, blood vessels, pituitary and pineal gland, skull, or spread from cancers primarily located in other organs.
Brain tumors are the tumors that originated in the brain and are named for the cell types from which they originated. They can be benign (non-cancerous), meaning that they do not spread elsewhere or invade surrounding tissues. They can also be malignant and invasive (spreading to neighboring area). In addition to the solid portion of the tumor, may have other associated parts such as edema and necrosis (see fir 4.1). By definition, brain edema is an increase in brain volume resulting from increased sodium and water content and results from local disruption of the blood brain barrier. Edema appears around the tumor mainly in white matter regions. Necrosis is composed of dead cells in the middle of the brain tumor and are seen hypointense in T1-weighted images

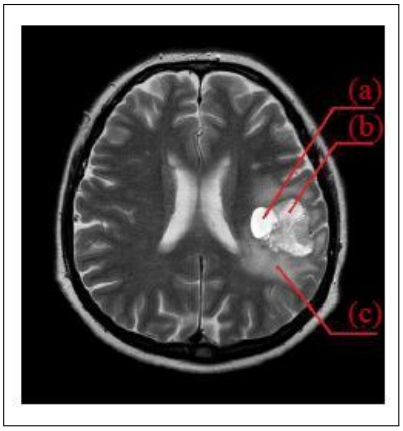

FIGURE 4.1 – T2-Weighted Image Showing the Brain Tumor. (a) Tumor. (b) Necrotic Tissue. (c) Edema

4. 1. 3 Brain Tumor Grades

Bailey and Cushing proposed the early classifications of brain tumors in 1926 [Doolittle, 2004]. They postulated 14 different forms of brain tumors, paid close attention to the process of cell differentiation, and dominated perceptions of gliomas until Kernohan and Sayre's revised categorization system was established in 1949. They provided the crucial insight, who hypothesized that various histopathologic manifestations could not signify distinct tumor types, but rather different degrees of differentiation of one tumor type.
They classified tumors into five sub-types: astrocytoma, oligodendroglioma, ependymoma, gangliocytoma, and medulloblastoma and very importantly added a four-level grading system for astrocytomas. The grading system was based on increasing malignancy and decreasing differentiation with increasing tumor grade. The addition of a grading system was a very important advance in classifying brain tumors, and provided information not only regarding tumors' biologic behavior but also information that could be used to guide treatment decisions [**?**].

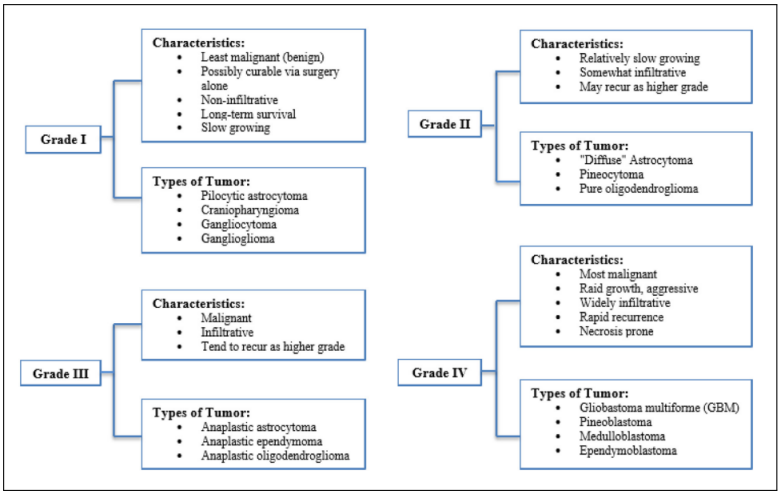

FIGURE 4.2 – Brain Tumors grades

4. 1. 4 MRI modalities

MRI offers several imaging modalities, each highlighting specific tissue properties and providing unique insights into the pathology. Some of the commonly used MRI modalities include:

— **T1-weighted (T1W) MRI**: It provide good anatomical detail and are useful for visualizing brain structures.

— **T2-weighted (T2W) MRI**: T2-weighted images are sensitive to water content, making them valuable for detecting edema or swelling associated with brain tumors.

— **Fluid-Attenuated Inversion Recovery (FLAIR)**: FLAIR sequences suppress the signal from ce-

rebrospinal fluid (CSF) while enhancing abnormalities like tumor lesions. FLAIR images are often used to visualize tumor infiltration and peritumoral edema, assisting in tumor segmentation.

the CE-T1w and FLAIR images: are sufficient for detection and segmentation of the majority of brain tumors and its components such as edema and necrosis.

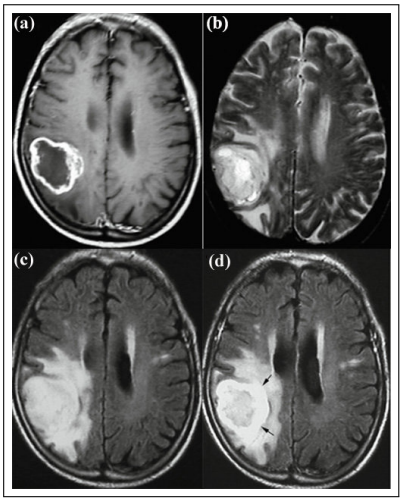

FIGURE 4.3 – Four imaging modalities: (a) T1-weighted MRI; (b) T2-weighted MRI; (c) FLAIR; and (d) FLAIR with contrast enhancement

4. 2 Problem Statement

Five years ago, learning based methods came around and tried to tackle image registration problem; These methods completely forgot the decades of research that had been done in registration and they said we're going to treat this as a black box problem (see figure below), just take these two images (Fixed and Moving), and put them through this black box CNN then you get a deformation field out, means that we need to get a really large dataset that has pairs of images and the ground truth deformation field between them and this treat as one big supervised regression problem (m, f, ϕ); and this was a great idea but the problem is we never really have this ground truth deformation field; you can't really get it from experts, because an expert would need hours or days to draw out all these little arrows, they could do landmarks but that's also extremely costly because this is all in 3d, so it's really hard to find this kind of correspondences.
VoxelMorph aims to address these challenges by employing a data-driven, unsupervised approach (m, f) based on convolutional neural networks (CNNs).

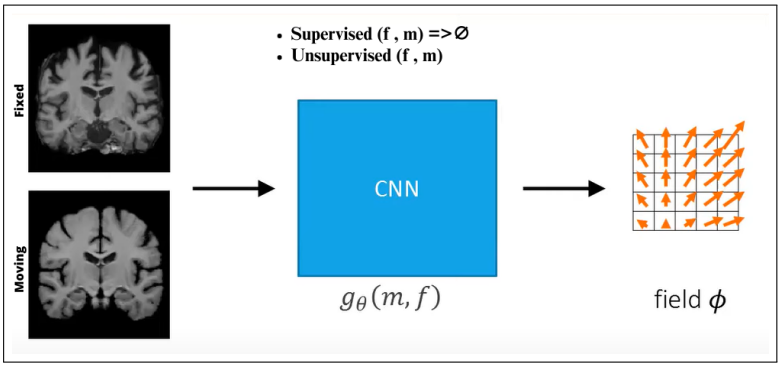

FIGURE 4.4 – Deformable Image Registration

4. 3 Architecture

The network(see fig 4.5) takes Fixed image f and Moving image m as input, and computes Registration field ϕ using a set of parameters θ. We warp m to $m \circ \phi$ using a spatial transformation function, enabling evaluation of the similarity of $m \circ \phi$ and f. Given unseen images f and m during test time, we obtain a registration field by evaluating $g_\theta(m, f)$ [**?**].

They use (single-element) stochastic gradient descent to find optimal parameters θ by minimizing an expected loss function using a training dataset.

The term "Auxiliary Information" refers to any additional information or data that is available during the training process and can be used to enhance the registration performance. The choice of auxiliary data depends on the specific application and the nature of the images being registered; For example, in medical imaging applications, anatomical labels or segmentation maps S_f, S_m can serve as auxiliary data.

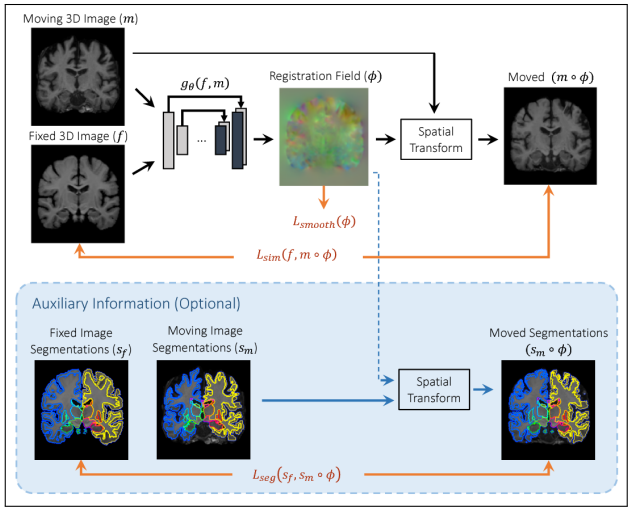

FIGURE 4.5 – Voxelmorph Architecture

4. 4 VoxelMorph CNN Architecture

The parametrization of $g_\theta(.,.)$ is based on a convolutional neural network architecture similar to UNet, which consists of encoder and decoder sections with skip connections. Figure 4.6 depicts the network used in VoxelMorph, which takes a single input formed by concatenating m and f into a 2-channel 3D image. The input is of size $160 \times 192 \times 224 \times 2$, but the framework is not limited by a particular size. It applies 3D convolutions in both the encoder and decoder stages using a kernel size of 3 and a stride of 2. Each convolution is followed by a **LeakyReLU** layer with parameter 0.2. The convolutional layers capture hierarchical features of the input image pair, used to estimate ϕ.

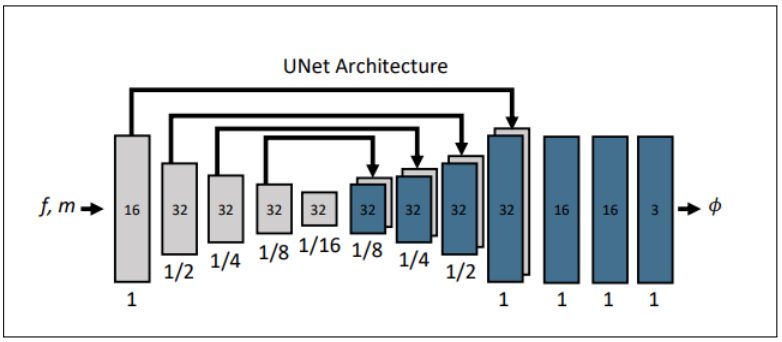

FIGURE 4.6 – VoxelMorph CNN Architecture

4. 5 Loss Function

VoxelMorph utilizes a novel unsupervised loss function called the "VoxelMorph Loss."(see Equation 4.1), designed to drive the network to learn meaningful image correspondences and deformation fields during training. It combines two essential components: the image similarity term and the smoothness term.

$$L_{us}(f, m, \phi) = L_{\text{sim}}(f, m \circ \phi) + \lambda L_{\text{smooth}}(\phi) \tag{4.1}$$

where ϕ represents the deformation field and $m \circ \phi$ represents m warped by ϕ, and λ is the regularization trade-off parameter.

4. 5. 1 Image Similarity Term

The image similarity term in the VoxelMorph Loss measures the dissimilarity between the warped source image and the target image. It encourages the registered image to align well with the target image. expe-

rimented with two often-used functions; The first is the mean squared voxel wise difference (see Equation 4.2), applicable when f and m have similar image intensity distributions and local contrast.

$$MSE(f, m \circ \phi) = \frac{1}{|\Omega|} \sum_{p \in \Omega} [f(p) - [m \circ \phi](p)]^2 \tag{4.2}$$

Where:

— f is a certain function.

— ϕ represents the deformation field.

— $m \circ \phi$ represents m warped by ϕ.

— $|\Omega|$ denotes the cardinality of the set Ω.

— p represents each voxel in the set Ω.

The second term is the local cross-correlation of f and $m \circ \phi$, which is more robust to intensity variations found across scans and datasets. Let $\hat{f}(p)$ and $[\hat{m} \circ \phi]$ denote images with local mean intensities subtracted out: $\hat{f}(p) = f(p) - \frac{1}{n^3} \sum_{p_i} f(p_i)$, where p_i iterates over an n^3 volume around p, with $n = 9$ in Voxelmorph experiments. The local cross-correlation of f and $m \circ \phi$ is written as:

$$\text{Cross-Correlation}(f, m \circ \phi) = \frac{\sum_{p \in \Omega} \left[\hat{f}(p) \cdot (\hat{m} \circ \phi) \right]}{\sqrt{\sum_{p \in \Omega} \left[\hat{f}(p)^2 \right] \cdot \sum_{p \in \Omega} \left[(\hat{m} \circ \phi)^2 \right]}} \tag{4.3}$$

Where:

— f is a certain function.

— ϕ represents the deformation field.

— $m \circ \phi$ represents m warped by ϕ.

— $\hat{f}(p)$ is the image f with the local mean intensity subtracted out at each voxel p.

— $[\hat{m} \circ \phi]$ is the image $m \circ \phi$ with the local mean intensity subtracted out at each voxel p.

— n is the side length of the cubic volume around each voxel p for local mean computation (here, $n = 9$ in Voxelmorph experiments).

— Ω denotes the set of voxels in the images f and $m \circ \phi$.

4. 5. 2 Smoothness Term

The smoothness term promotes spatially smooth deformation fields, preventing overly complex or unrealistic deformations. It encourages neighboring voxels to have similar displacement values, ensuring spatial coherence in the deformation field. The smoothness term can be formulated using different approaches, such as the regularization of deformation gradients or the Laplacian operator.

Voxelmorph encourages a smooth displacement field ϕ using a diffusion regularizer on the spatial gradients of displacement u:

$$L_{Smooth}(u) = \sum_{p \in \Omega} \|\nabla u(p)\|^2 \tag{4.4}$$

Where:

— ϕ is the deformation field.

— u represents the displacement field, which is the difference between the original and warped locations.

— Ω denotes the set of voxels in the displacement field.

— $\nabla u(p) = \left(\frac{\partial u(p)}{\partial x}, \frac{\partial u(p)}{\partial y}, \frac{\partial u(p)}{\partial z}\right)$ represents the spatial gradient of displacement u at voxel p.

The spatial gradients of the displacement field are approximated using differences between neighboring voxels. Specifically, we approximate the partial derivatives as follows:

$$\frac{\partial u(p)}{\partial x} \approx u((p_x + 1, p_y, p_z)) - u((p_x, p_y, p_z)), \tag{4.5}$$

$$\frac{\partial u(p)}{\partial y} \approx u((p_x, p_y + 1, p_z)) - u((p_x, p_y, p_z)), \tag{4.6}$$

$$\frac{\partial u(p)}{\partial z} \approx u((p_x, p_y, p_z + 1)) - u((p_x, p_y, p_z)). \tag{4.7}$$

These approximations allow us to calculate the spatial gradients using the differences between neighboring voxels in the displacement field, enabling the smoothness regularization to be applied effectively.

4. 5. 3 Auxiliary Data Loss Function

The Auxiliary Data Loss function utilizes extra information or data that is available during training to guide the learning process. This additional data can be in the form of anatomical labels, segmentation maps, or any other relevant information related to the images being registered.

Let S_f^k and $S_m^k \circ \phi$ be the sets of voxels corresponding to structure k in images f and $m \circ \phi$, respectively. The volume overlap for structure k is quantified using the Dice score:

$$\text{Dice}(S_f^k, S_m^k \circ \phi) = \frac{2|S_f^k \cap (S_m^k \circ \phi)|}{|S_f^k| + |S_m^k \circ \phi|} \tag{4.8}$$

Where:

— S_f^k represents the set of voxels corresponding to structure k in image f.

— $S_m^k \circ \phi$ represents the set of voxels corresponding to structure k in the image $m \circ \phi$, which is m warped by the deformation field ϕ.

— $|A|$ denotes the cardinality (number of elements) of set A.

— $\cap$ represents the intersection operation between two sets.

— ϕ is the deformation field used for warping m to align with f.

The Dice score is commonly used to measure the similarity or overlap between two sets, in this case, the overlap between structures in the reference image f and the deformed image $m \circ \phi$. A higher Dice score indicates better alignment and registration performance. A Dice score of 1 indicates that the anatomy matches perfectly, and a score of 0 indicates that there is no overlap.

The segmentation loss L_{seg} over all structures $k \in [1, K]$ is defined as:

$$L_{\text{seg}}(S_f, S_m \circ \phi) = -\frac{1}{K} \sum_{k=1}^{K} \text{Dice}(S_f^k, S_m^k \circ \phi) \tag{4.9}$$

Where:

— $\text{Dice}(S_f^k, S_m^k \circ \phi)$ is the Dice score quantifying the volume overlap between structure k in the reference image f and the deformed image $m \circ \phi$.

— S_f^k represents the set of voxels corresponding to structure k in image f.

— $S_m^k \circ \phi$ represents the set of voxels corresponding to structure k in the image $m \circ \phi$, which is m warped by the deformation field ϕ.

— K is the total number of structures being considered.

The segmentation loss is used to guide the registration process by penalizing misalignments between the segmented structures in the reference image and the deformed image. By maximizing the Dice scores for all structures, the registration algorithm aims to achieve accurate and consistent alignment. The segmentation loss L_{seg} alone does not encourage smoothness and agreement of image appearance, which are essential for good registration. Therefore, we combine L_{seg} with the VoxelMorph Loss L_{us} (see Equation 2) to obtain the objective:

$$L_{\text{a}}(f, m, S_f, S_m, \phi) = L_{\text{seg}}(f, m, \phi) + \lambda L_{\text{us}}(S_f, S_m \circ \phi) \tag{4.10}$$

Where:

— L_{seg} is the segmentation loss, quantifying the discrepancy between segmented structures in the reference image and the deformed image.

— L_{us} is the VoxelMorph Loss, as defined in Equation 4.2), which encourages smoothness in the deformation field and agreement of image appearance.

— λ is a parameter representing the trade-off between the two losses, controlling their relative importance in the overall objective.

By combining the segmentation loss with the VoxelMorph Loss, the registration algorithm aims to strike a balance between accurate alignment of structures and smoothness in the deformation field. This approach improves the robustness and quality of the registration process, leading to more effective and reliable results.

4. 6 Experiments and Result

The primary objective of our research is to achieve precise alignment between these synthetic moving images and the fixed images from the BraTS dataset. To accomplish this challenging task, we have chosen to employ VoxelMorph, a state-of-the-art deep learning model renowned for its prowess in medical image registration tasks. By leveraging the capabilities of VoxelMorph, we aim to not only align the images accurately but also expedite the registration process, which is crucial in clinical settings.

4. 6. 1 Experimental Setup

In our comprehensive brain image registration experiments, we leverage the extensive and invaluable BraTS 2021 dataset, renowned in the medical imaging community for its diverse collection of brain tumor images. This dataset serves as our fixed images, providing a robust foundation for evaluating the accuracy and efficacy of our registration techniques. To introduce moving images into our experiments, we employ the powerful SimpleITK library to generate synthetic brain images. These synthetic moving images are crafted with precision, allowing us to control various parameters and simulate different imaging scenarios, enhancing the versatility of our experiments.

4. 6. 2 Evaluation Metrics

To evaluate the quality of our registration results, we employ a set of robust metrics, including the Dice score, which assesses the spatial overlap between registered images, and negative cross-correlation (NCC), which quantifies the degree of alignment of image intensities in a localized context. In addition to these metrics, we also utilize the mean squared error (MSE) to assess the overall accuracy of registration. By carefully analyzing these metrics, we gain insights into the effectiveness and precision of our brain image registration approach.

4. 7 Result

Our brain image registration experiments yielded remarkable results that demonstrate the effectiveness of our approach. The MSE (Mean Squared Error) loss, a fundamental measure of registration accuracy, was impressively low at 0.002. This result indicates that the registered images closely align with their target counterparts from the BraTS dataset, showcasing the precision of our method.
Additionally, the NCC (Negative Cross-Correlation) loss, which gauges the similarity in image intensities, reached approximately -8.49. This negative value signifies a strong alignment of image intensities between the registered images and the BraTS dataset images. The magnitude of the NCC value can vary, but a higher

(closer to 1 or -1) NCC value indicates a better alignment of intensity patterns. Our result demonstrates a high degree of alignment in terms of image intensities, which is a positive outcome for the accuracy of our registration method.

To provide a visual representation of the registration process and its impact, we present the following plots (see 4.7 , 4.8).

In addition to these plots, we provide a representative test image that showcases the results of our registration

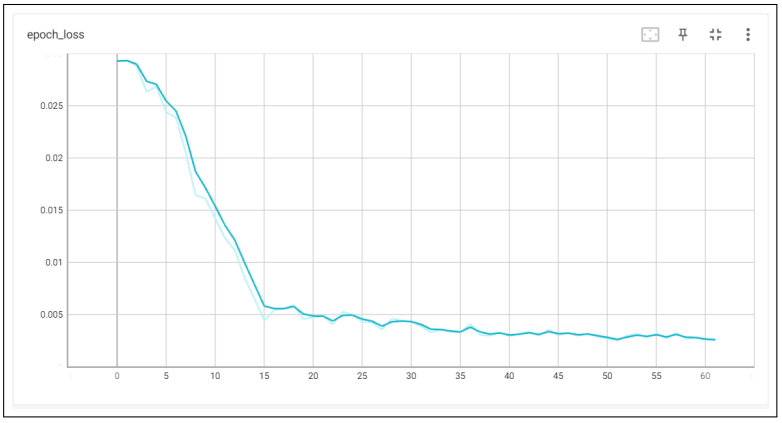

FIGURE 4.7 – This plot illustrates the decrease in the Mean Squared Error loss over the course of registration, showcasing the progressive improvement in alignment accuracy.

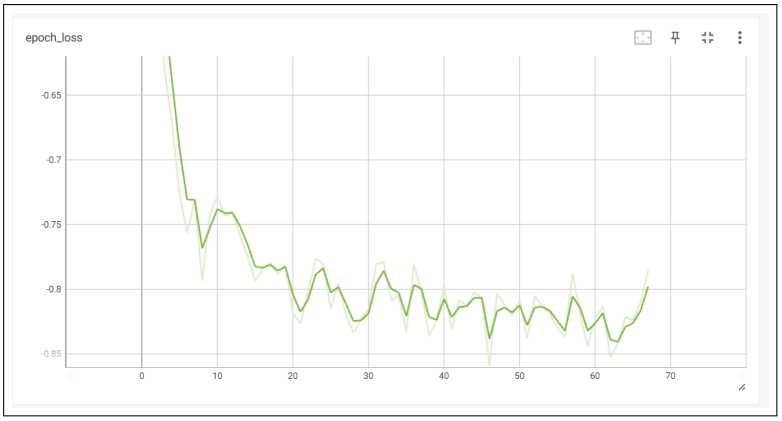

FIGURE 4.8 – The Negative Cross-Correlation plot highlights the increase in similarity between image intensities as registration proceeds, reinforcing the effectiveness of our approach.

process, emphasizing the successful alignment achieved (see 4.9 ,4.10).

These results underscore the proficiency of our brain image registration technique, which can play a pivotal role in medical imaging applications, enabling more accurate diagnosis and treatment planning by aligning images with exceptional precision.

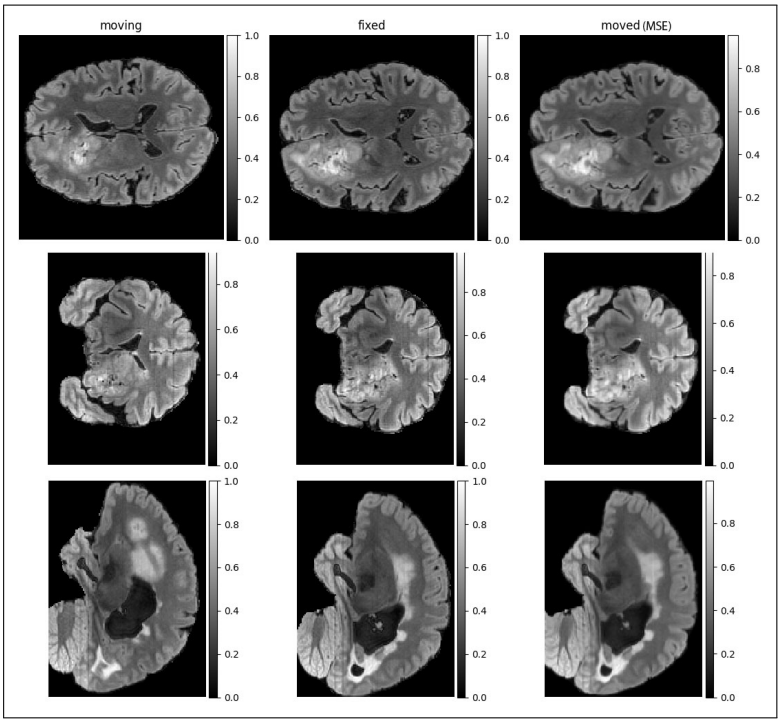

FIGURE 4.9 – The test result of voxelmorph algorithm using Mean Squared Error (MSE) loss function .

4. 8 Conclusion

In conclusion, our research presents a robust and efficient approach to brain image registration, a crucial task in the field of medical imaging. Leveraging the capabilities of VoxelMorph and a meticulously crafted experimental setup, we have demonstrated the precision and effectiveness of our registration technique. The low Mean Squared Error (MSE) loss of 0.002 and the strong Negative Cross-Correlation (NCC) alignment close to -8.49 underscore the success of our method in aligning synthetic moving images with the BraTS dataset. These results hold great promise for improving the accuracy and efficiency of medical image analysis, particularly in the context of brain tumor diagnosis and treatment planning. Our approach not only contributes to the field of medical imaging but also has the potential to positively impact clinical practices by providing clinicians with aligned images that aid in more accurate assessments and diagnoses. As we look ahead, further refinements and applications of our registration technique are poised to enhance medical image analysis and ultimately benefit patients and healthcare professionals.

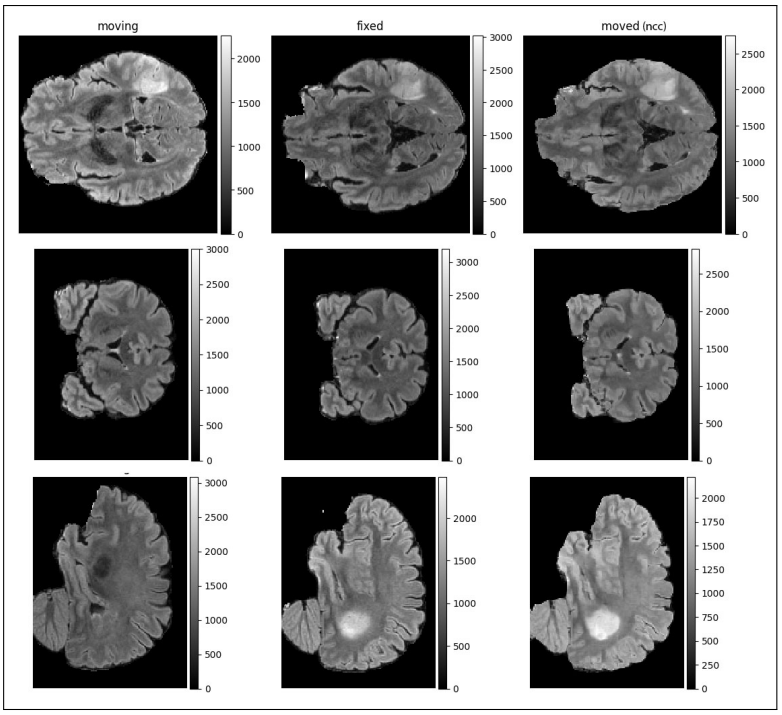

FIGURE 4.10 – The test result of voxelmorph algorithm using Negative Cross-Correlation (NCC) loss function

CHAPTER 5

3D U-NET: FOR IMAGE MEDICAL SEGMENTATION

Preambule

This chapter presents the 3D U-Net model's application in medical image segmentation, with a specific focus on tumor segmentation methods. It delves into traditional and deep learning-based techniques. The chapter discusses data preprocessing, model construction, and provides insights into experiments and results. Concluding remarks on 3D U-Net's effectiveness in medical image segmentation are also provided

5. 1 Introduction

Within the realm of medical image processing, the precise identification and segmentation of tumors emerge as critical pursuits. In this chapter, we embark on an exploration of tumor segmentation techniques, where we navigate both traditional methodologies grounded in mathematical models and algorithms and the cutting-edge frontier of deep learning. The journey encompasses a deep dive into various methods, their applications, and the pivotal role data preprocessing plays in enabling accurate segmentation. Ultimately, we culminate in the deployment of the U-Net model for brain tumor segmentation, unveiling its remarkable potential in revolutionizing medical image analysis. This chapter underscores the transformative impact of advanced technologies on healthcare, paving the way for enhanced diagnostics and treatment strategies.

5. 2 Tumor segmentation methods

Image segmentation and image classification are two leading pillars of image processing. There are number of techniques that have been used for the segmentation and classification purpose. Medical image

segmentation is a procedure to find the region of interest or dividing the image into different regions or distinguishing foreground and background based on the pixel similarities from the 2D or 3D images taken in different modalities; MRI, X-ray, CT, Microscopy, Endoscopy and many other. Due to the different human anatomy and high variability, medical image segmentation or labelling is a major challenge. General classification of image segmentation methods are :

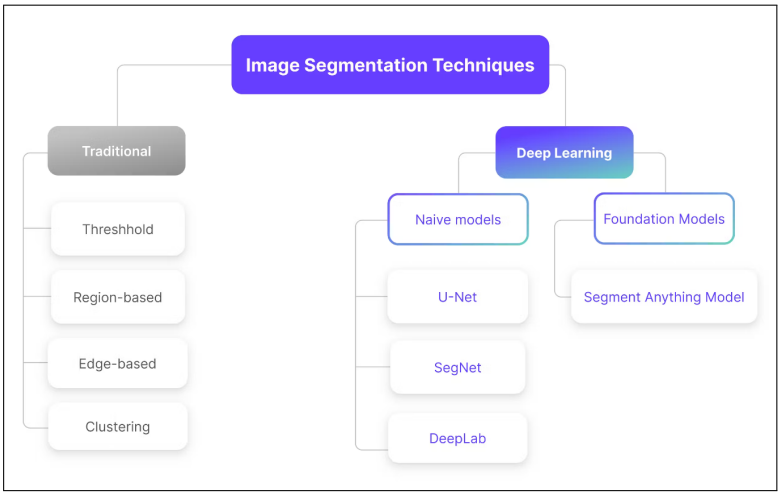

FIGURE 5.1 – Image Segmentation Techniques

5. 2. 1 Traditional Techniques

Traditional image segmentation techniques have been employed in the field of computer vision for decades to extract valuable information from images. These techniques are rooted in mathematical models and algorithms designed to identify regions within an image that share common characteristics, such as color, texture, or brightness. Traditional image segmentation methods are typically computationally efficient and relatively straightforward to implement. They are often utilized in applications that require rapid and accurate image segmentation, such as object detection, tracking, and recognition. In this section, we will explore some of the most commonly used techniques.

5. 2. 1. 1 Thresholding

Thresholding is one of the simplest image segmentation methods. Here, pixels are classified into different categories based on their intensity in the image histogram relative to a predefined threshold value. This method is suitable for segmenting objects when there is a significant difference in pixel values between the two target classes. In images with low levels of noise, a constant threshold value can be used, but for noisy images, dynamic thresholding performs better. In thresholding-based segmentation, the grayscale

image is divided into two segments based on their relationship to the threshold value, resulting in binary images. Commonly used thresholding methods include Global Thresholding and Adaptive Thresholding.

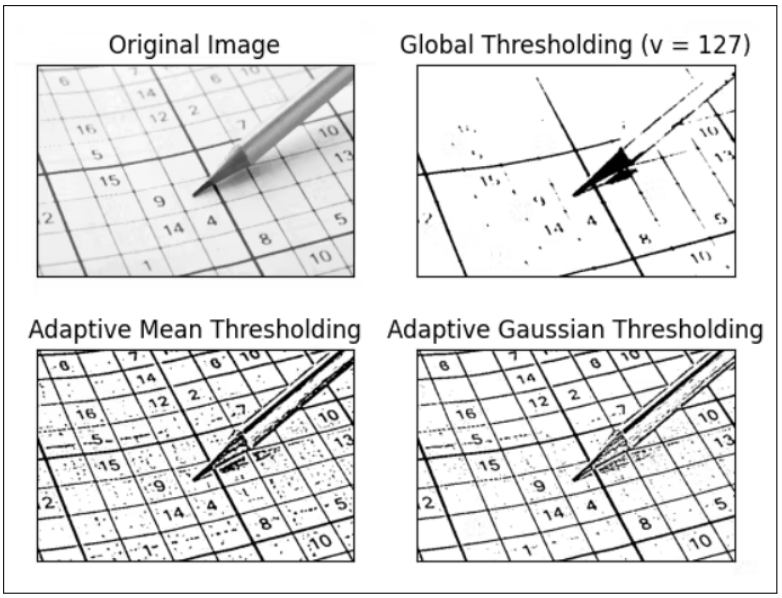

FIGURE 5.2 – Image showing different thresholding techniques.

5. 2. 1. 2 Region-based Segmentation

Region-based segmentation is a technique used in image processing to partition an image into regions based on similarity criteria, such as color, texture, or intensity. This method involves grouping pixels into regions or clusters based on their similarities and then iteratively merging or splitting regions until the desired level of segmentation is achieved. Two commonly used region-based segmentation techniques are Split and Merge Segmentation and Graph-based Segmentation.

5. 2. 1. 3 Edge-based Segmentation

Edge-based segmentation is a technique used in image processing to identify and isolate the edges of an image from the background. This method involves detecting abrupt changes in intensity or color values of pixels in the image and using them to delineate object boundaries. The two most common edge-based segmentation techniques are as follows:

— **Canny edge detection**: This is a popular method for edge detection that utilizes a multi-stage algorithm to detect edges in an image. The process includes smoothing the image using a Gaussian filter, computing the gradient magnitude and direction of the image, applying non-maximum suppression to refine the edges, and using hysteresis thresholding to remove weak edges5.4.

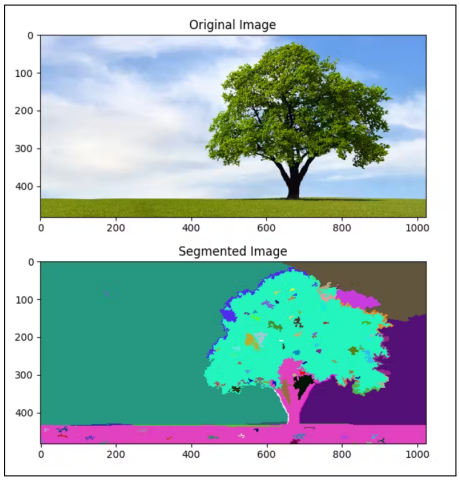

FIGURE 5.3 – Region-based Segmentation

— **Sobel edge detection**: This method employs a gradient-based approach to detect edges in an image. It calculates the gradient magnitude and direction of the image using a Sobel operator, which is a convolution kernel that separately extracts horizontal and vertical edge information5.5.

— **Laplacian of Gaussian (LoG) edge detection**: LoG edge detection is a method that combines Gaussian smoothing with the Laplacian operator. It involves applying a Gaussian filter to the image to reduce noise and then applying the Laplacian operator to highlight the edges. LoG edge detection is a robust and accurate method for edge detection, but it can be computationally expensive and may not perform well with images featuring complex edges5.6.

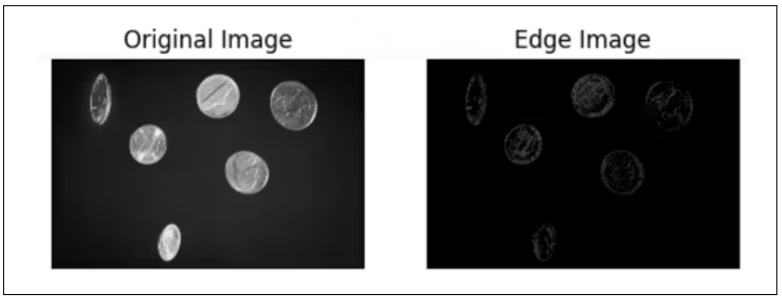

FIGURE 5.4 – Example of Edge-based Segmentation Techniques

5. 2. 1. 4 Clustering

Clustering is one of the most popular techniques used for image segmentation, as it allows the grouping of pixels with similar characteristics into clusters or segments. The core concept behind clustering-based segmentation is to group similar pixels into clusters, with each cluster representing a distinct segment. This

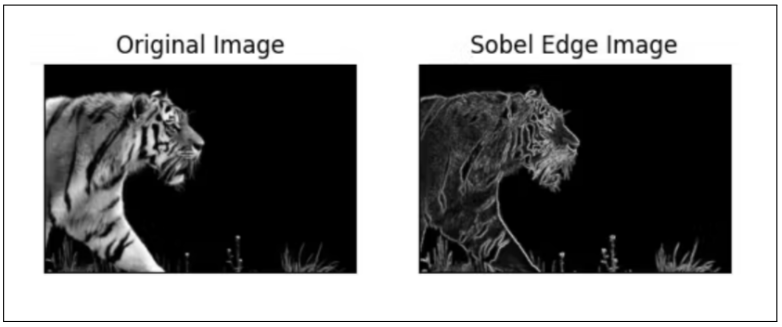

FIGURE 5.5 – Example of solde Edge-based Segmentation Techniques

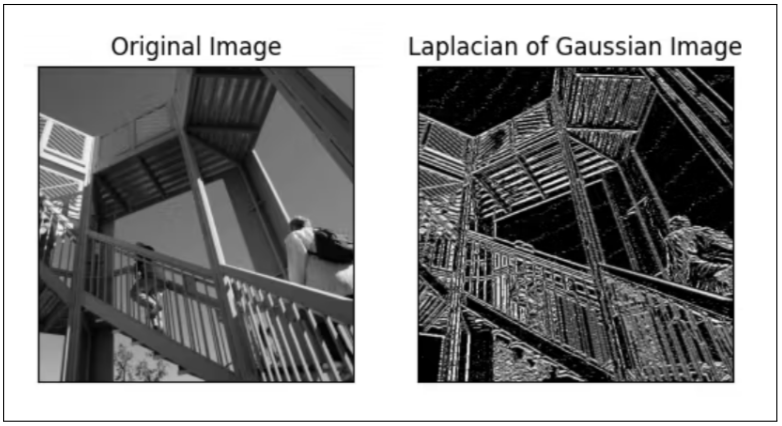

FIGURE 5.6 – Example of Laplacian Edge-based Segmentation Techniques

can be achieved using various clustering algorithms, including K-means clustering, mean-shift clustering, hierarchical clustering, and fuzzy clustering.

5. 2. 2 Deep Learning Techniques

Neural networks offer innovative solutions for image segmentation by training models to discern significant features within images, eliminating the need for custom functions as seen in traditional algorithms. Segmentation tasks in neural networks typically employ an encoder-decoder architecture.

The encoder efficiently extracts image features through deeper and narrower filters. If the encoder has been pretrained on tasks like image or facial recognition, it leverages that prior knowledge to extract features relevant to segmentation (a technique known as transfer learning). The decoder, on the other hand, expands the encoder's output into a segmentation mask that matches the input image's pixel resolution through a series of layers.

Numerous deep learning models excel in the realm of segmentation. Let's explore a few notable ones:

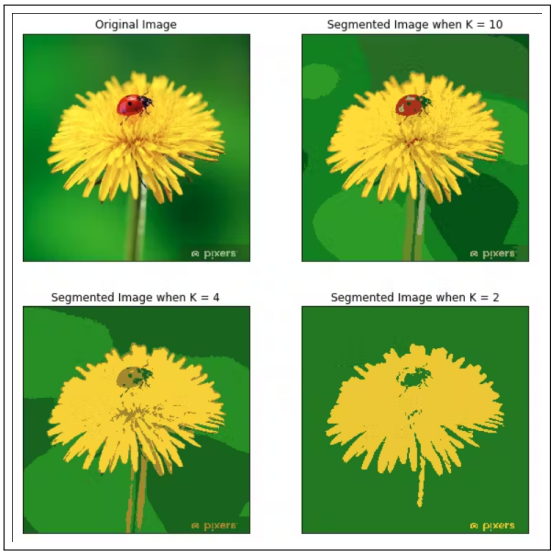

FIGURE 5.7 – Example of K-mean clustering

5. 2. 2. 1 U-Net

U-Net represents a modified [**?**], fully convolutional neural network initially developed for medical applications, specifically the detection of tumors in the lungs and brain. Notably, it shares the same encoder and decoder components. Unlike traditional fully convolutional networks that employ upsampling for feature extraction, U-Net incorporates shortcut connections in its architecture to mitigate information loss. These shortcuts allow high-level features to be concatenated with low-level ones, enabling the network to deliver more precise segmentation results5.8.

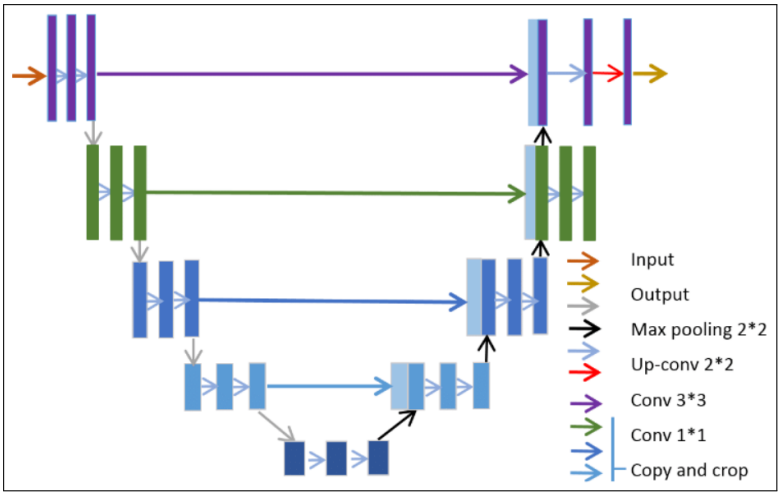

FIGURE 5.8 – U-Net Architecture

5. 2. 2. 2 SegNet

SegNet, another fully convolutional network designed primarily for semantic pixel-wise segmentation, also employs an encoder-decoder structure. What sets SegNet apart is its unique approach to decoder-based feature upsampling. It utilizes pooling indices computed during max-pooling, which, in turn, facilitates non-linear upsampling by the encoder. This innovative approach eliminates the need to learn the upsampling process, making SegNet particularly suitable for scene-understanding applications5.9.

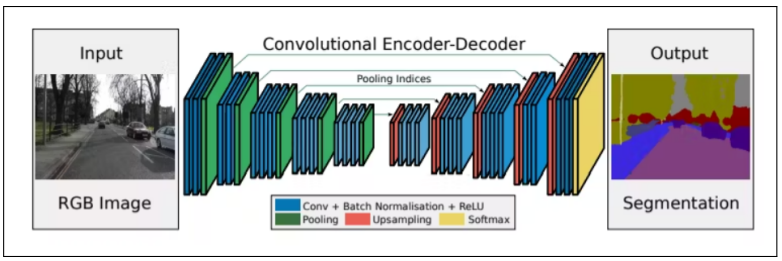

FIGURE 5.9 – SegNet Architecture

5. 2. 2. 3 DeepLab

DeepLab is primarily a convolutional neural network (CNN) architecture designed for image segmentation. Distinguishing itself from the previous models, DeepLab leverages features from every convolutional block and concatenates them with their corresponding deconvolutional blocks. The network utilizes the atrous convolution (dilated convolution) method for upsampling. This technique not only reduces computation costs but also captures more information, enhancing its segmentation capabilities5.10.

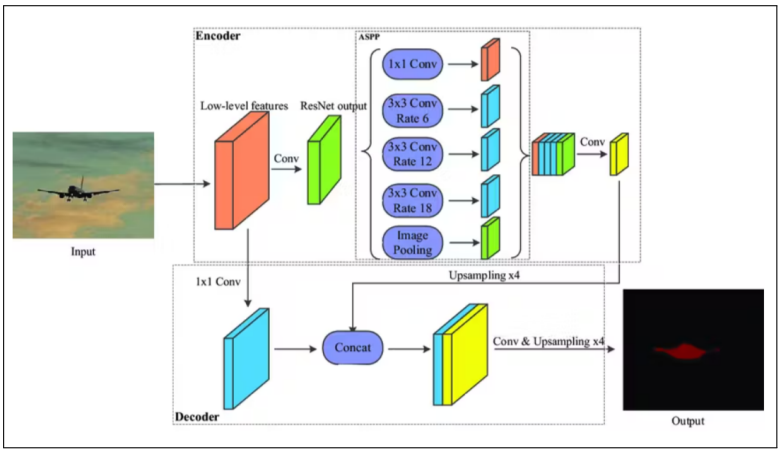

FIGURE 5.10 – DeepLab Architecture

5. 2. 2. 4 Foundation Model Techniques

Foundation models have also found application in image segmentation, a task that involves dividing an image into distinct regions or segments. Unlike language models, which typically rely on transformer architectures, foundation models for image segmentation predominantly employ convolutional neural networks (CNNs) tailored for handling image data.

5. 2. 2. 5 Segment Anything Model

The Segment Anything Model (SAM) stands as a pioneering foundation model designed specifically for image segmentation. SAM is trained on the most extensive segmentation dataset to date, encompassing over 1 billion segmentation masks. It is engineered to generate valid segmentation masks for a wide range of prompts, including foreground/background points, rough boxes or masks, freeform text, or general instructions indicating what to segment in an image. SAM's architecture combines an image encoder, which produces a one-time embedding for the image, with a lightweight encoder capable of converting any prompt into a real-time embedding vector. These two sources of information are then fused in a lightweight decoder that predicts segmentation masks5.11.

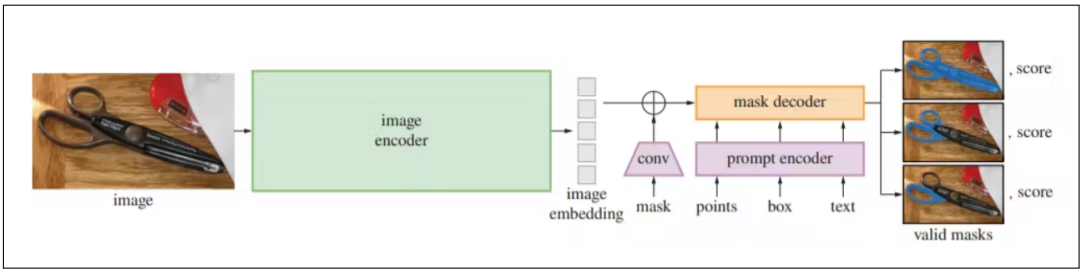

FIGURE 5.11 – Segment Anything Model Architecture

5. 3 Data Preprocessing

The preprocessing steps for the BraTS data involved several key operations. First, the data was understood, and then it was cropped from its original dimensions of (240, 240, 155) to (176, 176, 128). This cropping step was performed to reduce the computational complexity and focus on the relevant brain region. Next, a min-max scaling function was applied to normalize the intensity values of the MRI modalities (FLAIR, T1ce, T2). This step ensured that the intensity values were within a specific range, making the data more comparable and facilitating subsequent analysis. After scaling, the FLAIR, T1ce, and T2 modalities were combined into a single representation. This fusion of modalities allowed for a comprehensive analysis of the brain tumor, utilizing the information captured by each modality.

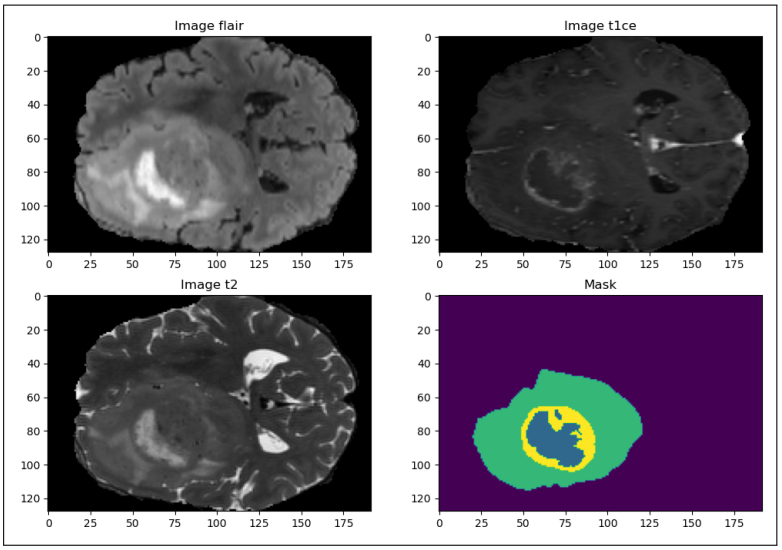

FIGURE 5.12 – BraTs 2021 Dataset after Preprocessing

Finally, the preprocessed data was saved with a .npy extension, indicating that it was stored in the NumPy array format. Additionally, the data was split, into training and validation sets, to facilitate the development and evaluation of machine learning models for tumor segmentation.
These preprocessing steps were performed to prepare the BraTS data for further analysis and segmentation of brain tumors. They helped optimize the data for subsequent modeling and classification tasks, enabling more accurate and efficient analysis of brain tumor images.

5. 4 Chose and build the model

For the training of the BraTS data, I chose to employ the U-Net model, a popular architecture commonly used for medical image segmentation tasks. The U-Net architecture is particularly suitable for this type of task due to its ability to capture both local and global information while preserving spatial details [**?**].
The U-Net model consists of an encoder-decoder structure. The encoder part of the network is responsible for capturing the high-level features from the input images. It consists of a series of convolutional and pooling layers, which progressively downsample the input, enabling the model to learn abstract representations of the data.
On the other hand, the decoder part of the U-Net performs upsampling to reconstruct the segmented output. It consists of a series of convolutional and upsampling layers, which gradually increase the spatial resolution while incorporating the learned features from the encoder. The skip connections between corresponding encoder and decoder layers help to preserve spatial information and aid in precise localization of the tumor boundaries

5. 5 Experiments and Result

In the pursuit of accurate brain tumor segmentation, a rigorous series of experiments were conducted using the preprocessed BraTS data and the U-Net model. The model was trained using the training dataset, and its performance was evaluated on the validation set. During training, the model exhibited remarkable learning capabilities, steadily improving its segmentation accuracy over epochs. Notably, the Dice score, a metric commonly used to assess the overlap between predicted and ground truth segmentations, achieved an impressive score of 89.3%. This high Dice score underscores the model's proficiency in accurately delineating tumor boundaries.

Moreover, the model displayed exceptional generalization to new data, with an accuracy of 98.8% on the validation set. The accuracy metric provides insights into the model's ability to correctly classify each voxel within the brain images, further validating its robustness. The loss function, a crucial indicator of model convergence, reached an impressively low value of 0.02, signifying that the model learned to minimize the error between predicted and ground truth segmentations effectively. To provide a visual representation of the training process, the following accuracy and loss plots are included: Additionally, the Mean Intersection

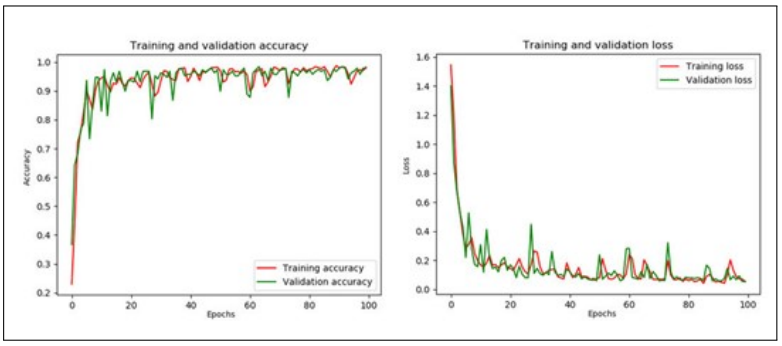

FIGURE 5.13 – These plots vividly illustrate the model's learning trajectory, showing a consistent decrease in loss and a corresponding increase in accuracy throughout the training process.

over Union (IoU) score, which measures the overlap between predicted and ground truth segmentations, achieved an outstanding score of 86.6%. This metric further reinforces the model's exceptional segmentation performance.

Below -Figure : 5.14- is an example images displaying the model's predictions (in color) overlaid on the corresponding test images (in grayscale) for a representative case; This images visually showcases the model's segmentation results, highlighting its ability to accurately identify tumor regions within the brain images. The combination of these outstanding results and the visual representation of the model's performance underscores the U-Net's efficacy in medical image segmentation tasks, particularly in the context of brain tumor analysis.

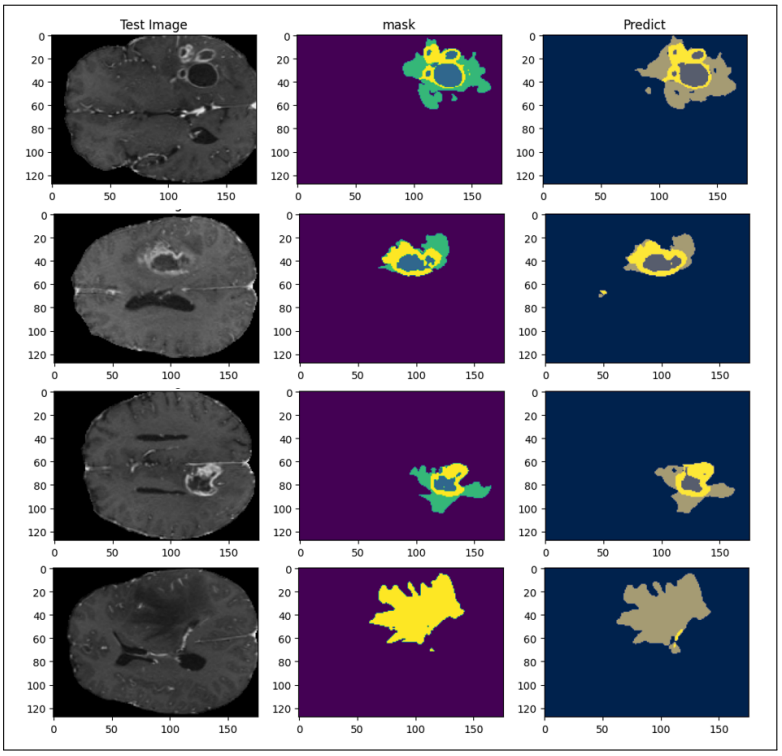

FIGURE 5.14 – Predictions vs Test Image

5. 6 Conclusion

In conclusion, this study highlights the potential of deep learning, particularly the U-Net architecture, in the critical task of brain tumor segmentation. The remarkable accuracy, low loss, and impressive IoU score validate the model's effectiveness in medical image analysis. Our findings hold promise for enhancing the diagnostic and treatment planning processes in neurology by providing accurate and efficient tools for brain tumor localization and characterization. As we move forward, further research and refinements in this domain are poised to have a transformative impact on healthcare, ultimately benefiting patients and medical professionals alike.

Bibliography