UNIVERSITAT POLITÈCNICA DE VALÈNCIA

ESCUELA TÉCNICA SUPERIOR INGENIERÍA INDUSTRIAL VALENCIA

TRABAJO FIN DE GRADO EN INGENIERÍA BIOMÉDICA

DESIGN AND DEVELOPMENT OF A SYSTEM FOR SEMANTIC SEGMENTATION OF ATRIAL CAVITY WITH A NEURAL NETWORK ARCHITECTURE ENCODER-DECODER.

AUTHOR: MARTA SAIZ VIVÓ

SUPERVISOR: VALERY NARANJO ORNEDO

SUPERVISOR: ADRIAN COLOMER GRANERO

Academic year: 2019-20

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

ESCUELA TÉCNICA SUPERIOR INGENIEROS INDUSTRIALES VALENCIA

Structure

I. Technical Report
II. Budget

Technical Report

Design and development of a system for semantic segmentation of atrial cavity with a neural network architecture encoder-decoder.

Author: Marta Saiz Vivó

Directors: Valery Naranjo Ornedo

Adrian Colomer Granero

July, 2020

Acknowledgements

I would like to express my gratitude to Dr. Valery Naranjo for introducing me to the exciting world of image analysis. I would also like to thank Dr. Adrian Colomer whose expert guidance throughout the project has been of great help for me. In general, I would like to thank the CVBLab team for welcoming me in their laboratory.

A special thank you to my university friends who have made my years spent at the university truly unforgettable. Thank you for brightening my days.

Finally, I would especially like to thank my family whose unwavering love and support has led me through any challenge encountered. I am eternally grateful; this project is dedicated to you.

Abstract

Atrial Fibrillation (AF) is a supraventricular arrythmia characterised by a chaotic and disorganised electrical activity which leads to an irregular contraction. AF is the most common cardiac arrythmia and its prevalence increases with the ageing population. It is estimated that there are currently 4.5 million cases in Europe. Currently, cardiac ablation is the main treatment procedure for AF treatment. To guide and plan this procedure it is essential for clinicians to obtain 3D anatomical reconstruction models of the atria which are patient specific. In clinical practice the atria is manually delineated which is a labour-intensive method. The aim of this project is to develop automatic algorithms with employment of Deep Learning (DL) techniques for left and right atrium segmentation from MRI volumetric images.

In this project neural networks have been used with encoder-decoder architecture to provide accurate pixel-wise segmentation of the atria from 3D MRI images. Whilst most studies in this area concentrate their efforts in left atrium (LA) segmentation the aim of this project was to obtain a segmentation model also capable of accurate right atrium (RA) segmentation from few annotated data. For this purpose, the network was trained with 2 databases. The first database consisted of 100 volumetric MRI images with LA ground-truth segmentations and was used to obtain a high-performance model for posterior deep fine-tuning techniques. The second database consisted of 19 volumetric MRI images of high variability with ground-truth segmentation of both atria cavities and was used to verify the segmentation accuracy of the network for LA and RA from few annotated samples. Furthermore, the high generalisation capacity obtained by the model trained with the second database was verified with an external database of only LA ground-truth. To compare the predicted segmentation with the labelled data a similarity coefficient was computed from them and was compared with the existing literature. In addition, volumetric visualization and reconstruction libraries were employed to obtain the anatomy of the atria cavity.

The results obtained are very promising. The second database model was capable of accurately segmenting the RA with a dice coefficient of 0.9160 when fine tuning techniques are implemented. Furthermore, the model created with database 2 demonstrated a very high generalisation capacity when tested with the third database obtaining a dice coefficient of 0.8515 for LA segmentation.

Keywords: Atrial Geometry; Semantic Segmentation; Deep Learning; Convolutional Neural Networks; Magnetic Resonance Imaging.

Resumen

La fibrilación auricular (FA) es una arritmia supraventricular caracterizada por una actividad eléctrica caótica y desorganizada que conduce a una contracción irregular. La FA es la arritmia cardíaca más común y su prevalencia aumenta con el envejecimiento de la población. Actualmente, la ablación cardíaca es el principal tratamiento de la FA. Para guiar y planificar este tratamiento es esencial que los médicos tengan acceso a modelos 3D de la anatomía auricular que sean específicos del paciente. Para ello, en la práctica clínica se segmentan manualmente las aurículas lo que requiere una intensiva labor. El objetivo de este proyecto es desarrollar algoritmos automáticos para la segmentación de la aurícula izquierda y derecha, a partir de imágenes volumétricas de Resonancia Magnética (RM), mediante el empleo de técnicas de *Deep Learning*.

En este proyecto se han utilizado redes neuronales con la arquitectura *encoder-decoder* para obtener una segmentación precisa, píxel a píxel, de la aurícula a partir de imágenes 3D de RM. Mientras que la mayoría de los estudios en esta área concentran sus esfuerzos en la segmentación de la aurícula izquierda (AI), el objetivo de este proyecto ha sido obtener un modelo de segmentación capaz también de segmentar la aurícula derecha (AD) a partir de pocas muestras anotadas. Con este fin, se entrenó a la red con 2 bases de datos. La primera base de datos está compuesta de 100 imágenes volumétricas de RM con segmentaciones manuales o *ground-truth* de AI y se utilizó para obtener un modelo de alto rendimiento para posterior aplicación de técnicas *deep fine-tuning*. La segunda base de datos está compuesta de 19 imágenes volumétricas de RM de alta variabilidad con segmentaciones anotadas de ambas cavidades auriculares y se utilizó para verificar la precisión de segmentación de la red para AI y AD a partir de pocas muestras anotadas. Además, la alta capacidad de generalización obtenida por el modelo formado con la segunda base de datos se verificó con una base de datos externa de sólo *ground-truth* de AI (tercera base de datos). Para comparar la segmentación predicha con los datos etiquetados se calculó un coeficiente de similitud, también denominado *Dice coefficent* que se comparó con la literatura existente. Además, se emplearon librerías de visualización volumétrica y reconstrucción para obtener la anatomía de la cavidad auricular.

Los resultados obtenidos son muy prometedores. El modelo desarrollado fue capaz de segmentar con precisión la AD con un *Dice coefficient* de 0,9160 cuando se implementan técnicas de *deep fine-tuning*. Además, el modelo obtenido con la base de datos 2 demostró una capacidad de generalización muy alta cuando se evaluó con la tercera base de datos obteniendo un *Dice coefficient* de 0,8515 para la segmentación de AI.

Palabras clave: Geometría Auricular; Segmentación Semántica, *Deep Learning*; Redes Neuronales Convolucionales; Imágenes de Resonancia Magnética.

Resum

La fibril·lació auricular (FA) és una arrítmia supraventricular caracteritzada per una activitat elèctrica caòtica i desorganitzada que condueix a una contracció irregular. La FA és l'arrítmia cardíaca més comuna i la seua prevalença augmenta amb l'envelliment de la població. Actualment, l'ablació cardíaca és el principal tractament de la FA. Per a guiar i planificar aquest tractament és essencial que els metges tinguen accés a models 3D de l'anatomia auricular que siguen específics del pacient. Per a això, en la pràctica clínica se segmenten manualment les aurícules el que requereix una intensiva labor. L'objectiu d'aquest projecte és desenvolupar algorismes automàtics per a la segmentació de l'aurícula esquerra i dreta, a partir d'imatges volumètriques de Ressonància Magnètica (RM), mitjançant l'ús de tècniques de *Deep Learning*.

En aquest projecte s'han utilitzat xarxes neuronals amb l'arquitectura *encoder-decoder* per a obtindre una segmentació precisa, píxel a píxel, de l'aurícula a partir d'imatges 3D de RM. Mentre que la majoria dels estudis en aquesta àrea concentren els seus esforços en la segmentació de l'aurícula esquerra (AE), l'objectiu d'aquest projecte ha sigut obtindre un model de segmentació capaç també de segmentar l'aurícula dreta (AD) a partir de poques mostres anotades. A aquest efecte, es va entrenar a la xarxa amb 2 bases de dades. La primera base de dades està composta de 100 imatges volumètriques de RM amb segmentacions manuals o *ground-truth* d'AE i es va utilitzar per a obtindre un model d'alt rendiment per a posterior aplicació de tècniques *deep fine-tuning*. La segona base de dades està composta de 19 imatges volumètriques de RM d'alta variabilitat amb segmentacions anotades de totes dues cavitats auriculars i es va utilitzar per a verificar la precisió de segmentació de la xarxa per a AE i AD a partir de poques mostres anotades. A més, l'alta capacitat de generalització obtinguda pel model format amb la segona base de dades es va verificar amb una base de dades externa de només *ground-truth* d'AE (tercera base de dades). Per a comparar la segmentació predita amb les dades etiquetades es va calcular un coeficient de similitud, també denominat *Dice coefficent* que es va comparar amb la literatura existent. A més, es van emprar llibreries de visualització volumètrica i reconstrucció per a obtindre l'anatomia de la cavitat auricular.

Els resultats obtinguts són molt prometedors. El model desenvolupat va ser capaç de segmentar amb precisió l'AD amb un *Dice coefficient* de 0,9160 quan s'implementen tècniques de *deep fine-tuning*. A més, el model obtingut amb la base de dades 2 va demostrar una capacitat de generalització molt alta quan es va avaluar amb la tercera base de dades obtenint un *Dice coefficient* de 0,8515 per a la segmentació d'AE.

Paraules clau: Geometria Auricular; Segmentació Semàntica, *Deep Learning*; Xarxes Neuronals Convolucionals; Imatges de Ressonància Magnètica.

Contents

List of Figures

List of Tables

Chapter 1. Motivation and Problem Statement

Atrial Fibrillation (AF) is the most common sustained cardiac arrythmia and has become one of the most important public health problems in the last 20 years [1]. AF affects approximately 1-1.5% of the general population and its prevalence is projected to rapidly increase due to the ageing population[2]. As a result, the number of cases estimated in 2030 ascends to 14-17 million in Europe[1]. AF is a cardiac pathology highly associated with morbidity and mortality factors such as heart failure, ischaemic and haemorrhagic strokes. Moreover, it is heavily associated with an increase in public health expenses due to the increase in healthcare resource consumption by the public [2].

At this moment, the most common method for treating AF patients is radiofrequency catheter ablation to produce scars and electrically isolate the pulmonary veins [3]. Due to the large variation in the pulmonary veins connection pattern to the left atrium (LA) there is a need for LA models in order to transfer the generic ablation strategy to a specific patient's anatomy. Furthermore, the viability assessment of the myocardium after these procedures is of utmost importance to quantify the patients scar burden and location and increase the success rate of these therapies [4].

Therefore, accurate segmentation of LA is highly desirable for patient specific scar characterization to select the appropriate ablation strategy. Furthermore, to plan the procedure 3D geometrical models of both the left and right atrium are required [5]. These 3D patient-specific models of the atria are obtained from the segmentation of MRI and CT images. In clinical routine this segmentation is normally performed manually by experts. However manual delimitation is a time-consuming, labour-intensive and error-prone method. The inter and intra variability in the observations is an important source of error [6]. In addition, the amount of slice annotations that have to be performed for each volumetric segmentation renders this method impractical.

Furthermore, the automatic segmentation of these structures, specially the LA remains a challenging task due to the complex geometry of the LA and the morphological variations between patients [6]. In addition, although MRI acquisition techniques may prove useful due to their high contrast to noise ration and the possibility of scar quantification from contrast enhancement methods (LGE-MRI), they also provide additional challenges such as the low imaging resolution [6].

An emerging machine learning technique known as deep learning (DL) is gaining popularity in the field of medical image segmentation and has shown promising results for LA and RA segmentation. DL algorithms, referred to as neural networks, avoid limitations of traditional machine learning methods by means of supervised learning algorithms that are capable of self-learning features of the image for a more accurate segmentation.

On the other hand, the use of these algorithms in routine clinical practice remains a challenge due to their high computational cost, their difficulties for generalisation and the fact that they require large amounts of segmented data, otherwise known as ground-truth, to be trained.

Therefore, in this project the aim is to develop DL segmentation algorithms capable of segmenting both LA and RA cardiac structures with increased accuracy and generalisation capacity.

Chapter 2. State of the Art

Deep learning (DL) methods consist in deep artificial neural networks that automatically extract discriminant features from input data for object detection and segmentation among other tasks. In the past decade, DL techniques, in particular Convolutional Neural Networks (CNN), have achieved great progress in computer vision tasks and have become the methodology of choice for medical image segmentation and classification [7].

CNNs consist of an input layer, an output layer and a stack of functional layers in between known as convolutional layers and pooling layers tasked with the feature extraction of the input data. These layers will be more thoroughly detailed in the following section. The output of the network is a fixed size vector where each element corresponds to a probabilistic score of each category (e.g. for image classification). CNNs were first introduced for medical image segmentation by Ciresan et al. for electron microscopy (EM) image segmentation [8]. However, this method required to divide the images into patches and train the CNN separately to predict the class for every centre pixel. This patch-based approach has a lot of redundancy due to the overlapping patches in the image and its highly inefficient. As a result, CNNs are mainly used for object localization to estimate the bounding box of an object and extract the region of interest. The bounding box is then cropped from the image, forming a pre-processing step to reduce computational cost of segmentation [9]. For efficient, pixel wise segmentation a variant of CNN, known as fully convolutional neural network (FCN) is more commonly used.

FCNs were first introduced by Long et al. for image segmentation [10]. FCNs are a type of CNN that do not have fully connected layers, in other words, the output of the network is an image (2D/3D) instead of a 1-dimensional vector. FCN have a structure known as encoder-decoder which takes input of an arbitrary size and produce an output of the same size. Given an input image the encoder or down-sampling path extracts the high features of the image through a series of convolution and pooling operations and then the symmetrical up-sampling path or decoder interprets the feature maps and recovers the spatial information producing a pixel wise prediction as an output. The precise layers and operations found in the encoder-decoder structure will be explained in more detail in the next section.

Contrary to CNN, FCN can be applied to entire images thus removing the need for patch selection [10]. On the other hand, FCN with simple encoder-decoder structure may be limited in capturing a precise image segmentation as some features may be eliminated during the pooling process in the encoder. Several FCN variants have been proposed that propagate features from encoder to decoder thus boosting segmentation accuracy, the most widely known variant for biomedical image segmentation is the U-net, first proposed by Ronneberger et al. [11].

The U-net recovers spatial context loss in the down-sampling path by employing skip connections between encoder and decoder, therefore yielding a more precise segmentation. Traditional 2D U-Nets have achieved good results in the field of medical image segmentation [11, 12]. However, as the convolution is performed in each 2D slice the spatial relationship between slices cannot be captured. On the other hand, the 3D extension of the U-Net, first proposed by Çiçek et al. [13], by expanding the filter operator into 3D space extracts image features in 3D and therefore considers the spatial continuity between the slices of the image.

Furthermore, Millietari et al. [14] proposed the 3D V-Net for volumetric segmentation wherein a novel loss function known as Dice coefficient was introduced and the learning of a residual function improved the training convergence.

Several state-of-the-art cardiac image segmentation methods that include the U-net or its 3D variants, the 3D U-Net and 3D V-Net, have achieved promising segmentation accuracy in different cardiac segmentation tasks [15-17] .

In the field of automated MRI atrial segmentation several traditional non-DL methods such as region growing [18], atlas-based label fusion [19] and non-rigid registration [20] have been applied in the past. However, these methods rely highly on good initialization and ad-hoc pre-processing methods which limits their potential adoption in the clinic.

More recently, Bai et al. [21] applied 2D FCN to segment LA and RA from 2D long axis images. Likewise, other groups have also applied 2D FCNs to segment the left atrium from 3D LGE-MRI images in a slice by slice fashion with feature learning enhancement [22, 23]. Furthermore, Mortazi et al. proposed a variant of the FCN known as multi-view 2D FCN where the network is trained with different planar views of cardiac MRI images for a more robust segmentation [24].

3D neural networks enable the full use of spatial information in 3D images and their feature extraction may better reflect the shape features of the anatomy in cardiac segmentation [25]. Several variants of the 3D U-Net have been proposed for 3D LGE-MRI atrial segmentation achieving very promising results. Vesal et al. [26] proposed a 3D U-Net with dilated convolutions in the lowest level of the network to extract features spanning a wider spatial range. Li et al. [27] proposed a 3D U-Net with hierarchical aggregation to obtain better spatial fusion information.

The network proposed by Jia et al. [25] for 3D LGE-MRI LA segmentation which will be employed in this project consist of a two-stage 3D U-Net. The ‘first’ U-Net is used to coarsely segment and locate the left atrium, based on this output the ‘second’ U-Net accurately segments the left atrium under higher resolution after the region of interest has been cropped.

Chapter 3. Objectives

The main objective of this end of degree project is to develop automatic algorithms based on CNNs for the segmentation of LA and RA cardiac structures with increased accuracy and generalisation capacity for a future application in routine clinical practice. For this, a series of specific objectives have to be met first:

1. To obtain LA and RA ground-truth databases of MRI images with high inter patient variability
2. To review the state of the art of neural networks for cardiac segmentation and select the appropriate network architecture.
3. To pre-process the images and ground-truth masks of these databases to select the ground-truth segmentation of interest: LA or RA
4. To train the network with a high variability database to increase its generalisation capacity and test it with an external database.
5. To use deep fine-tuning techniques to increase the segmentation accuracy of the network
6. To post process the output predictions of the network to integrate LA and RA.
7. To design a 3D reconstruction algorithm for a visual assessment of atrial geometry and use 3D morphological filters to improve the reconstruction
8. To review the existing segmentation methods in the literature and perform a comparison of the results obtained.
9. To identify and describe limitations encountered during the course of this project and propose future lines of work.

Chapter 4. Theoretical Framework

4.1 Heart Anatomy

The cardiovascular system is fundamentally composed by the heart and the blood vessels such as veins, capillaries and arteries. The function of the cardiovascular system is to deliver oxygen and other nutrients transported by blood to body tissues and to remove waste products such as carbon dioxide. The heart is the main organ of the cardiovascular system, it is a muscular organ which by its contraction pumps blood through veins and arteries to the whole organism.

Anatomically, the heart is located within the thoracic cavity, surrounded by both lungs medially and the diaphragm below. It is separated from other organs by a membrane known as the pericardium which surrounds the heart and conforms the pericardial cavity filled with serous liquid. As illustrated in ***Figure 1***, the upper surface of the heart, known as the base, is where the vena cavae, the aorta, the pulmonary trunk and the great arteries and veins are attached. The inferior tip part of the heart is known as the apex [28].

The internal cavity of the heart is divided in 4 chambers: the Left Atrium (LA), the Right Atrium (RA) located in the upper section of the heart, and the Left Ventricle (LV) and the Right Ventricle (RA) located in the lower section. The chambers are divided from each other by the septa.

Each atrium is connected to their corresponding ventricle through an atrioventricular valve that controls the blood flow and ensures a unidirectional flow. The LA is connected to the LV through the Mitral Valve (MV) which has a bicuspid structure and the RA is connected to the RV through the Tricuspid Valve (TV).

Furthermore, the LV is connected to the aorta through the aortic valve and the RV is connected to the pulmonary trunk through the pulmonary valve. These are known as the semilunar valves and they are bicuspid by nature.

Both the atrioventricular and semilunar valves are surrounded by a fibrotic ring known as the cardiac skeleton which provides mechanical stability to the heart during contraction and provides electrical isolation between the atria and ventricle [28].

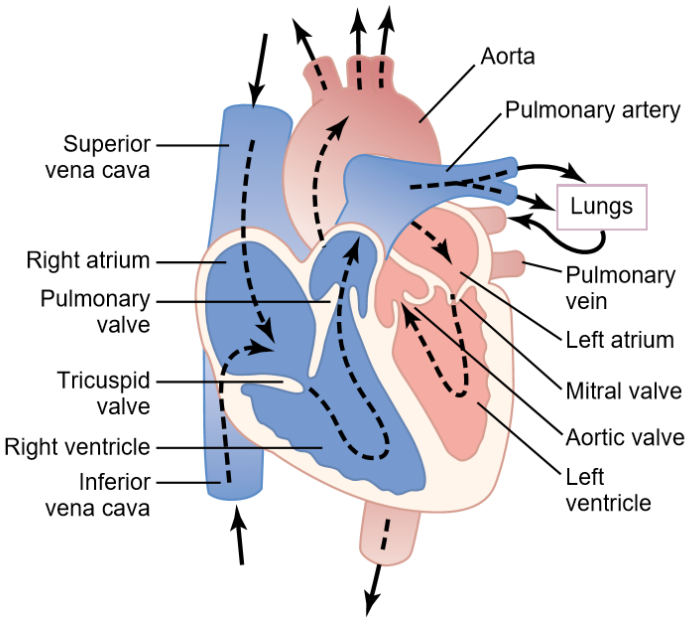


Figure 1 Heart structure and circulatory movement of blood through cardiac cavities [28].

Unlike the atrioventricular septum, the interatrial septum and the interventricular septum that divides both atria and both ventricles from each other have no openings, ensuring that there is no blood communication between both chambers [28].

All the structures mentioned above have a very specific function in the cardiovascular system. The LA and RA act as receiving chambers that contract and push the blood to the lower chambers (LV and RV). The LV and RV are the primary pump chambers that eject the blood from the heart towards the lungs or body, respectively [28].

The composition of the heart wall consists of 3 layers: the epicardium, the myocardium, and the endocardium. The innermost layer is known as the endocardium and the external layer is known as the epicardium or visceral pericardium. The middle and thickest layer, known as the myocardium, is a layer composed of collagenous fibres and muscle cells which allows the heart to contract. It surrounds both ventricles and atria however its thickness varies from one cavity to another. For example, the myocardium thickness around the LV is much greater than around the RV. This occurs because the systemic circuit is englobing greater number of organs than the pulmonary circuit, so the LV needs to contract with a higher driving force to surpass this higher resistance. The myocardial wall of both atria is much thinner than those of the ventricles as their contraction is needed to pump blood into the ventricles, not the external tissues of the heart. The myocardial walls of both atria are thinner compared to those of the ventricles as the necessary force needed to pump blood into the ventricles is much less compared to the one needed for blood to reach external tissues of the heart [28].

4.2 Atrial Anatomy

The geometry of the RA and LA varies in shape, however both atria are composed by the same basic structures: an appendage, an atrioventricular valve, a venous component, and the interatrial septum separating atria chambers from each other. As previously stated, the interatrial septum provides

electrical isolation between the atria. Surrounding the septum, the atrial walls are covered by conductive myocardium except for the region of the Fossa Ovalis (FO) located in the right atrial side of the interatrial septum where only peripherical ring is conductive. The Right and Left Atrial Appendages (RAA and LAA) are appendages that prolong to the anterior surface of the cardiac muscle. They are variable in shape and prominence between both atria. Whilst the RAA has a triangular shape, is less prominent and it is located in the posterior region of the atria, the LAA protrudes from the main body of the LA with a tubular shape (see ***Figure 2***).

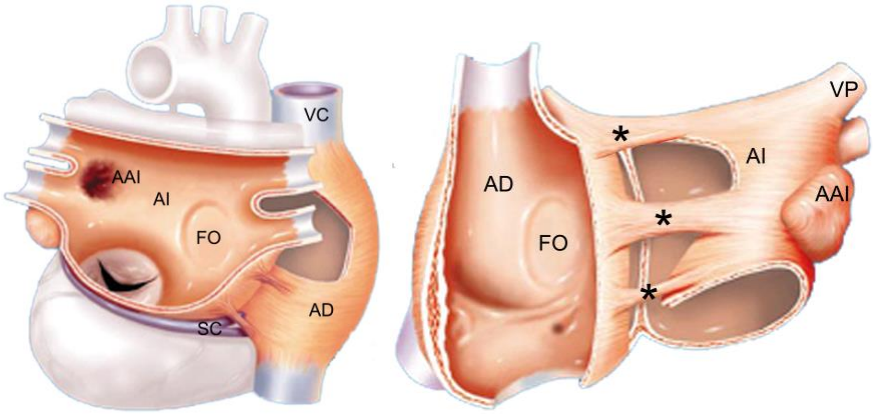


Figure 2 Coronal view of the atria. Left Atrium (AI) and Right Atrium (AD). The Fossa Ovalis (FO), Left Atrial Appendage (AAI), coronary sinus (SC), superior and inferior vena cava (VC) and pulmonary veins (VP) can be seen. Modified from [29].

The RA is a muscular structure composed of the RAA, the intercaval area, the vestibule surrounding the TV, the septum, and the atrioventricular orifice where the TV is located. The intercaval area forms the continuation between the superior and inferior orifice of the Superior Vena Cava (SVC) and the Inferior Vena Cava (IVC) located in the antero-superior region of the atrium. Between the atrioventricular orifice and the IVC orifice the Coronary Sinus (CS) is located. The CS deposits the venous blood from the coronary system in the RA, however it is also connected to the wall of the LA through muscular connections. The SVC, IVC and CS form the venous region of the RA [29].

The Crista Terminalis (CT) is the largest muscle bundle in the RA. As can be seen in ***Figure 3***, it is located in the internal surface of the atria, along the intercaval area. The PM spreads from the CT through the wall of the RAA. Near the CT the SinoAtrial Node (SAN) is located which is the natural pacemaker of the heart. Between the IVC and the TV the cavo-tricuspid isthmus is found. It is a region of slower electrical conductivity and a target region for atrial ablation due its relation to macroreentries in atrial flutter [30].

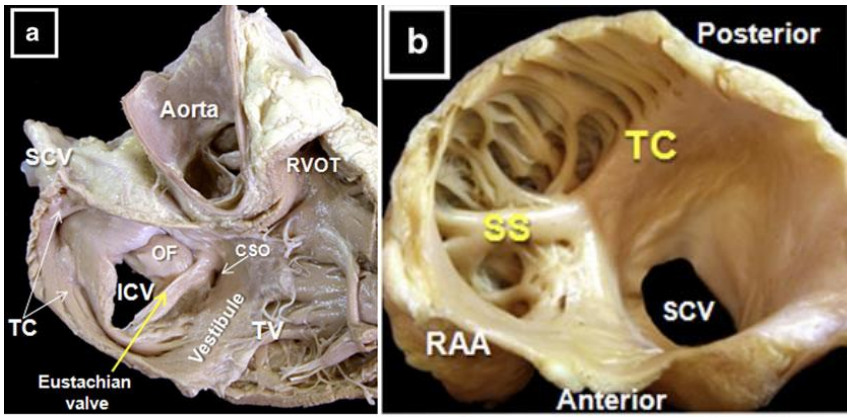


Figure 3 Right atrium views. TC, terminal crest; SCV, superior caval vein; ICV, inferior caval vein; OF, oval fossa; CSO, coronary sinus orifice; TV, tricuspid valve; SCV, superior caval vein [30].

As observed in Figure 4 the LA has a similar structure to the RA and it is composed by the LAA, the venous region that comprises the Pulmonary Veins (PV), the vestibule surrounding the orifice where the MV is found and the septum. The orifices corresponding to the 4 PV are found in the postero-superior region, the two corresponding to the LPV (superior LSPV and inferior LIPV) located laterally in the external region of the LA and the other two corresponding to the RPV (superior RSPV and inferior RIPV) located near the interatrial septum. In the inferior region of the LA, similarly to the RA, we find a great orifice where the MV is located. Between the MV and the left inferior Pulmonar Vein (LIPV) the mitral isthmus is located which can also be a target region for ablation in patients with Atrial Fibrillation (AF) [31].

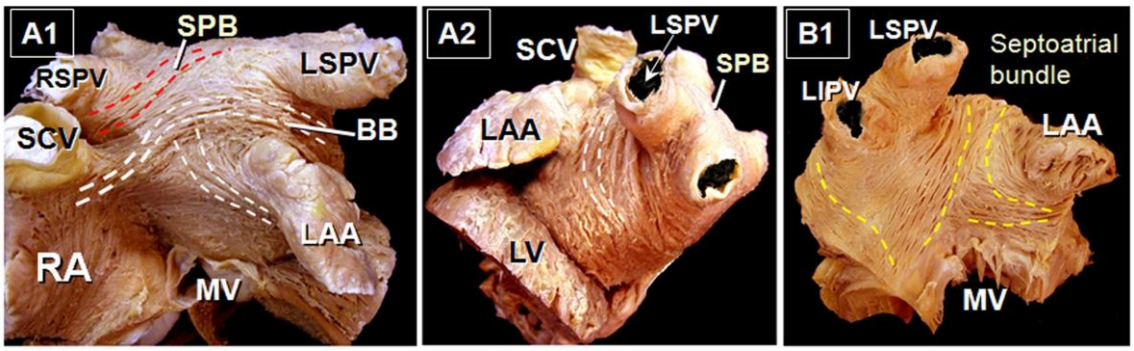


Figure 4 Left atrium views. SCV, superior cava vein; RSPV, right superior pulmonary vein; LSPV, left superior pulmonary vein; BB, Bachmann bundle; LAA, left atrial appendage; MV, mitral valve annulus, SPB, septopulmonary bundle [31] .

The thickness of the internal cavity wall of the atria depends on the presence of muscle fibres. In the most distant region to the septum is where a greater number of muscle fibres is found therefore where wall is thicker. In the RA these fibres are disposed irregularly though parallel to the longitudinal direction of the heart forming the Pectinate Muscle (PM). On the other hand, in the LA the internal cavity wall is smooth, and the PM is exclusively located in the LAA.

The thickness of the atria wall varies greatly between different parts of it. Thereby, the atrial septum and anterior LA are the thickest areas whilst the vena cava areas are the thinnest. The atria present a thickness in the range of 0.4 to 11.7mm [32].

4.3 Heart Electrophysiology

The contraction of the heart is a fundamental action required to pump the blood through the organism. The heart contracts at a regular rhythm known as the heartrate (HR) which in physiological and rest conditions is approximately 60 bpm (heartbeats per minute). The cardiac cycle occurs between one heartbeat and the other and it is comprised of 3 periods: atrial systole, ventricular systole and diastole. Systole is a period where the atria or ventricle contracts and ejects the blood inside their cavity, diastole corresponds to the atria or ventricle relaxation where they receive the influx of blood.

For the myocardial cells to contract they must be excited by an electrical stimulus. This stimulus is autonomously generated by the heart, specifically by a group of cells that have electrical automaticity, known as pacemaker cells. These pacemaker cells can be found in different regions of the heart, both in the atria and ventricles, with varying degrees of automaticity. Specifically, they can be found in the structures known as the sinus node (SA) node, the atrioventricular (AV) node and the Purkinje fibres. These structures along with others that have faster electrical conductivity than the normal myocardial tissue comprise what is known as the cardiac conduction system (CCS). The CCS is a pathway that allows the propagation of the electrical impulse through the 4 chambers of the heart in a synchronised manner which in return allows for a synchronised contraction. The CCS is illustrated in ***Figure 5*** [33].

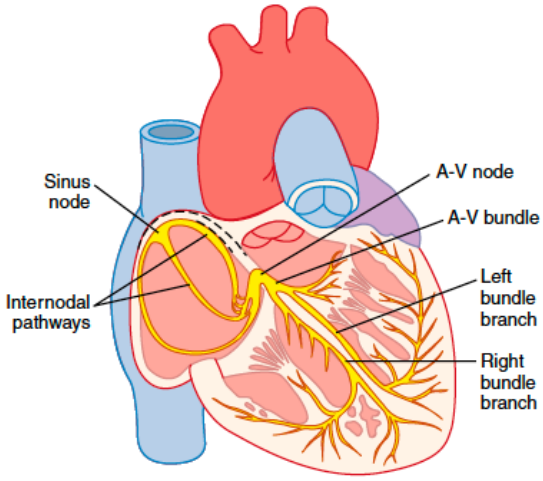


Figure 5 Specialized excitatory and conduction system of the heart [33].

In a normal heartbeat the electrical impulse is first generated automatically in the (SA) node which is located in the junction of the RA with the SVC. From there it propagates through a series of muscle bundles with preferential conductivity at the atrial level. The electrical impulse propagates faster along these bundles than the rest of the atrial myocardium. These bundles are the Bachmann´s Bundle (BB), found in the LA that connects the LA and RA, the crista terminalis (CT) and the pectinae muscles (PM) both found in the RA [33].

At the ventricular level, the AV node acts as the second pacemaker of the heart, after the SA node, as it has a lower degree of automaticity. From the AV the electrical impulse propagates to the structure known as the bundle of His, found in the base of the ventricle. The bundle of His then divides into right and left pathways to provide electrical stimulation to the right and left ventricle, respectively. When

they reach the apex, they divide into a finer fibre network that surrounds the ventricles. It is known as the Purkinje Fibres [33].

Therefore, the electrical impulse first generated in the SA travels through the atria via the BB, CT and PM until it reaches the AV node from where it propagates to the ventricles through the His bundle and the Purkinje Fibres. ***Figure 6*** illustrates this electrical propagation of electrical activity in the heart [34].

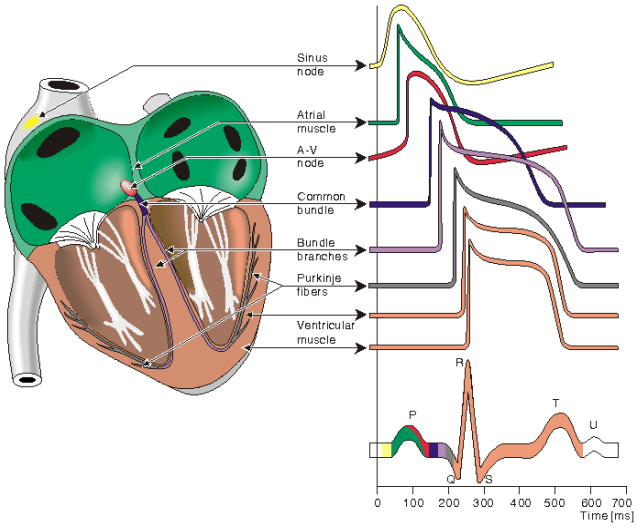


Figure 6 Electrophysiology of the heart. The different action potential for each of the specialized cells of the heart are shown [34].

4.4 Atrial Arrhythmias

Lesh et al. [35] grouped the atrial tachycardia, taking into account the electrophysiological mechanisms and anatomical substrate, in focal atrial tachycardia, macroreentrant tachycardia and atrial fibrillation.

Focal Atrial Tachycardias

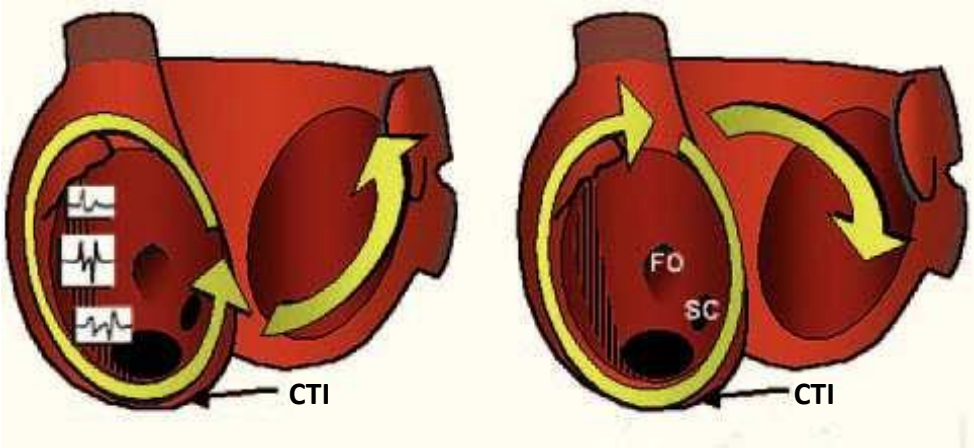
Focal Atrial Tachycardias are defined as non-physiological, regular, fast-paced atrial contractions at a rate higher than 100 bpm with the onset and maintenance limited to the atria. Ectopic foci, usually in the CT or the PVs, or an anatomical or functional reentrant current are usually the onset mechanisms for AT. They can be classified depending on their onset as: 1) AT with onset near the CT, 2) with onset near to the atrioventricular valves, 3) Septal tachycardias, 4) Left Atrial Tachycardia with onset near the PVs and 5) AT with other onsets [35].

Macroreentrant Tachycardias

These tachycardias depend on macroreentries to occur due to functional or intentional blocks in the atria. Atrial Flutter is by-far, the most common type of Macroreentrant Tachycardias, affecting both healthy and diseased hearts. Due to its high recurrence and refractoriness to pharmacological strategies, its clinical treatment is complex. Atrial flutter is defined as a regular AT characterized with having a mainly constant atrial cycle with frequencies between 250 and 300 bpm and with an atrioventricular conduction block of usually 2 to 1.

Animal and human studies have shown that flutter is an arrhythmia based on a reentry mechanism which depends on the conduction of the RA and has a zone of protected and forced conduction through the cavotricuspid isthmus (CTI). The CTI is delimited by the inferior vena cava posteriorly, by the tricuspid ring anteriorly, by the coronary sinus ostium and the border of the eustachian valve medially and by the CT laterally. The reentry mechanism in most of the flutter cases is anti-clockwise (Figure 7a). However, the electrical impulse can also follow the same path but in the inverse direction (clockwise): ascending through the lateral side of the RA and descending by the interatrial septum (see Figure 7b) [37].

Other reentrant arrythmias in the LA, with pathways that don't include the isthmus such as those around the coronary sinus or with other foci, are classified as atypical flutter [36].



(a) (b)

Figure 7 Characteristic macroreentry circuits in the RA observed in flutter. (a) typical flutter, (b) inverse typical flutter. The Cavotricuspid Isthmus (CTI) is labeled. Modified from [37].

Atrial Fibrillation

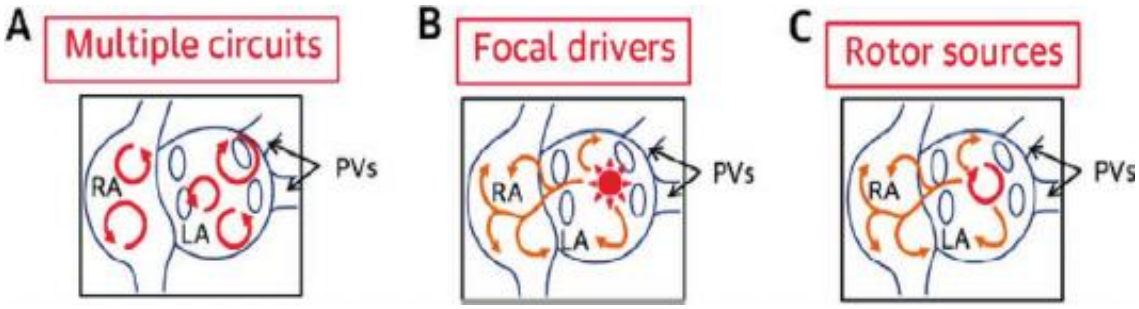
The most common type of atrial tachycardia is atrial fibrillation (AF) which affects around 2% of the general population, 10-17% of those being older than 80 years[1]. AF is characterized by having a fast, desynchronized and chaotic electrical activity unable to contract the atria effectively. The desynchronized electrical activity propagates in waves that rotate through both atria causing them to quiver or fibrillate at frequencies of 300 to 600 bpm.

Hypothesis on the mechanisms of atrial fibrillation.

In 1962, Moe, G. K. proposed the "multiple wavelet hypothesis"[38], which states that AF is generated by many simultaneous wavefronts propagating in a chaotic manner through the atria (***Figure 8***a). AF will be maintained as long as the number of wavefronts is high enough that the probability of every wave to terminate at the same time is low.

This hypothesis was challenged by Haissaguerre, M. et al., proposing the "focal hypothesis"[39], in which AF would be initiated and maintained by one or several high frequency foci originating in or around the PVs (***Figure 8***b). These activation fronts would fractionate and disorganize the neighboring tissues creating what is known as fibrillatory conduction.

In 2003, Jalife, J. proposed the "rotor hypothesis"[40] suggesting that the onset of AF is caused by a combination of ectopic beats originating mainly in the PVs whose wavefronts would fragment when reaching the curvature of the venoatrial junction generating 2 vortices rotating in opposite directions. Finally, one of these vortices would stabilize in the posterior wall creating a functional reentry or rotor, which would maintain the AF by activating the local tissue in a chaotic, high frequency manner thus generating wavefronts that would fragment and propagate through the atria (***Figure 8***c). According to Jalife, the PVs and the posterior wall play a fundamental role in the onset and maintenance of fibrillatory activity.



(a) (b) (c)

Figure 8 AF maintenance mechanisms: (a) multiple wavelet hypothesis, (b) focal hypothesis, (c) rotor hypothesis. Adapted from [41].

Experimental studies have found PV reentries as a possible onset of AF. Electrophysiological studies in the PVs have shown the presence of all the "ingredients" required to maintain the reentry circuits. The myocytes in the PVs have a transmembrane action potential with a lower velocity and duration in depolarization phase, which translates to a shortening of the early repolarization (ERP) and a slowing of the conduction velocity [42]. Furthermore, sudden changes in the disposition of the muscle fibers and their intertwinement in the venoatrial junction give rise to slower and discontinuous conduction [36]. Therefore, PVs constitute the ideal substrate

Atrial fibrillation can be classified as: paroxysmal, persistent and permanent (long-standing persistent). Paroxysmal AF is defined as recurrent episodes with spontaneous interruption generally within 7 days but mostly between 24-48 hours. Persistent AF is an arrhythmia that does not interrupt spontaneously but does with cardioversion, process by which the AF is converted back into sinus rhythm, or with pharmacological treatment. Permanent AF does not respond to treatment.

Around 25% of the diagnosed AF cases are paroxysmal and usually affects younger patients with fewer comorbidities than persistent AF patients. Approximately 20% of paroxysmal AF patients will progress into permanent AF in a period of 4 years as a result of atrial remodeling.

Atria are anatomical structures malleable to the different physio-pathological mechanisms that can experience. Studies carried out by Wijffels et al. in goats [43] showed that the longer the AF episodes the easier they were to induce and the longer they would become in the future (Figure 9). This lengthening in the episodes was associated with a shortening of the atrial fibrillatory cycle. Based on these findings, several experimental and clinical studies have been done where the fact that the alterations in the atrial properties induced by tachycardias that help maintain the phenomena that occur during the episodes remain even after the conversion to sinus rhythm, is proven. In short, "AF begets AF". These permanent physiological alterations are correlated with the duration of the previous

tachycardia episodes and were called “atrial remodeling” and have been described in several animal [43] and human models [44, 45].

The changes in atrial properties include electrical changes [43-45] , changes in the gap junctions [46], changes in cellular morphology and anatomical changes [47].

4.5 Atrial Arrhythmias Treatment

As in-depth knowledge of the physiopathology of the disease remains lacking, the treatment of AF is a complex area. The main strategies include anti arrhythmic drugs and atrial ablation. Anti-arrhythmic drugs aim to modify the electrophysiological alterations of the atrial myocardium. Ablation treatments, on the other hand, consist in performing small lesions to the atrial tissue resulting in a scar that interrupts the propagation of the wavefronts responsible for AF. Pharmacological treatment is normally the first treatment option although their effectiveness can be limited and have related side-effects. The ablation procedure is the alternative for patients with failed pharmacological treatment and it is normally more efficient in the long-term. However, although this technique has significatively improved in recent years it cannot be applied to every patient and its effectiveness depends on the characteristics of the patient.

The mainstay of current ablation strategies is the isolation of PVs in the LA due to their discovery as focal ectopy. Success rate from PV isolation (PVI) in patients with paroxysmal AF is approximately 70-75% however, it is significantly less effective in persistent AF where the success rate of single ablation procedure is approximately 50% [48]. Persistent AF patients develop structural remodeling with the appearance of fibrotic tissue areas which contribute to the maintenance of AF. For improving the success rate in these types of patients, the clinician tries to eliminate the AF drivers’ areas guided by complex navigating systems, such as Carto or Navex. These systems integrate the electro-anatomical information provided by the navigation systems with a 3D reconstruction of the atria using images obtained from MRI or CT scans. In ***Figure 9***, the pressure done by the ablation catheter on the 3D reconstructed atria can be observed.

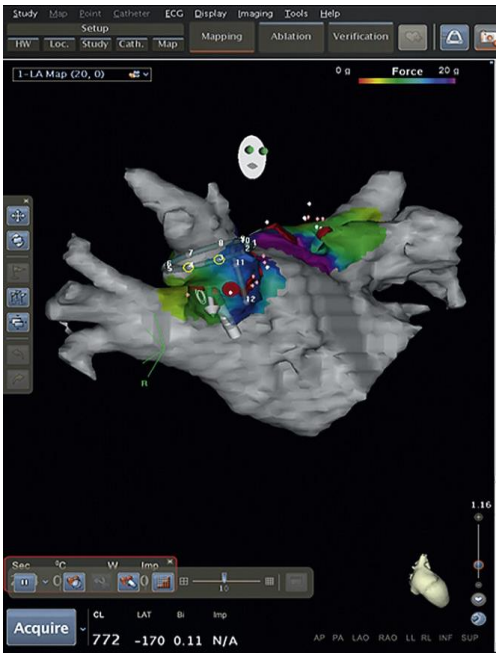


Figure 9. Three dimensional mapping (Carto system 3) with image fusion (3-T MRI) [48].

In the last years, the interest in using computational models of the atria to develop tools which would aid the clinicians in the ablation procedures has increased. Most sophisticated methods are based on patient specific 3D models obtained from the segmentation of MRI or CT images. Figure 10 shows the method developed by Dra. Trayanova [49]. It consists in collecting the MIR scan from the atria, image segmentation, atrial reconstruction and patient specific simulation which will suggest the clinician the optimum ablation area.

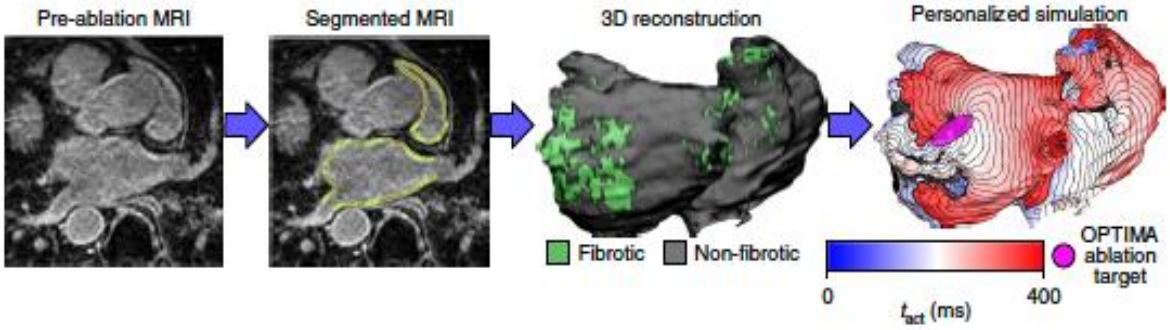


Figure 10 Personalized atrial simulation based on 3D reconstructed patient specific atrium [49].

To improve the effectiveness of this procedure, a more in-depth understanding is needed of the underlying atrial structural and functional substrates that sustain AF in human heart. Thus, the need for quantitative tools to evaluate the contribution of human atria 3D structural features in clinical and experimental settings.

Recent studies on the human atria imaged with LGE-MRI suggest that AF re-entrant drivers may be identified by distinct structural features, also referred to as “fingerprints”. These fingerprints consist of a combination of intermediate wall thickness, intermediate fibrosis and myofiber orientation.

Therefore, the quantification of the entire 3-dimensional atrial architecture may provide a novel method to predict AF driver location and improve patient-specific ablation procedures [32].

4.6 Cardiac Imaging Modalities

4.6.1 Magnetic Resonance Imaging

Signal generation

Magnetic Resonance Imaging (MRI) is a non-invasive imaging technique that produces detailed anatomical 3D images of the internal body structures based on the magnetic properties of the body. As ***Figure 11*** shows, for the image acquisition the patient is placed inside the MRI Scanner which consists of a large magnet that generates a strong and uniform magnetic field (1.5/3T). A set of radiofrequency coils apply radiofrequency (RF) pulses to the patient which excite the hydrogen protons found in the (water of) tissues. Upon relaxation the body will generate a signal that will be received by RF coils. Gradient coils cause local variations of the magnetic field and are used for the acquisition of spatially different "slices" of the body (image). This image modality is widely used for cardiac imaging as it provides high contrast of soft-tissue organs (they contain more water) such as the heart [50].

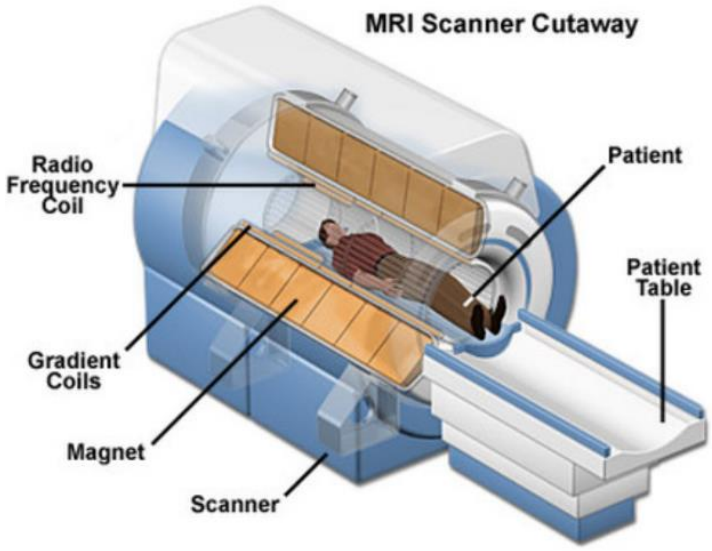


Figure 11 MRI Scanner [113]

This technique is based on the physical principle that nuclei with unpaired protons (e.g. H1 nucleus) possess a property known as quantum spin. As these nuclei have net positive charge, they generate a magnetic field as they spin represented by a dipole magnetic moment parallel to the direction of spin vector. At rest, these protons are randomly arranged with magnetic moments in different directions. However, in the presence of a strong magnetic field (Bo) the protons undergo a motion known as precession where they rotate around their axis at a specific frequency. This frequency is known as Larmor frequency and it is given by Equation (1). This frequency (w_0) is proportional to the magnetic

field strength. Furthermore, the nuclei align in a direction either parallel or antiparallel to the directions of the main magnetic field. The resultant magnetic vector is referred to as net magnetic vector (M_0) [2].

The frequency w_0 at which H nuclei precess is given by the Larmor equation:

$$w_0 = \gamma * B_0 \quad (1)$$

Where γ is the gyromagnetic constant.

As we can see the frequency of hydrogen precession depends on the magnitude Bo. MRI Scanners generate magnetic fields of 1.5T or 3T which correspond to H frequency of precession of 64MHz and 127.74MHz respectively.

When a RF pulse is applied, the energy transmitted as an electromagnetic wave, interacts with the magnetic vector of the protons. This interaction is due to resonance and only occurs if the frequency of the RF pulse equals the processional frequency of H nuclei. As shown in Figure 12 when a RF pulse is applied perpendicular to the direction of Bo (90º RF) the M_0 flips to the transverse plane (M_{xy}). Upon relaxation the protons return to their equilibrium state and lose the energy acquired in the form of an RF pulse otherwise known as MR signal, this signal consists of a magnitude and a phase which prior to relaxation is equal to 0 (in phase). If a coil is placed perpendicular to the transverse plane the RF pulse emitted induces a current (Faraday Law) that can be measured and processed representing the MR signal [50].

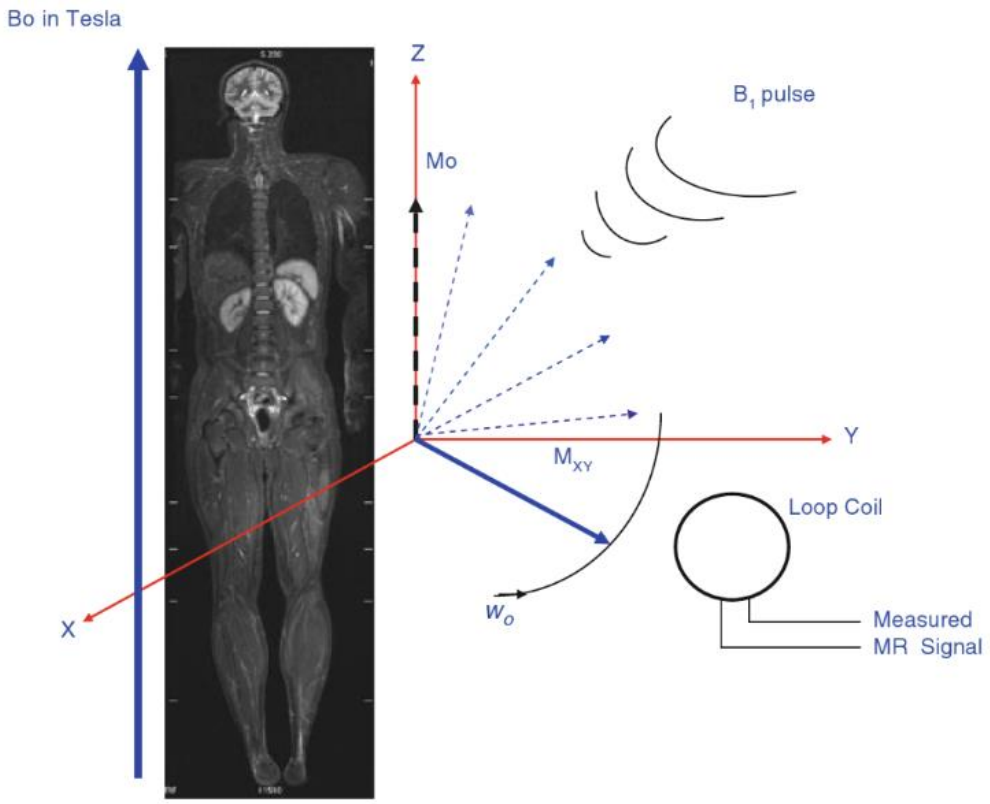


Figure 12 Acquisition process of MRI signals [50]

Furthermore, to encode the spatial location of the protons in the body a low-level spatially changing magnetic field is applied, otherwise known as gradient. This frequency encoding gradient causes a small frequency variation of the protons in each location that can be later decoded to create the MR image. The phase encoding gradient is applied in the opposite direction and further encodes the protons in a row. The frequency and phase encoding gradients label the protons in a specific 2D slice however, in order to select the slice a third spatial encoding gradient must be applied. The slice select gradient enables to choose the thickness and location of the 2D slice. In 3D imaging acquisition an additional gradient is applied to label the protons in the third dimension [50].

Signal properties

After the MRI signals are received by the coil, they are stored in a structure known as k-space. K-space is a complex space that represents the spatial frequency components of the signal. As shown in

Figure ***13***, both the frequency and phase gradient encoding (FE and PE) of the signal determine the position at which it is stored. The centre region represents low frequency components whilst the outer region encodes for high frequency components and have the lowest signal amplitude. An inverse Fourier transform is applied over the k-space to change from a frequency to a spatial domain therefore obtaining the image [50].

The properties of the k-space are related to the field of view (FOV) and resolution of the resultant image. The FOV represents the region of a slice acquired and it is dependant of the phase encoding gradient. An inadequate FOV result in aliasing which causes the image to wrap over itself after the inverse FT. To prevent this the FOV is chosen to be equal to $1/\Delta k$ where Δk represents the distance between subsequent points in k-space. Additionally, spatial extent of the k-space (FOV_k) is directly related to the resolution of the resultant image ($1/\Delta x$). As can be seen in Figure 13 a large FOV_k results in high-image resolution [51].

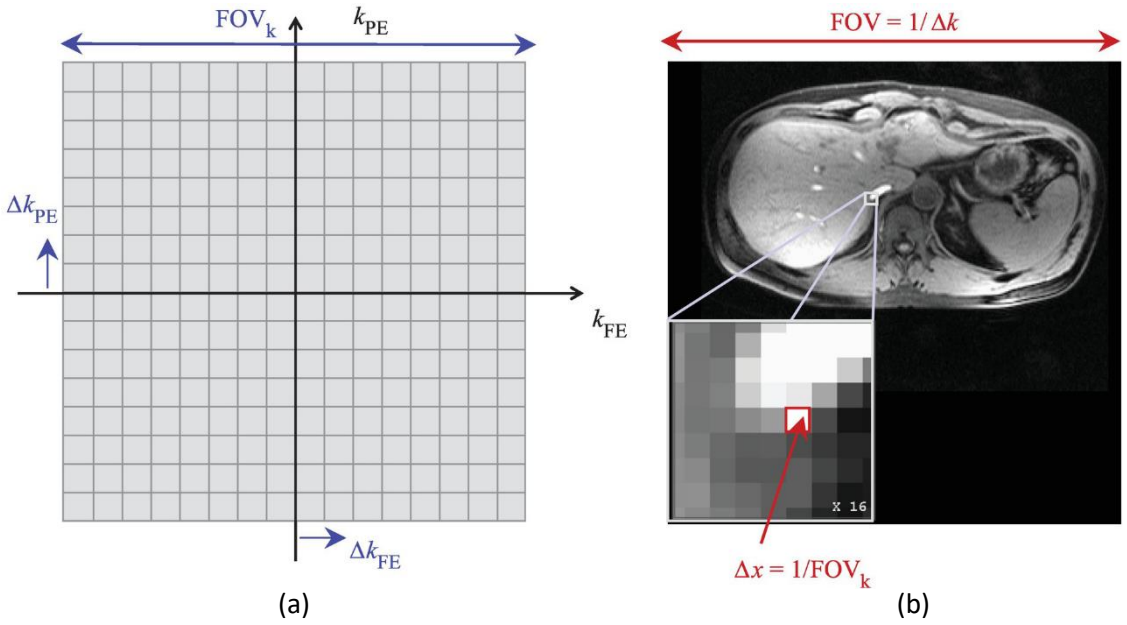


Figure 13 (a) k-space representation of the signal (b) spatial representation of the signal [52].

In 3D imaging the spatial encoding is performed in the 3 dimensions by an additional space encoding gradient resulting in a 3D k-space. As a result, the image is also encoded in the slice direction. 3D imaging must not be confused with multi-slice 2D imaging which consist of stacking multiple k-spaces. 3D encoding is slowly but surely entering the clinical practice as it provides the advantage of a higher SNR [51].

The contrast of the image is defined by the relaxation process. Relaxation is the process by which Mo returns to its initial state after the RF pulse excitation. It is governed by two processes: longitudinal and transverse relaxation, both are tissue specific [51].

Longitudinal relaxation refers to the recovery of the longitudinal direction of Mo due to spin-lattice interactions. The rate of recovery is governed by a time constant T1 which is tissue specific. Tissues with short T1 such as fat will appear hyperintense in T1-weighted images whereas tissues with long T1 such as water will appear as hypointense. In contrast, transverse relaxation occurs when spin-spin interactions between protons results in protons dephasing. This process is also defined by a time constant T2 where, contrary to T1, is larger in fat tissue (hypointense) and smaller in water (hyperintense) [50].

T1-weighted images especially useful for anatomical studies such as the evaluation of myocardial RV and LV wall thickness. Furthermore, T1 mapping has recently become of great interest in cardiac imaging as there is evidence to suggest that after contrast administration the T1 contrast can correlate with the amount of myocardial fibrosis. This technique is known as LGE-MRI and will be addressed further on.

Cardiac MRI in clinical practice

Cardiac MRI has proven useful for the evaluation of acquired and congenital cardiac related diseases such as cardiomyopathies, myocardial ischemia, coronary artery diseases, and congenital heart disease (CHD) among others. The availability of a large FOV, multiplanar acquisition and lack of ionizing radiation features of the MRI has led to its widespread use in routine cardiology clinical practice [53]. Furthermore, cardiac MRI encompasses various pulse sequences and protocols that can be applied to evaluate different aspects of the heart's anatomy and function.

The main coordinate systems used in 3D MRI cardiac acquisition include the body or scanner planes (Figure 14). These planes are oriented orthogonal to the main axis of the body and provide a qualitative overview of cardiac morphology. They consist of coronal, sagittal and axial plane. The sagittal plane can be used to track the aorta arising from the ventricles whilst the coronal and axial plane provide an overview of different cardiac structures such as the left and right atrium and the pulmonary veins [54].

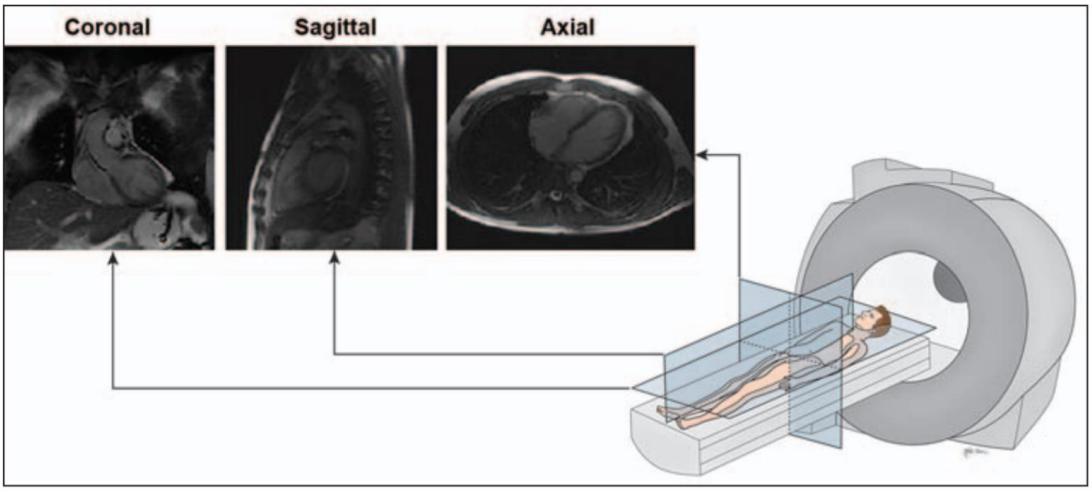


Figure 14 Schematic show of main body planes and their appearance in bright blood imaging technique [54].

With prior knowledge of T1 and T2, pulse timing parameters can be altered to provide specific tissue contrasts. This is done by software programs that determine the magnetic field gradient and the magnitude and timing of the RF pulses emitted by the MR scanner. By varying the sequence of RF pulses emitted and collected different contrast images can be produced. Repetition time (TR) is the time between two successive pulses and echo time (TE) is the time between the application of a RF pulse and the reception of the emitted signal [51].

In the following section the main pulse sequences along with the acquisition protocols currently used for cardiac imaging will be further detailed.

4.6.2 Balanced steady state free precession (b-SSFP)

Balanced steady state free precession (b-SSFP) sequences are bright blood imaging techniques that when applied generate images where fast-moving blood is represented with high intensity; this is normally done to evaluate cardiac function. bSSFP sequences are essentially gradient echo (GRE) sequences that for the signal production rely on steady state magnetization. The steady state magnetisation is achieved by keeping the TR short so that residual transverse magnetisation is present in subsequent excitations and is optimised by balancing the phase encoding gradients. bSSFP has recently replaced GRE sequences for cardiac imaging due to the improved myocardial-blood contrast and the high signal efficiency achieved [51].

Unlike GRE, bSSFP sequences are dependent on the square root of T2/T1 therefore blood provides a much higher signal than myocardium (blood: T1=1200ms, T2=200ms; myocardium: T1=867ms, T2=57ms). The signal from each tissue also depends on the flip angle, the optimum flip angle for blood is 45º whilst for myocardium is 30º. In clinical applications the flip angle is usually set between 50-80º this results in the blood signal intensity being approximately 2 times greater than the myocardium [51].

bSSFP sequences have a widespread use in both single and multi-phase cardiac image acquisition. In multi-phase or cine MRI where multiple k-spaces are acquired throughout the cardiac cycle, bSSFP is the main sequence used. This is due to the high myocardial-blood contrast but also the improved

temporal resolution they provide which enables an excellent evaluation of myocardial morphology and function. Many studies regarding cardiac function and morphology also use this technique. For example, Masamichi et al. uses 2D multiphase b-SSFP sequences for LA wall tracking to identify time-dependant changes in LA volume and strain rate and evaluate their function [55].

Unfortunately for 2D multiphase acquisitions the patient is required to hold their breath multiple times which can be difficult with non-cooperative and sedated patients. Furthermore, the scan must be carefully planned requiring expert knowledge on cardiac anatomy. To deal with these drawbacks without losing temporal resolution Uribe et al. proposed a 3D multiphase acquisition that incorporates self-respiratory gating [56]. Self-respiratory or navigator gating is a technique for respiratory motion correction where the diaphragmatic position is measured to restrict data acquisition to certain points of the respiratory cycle [51].

bSSFP sequences are also frequently used in 3D whole-heart techniques (Figure 15) that consist in single-phase acquisitions where a single volume of the heart is acquired at a certain point of the cardiac cycle. By acquiring the heart at specific points of cardiac cycle, otherwise known as ECG gating, the cardiac motion is compensated which improves image quality. Due to the high contrast ratio between blood pool and myocardium one of the applications of this technique is non-contrast MR Angiography however, because it must be cardiac gated the acquisition time is long therefore it cannot be performed during a single breath hold. As a result, navigator gating is normally used with 3D MRA to compensate for respiratory motion. One of the major drawbacks of this technique is the long acquisition times, 10-15 min, however this technique is still used in several studies as it provides excellent delineation of intra-cardiac structures. Hilbert et al. [57] acquired images with 1.5T 3D bSSFP free-breathing technique for segmentation and posterior reconstruction of left and right atria volume for catheter guidance in ablation.

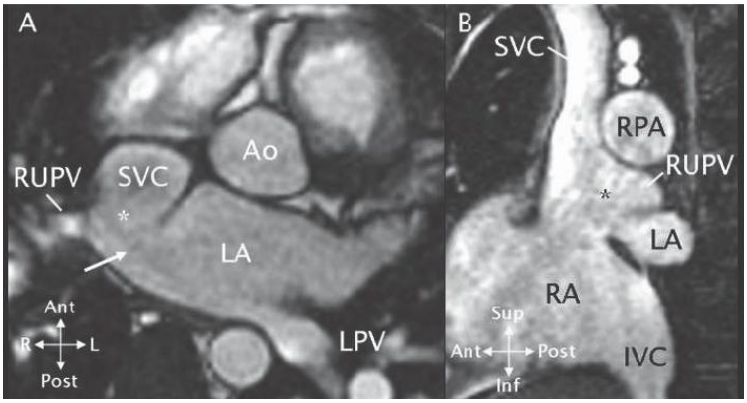


Figure 15 3D- bSSFP MRI A) axial view showing the superior vena cavae (SVC), the right upper pulmonary vein (RUPV), left pulmonary vein (LPV), left atria (LA) and aorta (Ao). B) sagittal view that shows the right atria (RA) and LA, the right pulmonary [58].

The high blood pool to myocardial contrast present throughout the cardiac cycle with no dependency on contrast enhancement has allow bSSFP to become one of the predominant sequences in cardiovascular imaging. However, one of its drawbacks is the high sensitivity to field inhomogeneities which results in dark band artifacts. Dark band artifacts are caused as field inhomogeneities induce dephasing of the magnetic vector until the signal collapses (180º). To prevent this local shimming can be applied to reduce magnetic field inhomogeneities. However, there are other sources of magnetic

field inhomogeneities such as metal inside the thoracic cavities (e.g. stents, clips), in these circumstances the use of bSSFP is not recommended. The amount of dephasing can also be reduced by keeping the TR short. In clinical practice, for example, optimal TR is kept around 2-3ms. Another drawback of bSSFP is the high energy deposition due to the large flip angle and short TR, this is not a significant problem with 1.5T field strength acquisitions [51].

4.6.3 Late Gadolinium Enhancement -MRI

Late gadolinium enhancement MRI (LGE-MRI) is a widespread technique for fibrosis detection in myocardial tissue. It is based on the increased volumetric distribution of gadolinium in fibrotic tissue. Gadolinium (Gd) is an element which serves as an extracellular agent that reduces T1 due to its paramagnetic properties. Thus, in T1 weighted sequences (e.g. FLASH) it will appear with high intensity. Furthermore, the acquisition is often performed with inversion recovery (IR) sequence to null the myocardial signal and increase contrast to noise ratio [59].

IR sequences consist in applying a 180º (inverse) pulse that flips the longitudinal magnetization, as it recovers (T1) it will pass through the transverse plane, if the signal is acquired at that specific time it equals to 0. Inversion time (TI) represents the time between applying the inverse pulse and image acquisition, if it is chosen carefully specific tissues can be abolished in the signal [51].

The acquisition protocol for LGE-MRI varies from study to study, however, there are some features that in general are kept the same. The LGE acquisition protocol consists in the intravenous injection of 0.1-0.2 mmol/kg bolus of Gd contrast and after 10-20 minutes the acquisition of T1-weighted image. The image is acquired with IR prepared pulse sequence. The IR preparation pulse is applied to null the normal myocardial signal and obtain a higher intensity signal from the fibrous tissue. Also, the acquisition of the T1-weighted image is normally acquired with GRE pulse sequence for single phase acquisitions with a TE chosen to supress fat tissue and improve delineation of LA wall [59-61] . In 3D acquisitions especially were the scan times are long, the image is acquired with navigator gating and ECG gating to compensate for cardio-respiratory motion and allow free breathing [60,61].

Myocardial suppression results in images with higher diagnostic quality where the endo-, mid- and epicardium can be easily differentiated. However, for this the optimal TI must be chosen and this value differs from patient to patient. The TI can be determined by Look-Locker sequences where multiple images with variable TI are acquired during one breath-hold after which the examiner picks the one with a visually better contrast. Furthermore, the LGE-MRI image acquisition is normally performed with navigation gating and ECG gating to compensate for cardiac and respiratory motion [51,59].

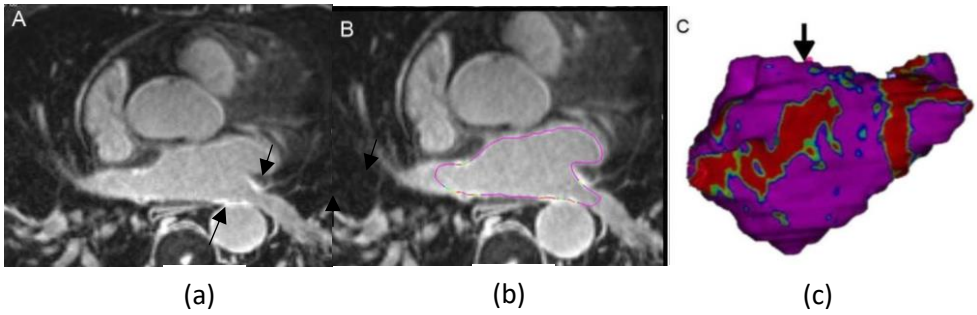
The physio pathological principle of Gd imaging is that Gd is a molecule that under normal conditions, after intravenous injection, distributes in the extracellular space without entering the myocardial cells. However, under certain pathological conditions, either by the disruption of the myocardial cell wall or the increase in extracellular space, the volume of Gd distribution is increased. Some of these pathological conditions include acute myocardial infarction where necrosis ruptures myocardial cell membranes leading to an increase in Gd distribution and chronic myocardial infarction where the fibrous scar produced leads to an increase in extracellular space. In T1-weighted images the increase in Gd distribution can easily be detected as the intensity of the image is increased (hyperenhancement), therefore the fibrous region is identified [62].

On the other hand, the threshold to identify hyperenhancement is not clearly defined. In clinical practice the enhanced fibrotic region is evaluated by visual estimation, measuring the percentage of thickness of infarcted myocardium related to global wall as a measure of transmural fibrosis [59]. The American Heart Association proposed a semi-quantitative approach to define extent of transmural fibrosis composed by a 17 segment model where each segment is given a score from 0-4 depending on the extent of the scar (0:no scar, 4:100%) [63].

Whilst LGE-MRI ventricular fibrosis assessment has achieved excellent results in numerous studies [64-66] it remains a challenge in atrial fibrosis assessment due to the thinness (1-2mm) and unpredictable shape of the LA wall, thus requiring greater spatial resolution [67].

Nevertheless, LGE-MRI has emerged as the most important independent predictor for AF post-ablation recurrence as well as a good technique for patient selection and procedural ablation strategy [60,68]. As ablation by RF induces myocardial fibrosis, LGE-MRI is a powerful method to define the extent of LA wall injury after this procedure. As a result, it can also detect gaps at ablation lines that may lead to AF post ablation recurrence [69] and guide ablation strategies. In another study Mc Gann et al. [61] validate the hypothesis that LGE-MRI can identify LA wall structural remodelling (SRM) and stratify patients who are likely to benefit from ablation. In this study LA wall enhancement was significantly greater in patients with AF versus patients with non-AF (control). In addition, Sanchis et al. [70] segment the LA from LGE-MRI images and quantify fibrosis for posterior 3D reconstruction to guide ablation procedures (

Figure ***16***).



(a) (b) (c)

Figure 16 (a) Axial view of 3D LGE-MRI left atrial acquisition (b) epicardial and endocardial traced borders for left atrium segmentation. The hyper enhanced region corresponding to fibrous tissue is indicated (c) 3D volume rendering of LA segmentation (purple) and the fibrous tissue projected over it (red) [70].

Despite these promising findings LGE-MRI has not been widely adopted in routine clinical practice for assessment in LA fibrosis. This is mainly due to the poor reproducibility of the results obtained by different scientific centres. The intensity of LGE-MRI image in LA is affected by parameters such as the coil proximity, contrast dosage and the time of acquisition after it is applied and patient characteristics such as body mass index or renal function. All these variables in the MRI acquisition make difficult the comparison of results between investigation groups and the lack of a consensus in the acquisition protocol further leads to different and sometimes contradicting results. For example, in LGE T1 mapping the optimal TI varies between different subjects. The selection of the optimal TI is of crucial importance as a sub optimal TI value may lead to over or under estimation of the fibrous scar [71].

In addition, all the previously mentioned studies as well as the current clinical practice quantify atrial fibrosis by performing manual segmentations of the LA chamber from LGE-MRI images which are time-consuming, labour intensive and error prone due to the complex anatomy of the atria. As a result, there is an interest in automatic segmentation algorithms of the left atrium for posterior scar quantification [67].

4.6.4 Computed Tomography

Computed Tomography (CT) is non-invasive imaging technique that allows fast acquisition of images at a high spatial resolution. This allows the high detailed delineation of cardiac structures even with cardiac and respiratory motion [72].

During a CT acquisition a narrow beam of X-rays is aimed at the patient and rotated around the body to produce 2D cross-sectional images or "slices" of the body. These computer-generated images can be digitally "stacked" to produce a 3D image. As the X-ray beams travel through the patient's body, they attenuate by a different degree depending on the tissue thus generating the contrast. Bone tissue produces the biggest attenuation and in the resulting image is visualized with the highest intensity whilst soft tissues produce a varying degree of attenuation and may be difficult to see in the resultant image. A detector receives the attenuated signals and sends them to a computer where the image is reconstructed and visualized [73].

One of the disadvantages of CT is the relatively low contrast resolution when compared with MRI which limits its cardiovascular applications. Contrast enhanced CT is a common technique to increase contrast resolution in cardiac imaging based in the injection of contrast agents into the blood stream to highlight specific structures of the circulatory system. The intravenous (IV) contrast agent is based on iodine which has a high degree of X-ray attenuation and is therefore shown with a high intensity in the resultant image, this technique is known as CT Angiography (CTA) [74]. The high spatial resolution and contrast achieved with CTA has led to its many applications in analysis of small and complex structures such as the LA and PVs [75]. Other applications in the field of AF include detection of LAA thrombi which is associated with AF [76] and exclusion of coronary artery disease in patients with AF [77].

However, its main disadvantage is that ionizing radiation (X-rays) must be applied to the body for image acquisition which are in the long-term harmful for the patient. Also, it has a lower temporal resolution than MRI which is an important feature for many cardiac imaging applications. On the other hand, new scans are being developed to reduce the temporal resolution [72].

There is a lack of consensus in the best image acquisition technique for 3D computational atria model generation. 3D single phase b-SSFP methods along with LGE-MRI are the most used sequences for cardiac segmentation although contrast-enhanced CT are also used in several studies. For atria segmentation which is the aim of this project all these techniques have been used in different state-of-the-art studies with good accuracy results [61,78,79]. CT has the advantage of higher spatial resolution than MRI, but it applies ionising radiation to the patient's body. All techniques either by contrast enhancement or by sequence of pulses used provide high blood-myocardium contrast which is important for image segmentation. LGE-MRI has the added advantage of enabling the analysis of transmural fibrosis with the intravenous injection of Gd [60,61].

Chapter 5. Materials

5.1 Databases

For the development of this project three different databases were used to test the generalisation ability of the proposed semantic segmentation model. The three of them comprised of MRI images of the heart from human subjects.

5.1.1 Database 1. Left Atria

The images from this database were provided by the STACOM 2018 Atrial Segmentation Challenge. A total of 100 3D LGE-MRIs from patients with AF were submitted for the purpose of this challenge. A large proportion of the challenge data was originally provided by the University of Utah (NIH/NIGMS Center for Integrative Biomedical Computing (CIBC)) while the rest came from several other institutes. The original resolution of the data is 0.625x0.625x0.625mm^3[80].

Each 3D LGE-MRI patient data was acquired with a whole-body MRI scanner and contains the raw MRI scan and a binary mask corresponding to the left atrial (LA) cavity. The ground-truth was manually generated by experts in the field. The raw MRIs are in grayscale and the segmentation labels in binary data were 255 means positive and 0 means negative. The spatial dimensions of the MRIs vary depending on each patient, some of the volumes had 640x640 dimension whilst others 576x576. However, all MRIs contain 88 slices in the Z-axis [80].

Figure 17 shows an example of the raw MR image with the LA segmentation label for one patient.

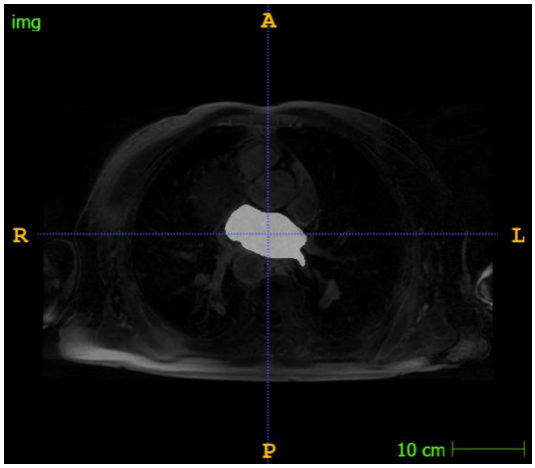


Figure 17 Axial slice view of raw MRI + LA label of dimension 640x640x88 and slice position equal to 60.

5.1.2 Database 2. Left Atria and right atria

The images from this database were provided by the Multi-Modality Whole Heart Segmentation (MM-WHS) challenge, in conjunction with MICCAI 2017. The database consists of cardiac MRI images that cover the whole heart from 20 patients. The data was acquired at a resolution of around 1.6-2x1.6-2x2-3.2 mm^3 and reconstructed to half its acquisition resolution, about 0.8-1x0.8-1x1-1.6mm^3.

The cardiac MRIs were acquired from two hospitals in London, UK. One set of data was acquired from St. Thomas Hospital on a 1.5T Philips Scanner and the other was acquired from Royal Brompton Hospital on a Siemens Magnetom Avanto 1.5T Scanner. Both acquisitions were made with a 3D balanced steady state free precession (b-SSFP) sequence for whole heart imaging which was navigator-gated for free-breathing.

The MRI data came from patients with an extensive range of cardiac diseases including myocardium infarction, atrial fibrillation (AF), tricuspid regurgitation, aortic valve stenosis, Alagille syndrome, Williams syndrome, dilated cardiomyopathy, aortic coarctation and Tetralogy of Fallot.

In this database different anatomical structures were manually delineated including the left ventricle cavity , the myocardium of the left ventricle, the right ventricle blood cavity, the left atrium blood cavity, the right atrium blood cavity, the ascending aorta and the pulmonary artery (Figure 18). For this project the main interest relays on the manual segmentation of the left and right atrium with label value equals to 420 and 550, respectively [81].

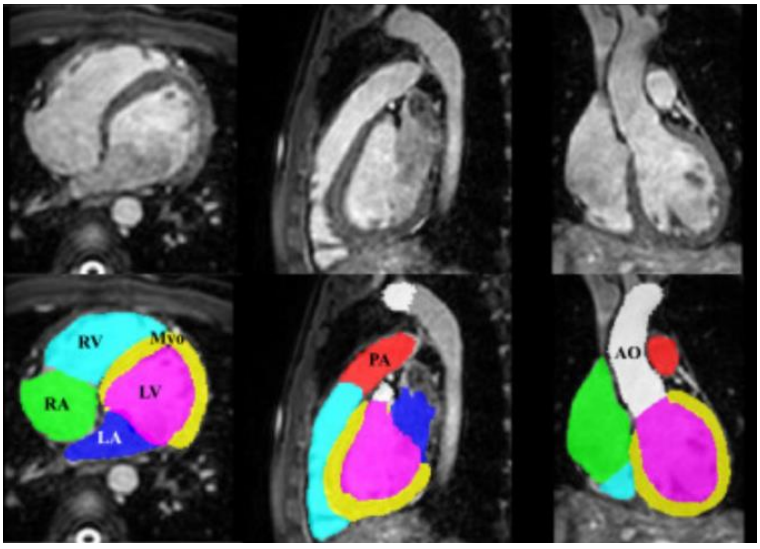


Figure 18 Example of raw cardiac MRI with labels of the cardiac structures. LV: left ventricle; RV: right ventricle; LA: left atrium; RA: right atrium; Myo: myocardium of LV; AO: ascending aorta; PA: pulmonary artery [81].

In this database there are different sources of variability. First, the shape of the heart varies greatly from subject to subject due to pathological and physiological changes. Secondly, the appearance of the image and the quality can also be variable (Figure 19). This is common in MRI images from clinical data where motion artifacts, poor contrast-to-noise-ratio and signal-to-noise-ratio can significantly deteriorate the image quality. Furthermore, there was also variability in terms of the spatial and slice dimensions of the volumes (

Table ***1***).

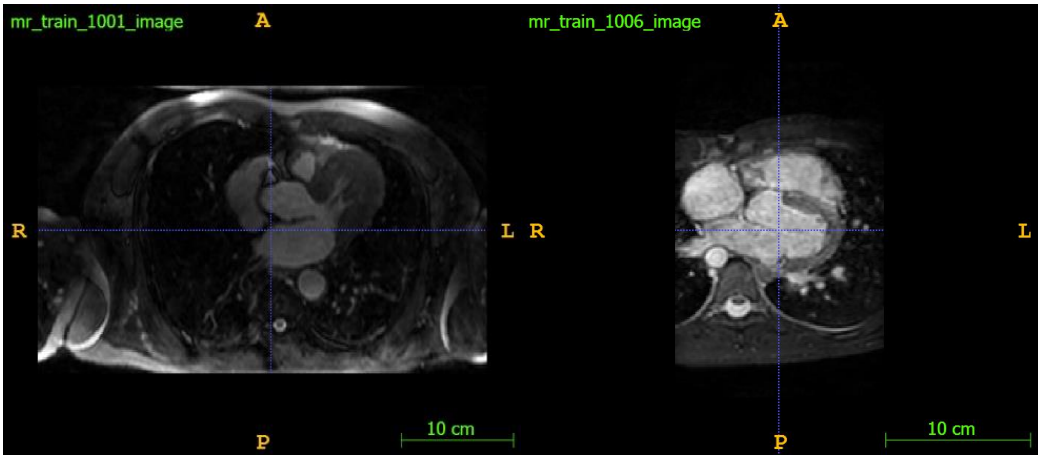


Figure 19 Axial view of patient 1001 (high quality) and patient 1006 (low quality)

ID Patient	Size (pixel)	ID Patient	Size (pixel)
1001	512x160x512	1011	161x288x288
1002	512x128x512	1012	512x128x512
1003	161x288x288	1013	512x112x512
1004	121x288x288	1014	512x160x512
1005	131x288x288	1015	201x340x340
1006	161x256x256	1016	131x288x288
1007	181x288x288	1017	141x288x288
1008	131x288x288	1018	151x288x288
1009	512x120x512	1019	136x288x288
1010	161x288x288	1020	136x288x288

Table 1 Spatial dimensions of the MRI acquisition per patient from database 2.

5.1.3 Database 3. Left Atria and other structures

The images from this database were provided by STACOM 2013 Left Atrial Segmentation Challenge in conjunction with MICCAI13. The data includes 30 MRI volumes provided by Philips Techhnologie GmBH, Hamburg and King´s College London, London, UK [82].

The MRI acquisition was performed on a 1.5T Achieva Scanner (Philips Healthcare,Best,The Netherlands) . The whole heart image was acquired using a 3D b-SSFP with navigator gating for free breathing and ECG gating for end diastole acquisition. The sequence acquired a non-angulated volume of the whole-heart with a voxel resolution of 1.25x1.25x2.7mm^3. Each dataset represents a single cardiac phase 3D volume image. The volumes of the datasets were provided with variety of quality levels in the following proportions: 9 high quality, 10 moderate quality, 6 local artifacts and 5 high noise datasets [83].

The ground-truth generation was performed by King´s college through an automatic segmentation of left atrial blood pool followed by manual corrections by experts. For 10 of these patients the ground-truth consisted solely of LA mask, however for the other 20 there were also other structures segmented and labelled such as the pulmonary veins (PV) and the left atrial appendage (LAA) (Figure 20). For these patients LA label equal to 36 [83].

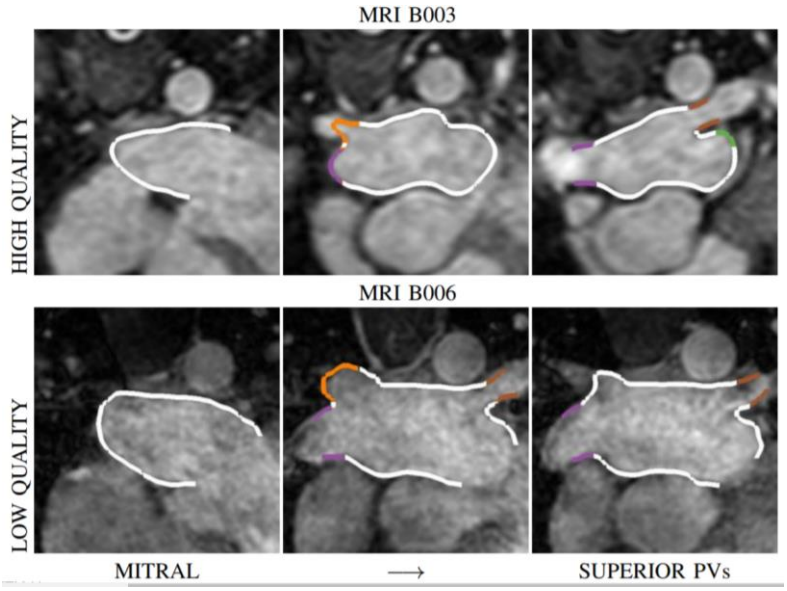


Figure 20 Examples of datasets provided with low- and high-quality data. Colour contours show the manual ground-truth (LA body=white, LAA=green, PVs= other colours) [83].

5.2 Software

For the neural network implementation and the evaluation of the resulting models the programming environment of PyCharm v2019.3.2 [84] has been employed. This Integrated Development Environment (IDE) supports the programming language of Python v.3.5.2 [85]. Python is an object-oriented, high-level programming language with dynamic semantics. It is very attractive for scripting due to its simple and easy to learn syntax. This, in addition to its open source characteristic makes Python the leading choice for data analytics and machine learning tasks. Python also integrates a set of libraries and modules for Data Science. Some of the libraries used in this project include Numpy, matplotlib, scikit-learn, SimpleITK, VTK and Keras. Keras is a high-level framework designed for training neural networks algorithms [86]. Keras also make use of backend libraries such as Tensorflow.

For external data processing and analysis of the results the programming environment MATLAB v.R2020a has been employed. MATLAB is platform designed for problem resolution in the field of mathematics and engineering that incorporates a high-level language based on matrix operations. It is used for signal and image processing and many other fields. Similar to Python it is available for different platforms such as Unix, Windows, Mac OS X and GNU/Linux [87].

Finally, the network training is performed in a remote server due to its high computational cost. To access this server the program MobaXterm [88] has been used. This tool allows an SSH (Secure Shell)

connection that provides an easy communication between the client (personal computer) and the server (where the algorithms are run).

5.3 Hardware

Most of this Project has been developed in a computer with a processor intel Core i5-7200U CPU @2.5 GHz and an operating system Windows 10 of 64 bits. The RAM memory capacity is of 8GB.

For deep CNN training, servers with very powerful processors and a larger storage capacity are required. As a result, in this project for the training processes a high-performance external server that belongs to the research group CVBLab has been employed. This server is composed of a processor Intel i7 @4.20GHz, 32GB de RAM and a graphic card NVIDIA Titan V.

Chapter 6. Methods

A deep neural network architecture proposed by [25] has been employed to develop an atrial segmentation model (LA and RA) with a high-level generalisation capability.

6.1 Deep Learning Fundamentals

6.1.1 CNN architecture

For an in-depth understanding of the 3D U-Net employed in this project we must first describe the network upon it is based on, the CNN. As can be observed in Figure 21, traditional CNNs consist of an input layer, in this case the image to segment, an output layer that shows the probability of an image to belong to a certain category and a series of functional hidden layers. The hidden layers are composed of convolution layers, batch normalization layers, pooling layers, flattening layers and fully connected layers.

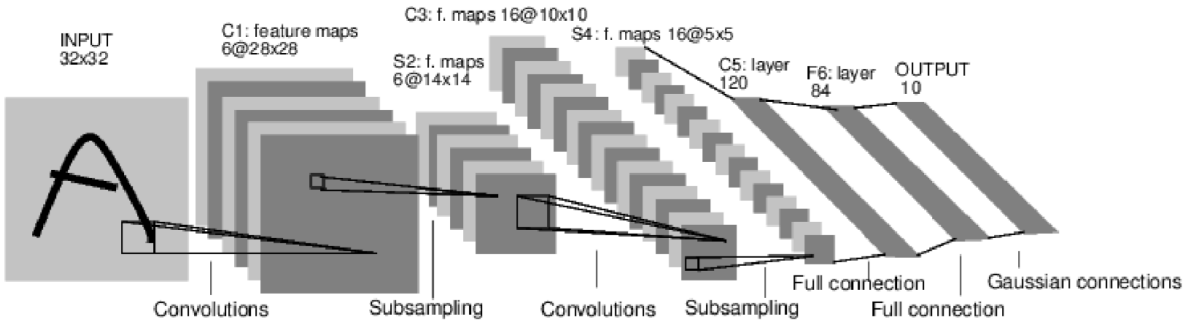


Figure 21 Generic example of CNN [89].

CNNs are composed of different levels. Each level consists, in general, of a convolutional layer followed by a normalization layer, after which the output is passed through a non-linear activation function to extract the feature maps of the image. These feature maps are down-sampled by a pooling layer at the end of each level. The last layer is connected to a fully convolutional layer to reduce the number of the features maps and generate the output vector of the network. This output vector in classification problems contains the probability of each pixel to belong to a certain class (e.g. Dog). The aforementioned layers along with their main functions will be described below.

Convolutional layers

Convolutional layers are the most important layers of the CNN. They are responsible for the automatic feature extraction and carry the main weight of the computational cost. Each convolutional layer consists of a series of filters, otherwise known as kernels, that are applied to the input data through a linear operation known as convolution. A convolutional operation involves the multiplication of an array or matrix of input data (image) with the 2D or 3D array of weights in the kernel. In other words,

each convolution uses a $n\ x\ n$ kernel (for 2D input) or $n\ x\ n\ x\ n$ (for 3D input). After each training process the weights of the kernels are updated.

Furthermore, the dimension n of the kernel is small to ensure that each neuron is only connected to a local region of the previous volume to improve computational efficiency. In deep learning approaches the popular choice is n=3; if n=5 or 7 the number of parameters and therefore the computational cost increases dramatically. The spatial extent of kernels is defined by an hyperparameter known as receptive field that represents the area of the kernel applied to the image ($n\ x\ n$). Although the dimension of the kernel is small by increasing the number of convolutional layers the receptive field also increases. Technically, the mathematical operation described as convolution in the use of CNNs is actually a "cross-correlation" given by Equation (2), however in deep learning field it is referred to as "convolution".

$$y[m,n] = x[m,n] * h[m,n] = \sum_k \sum_l x[k,l] * h[m+k, n+l] \qquad (2)$$

where x is the data input, h is the filter and y is the filter output.

Using a filter with smaller dimensions than the image allows the filter or set of weights to be multiplied by the input data at different points on the input. This systematic application of the same filter across the input array enables the discovery of specific features anywhere in the image. This capability is known as translation invariance. The number of pixels that the kernel shifts over the input matrix is another hyperparameter known as stride. It should also be considered that although in each stride the connectivity of the filter is local in terms of width and height, it extends over the full depth of the channel dimension in the input image. For example, for RGB images the channel dimension equals to three. Figure 22 illustrate the process of convolution with a kernel of stride equal to 1.

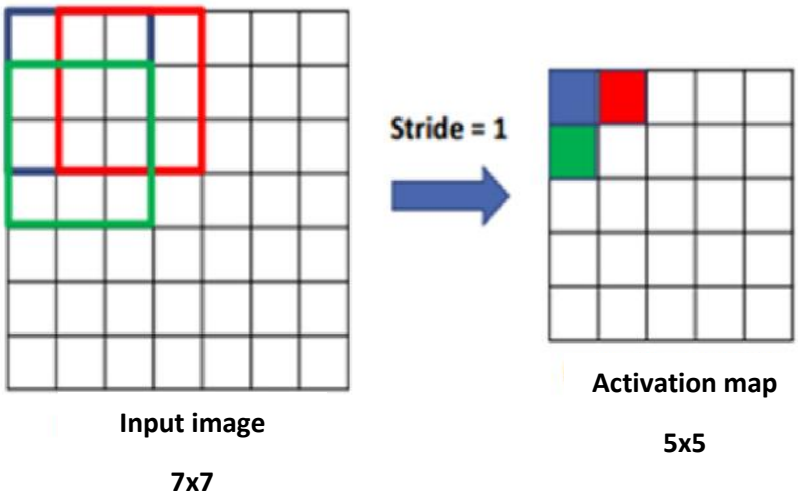


Figure 22 3x3 kernel (with stride equal to 1) sliding over input array.

As can be observed, after the convolution the dimensions of the resultant image are smaller than the input which leads to loss of information. To preserve the dimensions of the output image or volume a technique known as zero padding is used where the input volume is surrounded with zeros. The

dimensions of the padding is an hyperparameter that depends on the input image dimensions and kernel size. Figure 23 illustrates an example of applying zero-padding to an RGB image (3 channels). As we can observe, to maintain the image dimension of $M\ x\ N\ x\ 3$ a zero padding of dimensions $N\ +\ (K1-1)\ x\ M\ +\ (K2-1)\ x\ 3$ must be integrated to the image before convolution. $K1$ and $K2$ represent the dimensions of the kernel $(K1xK2)$ and M and N represent the row and columns of the image, respectively.

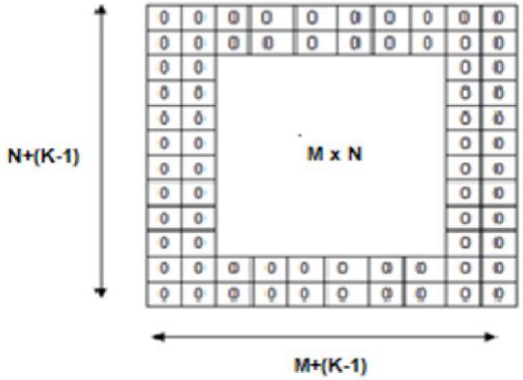


Figure 23 Zero padding applied over MxN array.

As each filter sweeps over the input image a bidimensional map that preserves the relationship between the pixels is formed. The map is known as feature map and each filter produces a different one. The feature maps are stacked in sequence and constitute the depth of the output volume, in other words, the output volume will have a channel dimension equal to the number of filters in the layer (Figure 24).

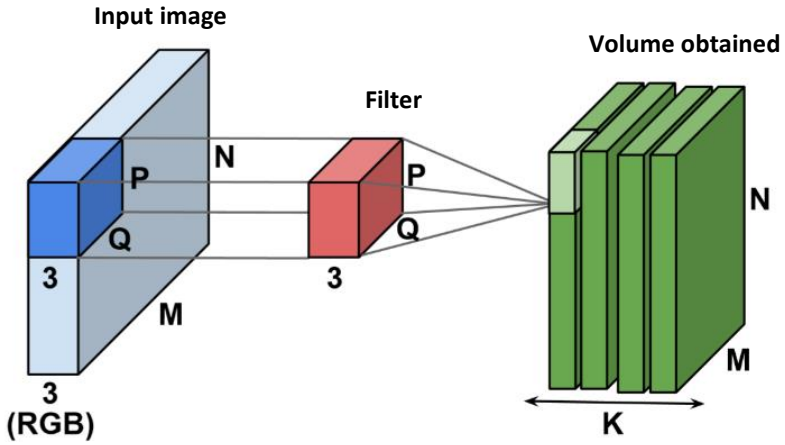


Figure 24 2D kernel (PxQ) applied over an input RGB image (MxN). Note that the depth or channel dimension of both the filter and the image coincide. The output volume is characterized by K channels referring to the number of filters established in the convolutional layer.

2D kernel (PxQ) applied over an input RGB image (MxN). Note that the depth or channel dimension of both the filter and the image coincide. The output volume is characterized by K channels referring to the number of filters established in the convolutional layer.

3D convolutions use a similar approach, the main difference is that the kernel moves in 3 directions. Following the previous idea 3D input images and kernels have 4 dimensions, 3 spatial dimensions and the channel dimension.

In a neural network, the filters from the first convolutional layer are applied to the input image or volume and extract the low-level feature of the image such as curves and edges. The resultant activation map is fed to the next convolutional layer as the input volume and the process is repeated each time obtaining more complex features. The subsequent deeper convolutional layers will extract image features of a higher dimension.

In addition, to down-sample the number of feature maps $1x1$ convolutions are normally used. If this convolutional kernel is systematically applied with a stride equal to 1 and no zero padding the resulting feature map has the same width and height of the input. As no neighbouring pixels are taken into account during convolution the output doesn´t contain any additional information and is considered a linear projection of the input. As each $1x1$ convolutional kernel will still generate a feature map, a convolutional layer with a specific number of $1x1$ convolutional kernels can therefore be used to control the number of feature maps [90].

Activation layer

In this layer an activation function is applied to introduce non linearities to the otherwise linear convolutional output, this enables the network to learn complex patterns in the data. They are applied at the end of convolutional layers and their output corresponds to the input of the next layer. Furthermore, they also preserve the input dimensionality to improve robustness. There are several existing activation functions that can be applied (Figure 25) however, the most used at this moment is the RELU (Rectified Lineal Unit) that transforms the negative activations to 0 by applying the function $f(x) = max(0, x)$. This accelerates the training process without compromising its accuracy. On the other hand, the RELU function is often described as fragile as it may become inactive no matter the input supplied. This is known as the dying neurons problem and it prevents the learning progress of the neural network [91]. In order to avoid this, a variant of the function, known as the leakyRELU is often used instead. This function was first introduced by Maas et al. [92] and it is characterised by allowing a small positive gradient for negative inputs thus extending the range of the function (Figure 25). The negative activations $(x < 0)$ will be transformed by the function $f(x) = \alpha * x$. Where α is a fixed parameter in range $(1, +inf)$.

The last layer of the network will have either a Softmax (more than 1 output) or a Sigmoid (1 output) activation function. As we can observe in Figure 25 the sigmoid function limits the output range between 0 and 1 thus these functions are normally used in the prediction of probabilities.

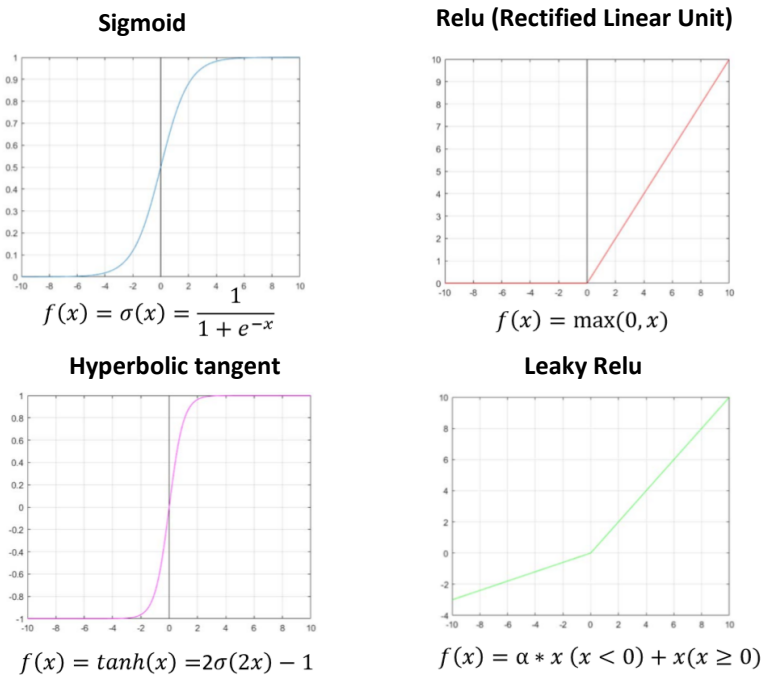


Figure 25 Common activation functions.

Pooling layer

Pooling layers are applied after the activation layer to reduce the spatial resolution of the feature maps in order to remove redundant features and improve statistical efficiency and model generalization. Furthermore, down-sampling the data reduces the number of parameters to adjust during the network training and helps to control overfitting. Pooling method consists of applying a filter of nxn dimensions to select pixels in a given neighbourhood with stride equal to 2 for 50% down-sample. There are two different pooling methods depending on the operation applied by the kernel to the selected pixels: average and max pooling. In average pooling the kernel applies an average function over the selected pixels whilst in max pooling a max function is applied to select the maximum pixels (Figure 26). In convolutional neural networks max pooling is the most employed method. Another common technique to reduce spatial dimensions is by applying a convolution function over the selected pixels. Strided convolutions have been increasingly used to down-sample in recent network architectures because they add extra parameters, the convolutional weights, and therefore improve the representational capability.

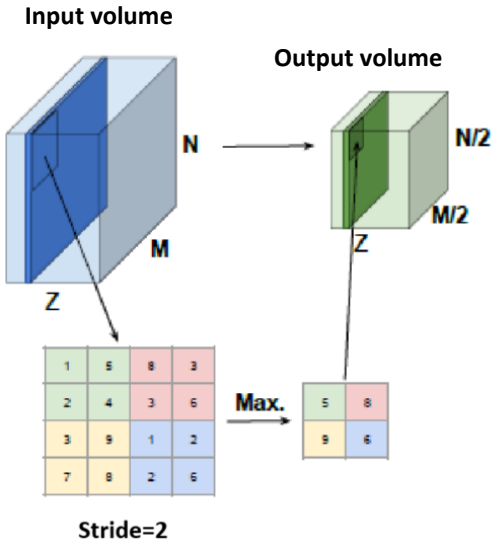


Figure 26 Down-sampling example with max pool with a 2x2 kernel with stride equal to 2.

Both methods are valid and used in state-of-the-art networks. The advantage of strided convolutions is that the acquisition of feature maps and the down-sample can be performed in one operation which reduces computational costs [93].

Once we reviewed the different layers that compose a CNN we must detail the process by which the network learns from labelled data in order to generate the output segmentation, this is known as training of the network. The main objective of the training process is to minimise the error between the output generated by the network (prediction) and the labelled data (ground-truth) by optimising the weights of the kernels. The training is performed by an algorithm known as backpropagation that adjusts the weights in order to find relevant features of the image. This process can be divided into 4 different sections: Forward pass, loss function computation, backward pass and weight update.

6.1.2 Training process

Forward pass

Forward pass refers to the process of the input image traversing through all the neurons (kernels) from first to last layer. In the first iteration the weights are randomly initialized therefore the output of the network (prediction) will not provide any information. In order to optimise the weights to provide useful information the prediction and the ground-truth must be compared via a loss function.

Loss function

The loss function refers to the error between the predicted output segmentation and the ground-truth. As our objective is that the prediction is as similar as possible to the ground-truth the loss function must be computed after each iteration in order to minimise it.

The network predictions, consisting of 2 volumes with the same resolution as the input data, are flattened and processed through a sigmoid function which as an output provides the probability of each voxel to belong to the foreground and to background. In medical volumes what is often the case is that the anatomy of interest occupies only a very small region of the scan and therefore there are much more background voxels than foreground. This may cause the learning process to get caught in

local minima resulting in predictions strongly biased toward the background and with partially detected foreground, this is known as class imbalance.

As a result, previous approaches inclined towards the use of loss functions with sample reweighting (e.g. cross-entropy loss function) where the regions with foreground pixels are given more importance than background during training. However, in this project the network implementation takes a different approach by applying a loss function based on dice coefficient, this idea was first introduced by Milletari et al.[14]. The dice coefficient is an evaluation metric that by the intersection of the prediction (p) and ground-truth (g) gives a metric of their similarity. This metric ranges between 0 and 1 the higher it is the most similar the 2 samples are. The dice loss function is of equal magnitude and opposite sign to the dice coefficient therefore, during training the aim is to minimize the dice loss function which is equivalent to maximizing the dice coefficient.

The binary class dice coefficient loss function is defined in Equation (3):

$$L_{DC} = -\frac{2\sum_i^N p_i g_i}{\sum_i^N p_i + \sum_i^N g_i + \sigma} \tag{3}$$

where the sum runs over the N voxels of the predicted binary segmentation (p_i) and the ground-truth binary volume (g_i) of each sample i. With a smoothness factor of σ = 0.00001 to avoid the division by 0.

The differentiation of the dice loss function is defined Equation (4):

$$\frac{\partial L_{DC}}{\partial p_j} = -2\left[\frac{g_j\left(\sum_i^N p_i^2 + \sum_i^N g_i^2 + \sigma\right) - 2p_j\left(\sum_i^N p_i g_i\right)}{\left(\sum_i^N p_i^2 + \sum_i^N g_i^2 + \sigma\right)^2}\right] \tag{4}$$

Which is computed with respect to the j-th voxel of the prediction. By using this formula there is no need to correctly balance the foreground and background voxels with weight assignation to samples of different classes. Furthermore, Millietari et al. states that the experimental results obtained by optimising with the dice loss function were much better than with the loss with sample re-weighting.

During the first training iterations the loss is extremely high therefore the objective during the training process is to minimize the loss function to generate a predicted segmentation as similar as possible to the ground-truth. In order to achieve this, the network must learn the weights that contribute in a higher proportion to the loss of the network or, in other words, the weights that minimize the loss function.

On the other hand, as we can observe in Equation (4), the gradient function is non-linear therefore it is not an easy task to find its minimum in order to minimize the loss function. As a result, an algorithm must be used for a spatial search of the local minima and the parameters that minimize the function, this is performed in the backward propagation.

Backward propagation

After computing the loss, the next step is to traverse the network backwards to adjust the weights in order to minimize the loss function, this is known as gradient descent.

Gradient descent is a method that minimizes an objective function $L(\theta)$, where θ are the model's parameters, by updating the parameters in the opposite direction of the gradient of the objective

function ∇L (θ). An hyperparameter known as the learning rate η determines the size of the steps taken to reach the local minimum. Simply put, it follows the slope of the surface generated by the loss function downhill until a valley (local minimum) is reached [94].

Gradient descent can be performed in different ways depending on the quantity of the data used to compute the gradient of the loss function. The most used gradient descent method at this moment is mini-batch Stochastic Gradient Descent (SGD). This optimisation algorithm divides the training set into batches (small sets of data) and performs the weight update one for every mini batch.

However, due to the relatively small training dataset used in this project the mini-batch size is set to 1. This means that the parameter update will be performed for every training sample as in traditional SGD techniques. The computational cost of this algorithm is low however the frequent updates causes the loss function to fluctuate heavily. This fluctuation enables the gradient descent to jump to potentially new and better local minima however it complicates the convergence to an exact minimum as SGD will keep overshooting.

Equation (5) illustrates the mathematical operation of the traditional SGD update.

$$\mathrm{W}(t+1) = \mathrm{W}(t) - \eta \nabla L \tag{5}$$

where W represent the weights.

Furthermore, there are other challenges that must be faced when dealing with SGD. The first one is that all parameter updates have the same learning rate however, if the data is sparse with features of varying frequency it may be better to perform a larger update for low frequency features. Another challenge for minimizing loss functions, common for neural networks, is getting trapped in suboptimal local minima [94].

As previously stated, one of the limitations of SGD is that it has trouble navigating through areas where the surface curve is steeper in one dimension than in another, which are common around local minima. As a result, SGD oscillates across the slope making slow progress towards the local optimum. Momentum is an extension of SGD algorithm that accelerates the convergence of SGD in the right direction. This is done by adding a fraction (lambda) of the previous parameter update to the current update vector [94].

The SGD update with momentum is represented in Equation (6).

$$\mathrm{V}(t+1) = \gamma \mathrm{V}(t) - \eta \nabla L, \qquad \mathrm{W}(t+1) = \mathrm{W}(t) + \mathrm{V}(t+1) \tag{6}$$

where γ represents a momentum term usually set around 0.9 and V a new variable initialized to 0.

Optimizer

Similar to the way the momentum of a ball increases as it rolls downhill, the parameter update increases for dimensions that have gradients pointing in the same directions and decreases for gradients of diverging directions. This results in a faster convergence. Figure 27 illustrates this phenomenon.

(a) SGD without momentum (b) SGD with momentum

Figure 27 SGD oscillation towards local minima[94]].

The optimisation algorithm used to train the neural network of this project is known as Adam optimizer. This is a variant of SGD that computes adaptative parameter updates for every time step t by performing larger updates for infrequent parameters and smaller updates for frequent ones which is useful when dealing with sparse data. The update of the parameter will depend on the computed gradients of the previous parameter. Specifically, Adam stores the exponentially decaying average of past gradients $v(t)$ for parameter update and uses an extension of momentum by storing the decaying average of past gradients $m(t)$.

The Adam optimizer function is defined in Equations (7)

$$m(t+1) = \beta_1 * \mathrm{m}(t) + (1-\beta_1) * \nabla L\,; \quad v(t+1) = \beta_2 * v(t) + (1-\beta_2) * \nabla L^2 \tag{7}$$

Where $m(t)$ and $v(t)$ correspond to the first moment (mean) and second moment (variance) of the gradients, respectively. As $m(t)$ and $v(t)$are initialized as vectors of 0 they are biased toward 0 so the authors counteract this by computing bias corrected $m(t)$ and $v(t)$ estimates.

Finally, these averages are used to update the parameters following the Adam update rule defined in Equation (8):

$$W(t+1) = W(t) - \frac{\eta}{\sqrt{\mathrm{c}(t+1)} + \varepsilon} * m(t+1) \tag{8}$$

where $c(t+1)$ is equivalent to the bias corrected $v(t)$

The CNN of this project uses the Keras implementation of Adam optimizer which by default has the parameter values of: $\beta 1 = 0.9, \beta 2 = 0.999$ and $\varepsilon = 1e-07$ which are the recommended values by the author.

Learning rate

Furthermore, one of the challenges in training neural networks is choosing an optimal learning rate value (n). If the learning rate is too small, it leads to very slow convergence but if it's too high it could cause the loss function to diverge and not reach the local minimum. Each iteration of the backpropagation over the complete set of training samples is known as an epoch. The best approach is to vary the learning rate each epoch. Normally it is set to be higher in the first epochs and smaller or more precise towards the final epochs of the training process. There are several techniques for learning rate optimisation. In this project we use the technique known as step decay where the learning rate is reduced by a specific factor every number of epochs [95].

Transfer learning & fine-tuning

Typically, the U-Net is trained from scratch with randomly initialized weights, however training a deep CNN from scratch has several complications. In the first place, for a successful training CNNs require a large amount of labelled data, a requirement often difficult to meet in the medical field where manual segmentation by experts is highly time consuming and expensive. Furthermore, training a deep CNN requires large computational and memory resources in order to reduce the computational time of the training. Finally, training a deep CNN can be challenging due to overfitting and convergence issues that to be resolved normally require frequent adjustments in the network architecture or learning parameters of the network. Therefore, deep learning from scratch for each new dataset is tedious and time-consuming [96].

Before the training process the weights are initialised randomly and during backpropagation they are updated after each epoch from the weights of the previous iteration. More specifically, they are normalized with a mean of approximately 0 and with a small standard deviation. However, large number of weights together with limited labelled data may lead to undesirable local minimums during the backpropagation of the loss function which results in convergence issues. In order to resolve this, the weights of the convolutional layers can be alternatively initialized to the weights of a pre-trained CNN with the same architecture, this is known as transfer learning [96].

With this technique, instead of training a CNN from scratch, the pre-trained weights of a CNN trained with an extremely large dataset are taken advantage of, typically the ImageNet dataset. Specifically, this method consists of freezing all the layers from the pretrained network preventing them to update their pretrained weights and training the top model of the network with images for a specific application. Iglovikov et al.[97] uses this approach by pre-training the encoder layers of a U-Net.

On the other hand, the transfer learning technique albeit being relatively simple does not always provide good results for complex problems as it is not very specific for the image dataset. Another alternative is to only freeze some of the initial convolutional blocks and retrain or fine tune the final ones instead of freezing all of them. In general, the early layers of a CNN extract low level features whilst the late layers learn high level features more specific for the target application. Therefore, an effective fine-tuning technique is to start by unfreezing the last layers and incrementally include more layers until the desired performance is reached. With this the weights of the final layers start the update based on the pretrained weights of the previous layers. Fine tuning only the last layers of the feature extractor is referred to as “shallow tuning” whilst fine tuning blocks of convolutional layers is referred to “deep tuning” (Figure 28). Tajbakhsh et al.[96] in their fine-tuning review stated that deep fine tuning outperforms shallow fine tuning for medical image applications.

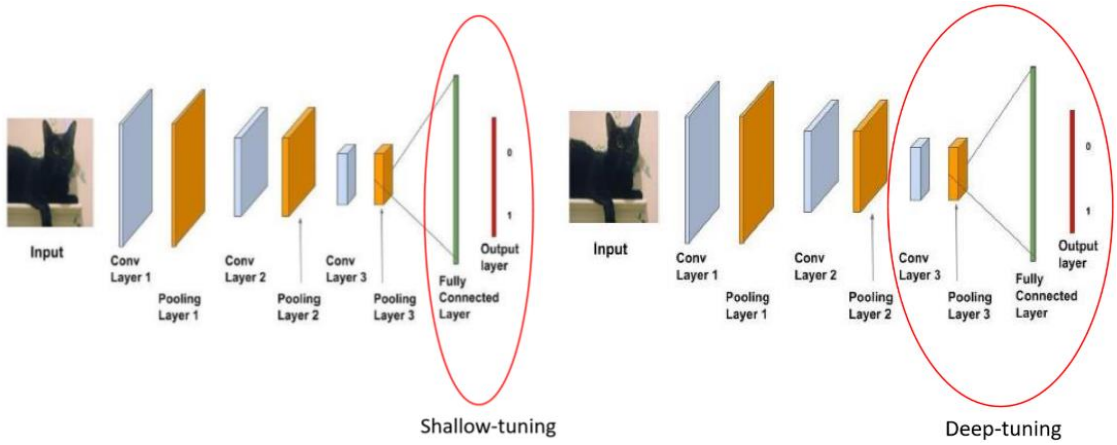


Figure 28 Shallow and Deep tuning example. The red circle represents the unfrozen layers

6.1.3 Normalization and regularization in deep learning

Dropout

Deep convolutional neural networks with large number of parameters and high complexity face the challenge of learning too well from the training dataset and are therefore, not able to generalize well when presented with other data, this is known as overfitting. This problem can be addressed with regularization methods. One of the regularization methods used in this project is dropout. Dropout regularization is a method where in each training iteration a random sample of outputs in the layers are deactivated with a probability p (probability of disconnecting outputs). This random disconnection discourages complex co-adaptations of learnt representations between layers which do not generalize for unseen data. For example, a spatial dropout of 0.5 means that in every iteration half of the layers output will be randomly disconnected [98].

Instance normalization

One of the challenges of training deep neural networks is the change in distribution of the deep layer´s input when the parameters of the previous layers are modified during training. This phenomenon, known as internal covariate shift, slows down the training process and makes it difficult to train models with saturating non linearities. The change in the input distribution of the deep layers will affect their output producing fluctuations in saturable regions of some activation functions such as the sigmoid. When this happens the neurons (kernels) are not able to update their weights as the gradient can´t travel backwards through the layers. This is known as vanishing gradients [99].

To mitigate the internal covariate shift and avoid the vanishing gradient phenomenon after each convolution the feature maps must be normalized. There are several normalization techniques. At this moment, the most popular technique is Batch Normalization where the activations are normalised across each mini-batch (set of input samples in each iteration), in other words, each feature maps is normalised by subtracting the mean and dividing by the standard deviation of the mini-batch [99].

On the other hand, the network used in this project replaces the traditional batch normalization layer with instance normalization. This is done because when working with small batch sizes, as occurs in this project, the stochasticity induced may destabilize batch normalization. The key difference between batch normalization and instance normalization is that in the latter the mean and standard

deviation are computed from a single sample instead of from a whole batch of images. Furthermore, due to its non-dependency to mini-batch, instance normalization can also be applied at inference time [99, 100].

Instance normalization is defined in Equation (9) [100]:

$$y_{tijk} = \frac{x_{tijk} - \mu_{ti}}{\sqrt{\sigma_{ti}^2 + \epsilon}};\ \mu_{ti} = \frac{1}{HW}\sum_{l=1}^{W}\sum_{m=1}^{H} x_{tilm};\ \sigma_{ti}^2 = \frac{1}{HW}\sum_{l=1}^{W}\sum_{m=1}^{H}(x_{tilm} - mu_{ti})^2 \quad (9)$$

Here x $(TxCxWxH)$ consists of an input tensor which contains a batch of T images. For the x_{tijk} th element, i is the feature channel, j and k are the spatial dimensions corresponding to the height (H) and width (W) of the image and t is the index of the image in the batch. Therefore, μ_{ti} and ${\sigma^2}_{ti}$ correspond to the mean and variance of each image sample, respectively. The ϵ is non-zero small constant to avoid division by 0.

6.2 3D Dual U-Net Architecture

6.2.1 Encoder-decoder CNN

In the following section the architecture of the network employed in this project will be described. This network was proposed by Jia et al. for the 2018 STACOM LA challenge. The challenge consists in the automatic segmentation of LA blood pool in LGE-MRI images and the output of the network is a binary mask with voxels corresponding to the blood pool volume of LA activated.

For the segmentation task Jia et al. proposed a dual 3D U-Net. The U-Net is a typical encoder-decoder network based on the work by Ronneberger et al. [11]. In addition, the network used in this project follows the 3D extension of the U-Net first proposed by Cicek et al. [13] where 3D volumes are taken as input and processed by 3D convolutional kernels. Training with volumetric images enable for a better generalisation of the model therefore the network can be effectively trained with sparsely annotated data which is the case in most biomedical applications.

The composition of the encoder and decoder pathway is detailed below, the implementation of the U-Net described follows the work of Isensee et al. [101].

Encoder

The encoder is a context aggregation pathway that extracts increasingly abstract representations as it goes deeper in the network. The feature maps in the encoder are computed by context modules. Each context module is in fact a pre-activation residual block [102] with additive residual connections that consists of 2 convolutional blocks with a dropout layer ($p = 0.3$) in between. Each convolutional block consists of a convolutional layer with 16 filters of kernel size $3x3x3$ and $stride = 1$ with zero padding, followed by instance normalization layer and a reLU activation function. The zero padding is applied to maintain the spatial resolution of the image after applying the stride=1 convolution. Furthermore, context modules are connected with each other by $3x3x3$ convolutions with $stride = 2$ that act as the down sampling layer, halving the resolution of feature maps whilst doubling their number.

As it can be observed in Figure 29 the network has a 5-level depth. Higher depth levels refer to higher dimensional feature representations but with lower spatial resolution, as seen, as the level of depth increases the size of the input is decreased whilst the number of feature maps increases. Furthermore, each level except the first one consists of a down-sampling convolution followed by a context module. In the first level a simple $3x3x3$ convolution is applied over the input volume to extract the 16 initial feature maps.

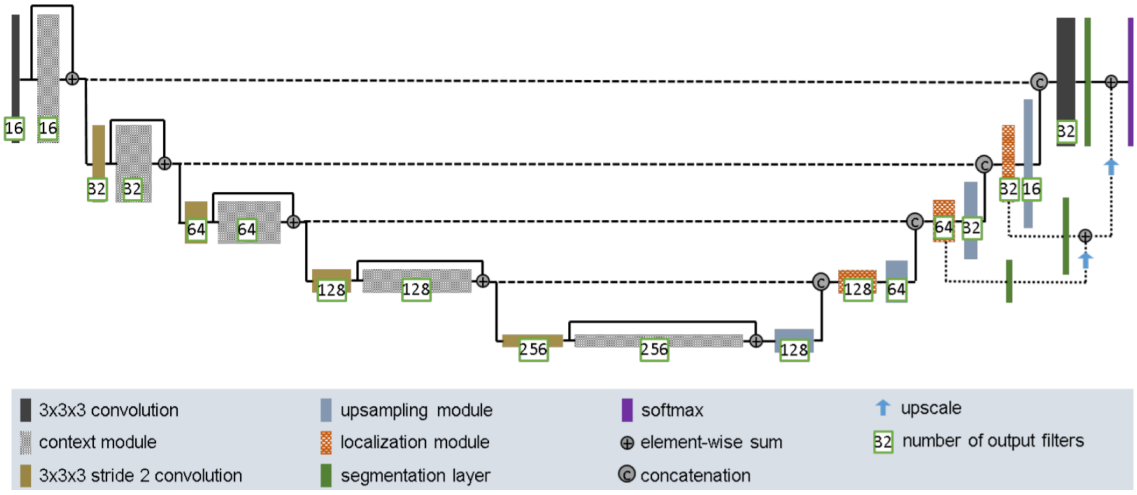


Figure 29 Network architecture from Isensee et al. [101].

Decoder

The decoder or localization pathway combines the high level representations extracted by the encoder with shallower features to reconstruct the output image at full resolution. The resolution is increased with the up-sampling module that consists of a 3D upscale where the feature voxels are repeated twice in each spatial dimension followed by a $3x3x3$ convolution that halves the number of feature maps. This is done instead of the more common approach of deconvolution to prevent checkerboard artefacts [103] in the network output. The up-sampling module is followed by a concatenation layer that recombines the up-sampled features with the corresponding level features of the encoder via concatenation, to generate what are known as 'skip connections'. Skip connections are an important building block of U-Nets that consist in shortcut connections between layers of equal resolution from encoder to decoder, this enables feature propagation between layers to provide high resolution features. Then the localization module which consists of a $3x3x3$ convolution followed by a $1x1x1$ convolution recombines these features together and posteriously halves the number of feature maps. Throughout the network the activation function used is leakyRELU with a negative slop of 10^{-2} for all convolutions computing feature maps.

Following the work of [104], segmentation layers are integrated at different levels of the decoder and combined to form the final output of the network in a method known as deep supervision. Kayalibay et al. [104] used this idea in the original FCN by Long et al.[10] to reduce the coarseness of the final segmentation. Since then, several works on 3D medical image segmentation have also reported creating multiple segmentation maps at different resolutions [105, 106]. As observed in ***Figure 29*** , three segmentation maps are created in this network: one in the final layer therefore having the same

size as the input, one in the second last layer with half the size of the input in each spatial dimension and one in the third last layer with one fourth of the size of the input. These are then combined as follows: first the segmentation map with the lowest spatial resolution is up-sampled by repeating the voxels twice in each dimension to have the same size as the second lowest resolution segmentation map. Then both are added via element-wise summation and the output is up-sampled in the same way and added to the highest resolution map to form the final network output. Since the decoder already receives high resolution features with the encoder-decoder skip connections combining the three segmentation maps does not have the purpose of feature refinement but rather to speed up convergence by "encouraging" early layers of the network to produce good segmentation results [104].

The successive dual 3D U-Net framework proposed by Jia et al. and followed in this project is illustrated in ***Figure 30***. It actually consists of one 3D U-Net that is trained twice with different inputs. Based on the order at which they are trained we will refer to them hereafter as first and second 3D U-Net.

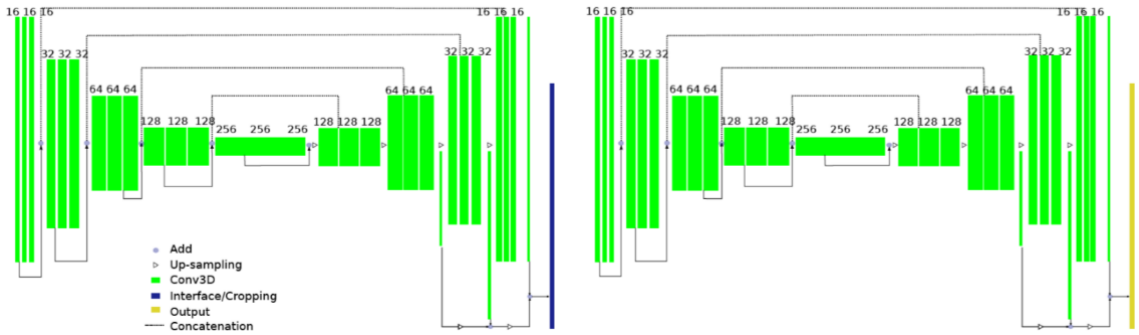


Figure 30 3D Dual U-Net structure proposed by [25]. Green blocks represent 3D features; Dark blue refers to the cropping interface to crop the region of interest of the first U-Net prediction.

The first 3D U-Net is tasked with locating and extracting the region of interest. The input of the first net are MR images resized to size (224,144,96) and its output is preliminary prediction masks of the left atrium. The crop interface keeps the largest connected components of the prediction masks and computes the spatial location of the region of interest (ROI) of the left atrium. The MR images and ground-truth masks are then cropped with a cuboid centred around the left atrium.

The second 3D U-Net performs a secondary training with the cropped images at full resolution. Cropping the images removes background noise and enables for a more precise segmentation. Specifically, the input of the second U-Net are MR images cropped around the predicted LA without resampling of size (224,114,96). The output is a prediction of the LA segmentation of cropped dimensions.

6.2.2 Residual blocks

Similarly to the work of Milletari et al. [14], the design of the U-Net is modified by residual blocks. First introduced by He et al. [107], residual blocks employ special additive skip connections to combat vanishing gradients during backpropagation. More specifically, the data flow is separated into 2 streams at the beginning of the residual block: the first carries the unchanged input of the block and the second goes through the residual block where weights and non-linearities are applied. These two streams are merged at the output of the block by element-wise addition [104]. Technically, the net by Jia et al. employs the modification introduced by He et al. [102] where after the element wise addition

no activation function is used, these are known as pre-residual activation blocks. The proposed variation of the residual block can be observed in Figure 30.

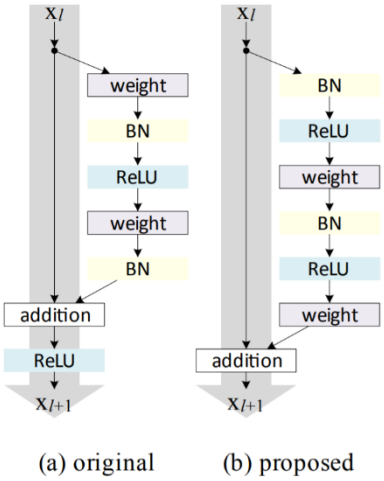


Figure 31 Left: (a) original Residual Unit in [107] ; (b) proposed Residual Unit [102]

6.3 Left and Right Atria Segmentation via 3D Dual U-Net

The main objective of this project is to generate a high generalisation network model capable of obtaining segmentation masks of both RA and LA from MRI images. The segmentation of both atria cavities is of great importance in AF treatment where to plan for ablation procedures a full 3D anatomical model of the atria is required, however most research groups concentrate their efforts in network models specific for LA segmentation only.

As previously seen, to obtain these segmentations a specialized CNN or 3D U-Net is used. These DL algorithms are based on supervised learning techniques, in other words, in order to generate the segmentation they need a set of training data that consists of the raw MRI image to be segmented and an image with the corresponding labels, otherwise known as ground-truth.

In this project the network was trained with different databases of varying image quality and ground-truth segmentations. Firstly, the network was trained with Database 1 to obtain accurate LA segmentation and with Database 2 to obtain both LA and RA prediction segmentations. The learning of LA segmentation from the first network was transferred as positive inference to increase segmentation accuracy. Database 3 was then used to test the generalisation ability of the network trained with Database 2 for the segmentation of LA and RA from images of high variability.

6.3.1 Image pre-processing

Before training the network, the raw MRI images must be pre-processed and we must ensure that the ground-truth labels do in fact label the region of interest.

The internal pre-processing of the algorithm proposed by [25] involves a resize to (224,144,96) and a normalization of the raw MRI image by setting the feature map mean to 0. We must highlight that

although all images are resized to have the same input dimension, during validation, the predictions generated by the network have the same size as the original dimensions of the image to maintain the resolution.

As seen in Chapter 5, Database 1 was provided by the 2018 segmentation challenge and the network form Jia et al. which used in this project was specifically created to enter this challenge therefore there was no need to externally pre-process the images more than the network internally already does.

Database 2 consists of 20 patients each containing a raw MRI volume with a label map that contains the manual segmentations of different cardiac cavities such as the left and right ventricle, the left and right atrium, the myocardium, the ascending aorta and the pulmonary artery. Due to the fact that in this project there is only interest in segmenting the atria we extract the left atria mask (label value = 420) and the right atria mask (label value=550) and save them with the raw MRIs in two distinct datasets (Figure 32).

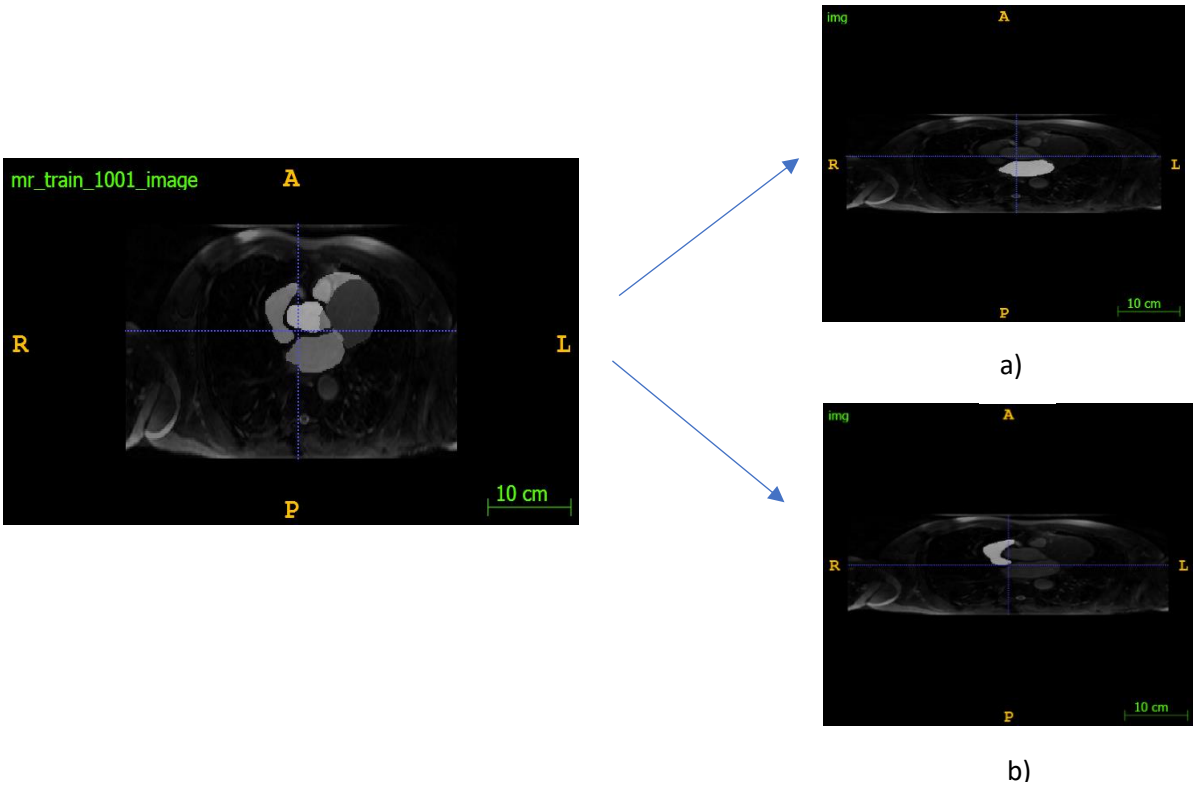


Figure 32 ground-truth label extraction, axial slice. a) LA label extraction b) RA label extraction

In addition, further external pre-processing involved permuting between the second and third dimension of the volume (from [x y z] to [x z y]) so that the 3rd dimension contained the axial slices, and rotating some patient volumes that had a different spatial orientation when the image was compressed in NIFTI format (Figure 33).

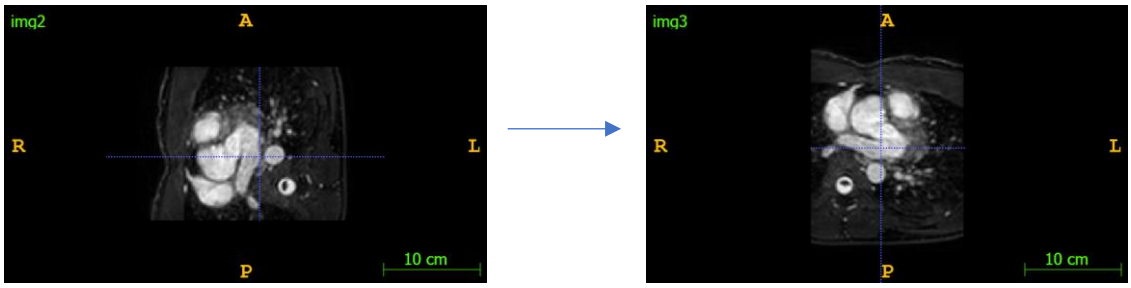
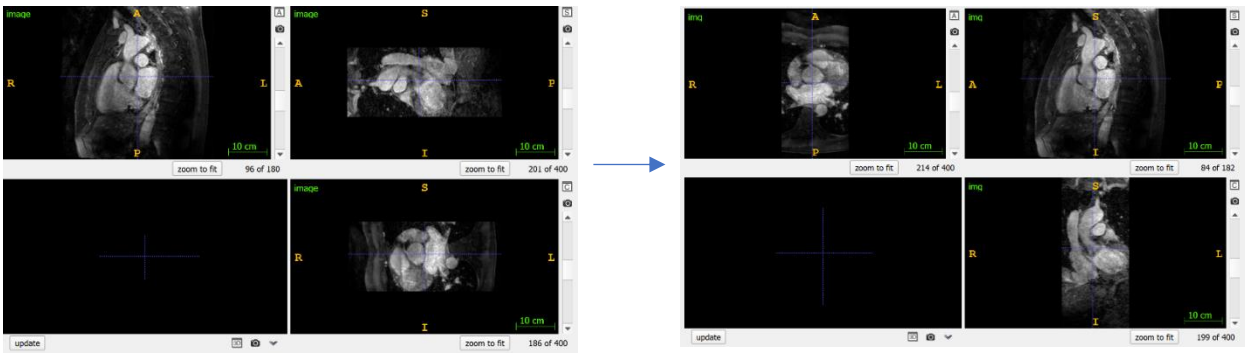


Figure 33 90º rotation. Axial slice

Furthermore, there is a high variability in terms of spatial dimension and slice number between patients (see

Table ***1***). Also, in terms of contrast and size of atria cavities as the 20 patients comprised a wide range of cardiac pathologies. Therefore, with this database the objective was to test the generalisation ability of the network in order to provide an accurate segmentation of LA and RA from MRI images of high variability.

Database 3 consists of 30 patients of varying image qualities. The images were pre-processed to have the same spatial orientation as the other databases. This was done with a MATLAB function in a series of volumetric rotations. Figure 34 shows the transformation which all images underwent.



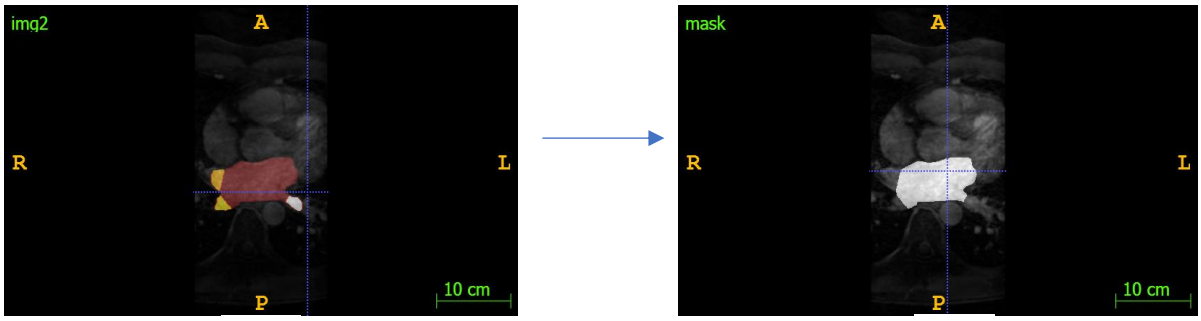
(a) (b)

Figure 34. (a) original orientation; axial view: y-dimension, coronal view: x-dimension, sagittal view: z-dimension. (b) transformed orientation; axial view: z-dimension, coronal view: y-dimension, sagittal view: x-dimension. In both images the same slice is viewed.

The dimensions of the volumes are similar except for 2 patients with dimension 182x400x400. The rest of the patients have dimensions that range 92-142x320x320.

In terms of the ground-truth, 10 patients (a1-a10) contained a ground-truth mask with the LA volume segmentation as a binary mask. 20 patients (b1-20) contained a ground-truth with other structures also labelled such as the pulmonary veins and the left appendage.

The pre-processing consisted in extracting other ground-truths segmentations to remain only with the body of the LA (Figure 35).



(a) (b)

Figure 35 Patient b006. (a) Original ground-truth, label red: LA, other colours: PV and LAA; (b) ground-truth of body volume LA.

6.3.2 Data partitioning and data augmentation

After pre-processing and ground-truth generation it is fundamental to determine the set of data that will be used for training. In the first place, to guarantee the robustness of the model, the dataset must be partitioned into a training set with which the network will be trained, and a test set to test the resulting predictive model. This partition is fundamental because if the same data is used for training and for testing the results obtained will be mistakenly good as the model has learnt the training dataset however what needs to be tested is the performance on a different sample.

The first partition divides the dataset intro training and test. This is done in a proportion of 80% training and 20% testing. Therefore, for the Database 1 the training dataset consists of 80 patients and 20 for the test set. For the Database 2 the training dataset consists of 16 patients and the test subset of 3 patients. Database 3 was not divided as it was only used for testing.

The training set then outgoes a validation split performed by the algorithm that randomly splits the training set in 80% training and 20% validation. The validation split is important to monitor the network´s performance throughout the training process. The random data split changes every time the algorithm is run therefore avoiding biased segmentation results.

Due to the small size of samples of the dataset used the model tends to 'memorise' the training data which limits the ability of the model to generalise when new data samples are passed through the network, this is known as overfitting. To prevent this, a technique known as data augmentation is used to increase the number of new image samples during training. Augmentation in medical imaging normally involves applying small geometrical transformations during training to create variety. In the Jia et al. algorithm these transformations involve rotations, height, and width translation shifts, horizontal and vertical flips and zoom. In this project we experiment with varying data augmentation options to generate various models [108].

6.3.3 Training and evaluation

The network was trained several times, for each training experiment the training parameters from [25] were kept the same. These hyperparameters consist of an initial learning rate equal to $5e^{-4}$ that reduces by half after 10 epochs when validation loss does not improve. After 50 epochs, convergence is defined as not improving and the maximum number of epochs is 500 with 200 steps per epoch. The input of

the network are normalized MR volumes resized to (224,144,96). The loss function employed is the previously described dice loss function to avoid the class imbalance issues. This function is computed with the Softmax output of the last layer of the network.

Each training process in the dual framework consists of an initial training and evaluation which provides a first prediction output that serves to identify the region of interest (ROI). This ROI is then extracted from the original image and ground-truth with an external cropping algorithm. The algorithm designed reads the prediction mask from each patient and by keeping the largest connected component returns the spatial coordinates of a 3D bounding box that represents the smallest cube containing the LA volume. Then the coordinates are stored in a vector and are used to crop the raw image and ground-truth volume of their corresponding patient. This is a sort of pre-processing done in this framework. The final model is obtained by training with the ROI cropped images and ground-truth. Once the network is trained the evaluation process generates the final prediction outputs from the testing images passed through the network. The evaluation metric used is the Dice coefficient. The dual framework proposed by [25] is represented in Figure 36.

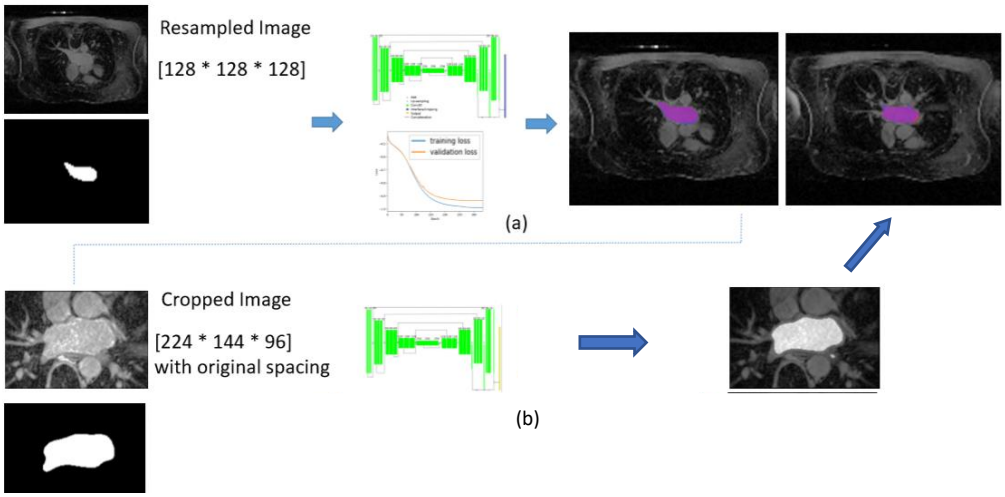


Figure 36 Successive U-Net training framework. a) the first U-Net for cropping b) the second U-Net for segmenting. Modified from [25].

The training and evaluation processes made throughout this project can be divided into 3 experiments.

In Experiment 1 the network was trained with images from Database 1 (80 volumes) and evaluated with testing images also from Database 1 (20 volumes). As previously stated, Database 1 was provided by the same challenge that the network of [25] was originally designed for therefore there was no need to further pre-process the images. Once the final prediction output was obtained it was evaluated with the Dice coefficient and the cropped predictions were post processed. Database 1 contains ground-truths of LA segmentation therefore the output prediction generated were LA binary masks. Due to the large number of training samples this was considered a high-performance model.

Figure 37 illustrates the steps taken in this experiment where N1 refers to the trained model obtained.

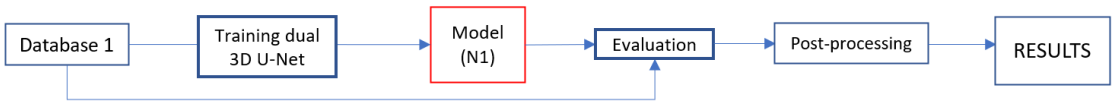


Figure 37 Flow chart Experiment 1.

In Experiment 2 the network was retrained (16 volumes) and tested (3 volumes) with data from Database 2 and the model N2 was obtained. As previously stated, Database 2 contains ground-truth of both LA and RA masks which had to be separated during pre-processing therefore the training was performed separately for each cardiac structure. After the RA and LA segmentation outputs were obtained in the evaluation, they were post processed and 3D reconstructed. As previously stated, Database 2 is a database with high inter patient variability therefore by training the network with this database the aim was to obtain a model of higher generalisation capacity.

Figure 38 illustrates the sequential processes followed in the first part of Experiment 2.

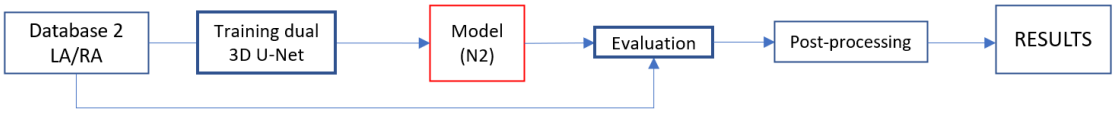


Figure 38 Flow chart Experiment 2.

Then the network was retrained again for LA and RA with the implementation of deep fine-tuning technique obtaining model N3. The weights of all convolutional layers were initialised with pretrained weights obtained from the high-performance model N1 as a positive inference in the learning of the network for LA and RA segmentation. The testing was also performed with images from Database 2 and the prediction outputs were postprocessed and 3D reconstructed.

Figure 39 details the sequential processes followed in the second part of Experiment 2.

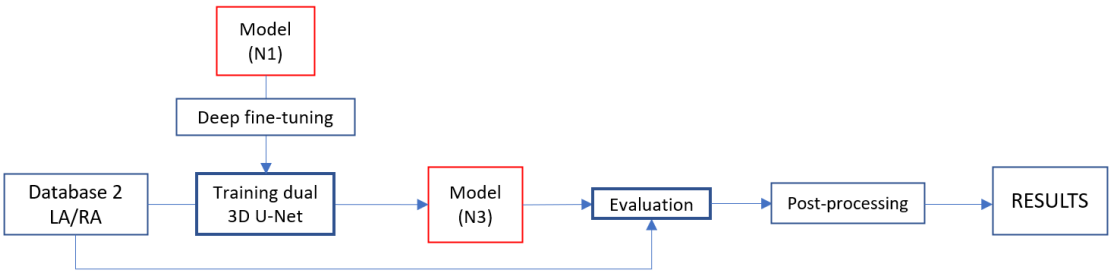


Figure 39 Flow chart Experiment 2.

In Experiment 3 the trained models obtained in the previous experiments: N1, N2 and N3 were evaluated with images from Database 3. Database 3 contains ground-truth of only the LA segmentation therefore the N2 and N3 models for RA were excluded from this experiment. 30 volumes were passed as testing images and the prediction output was post processed and 3D reconstructed. The aim was to test the generalisation capacity of the previously trained networks when they were evaluated, without retraining, with an external database.

Figure 40 illustrates the sequential processes followed in Experiment 3.

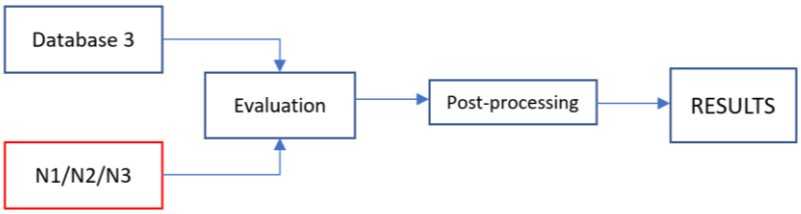


Figure 40 Flow chart Experiment 3.

6.3.4 Post processing

After the training and testing of the algorithm the segmentation predictions of the cropped LA and RA are obtained. Prior to the spatial reintegration of the cropped images a morphological filter was applied to the predictions to smoothen the edges and fill the holes of the anatomical inconsistencies registered in of the resulting predicted masks. These errors are propitiated due to the non-optimal thresholding of the sigmoid function in some pixels.

The morphological filter designed in MATLAB consisted of hole filler command which applies the following operation, defined in Equation 10:

$$f(x) = 255 - Rec(m, 255 - (x)) \quad (3)$$

where $Rec\ (m, 255 - (x))$ refers to the binary reconstruction of the negative of the original mask $(255 - x)$ with a marker m that corresponds to a black image with white edges.

Following this operation, an opening command was applied to smooth the edges based on the following operation:

$$\gamma_B(X) = \delta_B(\ \varepsilon_B(X)) \quad (4)$$

where $\gamma_B(X)$ represents the opening of an image X with a structural element B. This is performed with an erosion ε followed by a dilation δ with the same structural element B. The erosion eliminates thin gulfs and objects smaller than the structural element B and the following dilation recovers the original dimensions of the object therefore smoothing the edges. The structuring element chosen B consist in a sphere with a radius equals to 2.

After the filter was applied, to reintegrate the cropped prediction to the original spatial location with respect to the original dimensions of the image an algorithm was designed in MATLAB. This method stores the coordinates and dimensions of the cuboid extracted in the crop interface in a vector. This vector is then used to position the cropped mask into an empty matrix with the original image dimensions.

This process is illustrated in Figure 41 where the blue cube represents the cuboid computed by the crop interface and the red cube the cropped prediction mask.

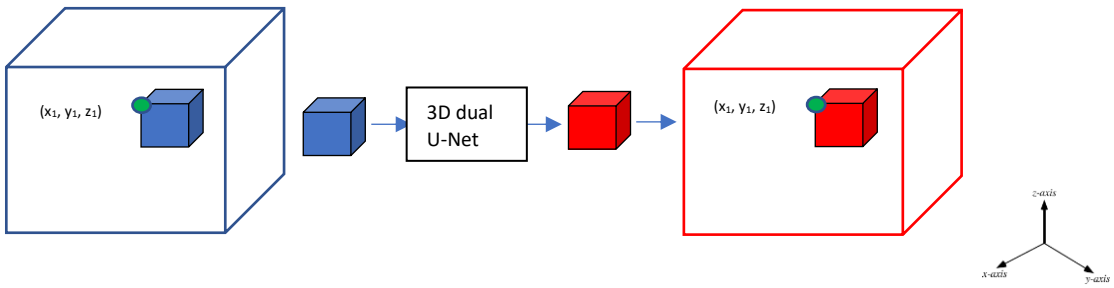


Figure 41 Spatial positioning of LA and RA cropped prediction masks to original dimension mask. The blue cube represents the cuboid in the original prediction and the red cube represents the cropped prediction acquired.

This process is performed twice, both for LA and RA prediction masks. It is of interest to spatially combine both segmentation in one 2D image. For this purpose, another algorithm was designed in MATLAB for assigning label equals 1 to LA and label equals 2 to RA and unify both prediction masks. As a result, we spatially combine both masks in one 2D image where the LA and RA are differentiated by different label values and the pixels which overlap have label equals3. Figure 42 illustrates the post processing pipeline.

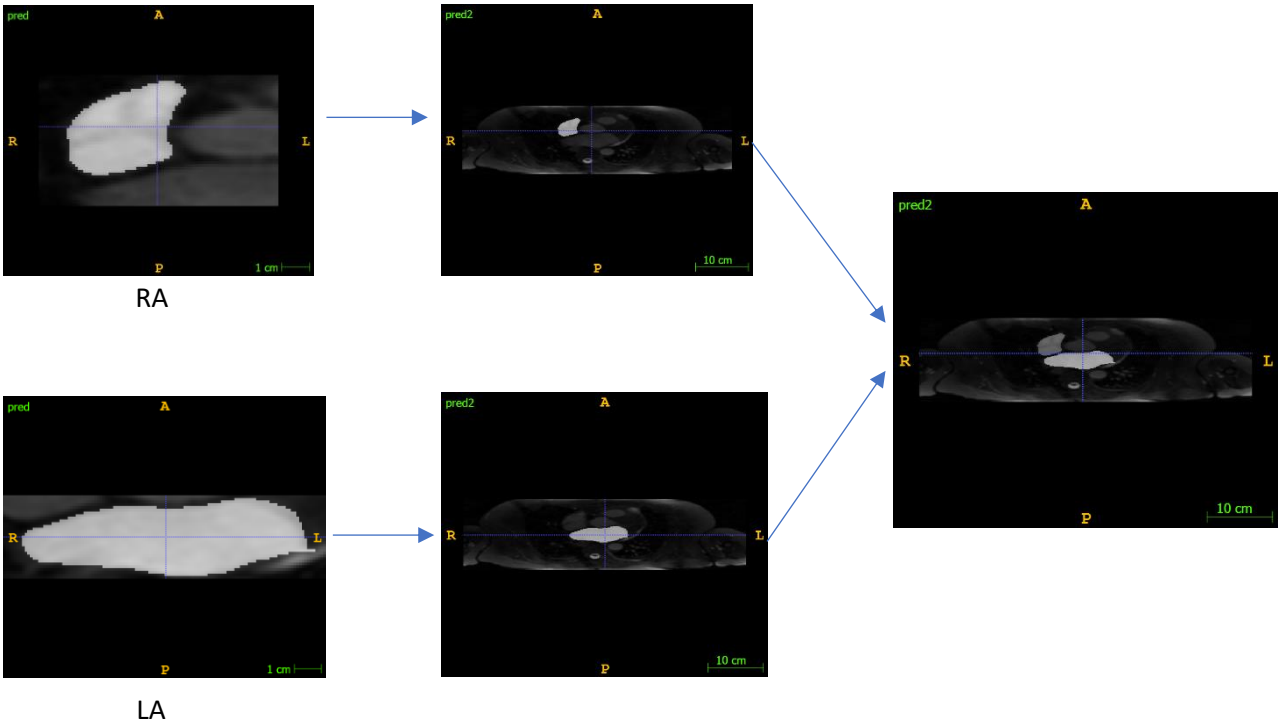


Figure 42 Post processing steps. Database 2

6.3.5 3D Atrial reconstruction

Once we have the 2D volume containing the prediction masks of both the left and right atria for application purposes it is interesting to visualize the 3D volume. For this we employ the Visualization Toolkit (VTK) extension in Python. VTK is an open-source object-oriented software system for 3D image processing and visualization. [109]

To visualize the data in VTK an algorithm was developed in Python following the pipeline in Figure 43. The source/reader step refers to the data loading and arrangement. Once the slices have been

arranged a filter is defined that customizes the data by adding colour, opacity, and transparency parameters to the volume property. Then the mapper volume is defined which maps the input data with graphic geometries that can be displayed by the rendered. Following that, an actor must be created that represents the volume object integrating the volume property and mapper previously defined. The rendering class then converts the 3D graphics geometry form the volume actor together with light properties into a 2D image that can be displayed on the screen. Finally, the render window creates a window for renderers to draw into and the interactor provides window interaction via mouse and keyboard.

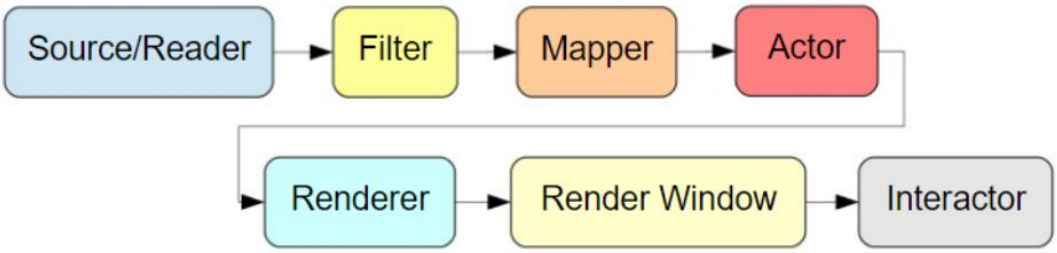


Figure 43 VTK pipeline. [110]

Chapter 7. Results and Discussion

As previously mentioned, the automatic segmentation algorithms trained in this project must be validated by testing their performance. For this the prediction masks obtained from the testing images were compared with the ground-truth provided in the database. As their name implies the ground-truth is a binary segmentation mask considered as the 'truth' of the anatomical segmentation. In this project the ground-truth from all databases were manually generated by experts. Therefore, by comparing the prediction output of the network with the ground-truth a measure of the performance of the network can be obtained.

There are different ways in which the binary prediction mask and ground-truth can be compared. In this project the metric used was the Dice similarity coefficient. The Dice coefficient is a statistical index that compares the similarity between two samples based on their spatial overlap. It can therefore be used to assess the pixelwise similarity between a prediction (p) and ground-truth (g) binary mask [111].

The formula of the dice coefficient is defined in Equation 12:

$$DC = \frac{2\sum_i^N (p_i\, g_i)}{\sum_i^N p_i + \sum_i^N g_i + \sigma} \tag{5}$$

Where i represents each pixel of the volumetric sample and σ a smoothing constant.

The dice coefficient is obtained for each testing patient and the mean and standard deviation are computed to represent the performance of the segmentation.

In the following section the quantitative and qualitative results obtained for each experiment along with the discussion will be presented. The quantitative results refer to the Dice coefficient obtained from the predicted segmentations and the qualitative results refer to 2D and 3D reconstructions of the predictions and ground-truth.

7.1 Experiment 1

As previously described in section 6, the first experiment consists of training and testing the network with patients from Database 1. Database 1 contains 20 testing volumes each of them composed of 88 slices.

7.1.1 Quantitative results

Table 2 *Dice coefficient of predicted segmentations for the testing subset of Database 1.* presents the average Dice coefficient of the 20 patients for testing along with the standard deviation. The results presented are computed with the final prediction output of the model.

Mean ± σ
0.9155 ± 0.0270

Table 2 Dice coefficient of predicted segmentations for the testing subset of Database 1.

7.1.2 Qualitative results

In Figure 42 the axial slices of 3 different patients from Database 1 can be observed, along with the superimposed binary mask and prediction obtained.

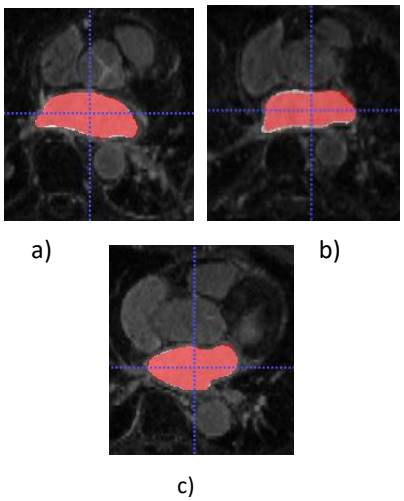


Figure 42. a) Patient 1, slice=42, Dice=0.9458 b) Patient 2, slice=54, Dice= 0.8962 c) Patient 3, slice=49, Dice=0.9060. Axial slices of MRI images overlapped with ground-truth of LA segmentation in white, prediction mask in red and intersection of the two in pink.

7.1.3 Discussion

As can be observed in Table 2 Dice coefficient of predicted segmentations for the testing subset of Database 1.the first experiment yields very accurate segmentation results. This was to be expected as Database 1 corresponds to the data provided by the 2018 LA segmentation challenge and the network proposed by [25] was specifically designed to solve this challenge. Therefore, by keeping the same learning parameters, the dice obtained coincides with the dice reported in [25] (0.91-0.92). Furthermore, the network was trained with a relatively large number of patients which added to the good performance of the network. This can be observed in Figure 43 where the ground-truth and prediction intersection (pink) is almost complete.

The reason for performing this experiment was firstly to test the capacity of the network to provide accurate segmentations and secondly to obtain a *high-performance* model as a starting point for LA segmentation through fine-tuning techniques.

7.2 Experiment 2

In the first part of this experiment the network was trained and tested with Database 2. This database consists of 16 volumes for training and 3 volumes for testing. As described in Chapter 3, the ground-truth of this data contained binary masks of both LA and RA. Therefore, the network was trained and tested twice, once for each ground-truth binary mask.

Furthermore, it should be considered that this data is characterized by high variability as the MRI volume were acquired from patients with different cardiac pathologies which may affect the size and shape of the atria. Furthermore, each MRI volume acquired had a different spatial dimension.

In the second part of this experiment the technique of deep fine tuning was carried out. This was done by initialising the weights from all convolutional layers with the pre-trained weights obtained from the network trained in Experiment 1.

7.2.1 Quantitative results

The following tables show the dice coefficient from network predictions with and without fine tuning implementation. Both network trainings were performed twice, once for LA segmentation (***Table 3***) and once for RA segmentation (***Table 4***)

Implementation	Patient 1	Patient 2	Patient 3	Mean ± σ
w/o fine tuning	0.8693	0.8756	0.8736	0.8729 ± 0.0032
with fine tuning	0.8781	0.8949	0.8707	0.8813 ± 0.0124

Table 3 Dice coefficients in LA predicted segmentations.

Implementation	Patient 1	Patient 2	Patient 3	Mean ± σ
w/o fine tuning	0.9162	0.8773	0.8892	0.8942 ± 0.0199
with fine tuning	0.9273	0.9043	0.9166	0.9160 ± 0.0115

Table 4 Dice coefficients in RA predicted segmentations.

7.2.2 Qualitative results

Figure 44 shows an axial view of the prediction segmentations and ground-truths superposition of LA and RA structures for two patients. Each row represents a different patient and the columns the implementation or not of fine tuning during the training of the network

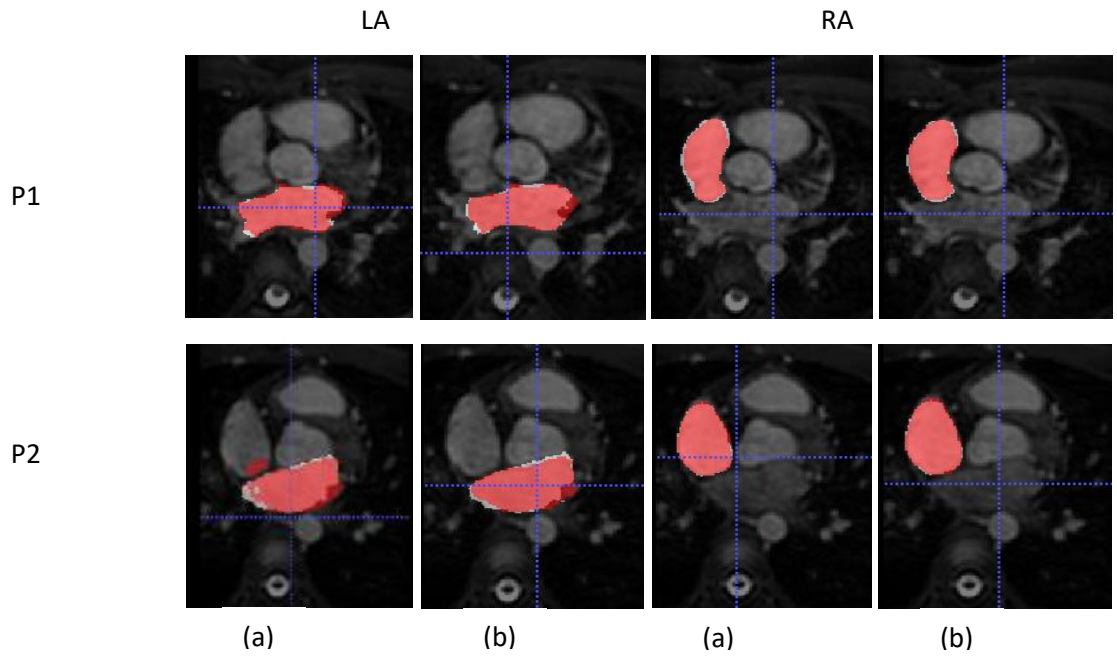


Figure 44 LA and RA segmentations for patient 1 (P1) and patient 2(P2). Axial slice=175. ground-truth mask in white, prediction mask in red and intersection in pink. (a) Prediction from network w/o fine tuning. (b) Prediction from network with fine tuning. (b) Prediction from network with fine tuning.

Furthermore, to qualitatively assess the geometry of the atria structures the left and right atria prediction masks where 3D reconstructed. ***Figure 45*** show the 3D reconstruction of LA and RA for patients 1 and 2 in three different views. Prior to the reconstruction a 3D morphological filter of structural element size 2 was applied, to smoothen the edges.

P1

P2

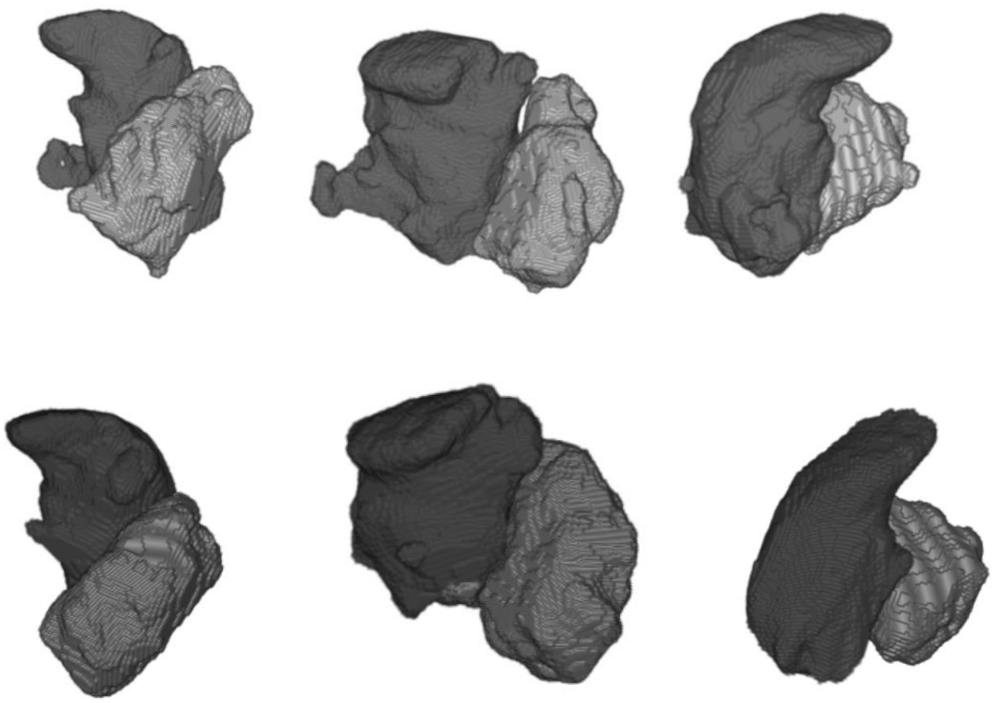


Figure 45 3D reconstruction with filter of patient 1 and 2. LA shown in light grey and RA in dark grey.

7.2.3 Discussion of Experiment 2 results

In this experiment, due to the limited number of data in Database 2 only 3 volumes were used to evaluate the resulting models, therefore, the results obtained serve only as a first view of the performance of these networks. An increased amount of testing samples would be needed to obtain more representative results.

In the first place, by training the network with Database 2 the aim was to obtain a segmentation model of both atria cavities as Database 1 didn´t provide RA ground-truth masks. With the model from scratch it was of interest to test the ability of the network to generate accurate segmentations when trained with small number of samples. Afterwards, it was considered convenient to apply deep fine-tuning to transfer the knowledge of LA segmentation from Experiment 1 network as a positive inference in the segmentation of RA.

Furthermore, due to the high variability of Database 2 a network trained with this database can be expected to have a high generalisation capacity. This will be further verified in Experiment 3.

From the results obtained in Table 3 and 4 it can be observed that the application of deep fine-tuning does increase the dice coefficient in both cardiac structures. As can be seen, the fine-tuning technique increased the Dice coefficient by 0.01 for LA segmentation and 0.02 for RA segmentation. This may not seem like a huge difference however this increase in accuracy becomes of importance in medical applications where lower accuracy results are associated with increased patient risk.

The results in Table 4 demonstrate that the network could successfully segment the RA. This generalisation capacity of the network is very important for clinical practice where it is of interest that the automatic segmentation algorithms are capable of segmenting different structures. Furthermore, as can be seen in Table 3 and 4, the dice obtained for RA segmentation was even higher than the one

obtained for LA segmentation in both training from scratch the network and fine-tuning the weights from Experiment 1. Upon a first consideration it would be expected that the segmentation results for LA were higher as the network was designed for this purpose, however several things must be previously considered. In the first place, as previously stated, the small testing sample may not be wholly representative of the network´s performance. Secondly the left atrium has a smaller and more complex structure than the right atrium therefore it is to be expected that the network finds a more challenging task the LA segmentation.

Furthermore, the dice coefficients obtained in both implementations were slightly lower than the dice values obtained in Experiment 1. As previously described, Database 2 contains different dimension volumes from patients with varying pathologies that can cause variations in the LA and RA anatomical dimensions. In addition, in the second experiment the network was trained with only 16 volumes whilst in the first experiment it was trained with 80, therefore, even with the application of fine tuning more samples would be needed for training to increase the network´s performance.

Nevertheless, the dice obtained in both LA and RA were significantly higher than the mean dice obtained by several research groups reporting results in the segmentation challenge of Database 2 (LA:0.82, RA:0.83) [81] which suggests that fine tuning technique is an effective method for LA and RA segmentation.

In terms of a qualitative evaluation, it is difficult to visually assess images with dice variations in the range of 0.01. Figure 43 shows the slight increase in superposition between prediction mask and ground-truth when the fine-tuning technique is applied for both LA and RA segmentations. This may serve as a qualitative validation of the dice results obtained. However, it must be highlighted that this superposition may vary from slice to slice therefore for a more extensive qualitative analysis more slices must be analysed in each experiment.

In Figure 44 the prediction segmentations are further assessed by means of a 3D reconstruction of the data where for each patient different views are shown to understand the 3D geometry and spatial relationship of both atria. As can be observed the left atrium (light gray) has a smaller and more complex geometry than the right atrium (dark gray) and, this is one of the reasons why it is normally a more challenging structure to segment. The geometry observed refers to the blood pool volume of both structures, other structures such as PVs and LAA in the case of LA cannot be qualitatively assessed.

7.3 Experiment 3

This experiment consisted in an extensive evaluation of the previously trained models using an external database (Database 3): the network trained with Database 1 (N1), the network trained with Database 2 (N2) and the network trained with Database 2 via fine tuning technique (N3). The aim of this experiment was to corroborate the generalization ability of the three models. Database 3 consists of 30 patients with only LA ground-truth therefore the networks trained with RA segmentations were excluded from the following experiments.

7.3.1 Quantitative results

Table ***5*** shows the average dice results obtained from each trained network in the evaluation with Database 3.

Network N	Mean ± σ
N1	0.7773 ± 0.0643
N2	0.8230 ± 0.0943
N3	0.8515 ± 0.0796

Table 5 Average dice for each trained network.

7.3.2 Qualitative results

Figure 46 shows the axial view of predictions obtained for 3 patients of Database 3 with the 3 different trained networks.

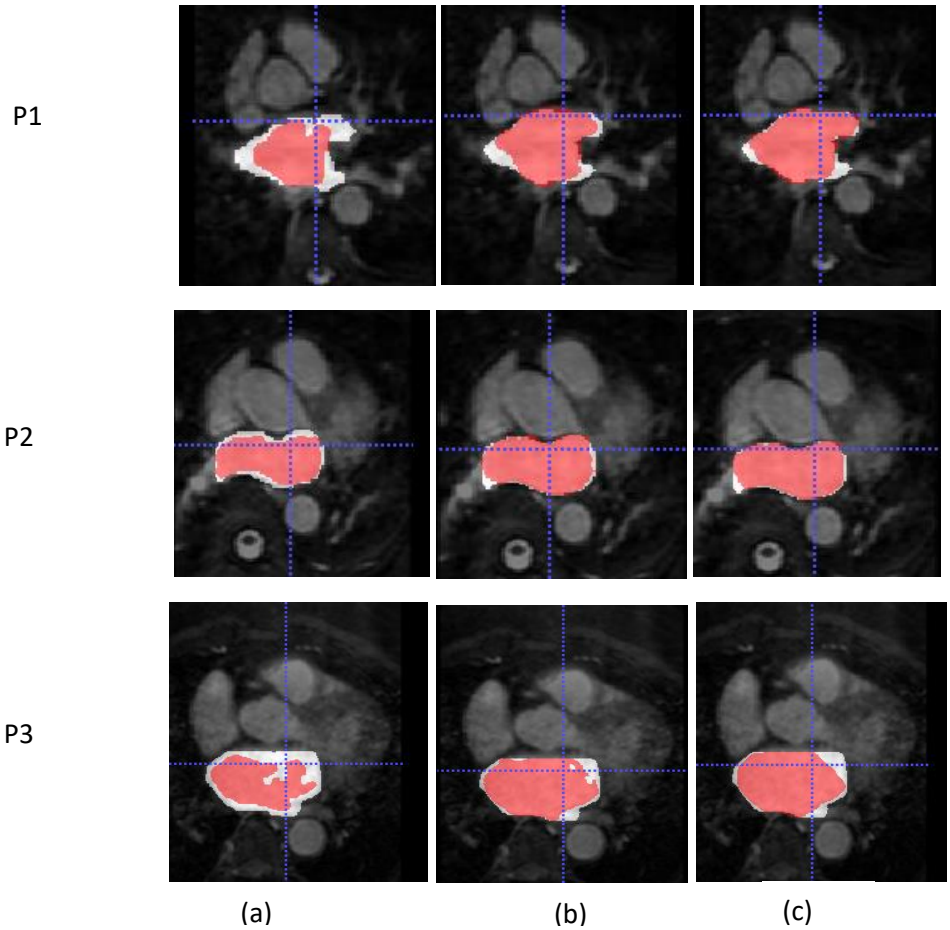


Figure 46 LA segmentations for 3 patients (P1,P2, P3) in networks (a) N1, (b) N2 and (c) N3. ground-truth mask in white, prediction mask in red and intersection in pink.

7.3.3 Qualitative results

Figure 47 illustrates the 3D LA reconstruction from the predictions of the three patients illustrated in Figure 45 along with the 3D reconstruction of the ground-truth.

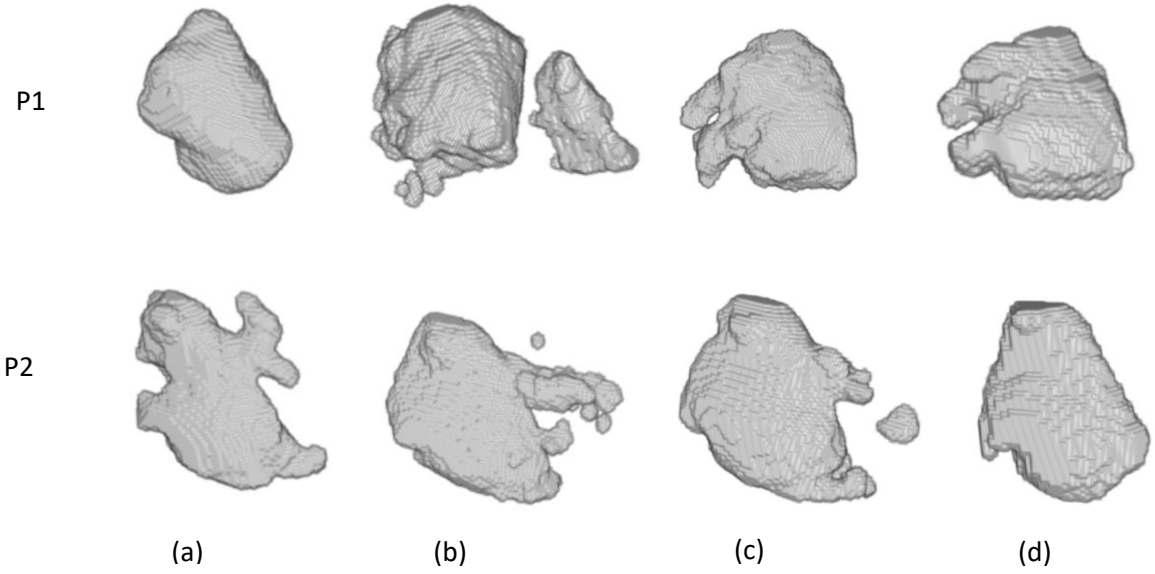


Figure 47 LA 3D reconstructions of predictions for patients P1, P2 and P3 passed through the networks (a) N1, (b) N2 and (c) N3. (d) 3D reconstruction of ground-truth.

7.3.4 Discussion of Experiment 3 results

To compare the generalisation ability of the 3 different trained networks, Database 3 images were inferred as testing through each of them. As can be observed in Table 5 the dice coefficients obtained where lower than in previous experiments, this is to be expected as the networks were not trained with images from this database.

As can be observed in Table 5 the higher dice coefficient was obtained with the network trained with Database 2+fine tuning (N3). This suggests that training a network with highly variable images by fine tuning the weights of a high-performance network could increase the generalisation capacity of the network and result in higher accuracy segmentations when evaluated with external databases. It should be remarked upon that a network trained with 16 volumes (N3) could provide much higher accurate segmentations than a database trained with 80 volumes. This is certainly interesting for medical applications when often the medical data available for training the networks (with ground-truth segmentations) is not extensive.

Additionally, the dice obtained in N3 was slightly lower than the average dice obtained from the research groups participating in the segmentation challenge of Database 3 (0.88) [112]. However, it must be highlighted that the algorithms proposed in the challenge were specifically tailored for Database 3 whilst in this experiment the results are obtained from a generalised trained network. This

is the case for many classical prediction algorithms where they end up overfitted for their specific database with a multitude of parameters fixed ad-hoc for those images.

In terms of a qualitative evaluation, when comparing the 3D reconstruction with the 3D ground-truth, it can be seen that the network trained with Database 1 (N1) was not capable of capturing the geometrical anatomy of the atria whilst N2 and N3 remained closer. Furthermore, in N2 from patient 1 and 2 several inconsistencies can be seen in the reconstruction which are not present when fine tuning is applied. For further anatomical identification a cardiac expert would be required. However, when comparing the geometry of the N3 prediction with the ground-truth several similarities can be seen especially in patient 1 (P1), which leads to believe that it provides a good approximation of the 3D atrial geometry. Furthermore, when compared to the literature [38] it seems like a good basis for a 3D anatomical model which could be used in procedural applications such as catheter ablation.

In view of the results achieved we think that if the model N3 could be validated with an external database containing RA ground-truths also and its performance could be compared with N1 and N2 then we could write an article for CASEIB and CARS.

Chapter 8. Conclusion

The main objective of this project was to experiment with state-of-the art automatic segmentation algorithms, such as deep neural networks, for LA and RA segmentation from MRI images. The aim was to obtain a trained neural network capable of accurately segmenting LA and RA cardiac structures from images of varying quality without the need of retraining. This is of interest in clinical practice where there is a need for automatic segmentation algorithms that accurately segment atrial structures and are easy to implement. LA and RA segmentation is especially useful in the field of AF treatment where atria segmentation for posterior 3D anatomical reconstruction is a widely used approach for guiding ablation procedures. Furthermore, in the field of research it is also useful to investigate the relationship of atrial geometry with mechanisms for sustaining AF.

In order to achieve this, the neural network architecture and learning parameters proposed by Jia et al. were used to perform several segmentation experiments over three different databases. The neural network proposed by Jia et al. consists of a 3D U-Net with residual and skip connections which is currently the state-of-the-art in automatic segmentation algorithms therefore it was of interest to test the performance of this network under different quality databases. Furthermore, the technique of deep fine tuning was performed to increase the segmentation accuracy of the network.

The databases used were provided by 3 different segmentation challenges. The first database was provided by the same challenge for which the network of Jia et al. was designed for therefore there was no need for further pre-processing. However, database 2 and 3 had image volumes of varying orientations and with different ground-truths which had to be dealt with prior to the training and testing of the network. Database 2, for example, was provided by a whole heart segmentation challenge therefore some ground-truth labels had to be extracted in order to generate the LA and RA ground-truth. Furthermore, this database contained images with different volume dimensions and from patients with varying pathologies which affected the volume of atria structures therefore it was considered as a database of high variability.

Once the images were pre-processed, following the dual stage framework proposed by Jia et al., for each training experiment an initial training was performed to obtain a first prediction from which the ROI could be extracted. This was used to crop the images and ground-truths and the cropped data was passed as input to a second training which obtained the final output segmentation.

The dual stage training process of the network was performed three times thus obtaining three different trained models. Once with images from the first database, the second with images from the second database for both LA and RA, and the third with images from the second database carrying out a deep fine tuning from the first model. This technique consisted of initialising the weights from all convolutional layer with were pretrained weights obtained from the first trained network.

The aim of training the network with Database 2, a highly variable database, was to test the ability of the network to segment a different cardiac structure, the RA, and to obtain a network with increased generalisation capacity that was able to accurately segment images from an external database without further retraining. This was tested with the third database. The third database contained LA segmentation masks and MRI volumes from 30 patients. This database was passed through the three previously described trained networks as testing images to test their ability to perform accurate segmentation.

For every experiment, the segmentation accuracy was evaluated with the dice coefficient. Once the training and testing were performed the predictions masks where postprocessed with a morphological filter to smoothen the edges and the cropped masks were relocated to their original dimension space to be superimposed with the original image for a qualitative evaluation, Finally, the predictions were assessed qualitatively with a 3D reconstruction using the VTK pipeline.

The results obtained show that a network trained with a highly variable database is indeed more capable of accurately segmenting an external database without retraining. Furthermore, with the additional use of fine tuning the segmentation accuracy can be improved. However, to further validate these results the network should be tested with more external databases. In addition, the third database only had LA ground-truth masks therefore further testing with RA ground-truths is required in order to test the capability of the network in segmenting a different cardiac structure without retraining.

Nevertheless, these results are promising in the field of image segmentation and are a first step towards the integration of automatic segmentation algorithms in the clinical practice.

Chapter 9. Future Work

The future line of work for medical image segmentation algorithms is to improve their generalisation capacity as they are normally tailored for a specific database which in clinical practice has few applications. Clinicians require accurate automatic segmentation model that are capable of segmenting images of various contrast and image quality.

The future line of work for this project would be to further validate the trained models obtained with external databases. It would be especially interesting if these databases contained RA ground-truth segmentations to test their ability of segmenting a different cardiac structure without retraining, especially for the trained network of database 2 with fine tuning which shows the most promising results for LA segmentation. It would be undoubtedly interesting to have generated a model capable of accurately segmenting the whole atria without retraining, however more data samples are needed to verify this.

In light of the promising results of applying deep fine-tuning techniques, more techniques should be reviewed and applied in this area to aim for a higher segmentation accuracy model.

The aim would be that this generalised model, once validated, could be integrated in the clinical practice for atria segmentation as a pipeline for AF management and treatment with the generation of 3D anatomical models.

Chapter 10. References

[1] M. Zoni-Berisso, F. Lercari, T. Carazza, and S. Domenicucci, “Epidemiology of atrial fbrillation: European perspective,” *Clin. Epidemiol.*, vol. 6, no. 1, pp. 213–220, 2014.

[2] G. Y. H. Lip, P. Kakar, and T. Watson, “Atrial fibrillation - The growing epidemic,” *Heart*, vol. 93, no. 5, pp. 542–543, 2007.

[3] L. Li *et al.*, “Atrial scar quantification via multi-scale CNN in the graph-cuts framework,” *Med. Image Anal.*, vol. 60, p. 101595, 2020.

[4] Y. Zheng, T. Wang, M. John, S. K. Zhou, J. Boese, and D. Comaniciu, “Multi-part Left Atrium Modeling and Segmentation in C-Arm CT Volumes for Atrial Fibrillation Ablation,” in *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2011*, 2011, pp. 487–495.

[5] S. Chen, T. Kohlberger, and K. J. Kirchberg, “Advanced level set segmentation of the right atrium in MR,” in *Medical Imaging 2011: Visualization, Image-Guided Procedures, and Modeling*, 2011, vol. 7964, pp. 906–911.

[6] T. Kurzendorfer, C. Forman, M. Schmidt, C. Tillmanns, A. Maier, and A. Brost, “Fully automatic segmentation of left ventricular anatomy in 3-D LGE-MRI,” *Comput. Med. Imaging Graph.*, vol. 59, pp. 13–27, 2017.

[7] et al. Litjens G, Kooi T, Bejnordi BE, “A survey on deep learning in medical image analysis,” *Med Image Anal.*, vol. 42, pp. 60–88, 2017.

[8] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, “Deep neural networks segment neuronal membranes in electron microscopy images,” *Adv. Neural Inf. Process. Syst.*, vol. 4, pp. 2843–2851, 2012.

[9] M. Avendi, A. Kheradvar, and H. Jafarkhani, “Fully automatic segmentation of heart chambers in cardiac MRI using deep learning,” *J. Cardiovasc. Magn. Reson.*, vol. 18, no. S1, pp. 2–4, 2016.

[10] J. Long, E. Shelhamer, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, 2017.

[11] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9351, pp. 234–241, 2015.

[12] B. Norman, V. Pedoia, and S. Majumdar, “Use of 2D U-net convolutional neural networks for automated cartilage and meniscus segmentation of knee MR imaging data to determine relaxometry and morphometry,” *Radiology*, vol. 288, no. 1, pp. 177–185, 2018.

[13] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-net: Learning dense volumetric segmentation from sparse annotation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9901 LNCS, no. June 2016, pp. 424–432, 2016.

[14] F. Milletari, N. Navab, and S. A. Ahmadi, “V-Net: Fully convolutional neural networks for volumetric medical image segmentation,” *Proc. - 2016 4th Int. Conf. 3D Vision, 3DV 2016*, pp. 565–571, 2016.

[15] Q. Tao *et al.*, “Deep learning-based method for fully automatic quantification of left ventricle function from cine MR images: A multivendor, multicenter study,” *Radiology*, vol. 290, no. 1, pp. 81–88, 2019.

[16] F. Isensee, P. F. Jaeger, P. M. Full, I. Wolf, S. Engelhardt, and K. H. Maier-Hein, "Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10663 LNCS, pp. 120–129, 2018.

[17] Q. Xia, Y. Yao, Z. Hu, and A. Hao, "Automatic 3D Atrial Segmentation from GE-MRIs Using Volumetric Fully Convolutional Networks," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11395 LNCS, pp. 211–220, 2019.

[18] R. Karim, R. Mohiaddin, and D. Rueckert, "Left atrium segmentation for atrial fibrillation ablation," *Med. Imaging 2008 Vis. Image-guided Proced. Model.*, vol. 6918, no. 69182, p. 69182U, 2008.

[19] Q. Tao, R. Shahzad, E. G. Ipek, F. F. Berendsen, S. Nazarian, and R. J. van der Geest, "Fully automated segmentation of left atrium and pulmonary veins in late gadolinium enhanced MRI," *J. Cardiovasc. Magn. Reson.*, vol. 18, no. S1, pp. 1–3, 2016.

[20] X. Zhuang, K. S. Rhode, R. S. Razavi, D. J. Hawkes, and S. Ourselin, "A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI," *IEEE Trans. Med. Imaging*, vol. 29, no. 9, pp. 1612–1625, 2010.

[21] W. Bai *et al.*, "Automated cardiovascular magnetic resonance image analysis with fully convolutional networks," *J. Cardiovasc. Magn. Reson.*, vol. 20, no. 1, pp. 1–12, 2018.

[22] A. Zhaohan Xiong, Vadim V. Fedorov, Xiaohang Fu, Elizabeth Cheng, Rob Macleod and J. Zhao*, "Fully Automatic Left Atrium Segmentation from Late Gadolinium Enhanced Magnetic Resonance Imaging Using a Dual Fully Convolutional Neural Network," *IEEE Trans Med Imaging*, vol. 38, no. 2, pp. 515–524, 2019.

[23] C. Bian *et al.*, "Pyramid Network with Online Hard Example Mining for Accurate Left Atrium Segmentation: 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers," 2019, pp. 237–245.

[24] A. Mortazi, R. Karim, K. Rhode, J. Burt, and U. Bagci, "CardiacNET: Segmentation of left atrium and proximal pulmonary veins from MRI using multi-view CNN," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10434 LNCS, pp. 377–385, 2017.

[25] S. Jia *et al.*, "Automatically Segmenting the Left Atrium from Cardiac Images Using Successive 3D U-Nets and a Contour Loss," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11395 LNCS, pp. 221–229, 2019.

[26] S. Vesal, N. Ravikumar, and A. Maier, "Dilated Convolutions in Neural Networks for Left Atrial Segmentation in 3D Gadolinium Enhanced-MRI," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11395 LNCS, pp. 319–328, 2019.

[27] and P. A. H. Caizi Li, Qianqian Tong, Xiangyun Liao, Weixin Si, Yinzi Sun, QiongWang, "Attention Based Hierarchical Aggregation Network for 3D Left Atrial Segmentation: 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11395, no. January, pp. 302–310, 2019.

[28] J. E. Hall and G. A. C., *Guyton and Hall Textbook of Medical Physiology (Guyton Physiology)*. 2015.

[29] S. B. Olsson, "Atrial fibrillation - Where do we stand today?," *J. Intern. Med.*, vol. 250, no. 1, pp. 19–28, 2001.

[30] D. Sánchez-Quintana, G. Pizarro, J. R. López-Mínguez, S. Y. Ho, and J. A. Cabrera, “Standardized review of atrial anatomy for cardiac electrophysiologists,” *J. Cardiovasc. Transl. Res.*, vol. 6, no. 2, pp. 124–144, 2013.

[31] J. A. Cabrera, F. Saremi, and D. Sánchez-Quintana, “Left atrial appendage: Anatomy and imaging landmarks pertinent to percutaneous transcatheter occlusion,” *Heart*, vol. 100, no. 20, pp. 1636–1650, 2014.

[32] J. Zhao *et al.*, “Three-dimensional integrated functional, structural, and computational mapping to define the structural ‘fingerprints’ of heart-specific atrial fibrillation drivers in human heart ex vivo,” *J. Am. Heart Assoc.*, vol. 6, no. 8, 2017.

[33] Hall and J. Edward, *Guyton and hall textbook of medical physiology thirteenth edition*. 2011.

[34] J. Malmivuo and R. Plonsey, *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. 2012.

[35] and J. E. O. Lesh, M.D., J.M.Kalman, “An electrophysiologic approach to catheter ablation of atrial flutter and tachycardia: frommechanism to practice,” in *Interventional Electrophysiology*, Baltimore: William and Wilkins, 1997, pp. 347–382.

[36] P. Jaïs *et al.*, “Mapping and ablation of left atrial flutter,” *Circulation*, vol. 101, pp. 2928–293, 2000.

[37] F. García Cosío, A. Pastor, A. Núñez, A. P. Magalhaes, and P. Awamleh, “Atrial flutter: An update,” *Rev. Esp. Cardiol.*, vol. 59, no. 8, pp. 816–831, 2006.

[38] A. J. Moe GK, Rheinboldt WC, “A computer model of atrial fibrillation,” *Am Hear. J*, vol. 67, pp. 200–220, 1964.

[39] M. Haïssaguerre *et al.*, “Spontaneous Initiation of Atrial Fibrillation by Ectopic Beats Originating in the Pulmonary Veins,” *N. Engl. J. Med.*, vol. 339, no. 10, pp. 659–666, 1998.

[40] J. Jalife, “Rotors and Spiral Waves in Atrial Fibrillation,” *J. Cardiovasc. Electrophysiol.*, vol. 14, no. 7, pp. 776–780, Jul. 2003.

[41] J. A. B. Zaman and S. M. Narayan, “Ablating Atrial Fibrillation: Customizing Lesion Sets Guided by Rotor Mapping,” *Methodist Debakey Cardiovasc. J.*, vol. 11, no. 2, pp. 76–81, 2015.

[42] R. Arora *et al.*, “Arrhythmogenic substrate of the pulmonary veins assessed by high-resolution optical mapping,” *Circulation*, vol. 107, no. 13, pp. 1816–1821, 2003.

[43] M. Wijffels, C. Kirchhof, R. Dorland, and M. Allessie, “Atrial Fibrillation Begets Atrial Fibrillation,” *Circulation*, vol. 92, pp. 1954–1968, 1995.

[44] R. F. Bosch, X. Zeng, J. B. Grammer, K. Popovic, C. Mewis, and V. Kühlkamp, “Ionic mechanisms of electrical remodeling in human atrial fibrillation,” *Cardiovasc. Res.*, vol. 44, no. 1, pp. 121–131, 1999.

[45] A. J. Workman, K. A. Kane, and A. C. Rankin, “The contribution of ionic currents to changes in refractoriness of human atrial myocytes associated with chronic atrial fibrillation,” *Cardiovasc. Res.*, vol. 52, no. 2, pp. 226–235, 2001.

[46] H. J. Jongsma and R. Wilders, “Gap junctions in cardiovascular disease,” *Circ. Res.*, vol. 86, no. 12, pp. 1193–1197, 2000.

[47] B. Al Ghamdi and W. Hassan, “Atrial remodeling and atrial fibrillation: Mechanistic interactions and clinical implications,” *J. Atr. Fibrillation*, vol. 1, no. 7, pp. 395–416, 2009.

[48] D. P. Zipes, J. Jalife, and W. G. Stevenson, *Cardiac Electrophysiology: From Cell to Bedside: Seventh Edition*. 2017.

[49] P. M. Boyle *et al.*, "Computationally guided personalized targeted ablation of persistent atrial fibrillation," *Nat. Biomed. Eng.*, vol. 3, no. 11, pp. 870–879, 2019.

[50] M. Elmaoğlu and A. Çelik, *MRI Handbook*. 2012.

[51] J. Bogaert, S. Dymarkovsky, A. M. Taylor, and V. Muthurangu, Eds., *Clinical cardiac MRI*. .

[52] D. W. McRobbie, E. A. Moore, and M. J. Graves, *MRI from picture to proton*. 2017.

[53] D. J. Pennell *et al.*, "Clinical indications for cardiovascular magnetic resonance (CMR): Consensus Panel report," *Eur. Heart J.*, vol. 25, no. 21, pp. 1940–1965, 2004.

[54] D. T. Ginat, M. W. Fong, D. J. Tuttle, S. K. Hobbs, and R. C. Vyas, "Cardiac imaging: Part 1, MR pulse sequences, imaging planes, and basic anatomy," *Am. J. Roentgenol.*, vol. 197, no. 4, pp. 808–815, 2011.

[55] M. Imai *et al.*, "Multi-ethnic study of atherosclerosis: Association between left atrial function using tissue tracking from cine mr imaging and myocardial fibrosis," *Radiology*, vol. 273, no. 3, pp. 703–713, 2014.

[56] S. Uribe *et al.*, "Whole-heart cine MRI using real-time respiratory self-gating," *Magn. Reson. Med.*, vol. 57, no. 3, pp. 606–613, 2007.

[57] S. Hilbert *et al.*, "Real-time magnetic resonance-guided ablation of typical right atrial flutter using a combination of active catheter tracking and passive catheter visualization in man: Initial results from a consecutive patient series," *Europace*, vol. 18, no. 4, pp. 572–577, 2016.

[58] A. Romfh and F. R. Pluchinotta, "Congenital Heart Defects in Adults : A Field Guide for Cardiologists," *J. Clin. Exp. Cardiolog.*, vol. 01, no. S8, 2012.

[59] A. Doltra, B. Amundsen, R. Gebker, E. Fleck, and S. Kelle, "Emerging Concepts for Myocardial Late Gadolinium Enhancement MRI," *Curr. Cardiol. Rev.*, vol. 9, no. 3, pp. 185–190, 2013.

[60] R. S. Oakes *et al.*, "Detection and Quantification of Left Atrial Structural Remodeling Using Delayed Enhancement MRI in Patients with Atrial Fibrillation," *Circulation*, vol. 119, no. 13, pp. 1758–1767, 2009.

[61] C. McGann *et al.*, "Atrial fibrillation ablation outcome is predicted by left atrial remodeling on MRI," *Circ. Arrhythmia Electrophysiol.*, vol. 7, no. 1, pp. 23–30, 2014.

[62] R. J. Perea *et al.*, "T1 mapping: characterisation of myocardial interstitial space," *Insights Imaging*, vol. 6, no. 2, pp. 189–202, 2015.

[63] J. J. Lee *et al.*, "Myocardial T1 and extracellular volume fraction mapping at 3 tesla," *J. Cardiovasc. Magn. Reson.*, vol. 13, no. 1, pp. 1–10, 2011.

[64] R. J. Kim *et al.*, "Relationship of MRI delayed contrast enhancement to irreversible injury, infarct age, and contractile function," *Circulation*, vol. 100, no. 19, pp. 1992–2002, 1999.

[65] W. KIm, "THE USE OF CONTRAST-ENHANCED MAGNETIC RESONANCE IMAGING TO IDENTIFY REVERSIBLE MYOCARDIAL DYSFUNCTION," pp. 1445–1453, 2000.

[66] M. Shenasa, "Fibrosis and Ventricular Arrhythmogenesis. Role of Cardiac MRI," *Card. Electrphysiology Clin.*, vol. 11, pp. 551–562, 2019.

[67] K. Jamart, Z. Xiong, G. D. Maso Talou, M. K. Stiles, and J. Zhao, "Mini Review: Deep Learning for

Atrial Segmentation From Late Gadolinium-Enhanced MRIs," *Front. Cardiovasc. Med.*, vol. 7, no. May, 2020.

[68] C. J. McGann *et al.*, "New Magnetic Resonance Imaging-Based Method for Defining the Extent of Left Atrial Wall Injury After the Ablation of Atrial Fibrillation," *J. Am. Coll. Cardiol.*, vol. 52, no. 15, pp. 1263–1271, 2008.

[69] F. Bisbal *et al.*, "CMR-guided approach to localize and ablate gaps in repeat AF ablation procedure," *JACC Cardiovasc. Imaging*, vol. 7, no. 7, pp. 653–663, 2014.

[70] L. Sanchis, S. Prat, and M. Sitges, "Cardiovascular Imaging in the Electrophysiology Laboratory," *Rev. Española Cardiol. (English Ed.*, vol. 69, no. 6, pp. 595–605, 2016.

[71] I. M. Khurram *et al.*, "Magnetic resonance image intensity ratio, a normalized measure to enable interpatient comparability of left atrial fibrosis," *Hear. Rhythm*, vol. 11, no. 1, pp. 85–92, 2014.

[72] E. Lin and A. Alessio, "What are the basic concepts of temporal, contrast, and spatial resolution in cardiac CT?," *J Cardiovasc Comput Tomogr*, vol. 3, no. 6, pp. 403–408, 2009.

[73] L. Zhong, R. S. Tan, E. Y. K. Ng, and D. N. Ghista, *Computational and Mathematical Methods in Cardiovascular Physiology*. 2019.

[74] L. Hrvoje and M. W. Greenstaff, *X-Ray Computed Tomography Contrast Agents*, vol. 113, no. 3. 2014.

[75] M. R. M. Jongbloed *et al.*, "Atrial Fibrillation: Multi–Detector Row CT of Pulmonary Vein Anatomy prior to Radiofrequency Catheter Ablation—Initial Experience," *Radiology*, vol. 234, no. 3, pp. 702–709, 2005.

[76] J. Hur *et al.*, "Left atrial appendage thrombi in stroke patients: Detection with two-phase cardiac CT angiography versus transesophageal echocardiography," *Radiology*, vol. 251, no. 3, pp. 683–690, 2009.

[77] A. J. Vorre MM, "Diagnostic Accuracy and Radiation Dose of CT Coronary Angiography in Atrial Fibrillation :," vol. 267, no. 2, 2013.

[78] Y. Zheng, D. Yang, M. John, and D. Comaniciu, "Multi-Part Modeling and Segmentation of Left Atrium in C-Arm CT for Image-Guided Ablation of Atrial Fibrillation," *IEEE Trans. Med. Imaging*, vol. 33, no. 2, pp. 318–331, 2014.

[79] J. M. Sohns *et al.*, "Right Atrial Volume is Increased in Corrected Tetralogy of Fallot and Correlates with the Incidence of Supraventricular Arrhythmia: A CMR Study," *Pediatr. Cardiol.*, vol. 36, no. 6, pp. 1239–1247, 2015.

[80] "http://atriaseg2018.cardiacatlas.org/." .

[81] X. Zhuang *et al.*, "Evaluation of algorithms for Multi-Modality Whole Heart Segmentation: An open-access grand challenge," *Med. Image Anal.*, vol. 58, p. 101537, 2019.

[82] "https://www.cardiacatlas.org/challenges/left-atrium-segmentation-challenge/." .

[83] C. Tobon-Gomez *et al.*, "Benchmark for Algorithms Segmenting the Left Atrium From 3D CT and MRI Datasets," *IEEE Trans. Med. Imaging*, vol. 34, no. 7, pp. 1460–1473, 2015.

[84] "https://www.jetbrains.com/es-es/pycharm/." .

[85] "https://www.python.org." .

[86] “https://keras.io/tle.” .

[87] “https://www.mathworks.com/products/matlab.html.” .

[88] “https://mobaxterm.mobatek.net/.” .

[89] Y. Lecun, L. Bottou, Y. Bengio, and P. Ha, “Gradient Based Learning Applied to Document Recognition,” *Proc. IEEE*, no. November, pp. 1–46, 1998.

[90] “https://towardsdatascience.com/a-comprehensive-introduction-to-different-types-of-convolutions-in-deep-learning-669281e58215No Title.” .

[91] A. F. Agarap, “Deep Learning using Rectified Linear Units (ReLU),” no. 1, pp. 2–8, 2018.

[92] B. Xu, N. Wang, T. Chen, and M. Li, “Empirical Evaluation of Rectified Activations in Convolutional Network,” 2015.

[93] J. Kuen *et al.*, “Stochastic Downsampling for Cost-Adjustable Inference and Improved Regularization in Convolutional Networks,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 7929–7938, 2018.

[94] S. Ruder, “An overview of gradient descent optimization algorithms,” pp. 1–14, 2016.

[95] “https://machinelearningmastery.com/learning-rate-for-deep-learning-neural-networks/o Title.” .

[96] N. Tajbakhsh *et al.*, “Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?,” *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[97] V. Iglovikov and A. Shvets, “TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation,” 2018.

[98] S. Cai, Y. Shu, W. Wang, M. Zhang, G. Chen, and B. C. Ooi, “Effective and Efficient Dropout for Deep Convolutional Neural Networks,” pp. 1–12, 2019.

[99] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *32nd Int. Conf. Mach. Learn. ICML 2015*, vol. 1, pp. 448–456, 2015.

[100] D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Instance Normalization: The Missing Ingredient for Fast Stylization,” no. 2016, 2016.

[101] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, “Brain tumor segmentation and radiomics survival prediction: Contribution to the BRATS 2017 challenge,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10670 LNCS, pp. 287–297, 2018.

[102] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9908 LNCS, pp. 630–645, 2016.

[103] C. Odena, Augustus; Dumoulin, Vincent and Olah, “Deconvolution and Checkerboard Artifacts,” *Distill*, 2016.

[104] B. Kayalibay, G. Jensen, and P. van der Smagt, “CNN-based Segmentation of Medical Imaging Data,” 2017.

[105] H. Chen, Q. Dou, L. Yu, and P.-A. Heng, “VoxResNet: Deep Voxelwise Residual Networks for Volumetric Brain Segmentation,” pp. 1–9, 2016.

[106] Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, and P. A. Heng, “3D deeply supervised network for

automatic liver segmentation from CT volumes," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9901 LNCS, pp. 149–157, 2016.

[107] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016.

[108] Z. Eaton-Rosen, F. Bragman, S. Ourselin, and M. J. Cardoso, "Improving Data Augmentation for Medical Image Segmentation," *1st Conf. Med. Imaging with Deep Learn.*, vol. 10670 LNCS, no. Midl, pp. 450–462, 2018.

[109] "VTK." [Online]. Available: https://vtk.org/.

[110] J. Lin, "Introduction to VTK," *Distrib. Comput.*, vol. 2007, no. March, 2008.

[111] K. H. Zou *et al.*, "Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index," *Acad. Radiol.*, vol. 11, no. 2, pp. 178–189, 2004.

[112] C. Tobon-Gomez *et al.*, "Left Atrial Segmentation Challenge: A unified benchmarking framework," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8330 LNCS, pp. 1–13, 2014.

[113] M. S. H. Al-Tamimi and G. Sulong, "Tumor brain detection through MR images: A review of literature," *J. Theor. Appl. Inf. Technol.*, vol. 62, no. 2, pp. 387–403, 2014.

Budget

Design and development of a system for semantic segmentation of atrial cavity with a neural network architecture encoder-decoder.

Author: Marta Saiz Vivó

Directors: Valery Naranjo Ornedo

Adrian Colomer Granero

July, 2020

Contents

List of Tables

Chapter 1. Budget

In the following section a detailed economic assessment of the research project development is described.

1.1 Partial budget

To quantify the budget this has been divided into three parts: a) Labour costs b) Hardware costs c) Software costs

1.1.1 Labour costs

In this section the labour costs are presented taking into account the human resources needed for the development of this project.

Table 1 shows the labour costs in terms of the remuneration per participant and the time dedicated to the project. Specifically, for the development of this project the following people have participated:

- Valery Naranjo Ornedo, full professor at the Universitat Politecnica de Valencia and director of this project.
- Adrián Colomer Granero, PhD, doctor at the Universitat Politecnica de Valencia and cotutor of this project.
- Marta Saiz Vivó, student of Biomedical Engineering Degree and author of this project.

Nº	Labour description	Units	Quantity	Unit Price	Total
1	*Biomedical engineering student*	h	350	12 €/h	4200.00 €
2	*PhD in charge of tutoring and supervising the work*	h	48	22 €/h	1056.00 €
3	*Professor in charge of tutoring and supervising the work*	h	56	48 €/h	2688.00 €
				TOTAL	**7944.00 €**

Table 1. Labour Costs

1.1.2 Hardware costs

The main part of this project has been developed in a personal computer. However, due to the high computational costs of training a neural network this algorithm has been trained in an external server provided by CVBLab. The network has been trained a total of 5 times with the

different databases. Due to the relatively low number of samples the training of the network was achieved during the span of a day. However, due to the different experiments and techniques tried this server has been used intermittently for 5 months. Furthermore, this server has also been used to store the images generated by the network.

Table 2 shows the hardware costs associated with the personal computer and Table 3 shows the cost of the different components associated with the CVBLab servers. Finally, Table 4 shows the total costs associated with hardware for the development of this project.

Personal computer

Nº	Hardware description	Quantity	Unit cost	Amort. period	Amort. ratio	Total
1	*XPS 13 9360* *Intel®Core™ i5-7200U CPU @2.5GHz* *8.00GB RAM*	1 u	800.00	48	8/48	133.33 €
					TOTAL	**133.33 €**

Table 2. Personal Computer Costs

Server CVBLab hardware

Nº	Hardware description	Quantity	Unit cost	Amort. period	Amort. ratio	Total
1	*Intel i7 @4.20GHz processor*	1 u	1200.00	48	8/48	200.00€
2	Graphics card NVIDIA Titan V	1 u	3300.00	48	8/48	550.00€
3	Disk SSD of 250 GB	1 u	77.00	48	8/48	12.83 €
					TOTAL	**762.83** €

Table 3. Server CVBLab Hardware Costs

Total Hardware Costs

Nº	Hardware description	Total
1	*Personal computer*	133.33€
2	Server CVBLab hardware	762.83€
	TOTAL	**891.16€**

Table 4. Total Hardware Costs

1.1.3 Software costs

In the following section the costs of the software used in this project will be detailed. The neural network algorithms were programmed with Python which is an open source programming language. However, as an Integrated Development Environment (IDE) for Python the professional version of PyCharm was used which has associated license costs. Furthermore, for external function design MATLAB was used which has also associated license costs. Finally, to draft this document the program Microsoft Office Word was used which has also license costs. These costs are described in Table 5.

Nº	Sotware description	Quantity	Unit cost	Amort. period	Amort. ratio	Total
1	*Microsoft Office prof. 2019*	1 u	579	12	8/12	386.00€
2	MATLAB R2019b	1 u	800	12	8/12	533.33€
3	Pycharm 2019.3.2	1 u	199	12	8/12	132.67€
					TOTAL	**1052.00€**

Table 5. Total Software Costs

1.2 Grand Total

Once the costs have been broken down into its different components of labour, hardware and software costs the Total Cost of the budget is computed in Table 6 by adding the costs previously described.

Nº	Description	Total
1	*Labour cost*	2688.00€
2	Hardware cost	891.16€
3	Software cost	1052.00€
	TOTAL	**4631.16€**

Table 6. Total Costs

The grand total of the development of this End of Degree Project is calculated by including the General Expenses (13% of the total cost) and the estimated profit (6% of the total cost). Finally, 21% of the total cost is included as value-added tax (VAT). The grand total is detailed in Table 7.

Nº	Description	Total
1	Total Cost	4631.16€
2	General Expenses (13%)	602.05€
3	Profit (6%)	277.87€
	SUBTOTAL	**5511.08€**
	VAT (21%)	1157.33€
	GRAND TOTAL	**6668.41€**

Table 7. Grand Total