Democratic and Popular Republic of Algeria
Ministry of Higher Education and Scientific Research
University A. Mira of Bejaïa
Faculty of Exact Sciences
Computer science Department

MASTER THESIS

Option : Artificial Intelligence

2D/3D medical image segmentation by embedding EfficientNet in Convolutional neural network: Application on BraTS challenge 2020

Presented by

M. ALLAOUI Mohamed Lamine M. ZETOUT Ahcene

Evaluated by the jury composed of

Supervisor :	M. BELAID A.	Dr	University of Béjaïa
Member :	M^s ALOUI S.	Dr	University of Béjaïa
Member :	M. KHANOUCHE M. E.	Dr	University of Béjaïa

2019 - 2020

Acknowledgements

First and foremost, we thank Allah Who helped us accomplish this work, and Who has been with us at all times of our lives.

We especially thank our mentor, Dr. BELAID Ahror, who gave us an excellent opportunity to carry out this work, which allowed us to acquire a lot of knowledge and experience in the field of machine learning and the deep learning. His patience, his motivation and his deep knowledge always guide us.

We also thank the members of the jury composed of Dr. ALAOUI Soraya and Dr. KHANOUCHE mohamed essaid for the honour to evaluate our work.

We would also like to thank all our professors and our comrades at the University of Bejaia.

A special thank you to our friends in particular "Idriss", "Mohand" for the moment that we shared together. Last but not least, thank you to our families for always being with us every step of the way.

Contents

List of Figures

List of Tables

List of Abbreviations

AI	Artificial Intelligence
BraTS	Brain Tumor Segmentation
CNN	Convolutional Neural Networks
CT	Computerized Tomography
DL	Deep Learning
ET	Enhancing Tumor
HGG	High Grade Glioma
LGG	Low Grade Glioma
ML	Machine Learning
MRI	Magnetic Resonance Imaging
NLP	Natural Language Processing
NN	Neural Network
TC	Tumor core
WT	Whole Tumor

Introduction

In recent years, especially thanks to the development of information technologies and artificial intelligence (AI), as well as through functional medical devices, the medical community has experienced a real rise in power in clinical practice. One of the main challenges today is the analysis and interpretation of medical images by experts. Indeed, the processing of medical images is a multidisciplinary research axis, involving fundamental sciences such as computer science, medical fields and mathematics. It is a very popular industry because it offers a wide range of possibilities that can help doctors improve the diagnosis and treatment of patients. Medical image segmentation is the process of identifying organs or lesions from medical imaging modalities as Computerized Tomography (CT) scans or Magnetic Resonance Imaging (MRI) images, which can convey basic information about the shape and volume of these organs. Previously, edge detection filters and mathematical methods were used to automate this process. However, with the advent of AI, deep learning has become a mainstream technology because of their powerful functions in image processing tasks. Deep learning for medical image segmentation has existed for a long time. Over the years, hardware improvements have made it easier for hospitals all over the world to use it. Convolutional Neural Network (CNN) is one of the DL architectures used for this process. CNN's are composed of layers which are not fully connected: they have filters, sets of cube-shaped weights that are applied throughout the image. The general idea of CNN is to use a given input data (2D, 3D) and apply a successive filters. In the context of data analysis, most machine learning algorithms are based on the assumption that the data set used for training and the test data set belong to the same descriptor space and follow the same distribution of probabilities. However, this is not the case in many applications, and Transfer Learning can be used to optimize data-intensive and computationally expensive model retraining, especially in deep learning. Transfer learning is a new method in which the model is trained by using the latest pre-trained model, and the last few layers are frozen to learn weights specific to the problem. Low-level functions are usually borrowed from ImageNet. Another method is to use 2.5D CNN, which can process a certain amount of spatial information. They can strike a good balance between performance and computational cost. Compared with 2.5 CNN, 3D CNN provides better performance and handles richer spatial information.

Clinical motivation

Today the biggest problem in the field of medical imaging is the lack of data because this data is often confidential and we cannot have access to this data. Applying deep learning algorithms to segment a few unlabeled images seems therfore impossible. Nowadays, MRI represents a very important source of information due to a large number of medical collections (mainly 2D and 3D). However, like any physical device, certain constraints will appear during MRI exams. Such as the inequality of intensity and the presence of noise. These phenomena are common and may embarrass the physician when delimiting the target area.

Technical Motivation

The aim of our work is to study how to reuse the powerful neural networks that exist in the literature. Indeed, there are many powerful and well-trained classification networks on a very large database. Through transfer learning, these pre-trained networks can be easily reused on other classification problems. However, transferring the learning and detection power of the functionality of a classification network to another type of problem such as segmentation is not easy; we will try to solve this problem through the transfer learning technique.

Chapter 1

Medical images

1.1 Introduction

Image processing is a very vast field which has known and continues to undergo significant developments in recent decades. However, digital image processing specifies all the techniques likely to modify digital images to improve or extract information from them.

In this chapter, we will define the image as well as its characteristics and types, then we will introduce ourselves into medical imaging which are the type of image that we will be dealing with throughout our work.

1.2 Definition of digital image

A digital image is an image made up of picture elements (also called pixels). The intensity or gray level of each picture element has a finite and discrete digital representation. This is the output of its two-dimensional function. The space represented by x and y on the x axis and y axis, respectively. Depending on whether the resolution of the image is fixed, it can be vector or raster. The term digital image generally refers to raster images or bitmaps (as opposed to vector images).

1.3 Types of image

In general, there are 3 main types of digital images:

1.3.1 Binary image

The binary image is a black and white image with pixel values taking 0 or 1. This kind of image is rarely used and is often called bit map image. Binary images are used as masks for indicating the pixels of interest in many image processing tasks [1]. Figure 1.1 is an example of binary image.

Figure 1.1: binary image.

1.3.2 Grayscale image

In digital photography, computer generated imaging and colorimetry, grayscale or grayscale images are images in which the value of each pixel is a single sample that represents only the amount of light. In other words, it only carries intensity information. Grayscale images (a monochromatic black and white or a gray) are made up of grayscale only. The contrast ranges from the weakest black to the strongest white [1].

Grayscale images are different from one-bit two-color black and white images, which are images with only two colors in the context of computer imaging: black and white (also known as two level or binary images). There are many shades of gray between grayscale images [1].

A grayscale image can be the result of measuring the light intensity of each pixel under a specific weighted frequency (or wavelength) combination. In this case, when "captures only one frequency (actually a small frequency). Frequency principle the rod can originate from any position in the electromagnetic spectrum (such as infrared, visible light, ultraviolet, etc.) [1]. An illustration in figure 1.2 showing a grayscale image.

Figure 1.2: grayscale image.

1.3.3 RGB image

RGB (Red, Green, Blue) refers to three shades that can be mixed to create different colors. Combining red, green and blue light is the standard method for generating color images on screens such as televisions, computer monitors, and smartphone screens.

The RGB color model is an "additive" model. When each color is mixed 100%, white light will be produced. When 0% of each color is combined, no light is generated and black is generated. It is sometimes contrasted with CMYK (cyan, yellow, magenta, and black), which is the standard color palette used to create printed images. CMYK is a "subtractive" color model because the colors get darker when combined. In CMYK color model, mixing 100% of each color will produce black, and 0% of each color will cause white [2].

The number of colors supported by RGB depends on how many possible values can be used for red, green, and blue. This is known as "color depth" and is measured in bits. The most common color depth is 24-bit color, also known as "true color." It supports eight bits for each of the three colors, or 24 bits total this provides 28, or 256 possible values for red, green, and blue [2].

Figure 1.3 is an illustration showing a RGB image.

Figure 1.3: RGB (colored) image.

1.4 Characteristics of an image

Quality of an image is greatly affected by many parameters. A few of the most common parameters deal with the quality of image are described below.

1.4.1 Dimension

This is the size of the image, which is presented in the form of a matrix whose elements are numerical values representative of light intensities (pixels). The number of rows of this matrix multiplied by the number of columns gives us the total number of pixels in a 2D image, this number is multiplied by the number of layers to have the total number of pixels in a 3D image.

1.4.2 Resolution

The resolution of remote sensed raster data can be characterized in several different ways. There are two primary types of "resolution" for an image.

1.4.2.1 Spatial Resolution

Spatial resolution is a term that refers to the number of pixels that are used to construct a digital image. When we say that a digital image has higher spatial resolution than another image, it means that the higher spatial resolution image is composed by more pixels than the lower spatial resolution image for the same dimensions of the imaging part.

Spatial resolution can determine the quality of an image and describe how detailed an object can be represented by the image. It is a measurement to determine how small an object should be in order for an imaging system to detect it. In medical imaging, the term spatial resolution can be used to describe the imaging resolution. Spatial resolution of a medical imaging system is the ability of the system to depict microstructures [3].

Figure 1.4 show the different spatial resolution.

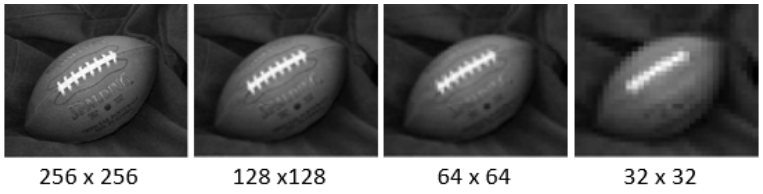

Figure 1.4: Different spatial resolution.

1.4.2.2 Tonal Resolution

Is meant the difference between the appearance of the feature and the background in brightness and/or color, there is an illustration in figure 1.5 showing that.

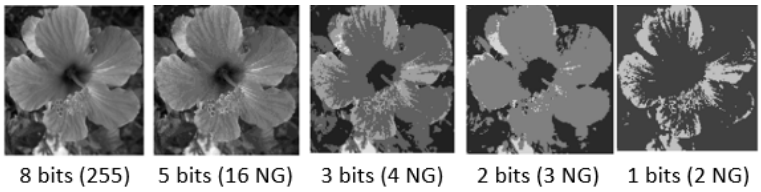

Figure 1.5: Different tonal resolution.

1.4.3 Image noise

Noise in an image is considered to be a phenomenon of sudden variation in the intensity of a pixel relative to its neighbors, it comes from the lighting of the optical and electronic devices of the sensor.

1.4.4 Histogram

The grayscale or color histogram of an image is a function that shows how often each grayscale (color) occurs in the image. It gives a great deal of information on the distribution of gray levels (color) and to see between which limits are distributed the majority of gray levels (color) in the case of an image that is too light or too dark.

1.4.5 Luminance

This is the degree of brightness of the points in the image. It is also defined as being the quotient of the luminous intensity of a surface by the apparent area of this surface, for a distant observer, the word luminance is substituted for the word brightness, which corresponds to the brightness of an object. Good luminance is characterized by: Bright (brilliant) images. Good contrast: images should be avoided where the range of contrast tends towards white or black; these images cause loss of detail in dark or bright areas.

1.4.6 Contrast

It is the marked opposition between two regions of an image, more precisely between the dark regions and the light regions of that image. Contrast is defined according to the luminance of two image areas. If L1 and L2 are the degrees of brightness respectively of two neighboring areas A1 and A2 of an image, the contrast C is defined by the ratio :

$$C = \frac{L1 - L2}{L1 + L2}$$

1.4.7 Entropy

It is a quantity which is used to describe the amount of information which must be coded by a compression algorithm, Low entropy images, such as those containing a lot of black sky, have very little contrast and large runs of pixels with the same or similar DN values. An image that is perfectly flat will have an entropy of zero. Consequently, they can be compressed to a relatively small size. On the other hand, high entropy images such as an image of heavily cratered areas on the moon have a great deal of contrast from one pixel to the next and consequently cannot be compressed as much as low entropy images [4].

And we can see an illustration of effect of some parameters on an image in figure 1.6.

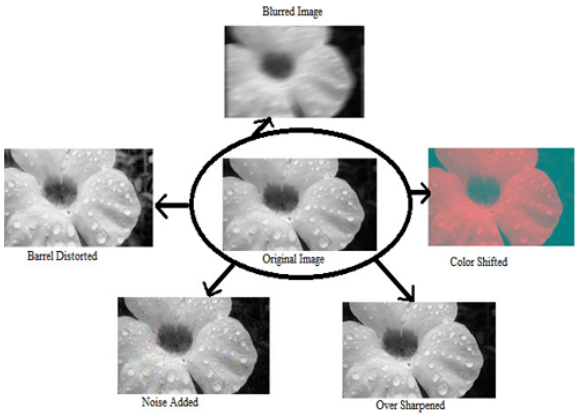

Figure 1.6: Effect of some parameters on an image.

1.5 Image processing system

We can see what digital image processing system is made up of in figure 1.7.

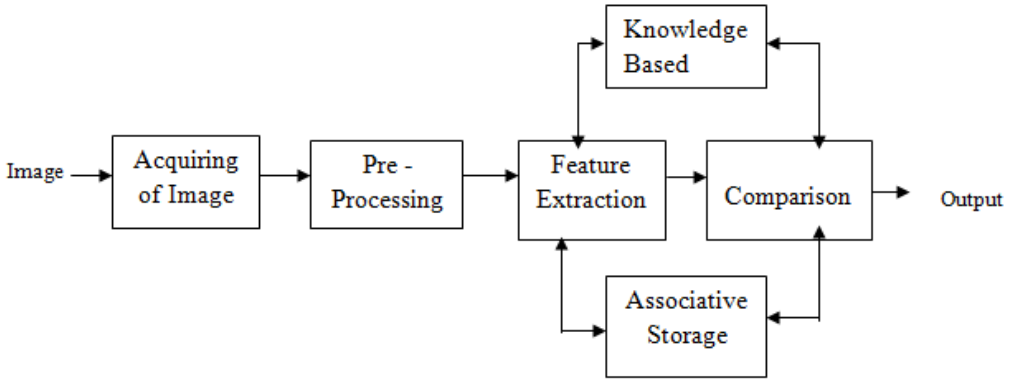

Figure 1.7: Digital image processing system diagram.

1.5.1 Image preprocessing

This stage is carried out immediately after the image is acquired, with the purpose of improving the quality of image segmentation. Processing time (CPU time) is very important. This is a decisive factor, and it must be as small as possible. This means that the operator must be local, that is, they must operate on a limited number of pixels, and actually must operate on pixels close to the current pixel.

We will present some of the most common preprocessing techniques and used in our work.

1.5.1.1 Segmentation

In this step, we are going to segment the image, separating the background from foreground objects.

1.5.1.2 Contour extraction

Edge detection is an initial step in many image analysis applications. Contours are indeed rich clues, just like points of interest, which can be used for subsequent image interpretation. The outlines in an image come from:

- Discontinuities of the reflectance function (texture, shadow);
- Depth discontinuities (edges of the object);
- Discontinuities of the intensity function in the images.

The classic principle of edge detection is based on the study of the derivative of the intensity function in the image: the local extremum of the gradient of the intensity function and the zero crossing of the Laplacian operator but there is one that interest us in our work and it is the canny edge detector an optimal filtering approach.

1.5.1.3 Canny edge detector

Optimal impulse response filter h(x) which satisfies the following three constraints:

- Good detection;
- Good localization;
- Sparsity of response.

There is a one parameter family of optimal filters, varying in the width of filter support σ. Detection improves and localization degrades as σ increases. As shown in figure 1.8.

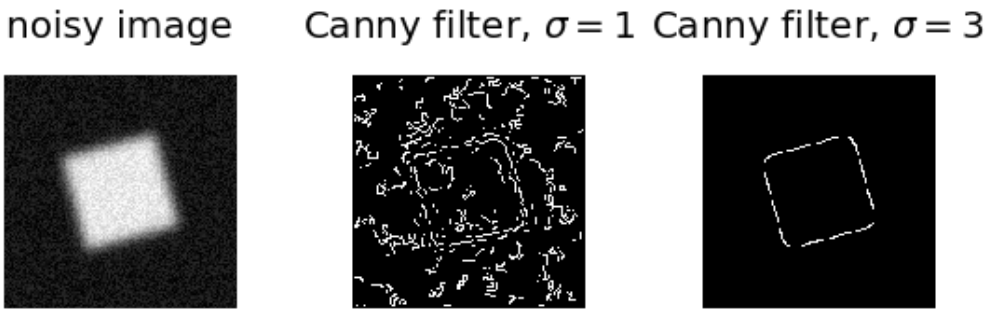

Figure 1.8: The application of the canny filter with two sigma values on an image.

1.6 Medical imaging

1.6.1 Definition of medical imaging

Medical imaging is often perceived to designate the set of techniques that noninvasively produce images of the internal aspect of the body. It is the techniques and processes used to create images of the human body for clinical purposes such as seeking to reveal, diagnose or examine injury, dysfunction or pathology. As a discipline and in its widest sense,it incorporates radiology, tomography, endoscopy, thermography, medical photography and microscopy (e.g. for human pathological investigations) [5].

In the clinical context, medical imaging is generally equated to radiology and the medical practitioner responsible for interpreting (and sometimes acquiring) the images is a radiologist. The radiographer or radiologic technologist is usually responsible for acquiring medical images of diagnostic quality, although some radiological interventions are performed by radiologists [5].

1.6.2 Types of medical images

1.6.2.1 Imaging using X-rays

X-ray imaging uses an X-ray beam that is projected on the body. When passing through the body, parts of the x-ray beam are absorbed. On the opposite side of the body, the X-rays are detected, resulting in an image [6].

1.6.2.2 Molecular Imaging

Molecular imaging provides detailed information of the biological processes taking place in the body at cellular and molecular levels and can indicate disease in its earliest stages [6].

1.6.2.3 Ultrasound imaging

Ultrasound imaging (sonography) is a diagnostic medical procedure that uses high-frequency sound waves to produce dynamic visual images of organs, tissues or blood flow inside the body. The sound waves are transmitted to the area to be examined and the returning echoes are captured to provide the physician with a live image of the area. Ultrasound does not require the use of ionizing radiation, nor the injection of nephrotoxic contrast agents [6].

1.6.2.4 Magnetic Resonance Imaging (MRI)

MRI systems use a powerful magnetic field and radiofrequency pulses to produce detailed images of the bodys internal structures as cross-sectional images or slices. Without exposing the patient or staff to ionizing radiation (X-rays), MRI provides high quality images with excellent contrast detail of soft tissue and anatomic structures such as gray and white matter in the brain. Its used in a wide range of examinations from brain tumors and inflammation of the spine to slipped discs, assessing blood flow and functioning of the heart. MRI does not emit any ionizing radiation [6].

1.6.3 Different formats of medical images

There are many format in medical imaging but we only going to define the 2 most popular.

1.6.3.1 Nifti

NIfTI (Neuroimaging Information Technology Project) is a data format used to store functional magnetic resonance imaging (fMRI) and other medical images. Nifti is a file format that was created by a committee of the National Institutes of Health in the early 2000s to create a neuroimaging format that maintains the advantages of the Analyze format but can solve its shortcomings. Nifti can actually be considered a revised Analyze format. This format fills in some unused/rarely used fields in the Analyze 7.5 header to store new information (such as image orientation) in order to avoid left-right ambiguity in brain research. In addition, Nifti also supports data types that are not considered in the analysis type, such as unsigned 16-bit. Although this format also allows the header and pixel data to be stored in separate files, the image is usually saved as a single ".nii" file in which the header and pixel data are combined. In the case of ".hdr" and ".img" data storage, the size of the header is 348 bytes, and in the case of a single ".nii" file, the size of the header is 352 bytes [7].

1.6.3.2 Dicom

The Dicom standard was established by the American College of Radiology and the National Electrical Manufacturers Association. Although the Dicom standard was born in 1993, it was only in the late 1990s that the Dicom standard was introduced into the imaging department. Today,

the Dicom standard has become the backbone of every medical imaging department. Generally, in terms of access, exchange, and availability of diagnostic medical images, the added value of using it is huge. Dicom is not only a file format, but also a network communication protocol, although these two aspects cannot be completely separated [7].

Below table 1.1 which summarizes the formats

Format	**Header**	**Extension**	**Data types**
Analyze	Fixed-length: 348 byte binary format	.img .hdr	Unsigned integer (8-bit), signed integer (16, 32-bit), float (32,64-bit), complex (64-bit)
Nifti	Fixed-length: 352 byte binary format a (348 byte in the case of data stored as .img and .hdr)	.nii	Signed and unsigned integer (from 8 to 64 bit), float (from 32 to 128 bit), complex (from 64 to 256 bit)
Minc	Extensible binary format	.mnc	Signed and unsigned integer (from 8 to 32 bit), float (32, 64 bit), complex (32, 64 bit)
Dicom	Variable length binary format	.dcm	Signed and unsigned integer, (8, 16 bit; 32 bit only allowed for radiotherapy dose), float not supported

Table 1.1: Summary of file formats characteristics[7].

1.6.4 Magnetic Resonance Imaging (MRI)

1.6.4.1 Definition

MRI (magnetic resonance imaging) is a noninvasive diagnostic test that takes detailed images of the soft tissues of the body. Unlike X-rays or CT, images are created by using a magnetic field, radio waves, and a computer. It allows your doctor to view your spine or brain in slices, as if it were sliced layer-by-layer and a picture taken of each slice. This test can help diagnose tumors, strokes, and disc herniation [8].

1.6.4.2 What does an MRI show

Nearly every part of the body may be studied with MRI. MRI gives very detailed pictures of soft tissues like the brain. Air and hard bone do not give an MRI signal so these areas appear black. Bone marrow, spinal fluid, blood and soft tissues vary in intensity from black to white, depending on the amount of fat and water present in each tissue and the machine settings used for the scan. The radiologist compares the size and distributions of these bright and dark areas to determine whether a tissue is healthy [8].

1.6.5 Brain

1.6.5.1 Definition of human brain

The human brain is the central organ of the human nervous system, and with the spinal cord makes up the central nervous system. The brain consists of the cerebrum, the brainstem and the cerebellum. It controls most of the activities of the body, processing, integrating, and coordinating the information it receives from the sense organs, and making decisions as to the instructions sent to the rest of the body. The brain is contained in, and protected by, the skull bones of the head [9].

1.6.5.2 Different part of brain

The human brain consists of three main parts:

- **The forebrain:**

  Greatly developed into the cerebrum, consists of two hemispheres joined by a bridge of nerve fibers, and is responsible for the exercise of thought and control of speech [10].

- **The midbrain:**

  The upper part of the tapering brainstem, contains cells concerned in eye movements [10].

- **The hindbrain:**

  The lower part of the brainstem, contains cells responsible for breathing and for regulating heart action, the flow of digestive juices, and other unconscious actions and processes. The cerebellum, which lies behind the brain stem, plays an important role in the execution of highly skilled movements [10].

Figure 1.9 is an illustration that shows the 3 parts of the brain.

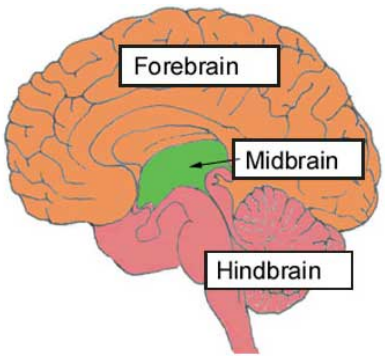

Figure 1.9: The main parts of the brain.

1.6.5.3 Brain tumor

A brain tumor is an abnormal growth of cells inside the brain or skull; some are benign, others malignant. Tumors can grow from the brain tissue itself (primary), or cancer from elsewhere in the body can spread to the brain (metastasis). Treatment options vary depending on the tumor type, size and location. Treatment goals may be curative or focus on relieving symptoms. Many of the 120 types of brain tumors can be successfully treated. New therapies are improving the life span and quality of life for many people [11].

1.6.5.4 What is a brain tumor

Normal cells grow in a controlled manner as new cells replace old or damaged ones. For reasons not fully understood, tumor cells reproduce uncontrollably [11]. A primary brain tumor is an abnormal growth that starts in the brain and usually does not spread to other parts of the body. Primary brain tumors may be benign or malignant [11]. A benign brain tumor grows slowly, has distinct boundaries, and rarely spreads. Although its cells are not malignant, benign tumors can be life threatening if located in a vital area [11]. A malignant brain tumor grows quickly, has irregular boundaries, and spreads to nearby brain areas. Although they are often called brain cancer, malignant brain tumors do not fit the definition of cancer because they do not spread to organs outside the brain and spine [11].

1.6.6 Skin

1.6.6.1 Definition of skin

The human skin is the outer covering of the body and is the largest organ of the integumentary system. The skin has up to seven layers of ectodermal tissue and guards the underlying muscles, bones, ligaments and internal organs [12].

1.6.6.2 The different layers of the skin

- **Stratum Corneum:**

  The stratum corneum (SC) or stratum corneum, is the outermost cell layer of the epidermis, the most superficial tissue of the skin. It is mainly composed of dead cells called corneocytes whose loss of the cell nucleus results from epidermal differentiation [12].

- **Epidermis:**

  The epidermis is the outermost of the three layers that make up the skin, the inner layers being the dermis and hypodermis. The epidermis layer provides a barrier to infection from environmental pathogens and regulates the amount of water released from the body into

the atmosphere through trans-epidermal water loss. The epidermis is composed of multiple layers of flattened cells that overlie a base layer (stratum basal) composed of columnar cells arranged perpendicularly [12].

- **Dermis:**

  The dermis or corium is a layer of skin between the epidermis (with which it makes up the cutis) and subcutaneous tissues that primarily consists of dense irregular connective tissue and cushions the body from stress and strain. It is divided into two layers, the superficial area adjacent to the epidermis called the papillary region and a deep thicker area known as the reticular dermis [12].

- **Hypodermis:**

  The hypodermis is the layer of tissue immediately below the dermis of the skin. It is generally not considered to be part of the skin as such (McKinley et al, Marieb and Hoehn). It is linked to the dermis towards the depth. It is a richly vascularized loose connective tissue which contains more or less adipose tissue depending on the location. It serves as an interface between the dermis and the mobile structures located below it such as muscles and tendons. It also protects the body from physical shocks, temperature variations and serves as a fat reserve [12].

Here is figure 1.10 which shows the 4 layers.

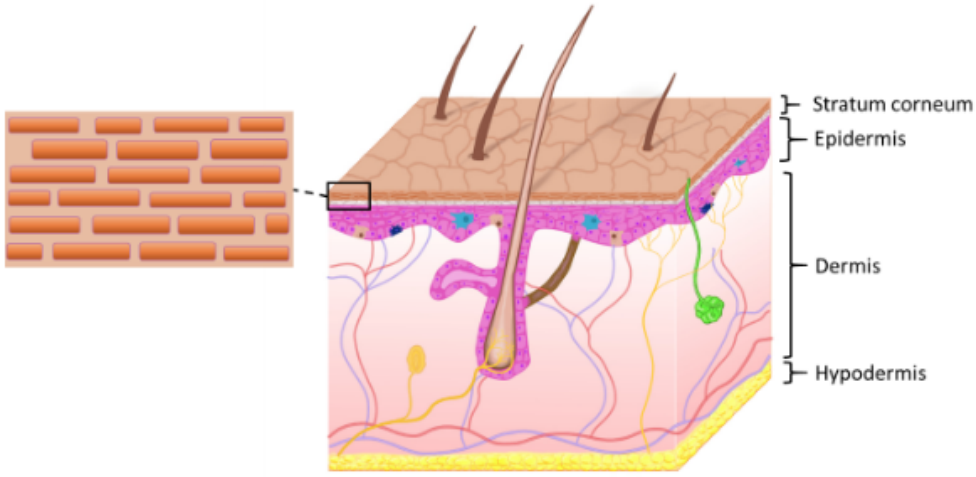

Figure 1.10: Skin layers.

1.7 Conclusion

In this chapter, we introduced the basic knowledge as a basis for understanding different image processing technologies. Some classic processing methods have been proposed in the literature. We have introduced some of the most common methods in image processing and analysis in our opinion, and these methods have been used throughout the work.

Image preprocessing makes it possible to consider subsequent processing to improve image quality. Two main methods can be considered to extract relevant regions from an image: We are looking for discontinuities in the scene, this is the contour method. And we are looking for areas with the same grayscale, this is the area method new research focuses on complex systems and their characteristics and the application of these methods in image processing.

In this chapter we have seen the characteristics of an image and the fundamental notions of medical images and we have defined the organs (the skin and the brain) that we will segment by applying deep learning, as we will see in the next chapters.

Chapter 2

Initiation to Deep learning

2.1 Introduction

Machine learning (ML) is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. but today saying IA or ML the first idea that comes to our mind is deep learning.

Nowadays, DL is the center of attention because its achievement is much more important than any other machine learning algorithm in such complex tasks as Image processing and object recognition in [13] which show us advances in the use of deep convolutional networks for object recognition, and the adoption of deep learning by the vision community by computer and also Speech recognition and signal processing in [14] which present the results obtained in phonetic classification for automatic speech recognition as the first industrial application of deep learning.

DL is one of the reasons behind the recent movements and advancements in AI, and this is the main reason why AI finally has the opportunity to become more realistic.

So what is DL?

According to the founders Yann LeCun, Yoshua Bengio Geoffrey Hinton in [15] Deep learning allows computational models made up of several processing layers to learn data representations with several levels of abstraction.

And we can see a representation of perceptron model in figure 2.1

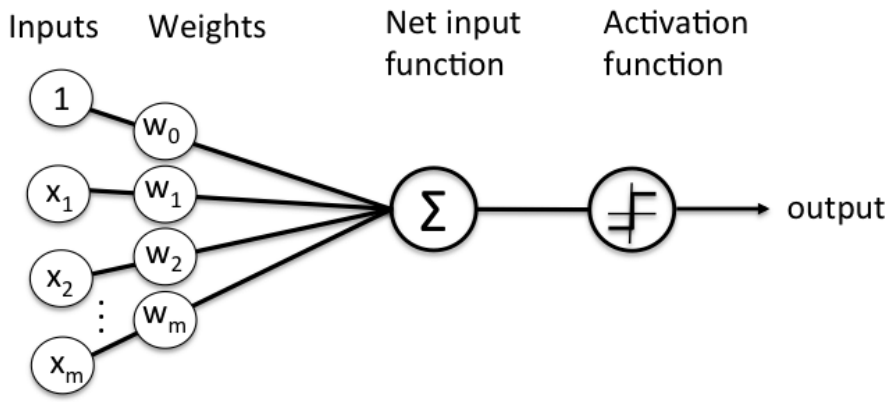

Figure 2.1: Perceptron

2.2 Definition of deep learning

Deep learning is a subfield of machine learning dealing with algorithms inspired by the structure and function of the human brain. In other words, it reflects how our brain works. Deep learning algorithms are similar to the way the nervous system structures itself where each neuron connects and transmits information [16].

Deep learning is a relatively new advance in neural network programming and represents a way to train deep neural networks essentially, any neural network with more than two layers is deep. The ability to create deep neural networks has been around since Pitts (1943) introduced the multilayer perceptron [17].

Figure 2.2 shows the structure of the multi-layer perceptron.

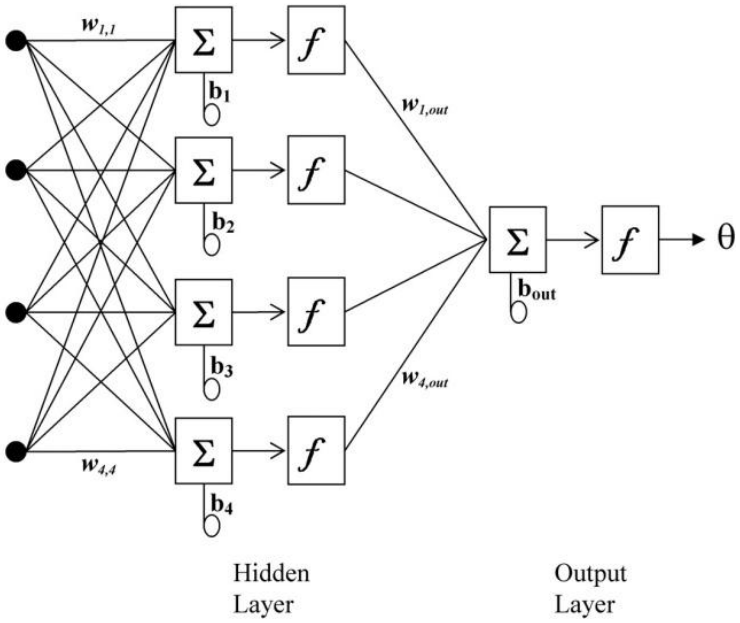

Figure 2.2: Multi-layer perceptron

2.3 History of Deep learning

The history of Deep Learning can be traced back to 1943, when Walter Pitts and Warren McCulloch created a computer model based on the neural networks of the human brain. They used a combination of algorithms and mathematics they called threshold logic to mimic the thought process. Since that time, Deep Learning has evolved steadily, with only two significant breaks in its development. Both were tied to the infamous Artificial Intelligence [18].

In 1960 Henry J. Kelley is given credit for developing the basics of a continuous Back Propagation Model. In 1962, a simpler version based only on the chain rule was developed by Stuart Dreyfus. While the concept of back propagation (the backward propagation of errors for purposes of training) did exist in the early 1960s, it was clumsy and inefficient, and would not become useful until 1985 [18].

The first convolutional neural networks were used by Kunihiko Fukushima. Fukushima designed NN with multiple pooling and convolutional layers and he developed in 1979 an artificial neural network, called Neocognitron, which used a hierarchical, multilayered design. This design allowed the computer the learn to recognize visual patterns. Additionally, Fukushimas design allowed important features to be adjusted manually by increasing the weight of certain connections [18].

1999 was a significant evolutionary step year for DL, when computers started becoming faster at processing data and GPU (graphics processing units) were developed. It gives faster processing. During this time, NN began to compete with support vector machines. While a NN could be slow compared to a support vector machine, NN offered better results using the same data and also have the advantage of continuing to improve as more training data is added [18].

In 2001, a research report by META Group (now called Gartner) described he challenges and opportunities of data growth as three-dimensional. In 2009, Fei-Fei Li, an AI professor at Stanford launched ImageNet, assembled a free database of more than 14 million labeled images. The Internet is, and was, full of unlabeled images. Labeled images were needed to train neural nets. Professor Li said, Our vision was that Big Data would change the way machine learning works. Data drives learning [18].

By 2011, the speed of GPUs had increased significantly, making it possible to train convolutional neural networks without the layer-by-layer pre-training. With the increased computing speed, it became obvious DL had significant advantages in terms of efficiency and speed. One example is AlexNet, a convolutional neural network whose architecture won several international competitions during 2011 and 2012. Rectified linear units were used to enhance the speed and dropout [18].

Currently, the processing of Big Data and the evolution of Artificial Intelligence are both dependent on Deep Learning. Deep Learning is still evolving and in need of creative ideas [18].

2.4 Deferent architecture of deep learning

There are mainly three types of deep learning architectures, each architecture has its own field of use and its characteristics.

2.4.1 DNN (Deep Neural Networks)

They are also well known by ANN (Artificial Neural Network). In many scientific disciplines, artificial neural networks appear to be a useful alternative to traditional statistical modeling techniques. Figure 2.3 shows the architecture diagram of DNN.

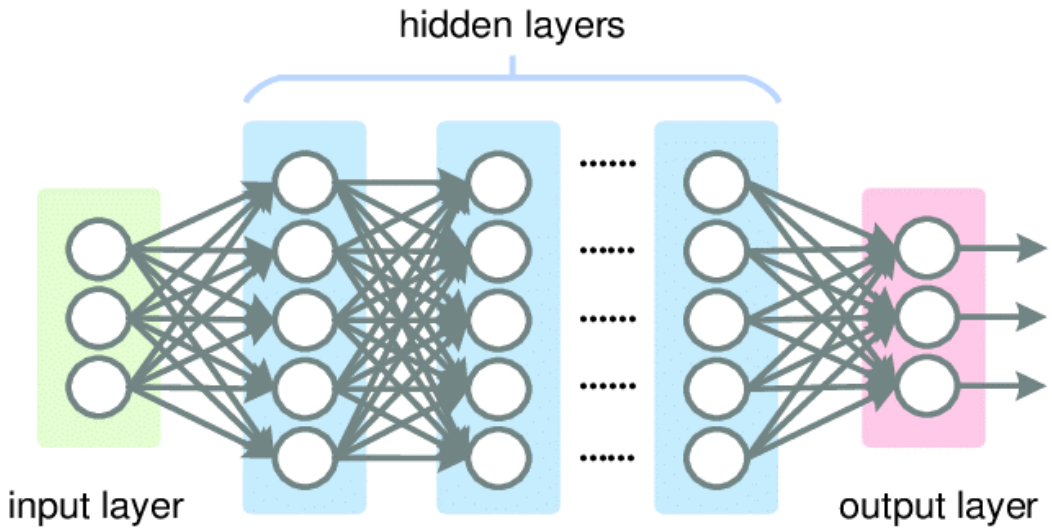

Figure 2.3: Architecture diagram DNN.

2.4.2 CNN (Convolutional Neural Networks)

The deep convolutional neural network (CNN) has always been at the heart of the rapid development of deep learning. Although CNN has been used to solve character recognition tasks since the 1990s (LeCun et al., 1997), their current widespread use is due to recent work. We can see the architecture diagram of CNN in Figure 2.4.

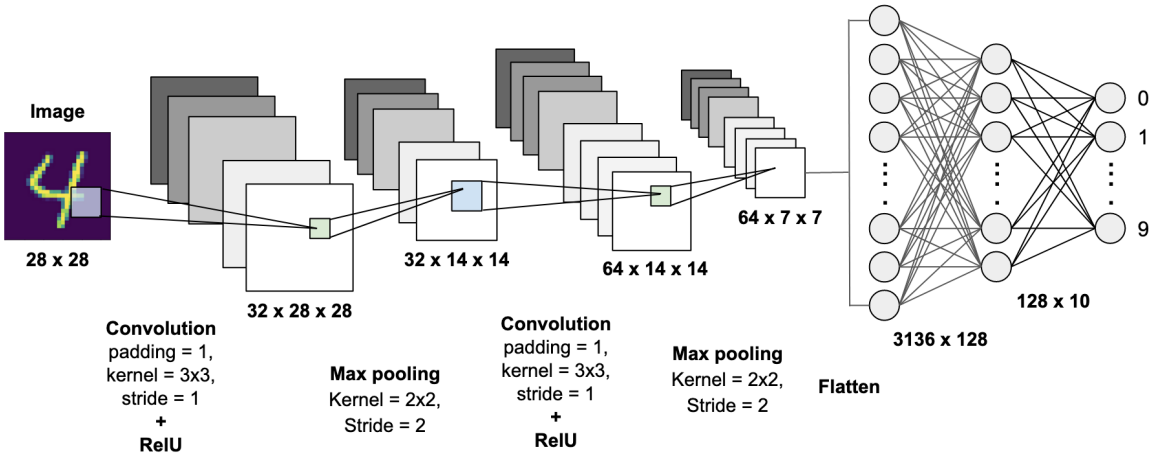

Figure 2.4: Architecture diagram CNN.

2.4.3 RNN (Recurrent neural networks)

The recurrent neural network is one of the main contents of deep learning, which enables the neural network to process sequences of data such as text, audio and video. They can be used to reduce sequences to advanced comprehension, annotate sequences and even generate new sequences from scratch.Figure 2.5 shows the architecture diagram of RNN.

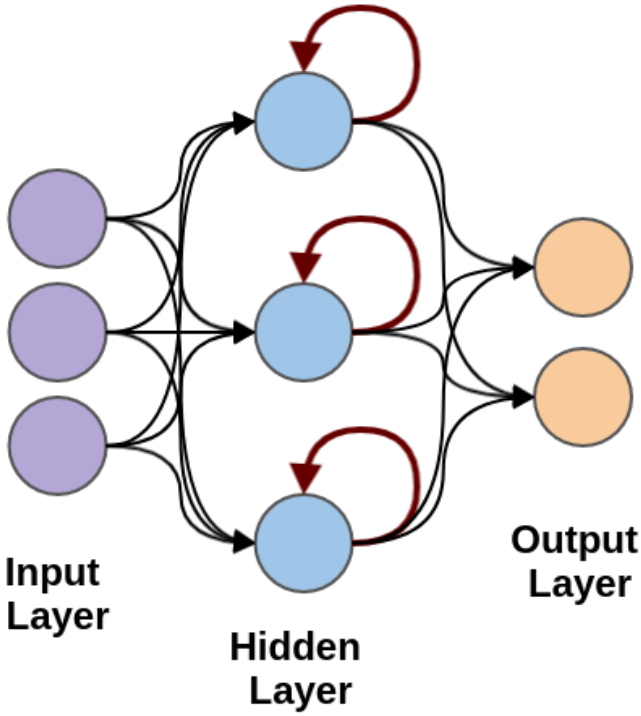

Figure 2.5: Architecture diagram RNN.

2.5 Deep learning use cases

Although deep learning is starting to be popular in several sectors, but in some of these sectors it is becoming essential.

We will define some of these essential sectors and we will end up with the area where our work has been devoted.

2.5.1 Financial Services

The use cases here powered by deep learning architectures specially CNNs include such as leveraging face detection and verification for Know Your Customer from documents applications relating to Regulatory Technology.

2.5.2 Marketing

Because of its lower regulatory barriers compared to other sectors and to its vast datasets available, marketing seems to us to be the first sector to be transformed by AI. The deep learning touch comes on hyper-personalization with customer segmentation and targeted marketing that is very relevant to the user as the future of marketing. NLP (Natural Language Processing) which is a very important branch of Machine Learning and therefore of artificial intelligence, help us harness the large amount of unstructured information present in social media for automated lead generation.

2.5.3 Security

This is a sector that will greatly benefit from the adoption of machine learning and deep learning techniques such as CNNs for face detection and recognition, behavioral recognition in videos and assist in detecting suspicious activity in cyber security.

2.5.4 Manufacturing and Industry

Automated agents (such as robots) that can automatically perform defect analysis and apply deep reinforcement learning will be able to achieve precision manufacturing at a higher speed and scale, and improve the efficiency of supply chain management.

2.5.5 Transportation

With the advent of driverless cars and drones, this is an industry facing fundamental changes. 5G will make the connected car an experience cabin for insiders to use augmented reality and virtual reality to power. In addition, areas such as "multi-agent deep reinforcement learning" will enable autonomous vehicles to communicate with each other and the surrounding environment,

thereby providing a strong foundation for communicating with well-defined standards. This will enable safe autonomous driving. In addition, autonomous drone technology will be able to provide logistical support for search and rescue teams in remote or difficult terrain areas, and assist farmers in automatic irrigation and harvesting.

As we said there is many other sector like Retail Sector and Insurance but now were going to end up with the one that interests as and which is none other than Health Care (medical) sector.

2.5.6 Health Care

With the passage of time, this field is likely to be the most transformed field of AI. Machine learning and deep learning will have a significant impact in fields such as medical imaging, mining electronic health data, and robotic surgery using 5G deep reinforcement learning. The application of machine learning technology on wearable devices allows us to process sensor data to achieve preventive health care. In addition, fields such as precision medicine also provide patients with improved results through technologies such as variational autoencoders.

2.6 Problems encountered in deep learning

There are two main problems encountered in deep learning and they are:

2.6.1 Overfitting

As the model learns the details and noise of the training data, overfitting occurs because it will negatively affect the performance of the model on the new data. This means that noise or random fluctuations in the training data will be absorbed by the model and learned as a concept. The problem is that these concepts are not applicable to new data and will have a negative impact on the generalizability of the model.

Nonparametric and nonlinear models are more likely to overfit, and they have more flexibility when learning the objective function. In this way, many nonparametric machine learning algorithms also include parameters or techniques to limit and constrain the amount of detail in training the models.

For example, a decision tree is a very flexible nonparametric machine learning algorithm, and it is easy to overfit the training data. This problem can be solved by pruning the tree after learning to remove some of the details it collects.

2.6.2 Underfitting

Underfitting refers to models that cannot be modeled on training data or that cannot be generalized to new data. An underfit machine learning model is not a suitable model because it performs poorly on training data, so it's obvious. The problem of underfitting is usually not addressed because it is easy to detect with good performance indicators. The solution is to keep trying other machine learning algorithms. However, this contrasts sharply with the problem of overfitting.

2.7 CNN's

Convolutional Neural Network (CNN) is a deep learning neural network designed to process structured data arrays such as images. Convolutional neural networks have been widely used in computer vision, and have become the latest technology in many vision applications (such as image classification), and have also achieved success in natural language processing for text classification.

Convolutional neural networks are very good at picking up patterns in the input image, such as lines, gradients, circles and even eyes and faces. It is this characteristic that makes convolutional neural networks so powerful for computer vision. Unlike earlier computer vision algorithms, convolutional neural networks can run directly on the original image without any preprocessing.

Convolutional neural networks are feed-forward neural networks, usually with up to 20 or 30 layers. The function of the convolutional neural network comes from a special layer, the convolutional layer.

2.7.1 The usefulness of CNNs

The use of CNNs for deep learning has grown for three important reasons:

- They eliminate the need for manual feature extraction, as they learn them directly;
- They produce excellent recognition results;
- They can be retrained to perform new reconnaissance tasks, allowing you to build on pre-existing networks.

CNN provides the best architecture for image recognition and pattern detection. Combined with advances in GPU and parallel computing, these are essential technologies that will drive new developments in autonomous driving and facial recognition.

For example, deep learning apps use CNN to review thousands of pathology reports to visually detect cancer cells. Convolutional neural networks also allow autonomous vehicles to detect objects and learn to distinguish between traffic signs and pedestrians.

2.7.2 How CNNs work

Convolutional neural networks can have dozens or even hundreds of layers, and each layer learns to recognize different characteristics of an image. Apply the filter to each image trained at a different resolution and use the output of each convolution image as the input for the next layer. The first filter can be very simple features like brightness or edges and then move on to more complex features to uniquely define the object.

CNNs perform the task of identifying and classifying images, text, sound and video.

A CNN always starts with an input layer then hidden layers on which it performs convolution operations followed by pooling and ends with a fully connected layer.

2.7.3 The convolution operation

In its most general form, convolution is an operation on two functions of a real-valued argument. The convolution product of two real or complex functions f and g is another function, which is generally denoted "f g" and which is defined by:

$$(f * g)(x) = \int_{-\infty}^{+\infty} f(x-t)g(t)dt = \int_{-\infty}^{+\infty} f(t)g(x-t)dt$$

There are several types of convolution but we are going to discuss only five that we have used in our architecture.

2.7.3.1 Simple Convolution

The convolution operation consists of applying a filter on the image matrix and it is up to us to choose the size of the filter as well as the stride and the padding if we need, after applying the filter the values pixels change so we will have feature maps and these feature maps represent the most relevant information of the image.

Figure 2.6 shows the convolution operation.

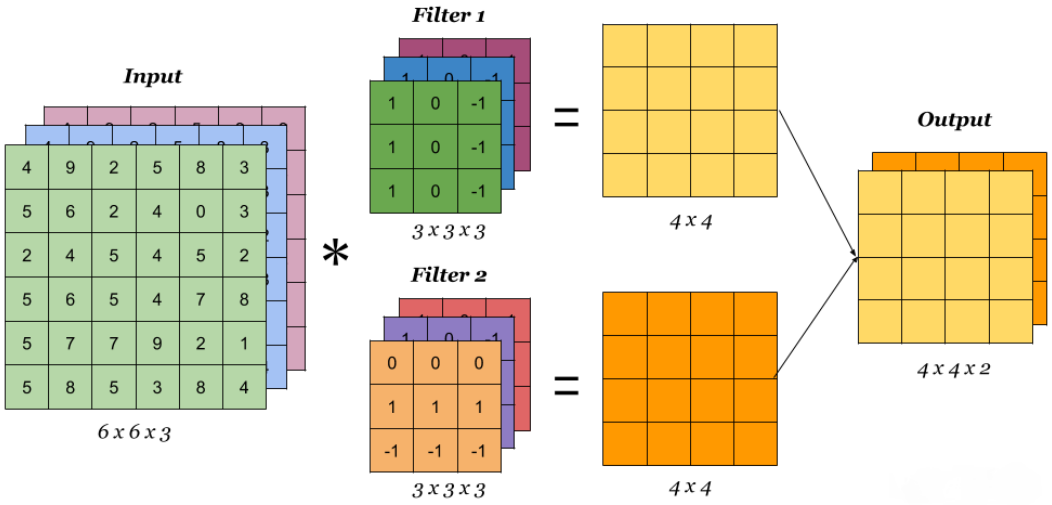

Figure 2.6: Convolution operation [20].

2.7.3.2 Transposed Convolution (Deconvolution)

Transposed convolution is also called deconvolution, which is inappropriate because deconvolution is all about removing the effects of convolution that we did not attempt to achieve. This is also called oversampling convolution, and it is intuitive for the task it is used to perform (i.e., oversampling of the input characteristic map).

Figure 2.7 shows the transposed convolution operation.

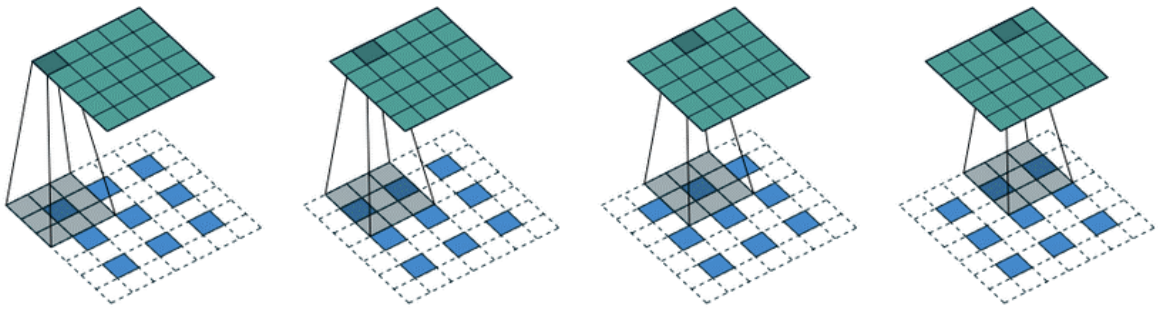

Figure 2.7: Transposed convolution [20].

2.7.3.3 Depthwise Convolutions

In depth-wise convolution, we use each filter channel only at one input channel. In the example, we have 3 channel filter and 3 channel image. What we do is break the filter and image into three different channels and then convolve the corresponding image with corresponding channel and then stack them back [18].

Figure 2.8 shows an example of a Depthwise Convolution:

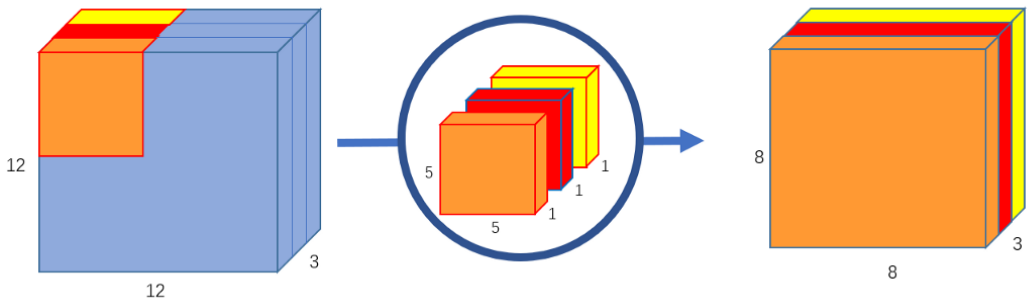

Figure 2.8: Depthwise Convolution [20].

2.7.3.4 Pointwise Convolution

Also known as point-to-point convolution is so named because it uses a 1x1x1 kernel or a kernel that iterates through each point. The kernel has the depth of the number of channels of the input image.

Figure 2.9 shows an example of a Pointwise Convolution.

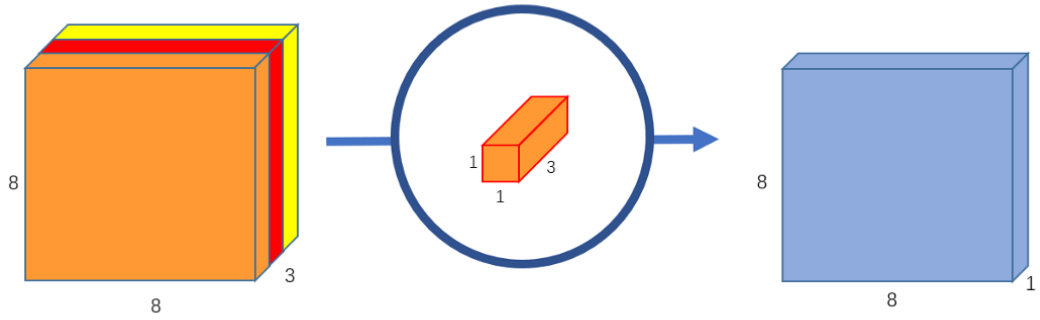

Figure 2.9: Pointwise Convolution [20].

2.7.3.5 Depthwise Separable Convolutions

The combination of Pointwise convolution and Depthwise convolution forms the Depthwise Separable Convolution.

The convolution separable in depth is so named because it involves not only the spatial size, but also the size of the depth - the number of channels -. The input image can have 3 channels: RGB. After several convolutions, the image can have several channels. You can think of each channel as a specific interpretation of that image. For example, the "red" channel explains the "red" of each pixel, the "blue" channel explains the "blue" of each pixel, and the "green" channel explains the "green" of each pixel. An image with 64 channels has 64 different interpretations of the image [20].

2.7.4 Activation functions in a CNN

The most common functions in deep learning are:

2.7.4.1 Sigmoid

The sigmoid activation function, also known as the logistic function, has traditionally been a very popular activation function for neural networks. The input to this function will be converted to a value between 0.0 and 1.0 [20].

Figure 2.10 is an illustration of sigmoid function.

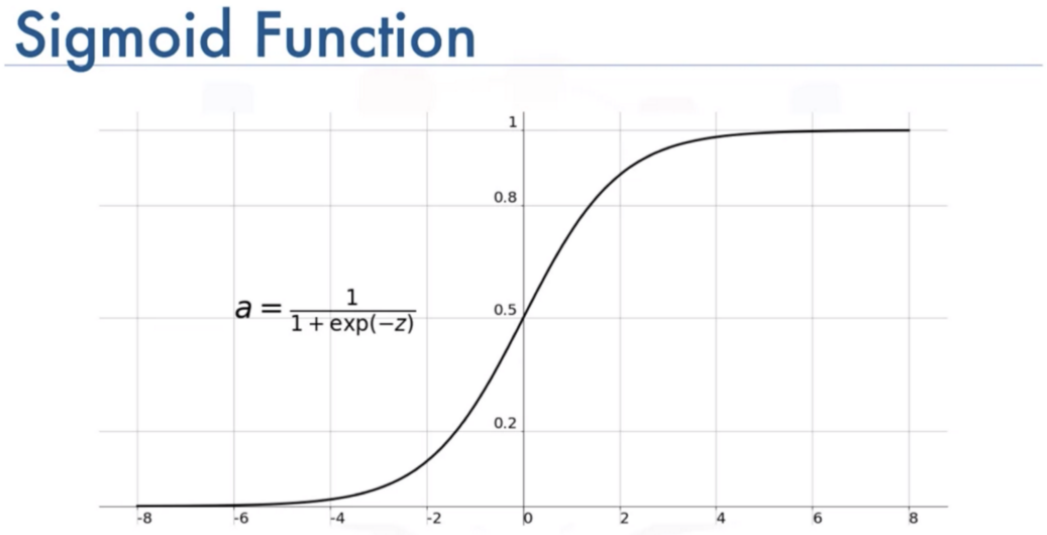

Figure 2.10: sigmoid function.

2.7.4.2 Hyperbolic tangent (tanh)

The hyperbolic tangent function, or tanh for short, is a similar shaped nonlinear activation function that outputs values between -1.0 and 1.0 [20].

Figure 2.11 is an illustration of hyperbolic tangent function

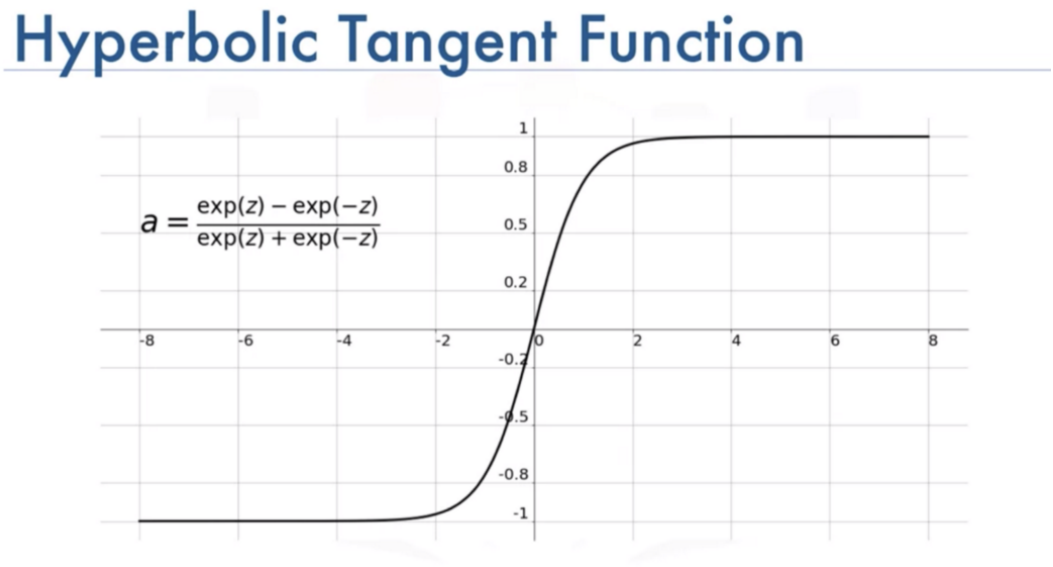

Figure 2.11: Hyperbolic Tangent Function.

2.7.4.3 The rectified linear unit (ReLU)

The rectified linear unit function or ReLU is the most widely used activation function in network design today. Besides being nonlinear, compared to other activation functions, the main advantage of using the ReLU function is that it does not activate all neurons at the same time. According to the diagram here, if the input is negative, it will be converted to 0 and the neuron will not be activated. This means that only a few neurons are activated at a time, which makes the network rare and very efficient. In addition, the ReLU function is one of the major advancements in the field of deep learning that overcomes the problem of gradual gradients.

We can see an illustration of ReLU function in figure 2.12.

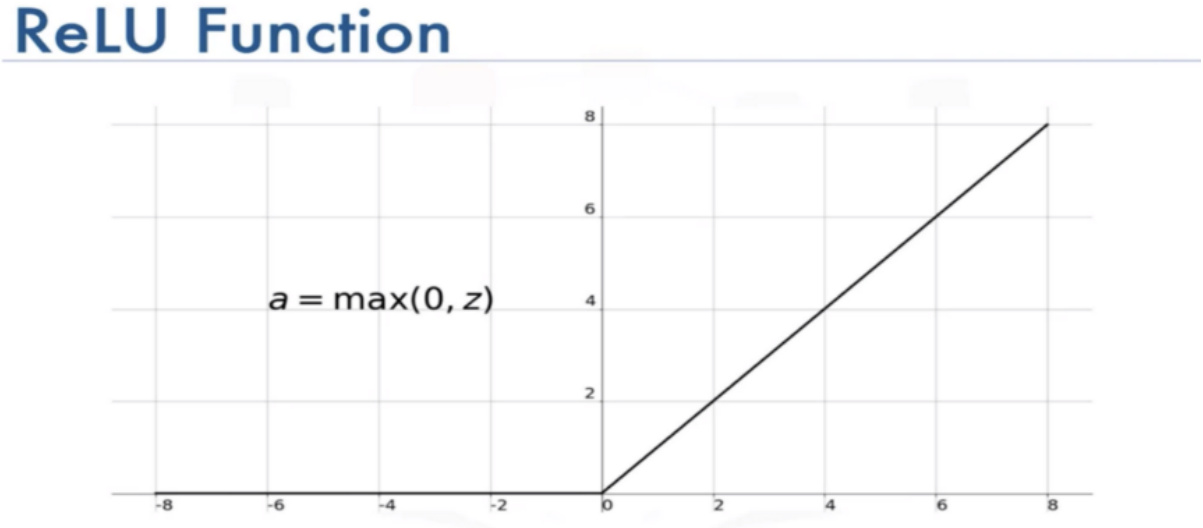

Figure 2.12: ReLU Function.

2.7.5 Transfer Learning

Transfer learning is one of the research areas of machine learning, which aims to transfer knowledge from one or more source tasks to one or more target tasks.

It can be seen as the system's ability to recognize and apply knowledge and skills acquired in previous tasks and apply to new tasks or areas of similarity. Figure 2.13 shows a diagram of the concept.

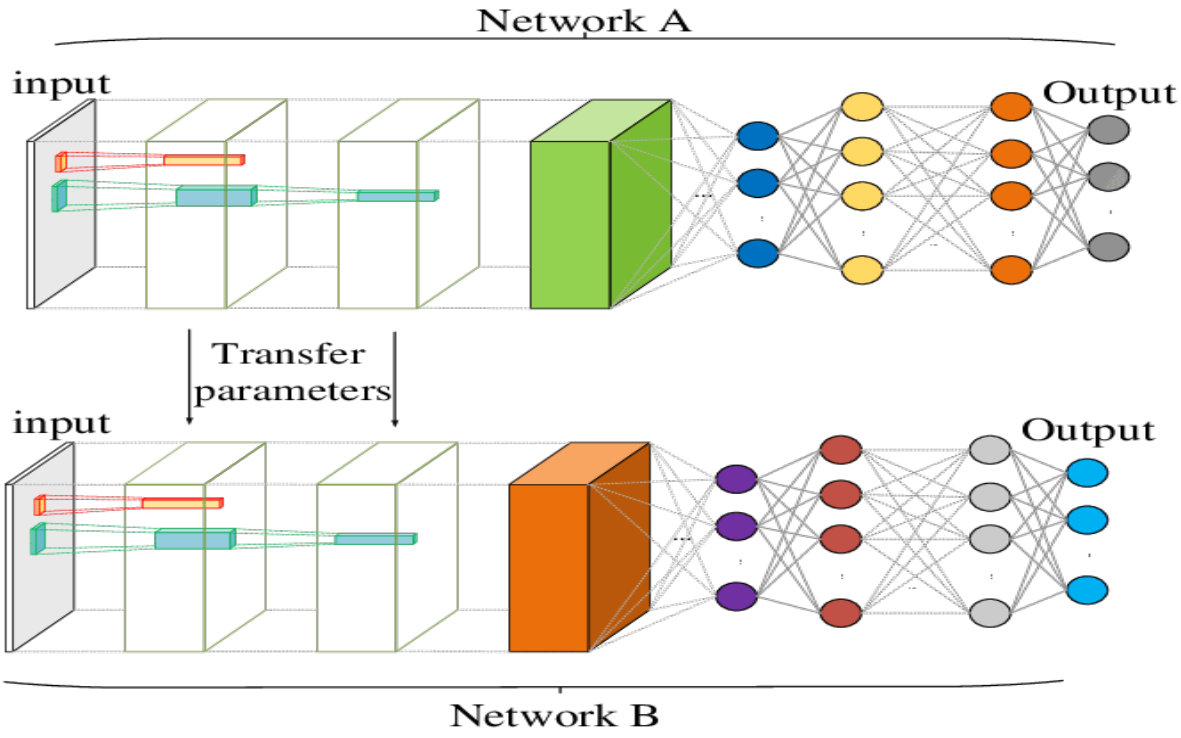

Figure 2.13: Transfer learning illustration.

2.8 Auto-encoders

2.8.1 Definition of Auto-Encoders

Autoencoding is an unsupervised learning technique used in neural networks. It learns efficient representation of data (encoding) by teaching the network to ignore signal "noise". The autoencoder network has three layers: an input layer, a hidden layer for encoding and an output decoding layer. Using backpropagation, unsupervised algorithms train continuously by setting the target output value to match the input. This forces the smallest hidden coding layer to use size reduction to remove noise and reconstruct the input.

2.8.2 Parts of the Auto-encoder

There are three main components in Autoencoder. They are Encoder, Decoder, and Code. The encoder and decoder are completely connected to form a feed forwarding mesh. The code act as a single layer that acts as per own dimension. To develop an Autoencoder, you have to set a hyperparameter that is you have to set the number of nodes in the core layer. In a more detailed manner, the output network of the decoder is a mirror image of the input encoder. The decoder produces the desired output only with the help of the code layer [21].

Figure 2.14 shows Parts of Auto-encoder.

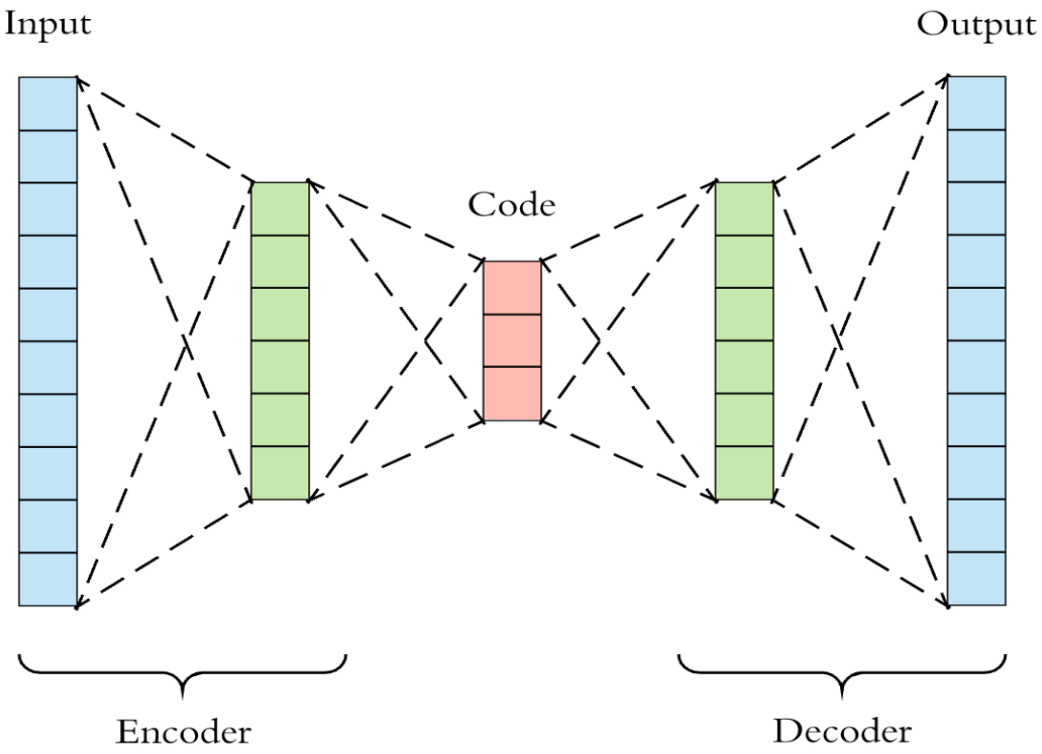

Figure 2.14: Parts of Auto-encoder [34].

2.9 Conclusion

although the field of deep learning remains vast and covers several sectors nevertheless we have defined in this chapter some fundamental notions as well as the advanced approaches of deep learning such as convolutional neural networks and autoencoders which will allow us to work in the sector that interests us which is none other than the medical domain for which the segmentation of images is our main concern.

Chapter 3

Medical image segmentation

3.1 Introduction

Transfer learning from 2D CNNs, trained on large-scale datasets (e.g., ImageNet), is a widely-used approach in 3D medical image analysis. To mimic the 3-channel image representation (i.e., RGB), prior studies follow either multi-planar or multi-slice representation of 3D images as 2D inputs. In these studies, pretrained 2D CNNs are usually fine-tuned on the target medical dataset.

Our job aim is to use a pre-formed network as an encoder and use a classic decoder to form a UNet and use it to segment 2D and 3D images.

3.2 MRI for Brain tumors

Magnetic Resonance Imaging (MRI) has been quickly imposed as being an essential tool of modern medical imaging and disease diagnosis. MRI is particularly useful for brain tumor diagnosis, patient follow-up, therapy evaluation and human brain mapping. The main advantage related to the use of MRI is its ability of acquiring non-invasive and non- irradiant medical images. It is also very sensitive to the contrast and provides an excellent spatial resolution which is entirely appropriate for the exploration of the brain tissues nature. In addition, the imaging derives easily 3D volumes according to brain tissue. In many practical cases, MRI is associated to conventional imaging of gliomas, unless indicated otherwise.

3.3 Unet architecture

UNET [26] was developed by Olaf Ronneberger et al. Used for segmentation of biomedical images. The architecture consists of two parts, the first part is the contraction part of the applet encoder, used to capture the context in the image. The encoder is just a traditional stack of maximum convolution and aggregation layers. The second part is the decoder part which used for symmetric expansion and used also for precise positioning using transposed convolution and these two parts are linked by skip connections which makes access from the encoder to the decoder possible.

The diagrams in Figure 3.1 show the Unet architecture.

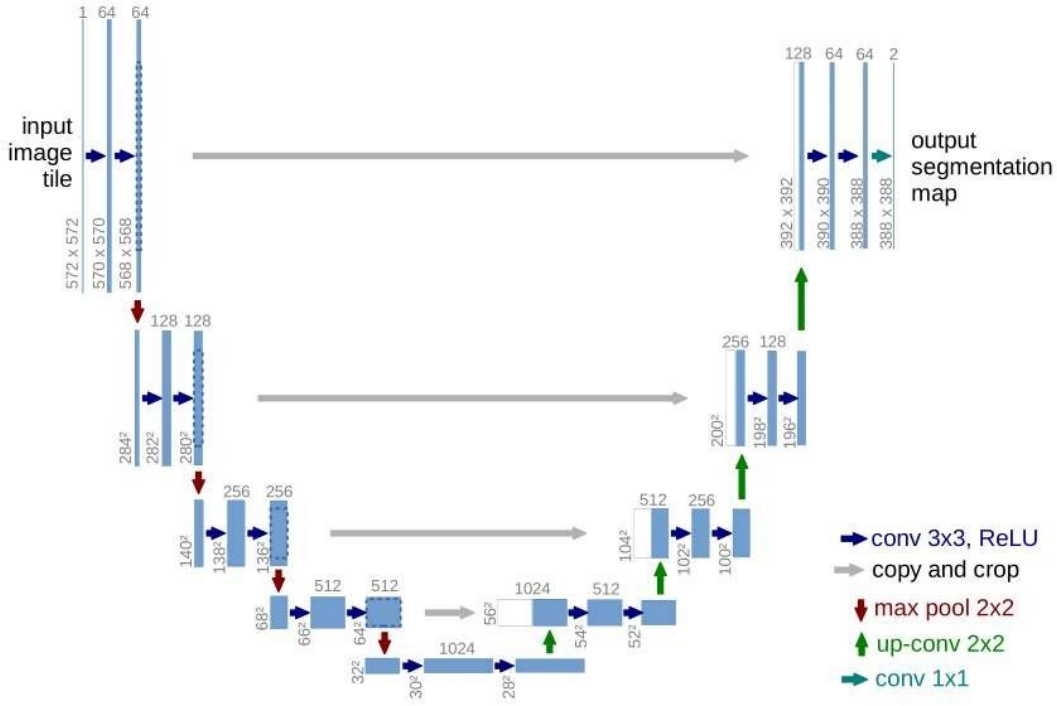

Figure 3.1: Unet architecture [26].

The network architecture consists of a contracting path (left side) and an expansive path (right side). The contracting path follows the typical architecture of a convolutional network. It consists of the repeated application of two 3x3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling. At each downsampling step we double the number of feature channels. Every step in the expansive path consists of an upsampling of the feature map followed by a 2x2 convolution (up-convolution) that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU. The cropping is necessary due to the loss of border pixels in every convolution. At the final layer a 1x1 convolution is used to map each 64 component feature vector to the desired number of classes. In total the network has 23 convolutional layers [26].

And although Unet is an old architecture (2015) but it remains difficult to beat in the field of segmentation of medical images and all this thanks to its skip connections.

3.4 EfficientNet architecture

The EfficientNet-B0 architecture is not developed by engineers, but by the neural network itself. They developed the model using research on a multi-objective neural architecture that optimizes precision and floating-point operations. Based on EfficientNet-B0, the author developed a complete

series of EfficientNets from B1 to B7, which achieved the highest accuracy on ImageNet dataset and were also very effective for competitors.

EfficientNet architecture is a very recent architecture (2019) it has proven its efficiency in terms of precision and speed, the secret of architecture comes from three main elements.

3.4.1 Depthwise Separable Convolution

By adopting the original two-step convolution depthwise and pointwise, the calculation cost can be greatly reduced, and the loss of precision can be reduced at the same time.

3.4.2 Inverse Res

Linear activation is used in the last layer of each block to avoid losing ReLU information. The main component of EfficientNet is MBConv, which is a reverse bottleneck conveyor belt, originally called MobileNetV2. By connecting a smaller number of channels (relative to the expansion layer) using a shortcut between the bottlenecks, combining it with a separable deep convolution, which reduces the amount of computation of nearly a thousand.

Figure 3.2 shows the architecture of EfficientNet.

Figure 3.2: EfficientNet architecture [22].

3.4.3 Linear bottleneck

Uses linear activation in the last layer in each block to prevent loss of information from ReLU.

3.5 Our architectures

In what follows we will present two kinds of architecture, they are very similar, but they are applied on different datasets.

3.5.1 2D architecture

According to [22], they designed a new core network using neural architecture research and extended it to obtain a series of models called EfficientNets that provide better accuracy and efficiency than other models. In particular, EfficientNet-B7 achieved a peak accuracy of 84.4% in the top 1/top 5 of ImageNet, while inference was 8.4 times smaller and 6.1 times faster. Better than the best ConvNet, where this EfficientNets can also transfer well with fewer parameters and reach the highest accuracy of an order of magnitude. This is why we decided to introduce the basic model B0 of EfficientNet into our work, which is certainly not as powerful as B7,but much faster and less in memory space.

In this first part we will deal with the problem of segmentation of skin images, i.e. segmenting each layer and extracting the edges using the clever algorithm and to segment we will use autoencoders.

3.5.1.1 Images preprocessing

First of all, we must prepare all images and adapt them to neural network and to do this we will perform several operations.

- **Create labels for images**:

  To create a label of an image we first of all have to study the region of interest so as not to make any errors, Then we used a software called Amira software to draw a label manually and save it in the desired file format Nifti in our case.

- **Data augmentation**:

  To remedy the problem of lack of data we have applied several image distortion techniques and these techniques are:

  - Zoom.
  - Flip (up-down, left-right).
  - Shift.
  - Rotation.

- **Data Normalization**:

  The problem with our data is that for each image we have the same min (0) but not the same max of voxels, to manage this we have extracted the max with respect to all the images and to normalize between 0 and 1 we have to divide all the matrices (images) on this max and we got the values to normalize between 0 and 1.

3.5.1.2 Method

Our method consists of using two different architectures. The first one is the classic Unet that remains very efficient in the field of medical image segmentation and the second one is an architecture that is known for its speed and power called EfficientNet.

EfficientNet architecture is dedicated to the classification of images and was formed on the ImageNet dataset and we took the opportunity to transfer learning to our images. We used EfficientNet as the encoder and we kept the Unet decoder and skip connections.As showed in Figure 3.3. After having programmed the callback, we started the training over 100 epochs.

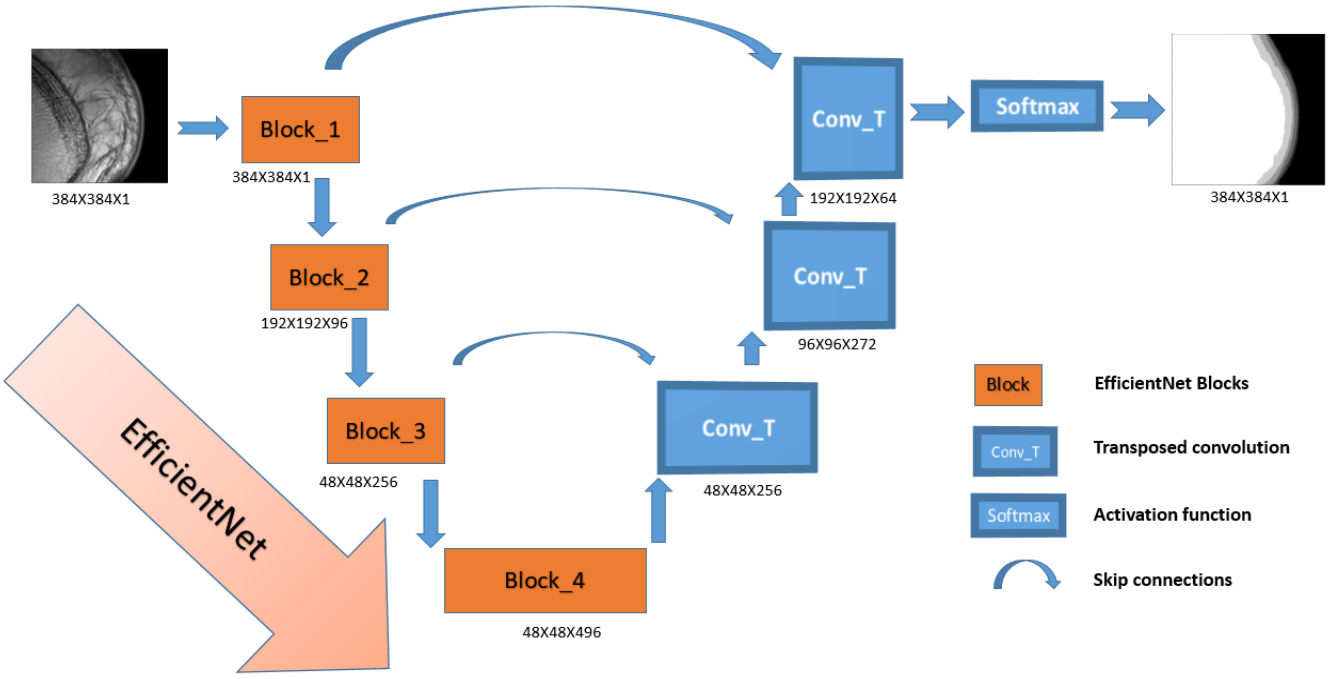

Figure 3.3: Our 2D architecture.

3.5.2 3D architecture

The proposed segmentation approach follows encoder-decoder Convolutional Neural Network architecture. It is built from an asymmetrically larger encoder to extract image features and a smaller decoder to reconstruct the segmentation mask. We embedded in the encoder part the recent efficient network called EfficientNet. Figure 3.4 shows Schematic illustration of the proposed network architecture. Input is a one channel 3D MRI crop. The inter- slice encoder as well as the decoder consist of a succession of blocks, each block is a Residual Network like with a GroupNorm normalization. The output of the decoder has three channels with the same spatial size as the input.

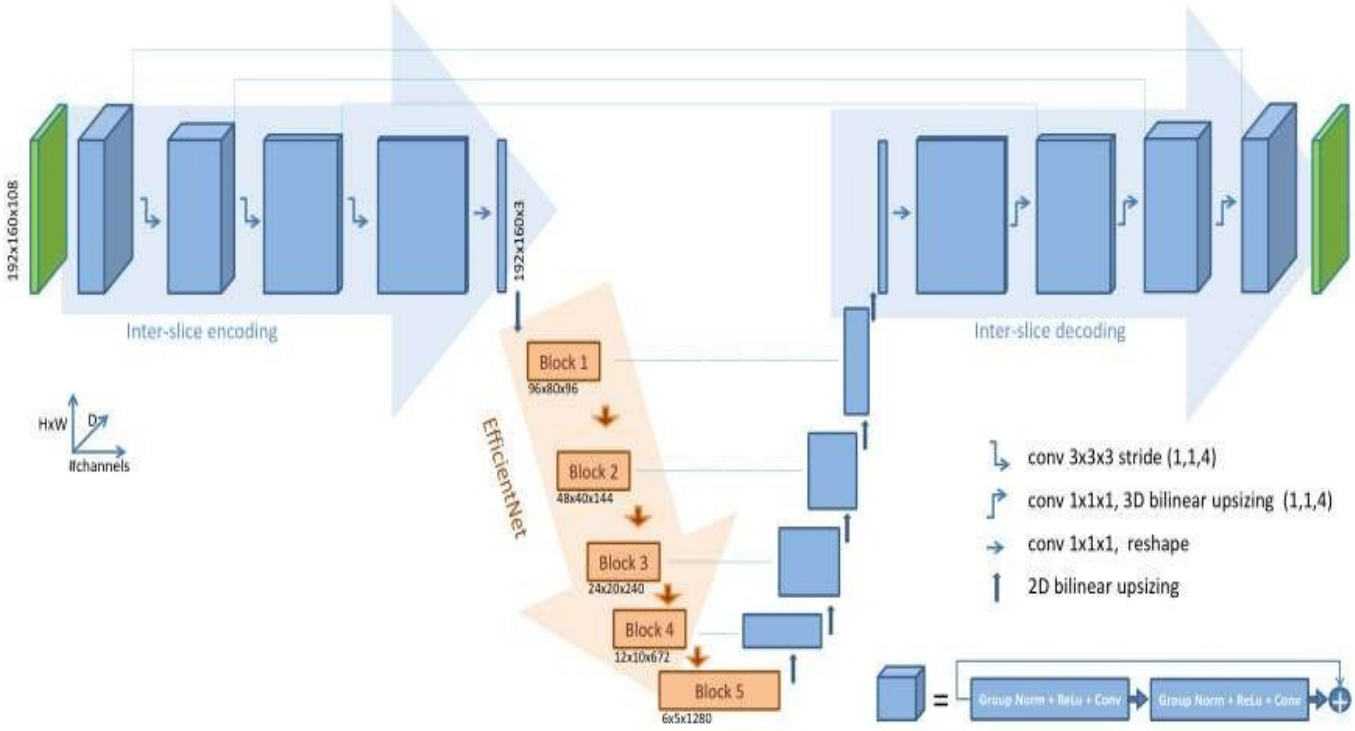

Figure 3.4: Our 3D architecture.

3.5.2.1 Data preprocessing and training

Because of the limitations in GPU memory and time-consuming computation, we were forced to take some precautions. We possessed each modality separately, resizing the dimensions of the images by reducing the background using the largest crop size of 192x160x108, and compromise the batch size to be 1. We did not use any additional training data and used only the provided training set.

3.5.2.2 Encoder part

The encoding process goes through two steps. First, encoding three-dimensional data to two-dimensional data, while keeping the height and width at their original size and compressing only the depth to 3 channels. Thus the data is ready to start the second encoding step, which is none other than the EfficientNet network without the fully connected layers.

As shown in Figure 3.4, the EfficientNet is represented as blocks as in its original version. However, only the blocks involved in skip connection layers are represented. Under each block, one can see outputs dimensions of features.

The inter-slices encoding part uses convolutional blocks which consists of two convolutions with normalization and ReLU, followed by identity skip connection. Following the works of [22]

we chose to use Group Normalization, which seems to be better than BatchNorm performance, especially when batch size is small.

3.5.2.3 Decoder part

Asymmetrically to the encoding part, the decoder is composed entirely of homogeneous blocks as shown in Figure 3.4. Obviously, the decoding part linked to the EfficientNet is a 2D decoder whereas the inter-slice decoding is a 3D decoder. Each decoder level begins by upsampling the spatial dimension, doubling the number of features by a factor of 2 followed by skip connections. The output of the decoder has three channels with the same spatial size as the input.

3.5.2.4 Loss

Many networks use cross-entropy loss functions for training, but based on Dice scores, the results may not be ideal. As an alternative, the soft dice loss function can be used to train the proposed network. Although there are several dice loss formulas in the literature, we prefer to use soft dice loss because according to the literature it is one of the best evaluation metrics. The soft dice loss function is differentiable and is given by:

$$L_{Dice} = \frac{2\sum P_{true} P_{pred}}{\sum P_{true}^2 + \sum P_{pred}^2 + \epsilon}$$

Where P_{true} and P_{pred} represent respectively the ground truth and the predicted labels. Brain MRI segmentation is a challenging task, partly because of the heavy class Imbalance and a large number of classes. Solve the problem of class imbalance the whole training only uses fixed loss function, cross entropy or dice Period is not the best strategy. Therefore, the linear combination of the two the loss function is generally considered to be the best practice and can bring more powerful functions and the best segmentation model. In fact, the final loss function is as follows:

$$L = L_{Cross} - L_{Dice}$$

3.5.2.5 Optimization

The proposed network architecture is trained with centered cropped patches of size 192x160x108 voxels, ensuring that all images content remains within the limits of the crop area. Constrained by the performance of the material, we set the batch size to 1. Training has been done using the Adam optimizer with an initial learning rate of $lr = 1e^{-4}$, and reduced by a factor of 10 whenever the loss has not improved for 50 epochs.

3.6 Conclusion

In this part of the work, we saw how we raised a problem of small dataset, also we put forward our different architecture and our methods including the transfer of skills from a powerful and pre-trained network that is not other than efficientnet on 2D images to an auto-encoder processing 3D images, and for that we have invested all our time to carry out this work and respond to the BraTS challenge 2020.

Chapter 4

Experimental Results

4.1 Introduction

In this chapter, we will introduce the data set and show the results of segmentation. the process of our work is to deal with two different data sets, Therefore two different problems, our first challenge is to segment different layers of skin (stratum corneum, epidermis, dermis, subcutaneous tissue). The second challenge is to segment brain tumors (whole tumor, tumor core, enhanced tumor) by participating in the Brats 2020 challenge, we must say though we lack material resources, but the effect is still good.

4.2 Skin dataset

The biggest problem with medical imaging in general is that there is not a lot of data available because the data is confidential, specially in our case the problem is that we have very little image and the other problem is that they are not labeled so segmenting very few unlabeled images using a powerful deep learning architecture is a difficult problem and avoiding overfitting is a miracle.

Figure 4.1 shows what the images look like.

The data consisted of 20 healthy subjects who underwent a 3T MRI (T2-weighted calculation with polycyclic echo) and a microscopic coil on the left heel. The outgoing images are of dimensions 384x384x1 (top, width, channel), these images are 2D MRI images encoded on 32 bits in Nifti format (.nii) and Figure 4.2 show the information of a sample.

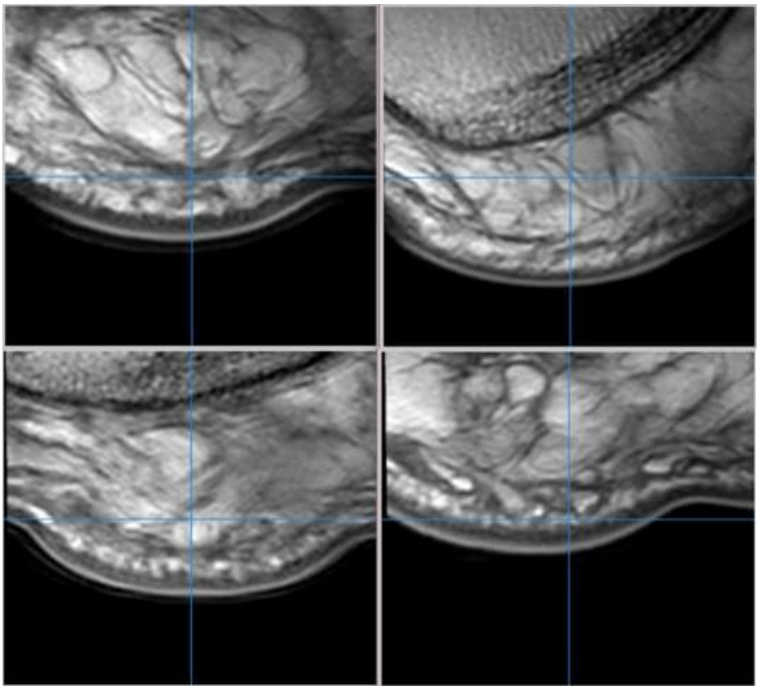

Figure 4.1: Some images from skin dataset.

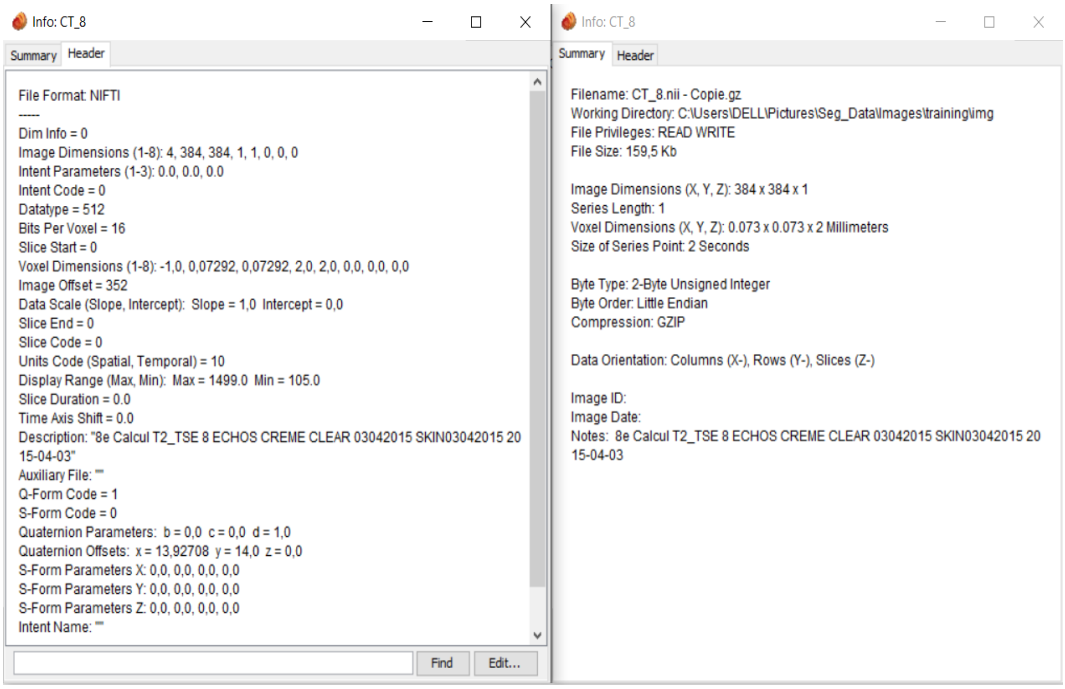

Figure 4.2: Skin image information.

4.3 BraTS challenge 2020

Multimodal Brain Tumor Segmentation Challenge (BraTS) aims at encouraging the development of state of the art methods for the segmentation of brain tumors by providing a large 3D MRI dataset of annotated Low Grade Glioma (LGG) and High Grade Glioma (HGG). BraTS 2020 training dataset included 369 cases, each with 4 modalities describing a) native (T1) and b) post-contrast T1- weighted (T1Gd), c) T2-weighted (T2), and d) T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) volumes, the data were acquired with different clinical protocols and various scanners from multiple institutions, using various MRI scanners. Each tumor was segmented into edema, necrosis and non-enhancing tumor and active/enhancing tumor. The annotations were combined into 3 nested sub-regions: whole tumor (WT), tumor core (TC) and enhancing tumor (ET).

4.4 BraTS dataset

Ample multi-institutional routine clinically-acquired pre-operative multimodal MRI scans of glioblastoma (GBM/HGG) and lower grade glioma (LGG), with pathologically confirmed diagnosis, this data are provided as the training, validation and testing data for this years BraTS challenge. Specifically, the datasets used in this year's challenge have been updated, since BraTS'19, with more routine clinically-acquired 3T multimodal MRI scans, with accompanying ground truth labels by expert board-certified neuroradiologists.

All BraTS multimodal scans are available as NIfTI files (.nii.gz) and describe a) native (T1) and b) post-contrast T1-weighted (T1Gd), c) T2-weighted (T2), and d) T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) volumes, and were acquired with different clinical protocols and various scanners from multiple institutions.

All the imaging datasets have been segmented manually, by one to four raters, following the same annotation protocol, and their annotations were approved by experienced neuro-radiologists. Annotations comprise the GD-enhancing tumor (ET label 4), the peritumoral edema (ED label 2), and the necrotic and non-enhancing tumor core (NCR/NET label 1). BraTS challenge provides 369 images of different patients for training with dimensions 240x240x155 (height, width, channel) these are very large images and this size requires powerful equipment (RAM memory and GPU memory).Figure 4.3 shows the 4 different modalities of one same brain and Figure 4.4 shows the label of the image in figure 4.3.

These images are 3D MRI images encoded on 32 bits in Nifti (.nii) format, figure 4.5 shows the information of an image from our dataset

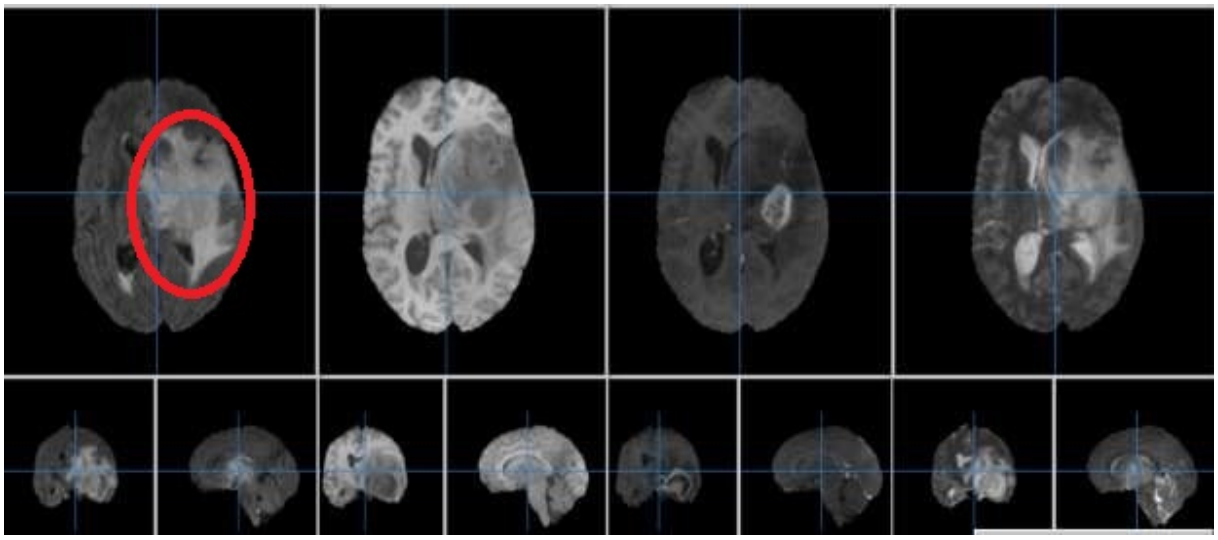

Figure 4.3: First image in dataset with its four modalities (flair, t1, t1ce, t2 from left to right).

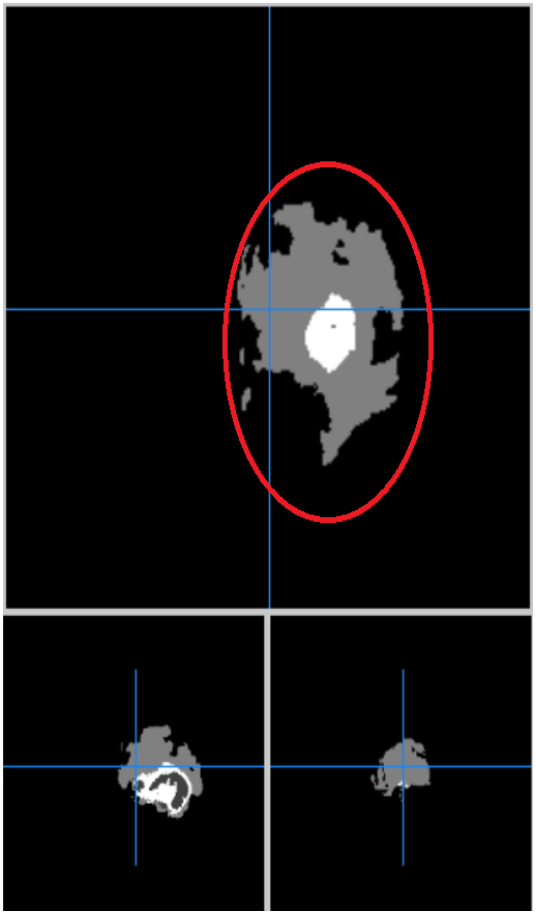

Figure 4.4: the ground truth of the first image in dataset.

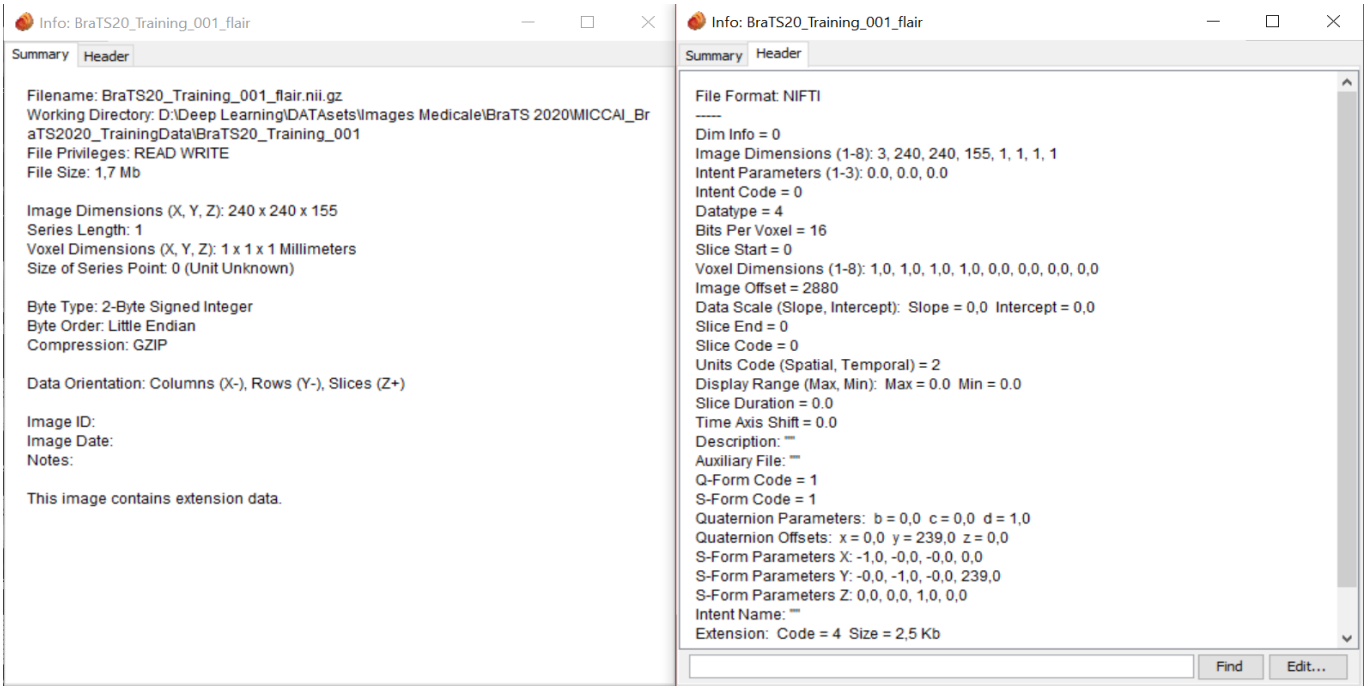

Figure 4.5: Brain image information.

4.5 Result of the 2D approach

4.5.1 Result of Skin data segmentation with Unet

For the first neural network (Unet) the accuracy reached is **81%**. Figure 4.6 shows the predicted mask from a test image (new image different from the training images).

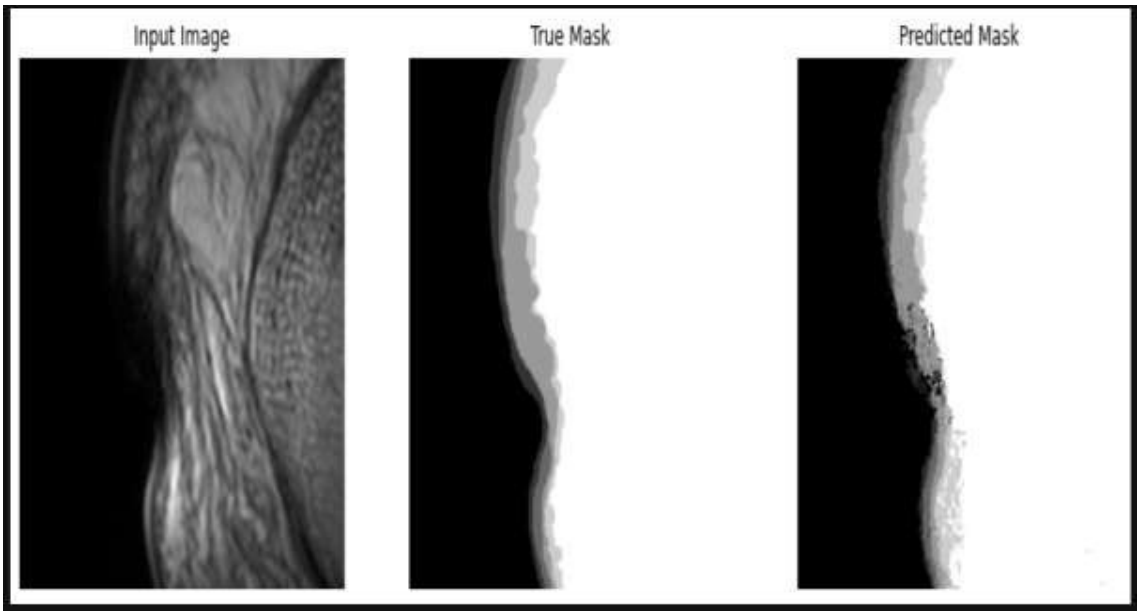

Figure 4.6: Result of Unet segmentation after 100 epochs.

After applying the canny filter on the predicted mask and extracting the borders, we superimposed it on the original image to better see the result (see figure 4.7).

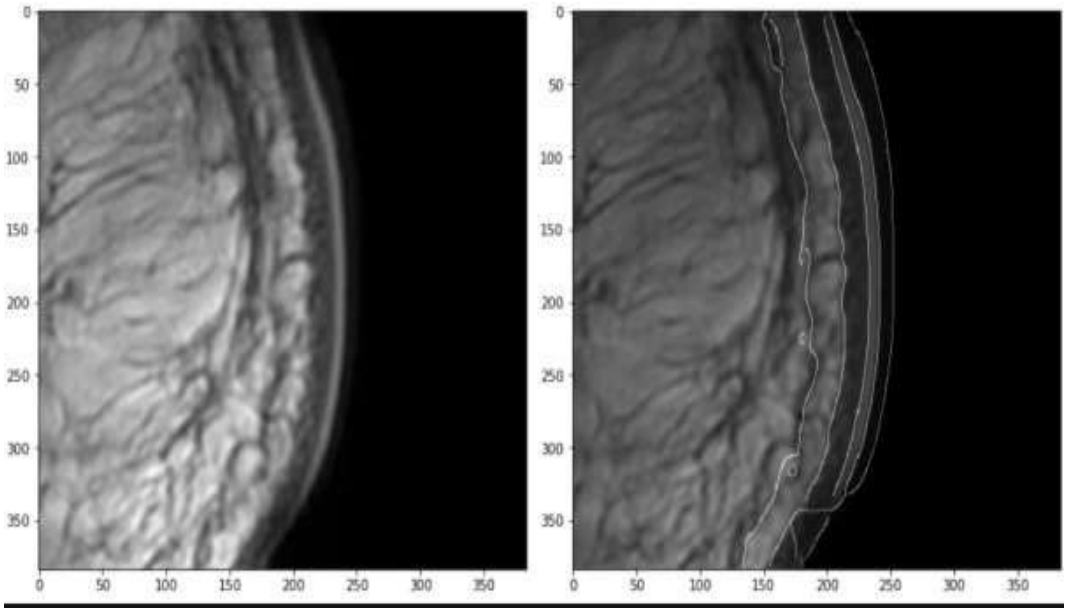

Figure 4.7: Canny filter result of Unet superimposed on the original image.

4.5.2 Result of Skin data segmentation with EfficientNet

For our second neural network (EfficientNet) the accuracy reached is **90%**. Figure 4.8 shows the predicted mask from a test image (new image different from the training images).

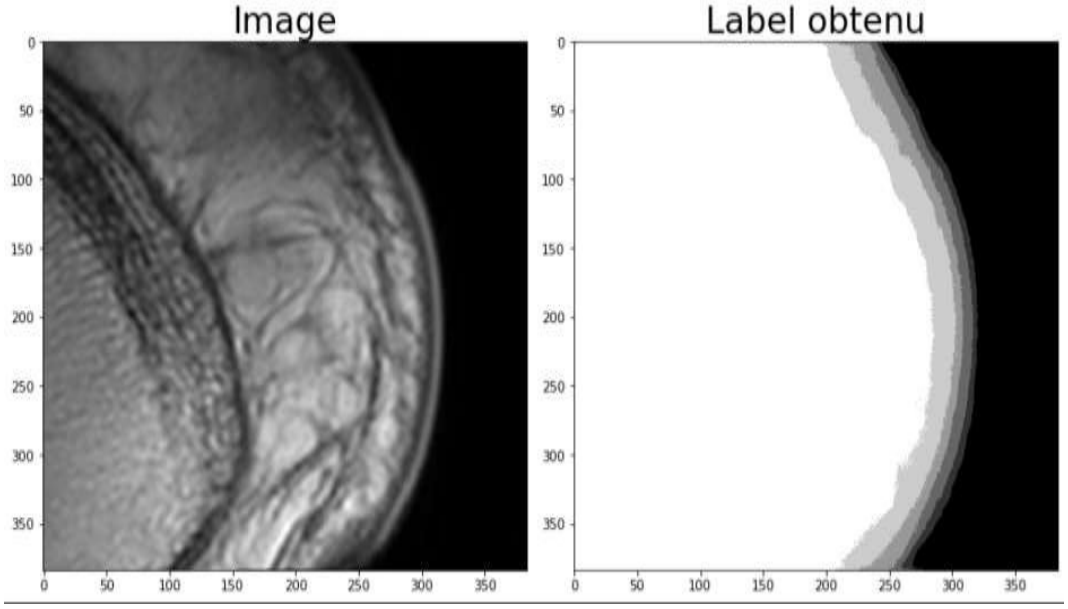

Figure 4.8: Result of EfficientNet segmentation after 100 epochs.

After applying the canny filter on the predicted mask and extracting the borders, we superimposed it on the original image to better see the result, figure 4.9 shows that.

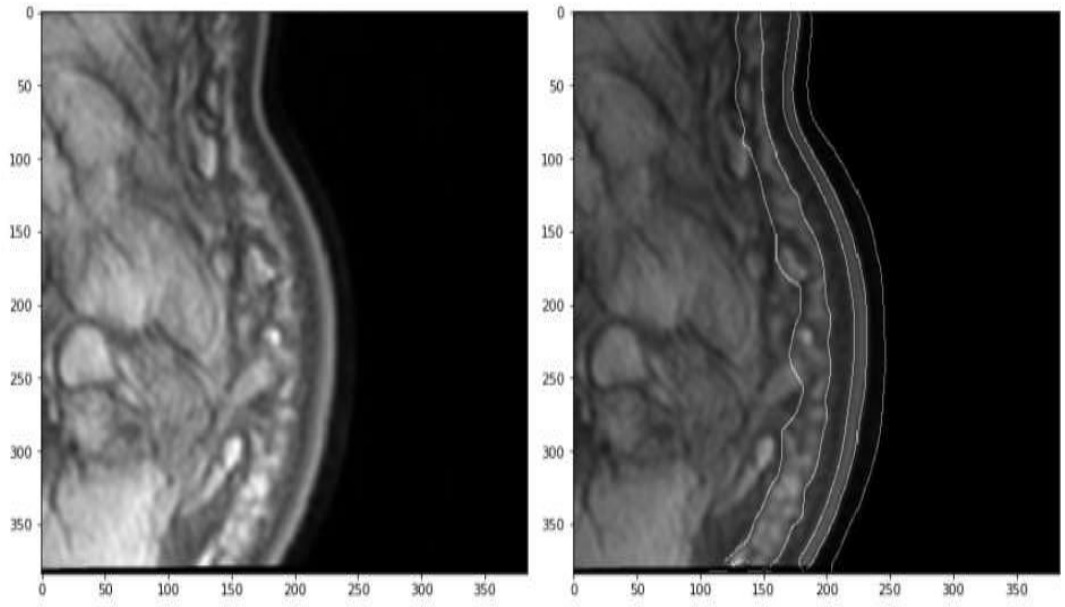

Figure 4.9: Canny filter result of EfficientNet superimposed on the original image.

The table 4.1 shows the 2 different results obtained from the 2 models applied to the skin data set.

Architectures	**Accuracy (after 100 epochs)**
Unet	81%
EfficientNet	90%

Table 4.1: Skin data results.

4.6 Result of 3D approach

Before seeing the results, we will first see the different sub-regions of the tumor through the figure 4.10 where each tumor was segmented into edema, necrosis and non-enhancing tumor, and active/enhancing tumor. Annotations were combined into 3 nested sub-regions: Whole Tumor (WT), Tumor Core (TC) and Enhancing Tumor (ET).

After the long training of our network on the 368 images, we downloaded the validation dataset that contains 125 new patients and we tested our network on these 125 images. Figure 4.11 shows the predicted mask of the validation image which gave the best result in achieving 89% Dice for enhancing tumor (ET) and 90% for whole tumor (WT) and 95% for tumor core (TC).

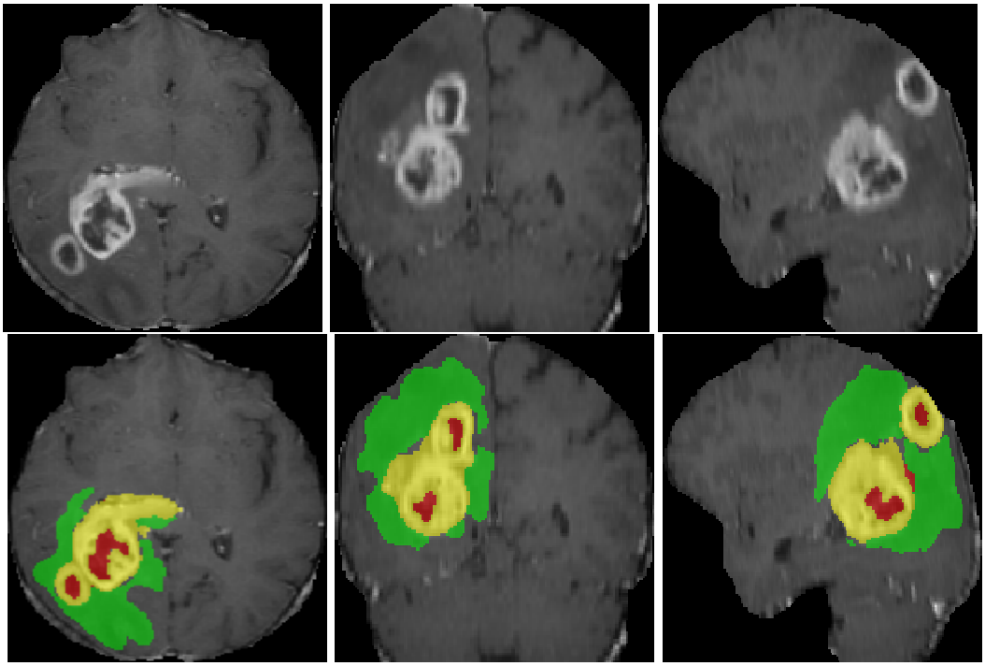

Figure 4.10: Typical result on the BraTS validation dataset. From left to right: axial, coronal and sagittal views in T1ce modality. Enhancing tumor is shown in yellow, necrosis in red and edema in green.

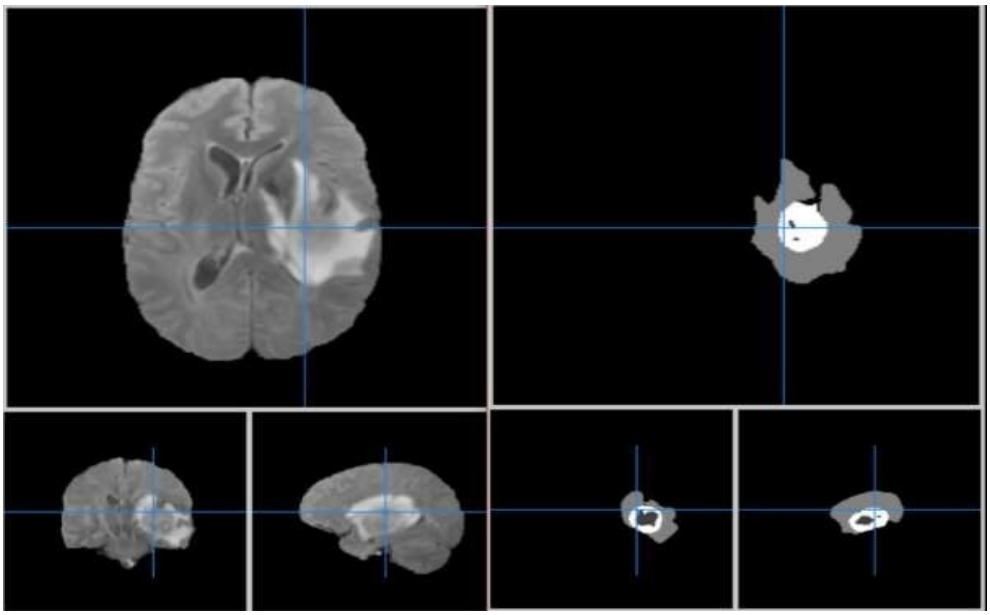

Figure 4.11: predicted mask from a validation image.

These results above show the precision of one image among 125 images, but the average of all results of the 125 images is as follows 65% Dice for enhancing tumor (ET) and 84% for whole tumor (WT) and 68% for tumor core (TC), and this after several modifications and the addition of a parameter alpha = [0.05 ... 0.4] which represents a slight noise added to the background class, we noticed that the network formed on the T1ce modality well segments the two tumors

(ET) and (TC) unlike the other networks formed on the other modalities, after that we consulted the opinion of a specialist (doctor) who informed us that the ET and TC tumors did not are visible only through the T1ce modality, and this is why we make sure that the network formed on T1ce focuses on the predictions of two tumors (ET) and (TC) and for the tumor prediction (WT) will be carried out by the network formed by the three combined modalities, so the modifications improved our results and we reached a better score and this despite of 20 mask predicted at 0% Dice for both (ET) and (TC) tumors, the following table 4.2 shows the improvements in our results after several tests.

Dice Coefficient	Mean			stdDev			Median		
Sub-regions	ET	WT	TC	ET	WT	TC	ET	WT	TC
Flair	0.30	0.84	0.53	0.23	0.13	0.25	0.37	0.89	0.61
T1ce	0.65	0.74	0.68	0.32	0.17	0.31	0.80	0.78	0.79
T2	0.35	0.81	0.57	0.25	0.13	0.25	0.42	0.85	0.65
First combination	0.65	0.84	0.68	0.31	0.10	0.31	0.81	0.87	0.78
Second combination	0.46	0.78	0.62	0.29	0.12	0.26	0.58	0.81	0.73
Best one (first combination)	0.65	0.84	0.68	0.31	0.10	0.31	0.81	0.87	0.78

Table 4.2: Results where first combination we used the weight of T1ce modality to predict ET and TC sub-regions and to predict WT we used the mean of the prediction get from 3 modalities which are (Flair, T1ce, T2). For the second combination, we also used the weight of T1ce modality to predict ET and TC sub-regions but to predict WT we used the mean of the prediction get from all modalities.

4.7 Conclusion

In this chapter, we have presented the results of different architectures applied to different datasets (Skin 2D dataset and Brain 3D dataset), and learned how the EfficientNet network has proven its effectiveness as an encoder by providing good results. (90% accuracy on the skin dataset and 65%, 84% and 68% of dice respectively for ET, WT and TC on the brain dataset). For the BraTS challenge, our goal is not to get the best score, but to adapt a network designed to classify 2D RGB images to another more complicated task which is the segmentation of 3D MRI images. It can be seen that the learning transfer of trained weights on 2D natural images can be used for processing 3D medical images.

Conclusion

Part of the work carried out within of this thesis involves the automatic segmentation of skin layers, and more importantly, in another part, we involve the automatic segmentation of brain tumors on volumetric MRI images.

This research represents an important issue in the field of medical diagnostic monitoring and assistance. The difficulties encountered by physicians in the process of quantification and 3D modeling indicate that manual segmentation may have direct neurological consequences for patient survival, and therefore, the main interest of the work carried out.

The first work concerns the segmentation of the cutaneous layers we have used two different architectures. The first one is Unet that is renowned for its power in the medical field, which reached 81% of precision. But despite its power, it remains weak compared to EfficientNet. the latter was introduced as being the encoder part of the network and keeping Unet as the decoder part, this network has exceeded the Unet in terms of precision and speed by reaching 90% of precision.

The second work is a participation in one of the most famous challenges in the field of medical image segmentation called Brain Tumor Segmentation "BraTS" 2020. This challenge proposes to segment 3D images of brain tumors. To do so, we introduced a generic 3D U-Net architecture that allows a performance transfer, by embedding 2D classifier network (efficientnet). The encoder as well as the decoder are composed of two stages. The 3D input data goes through a process of depth shrinking in order to transform the 3D data into 2D data. Moreover, decoding also goes through a 2D decoding phase followed by a 3D decoding procedure, we managed to segment the three existing sub-regions of a brain tumor ET, WT and TC for each image after several tests and our network reached 65%, 84% and 68% of dice coefficient respectively for ET, WT and TC.

As perspectives we want to integrate the concept of Variational Auto-Encoder (VAE) and the concept of regularization after each convolution of our network and will see if the results will improve.

Appendix A

Tools and software

A.1 Tensorfow

Created by the Google Brain team in 2011, as a proprietary system dedicated to deep learning neural networks, TensorFlow was originally called DistBelief. Subsequently, the source code of DistBelief was modified and this tool became an application-based library. In 2015 it was renamed TensorFlow and Google made it open source. Since then, it has undergone over 21,000 communication changes and moved to version 1.0 in February 2017. To put it simply, TensorFlow is a machine learning library, it is a toolkit for solving problems. extremely complex math with ease. It allows researchers to develop experimental learning architectures and turn them into software [23].

A.2 Mango Software

Mango software is a tool for viewing medical images, whether 2D or 3D or other, mango provides image viewing and provides all the information related to this image.

A.3 Amira Software

Amira Software is a powerful, universal 2D-5D solution for visualizing, analyzing and understanding life science and biomedical research data from many image modalities, including Optical and Electron Microscopy, CT, MRI and other techniques. With incredible speed and flexibility, Amira Software supports advanced 2D-5D bioimaging workflows in research areas ranging from structural and cellular biology to tissue imaging, neuroscience, preclinical imaging and bioengineering [24].

We use this software to create labels for our skin images.

A.4 Google Colab

Google Colab or Colaboratory is a cloud service provided by Google (free), based on Jupyter Notebook, designed for training and research in machine learning. This platform allows you to train machine learning models directly in the cloud. Therefore, there is no need to install any software on our computer except the browser [25].

Appendix B

Acceptance letter

De : Ahror Belaid
Envoyé le :mardi 20 octobre 2020 09:53
À : mohamed lamine allaoui; ahcene zetout
Objet : Fwd: BraTS 2020 results and manuscript publication

---------- Message transféré ----------
De : **Bakas, Spyridon** <Spyridon.Bakas@pennmedicine.upenn.edu>
Date : mardi 20 octobre 2020
Objet : BraTS 2020 results and manuscript publication
À : "belaid.ahror@gmail.com" <belaid.ahror@gmail.com>, "ahror.belaid@univ-bejaia.dz" <ahror.belaid@univ-bejaia.dz>
Cc : "brats2020@cbica.upenn.edu" <brats2020@cbica.upenn.edu>

Dear BraTS 2020 participant,

I am writing to thank you for your participation in this year's challenge and also inform you of a few points:

1. **BraTS proceedings (LNCS) paper**
    - PAPER UPDATE
        - Attached you can find the performance evaluation of your method during the testing phase. Feel free to include these numbers in your updated paper.
        - Extend your paper submission to be between 10 and 14 pages. Extending by using more and/or larger figures and tables is allowed. Do not worry if references extend your paper beyond 14 pages.
        - For this update you are asked to edit your current submission in CMT.
        - This **submission deadline is October 30**.
    - UPDATED PAPER REVIEWS
        - Once the submissions are complete you will be invited to lightly review 2-3 papers for any potential errors or inappropriate language.
        - You need to review these papers while having in mind the following 2 points:
            - Is there anything missing in the methodological description?
            - Is there anything that is currently missing and the authors can do in a week to improve their paper?
        - Your paper will also be reviewed by another BraTS participant.
        - Make sure that the BraTS references are cited in all papers. (Note that you will be listed as a co-author in the arxiv paper, which I am currently updating)
        - You will have to **complete your reviews by November 15**.
    - FINAL (CAMERA-READY) PAPER
        - At this point you will have a week to finalize your paper by considering the reviews you received and upload your final camera-ready submission. This submission will be the paper that will finally be printed in the LNCS

proceedings. To complete the upload of the camera-ready submission, please note the following:

- Upload your paper as a pdf.
- Upload your paper source files. Either word or a zip file with the latex text will work.
- Upload the signed copyright form (https://tinyurl.com/ybpfodrg). When filing this in make sure to include in the "Title of the book": "BrainLes 2020", in the "Volume editors": "Alessandro Crimi and Spyridon Bakas". Finally, please tick "US" or "Crown" option only if you working directly in one of those institution, not if your university is in the US.

- In case you do not want your paper published please let us know and we will not ask you to review any other papers. In case you fail to inform us we will use your previously submitted paper for the LNCS proceedings. Please note that if you do not want your paper published as part of the LNCS proceedings then you will also not be included as a co-author of the BraTS journal manuscript submission.

2. **Rankings**
   - The top-ranked teams are now available in the BraTS 2020 website: https://www.med.upenn.edu/cbica/brats2020/rankings.html

2. **ArXiV endorsement**
   - Feel free to upload your individual LNCS paper on arxiv or researchgate (or any other repository), and let us know if you need an endorsement to enable you make this upload.

4. **BraTS journal manuscript**
   - We are in the process of updating the current arxiv paper with both the 2019 and the 2020 results and include all of you as co-authors. I will update you once the arxiv version is updated.
   - We are also in the process of finalizing the submission of the journal manuscript including a collective analysis over all the BraTS instances (much bigger than expected). We anticipate to be submitting this before the end of the year.

Feel free to email me if you have any questions.

Best wishes,

Spyros

Spyridon Bakas, Ph.D.

Instructor

Center for Biomedical Image Computing and Analytics (CBICA)

Jointly Appointed in the Departments of:

- Radiology

- Pathology & Laboratory Medicine

Perelman School of Medicine

University of Pennsylvania

http://www.med.upenn.edu/cbica/sbakas/

Bibliography

[1] medium, https://medium.com/image-processing-in-robotics/grayscale-binary-and-histogram-2548ffcee6c5, viewed 04/08/2020.

[2] Techterms, https://techterms.com/definition/rgb, viewed 04/08/2020.

[3] S. Lambros & al, Propagation of Segmentation and Imaging System Errors 2017, Atherosclerotic Plaque Characterization Methods Based on Coronary Imaging, 2017.

[4] M. A. Minakshi Gogoi, Image Quality Parameter Detection : A Study, 2016.

[5] Kasban, H., et al. A Comparative Study of Medical Imaging Techniques. 2015.

[6] Medicalradiation, http://www.medicalradiation.com/types-of-medical-imaging/imaging-using-x-rays/, viewed 10/08/2020.

[7] M. Larobina, L. Murino, Medical Image File Formats, Journal of Digital Imaging, 2013.

[8] mayfieldclinic, https://mayfieldclinic.com/pe-mri.htm, viewed 14/08/2020.

[9] wikipedia, https://en.wikipedia.org/wiki/Human_brain. viewed 14/08/2020.

[10] lexico, https://www.lexico.com/definition/brain, viewed 15/08/2020.

[11] M. P. Yair Gozal & al, Brain tumors: an introduction, Mayfield Clinic, 2018.

[12] wikipedia, https://en.wikipedia.org/wiki/Human_skin, viewed 16/08/2020.

[13] A. Krizhevsky & al, ImageNet Classification with Deep Convolutional Neural Networks, nipps, 2012.

[14] G. e. a. Hinton, Deep Neural Networks for Acoustic Modeling in Speech Recognition, DRAFT, 2012.

[15] L. Yann & al, Deep learning, nature, vol. 521, p. 436, 2015.

[16] medium, https://medium.com/intro-to-artificial-intelligence/deep-learning-series-1-intro-to-deep-learning-abb1780ee20, viewed 20/08/2020.

[17] theconversation,https://theconversation.com/deep-learning-and-neural-networks-77259, viewed 20/08/2020.

[18] dataversity, https://www.dataversity.net/brief-history-deep-learning/, viewed 21/08/2020.

[19] Y Guo & al, Depthwise Convolution Is All You Need for Learning Multiple Visual Domain, AAAI, 2019.

[20] towardsdatascience, https://towardsdatascience.com/a-basic-introduction-to-separable-convolutions-b99ec3102728, viewed 27/08/2020.

[21] educba, https://www.educba.com/autoencoders/, viewed 29/08/2020.

[22] M Tan, Q V.Le, EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, arXiv, 2019.

[23] lebigdata, https://www.lebigdata.fr/tensorflow-definition-tout-savoir, viewed 19/10/2020.

[24] Brochure, Amira Software for Life and Biomedical Sciences, Thermo Fisher Scientific.

[25] ledatascientist, https://ledatascientist.com/google-colab-le-guide-ultime,viewed 02/09/2020.

[26] O Ronneberger & al, U-Net: Convolutional Networks for Biomedical Image Segmentation, arXiv, 2015

[27] Towards Data science, https://towardsdatascience.com/introduction-to-images-c9c7abe6bfd2, viewed 15/09/2020.

[28] lifepixel, https://www.lifepixel.com/tag/monochrome, viewed 15/09/2020.

[29] br.pinterest, https://br.pinterest.com/pin/672303050602164808/, viewed 15/09/2020.

[30] researchgate, https://www.researchgate.net/figure/Figure-2-Digital-image-processing-system-The-Digital-Image-Processing-System-consistsfig2324827420, viewed 15/09/2020.

[31] medium, https://medium.com/@ssatyajitmaitra/what-canny-edge-detection-algorithm-is-all-about-103d94553d21, viewed 15/09/2020.

[32] goodtherapy, https://www.goodtherapy.org/blog/psychpedia/midbrain, viewed 15/09/2020.

[33] trainermagazine, https://trainermagazine.com/european-trainer-articles/skin-deep-overcoming-barriers-for-effective-transdermal-drug-delivery/2019/3/21, viewed 15/09/2020.

[34] theappliedarchitect, http://www.theappliedarchitect.com/tensorflow-docker-mnist-classifier-project/, viewed 15/09/2020.

nopageno
article
[18pt,a4paper,twoside,openright,titlepage]book

Abstract

3D medical image processing with deep learning greatly suffers from a lack of data. Thus, studies carried out in this field are limited compared to 2D image analysis related works, where very large datasets exist. As a result, powerful and efficient 2D convolutional neural networks have been developed and trained. In this work, we investigate the way to transfer the performance of a two-dimensional classification network for the purpose of three-dimensional semantic segmentation of brain tumors. We propose an asymmetric U-Net network by integrating the EfficientNet model as part of the encoding branch. As the input data is in 3D, the first layers of the encoder are devoted to the reduction of the third dimension in order to fit the input of the EfficientNet network. Experimental results on validation data from the BraTS 2020 challenge demonstrate that the proposed method achieve promising performance.