E.T.S. de Ingeniería Industrial, Informática
y de Telecomunicación

3D Multi-Class CNN Implementation for Abdominal Aortic Aneurysm Segmentation

Máster Universitario
en Ingeniería Biomédica

Trabajo Fin de Máster

Anne Oyarzun Domeño

Rafael Cabeza Laguna

Pamplona, Marzo 2021

upna
Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

Abstract

An Abdominal Aortic Aneurysm (AAA) is a progressive dilation produced as a consequence of the weakening and degradation of the abdominal region of the aorta. It is a life-threatening condition which requires monitoring and treatment, in the majority of cases. The most common procedure for AAA's treatment is the Endovascular Aneurysm Repair (EVAR), which involves the insertion of a stent graft in the aneurysm through a catheter. Despite the great progression of the technique, there is a lack of standardization in terms of determining the degree and rate of disease progression. This project aims at providing an automatic aneurysm segmentation algorithm, that would let the evaluation of the AAA surveillance. Firstly, the clinical framework regarding the AAA, the EVAR procedure and its complications have been studied, focusing on the main unfulfilled clinical needs. Secondly, a deep analysis into Computerized Tomography Angiography (CTA) has been performed, in which the characteristics of pre and post-operative CTA scans have been evaluated. Thirdly, Deep Learning techniques for medical image segmentation have been studied, introducing the concept of Convolutional Neural Network. Then, the materials and methodology carried out for the implementation of the algorithm has been developed, describing the characteristics of the network and the dataset, emphasizing on the generation of the ground truth mask and the consideration of uncertainty in the annotation process. Once the predictions of the volumes have been obtained, their evaluation has been performed using Jaccard Index and Hausdorff Distance scores. Finally, general conclusions about the obtained results and general overview of the project is presented.

Key words: Abdominal Aortic Aneurysm (AAA), surveillance, Computerized Tomography Angiography (CTA), segmentation, Convolutional Neural Network (CNN).

Resumen

El Aneurisma de Aorta Abdominal (AAA) es una dilatación progresiva de la aorta abdominal producida por el debilitamiento y la degradación de la región. Es una condición que compromete a la vida y requiere constante monitorización y tratamiento, en la mayoría de los casos. El procedimiento más común es la Reparación Endovascular (en inglés, EVAR), que consiste en la colocación de un stent a través de un catéter. A pesar de los grandes avances en torno a esta técnica, existe una falta de estandarización en cuanto a la determinación del grado y ratio de progresion de la patología. Este proyecto tiene como objetivo la implementación de un algoritmo de segmentación automática, que pueda ser útil en la monitorización de la patología. En primer lugar, se ha realizado un estudio del caso clínico del AAA y de su tratamiento, y se han analizado las principales necesidades clínicas que han motivado el desarrollo de herramientas automáticas. En segundo lugar, se han estudiado las características de las Tomografías Computerizadas de Angiografía (en inglés, CTA), y las principales diferencias que presentan los estudio pre y post-operatorios. En tercer lugar, se han descritos algunas técnicas de Aprendizaje Profundo para la segmentación de imágenes médicas y se ha introducido el concepto de Redes Neuronales Convolucionales. En cuarto lugar, se ha analizado el conjunto de volúmenes disponibles y se han generado las máscaras para el entrenamiento. Una vez obtenidos los resultados de la segmentación, se han utilizado métricas como el Índice de Jaccard y la Distancia de Hausdorff para su evaluación. Finalmente, se han expuesto las conclusiones generales obtenidas tras la implementación del algoritmo automático.

Palabras clave: Aneurisma de Aorta Abdominal (AAA), monitorización, Tomografía Computerizada de Angiografía (CTA), segmentación, Red Neuronal Convolucional (CNN).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The *Abdominal Aortic Aneurysm (AAA)* is a localized enlargement of the abdominal region of the aorta, the main blood vessel in the human body. If none treatment is carried out, the aneurysm tends to grow, and sometimes, it ruptures causing a life-threatening bleeding. The mechanisms of AAA formation are unknown and the majority of patients are asymptomatic. The diagnostic of the AAA is mainly made by a *Computerized Tomography Angiography (CTA)*, that evaluates the anatomical shape of the aneurysm and enables the measurement of aneurysm's diameter.

Regarding its treatment, open surgical repair has been the gold standard for many years. Now, minimally invasive *Endovascular Aneurysm Repair (EVAR)* is used in the majority of interventions, which requires individualized pre-operative planning and post-operative follow-up. In addition to the importance of correct planning, post-operative follow-up is considered of vital importance to ensure continued success of the surgery and possible rupture prevention. Despite the great progression in the diagnostic and treatment procedures, there are some unfulfilled needs and requirements. Imaging modalities suffer from a lack of standardization in terms of determining the rate of disease progression and a clear variability exists in measuring aneurysm diameter. Furthermore, there is a notable need of automatic aneurysm segmentation tools that would enable precise measures.

Image segmentation has to do with the technique of dividing an image or volume into characteristic areas, to extract objects of interest. This identification process results of vital important in the efficiency of diagnostic, treatment, and even more, in the assessment of any condition or pathology, as the AAA can be.

Regarding the image-based analysis of the AAA, post-operative images are used for the evaluation of EVAR, detection of possible endoleaks and assessment of AAA rupture risk. At this stage, a comparison of CTA scans taken at different time points is required, in order to evaluate the changes over time. However, there is a lack of image-based computerized tools that work with post-operative images, as the majority of them are developed for the pre-operative stage. In addition, direct application of pre-operative algorithms seems not to work properly, mainly due to the presence of the stent-related artifacts in the image.

For many years, image segmentation tools have been built on traditional methods such as thresholding or active contours, but now, Deep Learning approaches have became dominant. These functions simulate the performance of the human brain in data processing and pattern creation tasks. In this field, the main architecture used for image processing is the *Convolutional Neural Network (CNN)*, that automatically detects the important features of an image, without any supervision.

Thus, this project proposes the implementation of a 3D CNN for the multi-class segmentation of lumen and aneurysm, in order to carry on the follow-up of the pathology using post-operative images. This approach enables to quantitatively analyse obtained results and may aid clinicians during the post-operative surveillance.

Chapter 2

Objectives

The main objective of this Final Master Project has been the implementation of a 3D Convolutional Neural Network for the segmentation of Abdominal Aortic Aneurysm, as a CTA image based automatic approach, for the assessment of rupture risk and progression of the aneurysm. The goals of the implementation of the automatic tool have been:

1. The development of post-operative specific Computerized Tomography Angiography based automatic tool, that would face the segmentation problems arising at this stage, regarding the presence of the stent.
2. The extension of Convolutional Neural Network's capabilities in order to achieve the multi-class segmentation of the Abdominal Aortic Aneurysm, in which both the lumen and the aneurysm are segmented.

Part I

Clinical Framework

Chapter 3

The Aorta

This part provides the clinical framework required for better understanding the context of the project. It includes the anatomical description of the aorta and focuses on the most relevant section of this artery, that is the abdominal aorta. Secondly, it analyses the Abdominal Aortic Aneurysm (AAA) condition and its symptomatology. Once the pathology is described, this part focuses on its treatment, specifically on Endovascular Aneurysm Repair (EVAR), since it is the point of emphasis of the project. Finally, some unmet clinical needs are studied, which their support actually motivates the development of this project.

The aorta is the largest artery in the organism and it receives the cardiac output from the left ventricle and supplies the cells with oxygenated blood by the systemic circulation [1]. This artery is divided into four sections: the ascending aorta, the aortic arch, the thoracic (descending) aorta and the abdominal aorta, and ends at the level of L4 vertebra by bifurcating [1] into the left and right iliac arteries.

3.1. Ascending Aorta

The Ascending Aorta arises from the left ventricle and ascends to become the aortic arch. At the level of the aortic valve, the left and right aortic sinuses are located and they give rise to the left and right coronary arteries that supply myocardium (see Figure 3.1).

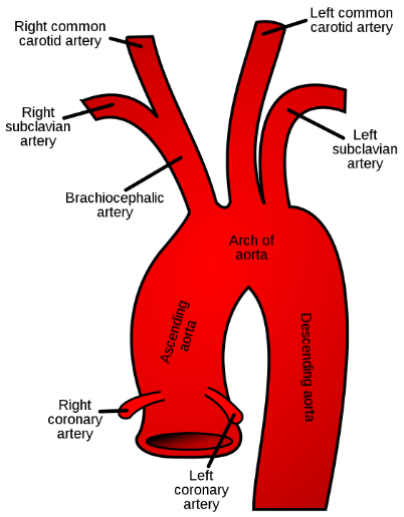


Figure 3.1: Ascending Aorta. Gray, H. (1918), Anatomy of the Human Body. Accessed on 02-02-2021.

3.2. Aortic Arch

The next section of the aorta is the Aortic Arch, that begins at the level of the second sternocostal joint and ends at the level of the T4 vertebra [1] (see Figure 3.1). There are three main branches that arise from this section (proximal to distal) [1]:

– Brachiocephalic trunk. The first and the largest branch that splits into the right carotid and right subclavian arteries. They supply the right side of head and neck and the right upper limb.

– Left carotid artery. It supplies left side of head and neck.

– Left subclavian artery. It supplies the left upper limb.

3.3. Thoracic Aorta

The descending aorta is also known as Thoracic Aorta, that spans from the level of T4 to T12 vertebra. It begins from the left of the vertebral column, approaching the mid-line as it descends. It leaves the thorax via the aortic hiatus and becomes the abdominal aorta [1]. The Figure 3.2 shows its location.

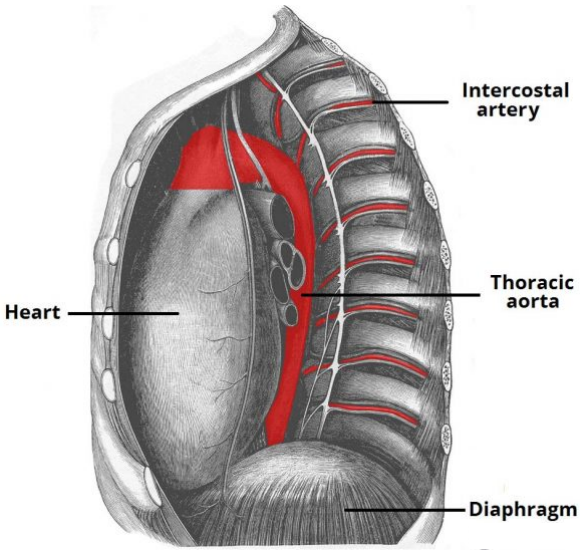


Figure 3.2: Lateral view of the thoracic aorta, with the intercostal branches shown. TeachMeSeries Ltd (2020). Accessed on 02-02-2021.

The main branches arising from Thoracic Aorta are [1]:

– Bronchial arteries that supply bronchial and peribronchial tissue and visceral pleura.

– Mediastinal arteries that supply the lymph glands.

– Oesophageal arteries that supply oesophagus.

– Pericardial arteries that supply dorsal portion of the pericardium.

– Superior phrenic arteries that supply the superior portion of the diaphragm.

– Intercostal and subcostal arterial. Intercostal arteries supply the intercostal spaces and subcostal arteries supply the flat abdominal wall muscles.

3.4. Abdominal Aorta

Continuing from the Thoracic Aorta, the Abdominal Aorta begins at the level of the T12 and ends on L4 vertebra. It is approximately 13 cm long and at the end, it bifurcates into the right and left iliac arteries that supply lower body [1]. The abdominal aorta is represented in the Figure 3.3.

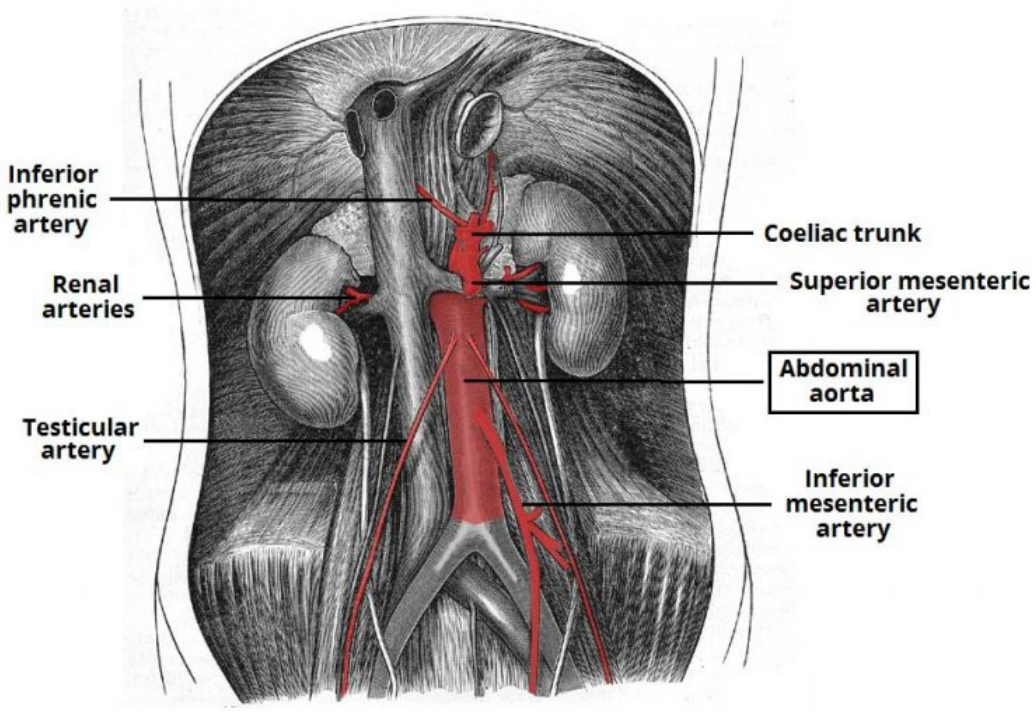


Figure 3.3: Abdominal Aorta. TeachMeSeries Ltd (2021). Accessed on 02-02-2021.

Chapter 4

Abdominal Aortic Aneuysm

An *Abdominal Aortic Aneurysm (AAA)* is a progressive balloon-like dilation produced as a consequence of weakening and degradation of the abdominal region of the aorta. The AAA is a life-threatening condition that requires monitoring and treatment, in the majority of cases. The incidence of this condition increases after age 60 and peaks in the seventh and eight decades of life [2].

In general, the mechanisms of AAA formation are, to a great extent, unknown. It tends to arise due to a failure of the structural proteins of the aorta; such as elastin and collagen. Even if the cause of protein failure is still unspecified, they cause the gradual weakening and loss of structural integrity of the aorta. The aortic wall is composed of collagen lamellar units. The Infrarenal Aorta presents higher incidence than Thoracic Aorta (see Figure 4.1). This is because of the lower number of lamellar units in the Infrarenal Aorta [2]. Besides, there are other factors that contribute to the development of the aneurysms, including genetics or proteolytic degradation of the connective tissue in the aortic wall [3].

AAA is defined as a localized dilatation of the vessel, at least 150 % compared to a relative normal adjacent diameter of that artery [4]. Larger aneurysms have higher risk of rupture due to the *Law of Laplace*; wall stress is proportional to the radius of the aneurysm [2]. Hypertension also increases rupture risk.

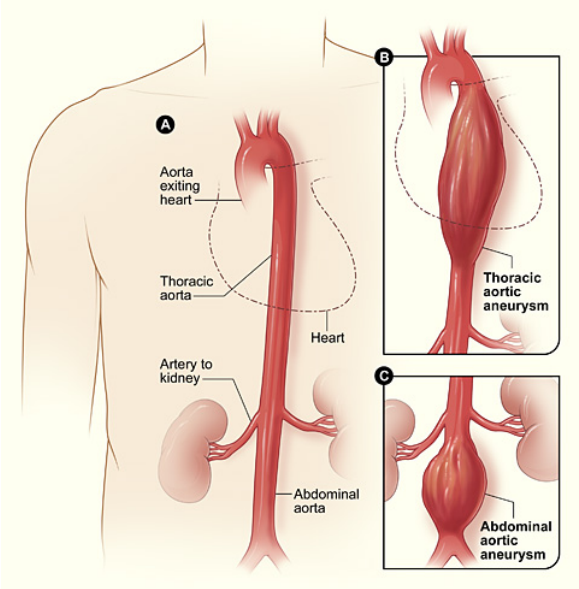


Figure 4.1: A) Aorta and surrounding organs. B) Thoracic aortic aneurysm. C) Abdominal aortic aneurysm. Extracted from [1]. Accessed on 02-02-2021.

Related to symptomatology, most patients are asymptomatic and they are mostly identified incidentally during examination for another unrelated pathology [2]. Sometimes, palpation of the abdomen can revel a pulsatile abdominal mass or some abdominal, flank or back pain can appear [2]. Its mayor risk is its rupture, which is life-threatening. In fact, most patients with AAA rupture die before hospital arrival. The localization of these aneurysms can vary, but most of them are located below the origin of renal arteries [2]. They are also classified into saccular (localized) or fusiform (circumferential), being the fusiforms the 90 % of the total (see Figure 4.2).

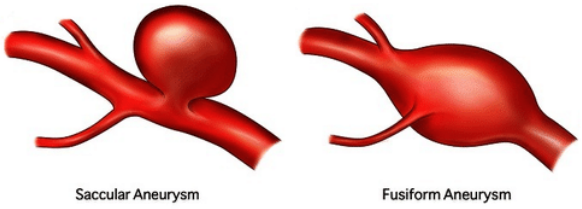


Figure 4.2: Types of AAA. Left: Saccular Aneurysm. Right: Fussiform Aneurysm. Source: https://openi.nlm.nih.gov/. Accessed on 02-02-2021.

4.1. Diagnostic of AAA

Ideally, the diagnosis of AAA should be made before any development of symptoms, in order to prevent its rupture. As mentioned before, a high percentage (30 %) of asymptomatic AAAs are found accidentally [5]. Besides, physical examination is sometimes practised, even if it presents high inter-observer variability.

Because of the limitation of certain techniques, diagnosis of the condition is usually made by *abdominal ultrasound (US)* or *ultrasonography (USG)*. This image modality is considered the screening modality of choice due to its high sensitivity and specificity between 95 and 100 % and its low cost [5]. US is less accurate for diagnostic of aneurysm above the renal arteries because of the overlying air-containing lung and viscera.

In addition to US screening, in order to determine the exact location, size and mapping of other vessels, a *Computerized Tomography (CT)* or a *Computed Tomography Angiography (CTA)* scan is required. It evaluates the anatomical shape of the aneurysm, including tiny details of mesenteric or iliac arteries [5]. It is also preferred to monitored the growth. Disadvantages of CTA include high costs, requirement of contrasts, exposure of radiation and the presence of lower accuracy in determining the shape of the AAA comparing with contrast angiography [5]. Figure 4.3 depics axial CTA and US slice from different AAA cases.

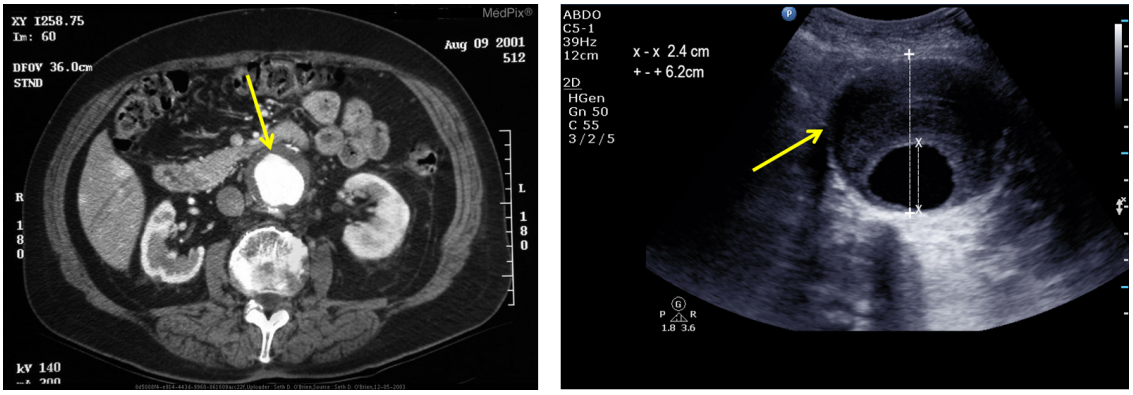


Figure 4.3: Imaging diagnosis. Left: Abdominal CTA with AAA pointed. World Federation for ultrasound in medicine and biology (WFUMB). Right: Abdominal US with AAA pointed. MedPix (2013). Accessed on 02-02-2021.

For those patients with allergies to contrast, *Magnetic Resonance Imaging (MRI)* is an option. Besides, echocardiogram is suggested if the patients tends or presents associated heart disease.

4.2. Rupture risk

Prevention of rupture of AAA is crucial and it is commonly determined by its diameter, considered the strongest predictor of this fatal event. A statement from the *Joint Council of the American Association for Vascular Surgery and Society for Vascular Surgery* estimated (see Table 4.1) the annual rupture risk according to the diameter of the AAA [5]. Summarizing, it has been shown that in the 5 years following the diagnosis of the aneurysm, rupture occurs in approximately the 2 % of AAAs less than 4 cm in diameter and more than 25 % larger than 5 cm.

Table 4.1: Estimated annual rupture risk according to the diameter of the AAA. Extracted from [5].

Diameter of AAA (cm)	Rupture risk (%)
<4	0
4 - 4.9	0.5 - 5
5 - 5.9	3 - 15
6 - 6.9	10 - 20
7 - 7.9	20 - 40
>8	30 - 50

In the actual clinical context, the tendency is to surgically repair large AAAs (with diameter larger than 5.5 cm) and monitor smaller ones (those with diameter smaller than 5.5 cm) [6]. Besides, recent researches have doubt over the suitability of surgical AAA repair based solely on the size of the aneurysm and that factors other than size [6] –as the measurement of wall stress– should be considered in this decision-making process. As it is known, small AAAs can also be ruptured and large AAAs can remain stable. Figure 4.4 depicts an EVAR approach.

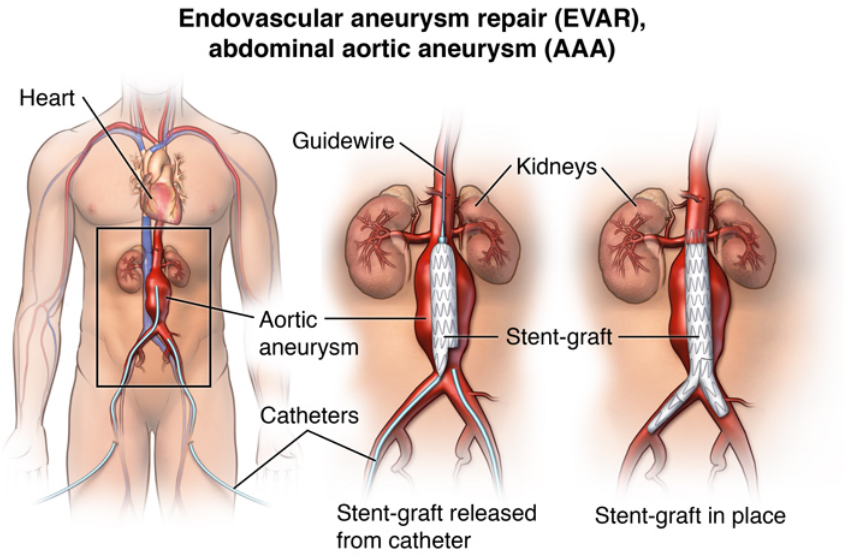


Figure 4.4: Endovascular aneurysm repair (EVAR) for abdominal aortic aneurysm (AAA). Burkett, M. et al. (2017). Accessed on 08-02-2021.

So, in addition to the size of the AAA, its expansion rate is also considered a good predictor of rupture risk. For example, a small AAA that expands 5 cm or more over six months is considered of high risk [5]. Not only that, but other factors as continued smoking, uncontrolled hypertension and increased wall stress are also predictors for rupture risk of AAA [5]. Nevertheless, the size of AAA still continues being the strongest predictor in the determination of the rupture risk.

Chapter 5

Treatment and management of AAA

The treatment of unruptured AAA has changed over years. As previously said, it mainly depends on the size or diameter of the aneurysm and its application is indicated when the risk of rupture exceeds the risk of surgery [7].

For many years open surgical repair via transabdominal or retroperitoneal approach has been the gold standard. Now, the minimally invasive alternative known as *Endovascular Aneurysm Repair (EVAR)* is used in the majority of treatment approaches. In the EVAR procedure, a stent graft is inserted in to the aneurysm through a catheter into the femoral artery and is then expanded inside the aorta and fastened in place to form a stable channel to blood flow [8], preventing the aneurysm also from rupturing.

Despite the great benefits EVAR presents over traditional open surgery, in EVAR, the aneurysm is excluded from blood circulation but it is not removed [9], what can lead to long term complications. In addition, patients may require lifelong surveillance to control their progress and prognosis postoperatively, consisting on yearly CTA scan.

In those patients in which surgical approach is selected, it is important to properly select suitable patients and timings in order to prevent aneurysm rupture. If rupture occurs, emergency repair is indicated, as mortality is very high. The 2005 ACC/AHA guidelines recommend surgical repair when AAA is 5.5 cm in diameter or greater, for asymptomatic patients [5]. In case of symptomatic aneurysms that grow by 0.5 or more in diameter in six months, should carry on repair, regardless of aneurysm diameter [5].

For asymptomatic patients with aneurysms between 4 and 5 cm in diameter, regular imaging monitoring is required, as sometimes their frequency of surveillance after surgical intervention is difficult to evaluate. The 2005 ACC/AHA guidelines indicates that AAAs with 3 cm to 4 cm in diameter should be monitored by US examination every two to three years, and AAAs with 4 cm to 5.4 cm should be monitored by US or CTA every 6 to 12 months [5].

In addition, in patients with small to medium sized aneurysm that are not treated surgically, strategies like pharmacological management, cessation of smoking and management of risk factors as hypertension [5] can be performed.

5.1. EVAR planning and follow-up

EVAR requires individualized pre-operative planning and post-operative follow-up.

Regarding the planning, this is done on the basis of CTA imaging and consists on quantifying diameters and lengths along the aorta [10], to select the stent that better suits anatomical features of the patient.

Conversely, the post-operative follow-up is important to ensure continued success of the surgery, as it can detect and prevent of endoleaks and aneurysm sac expansion [10]. When the aneurysm diameter decreases, it means that the progression is positive, whereas a stable or enlarged aneurysm requires further monitoring or even reintervention. The recommended protocols state that CTA

imaging is required at 1 month, 6 months and 1 year after the intervention [10]. Nowadays, post-operative follow-up is based on visual assessment of CTA imaging and measurement of maximum aneurysm diameter [10].

5.1.1. Complications: endoleaks

An endoleak consists of aneurysm intra-cavitatory blood flow accumulation and affects about 15-25 % of patients who have EVAR [11]. If not treated, the aneurysm can expand due to pressure increase and rupture risk may augment, which sometimes leads to reintervention [10]. Endoleaks are diagnosed during or after EVAR intervention by US, CTA or MIR imaging, testing whether the procedure was successful or not. This procedure allows to detect endoleaks. Some types of endoleaks are defined, depending on the site where appear (see Figure 5.1).

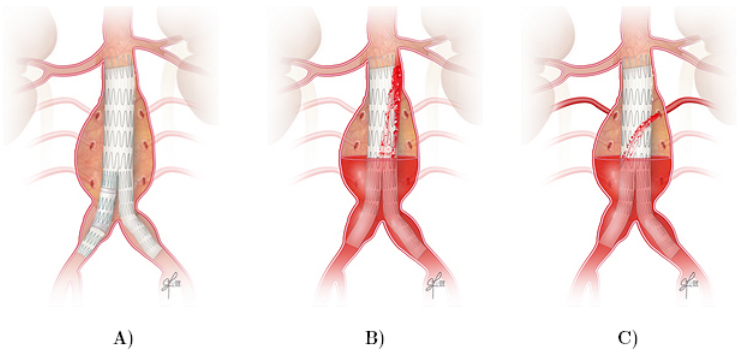


Figure 5.1: Types of endoleaks. A) None endoleak present. B) Endoleak present around the top or bottom of the stent graft (Type I). C) Endoleak present when blood flows into the aneurysm sac cavity through small branches (Type II). Cleveland Clinic (2019). Accessed on 08-02-2021.

5.2. Unfulfilled clinical needs and requirements

Despite the great progression in AAA diagnostic and treatment, there are some unfulfilled needs that should be pointed out.

As stated by *The Society for Vascular Surgery practice guidelines on the care of patients with an abdominal aortic aneurysm* both modalities for aneurysm imaging – US and CTA– suffer from a lack of standardization in terms of determining the degree and rate of disease progression [12].

Besides, it exists clear variability in measuring aneurysm diameter. Some studies included diameter measurements based on outer wall to outer wall, inner wall to inner wall or anterior outer wall to posterior inner wall and some intra and inter-observer variability [12]. As said before, other factors other than size of the aneurysm could be used together with it for better determining and establishing the risk of rupture of the aneurysm and evaluating the best treatment approach.

Furthermore, there is a high need of automatic aneurysm segmentation algorithms that would let proper and precise diameter quantitative measurement. This is of high importance in the post-operative stage, as the follow-up still lacks of optimized image analysis tools that enable personalized evaluation and monitoring of the patient.

Part II

CTA Imaging Surveillance

Chapter 6

Computed Tomography Angiography (CTA)

CT angiography (CTA) is considered the gold standard for diagnosis and evaluation of AAAs, and it is superior to US modality in detecting aorta's branches. The *Houndsfield Unit (HU)* is a semi quantitative measure that describes radio-density in medical CT and provides an accurate density for the type of tissue [13]. On the Hounsfield Scale, air is represented by a value of -1000 HU (black on the grey scale) and bone between +700 HU (cancellous bone) and +3000 HU (dense bone). Soft tissues and organs represent narrow HU ranges and are therefore, more difficult to differentiate between adjacent structures. Thus, artificial contrast agents are injected into patients' blood flow, that make some structures stand out more strongly, as they absorb X-ray energy.

6.1. Pre-operative imaging of the aorta

In the preoperative stage, CTA images are mainly used to determine the rupture risk of the AAA, measuring the maximal AAA diameter in axial slice planes, to ensure patient suitability for EVAR intervention and to select the implantable

device that better suits them. As said before, the use of contrast media is required, as it enhances blood flow and improves the visualization of the aortic lumen [10]. When deciding if an AAA is elective for EVAR, the whole aneurysm diameter –including luminal area– is measured [10]. Conversely, for endograft sizing at pre-operative planning, the relevant structure is the lumen [10] and its corresponding dimensions. Because of that, correct segmentation and lumen isolation is of vital importance for properly quantifying aortic diameters and sizes.

6.2. Post-operative imaging of the aorta

Post-operative CTA images are used for evaluation of EVAR, detection of possible endoleaks and other complications and assessment of AAA rupture risk after intervention. This stage requires the comparison of CTA scans taken at different time points to evaluate the changes along time [10]. Even the importance of this follow-up step, there is a lack of image-based computerized monitoring tools.

The difference between pre and post-operative CTA images is that these last present an endograft. In fact, the presence of these metallic devices represent one of the major problems in post-operative surveillance, as they generate stent-related beam-hardening artifacts (see Figure 6.1).

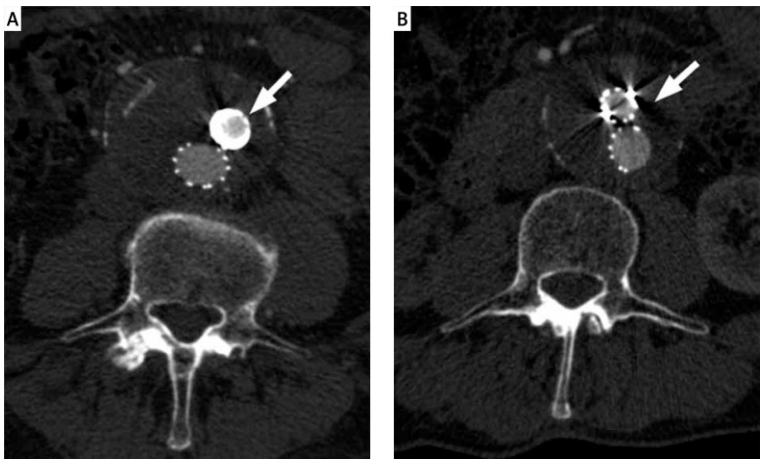


Figure 6.1: A) Stent-related blooming artefacts. B) Photon-starving artifacts. Wideochir Inne Tech Maloinwazyjne (2019). 14(1):1-11. Accessed on 08-02-2021.

Chapter 7

Lumen and aneurysm segmentation

This chapter provides a general overview of the existing algorithms for lumen and aneurysm segmentation. The lumen is relatively easy to segment, because the contrast agent injected to patient during the examination highlights blood flow on the image [14]. In contrast, aneurysm does not receive any contrast media, so it is not highlighted. Besides, the HU of the thrombus is similar to many tissues –usually between 0 and 100 HU [14]– which results in low contrast between the aneurysm and surrounding tissues. Also, conversely to organ segmentation, position and shape of the aneurysm is heterogeneous among the population [14]. In addition to stent-related artifacts presented in the image, there could be calcium deposits that may become the task of aneurysm segmentation even more difficult. Regarding the available tools, there are many segmentation techniques that have traditionally been used for lumen and aneurysm segmentation. This chapter briefly describes some of them.

7.1. Lumen segmentation

Regarding lumen segmentation algorithms, most of them are implemented in the pre-operative stage. These techniques include graph-based methods, that extract volume contours by applying intensity and shape related criteria based on the strong gradient generated between the contrasted lumen and surrounding structures [10]. Besides, centerline-based approaches aim at obtaining the vessels' centerline, to lately extract the lumen contour (see 7.1.1).

In the case of post-operative stage, segmentation and extraction techniques of the lumen are barely studied and implemented. In addition, direct application of pre-operative algorithms to the post-operative stages seems not to work properly [10], due to the presence of the metallic stent.

Some problems arising at the post-operative stage are:

- Stent-related artifacts, that hinder the performance of intensity-based algorithms [10], as a consequence of increased contrast or presence of obscured zones.
- Creation of unique lumen [10] because of the close proximity of both iliac arteries, causing difficulties in boundary detection.

7.1.1. *eVida Vascular* application

Among the commercial solutions for pre-operative stage lumen segmentation, *Vicomtech Foundation* –which is specialised in digital technologies related to Visual Computing and Interaction and Artificial Intelligence–, has developed *eVida Vascular* application. It consists of 3D Aorta inspection and analysis software. The tool contains a *segmentation module* that starts by defining a volume of interest and a seed point. Then, the computation of the segmentation is automatically proceed. Control parameters are set to default, but the user can modify them in order to

obtain better segmentation results; such as window-level adjustment. It also offers the possibility of eroding and dilating the segmentation results [15].

In addition to the segmentation module, it offers an *EVAR planning module* that involves a specialized interface with specific views and interactive tools adequate for endograft sizing. The module also includes tools to interactively edit the centerline and to deal with difficult datasets or complex cases involving large curvatures [15]. The Figure 7.1 depicts the different modules of the application.

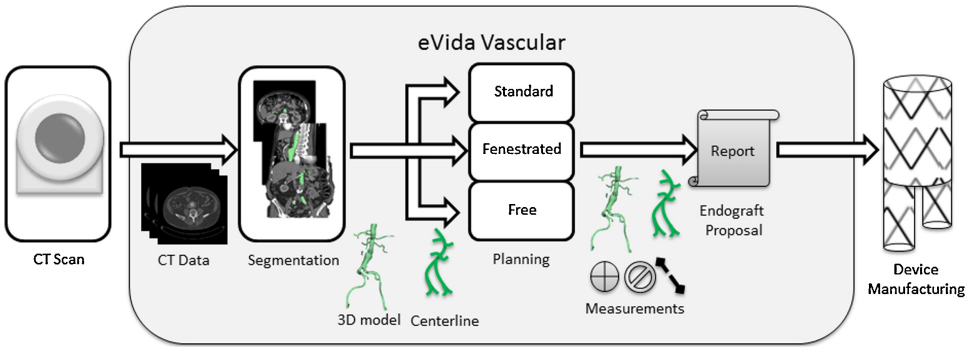


Figure 7.1: Vascular analysis in eVida Vascular application. Extracted from [15]. Accessed on 09-02-2021.

7.2. Aneurysm segmentation

Segmentation of the AAA is of vital importance, as its maximal diameter measurement is used as a predictor of aneurysm rupture risk. Besides, other factors as wall-stress, volume and morphology changes of the aneurysms are correlated with AAA rupture risk. Thus, there is a significant need in the development of automatic aneurysm segmentation algorithms.

However, there are some characteristics that difficult their development [10]

- Similarity of intensity values of the aneurysm and surrounding tissues, that causes segmentation algorithms to fail due to fuzzy boundaries.
- Obscured surface of the aneurysm since it is not a contrasted area.

– Stent-related artifacts that may cover some of the thrombus region.

– The geometric structure of the aneurysm is irregular, so it is not possible to approximate it using geometrical models.

There are many segmentation techniques that have been used traditionally, such as intensity-based semi-automatic algorithms combined with gradient-based segmentation [16]. Besides, [17] used a level-set approach with region and statistical information and [18] segmented the AAA using active learning and supervised random forest classifiers; in which active learning techniques select the optimal feature sets, a random forest classifier is trained with these features and finally, a voxel-based segmentation is performed. These all are merely a few examples of all the techniques and tools available for pre and post-operative AAA segmentation. Even if these tools yield good results, they require the optimization of some parameters that reduces the robustness and reproducibility [10] required in a real clinical scenario.

In recent years, deep learning-based approaches are achieving considerable strength in computer vision task solutions, including object recognition, classification and segmentation [19]. All these solutions are being efficiently addressed by various types of deep neural networks, such as convolutional neural networks, recurrent networks or autoencoders [19].

7.3. Motivation of the work

The motivation of this work is the implementation of a Convolutional Neural Network for post-operative multi-class 3D segmentation of lumen and aneurysm from CTA images. In the following sections, a complete description of the proposed method has been carried out.

Part III

Deep Learning for Medical Image Segmentation

Chapter 8

Introduction to Neural Networks

Artificial Neural Networks (ANN) are computational processing systems which are inspired by the way biological nervous system works [20] and consist of high number of interconnected nodes or neurons that collectively learn from the input in order to optimise the final output [20]. Figure 8.1 depicts a comparative between ANN and biological nervous system.

8.1. A single neuron

A *neuron*, node or unit is the basic computational unit of the brain and the basic unit of a *Neural Network (NN)*, that takes some inputs, maths with them and produces an output [21]. Each input has an associated *weight (w)*, that is assigned depending on its importance to other inputs. In a biological system, the dendrites carry the signal to the cell body and they are all summed. If this sum is above a certain threshold, the neuron can *fire* and sends an spike through the axon. In the computational model, the *firing* of the neuron is represented with the *activation function*, that represents the frequency of the spikes along the structure of the axon [21].

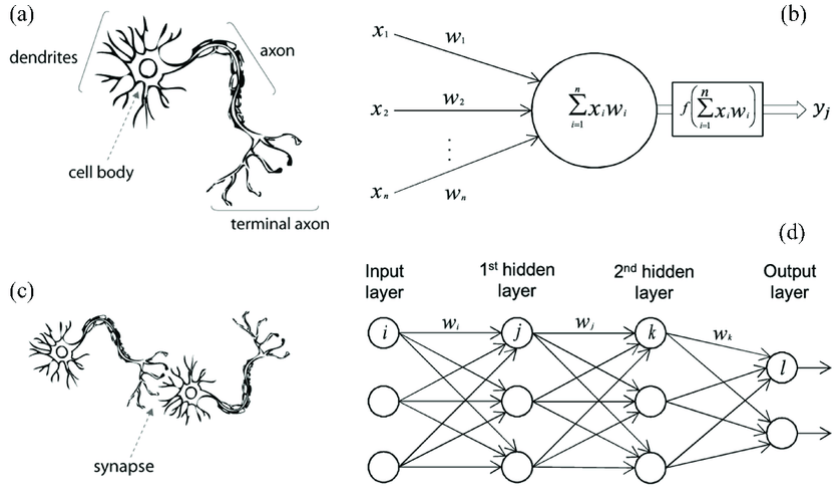


Figure 8.1: A biological neuron in comparison to an artificial neural network. A) Human neuron. B) Artificial neuron. C) Biological synapse. D) ANN synapses. Extracted from [22]. Accessed on 14-02-2021.

There are some important elements to consider in the NN architecture. In the network represented in Figure 8.1B, [*x1*, ..., *xn* are the inputs and have weights [*w1*, ..., *wn*] associated with them. *Y* is the output and *f* the activation function. The neural network is a connection system where the output of a predecessor *i* is transferred to the input of a successor neuron *j*. A weight *Wij* is assigned to each connection. An important element in the ANN is the *activation function* that defines the output of that node given an input or a set of them [21], similar as a chip circuit that can be *on* (1) or *off* (0), depending on the input. An artificial neuron calculates a weighted sum of its input, adds a bias and then decides whether it should be fired or not [23]. The *bias* is an element that provides every node with a trainable constant value, in addition to normal inputs [21]. It is represented as an input *i* with weight *b*.

Non-linear activation functions in Neural Networks

The *non-linearity* means that the neural network can approximate functions that do not follow linearity or that it can successfully predict the class of a function

that is divided by a decision boundary which is not linear, that will allow the model to create complex mappings between the network's inputs and outputs [24]. They are preferred as they allow the nodes to learn more complex structures in the data [25], comparing with linear activation functions that are easy to train but cannot learn complex mapping functions [25]. The main non-linear activation functions are described below.

Sigmoid and Hyperbolic Tangent (TanH)

There are two widely used non-linear activation functions: Sigmoid or logistic function and hyperbolic tangent (TanH) function.

In the case of *Sigmoid* activation function, the input to the function is transformed into a value between 0 and 1. Inputs that are larger than 1 are transformed to the value 1, same thing happens with values smaller than 0, which are snapped to 0 [25]. Sigmoid is usually used for binary classification or segmentation problems in logistic regression model. *Hyperbolic tangent* function or *TanH* is a shifted version from the Sigmoid function with range between -1 and 1 –zero centered–, which makes the learning for the next layer easier and faster.

Both Sigmoid and TanH functions present a saturation problem, which means that large values snap to 1 and small ones snap to 0 or -1, for Sigmoid and TanH, respectively. Besides, these functions are only sensitive to changes around their mid-point of their input [25]; 0.5 for Sigmoid and 0 for TanH. Once saturated, it is challenging for the learning algorithm to continue adapting the weighs to improve the performance of the model [25]. The error is back propagated through the network and used to update the weights. Here, it arises the problem of *vanishing gradient*; the amount of error decreases with each additional layer through which is propagated, given the derivative of the chosen activation function. This problem prevents deep networks from learning effectively as they make difficult to know which direction the parameters should move to improve the cost function [25]. Figure 8.2 represents the Sigmoid and TanH activation functions.

ReLU

Rectified Linear Unit (ReLU) is a linear function that will output the input directly if it is positive, otherwise, it will output zero. It acts as a linear function, but is, in fact, a non-linear function allowing complex relationships to be learned [25]. It is highly used as it is easier to train an often achieves better performance, as it is nearly linear, preserving many of the properties that make models easy to optimize with gradient-based methods [25]. In the Figure 8.2 there is a comparative between ReLU and Sigmoid.

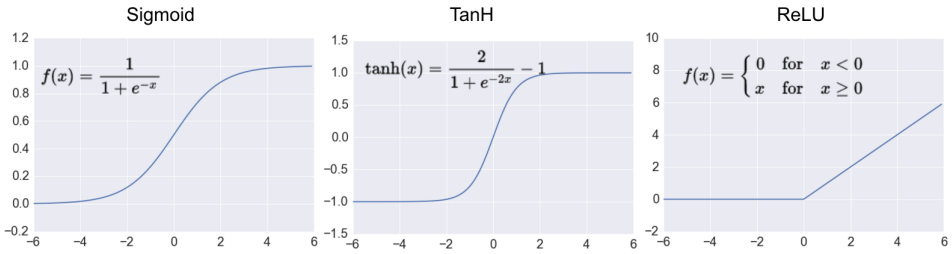
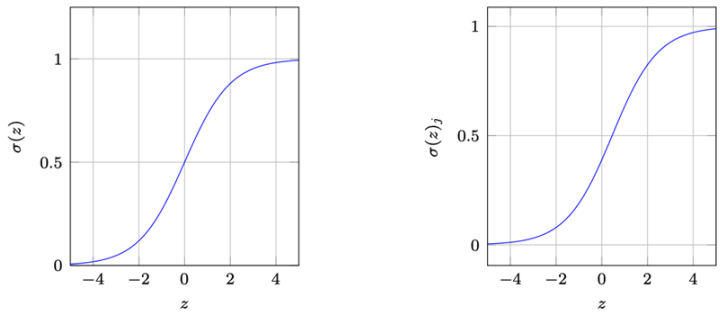


Figure 8.2: Sigmoid, TanH and ReLU. Extracted from [26]. Accessed on 15-02-2021.

Softmax

Softmax activation function returns a probability distribution over the target classes in a multi-class classification problem. The main difference between Sigmoid and Softmax is that even both give output in [0, 1] range, Softmax ensures that the sum of outputs along channels is 1, while Sigmoid makes each class output between 0 and 1 [24] (see Figure 8.3).



(a) Sigmoid activation function. (b) Softmax activation function.

Figure 8.3: Sigmoid and Softmax. Extracted from [27]. Accessed on 15-02-2021.

8.2. Feedforward and Backward Neural Networks

Feedforward network is the simplest type of ANN, where the information moves in only one direction (see Figure 8.4), from the input nodes, through the hidden nodes to the output nodes [21]. It computes the result of an operation and saves any intermediates needed for gradient computation in memory [28]. A feedforward network can be made of *single layer perceptron (SLP)* or *multi layer perceptron (MLP)*. MLPs have one or more hidden layers and are more useful than SLP, as they can also learn from non-linear functions. The process by which a MLP learns is called the *backprogation algorithm* or *backward propagation of errors*. It follows a supervised training scheme. For every input in the training dataset, the ANN is activated and its output is observed. This output is then compared with the desired output, and the error is propagated (error propagation) back to the previous layer (see Figure 8.4). The error is noted and the weights are adjusted until the output error is below a certain threshold [21].

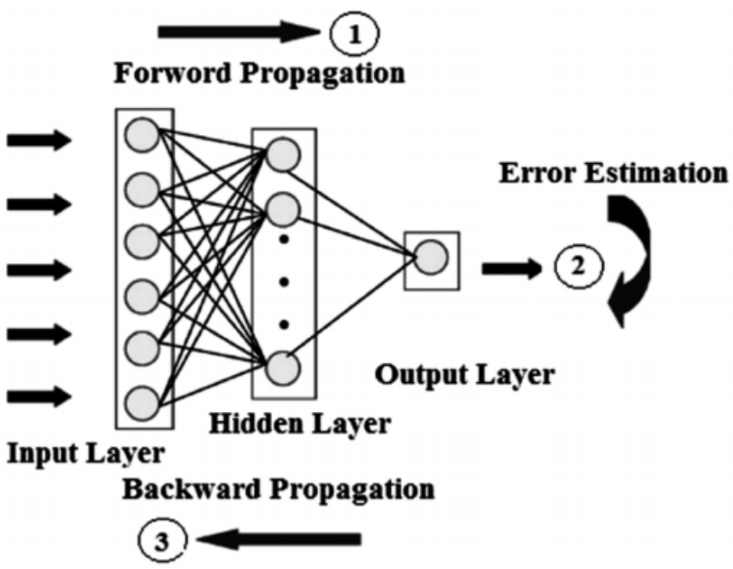


Figure 8.4: Forward and backward neural networks representation. Extracted from [29]. Accessed on 14-02-2021.

Chapter 9

Deep Learning in Healthcare Imaging

This chapter aims at providing a general perspective of deep learning-based tools for medical image segmentation.

For many years, image segmentation systems were built on traditional methods such as thresholding [30], histogram-based bundling [31], region-growing [32], k-means clustering [33], watersheds [34], active contours [35], conditional and Markov random fields [36]. Then, machine learning approaches became a dominant technique for long period, which aim to extract hand-crafted features. In the 2000s, with the arrival of hardware improvement, deep-learning approaches started to demonstrate their capabilities in image processing tasks [37]. *Deep Learning (DL)* is an *Artificial Intelligence (AI)* function that simulates the performance of human brain in data processing and pattern creation tasks, for decision making process. It is a subset of *Machine Learning (ML)* that uses hierarchical level of ANN to perform the process of ML. Many computer vision tasks require intelligent segmentation of an image. Thus, DL for computer vision models are used as today's image segmentation techniques, as they can learn patterns from visual inputs to predict object classes that make up an image [38]. The main DL architecture used for image processing is *Convolutional Neural Network (CNN)* [38].

9.1. Image Segmentation

Image segmentation has to do with the technique and process of dividing an image into characteristic areas or segments and extracting objects of interest [39]. The identification and analysis of certain regions of interest is of vital importance in the efficiency and accuracy of prognostic, diagnostic and treatment procedures.

There are three levels of image analysis [38]. First, the *classification* level consists of categorizing the entire image into a class, such as *animals*. Second, *object detection* consists of detecting objects within an image and drawing a bounding box around them. Finally, in the *segmentation* level, the objective is to identify parts of the image and understand what object they belong to, so that it fuses the process of detection and classification.

There are currently two categories of image segmentation taks; semantic segmentation and instance segmentation. *Semantic Segmentation* classifies all the pixels of an image into meaningful classes of the object [38]. In the case of *Instance Segmentation*, it does not only need to achieve pixel level classification, but also needs to distinguish instances on the basic of specific categories [40]. For example, if there is an image that contains three cars, semantic segmentation will classify all them as a one instance, whereas instance segmentation will identify each individual car. In fact, due to the similarity between organs and tissues, there are few reports [40] on instance segmentation in medical image segmentation.

Chapter 10

Convolutional Neuronal Networks

For medical image segmentation tasks, supervised learning methods are the most popular ones [40], which directly learn from labeled training samples, extracting features and context information in order to perform a dense pixel or voxel-wise classification [10].

Convolutional Neural Networks (CNN) consist of neurons that self-optimise through learning [20], as ANNs do. The main difference between CNNs and traditional ANNs is that CNNs are mainly used in the field of pattern recognition within images [20], that take advantage of the spatial nature of it [10]. With ANN, concrete data points must be provided [41]. For example, in a model where dogs and cats are classified, the width of noses and length of ears must be provided as data points. Conversely, when using CNN, these spatial features are extracted from image input [41]. The use of ANN for image classification problems is difficult because 2-dimensional images need to be converted to 1 dimensional vectors, increasing the number of trainable parameters. CNNs automatically detect the important features without any human supervision. Another difference is that CNNs are comprised of neurons organised into three dimensions; the spatial dimensionality of the input (height and width) and the depth [20]. The depth refers to the activation volume. Unlike ANNs, the neurons within any layer of the CNN will only connect to a small region of the layer preceding it [20].

Besides, CNNs involve three types of layers: convolutional layers, pooling layers and fully-connected layers (see Figure 10.1). First of all, *Convolution Layers* apply a convolutional operation to the input image, producing feature maps corresponding to image edges or specific shapes [10], among others. Secondly, *pooling Layers* are usually applied after convolutional operations, to simplify or summarize the information from the convolutional layers by performing statistical operations, such as averaging, that produces down-sampled feature maps [10]. The convolutional and pooling layers learn about the local spatial context of the input image. Finally, *fully-connected layers* learn at more abstract level and integrate the knowledge from the previous layers to perform the final classification, defining the image content [10]. Through this method of transformation, CNNs are capable of transforming the original input, layer by layer using convolutional and downsampling techniques to produce class scores for classification and regression purposes [20].

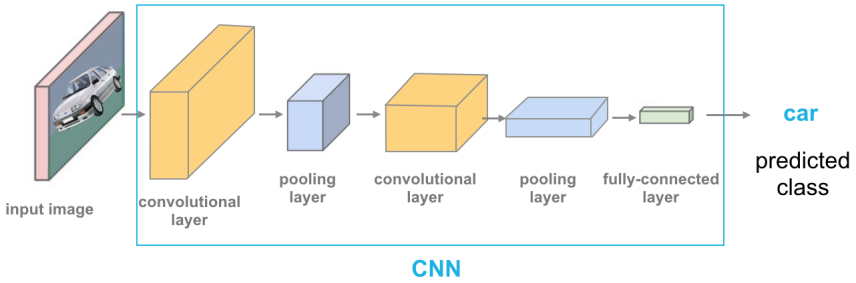


Figure 10.1: A simple CNN architecture, comprised of five layers. Extracted from [42]. Accessed on 15-02-2021.

10.1. Convolutional Neural Network Elements

In order to obtain a good grasp of CNN, the following section aims at providing a general description of the basic elements and characteristics of CNNs.

Convolutions

Let's say that the input of the neural network is an image of shape 32 x 32 x 3 (see Figure 10.2A), with a depth of 3, corresponding with RGB channel. Assuming that the network receives raw pixels as inputs, to connect the input layer to only one neuron, there should be 32 x 32 x 3 connections for the dataset. If adding one more layer, more than 6000 weight parameters will be used. However, it is clear that two neurons might not be enough for any image processing work. In order to make it more efficient, it is possible to connect the input image to the neurons of the next layer with the same values for the height and width [43], needing 32 x 32 x 3 by 32 x 32 connections.

There is a way to make it more efficient, that is looking for local regions in the image instead of in the whole image, as shown in Figure 10.2B. Thus, hidden neurons in the next layer only get inputs from the corresponding part of the previous layer [43], such as 5 x 5. If we want to have 32 x 32 neurons in the next layer, then there will be 5 x 5 x 3 by 32 x 32 connections, which is 76800 connections compared to the more than 3 million connection in the full connectivity. However, it still leaves so many parameters to solve. For further simplification, it is possible to keep the local connection weights fixed for the entire neurons of the next layers, what reduces the weights to 5 x 5 x 3 to connect 32 x 32 x 3 neurons to 32 x 32 in the next layers.

In addition to the reduction of parameters that this approach involves, an interesting concept is that fixing the weights for the local connections is similar to sliding a window of 5 x 5 x 3 in the input neurons and mapping the generated output to the corresponding place [43]. It provides the option to detect and recognize features regardless of their positions in the image and this is the reason why they are called convolutions [43]. There are different convolution matrices that act like a filter. In CNNs, these filters are initialized, followed by the training procedure. With the aim of making it more beneficial, it is possible to add layers after the input layer and each layer can be associated with different filters, extracting different features from the given image looking at the same part of the input image (see Figure 10.2C).

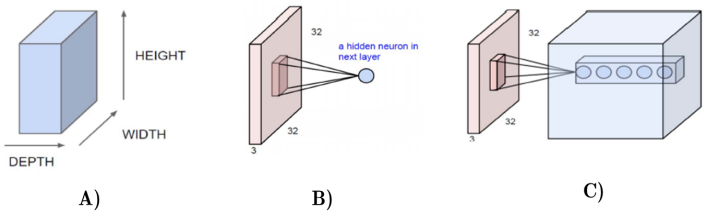


Figure 10.2: A) Three dimensional input. B) Convolution as an alternative for fully connected network. C) Multiple layers that correspond to different filters but looking at the same region of the given image. Extracted from [43]. Accessed on 11-02-2021.

Convolutional formula

The convolution for one pixel in the next layer is calculated according to the Equation 10.1.

$$net(t,f) = (x * w)[t,f] = \sum_m \sum_n x[m,n]w[t-m, f-n] \tag{10.1}$$

Where *net(t,f)* is the output in the next layer, x is the input image and w is the kernel or filter matrix and * is the convolution operation. Figure 10.3 shows how the convolution layer works; how the element by element product of the input and kernel is aggregated, and then how the corresponding point in the next layer is represented [43].

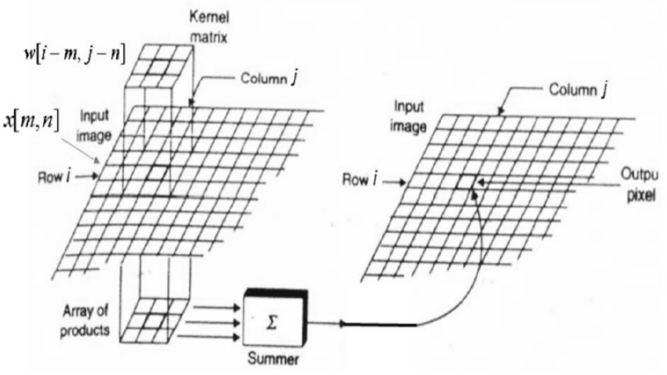


Figure 10.3: Details on Convolution Layer. Extracted from [43]. Accessed on 11-02-2021.

Stride

Another option for paramaters' reduction is the *stride*; that is used to control the overlap. Figure 10.4 depicts an 7 x 7 image; if the filter is moved one node every time, an output of 5 x 5 is achieved (note that the output of the three left matrices have an overlap, as same as three middle and three rights ones). If the stride is set to 2, the output will be 3 x 3 [43].

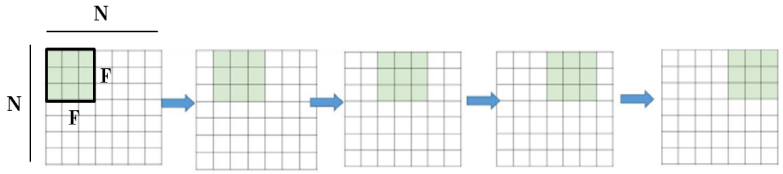


Figure 10.4: Effect of Stride 1, the filter window moves only one time for each connection. Own elaboration based on [43]. Accessed on 11-02-2021.

The following Equation 10.2 represents output size O, based on NxN dimension and filter size of FxF, where N is the input size, F is the filter size and S is the stride size.

$$O = 1 + \frac{N - F}{S} \tag{10.2}$$

Padding

Zero-padding is a a very simple method to resolve the issue of information loss on the border of the image, also giving the chance of managing the output size. For example, adding one zero-padding on the example of the Figure 10.4, provides an output of 7 x 7, the same as the original input (N = 9). The padding helps at preventing network output size from shrinking with depth [43]. The modified formula including zero-padding is represented in Equation 10.3.

$$O = 1 + \frac{N + 2P - F}{S} \tag{10.3}$$

Where P is the number of the layers of the zero-padding.

Pooling Layers

The main idea behind *pooling* is down-sampling in order to reduce the complexity for further layers. *Max-pooling* is the most common type of pooling; it partitions the image to sub-region rectangles, and it only returns the maximum value of that sub-region [43]. Figure 10.5 depicts a common 2 x 2 max-pooling with stride 2. It should be considered that this process needs to be applied when the presence of information (whether a filtered feature was encountered or not) is important rather than spatial information [43].

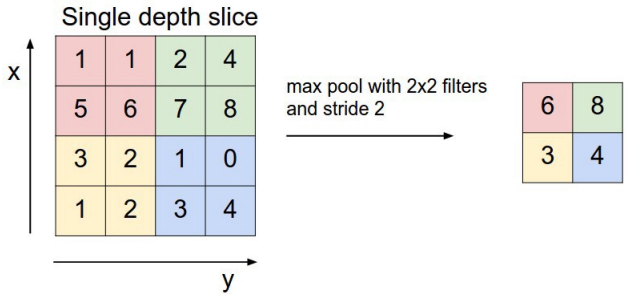


Figure 10.5: Max-pooling with 2 x 2 filter and stride 2. It down-samples each 2 x 2 blocks and the result is mapped into 1 block or pixel. Extracted from [43]. Accessed on 12-02-2021.

Fully Connected Layers

Each node in a fully-connected layer is directly connected to every node in both the previous and in the next layer as shown in Figure 10.6. Its input is the output of the pooling layer, where spatial information is ignored to learn correlation

between different locations [44]. They integrate the knowledge from the previous layers to perform the final classification defining the image content [10].

With the aim of transforming these classification CNNs to perform semantic segmentation recovering, [45] introduced one of the first *Fully Convolutional Networks (FCN)*. These networks adapt classifiers for dense pixel-wise prediction, replacing the final fully connected layers with convolutional layers to preserve local image regions [10]. The concept of *skip connection* was also presented. It allows to transfer local information by summing feature maps from the down-sampling path with feature maps from the up-sampling path [10]. Hence, a skip-net architecture centres on a primary stream and skips connections that incorporate feature responses from different scales in a shared output layer [10]. Examples of CT image segmentation approaches using FCNs include liver and lesion segmentation or pancreas segmentation [10].

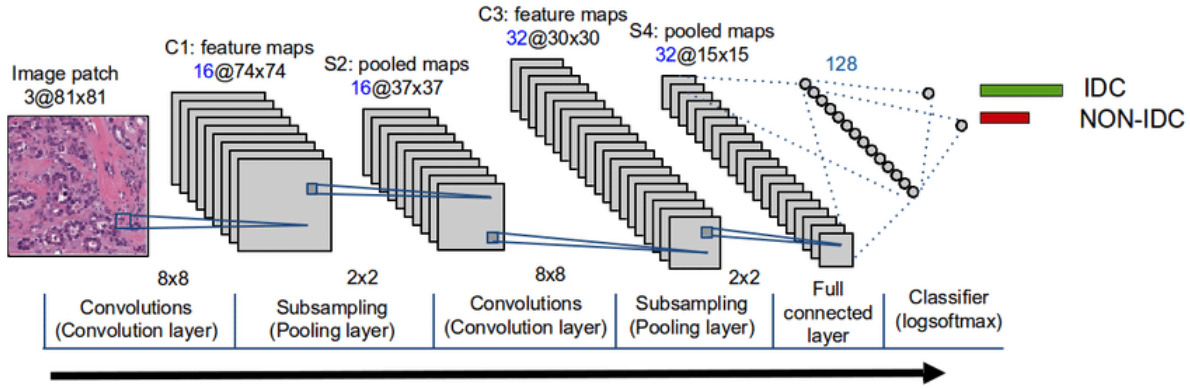


Figure 10.6: 3-layer CNN architecture composed by two layers of convolutional and pooling layers, a full-connected layer and a logistic regression classifier to predict if an image patch belongs to a Invasive Ductal Carcinoma (IDC) tissue or not. Extracted from [44]. Accessed on 12-02-2021.

Chapter 11

Encoder-decoder architectures

As previously seen, semantic segmentation aims to achieve pixel classification of an image. In 2015, Ronneberger et al. [46] proposed a new CNN architecture, designed for biomedical image segmentation; the *U-Net*, which has a encoder-decoder structure.

In these structures, an *encoder* or *contraction path* [47] is used to restore the extracted features and it consists of a traditional stack of convolutional and max pooling layers [47]. A *decoder* or *symmetric expanding path* [47] is used to restore extracted features to the original image size and to output the final segmentation result [20]. It has a symmetric U-shape, hence the name U-Net. Besides, it is an end-to-end FCN; it just contains convolutional layers and does not contain any dense layers [47]. Figure 11.1 depicts the original architecture of the U-Net.

Usually, medical images contain noise and present blurred boundaries, so that it is difficult to detect or recognize objects. U-Net fuses low-level and high-level image features by combining low-resolution and high-resolution feature maps through skip connections, so that these kind of networks deal with the disadvantage of medical images mentioned above.

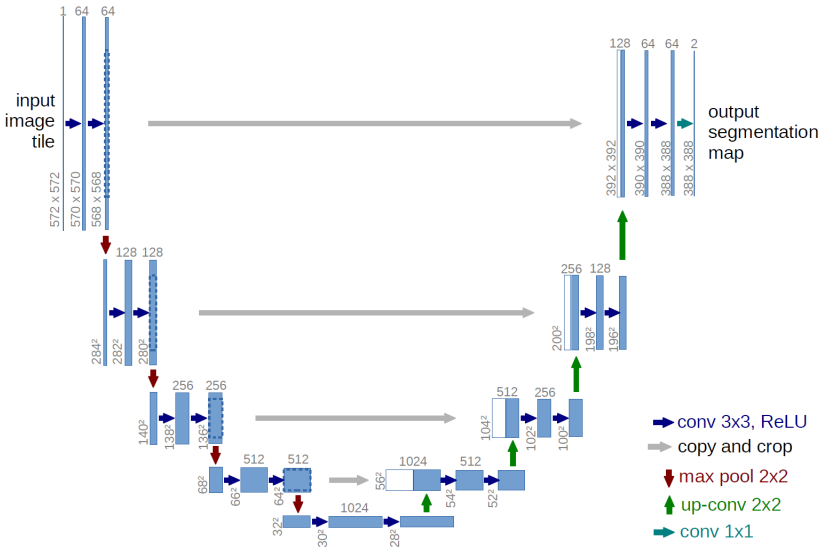


Figure 11.1: Original U-Net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y size is provided at the lower left edge of the box. White boxes represent copied feature maps and the arrows represent the different annotations. Extracted from [46]. Accessed on 12-02-2021.

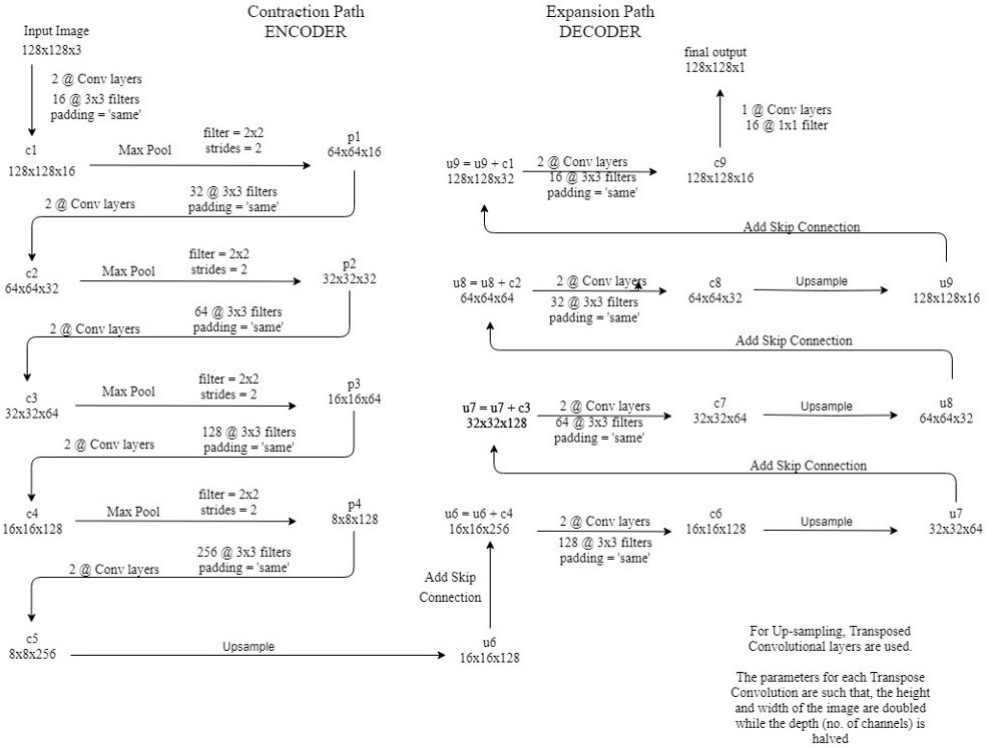


Figure 11.2: Detailed U-Net Architecture. Extracted from [46]. Accessed on 12-02-2021.

The detailed explanation of the architecture is in the Figure 11.2. The left hand side is the encoder, which includes convolution and max pooling layers. In the encoder, the size of the image gradually reduces while the depth gradually increases (starting from 128 x 128 x 3 to 8 x 8 x 256). This basically means that the network learns the *what* information in the image. However, it has lost the *where* information.

The right hand side is the decoder, which contains transposed convolutions along with regular convolutions. In the decoder, the size of the image gradually increases and the depth gradually decreases (starting from 8 x 8 x 256 to 128 x 128 x 1) [47]. Intuitively, the decoder recovers the *where* information by gradually applying up-sampling. To get precise locations, at every step, expansion path uses skip connections by concatenating the output of the transposed convolution layers, with the feature maps from the encoder at the same level (see Equation 11.1). After every concatenation, two consecutive regular convolutions are applied so that the model can learn to assemble a more precise output [47].

$$u6 = u6 + c4 \;\mid\; u7 = u7 + c3 \;\mid\; u8 = u8 + c2 \;\mid\; u9 = u9 + c1 \tag{11.1}$$

Figure 11.2 includes some important elements to note: *2@conv* layers means that two consecutive convolution layers are applied; *c1, c2,....c9* are the output tensors of convolutional layers; *p1, p2, p3* and *p4* are the output tensors of max pooling layers and *u6, u7, u8* and *u9* are the output tensors of up-sampling (transposed convolutional) layers.

11.1. From 2D to 3D

According to the dimension of the convolutional kernel used, CNNs can be categorized into 2D and 3D CNNs.

2D CNNs use 2D convolutional kernels to predict the full segmentation volume map taking one slice at a time [48].

As they are only able to leverage context across the height and width of the slice, they fail to comprehend context from adjacent slices, so that accuracy is reduced. See Figure 11.3.

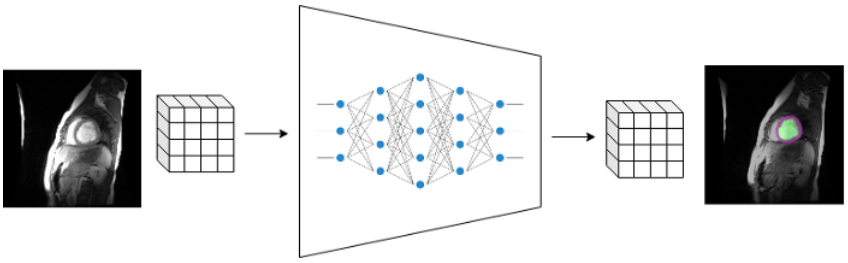


Figure 11.3: 2D CNNs predict segmentation maps for MRI slices in a single anatomical plane. Extracted from [48]. Accessed on 13-02-2021.

On the contrary, 3D CNNs are able to address this issue as they use 3D convolutional kernels to make segmentation predictions for a volumetric patch of a scan, taking advance of inter-slice context leverage [48]. However, such an improvement comes with a high computational cost and increased number of parameters (see Figure 11.4). Related with 3D CNNs, in 2016, Ciçec et al. [49] proposed a 3D U-Net that deals with 3D medical data. This networks also required high computational cost and GPU memory usage, due to the high number of parameters but was successfully trained to segment prostate from 3D MRI images.

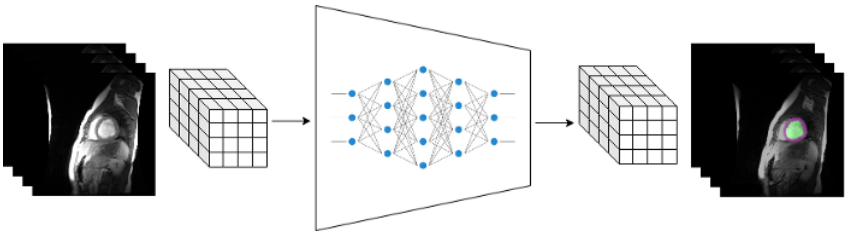


Figure 11.4: 3D CNNs use 3D convolutional kernels to predict segmentation maps for volumetric patches of an MRI volume. Extracted from [48]. Accessed on 13-02-2021.

Chapter 12

Challenges and Limitations

Deep-learning based image segmentation is by now a well-established and robust technique in image segmentation. This chapter summarizes the most common limitations and challenges compared to computer vision techniques. These limitations have been divided into dataset and training-related.

12.1. Dataset-related limitations

– Medical image databases are difficult to directly be exploited as there are many ethical, privacy, security and legal issues. The transmission, storage and usage of medical data is subject to specific regulations and it is frequently necessary to have patient's consent and data anonymization [10].

– Medical image datasets are relatively small compared to general-purpose images. In order to increase the size of the training dataset, the most common method is data augmentation, using simple operations such as translations or complex operations such as elastic deformations.

– There is a lack of expert annotations for the images. Networks, in order to gain the capability to handle complex conditions, usually require a large number of annotated samples. Performing annotated cases is needed to be done

by experts and it is mostly always tedious, expensive and time-consuming. Annotation process is human-dependent and usually causes label-noise, which compromises on the quality of image segmentation result. The *sparse annotation* techniques or weakly supervised learning [37] approaches are capable of dealing with such an issue.

– In some applications, instead of using the whole 3D volume, medical images are converted to stacks of independent 2D images or image patches [10] extracted from them, in order to increase the amount of data. Sample re-weighting is considered a proper solution, where a higher weight is applied to the foreground [37] objects. Besides, loss function redesigning can also be a great solution.

12.2. Training-related limitations

– *Overfitting* occurs when a model can capture the patterns and regularities in the training set with reasonably higher accuracy compared with unprocessed instances of the problem [50], that is usually caused by the small size of the training data. Data augmentation can be a solution to increase the dataset size and avoid overfitting. Besides, *dropout* application can be another solution to combat overfitting. It is used to discard the output of a random set of neurons in each iteration from the fully connected layers [37].

– Training time is another issue in deep-learning based image segmentation techniques. The application of pooling layers and batch normalization can help in having faster convergence. *Batch normalization* is a layer that is applied after the convolutional layers and before the activation layers, through which the output of the convolutional layers is normalized with additional scaling and shifting, following by passing the output to the activation layer and subsequently to the next convolutional layers [51]. The introduction of this layer allows network to give better accuracy of test data and faster training, as it reduces internal covariance shift [51]. Besides, the use of 3D volumes also increases the training time, comparing with networks that handle 2D images.

Part IV

Materials and Methods

Chapter 13

Employed datasets

The following chapter aims at describing the dataset used in the implementation of the CNN for multi-class 3D AAA. Post-operative CTA scans used in this project have been provided by *Biodonostia Health Research Institute* and consist of 43 post-operative contrast enhanced CTA studies of 17 patients.

13.1. Information Model for DICOM Images

All images are *Digital Imaging and Communications in Medicine (DICOM)* format, that is the standard for the management and communication of medical imaging information and related data, commonly used for storing and transmitting medical images enabling the integration of medical imaging devices and *picture archiving and communication systems (PACS)* from different manufacturers [52].

The *DICOM Information Model* defines the structure and organization of the information related to the communication of medical images [52]. The DICOM Information Model is divided into four levels, as it it shown in the Figure 13.1.

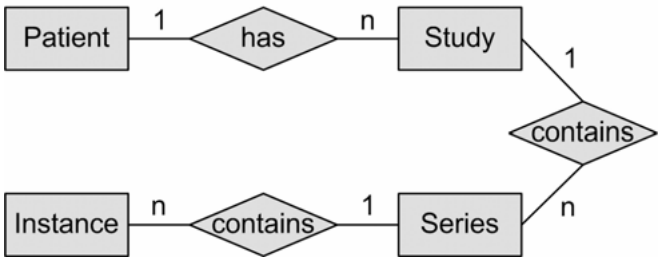


Figure 13.1: DICOM Standard workflow. Extracted from [52]. Accessed on 02-02-2021.

1. *Patient Level.* It contains identification and demographic information of the patient to which the study belongs. It is more common to use the study level for collecting information that comes from various information systems.

2. *Study Level.* It is the most important level within the information model and it is considered as a results of a certain medical examination request. This request can involve examination procedures for different modalities, that will exist on series of one or more images. Each patient can have different studies.

3. *Serie Level.* Image series are gathered at a lower level, which describes the modality of images, the date of image creation, details about the type of examination and used equipment. Commonly, the criteria for image grouping is the temporal and spatial relation between them.

4. *Image Level.* It is the lower level in the Information Model. Each image contains acquisition and position information and personal data. Depending on the modality of the image, this level contains data for just one image, two images (double plane system) or multi-frame images. In DICOM, images and instances are the same.

13.2. CTA Datasets

Regarding the materials needed to the implementation of the algorithm, each CTA scan spans similar body area, with different axial, sagittal and coronal slice number. The available data consist of 17 datasets from 17 different patients.

Each patient has a different number of studies (2 as minimum, 4 as maximum), corresponding with the number of post-operative image acquisition.

With respect to the naming and storage of the image, all CTA scans are pseudo-anonymized DICOM-format images, specifically saved with the number of the patient, the number of the post-operative image and the acquisition date (month and day). For example, the study named *46post_jul_18* means that the study corresponds to the first post-operative evaluation of the patient number 46 acquired on 18th July. Table 13.1 shows a summary of the available CTA dataset.

Table 13.1: Summary of CTA dataset.

Number of patient	post	post1	post2	post3
46	x	x		
47	x	x	x	
48	x	x		
49	x	x	x	x
50	x	x	x	
51	x	x	x	
52	x	x	x	
53		x	x	
54	x	x	x	
55		x	x	
56	x	x	x	
57	x	x	x	
58	x	x		
59	x	x		
60	x	x		
61	x	x		
62	x	x		

Chapter 14

Software and applications for Image Processing and Neural Network Implementation

This chapter aims at providing a general description of the tools and softwares used for image loading, reading and processing and the implementation of the neural networks for the purpose of this project.

14.1. Software and tools for Image Management and Processing

In order to properly manage and process postoperative CTA scans, some software and tools have been used. This section contains a description of the tools for image loading, management and processing: ITK-SNAP interactive software and SimpleITK image reading and analysis library.

14.1.1. ITK-SNAP

ITK-SNAP was first developed in the early 2000 and it is an interactive tool for image visualization, manual segmentation, and semi-automatic segmentation [53]. It is well adapted to biomedical image management and provides an environment for visualizing 3D and 4D imaging sets and a set of tools for an efficient labelling of structures of interest [53].

ITK-SNAP is an easy-to-learn and suitable tool written in C++, with pre-compiled binaries, available for Windows, Linux and MacOS [53]. It is able to read and write a variety of commonly used 3D image formats, and has built-in support for reading data in DICOM format [53]. Besides, ITK-SNAP supports working with multiple image data sets simultaneously, even if they are of different resolution, orientation or imaging modality [53].

The application contains different modules as: visualization, manual and semi-automatic segmentation tool with supervised learning and level sets and image registration, among others. In this project, the tool has primarily been used for image and label visualization and for editing of segmentation volumes.

Visualization Module

ITK-SNAP allows the user to load image volumes using common 3D medical image formats, including DICOM, NIFTI, MetaImage and NRRD [54], and it recognizes the encoded information in the image header on the spacial position and orientation of image volumes relative to the scanner physical coordinate system [54]. The first image loaded into ITK-SNAP is the *main image* and additional images or labels can be loaded above the main image (see Figure 14.5).

3D volumes are visualized as three orthogonal slices, that correspond to the axial, coronal and sagittal planes in physical space. Moving the 3D cursor in one slice view adjusts the slices visualized in other views, that provides a proper way to navigate through the volumes. In the Figure 14.1, there is a CTA image

in mhd format loaded into the main window of the application, with marked boxes representing interesting tools and label's visualization window. Figure 14.2 represents the different options that the *Main Toolbar* includes.

It is possible to analyse the intensity histogram of the images and its meta-data, which contains the information regarding its origin, direction, size and spacing, which are of special interest when managing CTA images (see Figure 14.3). Besides, ITK-SNAP also has a great tool that enables the automatic contrast-adjustment of the image. Figure 14.4 depicts how the contrast-adjustment works based on the image of the Figure 14.1.

Regarding the visualization of segmentation volumes, they are represented as 3D images with the same dimension and orientation as the main image. Each voxel of the segmentation volume is assigned a discrete integer label [54]. It simplifies three-dimensional editing of the segmentations, as a change made in one slice is translated into changes in the other slice views [54]. There is also the possibility to 3D volume visualizing of the volume of segmentation, apart from slice-viewing.

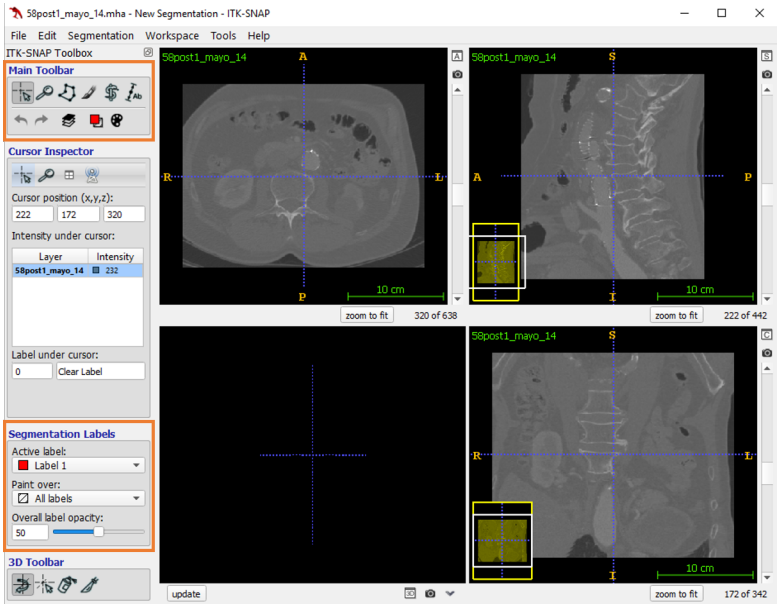


Figure 14.1: Main window of ITK-SNAP with .mhd format image loaded .

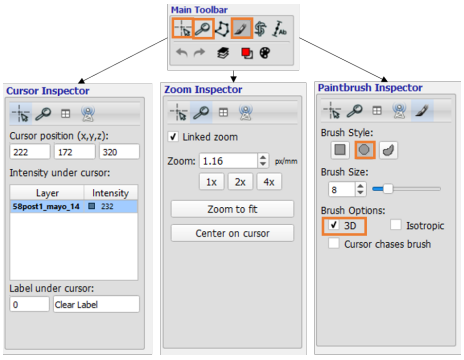


Figure 14.2: Main Toolbar options.

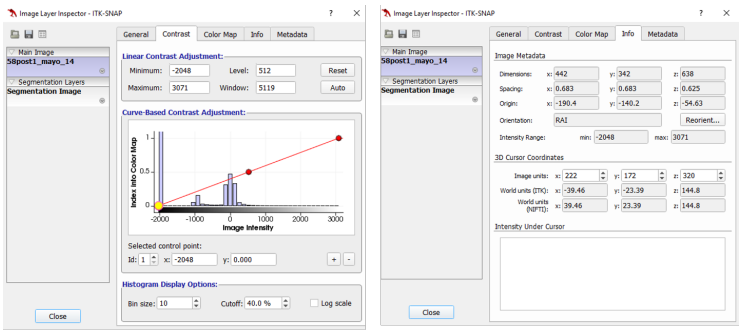


Figure 14.3: A) Contrast analysis window. B) Image Metadata information.

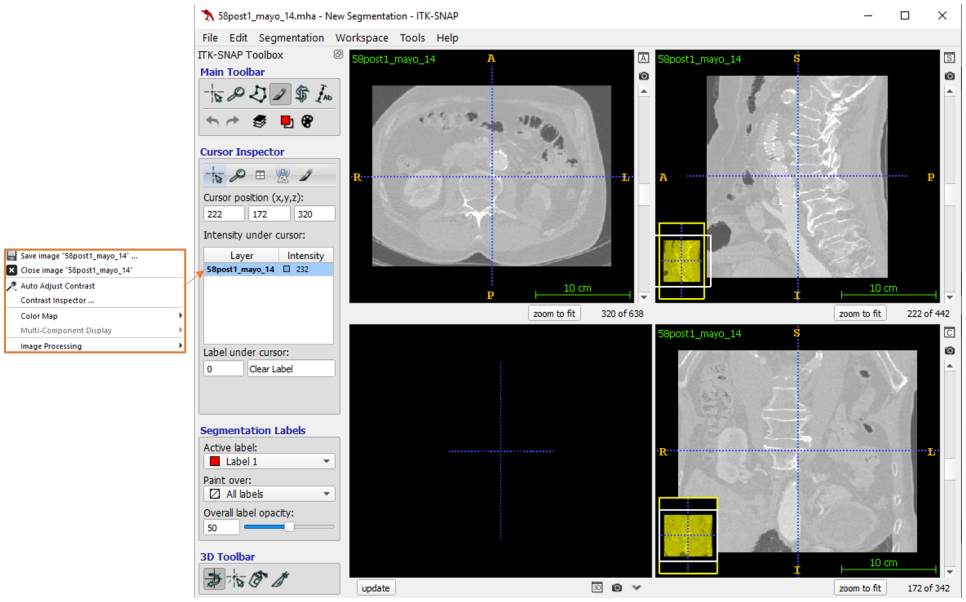


Figure 14.4: A) Contrast analysis window. B) Image Metadata information.

Edition Module

As automated segmentation in medical images may still include errors, ITK-SNAP has a variety of options to manually edit the segmentations or even generate them from the scratch. These tools include 2D outlining and paintbrush tools and 3D mesh editing tools, such as the 3D scalpel [53]. ITK-SNAP also includes an interpolation module that allows the user to trace a structure in just a handful of slices, with the algorithm filling in the intermediate slices [53] (3D brush option). The tool also allows the user to make annotations on top of segmentations and also to generate reports on the volume and intensity characteristics of segmented structures.

There are some tools that enable the manual segmentation and editing of semi-automatic segmentations. In this project, the *paintbrush* has been used (see Figure 14.2, which allows quick drawing and touch-up editing using the mouse, with masks of different shape and size. When manual segmentation is done, user should select the active label used [54] to perform the painting operation and where to paint them (over all existing labels, only the clear label or over an specific label) (see marked box at Figure 14.5. ITK-SNAP also includes a semi-automatic segmentation module, but this option has not been used for the present project.

Figure 14.5: Main image with superposed segmentation volume.

14.1.2. SimpleITK

SimpleITK is a simplified programming interface to the algorithms and data structures of the *Insight Segmentation and Registration Toolkit* [55]. It can be implemented in multiple programming languages including C++, Python, R, Java, and others. This library contains complex medical image analysis tools. In this project SimpleITK library has been used within *Pycharm Proffesional integrated development environment (IDE)*. Section 14.2 makes a description of basic tools and libraries used for Python code generation and training of the neural network.

Images

SimpleITK has a unique feature; it views images as physical objects occupying a bounded region in physical space [56]. This is a significant difference versus other image analysis libraries that treat with images and it comprises: 1) pixel/voxel spacing is assumed to be isotropic and 2) there is no notion of an image's location in physical space [57]. Images can have different spacing between pixels along each axis, and the axes are not necessarily orthogonal. Figure 14.6 illustrates these concepts.

The region in physical space which an image occupies is defined by the image's:

- The *size* of the image (vector like type) is the number of pixels/voxels in each dimension that define its dimension.
- The *origin* (vector like type) that consists of the coordinates of the pixel/voxel with index (0,0,0) in physical units as millimetres.
- The *spacing* (vector like type) that is the distance between adjacent pixels/voxels in each dimension given in physical units.
- The *direction* matrix (vector like type representing matrix in row major order) that can be mapping, rotation between direction of the pixel/voxel axes and physical directions.

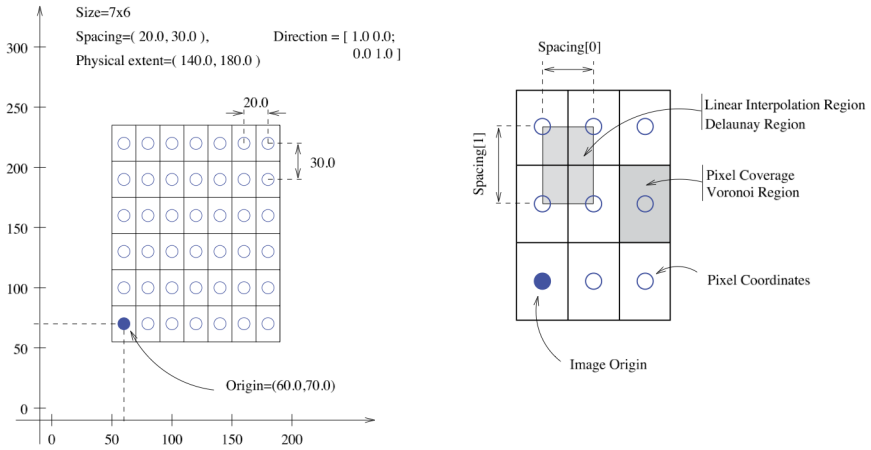


Figure 14.6: Spacing, origin and size concepts of SimpleITK. Extracted from [56]. Accessed on 16-02-2021

Besides, there are different pixel types in SimpleITK, that is represented as an enumerated type. Some examples are: *sitkUint8* that is an unsigned 8 bit integer or *sitkFloat32* that is a 32 bit float. The images used in this project have been set as *sitkFloat32* and masks as *sitkUint8*.

In order to load images, *sitk.ReadImage* function has been used. For its construction *sitk.Image* has been used, where some components described above are necessary for a complete definition of an image (see Figure 14.6). Pixel type description is also needed to create the image. All these features can be changed, but it has to be done cautiously, as the features of the images change completely.

For accessing pixels *GetPixel* and *SetPixel* has been used; which takes 2 parameters in 2D images (x and y index) and 3 parameters in 3D images (x, y and z indexes).

There is a fact that has to be considered when using SimpleITK and *Numpy* (fundamental package for scientific computing in Python, that provides a multidimensional array object, various derived objects and an assortment of routines for fast operations on arrays) both together in the same code. This is because the libraries' indexing access is in the opposite order; whereas SimpleITK has [x, y, z] convention, Numpy uses [z, y, x]. Firstly, there are two options for

converting from SimpleITK to Numpy: *GetArrayFromImage* that returns a copy of the image and then can be freely modified as it has no effect on the original image and *GetArrayViewFromImage* that returns a view on the image data which is useful for display in a memory efficient manner. Data cannot be modified and the view will be invalid if the original image is deleted. To complete the opposite step *GetImageFromArray* is used.

For getting some images' features, SimpleITK supports basic arithmetic operations between images, taking into account their physical space [56] as: *sitk.GetSize*, *sitk.GetSpacing* functions and for setting new parameters functions as *sitk.SetSize* or *sitk.SetSpacing* can be used, among others.

Channels

In biomedical domain, some image types have a single scalar channel (e.g. CT, US) or a three channel image where each channel has scalar values in range [0, 255] or RGB images. It should be considered that gray scale images are not always stores as a single channel image.

Mask Image Types and Default Values

The default mask image type is a scalar image of *sitkUint8* and its defaults values are 0 and 1, with 1s represents the mask. These defaults values are for thresholding, segmentation filters and for the binary morphology filters, so that they can be applied after segmentation. This has been an important consideration to take into account, as many post-segmentation filters have been used in order to create the ground-truth mask (see Section 15). In addition to default values, the mask can contain different labels (multi-class mask), corresponding to different classes.

14.2. Software used for Neural Network Implementation

The implemented CNN for this project, has been developed in Python language. Python is an interpreted, object-oriented, high-level programming language with dynamic semantics [58]. It supports modules and packages for neural networks' development. For this project Python version 3.7.9 has been used.

14.2.1. Pycharm Professional

Pycharm Professional is an integrated development environment (IDE), used in computer programming, specifically for the Python language and developed by the company JetBrains. This IDE has been used for pre-processing and mask generation and the design of the neural network. For these purposes, some essential libraries and packages have been utilized.

14.2.2. Tensorflow and Keras

Tensorflow 2 is an open source library for numerical computation and large-scale machine learning [59]. It uses Python to provide a convenient front-end API for building applications with the framework [59] and can run high number of neural networks and classification models. Tensorflow allows to create *dataflow graphs*, where each node in the graph represents a mathematical operation and each connected or edge between nodes is a multidimensional data array, or *tensor* [59]. Nodes and tensors in Tensorflow are Python objects. Although using TensorfFlow directly can be challenging, the modern tf.keras API beings the simplicity and ease of use of Keras to the Tensorflow project, that allows to design, fit, evaluate and use deep learning models to make predictions in few lines [60]. It makes common deep-learning tasks accessible to average developers [60].

Together with TensorFlow, it is automatically installed an interesting visualization tool called TensorBoard. It reads summary files from TensorFlow and helps the user visualizing, understanding and debugging its graphs [61]. Moreover, it shows the plots of quantitative metrics of the models and additional data that gives an insight into the training process of the model [61].

14.3. Other applications

Another application that has been used in the project is *GitLab*. It is an open source code repository and collaborative software development platform. It has been used for neural network codes' checking and uploading [62].

Chapter 15

Ground truth mask generation

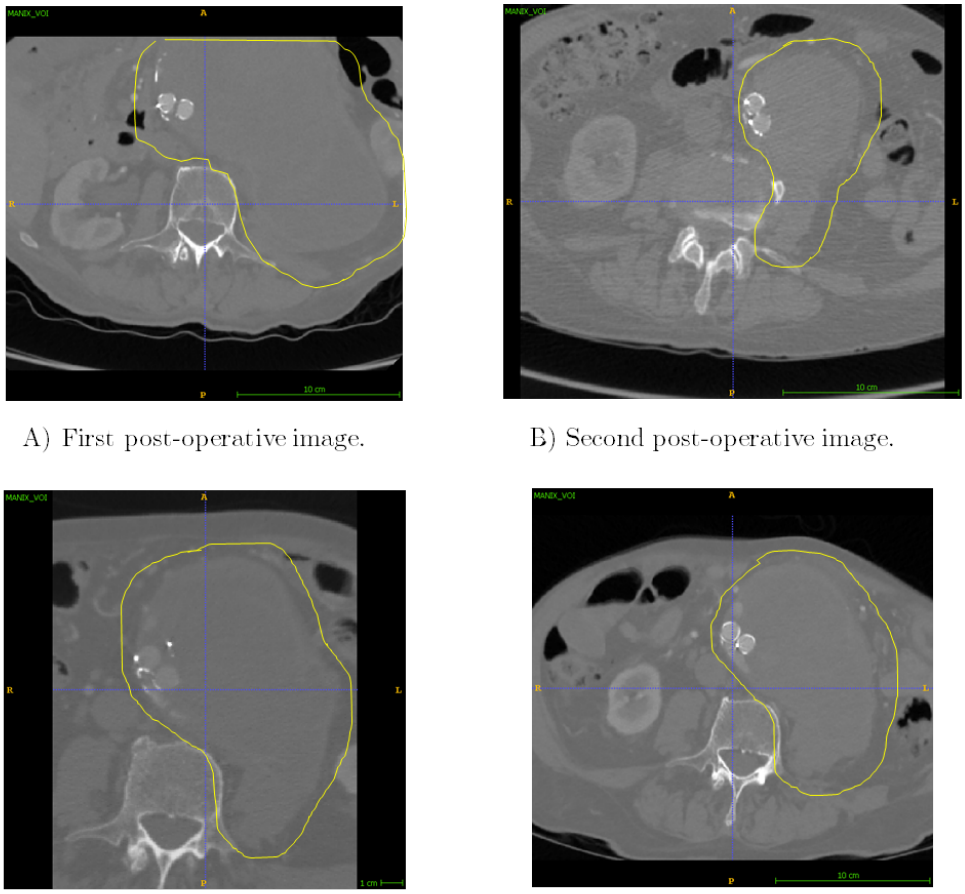
This chapter includes the initial analysis that has been carried out in order to select valid images, the procedure followed to manage preliminary segmentation volumes and its edition. Finally, an explanation of how the ground-truth masks have been generated is included.

15.1. Previous considerations about the dataset

Prior to start with the annotation process, all CTA images have been analysed in order to discard any problematic case and to have a complete and valid dataset.

From the 43 studies that were available, finally 37 have been considered. Two studies from different patients have been deleted due to the poor contrast of the image, that does not enable the correct segmentation and recognition of the structures of both lumen and aneurysm.

There has been an special case (patient 49) in which the aneurysm appears broken, so that its delineation results impossible. In total, 4 studies of that patient were deleted (see Figure 15.1).



A) First post-operative image.

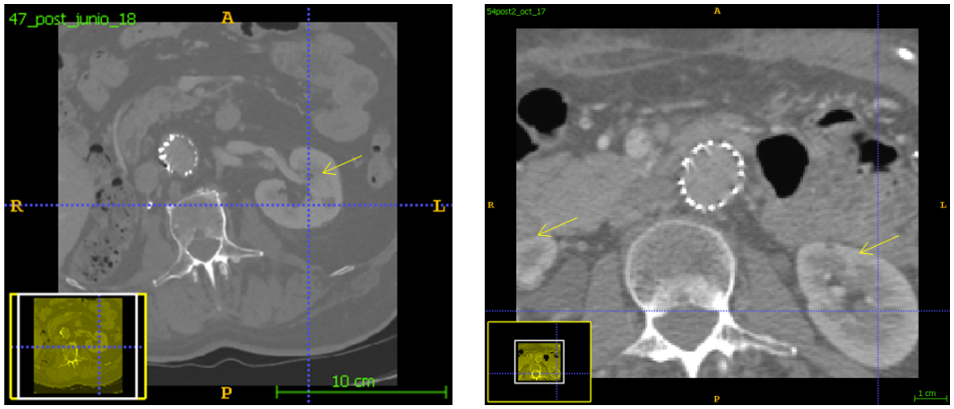
B) Second post-operative image.

C) Third post-operative image.

D) Fourth post-operative image.

Figure 15.1: Post-operative images of broken-aneurysm patient.

Within the valid dataset, there are cases in which the visualization of renal arteries is limited. In the first case (see Figure 15.2A), there is a kidney missing and in the second case, even if two kidneys are present, any of renal arteries are visible (see Figure 15.2B). The images with special characteristics, as the mentioned above, have been set as special images for testing the model trained, as will be explained in Chapter 18.



A) Patient 47 – post 1 B) Patient 54 – post 2

Figure 15.2: Limitations in the visualization of renal arteries of post-operative images from two different patients.

DICOM images have been converted into *MetaImage*, that allows raw data to be contained in a separated file (*mhd* and *raw*) or embedded in the same file (*mha*). This conversion is made through eVida Vascular application and it allows to read and load the dataset in ITK-SNAP software for image processing and in Pycharm Professional for neural network training (see 14.1). Initially, CTA images have been treated in mhd format, then, they have been embedded into mha files in order to avoid to work with two files (raw and mhd). In the case of segmented volumes, they all are in mha format.

The first step in order to train the CNN with CTA post-operative images, has been the generation of ground-truth mask database for both lumen and aneurysm. The proper annotations for mask generation is essential in order to have great dataset for neural network training. The methodology that has been carried out, has been similar for the structure of the lumen and the aneurysm: first, the process has been initiated with the help of automated internal tool that provided preliminary segmentation volumes for lumen and aneurysm and then, some edition work has been done in order to properly generate the masks.

15.2. Preliminary Lumen Segmentation

The first step in the process of mask generation has been the segmentation of the lumen. For that purpose, an internal 3D CTA image-based software for AAA inspection and EVAR sizing tool has been used, specifically designed for preoperative stage image analysis. This tool has been described in the Subsection 7.1.1.

First, eVida Vascular tool has been used to generate a cropped version of the images, as original abdominal CTA images span a large body area. This way, images will contain the region of interest (ROI); from descending aorta to final region of iliac arteries, obviously involving aneurysm region. In order to obtain a proper ROI, the x and y dimensions of the images have been maintained as in the original ones, while almost all of them have been reduced in z dimension. Once the ROI has been achieved, an initial lumen segmentation has been done using the segmentation module of the application, where all parameters have been used as they were by default. All segmented volumes were generated as mha format images.

Although this tool contains a module for centerline extraction of the lumen, this option appeared not to work for post-operative images, so that it has been discarded.

15.2.1. Lumen Segmentation results taken from the semi-automated tool

Once the semi-automatic tool has generated segmented volumes of the lumen, each of them have been analysed, in order to look for incoherences and points to improve, as a main objective of the implementation of CNN for lumen and aneurysm segmentation. As explained in the Section 7.1, there are certain limitation when directly applying pre-operative image based algorithms to post-operative images. There are two noteworthy limitations. The first one is that the iliac arteries that are part of the lumen, appear as a unique lumen, or even one or none of them are segmented. Table 15.1 collects the results of the semi-automated tool's performance

on the iliac arteries of the aorta denoting if: 0) none iliac arteries have been segmented , 1) just of the iliac arteries has been segmented (right or left), 2) both iliac arteries have been segmented but appear stuck one with the other and 3) both iliac arteries have been segmented correctly (separately).

Table 15.1: Results of the semi-automated tool's performance in the segmentation of the iliac arteries.

	Segmentation result (%)
0	35.1 % (13)
1	27.0 % (10)
2	27.0 % (10)
3	10.8 % (4)
TOTAL	37

The second limitation is that there is a difficulty in properly segment aorta's surrounding vessels. In order to properly carry out the follow-up of the EVAR and disease progression, it is of importance to properly segment these branches: the Celiac Trunk, the Superior and inferior (not as useful) Mesenteric artery, renal arteries and both iliac arteries (see Figure 15.3).

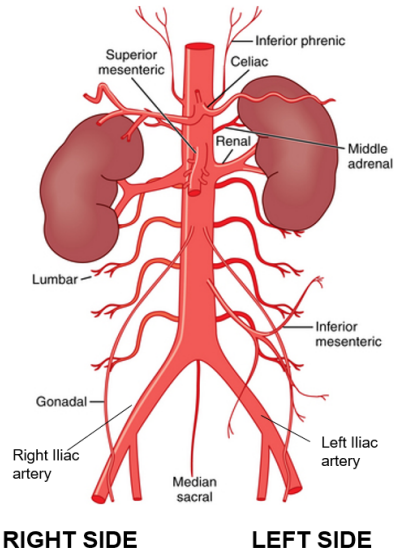


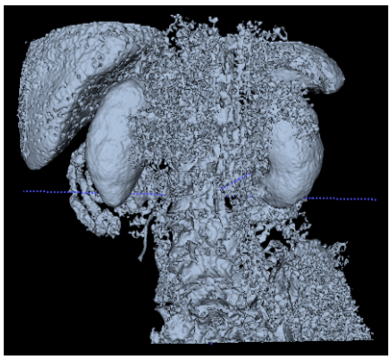
Figure 15.3: Representation of aorta's main branches. Adapted from Valki, K (2016), Chapter 7 of Radiology Keys. Accessed on 18-02-2021

In the Table 15.2, it is also indicated if aorta's surrounding vessels have been properly segmented, for each patient's study.

Table 15.2: Results of pre-operative semi-automated tool performance in aorta's vessels segmentation. CT - Celiac Trunk, SM - Superior Mesenteric artery, IM - Inferior Mesenteric artery, R - renal arteries and I - iliac arteries.

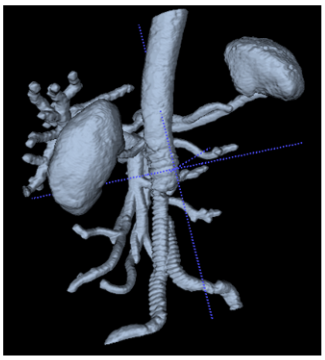
	Left	Right	Whole structure	Not segmented/valid
CT	-	-	59.5 % (22)	40.5 % (15)
SM	-	-	75.8 % (28)	24.3 % (9)
IM	-	-	0 % (0)	100 % (37)
R	5.4 % (2)	0 % (0)	37.8 % (14)	56.8 % (21)
I	13.5 % (5)	10.8 % (4)	37.8 % (14)	37.8 % (14)

The Figure 15.5 below represents some examples of problems regarding iliac arteries' segmentation using the semi-automate tool and Figure 15.6 represents some examples of limitations regarding aorta's branches' segmentation. In addition to these limitations, some segmentations also present large surrounding anatomical structures segmented, such as kidneys and other vessels, that hinders the identification and proper manual edition of the volume (see Figure 15.4). Besides, other segmentation volumes required to generate them from the scratch, as the tool has not been able to segment anything.



A)

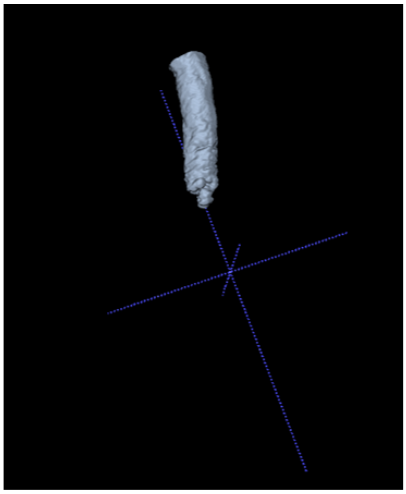
Segmentation presenting huge segmented area, that even includes part of the heart, liver and kidneys.



B)

Segmentation presenting kidneys segmented.

Figure 15.4: Noisy lumen segmentations.

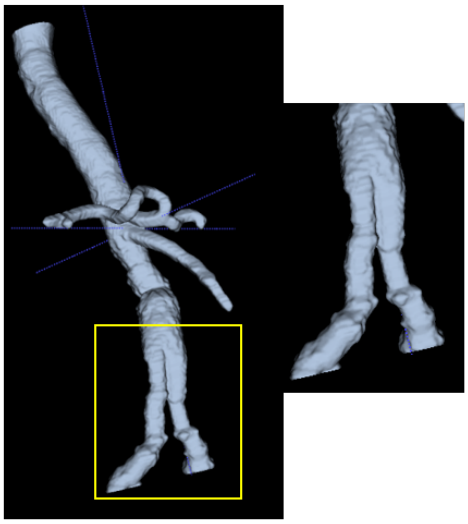


A)

Segmentation presenting only the trunk of the aorta, without any branches.

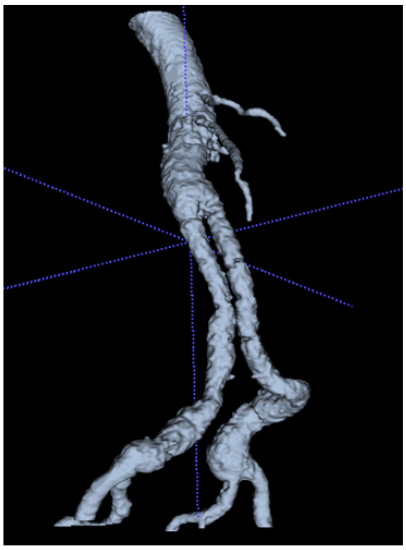
B)

Segmentation presenting just one iliac artery segmented (left iliac).



C)

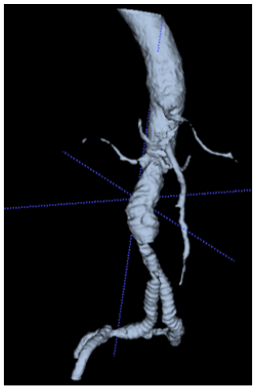
Segmentation presenting both iliac arteries segmented, but stuck.



D)

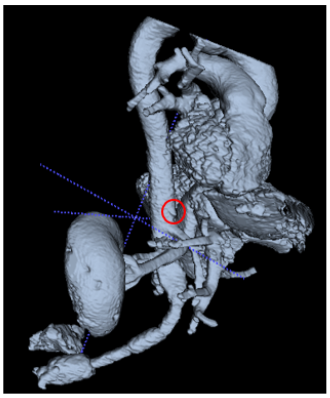
Segmentation presenting both iliac arteries separately segmented.

Figure 15.5: Lumen segmentation analysis focused on iliac arteries result.



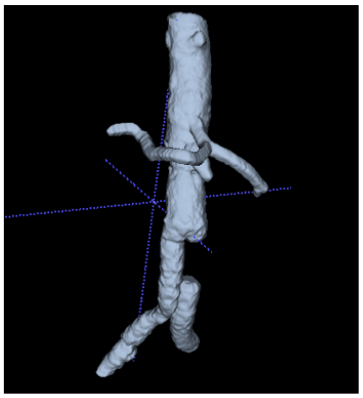
A)

Segmentation of the lumen with all the branches segmented (even if eroded).



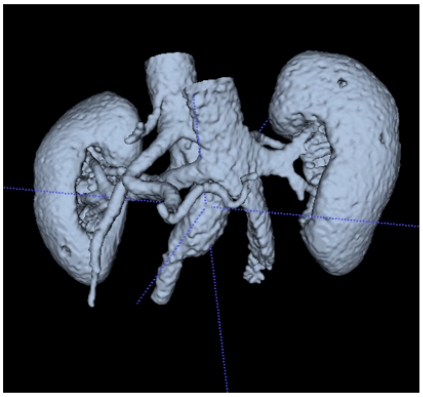
B)

Segmentation of the lumen with celiac trunk artery shortened (red circle) an left iliac missing.



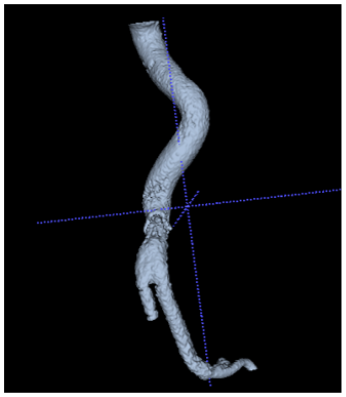
C)

Segmentation of the lumen with renal arteries missing and left iliac artery not fully segmented.



D)

Segmentation of the lumen with iliac arteries missing.



E)

Segmentation of the lumen with just right iliac artery segmented.

Figure 15.6: Lumen segmentation analysis focused on aorta's branches.

According to the performance of the semi-automated tool on iliac arteries, it is shown that in just the 10 % of the cases both iliac arteries are well segmented, whereas in the 27 % of the cases both iliac are segmented, but appear stuck. In the majority of cases (around 35 %) any iliac artery is segmented and in the 27 % of the cases just one iliac has been segmented. This performance is due to the limitations regarding the segmentation of the lumen from post-operative CTA images, when directly applying pre-operative based semi-automated lumen segmentation tool. These limitations arise because the stent is positioned at the level of iliac arteries and it is complicated to distinguish between those two right and left bifurcations as they usually not appear differentiated in the image. Besides, some metal-related artefacts appear due to the presence of the stent, which difficulties the segmentation of iliac arteries. Thus, almost all segmented lumen required high post-processing and editing tasks.

Regarding vessel segmentation results, it is seen that in more than half of the cases, the Celiac Trunk and Superior Mesenteric artery are well segmented. However, in more than 60 % of the cases, renal and iliac arteries are not segmented, which are considered of high importance in the evaluation of AAA. Besides, any Inferior Mesenteric artery has been segmented. This bifurcation artery is small and difficult to visualize, so that its segmentation is very challenging. In general, it is seen that the accuracy of the segmentation volumes obtained from the semi-automatic tool is low, so that there is high need to optimize its results by implementing new segmentation methods.

15.3. Preliminary AAA Segmentation

In the case of the AAA, preliminary segmentation volumes have been acquired from a pre-designed 3D CNN network for pre and post-operative images. This network seems to work fine as it has been specifically designed for both pre and post-operative images. It consist of an adapted *Hollistically-Nested Edge Detection (HED)* network and uses Weighted Dice Loss function (see Section 16.3.1) and accuracy, as a metric. It uses volumes of size 128 x 128 x 64.

However, due to stent-related artefacts or poor contrasted images, some of the thrombus has not been segmented preliminary, so that they have been delineated manually. Table 15.3 collects the results regarding the segmentation of the AAA by the pre-designed CNN.

Table 15.3: AAA segmentation results of the pre-designed CNN.

	Segmentation result (%)
Not segmented	32.4 % (12)
Incomplete	54.1 % (20)
Segmented	13.5 % (5)
TOTAL	37

A qualitative and quantitative evaluation of the results is gathered in the Chapter 18, where preliminary obtained AAA segmentations have been compared with final results of the implemented CNN network.

15.4. Edition of segmented volumes

Once the initial segmentation of both lumen and AAA is done, it is necessary to make some edition work in order to obtain proper annotated masks, that consisted of 1) the manual edition and annotation and 2) the post-processing of edited volumes.

15.4.1. Manual edition and annotation

In the manual annotation of the lumen the corrections that have been carried out are: the filling of the areas corresponding to missing branches, including iliac arteries and their separation, and filling of missing regions. In the Figure 15.7 below, there is an example of lumen edition process where both renal and right iliac arteries has been edited.

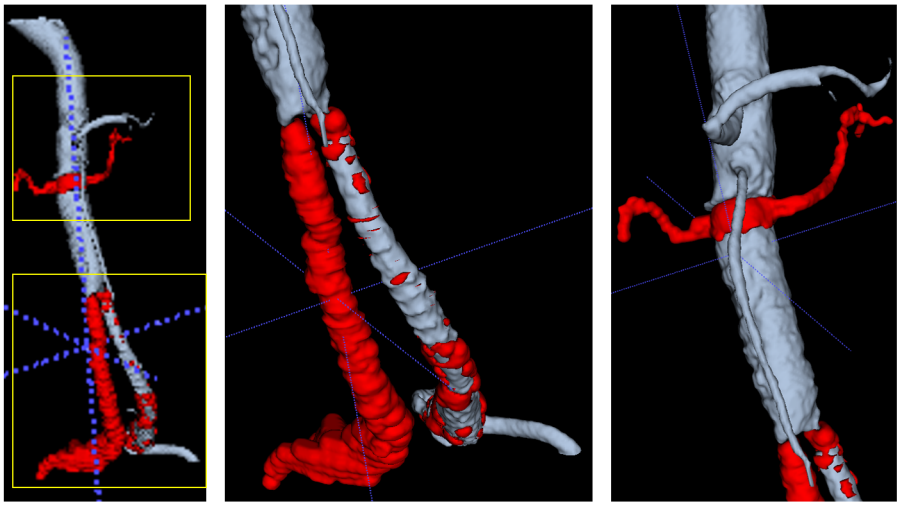


Figure 15.7: Results of manual edition of the lumen.

In the case of the aneurysm, the start and end point of the thrombus has been completed, the external contour of the aneurysm has been delineated and the hole corresponding with the lumen of the aorta has been corrected, emphasizing on the location of the stent and the separation of the iliac arteries. In the Figure 15.8 below, there is an example of AAA edition process, where some regions have been filled. The axial view of the point marked in Figure 15.7 and Figure 15.8 are represented in the Figure 15.9. Both CTA images correspond with the same patient.

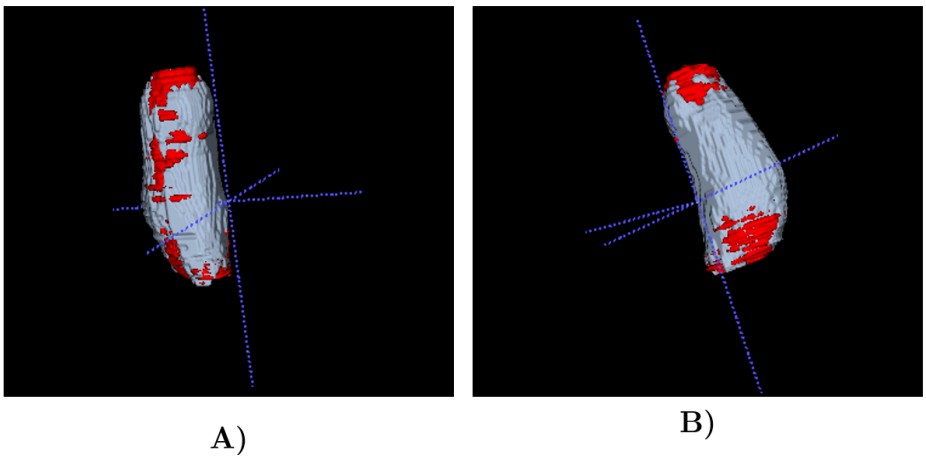


Figure 15.8: Results of manual edition of the AAA.

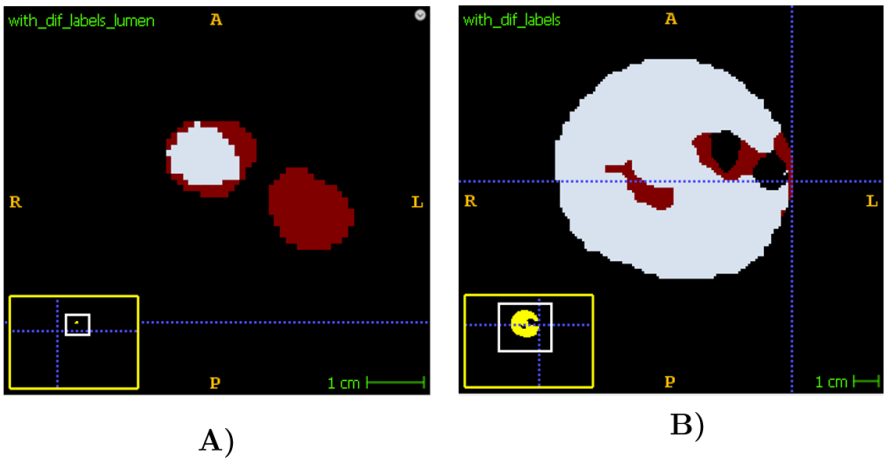


Figure 15.9: Axial slice-view of lumen and AAA.

As it is seen in all examples above, the segmented volumes present different label values. This is just to show how the process of manual editing has been done. In order to generate the masks, both the original and the edited segmented volume have been fused and annotated with same label value 1.

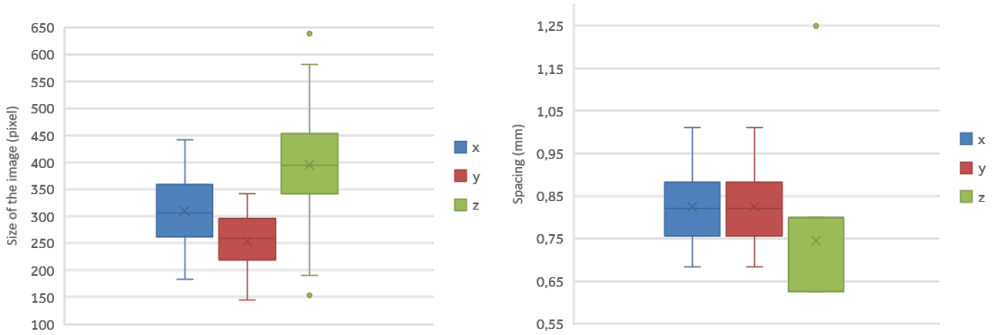
15.4.2. Post-processing of segmented volumes

Once the manual annotations of the volume have been completed, it is necessary to make some post-processing work. This subsection aims to include some important considerations about the dataset, an explanation of the post-processing of the images and some examples regarding the application of the filters employed in the post-processing step. After the post-processing work, the dataset is used to generate the multi-class ground-truth mask.

Volume and Spacing variability

Before proceeding with the application of post-processing filters to the volumes, it is of vital importance to consider how CTA image's characteristics differ one from each other. In other words, it is essential to analyse the size, origin and spacing of the images. As it has been explained in previous sections, abdominal

CTA images span similar body area but have different axial, sagittal and coronal slice number. Once the volumes have been cropped in order to select the ROI, the number of slices across axial, sagittal and coronal axes have changed, specially in z dimension. In addition to variability in size, there is a notable variability in the spacing value of the images. To quantitatively visualize that, the graphs shown in the Figure 15.10 have been generated. As it is shown, the variability in size and spacing in z dimension is larger that in x and y, just because the images have been principally reduced along z dimension.

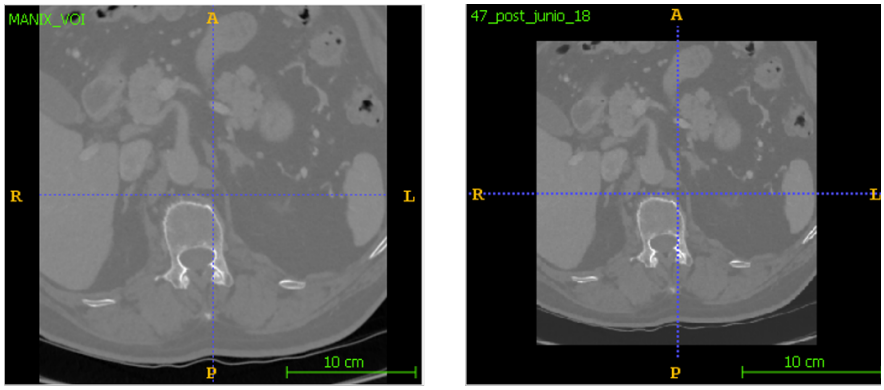


A) Variability in size B) Variability in spacing

Figure 15.10: Variability in size and spacing of CTA images.

In order to face the issue of size and spacing variability among images, the common and easiest strategy has to do with a resize of the images in which the spacing remains the same. However, the information related with the real size of the organs (e.g. its spacing or voxel size in the image) is lost in the training procedure. In order to avoid that information loss, the strategy that has been followed is to modify all images into a common spacing and resample them to a common size, applying a zero-padding. This way, the size of all voxel/pixels will be the same in all the new images, but bigger volumes will be occupy more voxels and vice versa. This approach will make the training slower as it has to learn from such a variability, but there would not be any information loss regarding the real dimension of the volumes. The selected value for the common spacing was [0.683, 0.683, 0.625], that corresponds to the minimal spacing encountered in the analysis of the images and a common size of [442, 342, 638], corresponding to the maximum size value. In order to accomplish this transformation, first, all images have been modified into a common spacing with

the function *SetSpacing* of SimpleITK library. Then, a function has been designed to achieve common-size images by zero-padding method. This function takes the original image, the dimension of the zero-padding to achieve the final size (lower and upper bound) and a scalar value that corresponds with the label of the pixels in the zero-padding region, usually specified as 1024 or 2048. The main command that is within the function is *sitk.ConstantPadImageFilter*. In the Figure 15.11, there is an example of the transformation in spacing and size of a patient's study.



A) Original-size image. B) Zero-padding size image.

Figure 15.11: Spatial transformation of an image. Original size image (A), with dimensions of [293, 313, 638] converted into [442, 342, 638] image (B). The zero-padding in the case A corresponds to the padding that is applied in order to fully visualize the image within the window of ITK-SNAP.

Post-processing of the segmentations by SimpleITK

All the tasks explained below have been applied on both aneurysm and lumen segmentation volumes. The post-processing of these volumes has been done together with the zero-padding application and the mask generation. In order to better explain both steps, they have been described separately. The post-processing consisted of:

- Tiny hole's filling using *sitk.BinaryFillholeImageFilter* and the parameter *SetFullyConnected* set to *False* and *SetForegroundValue* to 1.
- Surface Smoothing with *sitk.BinaryMedianImageFilter* and smoothing radius

of 2, a fine surface smoothing has been obtained.

– Verification of spacing, size and origin of the images and masks. It is essential that the CTA cropped image and the segmented volumes of the same study maintain the same origin, so that they could be overlaid.

15.5. Multi-class mask generation

Once the post-processing of the lumen and thrombus segmentation has been completed, the following step aims to generate a label-map or multi-class mask for both structures, for lately combine them.

First, a new image has been created using the function *sitk.Image* with *sitkUint8* pixel type. In order to access to pixels and define a label to them, it has also been converted into an array.

15.5.1. Uncertainty mask generation

As it logical to think about, the annotation work for any anatomical structures it is complex if the annotator is not a specialist. Thus, the probability of manual segmentation error is high enough to consider that the segmentations can contain some uncertainty. In this project, the uncertainty in annotation process has been used and work with, in order to obtain a more precise prediction in the aneurysm area.

Prior to uncertainty mask generation, the hard-label mask has been created, in which the pixel value of 1 has been set for lumen region and pixel-value of 2 for aneurysm. Then, two uncertainty labels have been generated. One uncertainty label surrounding the lumen and one uncertainty label corresponding with the external contour of the aneurysm.

The steps corresponding to uncertainty mask creation are:

– Surface contouring. This previous step has been done in order to generate external contours of the thrombus for later filling and dilating them. First, the function *sitk.BinaryContour* creates and external contour of the structure and *sitk.BinaryFillhole* fills it in order to obtain an unified mask. Then, in order to join those slice-by-slice structures, *sitk.JoinSeries* has been applied. In the case of the uncertainty mask of the lumen region, this step was not required as it has none holes in it.

– Dilatation of the volume. In the case of aneurysm' uncertainty, to obtain its external contour, once the structure is filled in, a dilatation has been applied using *sitk.BinaryDilateImageFilter* with a dilatation radius of 1. The external contour is the structure subtracted from the dilated and filled image. In the case of the lumen's uncertainty, the volume of the lumen has dilated as in the case of aneurysm's uncertainty and them the external contour of it has been obtained subtracting it from the original volume.

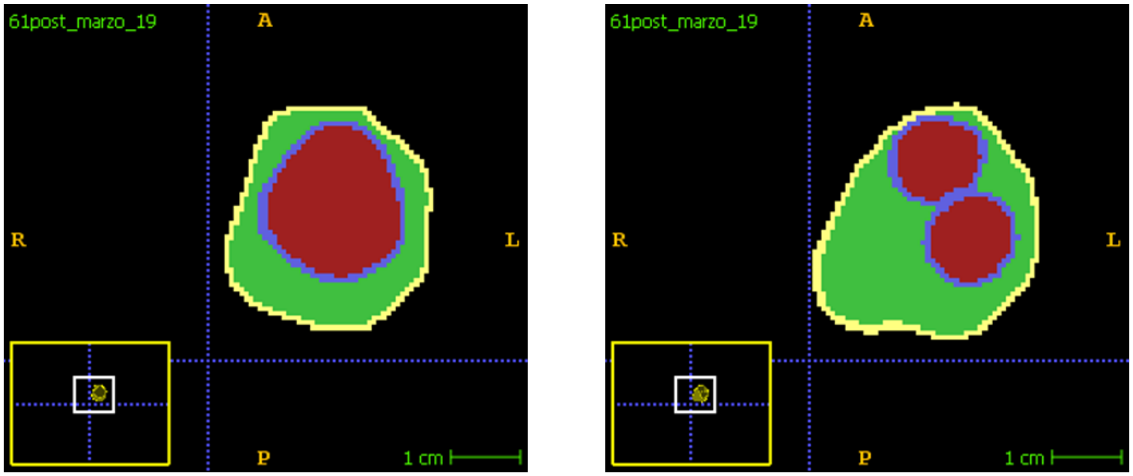
– Label definition and combination. As explained above, two uncertainty-related labels have been applied instead of a unique pixel-value for that region. This is because the uncertainty class pixels around the lumen and the uncertainty class pixels around the aneurysm have different correspondence probabilities, it is logical that the uncertainty mask would be define different on these regions. Thus, an uncertainty value of 3 has been set around the lumen, for both class correspondence of the lumen and aneurysm. Thus, the pixels of the mask that have 3 value can belong to lumen or aneurysm. For external contour uncertainty, the pixel value has been set to 4, so that those pixels can belong to background or aneurysm. At this point, it should be noted that the region of the stent-graft has been somehow ignored in the annotation process.

In summary, the generated mask will contain 5 classes, as represented in the Figure 15.12.

0	1	2	3	4
Background	Lumen	AAA	Lumen/AAA uncertainty	AAA uncertainty

Figure 15.12: Labels of the uncertainty mask.

In the Figure 15.13, there is a representation of two axial slices of the uncertainty mask of the post-operative image post-1 from patient 61.



A) B)

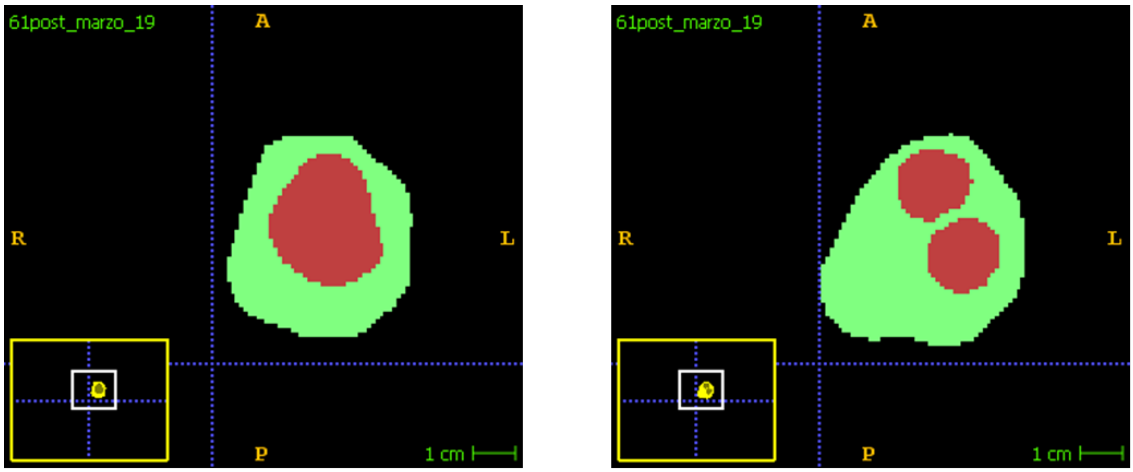
Figure 15.13: Uncertainty mask of post-1 from patient 61. AAA located at aorta's trunk level (A) and at iliac arteries' level (B).

15.5.2. Conversion of the uncertainty mask into hard-label mask

Even if the main implementation approach in this work includes the use of the uncertainty-labeled mask, it has been necessary to previously test the performance of the model using a hard-labeled mask. The goal of this mask's transformation is to ensure that the model is able to learn the structures of lumen and aneurysm from a mask that only contains the hard annotations of their corresponding regions (label values are represented in Figure 15.14). To remodeling the masks, the uncertainty region surrounding the aneurysm (label 4) was converted into aneurysm label (label 2) and the uncertainty around the lumen (label 3) also was converted into aneurysm region (label 2), in order to avoid that the lumen region located at iliac arteries' level get stuck and maintain both iliac arteries separated. The hard-labeled mask is represented in Figure 15.15.

0	1	2
Background	Lumen	AAA

Figure 15.14: Labels of the hard mask.



A) B)

Figure 15.15: Hard mask of post-1 from patient 61. AAA located at aorta's trunk level (A) and at iliac arteries' level (B).

Part V

Convolutional Neural Network Implementation for 3D Multiclass Image Segmentation

Chapter 16

Proposed Method

This chapter defines the principal objectives that motivated the development of the project, the segmentation pipeline and the network architecture.

In this work, a 3D multi-class CNN has been proposed in order to achieve the segmentation of the lumen and the AAA. The goal is to implement an automated tool that; first, will be able to operate with post-operative CTA scans and second, will segment both lumen and aneurysm volumes. Thus, proposed method aims at optimizing segmentation results obtained by pre-operative based CTA images, focusing on iliac arteries and aorta's surrounding vessels, and has the objective of generating 3D multi-class segmentation volumes. This way, predicted volumes will be clinically functional, as they would enable the follow-up of the disease after the EVAR intervention, in order to prevent and avoid the rupture of the AAA.

The following pipeline is applied to segment the AAA from CTA images: 1) image pre-processing which includes window-level adjustment and resizing, 2) segmentation of the aneurysm using CNN based U-Net and 3) image post-processing to binarize the prediction of the network and refine the results by removing small objects. Figure 16.1 depicts the pipeline.

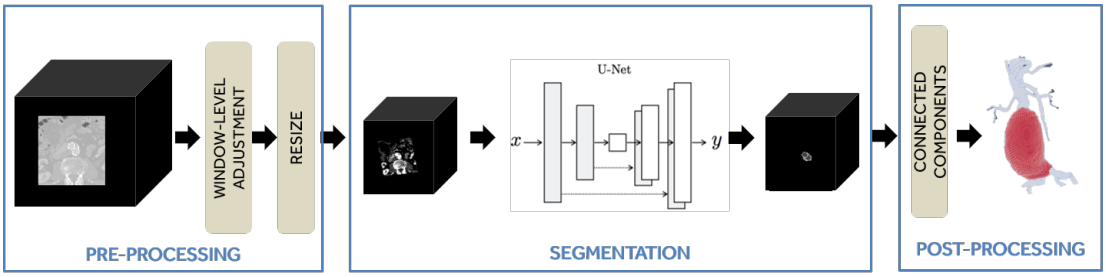


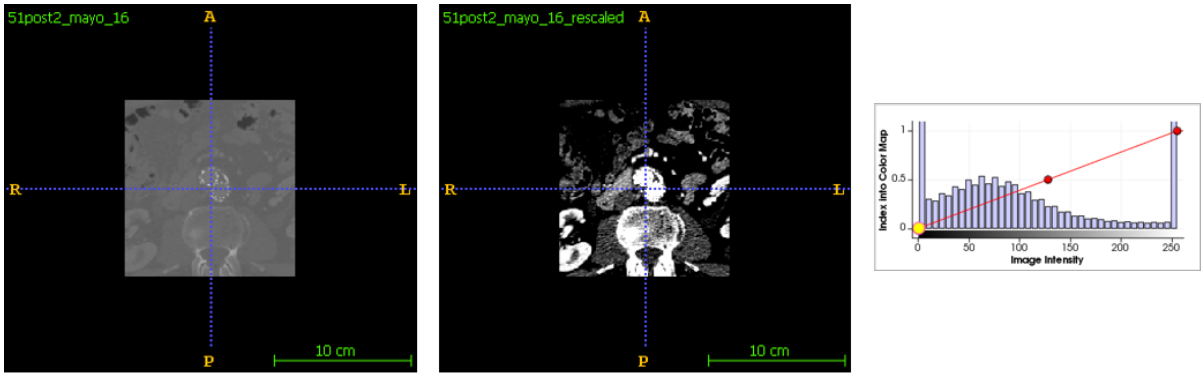
Figure 16.1: Pipeline for AAA 3D Multi-class segmentation using a CNN-based U-Net.

16.1. Dataset Preparation

As mentioned before, the imaging data for this project have been provided by Biodonostia Health Research Institute. From the 43 contrast-enhanced CTA scans initially available, 37 of them have been selected as valid data. This section includes the pre-processing and data-augmentation methods performed to achieve the final dataset.

Data pre-processing

The pre-processing of the input data consists of two steps. First, image contrast has been enhanced using a window-level adjustment and rescaling the intensity between 0 and 255. It has been done using *IntensityWindowing* filter that pixel-wise linear transformation to the intensity values of input image pixels [63]. The linear transformation is defined by the user in terms of the minimum and maximum values that the output image should have and the lower and upper limits of the intensity window of the input image [63]. The chosen values for minimum and maximum values for the output image have been 0 and 255 (default values), respectively. Then, image intensity has been rescaled between 0 and 255, using *RescaleIntensity* filter. This pre-processing filtering has only been applied to images of the training, validation and testing dataset. In the case of the masks, just a resize has been applied. In the Figure 16.2, there is an example of the resultant image after pre-processing.



A) **B)** **C)**

Figure 16.2: Image preprocessing results, showing the original image (A) and the result of window-level adjustment and rescaling (B). Graph C shows the intensity-histogram of the rescaled image.

Finally, 3D volumes and ground-truth masks have been reduced from 342 x 442 x 638 size to 128 x 128 x 256 or 128 x 128 x 128 resolution, to deal with the massive memory requirements of the CNN. It has been intended to preserve the information in z dimension and volume dimension's coherence as much as possible.

It is important to consider that, prior to the pre-processing stage, CTA images have been cropped in order to obtain the volume of interest and then, they have been transformed into common spacing and size, as described in Section 15.2 and Subsection 15.4.2.

Data augmentation

Data Augmentation is the process of generating additional viable training data applying some transformation to the initial training dataset [61]. In this work data augmentation was required due to the limited number of labelled scans, i.e. 27 training and validation scans. Usually, the use of small datasets arises the problem of over-fitting and the accuracy of validation dataset is lower.

First, a set of 8 rotations using fine angles (see Figure 16.3) an a set of 10

translations (see Figure 16.4) has been applied into images. Translation consist of moving the image along the x, y and z direction (or all). As images contain a zero-padding, rotation and translation do not change the final size of the image.

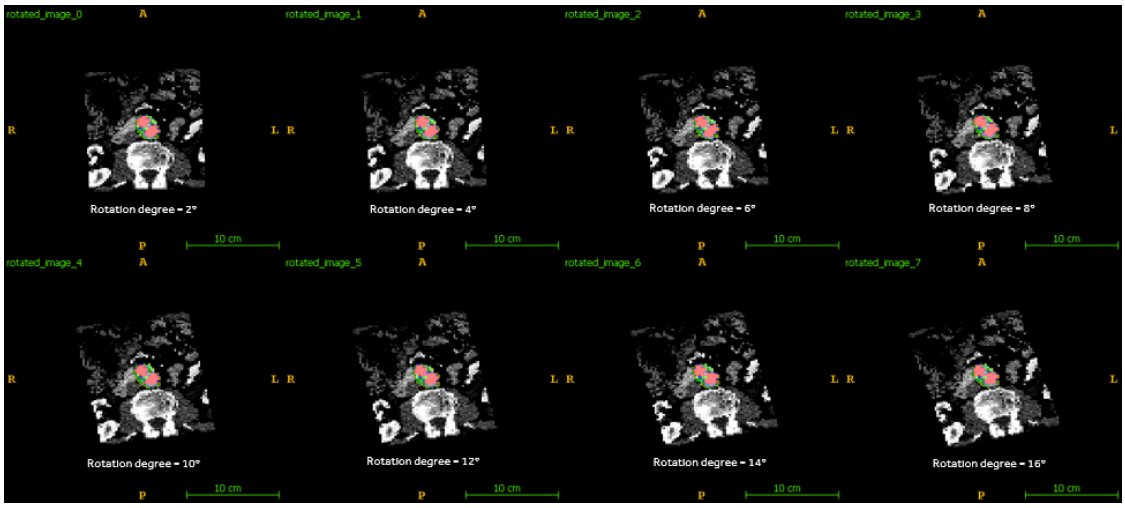


Figure 16.3: Result of data augmentation using a set of 8 rotations. Rotation degrees are indicated in each image.

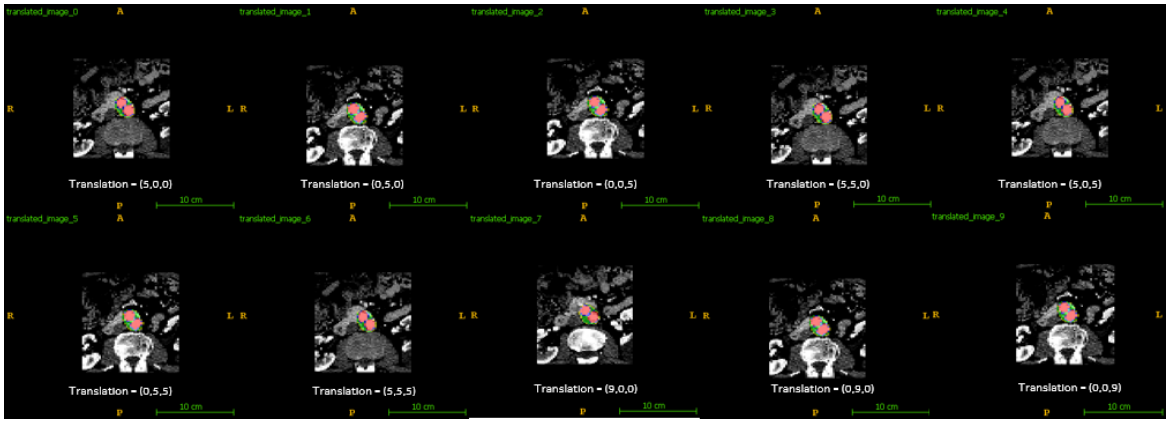


Figure 16.4: Result of data augmentation using a set of 10 translations. Translation values are indicated in each image .

Creating HDF5 Dataset

The amount of synthetically generated data is 18 x 27 = 486 volumes. In total, there are 486 + 27 = 513 trainable volumes. Ground-truth masks and testing

images are not augmented, so there are 10 scans available. In order to deal with such a big dataset, a *Hierarchical Data Format Version 5 (HDF5)* for Python (*h5py* package) has been used. It is an open source format that supports large, complex and heterogeneous data [64]. It uses a file directory-like structure that allows the user to organize data within the file in many structured ways and enables embedding of Metadata making it *self-describing* [64]. HDF5 files consists of *datasets* that are multidimensional arrays of homogeneous type, and *groups* that are container structures which can hold datasets and other groups. The structure for HDF5 files in this project are represented in the Figure 16.5, in which groups correspond to images and labels' datasets. For lately load data from the HDF5 files, the image and label datasets are retrieved.

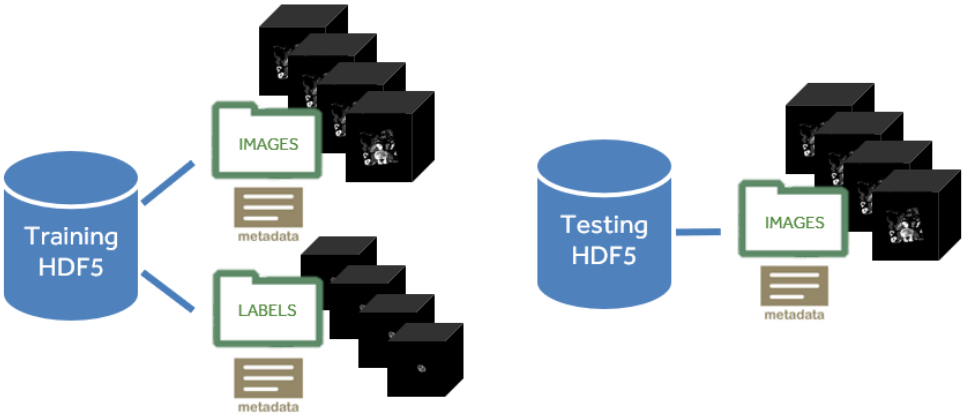


Figure 16.5: Structure of HDF5 files used in the project .

As it will be explained in further sections, different experiments have been performed using different dimensions of the images and different number of classes of for ground-truth mask. For that reason, a different HDF5 was created that contains the correspondence characteristics.

Final dataset

Original CTA scans have been divided into two groups: 27 scans from 12 patients for training and validation the algorithm and 10 scans from 5 patients for

testing it. Table 16.1 summarizes the employed datasets.

Table 16.1: Summary of the employed data for the 3D AAA segmentation.

	Number of patients	Number of volumes/masks
Training and Validation	12	27
Testing	5	10
TOTAL	17	37

Once the pre-processing and data-augmentation has been performed, the dataset for training and validation has been prepared. The CTA scans used for training and validation of the algorithm have been randomly split into subsets, where 75% of the scans correspond to the training dataset and the 25% of the scans correspond to the validation dataset. Table 16.2 summarizes the dataset used in this work.

Table 16.2: Training and validation datasets after augmentation and testing dataset.

	Number of volumes
Training	385
Validation	128
Testing	10

16.2. Network Implementation

The network architecture is based on the U-Net, explained in the Chapter 11. This network presents a contracting path where the resolution of the image decreases and the number of feature channels increases and an expanding path where the resolution of the image increases and the number of feature channels decreases. It is based on *Supervised Learning*, which trains the models with labelled data.

In this project, a *Modified U-Net* has been used, based on the approach

made by [65], which is inspired by the 3D V-Net [66] with modifications introduced from the 2D Fully Convolutional DenseNet (FC-DenseNet) [67] and the 2D Efficient neural network (ENet) [68]. Each of these networks consist of:

- The *V-Net* network is composed of convolution, deconvolution and pooling layers [65] displayed in an encoding and decoding path. A down-sampling is applied in every couple of layers of the encoding path, and in every pooling layer the number of feature maps is doubled. Before each down-convolution, a skip-connection is applied to pass higher resolution maps to the decoding path [65]. In the decoding path, an up-sampling is implemented in every couple of layers and feature fusion via the skip connection is applied [65].
- The *FC-DenseNet* has the same encoding-decoding architecture as the V-Net. However, opposed to V-Net, FC-DenseNet uses many convolutional layers with few channels each. Each layer is connected in a feed-forward fashion and batch normalization is applied before the convolutional layers, which helps to control over-fitting [65].
- The *ENet* provides real-time semantic segmentation by using low amount of filters, squeezing in as much information as possible in every parameter [65]. This network introduces a down-sampling block that combines max-pooling and strided convolution to avoid representation bottlenecks [10].

The Modified U-Net proposed in [65] is displayed in Figure 16.6.

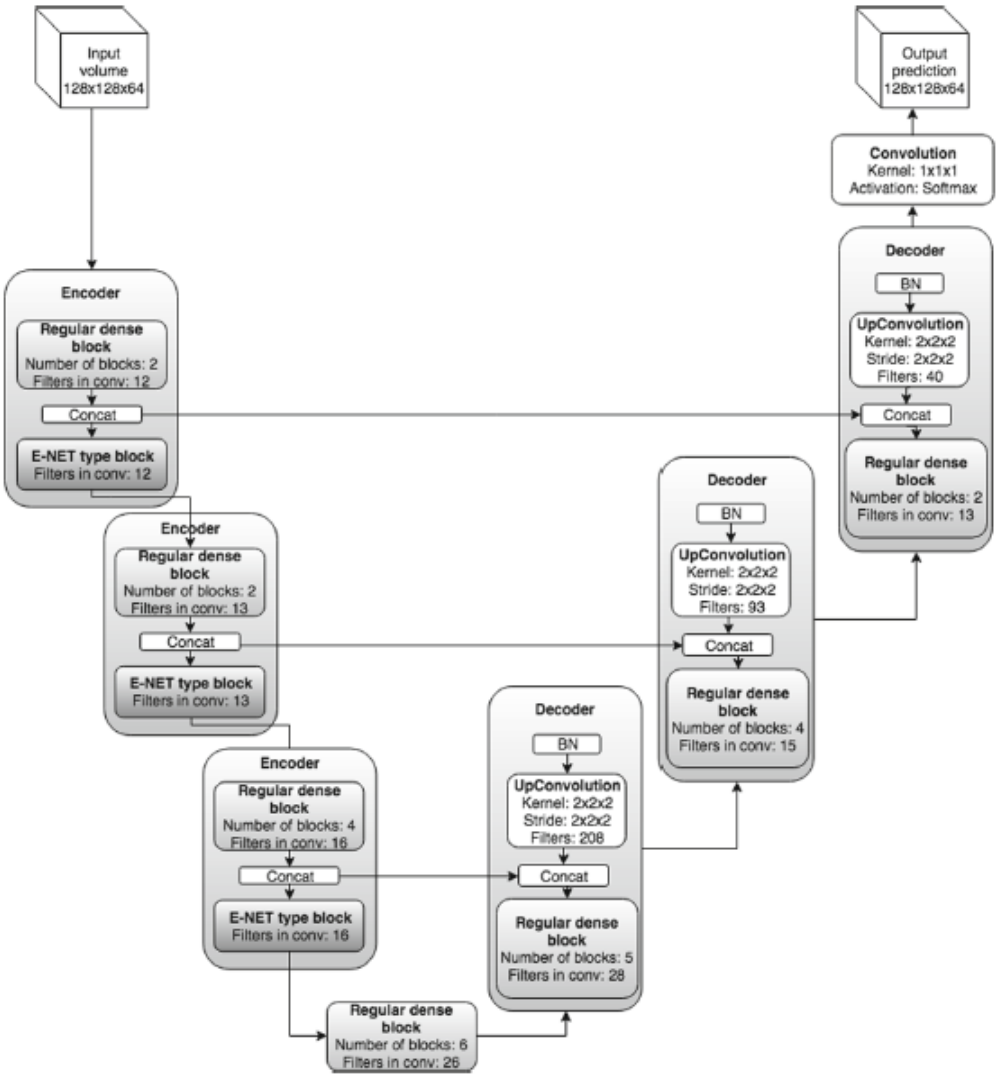


Figure 16.6: Scheme of the Modified U-Net. Extracted from [65]. Accessed on 25-02-2021.

In the following Figure 16.7 , there is a representation of the transformations that are applied to the 3D images during the training process in the network.

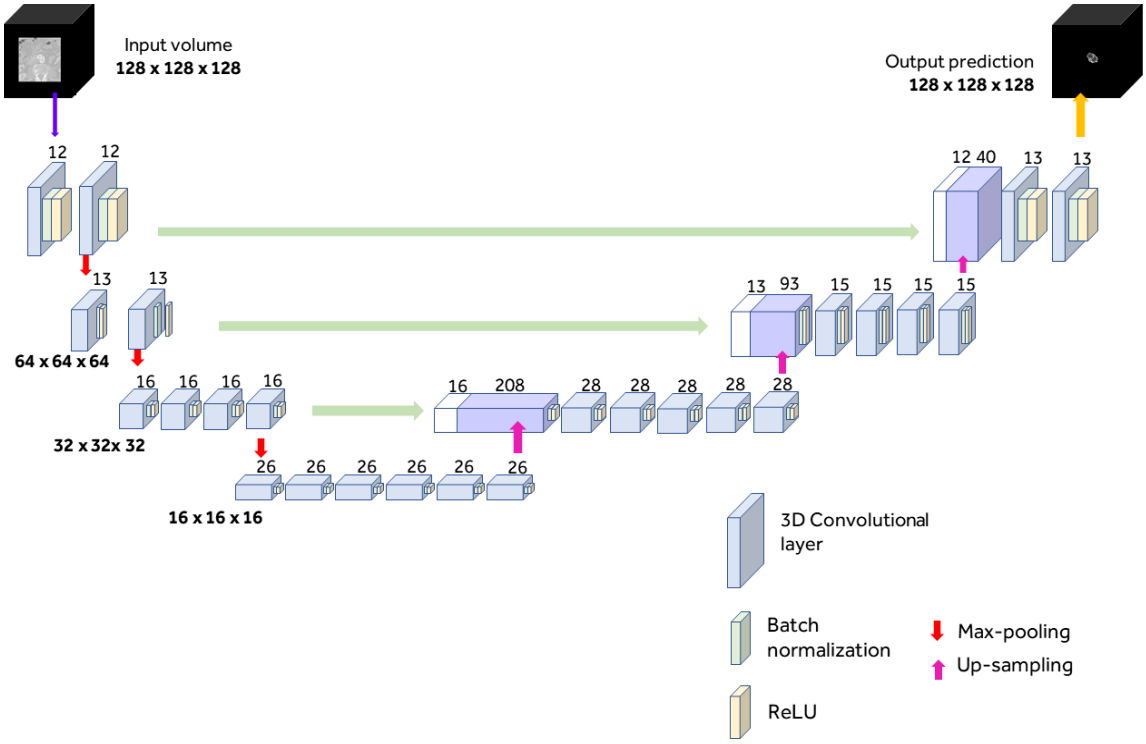


Figure 16.7: Scheme of the Modified U-Net.

The blue boxes represent the feature maps and black cubes represent the input and the output image, respectively. There are four depth levels and skipped connections (green arrow) between the ones on the same level. This is a modification with respect to the original U-Net, which has five depth levels and skipped connections. In the contracting path, each level has different number of convolutional layers, all of them using a 3 x 3 x 3 filter size, padding parameter set as *same*. Each of the convolutional layer is followed by a batch normalization layer before its ReLU activation function. Then, a 2 x 2 x 2 size and 2 x 2 x 2 stride max-pooling operation has been applied before the next level. In the Table 16.3 above there is a summary of the convolutional and max-pooling operations of the contraction path.

Table 16.3: Description of the contracting path of the implemented network, indicating the number of convolutional layers, number of filters and the final shape of the volumes after max-pooling.

	Number of Convolutional Layers	Number of filters	Output shape after max-pooling
Level 1	2	12	64 x 64 x 64
Level 2	2	13	32 x 32 x 32
Level 3	4	16	16 x 16 x 16
Level 4	6	26	-

The expanding path has a similar structure but applies an up-sampling or deconvolution that doubles the features map's size. Besides, the concatenation with the features maps of the corresponding contracting level takes place. The transposed convolution operations use 2 x 2 x 2 kernel size and 2 x 2 x 2 stride. In this path, convolution layers has the same characteristics as in contraction path. The Table 16.4 above summarizes the details of the expanding path. The last layer is a 1 x 1 x 1 convolution with 3 or 5 feature channels (depending on the experimental setting) and a Softmax activation function.

Table 16.4: Description of the expanding path of the implemented network, indicating the number of convolutional layers, number of filters and the final shape of the volumes after the deconvolution operation

	Number of Transposed Convolutional Filters	Number of Convolutional Layers	Number of filters	Output shape after max-pooling
Level 4	208	5	28	16 x 16 x 16
Level 3	93	4	15	32 x 32 x 32
Level 2	40	2	13	16 x 16 x 16
Level 1	-	1	Classes	128 x 128 x 128

16.3. Loss Functions

The *loss function* is a method of evaluating how well the algorithm models the dataset and helps to understand how much the predicted value differs from the ground-truth value. As it has been explained in Section 8.1 the non-linear Softmax activation function is a generalized logistic function widely applied to the output of a classifier in CNNs [61]. It converts a vector of activations values to a vector of real values in the range [0, 1] which all sum up to 1.

The Softmax is defined as follows:

$$p_k(x) = \frac{e^{a_k(x)}}{\sum_{n=1}^{N} e^{a_n(x)}} \tag{16.1}$$

where x refers to the pixel considered, $a_k(x)$ are the activation values of channel k and n for that pixel, N is the number of classes and p_k is the Softmax of class k [61].

In the original U-Net, after applying a pixel-wise Softmax, a weighted cross entropy is computed as loss function [46] and in [65], binary cross-entropy loss has been used.

In this project, two main experimental settings have been performed. First, the functioning of the network and its segmentation capability has been probed, using the hard-labeled mask (background, lumen and aneurysm). For this task, *Generalized Weighted Dice Loss* function has been used. Second, to perform the segmentation of both lumen and aneurysm structures including the uncertainty of the annotation process, the 5-labeled mask has been used. For that, a new *Soft Weighted Dice Loss* function has been designed, based on the *Soft Dice Loss* proposed in [69].

Weight map

In this work, the weight map refers to a matrix, with the same size of the input image, where each element represents a specific weight [61]. The weight mask is introduced in the computation of the loss function in order to cope with the problem of unbalanced classes [61]. In [46] Ronneberger, O et.al., proposed a weighted loss function where the background labels between touching cells obtained a large weight in the loss function. In Table 16.5, the number of labeled pixels of the training dataset for each class and the relative percentage over the total number, is shown. There is a relevant difference between the background and anatomical structures of lumen and aneurysm. In order to reduce the negative impact of the class imbalance, a *weight* has been applied to each class, which is calculated by the inverse of the class's volume. Inverting these probabilities enables to give more weight to classes that appear much less frequently than others. In the case of the background, the initial weight calculated was nearly 0 (0.002), so it was decided to set it as 0.1. Taking this into consideration, the respective weights for lumen and aneurysm classes have been adapted and a final weight map of [0.1, 0.53, 0.37] has been achieved (see Table 16.5).

Table 16.5: Number of labeled pixels for each class in the training dataset and relative percentage over the total number of pixels. The original training dataset includes 27 (without augmentation) of 342 x 442 x 638.

Class	Number of pixel	Percentage	Calculated weights	Final weights
Background	95539449	99.06	0.002	0.1
Lumen	373484	0.39	0.59	0.53
Aneurysm	529698	0.55	0.41	0.37

16.3.1. Dice Coefficient Loss

The *Dice Score Coefficient* (DSC) for multi-class segmentation is a measure of overlap between two samples: the ground-truth G and the prediction P [70]. The measure ranges from 0 to 1, where 1 denotes perfect overlap. It was originally

developed for binary data [71] and is represented as:

$$DSC = \frac{2\,|P \bigcap G|}{|G| + |P| + smooth} \tag{16.2}$$

where $|P \bigcap G|$ represents the common elements between sets G and P, $|G|$ represents the number of elements in set G (and likewise for set P) and *smooth* insures against zero denominator.

Because the ground-truth is binary (considering each class as one-hot-encoded), any pixels which are not activated in the ground truth mask are zero-out and for the remaining pixels, low-confidence predictions are penalized. Thus, higher values for $|P \bigcap G|$ expression lead to a better DSC. As the denominator of the expression double counts the common elements between the two sets, the number *2* is located in the numerator. Besides, in order to formulate a loss function which can be minimized [71], 1 - DSC is used as a DSC loss function (DL). In other words, the numerator is related with the common activations between the masks and the denominator represents the quantity of activation in each mask, separately [71].

Generalised Weighted Dice Loss

In [72], the *Generalised Dice Score (GDS)* was proposed as a modified version of the DSC, as a way of evaluating multiple class segmentation with a single score:

$$GDS = \frac{2\sum_{n=1}^{N} a_n(P_n G_n)}{\sum_{n=1}^{N} a_n(P_n + G_n)} \tag{16.3}$$

where G_i is the ground-truth of class n, P_i is the predicted probability of class n, N is the total number of classes and a_n corresponds to the weight of class n.

This way, weighting the DSC by the inverse of the class volume corrects the contribution of each label and reduces the correlation between the region size and the DSC [73].

16.3.2. Soft Dice Loss

A challenge facing delineation of anatomical structures is that boundaries are sometimes not well defined, leading to much ambiguity in the marking process along the contour of interest. In [69], they explored the use of a soft ground-truth mask to train a *Fully Convolutional Neural Network (FCNN)* for segmentation of *Multiple Sclerosis (MS)* lesions [69]. In this paper, as most of the inter-rater variability can be found along MS lesion contour voxels (see Figure 16.8), they proposed to modify the true delineations at those pixels by assigning soft class probabilities, and they present a loss function for training the FCNN that uses soft labeling within the framework of a Dice measure [69]. As they demonstrated, their new loss function provides additional information for the training process beyond the ground truth mask obtained by the expert.

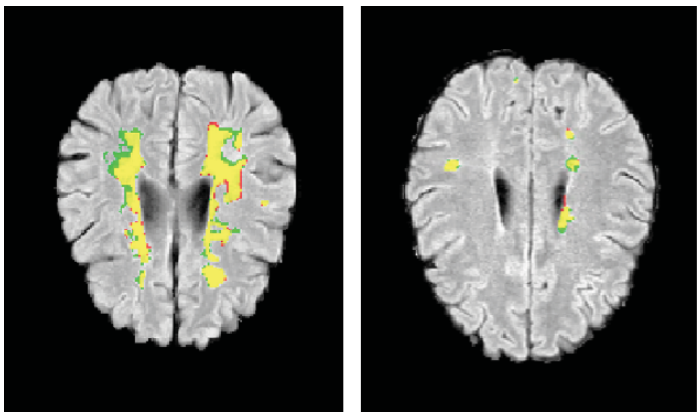


Figure 16.8: Examples of expert ground-truth binary masks: yellow - lesion voxels as defined by both raters, green - voxels delineated only by rater 1, red - voxels delineated only by rater 2. Extracted from [69]. Accessed on 20-11-2020.

In this paper, they modified the manual delineations of the experts, by expanding the original binary mask by 3D morphological dilation, as depicted in

the Figure 16.9. The voxels within the dilated region are assigned a soft label $0 < \gamma < 1$. In this project, as explained in Section 15.5.1, the uncertainty map has been generated from a dilatation of the hard-labeled mask, as done in the mentioned paper.

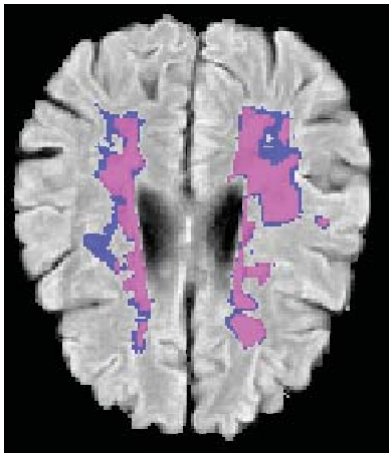


Figure 16.9: Examples of soft mask dilated up to 140 % of the original size that we used as ground truth in FCNN training: purple - original binary mask region, blue - soft labeled voxels. Extracted from [69]. Accessed on 20-11-2020.

DSC is usually used for FCNN training. As the mask contains an additional dilated region, it should has to be considered in the loss function. This function is based on the definition of DSC, studied in 16.3.1. Thus, training FCNN with the soft labeled mask is equivalent to training with the following loss function.

$$SoftDiceLoss = -\frac{\sum_i (T_i + \gamma D_i) P_i}{0.5 \sum_i P_i + 0.5 \sum_i (T_i + \gamma D_i)} \tag{16.4}$$

where matrix D is the binary mask of the dilated region, γ is the soft label assigned to voxels of the dilated region. Thus, soft labeled mask can be represented as $T_i + \gamma D_i$.

Table 16.6 shows obtained results in [69]. The soft value $\gamma = 0.3$ provided valuable information about near contour voxels during the training phase and achieved the highest Dice measure. Besides, Table 16.7 shows that using the soft

mask based loss function a clear improvement in both Dice, precision and recall measures is gained [69].

Table 16.6: Cross-validation results on the training set for different soft label values. Extracted from [69]. Accessed on 26-02-2021.

Soft label	Dice		Precision		Recall	
	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2
0.0	67.3	59.0	81.3	83.3	61.1	46.9
0.2	68.3	59.8	81.4	83.7	60.5	47.5
0.3	**69.9**	**62.0**	80.5	81.5	63.1	50.6
0.4	68.9	61.8	76.7	80.0	65.2	52.3

Table 16.7: Test results for the optimal combination of parameteres. Extracted from [69]. Accessed on 26-02-2021.

Method	Dice	Precision	Recall
Binary mask	56.0	82.9	44.6
Soft labeled mask	**57.8**	**83.8**	**46.6**

Results show that utilizing the soft labeled masks for FCNN training leads to a better precision-recall result, with a dependence on the volume of the soft labeled region and the soft class probability value [69].

Weighted Soft Dice Loss

Based on the concept of soft labeled mask, a new loss function has been designed in this project, for training with the 5-classes uncertainty mask (see Figure 16.10).

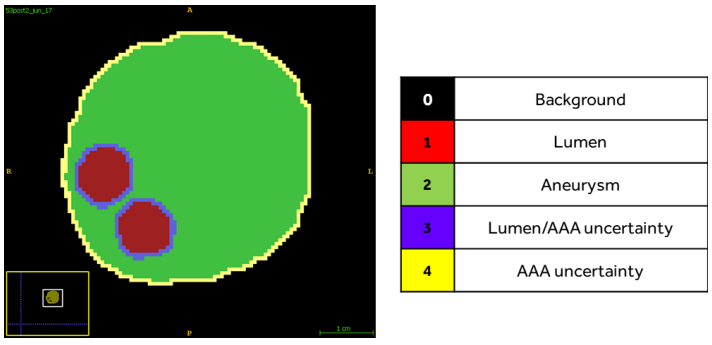


Figure 16.10: Example of 53post2 CTA post-operative 5-class uncertainty mask.

In this mask, labels 3 and 4 correspond to lumen-region uncertainty and aneurysm-region uncertainty, respectively. There are some points to consider: label = 3 voxels can correspond to lumen or aneurysm regions, they have different belonging probabilities; and label = 4 voxels can correspond to background or aneurysm regions, they have different belonging probabilities. As seen in 16.3.2, soft labeled mask is represented as:

$$SoftMask(Sm) = G_n + \gamma D_i \tag{16.5}$$

where matrix G is to the ground-truth mask , matrix D is the uncertainty mask, γ is the soft label assigned to voxels of the uncertainty mask and n is the label of the class. Thus, the soft labeled masks that correspond to each labels are:

$$BackgroundSoftMask = G_0 + \gamma D_4 \tag{16.6}$$

$$LumenSoftMask = G_1 + \gamma D_3 \tag{16.7}$$

$$AneurysmSoftMask = G_2 + \gamma(D_3 + D_4) \tag{16.8}$$

In addition, the weight map has been included in the *Weighted Soft Dice Coefficient (WSDC)*, which is represented as:

$$WSDC = \frac{2\sum_{n=1}^{N} a_n \sum_i Sm_i P_i}{\sum_{n=1}^{N} a_n \sum_i Sm_i + P_i} \tag{16.9}$$

The *Weighted Soft Dice Loss (WSDL)* function is then, *1 - WSDC*.

This loss function provides additional information for the training process, as the model will be able to predict lumen and aneurysm structures taking into

account undefined boundaries which can contain additional anatomical information about the lesion structure.

16.4. Output probability maps

The fully connected layer of a network produce values that are not normalized and cannot be interpreted as probabilities. Adding the Softmax function to the network enables to translate those values into a probability distribution [74], guaranteed to lie between 0 and 1. Softmax approach needs to ensure the output map resulting from the network architecture has multiple channels (that matches the number of classes). In this work, each channel in the output map would be treated as a probability map for a given class

Chapter 17

Training

This chapter defines the optimization parameters applied in this work, that allow the network to learn efficiently. Besides, it introduces the metrics chosen to evaluate the network performance and illustrates the training process and the obtained learning curves.

17.1. Metrics

The metrics considered to monitor the training and evaluate the test dataset are the Jaccard Coefficient and Hausdorff Distance.

17.1.1. Overlap based metrics

The DICE or overlap index is the most used metric in validating medical volume segmentations [75] and it is a statistical parameter that shows the similarity between two samples [61]. Its range of values is between 0 and 1, where 0 indicates that there is not overlapping and 1, that indicates that they are completely similar. This score is represented in the following expression:

$$DICE = \frac{2TP}{2TP + FP + FN} \tag{17.1}$$

where TP stans for the true positive, FP false positive and FN defines the false negative (see Figure 17.1).

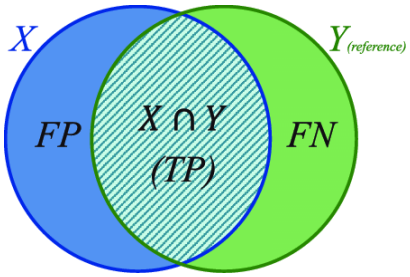


Figure 17.1: Elements used to compute the DSC, recall, and precision validation metrics. Extracted from [76]. Accessed on 27-02-2021.

Another common metric for these tasks is the *Jaccard Index (JAC)*, defined as the intersection between two samples divided by their union [75], that is

$$JAC = \frac{TP}{TP + FP + FN} \tag{17.2}$$

These expressions denote that JAC is always larger than DICE except at the extrema [0, 1], where they are equal [75]. Furthermore, DICE and JAC metrics are related according to

$$JAC = \frac{DICE}{2 - DICE} \tag{17.3}$$

$$DICE = \frac{2JAC}{1 + JAC} \tag{17.4}$$

Thus, it means that both DICE and JAC metrics measure the same overlapping aspects [75], and monitoring both of them does not provide any further

information. However, DICE does not satisfy the triangle inequality, so that is considered a semimetric [77]. This is why JAC metric has been employed in order to evaluate segmentation results.

In this project, the JAC is monitored for each class and background is not considered. This index has been calculated using *LabelOverlapMeasuresImageFilter* and*GetJaccardCoefficient* functions from SimpleITK.

17.1.2. Hausdorff Distance

Hausdorff Distance (HD) is one of the most informative metrics, as it is an indicator of the largest segmentation error [78]. In the task of segmentation of an organ or lesion of interest, the largest segmentation error quantified by HD can be a good measure of the usefulness [78] of the segmentation.

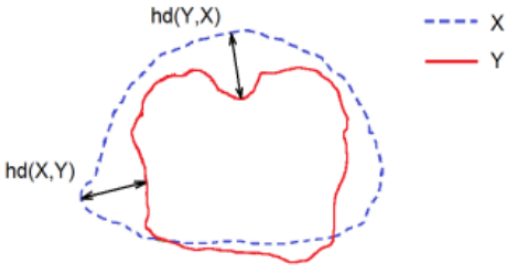


Figure 17.2: A schematic showing the Hausdorff Distance between points sets X and Y. Extracted from [78]. Accessed on 27-02-2021.

As depicted in the Figure 17.2, for two point sets X and Y, the one-sided HD from X to Y is defined as [78]:

$$hd(X,Y) = max_{x\epsilon X} min_{y\epsilon Y} \|x - y\|_2 \tag{17.5}$$

and for $hd(Y,X)$:

$$hd(Y,X) = max_{y\epsilon Y} min_{x\epsilon X} \|x-y\|_2 \tag{17.6}$$

The bidirectional HD between these two sets is:

$$HD(X,Y) = max(hd(X,Y), hd(Y,X)) \tag{17.7}$$

In all these definitions, the Euclidean distance [78] is used, but other metrics can be also used. *HD(X, Y)* is the longest distance from a point in on of the two sets to its closest point in the other set, and specifically in image segmentation, HD is computed between boundaries of the estimated and ground-truth segmentation, which consist of surfaces in 3D [78].

It may happen that large errors occur due to isolated locations, weak or missing edges or artifacts [78], even if the DICE score obtained high. This is something to take into consideration when evaluating the results.

This error has been calculated using *HausdorffDistanceImageFilter* and *GetHausdorffDistance* functions from SimpleITK.

17.2. Experimental set-up and optimization hyperparameters

In this work, 6 experiments have been carried out. First, the performance of the network has been tested using the hard-labeled mask and WDL function. For that, two image sizes have been used, [128, 128, 128] and [128, 128, 256], aiming to improve and compare the obtained results. Due to memory limitations, once the z dimension is increased, the batch size of the experiment must be limited. Thus, this pair of experiments are not directly comparable as the experimental set-up is different.

Second, once the network is able to segment hard volumes (lumen and aneurysm), soft labeled mask and WSDL function has been used to include the uncertainty information in the learning process. A γ values of 0.3 and 0.1 have been selected for image size of [128, 128, 128] and [128, 128, 256].

In the Table 17.1 below, there is a description of the experimental set-up implemented for this project.

Table 17.1: Experimental set-up of the project.

Experiments	Loss	Gamma	Classes	Image Size	BS	Epochs
1	WDL	-	3	128 x 128 x 128	3	100
2	WDL	-	3	128 x 128 x 256	1	250
3	WSDL	0.3	5	128 x 128 x 128	3	100
4	WSDL	0.3	5	128 x 128 x 256	1	250
5	WSDL	0.1	5	128 x 128 x 128	3	100
6	WSDL	0.1	5	128 x 128 x 256	1	250

In any deep learning algorithm, hyperparameters need to be initialized before training a model, as they directly control the behaviour of the training algorithm and have a significant impact on the performance of the model is being trained [79]. They can be divided into two categories: optimizer hyperparameters and model specific hyperparameters. These last have been explained within the description of the U-Net. Optimization hyperparameters are more related to the optimization and training process and can be summarized as following:

- *Learning rate* or *step size* refers to the amount of weights (usually in the range [0, 1]) that are updated during training and controls how quickly the model is adapted to the problem [80]. A learning rate that is too large can cause the model to converge too quickly, whereas a learning rate that is too small can cause the process to get stuck [80]. Thus, the proper tuning of this hyperparameter could avoid convergence problems.
- The *Number of epochs* is the number of complete passes through the training

dataset.

– The *Batch Size (BS)* defines the number of samples that will be propagated though the network.

In this work, Adam optimizer has been used in order to train the model. This is a significant difference compared to the original U-Net, where *Stochastic Gradient Descent (SGD)* is used [61]. *Adaptative Moment Estimation (Adam)* is a gradient-based optimization method that automatically computes adaptative learning rates for individual parameters [61].

Besides, a learning rate of 1e-04 and plateau learning rate decay with a factor of 0.02 is set, when the validation loss is not improved after 3 epochs, with a minimum learning rate of 1e-05. *categorical accuracy* has been used as a metric and it has been tried to minimize the weighted dice loss function and the weighted soft dice loss function (for each training experiments). Early stopping is also applied to avoid overffiting, thus, stopping the learning process after 8 epochs when there is no improvement. It has been set a *min-delta* of 0.001, that is the minimum change in the monitored quantity to qualify as an improvement.

The model is trained on a TITAN X Pascal (NVIDIA) GPU card with 11.91 GB and Tensorboard's graphs have been used in order to monitorize and measure how the models are improved.

17.3. Training and validation

As seen in the previous chapter, the training dataset is made up of 385 post-operative CTA scans, in which each of the image covers the thoracic area of the patient. For each sample there is the respective ground-truth image (soft-labeled or hard-labeled, depending on the experiment), with the same size, origin and spacing as images. On the other hand, 128 scans been used as validation dataset, together with 128 ground-truth masks.

17.3.1. Learning Curves

A learning curve is a plot of model learning performance and are widely used for algorithms that learn from a training dataset incrementally. It allows to diagnose problems such as underfitting or overfitting; the current state of the model at each step of the training algorithm can be evaluated on the training dataset, to give an idea of how well the model is learning, or it can be evaluated on the validation dataset to have an idea of how well the model is generalizing [81]. In this project, dual learning curves are used for model evaluation on both training and validation datasets. These learning curves have been created for two metrics: the model is optimized according the selected loss function and model performance is evaluated using categorical accuracy. The categorical accuracy is a score to maximize and loss function is a score to minimize. Thus, better scores for categorical accuracy indicate more learning and smaller values over time for loss means more learning.

For 1 and 2 experiments, obtained learning curves are depicted in the Figure 17.3 and Figure 17.4.

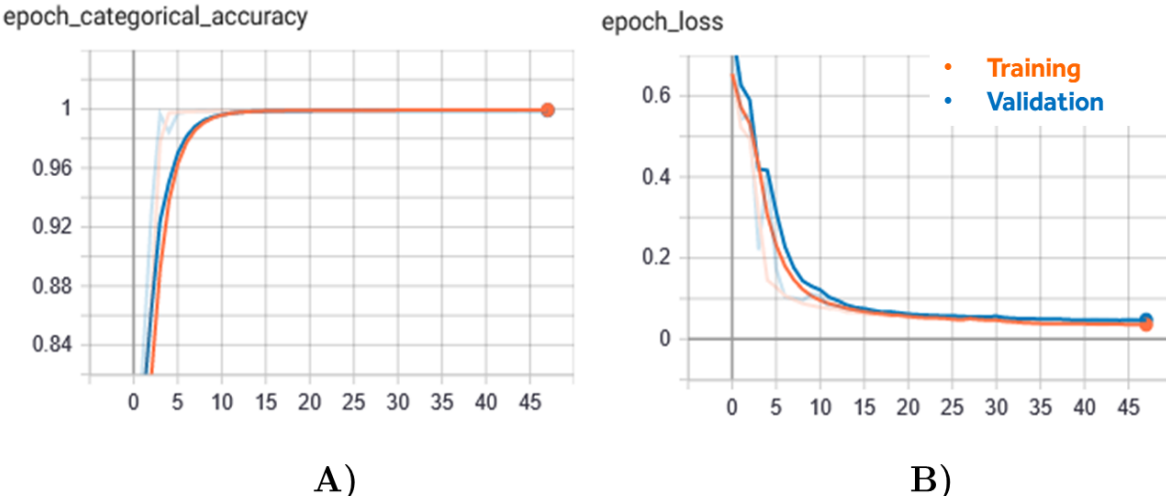


Figure 17.3: Learning curves for Weighted Dice Loss and hard-labeled masks, using an image size of 128 x 128 x 128. A) Categorical Accuracy evaluation over epochs. B) Weighted Dice Loss evaluation over epochs.

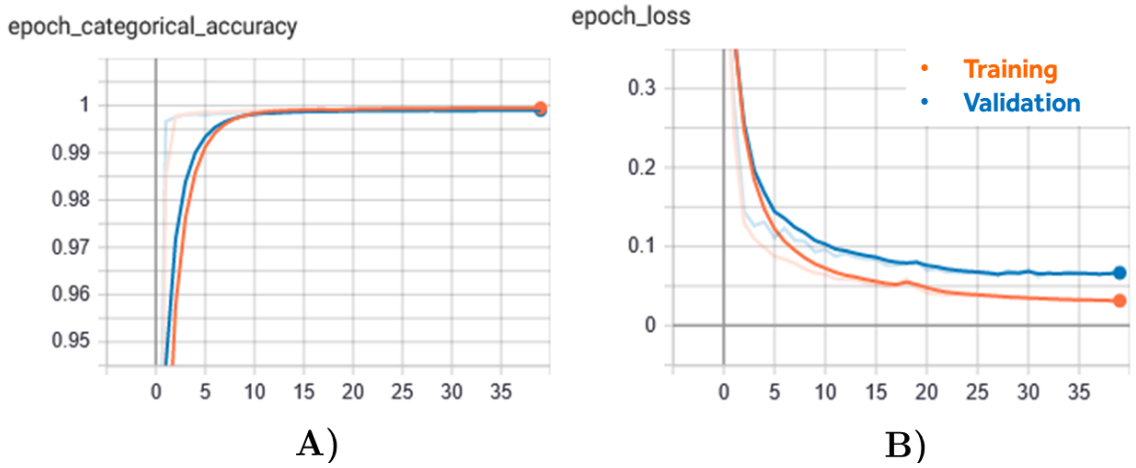


Figure 17.4: Learning curves for Weighted Dice Loss and hard-labeled masks, using an image size of 128 x 128 x 256. A) Categorical Accuracy evaluation over epochs. B) Weighted Dice Loss evaluation over epochs.

For 3 and 4 experiments, obtained learning curves are depicted in the Figure 17.5 and Figure 17.6.

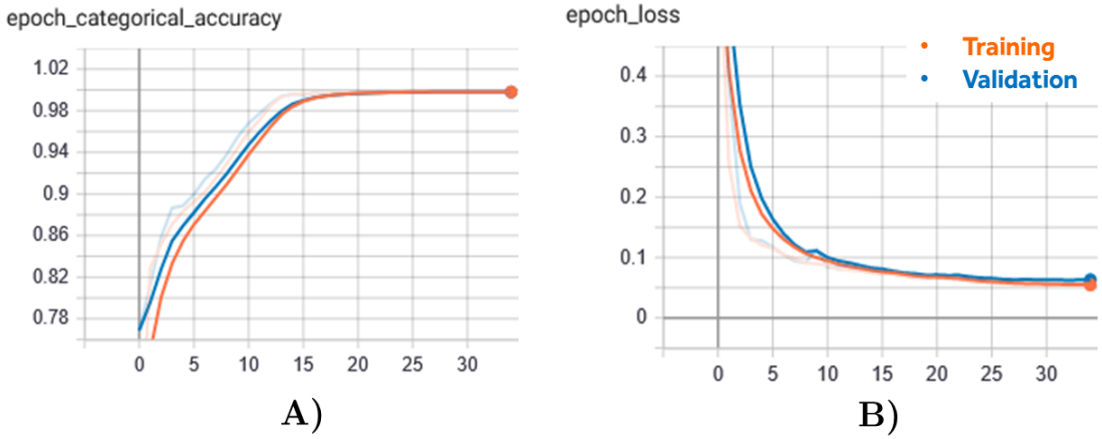


Figure 17.5: Learning curves for Weighted Soft Dice Loss and soft-labeled masks, using an image size of 128 x 128 x 128 and $\gamma = 0.3$. A) Categorical Accuracy evaluation over epochs. B) Weighted Dice Loss evaluation over epochs.

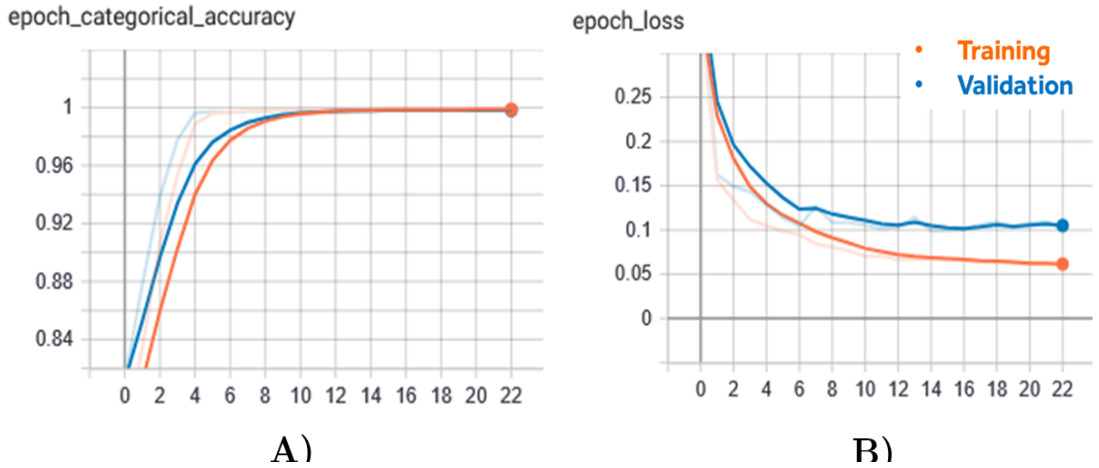


Figure 17.6: Learning curves for Weighted Soft Dice Loss and soft-labeled masks, using an image size of 128 x 128 x 256 and $\gamma = 0.3$. A) Categorical Accuracy evaluation over epochs. B) Weighted Dice Loss evaluation over epochs.

For 5 and 6 experiments, obtained learning curves are depicted in the Figure 17.7 and Figure 17.8.

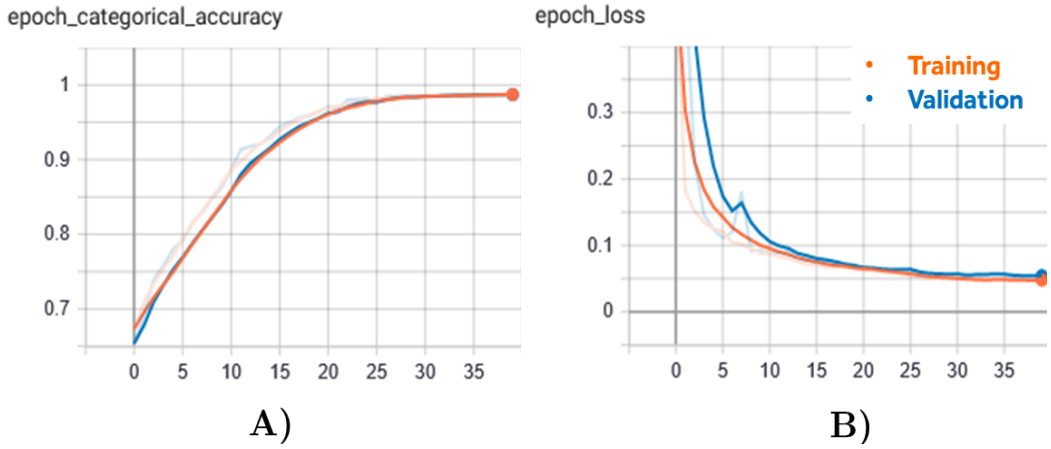


Figure 17.7: Learning curves for Weighted Soft Dice Loss and soft-labeled masks, using an image size of 128 x 128 x 128 and $\gamma = 0.1$. A) Categorical Accuracy evaluation over epochs. B) Weighted Dice Loss evaluation over epochs.

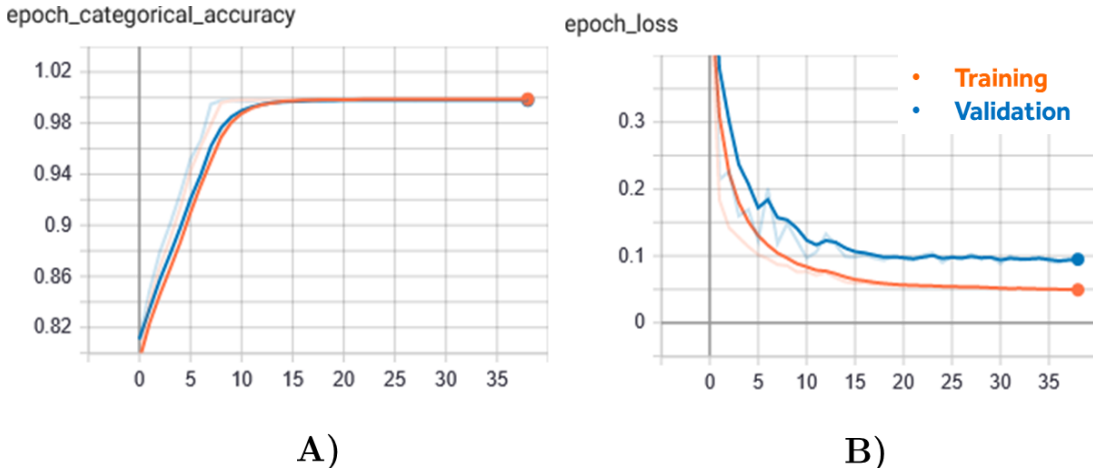


Figure 17.8: Learning curves for Weighted Soft Dice Loss and soft-labeled masks, using an image size of 128 x 128 x 256 and $\gamma = 0.1$. A) Categorical Accuracy evaluation over epochs. B) Weighted Dice Loss evaluation over epochs.

In general, in those experiments in which the batch size is set to 3, a good fit is observed, as training and validation loss decreases to a point of stability with a minimal gap between the two final loss values. Almost always, the loss of the model is lower on the training dataset than in the validation dataset [81], having a little gap between the train and validation loss curves (*generalization gap*).

In the experiments where the batch size is limited to 1, a bit of over-fitting is observed. Smaller batches mean more weight updates, as more batches fit into each epoch and the model is not guaranteed to converge to the global optima [82]. This fact is then translated into quantitative results, that will be analysed in Part 18. Besides, higher batches leads to lower asymptotic test accuracy [82].

Chapter 18

Results

Considering the training and validation results, the best models obtained are saved and then evaluated with the testing dataset, previously unseen by the network. The testing dataset is made up of 10 post-operative CTA scans. All these images are labeled for lately obtain quantitative analysis using explained metrics and qualitative study.

The predictions are 3D probability maps for every class, where the intensity of each pixel is the probability of it being (background), lumen or aneurysm, respectively. First, a resize to original size is applied to the grayscale output image, followed by an Otsu's thresholding that aims at selecting an optimal case-specific threshold [65]. Then, the intensity of the images has been inverted and small components (not vessels) have been removed, in order to reduce high Hausdorff Distance values that small components or artifacts could generate. Finally, lumen and aneurysm predicted masks have been combined in order to have a multi-class segmentation result.

18.1. Qualitative evaluation

This section includes a qualitative analysis of the obtained results; comparing the segmented volumes for lumen region achieved from eVida Vascular (eVida) tool and those for aneurysm region obtained with the previously designed HED network, with the results produced by the proposed method.

First, the analysis has been focused on the performance of aorta's surrounding vessels' segmentation an then, a comparative study of aneurysms' segmentations has been performed.

Evaluation on Iliac Arteries

As it has been mentioned, a correct segmentation of iliac arteries results complicated, mainly due to stent-related beam-hardening artifacts, that unable differentiation of these vessels in the region where the stent is implanted. As it has been analysed, this issue leads to the impossibility of having separated right and left iliac arteries, and in the worst case, to not be able to segment one or any of the branches. Hence, one of the objectives of the implementation of this method was to have a tool that enables facing these kind of limitations.

Table 18.1 collects a comparative of the results of segmentations achieved by eVida Vascular semi-automated tool and the segmentation volumes obtained by the proposed method. As it is seen, almost every segmented lumen volume achieved better results with the proposed method. It can be observed that in the cases in which the eVida Vascular tool was not able to segment any iliac arteries, the implemented network matches or improves the result.

Table 18.1: Comparative results between eVida Vascular semi-automated tool and the experiments performed in this project. 0) any iliac arteries have been segmented , 1) just one of the iliac arteries has been segmented (right or left), 2) both iliac arteries have been segmented but appear stuck one with the other and 3) both iliac arteries have been segmented correctly (separately).

		Modified U-Net					
	eVida Vascular	1	2	3	4	5	6
47 post	1	1	1	1	1	1	1
47 post1	0	1	1	1	1	3	1
54 post1	0	2	2	2	2	2	2
54 post2	0	2	2	2	0	2	2
58 post	2	3	3	2	2	2	2
58 post1	3	2	2	2	2	2	2
59 post	2	2	2	2	2	2	2
59 post1	0	2	2	2	2	2	2
61 post	2	2	2	2	2	2	2
61 post1	2	2	2	2	2	2	2

More generally speaking, in the Table 18.2, there is a comparative of the performance of the tools in the whole testing dataset. In the case of eVida Vascular tool, in the 40 % of the cases, none iliac was segmented. This percentage has declined to just the 2 % with the proposed method, which demonstrates that the segmentation has improved to a large degree. Besides, the percentage of cases in which both iliac have been segmented (even if stuck) has increased from 40 % to almost 75 %. However, the percentage of totally segmented iliac arteries (not stuck) has not increased; eVida Vascular tool was able to completely segment a unique case. In the proposed method, in just 3 of the 6 experiments this case was achieved to be fully segmented.

Table 18.2: General comparative results between eVida Vascular semi-automated tool the experiments performed in this project, for the segmentation of iliac arteries.

		Modified U-Net					
	eVida Vascular	1	2	3	4	5	6
0	40 % (4)	0 % (0)	0 % (0)	0 % (0)	10 % (1)	0 % (0)	0 % (0)
1	10 % (1)	20 % (2)	20 % (2)	20 % (2)	20 % (2)	10 % (1)	20 % (2)
2	40 % (4)	70 % (7)	70 % (7)	80 % (8)	70 % (7)	80 % (8)	80 % (8)
3	10 % (1)	10 % (1)	10 % (1)	0 % (0)	0 % (0)	10 % (1)	0 % (0)

Evaluation on Celiac Trunk and Superior Mesenteric artery

Regarding the segmentation results of the Celiac Trunk (CT), Table 18.3 shows a general comparative of the results. It is shown that, in general, the segmentations of the CT are worse with the implemented network, as the percentage of completely segmented cases has dropped from 80 % to just the 40 %. The part corresponding to the 40 % remaining, corresponds to incomplete CT segmentation.

Table 18.3: General comparative results between eVida Vascular semi-automated tool the experiments performed in this project, for the segmentation of Celiac Trunk.

		Modified U-Net					
	eVida	1	2	3	4	5	6
Not segmented	20 % (2)	40 % (4)	30 % (3)	40 % (4)	30 % (3)	20 % (2)	20 % (2)
Incomplete	0 % (0)	30 % (3)	30 % (3)	30 % (3)	20 % (2)	40 % (4)	40 % (4)
Segmented	80 % (8)	30 % (3)	40 % (4)	30 % (3)	50 % (5)	40 % (4)	40 % (4)

The same happens in the case of the Superior Mesenteric artery (SM), where the percentage of correct segmentations of the branch has declined, and the percentage of incomplete cases, has increased (see Table 18.4). It should be pointed that the Inferior Mesenteric (IM) artery has not been considered in this analysis, as none of them has been segmented using the eVida Vascular tool nor the implemented network in this project.

Table 18.4: General comparative results between eVida Vascular semi-automated tool the experiments performed in this project, for the segmentation of Superior Mesenteric artery.

		Modified U-Net					
	eVida	1	2	3	4	5	6
Not segmented	10 % (1)	40 % (4)	30 % (3)	40 % (4)	30 % (3)	20 % (2)	20 % (2)
Incomplete	0 % (0)	20 % (2)	30 % (3)	30 % (3)	20 % (2)	30 % (3)	40 % (4)
Segmented	90 % (9)	40 % (4)	40 % (4)	30 % (3)	50 % (5)	50 % (5)	40 % (4)

Evaluation on Renal arteries

Finally, regarding the segmentation performance on aorta's surrounding vessels, the produced segmentation of renal arteries has been analysed. As shown in Table 18.5, in nearly 30 % of the cases, any renal arteries have been segmented. This is an increment, comparing with the 20 % of non-segmented renal arteries achieved by eVida Vascular. Regarding the incompletely segmented renal arteries, the percentage has risen from 20 % to almost 60 % in the results produced by the proposed method. This is why the percentage of correctly segmented renal arteries has dropped from 60 % to just a % 10.

Table 18.5: General comparative results between eVida Vascular semi-automated tool the experiments performed in this project, for the segmentation renal arteries.

		Modified U-Net					
	eVida	1	2	3	4	5	6
Not segmented	20 % (2)	10 % (1)	40 % (4)	20 % (2)	40 % (4)	40 % (4)	40 % (4)
Incomplete	20 % (2)	90 % (9)	50 % (5)	80 % (8)	50 % (5)	50 % (5)	30 % (3)
Segmented	60 % (6)	0 % (0)	10 % (1)	0 % (0)	10 % (1)	10 % (1)	30 % (3)

Figure 18.1 and Figure 18.2, depicts a comparative of the segmentation results from 59post and 58post scans. In the case of Figure 18.1, it is seen that the predicted

lumen from the HED networks has a smoother surface (due to smoothing processing that is automatically performed in the tool) and branches are more accurately segmented. However, it is seen that, in the case of 58post, iliac arteries appear broken. In the case of the predicted lumen outputs by the proposed method, aorta's trunk is preserved and iliac does not appear broken, even if the accuracy of branches' segmentation seems not to be high enough.

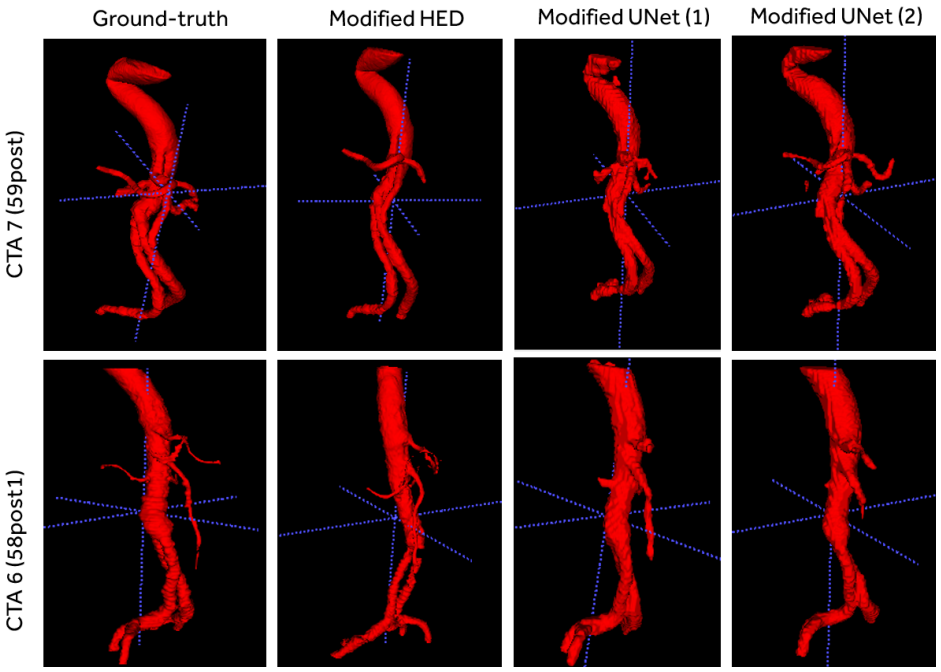


Figure 18.1: Comparative results of the pre-designed vs. Modified U-Net, using the Weighted Dice Loss function.

In the case of Figure 18.2, in the case of 59post, results are better regarding vessel segmentation, with the proposed method. In the 58post, aorta's trunk is well preserved, but surrounding branches are not completely segmented, specially in the case of renal arteries, which are not segmented.

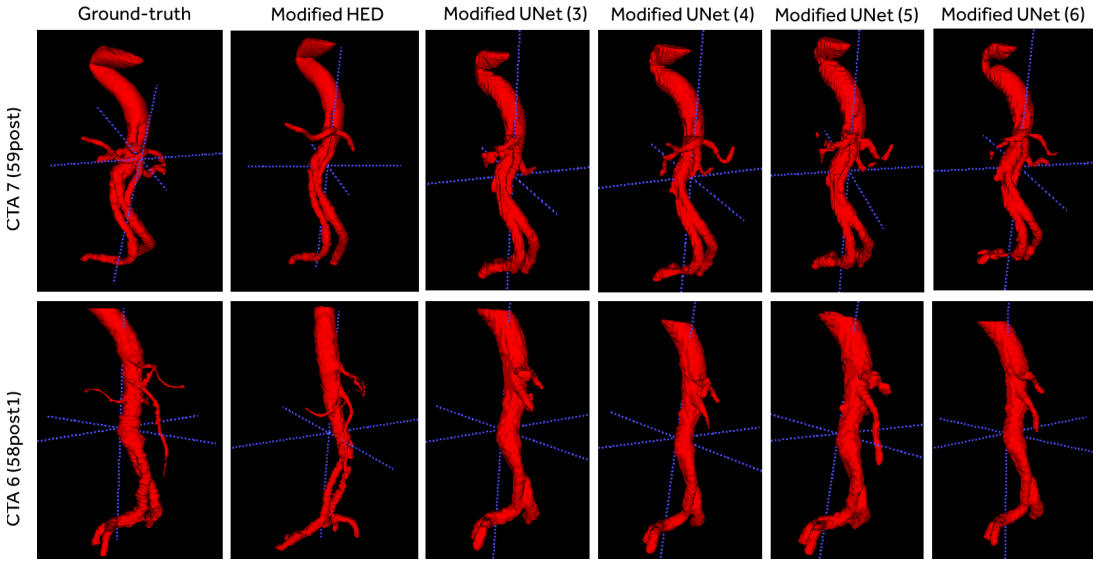


Figure 18.2: Comparative results of the pre-designed vs. Modified U-Net, using the Weighted Soft Dice Loss function.

Evaluation on Aneurysm

Regarding the evaluation of aneurysm's segmentation, the volumes obtained from the pre-designed network have been compared with those acquired from the implemented Modified U-Net. It is worth mentioning that the aneurysm belonging to the third post-operative scan of patient number 54 (54post2), was not initially segmented by the pre-designed network, so that it was generated from the scratch.

In the Figure 18.3 and Figure 18.4, there are two segmentation examples from two scans, 47post and 58post. Figure 18.3 depicts a comparative between the ground-truth aneurysm, the result obtained from the pre-designed network and results obtained from the proposed method in the experiment 1 and 2. Even if it is visually complicated to analyse, it is seen that all of the approaches have obtained similar aneurysm volume segmentations. In the case of HED network, the structure of the aneurysm seems to be partially broken and the start and end points of the thrombus appears not well delineated. In the second case (58post), the difference between obtained results are much clearer. It can be seen that the proposed method obtains similar volumes as the ground truth, whereas HED network captures an

extra volume that does not correspond with the aneurysm region.

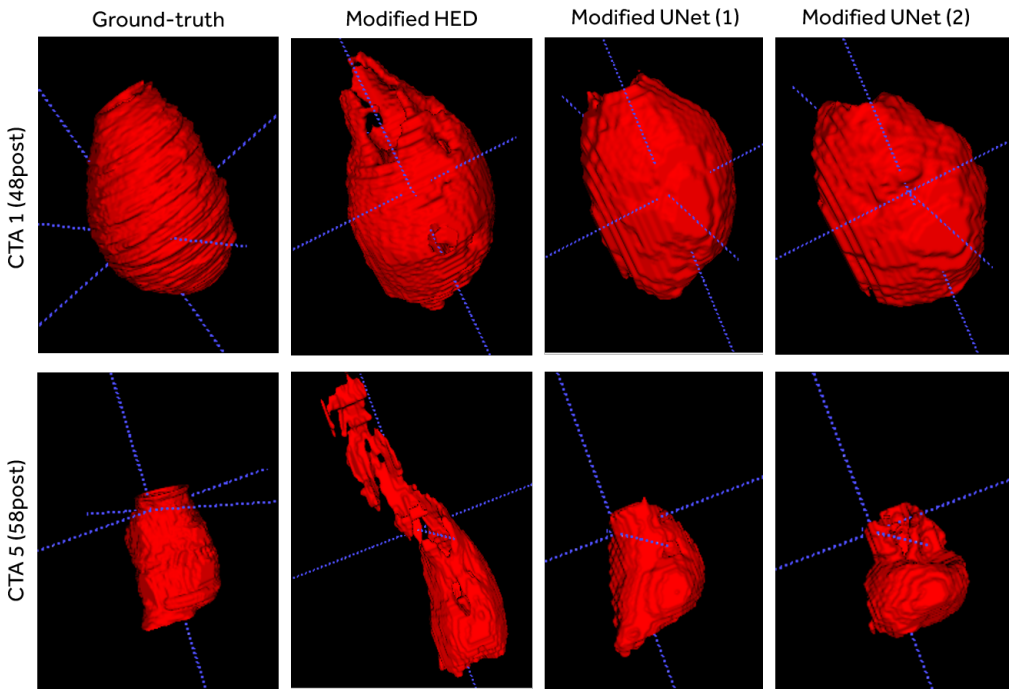


Figure 18.3: Comparative results of the pre-designed vs. Modified U-Net, using the Weighted Dice Loss function.

In the Figure 18.3, there is a comparative between the ground-truth aneurysm, the result obtained from the pre-designed network and results obtained from the proposed method in the experiment 3, 4, 5 and 6. As in the previous explanation, it is seen that the structure of the aneurysm has been somehow well segmented, even if in some cases (as in the experiment 4 and 6) the region above the iliac arteries has not been well-defined. As mentioned before, segmentations produced by the Modified U-Net obtain qualitatively better results that the HED, even if in experiments 4, 5 and 6, there is a part that appears broken. Anyway, the proposed method produces better results than the initial HED network.

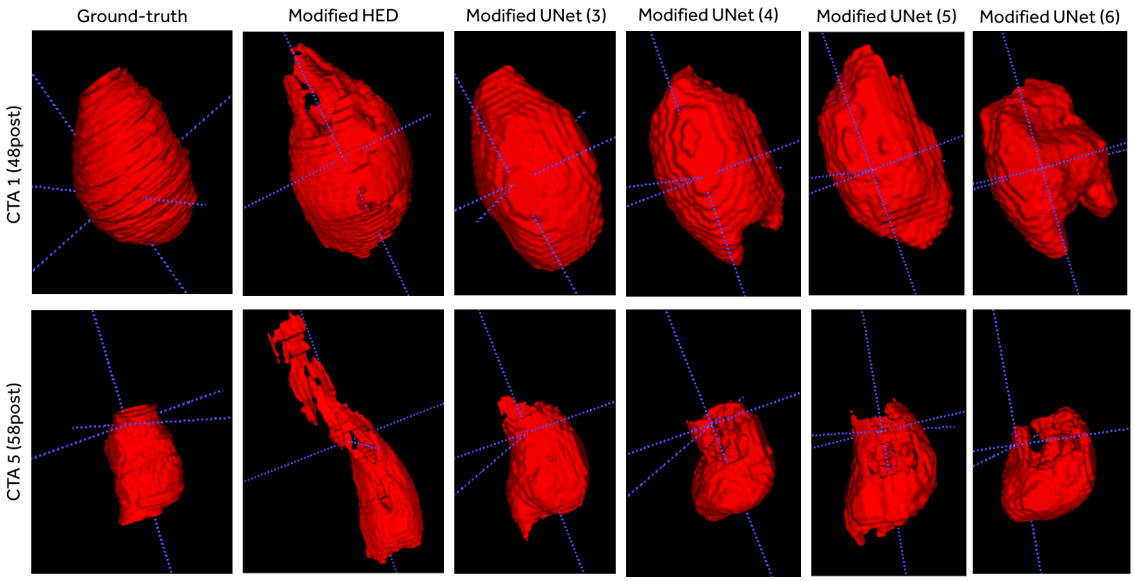


Figure 18.4: Comparative results of the pre-designed vs. Modified U-Net, using the Weighted Soft Dice Loss function.

Evaluation on multi-class volumes

In addition to the objective of optimizing initial segmentation results, another principal objective of the project has been the implementation of a multi-class segmentation network, that will predict both the lumen and the aneurysm from the CTA post-operative images. With the aim of qualitatively analyse some of the predictions, examples of 3 patients have been depicted in the following figures.

Figure 18.5 depicts the multi-class predictions obtained by the proposed method. It is seen that in those experiments in which the batch size is set to 3, the result is better. Regarding the segmentation of the vessels, experiment 1 and 3 obtained similar results as the ground-truth and in all cases iliac arteries are well-segmented.

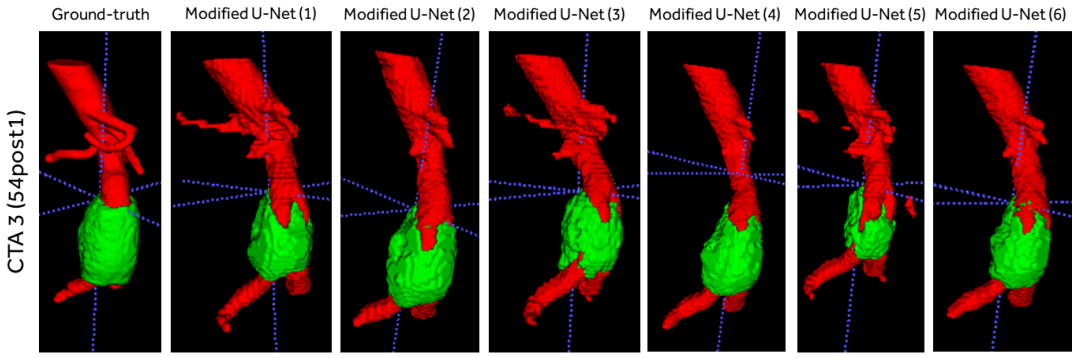


Figure 18.5: Prediction results on 54post1 volume.

For every predicted volumes by the implemented CNN, an axial slice has been extracted to analyse the performance of the network more in depth (see Figure 18.6). As it is seen, the experiments in which the delineation of the region is more accurate are 2, 3 and 4. Besides, iliac arteries have been separately segmented in all cases.

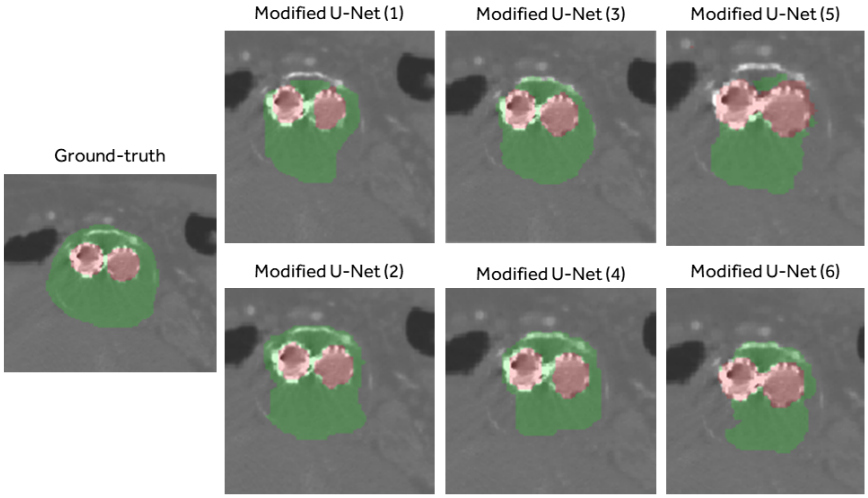


Figure 18.6: Axial view of the prediction results on 54post1 volume.

Another performance example is shown in the Figure 18.7. The lumen structure is larger in this case and every experiment has segmented it almost completely, including branches. In addition to that, experiment 1, 3 and 5 achieved better results regarding aneurysm volume.

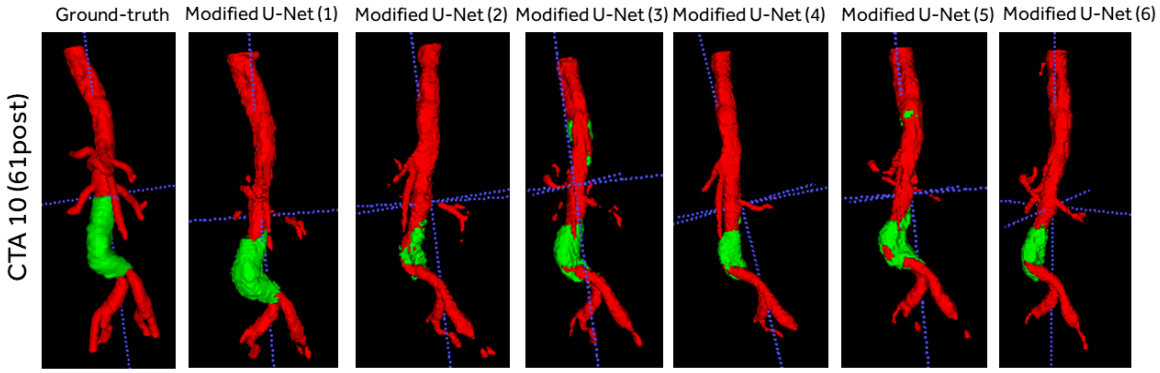


Figure 18.7: Prediction results on 61post volume.

As previously mentioned, the aneurysm volume of 54post2 scan, was not segmented by the pre-designed tool and the ground-truth mask was generated from the scratch. Results from the proposed method have not been accurate, as just the iliac arteries' trunk is segmented. According to the segmentation of the aneurysm, its structure is not well defined in the output results.

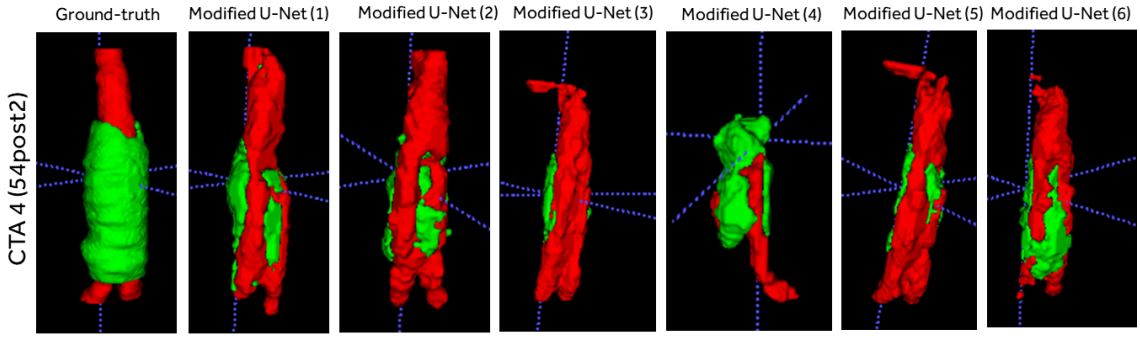


Figure 18.8: Prediction results on 54post2 volume.

18.2. Quantitative evaluation

As it can result difficult to qualitatively measure the difference between the obtained volumes, JAC and Hausdorff Distance (HD) metrics have been used to quantitatively compare the results againts the ground-truth masks.

Table 18.6 shows the average quantitative results for all the datasets, for the lumen and the aneurysm segmentation. According to the JAC, it is seen that the Modified U-Net that uses the WSDL function (experiments 3, 4, 5 and 6), performs better than the network with the WDL function (experiments 1 and 2), as higher overlap index is obtained. Besides, lumen segmentations seem to be more precise than aneurysm segmentation. In the case of HD metric, it should be cautiously interpreted. As it indicates, 1 and 2 experiments achieve better values for lumen volume and the opposed happens in the case of the aneurysm. As the structure of the lumen contains branches (or parts of them), the HD measured could be higher in presence of artifacts or unconnected segmented volumes, which would explain the distorted values. In the case of the aneurysm, as it is considered a closed anatomical structure (with no branches or external components), the probability of segmenting external volumes or artifacts is reduced, so that the HD is smaller.

Table 18.6: General evaluation of results. JAC: Jaccard Index, HD: Hausdorff Distance.

	JAC		HD	
	Lumen	Aneurysm	Lumen	Aneurysm
1	0.63	0.58	47.22	69.89
2	0.64	0.51	47.27	46.06
3	0.79	0.67	45.80	39.90
4	0.77	0.67	43.47	38.62
5	0.79	0.68	50.11	46.42
6	0.72	0.69	53.58	58.45

In the Figure 18.9, there is a box plot that shows the JAC and HD scores for experiments 1, 3 and 5, in which the size of the image has been set to 128 x 128 x 128 and the batch size is limited to 3. Figure 18.9A shows that experiments 3 and 5 obtained better overlap values, specially for the lumen segmentation, where the variability of results is smaller. The best result has been obtained when $\gamma = 0.3$. Comparing with the experiment 5, in which $\gamma = 0.1$, it seems that higher consideration of uncertainty region enables a better learning of the structures. Regarding HD values (Figure 18.9B), as explained before, results arise that there is a better performance in the aneurysm. As mentioned above, these results should be

cautiously interpreted. Even if results show lower mean distance values for aneurysm segmentation, their variability is also higher. The best result is shown in experiment 4, where $\gamma = 0.3$.

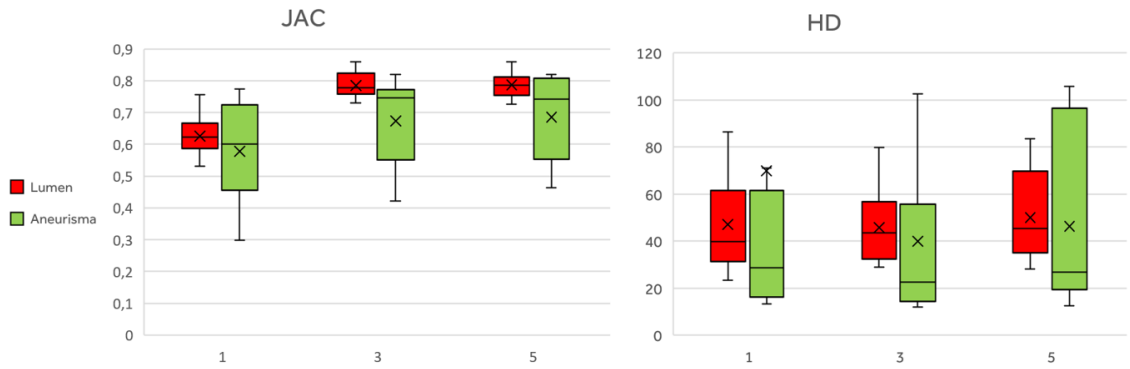


Figure 18.9: Box-plots showing the Jaccard (A) and Hausdorff Distance (B) score for experiments 1, 3 and 5, in which batch size was set to 3. Outliers' representation has been discarded.

In the case of the Figure 18.10, there is a box-plot that presents the JAC and HD scores for experiments 2, 4 and 6, in which the size of the image has been set to 128 x 128 x 256 and the batch size is limited to 1. In Figure 18.10A, JAC values show that experiments 4 and 6 achieved better results. Experiment 6 ($/gamma = 0.1$) presents slightly better results for aneurysm and the opposite happens in experiment 4 ($/gamma = 0.3$), where the lumen segmentation is better. As happens in the previous analysis, the HD shows the best results in experiment 4 (equivalent to experiment 3 in Figure 18.9) and predictions are better for aneurysm segmentation than for lumen segmentation. One more time, HD related results should be cautiously interpreted.

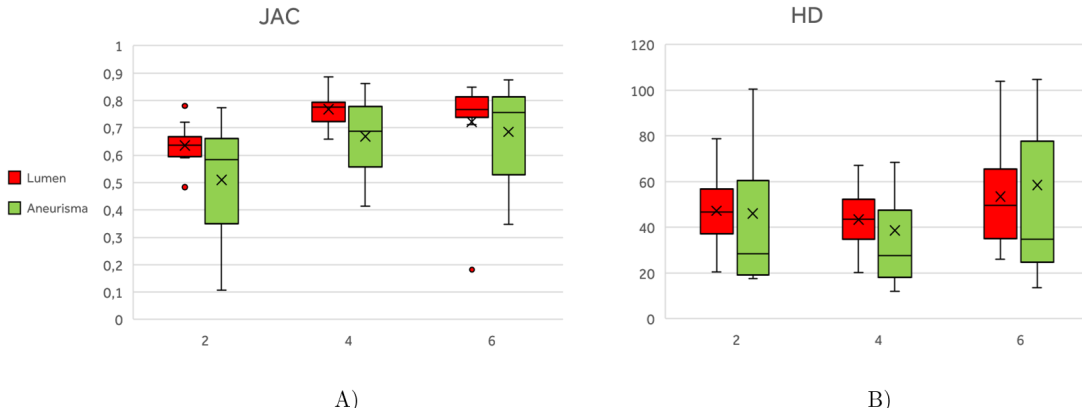


Figure 18.10: Box-plots showing the Jaccard (A) and Hausdorff Distance (B) score for experiments 2, 4 and 6, in which batch size was set to 1.

At this point, it should be considered how the evaluation of the results has been carried out in the experiments that use the WSDL (experiments 3, 4, 5 and 6), i.e. in those in which the uncertainty of region boundaries has been learnt (5-classes mask). In these cases, the predicted lumen output has been evaluated against: A) the mask that contains just the hard-labeled lumen (label 1), B) the mask that contains the hard-labeled lumen(label 1) and its uncertainty (label 3) and the predicted aneurysm output has been evaluated against: C) the mask that contains just the hard-labeled aneurysm (label 2), D) the mask that contains the hard-labeled aneurysm and its uncertainty (label 4), E) the mask that contains the hard-labeled aneurysm (label 2) and lumen-uncertainty (label 3) and F) the mask that contains the hard-labeled aneurysm (label 2) and both lumen and aneurysm uncertainty (label 3 and label 4). Displayed results in this section contain the best JAC and HD of these evaluations for the lumen and the aneurysm (see Table 18.7). In the case of the evaluation of the experiments that use the WDL (experiments 1 and 2), the predicted output of the lumen has been evaluated against the hard label of the lumen (label 1) from the 3-class mask and the output of the aneurysm has been evaluated against the hard label of the aneurysm (label 2).

As it is seen in the Table 18.7, there is a comparative of the effect of additional uncertainty information. In the case of the lumen segmentation evaluation (A and B), it is clear that the uncertainty region surrounding it includes additional information of the lumen, so that the metrics are higher. In the case of the aneurysm

segmentation evaluation (C, D, E and F), the case that includes the uncertainty of the lumen adds anatomical information. This means that a part of the lumen region's uncertainty belongs to the aneurysm. Otherwise, it is clear that in the case in which also the uncertainty of the aneurysm has been considered, good results have been obtained, meaning that it includes some additional information belonging to the aneurysm.

Table 18.7: Comparative results of the influence of the uncertainty region in the evaluation. JAC: Jaccard Index.

	JAC					
	A	B	C	D	E	F
3	0,65	0,79	0,51	0,63	0,67	0,64
4	0,63	0,77	0,51	0,62	0,67	0,63
5	0,65	0,79	0,53	0,65	0,68	0,66
6	0,59	0,72	0,54	0,63	0,69	0,65

Continuing with the previous analysis performed on the aneurysms' predictions, Table 18.8, shows a comparative mean JAC and HD values for the whole dataset, that compares the predicted aneurysm from the HED and the proposed method. As it is shown, predicted results obtained from the proposed method are better than those obtained by the pre-designed network, that shows low overlap value and very high HD value, which shows that the performance of that network is less accurate.

The HED network was trained using volumes of size 128 x 128 x 64, lower z dimension than the volumes used to train the experiments in this work. This may be a reason why the performance of the model presents lower accuracy. The most comparable results are those reached by the experiment 1, in which the volume size is 128 x 128 x 128 and the batch size is set to 3. In the case of HED network, the z dimension is lower than in experiment 1 and the batch size is limited to 2. As well as in experiment 1, HED network uses WDL as loss function and hard-labeled masks. Analysing obtained results, produced segmentations by the Modified U-Net present a 15 % higher JAC score. In addition to overlap measure, the HD index

shows that the largest segmentation error of HED network predictions is nearly 190, three times higher than the largest segmentation error obtained in the experiment 1 of the proposed method.

Besides, the HED network was trained using a binary mask, as opposed to the multi-class masks used in the implementation of the Modified U-Net. The use of a multi-class labeled masks may provide higher contextual information, which enables better segmentation of the aneurysm. Furthermore, the HED network is designed for both pre and post-operative scans, so that it uses common labeled masks for model training. The proposed method employs post-operative specific masks, which could increase the accuracy of the segmentation result.

Table 18.8: Comparative results of aneurysm segmentation. JAC: Jaccard Index, HD: Hausdorff Distance.

	JAC	HD
HED	0,49	186,81
1	0.58	69.89
2	0.51	46.06
3	0.67	40.18
4	0.67	39.34
5	0.68	47.24
6	0.69	58.73

18.3. Summary and conclusions

As a summary, it has been seen that uncertainty region surrounding fuzzy boundaries, gained additional information to the training process, which improves the final results. The uncertainty voxels around the lumen region are those which contain the most useful information. Otherwise, aneurysm surrounding uncertainty seems not to add any extra information for the learning process, which means that the manual annotation of the aneurysm in the ground truth is very accurate. In general, it resulted more difficult to delineate the interface between lumen and

aneurysm, due to the presence of the stent graft.

Regarding the differences observed in the segmentation volumes of lumen and aneurysm, in almost all the experiments, the lumen structure has been more accurately segmented. This would be due to the low contrast (aneurysm does not receive any contrast agent) of the aneurysm, as in that cases, it results impossible to manually delineate it.

As both qualitative and quantitative results arise, in those experiments where the batch size is set to 3, the segmentation results are better. This is logical as more number of samples are being propagated through the network, having more efficient learning process. The reason why the batch size of experiments 2, 4 and 6 has been set to 1 is due to memory limitation, as the dimension of the images has been increased in z.

It should also be pointed that the high dimension reduction that images suffered may have caused worse predictions. It is notable that when images of size 128 x 128 x 256, aorta's surrounding branches are better segmented, as it is capable of learning more spatial information along longitudinal dimension, even if the overall result is worse.

Finally, regarding the available tools, the proposed method enables to automatically segment both the lumen and the aneurysm, without having to manually set too many parameters, as in the case of eVida Vascular semi-automated tool.

Part VI

Conclusions and future work

Chapter 19

Conclusions

The general objective of this project was the development of an automatic segmentation tool that enables the follow-up and assessment of the aneurysm after the EVAR intervention, which was motivated by the lack of computerized image analysis tools to aid the clinicians in the post-operative stage of the patient.

The AAA is an abnormal dilatation of the main vessel of the human body which commits the life of the patients, in rupture case. For that reason, it is of vital importance to prevent and control its progression. Besides, in those patients in which surgical approach is selected, it is essential to lately evaluate and assess its evolution. As it has been demonstrated, post-operative based segmentation and extraction techniques are barely studied and implemented and the performed analysis in this work reflects the need to develop post-operative specific automatic tools, that consider the main differences in CTA scans between those and pre-operative scans.

As obtained results show, even if the performance of the implemented U-Net needs to be refined, it is clear that it was capable of matching or even increasing the segmentation results obtained by eVida Vascular semi-automated tool, which is a solution for pre-operative stage lumen segmentation. Thus, it has been seen that the proposed method that was specifically implemented for the post-operative CTA scans, optimizes lumen segmentation results. With regard to aneurysm's

segmentations, produced results are acceptable, even if the low contrast of the aneurysm region difficulties its detection.

What is more, the segmentation of both the lumen and the aneurysm served in order to have a multi-class map of the patient. This way, in addition to the volume of the aneurysm, the follow-up evaluation of AAA includes the structure of the lumen, which takes advance of anatomical context. As the learning requirements are higher, a multi-class segmentation requires higher computational cost comparing with binary segmentations. In general, it is also a time saving approach, as provides both volumes at once.

Besides, the performance of the Modified U-Net network has been analysed training it using two different loss functions. A novel approach in this project was the redesigning of the original Weighted Dice Loss to adapt it to an uncertainty-including mask, named as Weighted Soft Dice Loss function. Training the model with this loss function produced better results for both lumen and aneurysm, as the network was able to capture additional information regarding the uncertainty regions in manual annotation process. Concerning the idea of medical image annotation, the construction of uncertainty labeled mask reveals the importance of precise annotation, that is preferred to be made by clinical experts. The fact of having accurate, consistent and coherence ground-truth mask will make the difference between a good and poor segmentation result.

Broadly speeking, it seems that the implemented network achieved good results in comparison with available tools and networks previously designed. However, the relative volume difference between the automatic segmentation and the ground truth was large for some cases; sometimes a sub-segmentation occurred, as some anatomical regions were not segmented and sometimes over-segmentation occurred, as surrounding organs and structures were detected.

Finally, in relation of the applicability of this kind of automatic tools for the disease evolution assessment during follow-up, they should be considered as additional tools that would serve to aid and guide clinicians in the prevention, diagnostic, treatment and monitoring processes.

Chapter 20

Future work

The implementation done in this project could provide a good starting point for discussions and further research.

Firstly, extraction of the centerline, relying on previous segmentation of the lumen could be performed. The aortic centerline corresponds to the medial line or axis of vessels, whose loci correspond to the centroid of succesive cross-sections [10]. The extraction of the centerline would allow to measure the real diameters and lengths along the aorta.

Secondly, registration of post-operative aneurysm volumes over time would help in the follow-up assessment, as it would be possible to qualitatively and quantitatively measure the progression of the disease. It could be made based on lumen segmentation and graph matching method, as proposed in [83].

Moreover, regarding the stent-related beam-hardening of CTA scans, *Generative Adversarial Networks (GANs)* could be used in order to reduce these artifacts and keep the original information and details of the original scan, as in the project presented by [84].

Regarding the generation of ground-truth masks, the boundaries of the endograft have been omitted. Further developments could annotate stent region,

in order to have more precise segmentation result, and detect the existent interface between lumen and aneurysm.

Furthermore, the implementation of the automated segmentation tool could integrate image characteristics from both pre and post-operative stage. This way, a unique algorithm would be able to treat pre and post-operative image and segment each of them, for the planning and monitoring tasks, respectively.

Besides, in feature works, the segmentation of the AAA would be extended to the segmentation of another type of aneurysms, as suprarenal aneurysms or thrombus that extend to the iliac arteries.

Bibliography

[1] Fitzgerald, G., "The Aorta - Branches - Aortic Arch - TeachMeAnatomy," Accessed on 09/02/2021 to https://teachmeanatomy.info/abdomen/vasculature/arteries/aorta/, 2018. [Online]. Available: https://teachmeanatomy.info/abdomen/vasculature/arteries/aorta/

[2] P. M. Shaw, J. Loree, and R. C. Gibbons, "Abdominal Aortic Aneurysm," in *StatPearls*. Treasure Island (FL): StatPearls Publishing, 2020. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK470237/

[3] H. Jiang, T. Sasaki, E. Jin, M. Kuzuya, and X. W. Cheng, "Inflammatory Cells and Proteases in Abdominal Aortic Aneurysm and its Complications," *Current Drug Targets*, vol. 19, no. 11, pp. 1289–1296, 2018.

[4] K. W. Johnston, R. B. Rutherford, M. D. Tilson, D. M. Shah, L. Hollier, and J. C. Stanley, "Suggested standards for reporting on arterial aneurysms. Subcommittee on Reporting Standards for Arterial Aneurysms, Ad Hoc Committee on Reporting Standards, Society for Vascular Surgery and North American Chapter, International Society for Cardiovascular Surgery," *Journal of Vascular Surgery*, vol. 13, no. 3, pp. 452–458, Mar. 1991.

[5] S. Aggarwal, A. Qamar, V. Sharma, and A. Sharma, "Abdominal aortic aneurysm: A comprehensive review," *Experimental & Clinical Cardiology*, vol. 16, no. 1, pp. 11–15, 2011. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3076160/

[6] McGloughlin Timothy M. and Doyle Barry J., "New Approaches to Abdominal Aortic Aneurysm Rupture Risk Assessment," *Arteriosclerosis,*

Thrombosis, and Vascular Biology, vol. 30, no. 9, pp. 1687–1694, Sep. 2010, publisher: American Heart Association. [Online]. Available: https://www.ahajournals.org/doi/10.1161/ATVBAHA.110.204529

[7] D. Carino, T. P. Sarac, B. A. Ziganshin, and J. A. Elefteriades, "Abdominal Aortic Aneurysm: Evolving Controversies and Uncertainties," *The International Journal of Angiology: Official Publication of the International College of Angiology, Inc*, vol. 27, no. 2, pp. 58–80, Jun. 2018.

[8] Conte, M., Eichler, C., and Warren, G., "Department of Surgery - Endovascular Aneurysm Repair," 2020. [Online]. Available: https://surgery.ucsf.edu/conditions--procedures/endovascular-aneurysm-repair.aspx

[9] López-Linarez, K., García, I., Garcia-Familiar, A, Macía, I., and González Ballester, M.A., "3D convolutional neural network for abdominal aortic aneurysm segmentation," 2019. [Online]. Available: https://arxiv.org/abs/1903.00879

[10] K. López-Linares, "Image analysis and deep learning to support endovascular repair of abdominal aortic aneurysms," Ph.D. Thesis, Universitat Pompeu Fabra, Jun. 2019, accepted: 2019-07-03T11:23:19Z Publication Title: TDX (Tesis Doctorals en Xarxa). [Online]. Available: http://www.tdx.cat/handle/10803/667102

[11] Cleveland Clinic, "Endoleak Treatments," Accessed on 08-02-2021 to https://my.clevelandclinic.org/health/treatments/16960-endoleak-treatments., 2019. [Online]. Available: https://my.clevelandclinic.org/health/treatments/16960-endoleak-treatments

[12] E. L. Chaikof, R. L. Dalman, M. K. Eskandari, B. M. Jackson, W. A. Lee, M. A. Mansour, T. M. Mastracci, M. Mell, M. H. Murad, L. L. Nguyen, G. S. Oderich, M. S. Patel, M. L. Schermerhorn, and B. W. Starnes, "The Society for Vascular Surgery practice guidelines on the care of patients with an abdominal aortic aneurysm," *Journal of Vascular Surgery*, vol. 67, no. 1, pp. 2–77.e2, Jan. 2018, publisher: Elsevier. [Online]. Available: https://www.jvascsurg.org/article/S0741-5214(17)32369-8/abstract

[13] Bibb, R. and Paterson, A., "Hounsfield Scale - an overview | ScienceDirect Topics," 2015. [Online]. Available: https://www.sciencedirect.com/topics/medicine-and-dentistry/hounsfield-scale

[14] F. Lalys, V. Yan, A. Kaladji, A. Lucas, and S. Esneault, "Generic thrombus segmentation from pre- and post-operative CTA," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 9, pp. 1501–1510, Sep. 2017. [Online]. Available: https://doi.org/10.1007/s11548-017-1591-8

[15] I. Macía, M. de Blas, J. H. Legarreta, L. Kabongo, s. Hernández, J. M. Egaña, J. I. Emparanza, A. García-Familiar, and M. Graña, "Standard and fenestrated endograft sizing in EVAR planning: Description and validation of a semi-automated 3D software," *Computerized Medical Imaging and Graphics: The Official Journal of the Computerized Medical Imaging Society*, vol. 50, pp. 9–23, Jun. 2016.

[16] T. Siriapisith, W. Kusakunniran, and P. Haddawy, "Outer Wall Segmentation of Abdominal Aortic Aneurysm by Variable Neighborhood Search Through Intensity and Gradient Spaces," *Journal of Digital Imaging*, vol. 31, no. 4, pp. 490–504, Aug. 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6113142/

[17] F. Zhuge, G. D. Rubin, S. Sun, and S. Napel, "An abdominal aortic aneurysm segmentation method: level set with region and statistical information," *Medical Physics*, vol. 33, no. 5, pp. 1440–1453, May 2006.

[18] J. Maiora, B. Ayerdi, and M. Graña, "Random forest active learning for AAA thrombus segmentation in computed tomography angiography images," *Neurocomputing*, vol. 126, pp. 71–77, Feb. 2014. [Online]. Available: https://doi.org/10.1016/j.neucom.2013.01.051

[19] Gosh, S., Das, N., Das, I., and Maulik, U., "Understanding Deep Learning Techniques for Image Segmentation," *ACM Computing Surveys*, vol. 52, 2019.

[20] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," *arXiv:1511.08458 [cs]*, Dec. 2015, arXiv: 1511.08458. [Online]. Available: http://arxiv.org/abs/1511.08458

[21] Karn, U., "A Quick Introduction to Neural Networks," Aug. 2016. [Online]. Available: https://ujjwalkarn.me/2016/08/09/quick-intro-neural-networks/

[22] Z. Meng, Y. Hu, and C. Ancey, "Using a Data Driven Approach to Predict Waves Generated by Gravity Driven Mass Flows," *Water*, vol. 12, Feb. 2020.

[23] A. S. V, "Understanding Activation Functions in Neural Networks," Mar. 2017. [Online]. Available: https://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-9491262884e0

[24] E. Amor, "Understanding Non-Linear Activation Functions in Neural Networks," May 2020. [Online]. Available: https://medium.com/ml-cheat-sheet/understanding-non-linear-activation-functions-in-neural-networks-152f5e101eeb

[25] J. Brownlee, "A Gentle Introduction to the Rectified Linear Unit (ReLU)," Jan. 2019. [Online]. Available: https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/

[26] Moujahid, A., "A Practical Introduction to Deep Learning with Caffe and Python // Adil Moujahid // Data Analytics and more." [Online]. Available: http://adilmoujahid.com/posts/2016/06/introduction-deep-learning-python-caffe/

[27] S. Sharma, "Activation Functions in Neural Networks," Accessed on 15-02-2021 to https://towardsdatascience.com, Feb. 2019. [Online]. Available: https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6

[28] Coy, MA, "Forward and Back Propagation over a CNN... code from Scratch!!" 2020. [Online]. Available: https://www.linkedin.com/pulse/forward-back-propagation-over-cnn-code-from-scratch-coy-ulloa

[29] J. Leonel, "Multilayer Perceptron," Accessed on 14-02-2021 to https://medium.com/@jorgesleonel/multilayer-perceptron-6c5db6a8dfa3, Mar. 2019. [Online]. Available: https://medium.com/@jorgesleonel/multilayer-perceptron-6c5db6a8dfa3

[30] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, Jan. 1979, conference Name: IEEE Transactions on Systems, Man, and Cybernetics.

[31] R. Nock and F. Nielsen, "Statistical region merging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1452–1458, Nov. 2004, conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[32] N. Dhanachandra, K. Manglem, and Y. J. Chanu, "Image Segmentation Using K -means Clustering Algorithm and Subtractive Clustering Algorithm," *Procedia Computer Science*, vol. 54, pp. 764–771, Jan. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050915014143

[33] L. Najman and M. Schmitt, "Watershed of a continuous function," *Signal Processing*, vol. 38, no. 1, pp. 99–112, Jul. 1994. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0165168494900590

[34] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, Jan. 1988. [Online]. Available: https://doi.org/10.1007/BF00133570

[35] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001, conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[36] N. Plath, M. Toussaint, and S. Nakajima, "Multi-class image segmentation using conditional random fields and global classification," in *Proceedings of the 26th Annual International Conference on Machine Learning*, ser. ICML '09. New York, NY, USA: Association for Computing Machinery, Jun. 2009, pp. 817–824. [Online]. Available: https://doi.org/10.1145/1553374.1553479

[37] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges," *Journal of Digital Imaging*, vol. 32, May 2019.

[38] "Image Segmentation in Deep Learning: Methods and Applications," 2019. [Online]. Available: https://missinglink.ai/guides/computer-vision/image-segmentation-deep-learning-methods-applications/

[39] N. Madali, "Introduction To Image Segmentation," Jul. 2020. [Online]. Available: https://towardsdatascience.com/introduction-to-image-segmentation-9636fa95922

[40] T. Lei, R. Wang, Y. Wan, B. Zhang, H. Meng, and A. K. Nandi, "Medical Image Segmentation Using Deep Learning: A Survey," *arXiv:2009.13120 [cs, eess]*, Dec. 2020, arXiv: 2009.13120. [Online]. Available: http://arxiv.org/abs/2009.13120

[41] Meel, V., "ANN and CNN: Analyzing Differences and Similarities," Nov. 2020. [Online]. Available: https://viso.ai/deep-learning/ann-and-cnn-analyzing-differences-and-similarities/

[42] C. Camacho, "Convolutional Neural Networks," 2018, accessed on 15-02-2021 to https://cezannec.github.io. [Online]. Available: https://cezannec.github.io/Convolutional_Neural_Networks/

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

[44] Cruz-Roa, A., "Automatic detection of invasive ductal carcinoma in whole slide images with Convolutional Neural Networks," 2014. [Online]. Available: https://www.researchgate.net/publication/263052166_Automatic_detection_of_invasive_ductal_carcinoma_in_whole_slide_images_with_Convolutional_Neural_Networks

[45] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 3431–3440, iSSN: 1063-6919.

[46] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv:1505.04597 [cs]*, May 2015, arXiv: 1505.04597. [Online]. Available: http://arxiv.org/abs/1505.04597

[47] H. Lamba, "Understanding Semantic Segmentation with UNET," Feb. 2019. [Online]. Available: https://towardsdatascience.com/understanding-semantic-segmentation-with-unet-6be4f42d4b47

[48] W. Burton, "2D or 3D? A Simple Comparison of Convolutional Neural Networks for Automatic Segmentation of...," Sep. 2019.

[49] z. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," *arXiv:1606.06650 [cs]*, Jun. 2016, arXiv: 1606.06650. [Online]. Available: http://arxiv.org/abs/1606.06650

[50] R. Golan, C. Jacob, and J. Denzinger, "Lung nodule detection in CT images using deep convolutional neural networks," in *2016 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2016, pp. 243–250, iSSN: 2161-4407.

[51] S. Ioffe, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," p. 9, 2015.

[52] C. E. Kahn, C. P. Langlotz, D. S. Channin, and D. L. Rubin, "Informatics in Radiology: An Information Model of the DICOM Standard," *RadioGraphics*, vol. 31, no. 1, pp. 295–304, Jan. 2011, publisher: Radiological Society of North America. [Online]. Available: https://pubs.rsna.org/doi/full/10.1148/rg.311105085

[53] P. Guido, "ITK-SNAP – EMBS," Accessed on 16-02-2021 to https://www.embs.org/pulse/articles/itk-snap/., 2017. [Online]. Available: https://www.embs.org/pulse/articles/itk-snap/

[54] P. A. Yushkevich, A. Pashchinskiy, I. Oguz, S. Mohan, J. E. Schmitt, J. M. Stein, D. Zukić, J. Vicory, M. McCormick, N. Yushkevich, N. Schwartz, Y. Gao, and G. Gerig, "User-Guided Segmentation of Multi-modality Medical Imaging Datasets with ITK-SNAP," *Neuroinformatics*, vol. 17, no. 1, pp. 83–102, Jan. 2019. [Online]. Available: https://doi.org/10.1007/s12021-018-9385-x

[55] Z. Yaniv, B. C. Lowekamp, H. J. Johnson, and R. Beare, "SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and

Reproducible Research," *Journal of Digital Imaging*, vol. 31, no. 3, pp. 290–303, Jun. 2018. [Online]. Available: https://doi.org/10.1007/s10278-017-0037-8

[56] "SimpleITK conventions," Accessed on 16-02-2021 to http://insightsoftwareconsortium.github.io/SimpleITK-Notebooks. [Online]. Available: http://insightsoftwareconsortium.github.io/SimpleITK-Notebooks

[57] Insight Software Consortium, "SimpleITK Documentation," p. 209, 2021.

[58] "What is Python? Executive Summary," accessed on 21/02/2021 to https://www.python.org/doc/essays/blurb/. [Online]. Available: https://www.python.org/doc/essays/blurb/

[59] S. Yegulalp, "What is TensorFlow? The machine learning library explained," Jun. 2019. [Online]. Available: https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html

[60] J. Brownlee, "TensorFlow 2 Tutorial: Get Started in Deep Learning With tf.keras," Dec. 2019. [Online]. Available: https://machinelearningmastery.com/tensorflow-tutorial-deep-learning-with-tf-keras/

[61] R. Campanini, "A deep learning approach to bone segmentation in CT scans," 2017. [Online]. Available: /paper/A-deep-learning-approach-to-bone-segmentation-in-CT-Campanini/2d25f947b482496a4bb093fb7fdeab332f9b1a9c

[62] "What is GitLab?" Accessed on 21-02-2021 to https://about.gitlab.com/what-is-gitlab/. [Online]. Available: https://about.gitlab.com/what-is-gitlab/

[63] "Intensity Windowing — v5.2.0," 2020. [Online]. Available: https://itk.org/ITKExamples/src/Filtering/ImageIntensity/IntensityWindowing/Documentation.html

[64] S.-H. Tsang, "Tutorial: Creating HDF5 Dataset," Jul. 2020. [Online]. Available: https://sh-tsang.medium.com/tutorial-creating-hierarchical-data-format-hdf5-dataset-79cfc99d2613

[65] K. López-Linares Román, I. de La Bruere, J. Onieva, L. Andresen, J. Qvortrup Holsting, F. N. Rahaghi, I. Macía, M. A. González Ballester, and R. San José Estepar, "3D Pulmonary Artery Segmentation from CTA Scans Using Deep Learning with Realistic Data Augmentation," in *Image Analysis for Moving Organ, Breast, and Thoracic Images*, ser. Lecture Notes in Computer Science, D. Stoyanov, Z. Taylor, B. Kainz, G. Maicas, R. R. Beichel, A. Martel, L. Maier-Hein, K. Bhatia, T. Vercauteren, O. Oktay, G. Carneiro, A. P. Bradley, J. Nascimento, H. Min, M. S. Brown, C. Jacobs, B. Lassen-Schmidt, K. Mori, J. Petersen, R. San José Estépar, A. Schmidt-Richberg, and C. Veiga, Eds. Cham: Springer International Publishing, 2018, pp. 225–237.

[66] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," *arXiv:1606.04797 [cs]*, Jun. 2016, arXiv: 1606.04797. [Online]. Available: http://arxiv.org/abs/1606.04797

[67] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation," *arXiv:1611.09326 [cs]*, Oct. 2017, arXiv: 1611.09326. [Online]. Available: http://arxiv.org/abs/1611.09326

[68] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation," *arXiv:1606.02147 [cs]*, Jun. 2016, arXiv: 1606.02147. [Online]. Available: http://arxiv.org/abs/1606.02147

[69] E. Kats, J. Goldberger, and H. Greenspan, "Soft labeling by Distilling Anatomical knowledge for Improved MS Lesion Segmentation," *arXiv:1901.09263 [cs]*, Jan. 2019, arXiv: 1901.09263. [Online]. Available: http://arxiv.org/abs/1901.09263

[70] L. Fidon, W. Li, L. C. Garcia-Peraza-Herrera, J. Ekanayake, N. Kitchen, S. Ourselin, and T. Vercauteren, "Generalised Wasserstein Dice Score for Imbalanced Multi-class Segmentation using Holistic Convolutional Networks," *arXiv:1707.00478 [cs]*, vol. 10670, pp. 64–76, 2018, arXiv: 1707.00478. [Online]. Available: http://arxiv.org/abs/1707.00478

[71] Jordan, J., "An overview of semantic image segmentation." May 2018. [Online]. Available: https://www.jeremyjordan.me/semantic-segmentation/

[72] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations," *arXiv:1707.03237 [cs]*, vol. 10553, pp. 240–248, 2017, arXiv: 1707.03237. [Online]. Available: http://arxiv.org/abs/1707.03237

[73] C. B. Rosselló, "Brain lesion segmentation using Convolutional Neuronal Networks," p. 50, 2018.

[74] Wood, T., "Softmax Function," May 2019. [Online]. Available: https://deepai.org/machine-learning-glossary-and-terms/softmax-layer

[75] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool," *BMC Medical Imaging*, vol. 15, no. 1, p. 29, Dec. 2015. [Online]. Available: http://bmcmedimaging.biomedcentral.com/articles/10.1186/s12880-015-0068-x

[76] M. Shahedi, D. Cool, C. Romagnoli, G. Bauman, M. Bastian-Jordan, G. Rodrigues, B. Ahmad, M. Lock, A. Fenster, and A. Ward, "Postediting prostate magnetic resonance imaging segmentation consistency and operator time using manual and computer-assisted segmentation: Multiobserver study," *Journal of medical imaging (Bellingham, Wash.)*, vol. 3, Oct. 2016.

[77] E. G. Associate and E. D. Gallagher, *Compah Documentation*, 1999.

[78] D. Karimi and S. E. Salcudean, "Reducing the Hausdorff Distance in Medical Image Segmentation with Convolutional Neural Networks," *arXiv:1904.10030 [cs, eess, stat]*, Apr. 2019, arXiv: 1904.10030. [Online]. Available: http://arxiv.org/abs/1904.10030

[79] Prabhu, "Understanding Hyperparameters and its Optimisation techniques," Jul. 2018. [Online]. Available: https://towardsdatascience.com/understanding-hyperparameters-and-its-optimisation-techniques-f0debba07568

[80] J. Brownlee, "Understand the Impact of Learning Rate on Neural Network Performance," Jan.

2019. [Online]. Available: https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/

[81] Brownlee, J., "How to use Learning Curves to Diagnose Machine Learning Model Performance." 2019. [Online]. Available: https://machinelearningmastery.com/learning-curves-for-diagnosing-machine-learning-model-performance/

[82] K. Shen, "Effect of batch size on training dynamics," Jun. 2018. [Online]. Available: https://medium.com/mini-distill/effect-of-batch-size-on-training-dynamics-21c14f7a716e

[83] J. Czajkowska, C. Feinen, M. Grzegorzek, M. Raspe, and R. Wickenhöfer, "A New Aortic Aneurysm CT Series Registration Algorithm," vol. 283, Jun. 2014, pp. 15–26.

[84] S. Xie, H. Xu, and H. Li, "Artifact Removal Using GAN Network for Limited-Angle CT Reconstruction," in *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, Nov. 2019, pp. 1–4, iSSN: 2154-512X.