DUBLIN CITY UNIVERSITY (DCU)
SCHOOL OF ELECTRONIC ENGINEERING

Generalisable Cardiac MRI Analysis with Deep Learning

Author:
Carles Garcia Cabrera, B.E., M.E.

Supervisor:
Dr. Kevin McGuinness

Co-supervisors:
Prof. Noel E. O'Connor
Dr. Kathleen M. Curran (UCD)

A Dissertation submitted in fulfilment of the requirements for the award of Doctor of Philosophy (Ph.D.)

August 2024

Declaration of Authorship

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, and that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Student number: 19214004

Signed:  *Carles Garcia Cabrera*

Date: Sunday 25th August, 2024

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Kevin McGuinness, whose guidance, support, and encouragement were invaluable throughout this research. His profound impact on my academic journey and his dedication to advancing knowledge will always be remembered. My heartfelt thanks also extend to my co-supervisors, Noel E. O'Connor and Kathleen M. Curran, for their invaluable contributions, insights, and unwavering support throughout this project. In difficult times, they have always provided support beyond any expectation. Their expertise and encouragement have been instrumental in shaping the outcome of this work. I am profoundly grateful for their mentorship and the opportunities they have provided.

The completion of this dissertation would not have been possible without the help of Angela Lally and Eric Arazo. Thank you, Angela, for being like a second mum throughout my PhD journey. Your unwavering support and guidance provided me with the strength and encouragement needed to navigate the challenges. I am deeply grateful for your constant presence and the comfort you offered during this time. Eric, your assistance with research-related problems and your mentorship have been invaluable. Your readiness to help and your insightful advice have greatly contributed to my progress as a researcher. I am truly appreciative of your support and collaboration over these years.

I also do not want to forget David Azcona, who was there during difficult times and served as a role model both as a researcher and as a person. His guidance and example have been deeply inspiring. Additionally, I am grateful to Enric Moreu, who provided the best onboarding experience when I arrived in Dublin. His support was instrumental in helping me acclimate to a new environment and get started on my research journey.

I would also like to extend my heartfelt thanks to Ferran Marques and Xavi Giró, who placed their trust in me early in my career. Their confidence and support have been pivotal to many of my achievements, and I owe much of my progress to their belief in my potential. Additionally, I am grateful to Pau Medrano, Alistair Young, and Oscar Camara for introducing me to the fascinating world of cardiac MRI. Their insights and guidance have been instrumental in shaping my understanding and enthusiasm for this field.

I am profoundly grateful to Ireland for hosting me throughout this journey and providing a welcoming environment for my research. My sincere thanks go to Science Foundation Ireland for sponsoring my research and supporting my academic endeavours. I also extend my appreciation to my research centre, ML-Labs, and to Dublin City University for their invaluable support and resources. The opportunities and facilities provided by these institutions have been crucial to the success of my work.

Lastly, I would like to thank my parents, Eva and Miquel, for their unwavering love and support throughout this journey. Additionally, I extend my heartfelt appreciation to the rest of my family for their support and encouragement, which has been a constant source of strength throughout my PhD journey. I am also deeply grateful to all my friends who cheered me on during this process, ranging from the "Quan quedems" to Gerard and Guillem, and the village of Sant Esteve. A special thanks goes to Celia, my girlfriend, who has been by my side through the ups and downs, providing constant encouragement and understanding.

Publications

- **Carles Garcia-Cabrera**, Kathleen M. Curran, Noel E. O'Connor, and Kevin McGuinness. 'Semi-supervised learning of cardiac MRI using image registration' [1].

  In *Irish Machine Vision and Image Processing Conference (IMVIP)*. September 2021.

- **Carles Garcia-Cabrera**, Eric Arazo, Kathleen M. Curran, Noel E. O'Connor, and Kevin McGuinness. 'Cardiac Segmentation using Transfer Learning under Respiratory Motion Artifacts' [2].

  In *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI). Statistical Atlases and Computational Models of the Heart (STACOM) proceedings. CMRxMotion Challenge*. September 2022.

- **Carles Garcia-Cabrera**, Enric Moreu, Eric Arazo, Kathleen M. Curran, Noel E. O'Connor, and Kevin McGuinness. 'Cardiac MRI Dataset Balancing using Atlas Deformations and Style Transfer'.

  In *Scientific Assembly and Annual Meeting of the Radiological Society of North America (RSNA). November 2023.*

- **Carles Garcia-Cabrera**, Kathleen M. Curran, Noel E. O'Connor, and Kevin McGuinness. 'Cardiac Magnetic Resonance Phase Detection Using Neural Networks' [3].

  In *Irish Conference on Artificial Intelligence and Cognitive Science (AICS)*. December 2023.

- **Carles Garcia-Cabrera**, Enric Moreu, Eric Arazo, Kathleen M. Curran, Noel E. O'Connor, and Kevin McGuinness. 'Synthetic Balancing of Cardiac MRI Datasets' [4].

  In *Medical Image Understanding and Analysis (MIUA)*. July 2024.

Other Publications:

- James Callanan, **Carles Garcia-Cabrera**, Niamh Belton, Gennady Roschchupkin, Kathleen M. Curran. 'Integrating feature attribution methods into the loss function of deep learning classifiers' [5].

  In *Irish Machine Vision and Image Processing Conference (IMVIP)*. September 2022.

- Julia Dietelmeier, **Carles Garcia-Cabrera**, Anam Hashmi, Kathleen M. Curran and Noel E. O'Connor. 'Cardiac MRI reconstruction from undersampled k-space using double-stream IFFT and a denoising GNA-UNET pipeline' [6].

  In *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI). Statistical Atlases and Computational Models of the Heart (STACOM) proceedings. CMRxRecon Challenge*. September 2023.

- Jun Lyu, Chen Qin, Shuo Wang, Carles Garcia-Cabrera and Others. 'The state-of-the-art in Cardiac MRI Reconstruction: Results of the CMRxRecon Challenge in MICCAI 2023' [7].

Contents

List of Figures

List of Tables

Abbreviations

ED	**E**nd **D**iastole
ES	**E**nd **S**istole
EF	**E**jection **F**raction
CHD	**C**ardiac **H**eart **D**isease
CVD	**C**ardio**V**ascular **D**isease
GLS	**G**lobal **L**ongitudinal **S**train
RV	**R**ight **V**entricle
LV	**L**eft **V**entricle
MYO	**M**yocardium
CMR	**C**ardiac **M**agnetic **R**esonance
MRI	**M**agnetic **R**esonance **I**maging
MR	**M**agnetic **R**esonance
GT	**G**round **T**ruth
ECG	**E**lectro**C**ardio**G**ram
LA	**L**ong **A**xis
SA	**S**hort **A**xis
aFD	**a**verage **F**rame **D**ifference
NN	**N**eural **N**etwork
DNN	**D**eep **N**eural **N**etwork
CNN	**C**onvolutional **N**eural **N**etwork
SGD	**S**tochastic **G**radient **D**escent
GPU	**G**raphics **P**rocessing **U**nit

Abstract

"Generalisable Cardiac MRI Analysis with Deep Learning"

by Carles Garcia Cabrera

Cardiac Magnetic Resonance Imaging (CMR) is a powerful diagnostic tool for assessing cardiac structure and function, providing detailed information crucial for clinical decision-making. The complexity and variability of cardiac images, however, poses challenges for accurate and efficient analysis. This PhD research aims to develop a novel deep learning framework for generalisable CMR analysis, addressing the limitations of existing methods and enhancing the clinical utility of cardiac imaging. The novel deep learning methods developed in this thesis leverage the capacity of neural networks to automatically learn hierarchical representations from raw image data. The framework is designed to advance state-of-the-art in its ability to generalise across diverse patient populations, imaging protocols, and scanner types, ensuring robust performance in real-world clinical settings. To achieve this, a large and diverse dataset of cardiac MRI scans was curated from open-source data banks, incorporating variations in anatomy, pathology, and acquisition parameters.

Areas of investigation included semi-supervised learning techniques, pre-training and transfer learning, transformers and sequential networks, architectural refinements, and development of novel data augmentation strategies to mitigate respiratory artifacts and advance model generilisability. The key innovations include: (1) the development of novel synthetic label propagation techniques for precise time frame detection within the cardiac cycle to optimise the model's performance on limited annotated data; (2) achieving well-balanced data sets with the use of synthetic labels derived from image registration of intermediate time frames, thereby fortifying the model's adaptability to analyse scans from unseen vendors during training; (3) leveraging Pre-trained Models and MRI Specific Augmentations to mitigate respiratory artifacts and (4) synthetic balancing of CMR data sets using style transfer through deformations applied to an atlas.

Notably, the proposed framework enhances predictions for previously unseen pathologies, underscoring the positive impact of these techniques across diverse settings. Collectively, these advancements represent a significant step forward in the evolution of foundational models in the realm of cardiac MR analysis.

Chapter 1

Introduction

Cardiac Magnetic Resonance Image (CMRI) segmentation is an important step during quantitative analysis for the diagnosis of cardiovascular diseases. During the process, the image is partitioned into meaningful regions. The labelling of this information helps clinicians calculate and understand important features, such as the ejection fraction, which are then used to determine whether subjects have a particular pathology or if surgery is necessary [8]. The ejection fraction (EF) is a measure of the percentage of blood ejected from the heart's left ventricle during each contraction, providing insight into the heart's pumping efficiency. Other functional analyses, such as strain, flow, or torsion, also require or benefit from the segmentation being precise.

The aim of this thesis is to explore the utilisation of neural networks for automating the delineation of meaningful regions for the calculation of the ejection fraction. The thesis focuses on the study of alternatives to address an important challenge in current cardiac MRI analysis tools based on neural networks: the need for balanced datasets that represent the entirety of the population.

In the following chapters, methods are proposed and their impact analysed in machine learning tasks that directly relate to the calculation of the ejection fraction. In particular, most of the work tackles specific improvements in the segmentation of the ventricular cavities of the heart. Additionally, the correct

labelling of the key time frames (end-diastolic and end-systolic) is explored, given its importance and impact in understanding the ejection fraction.

This chapter presents an overview and provides the motivation behind the research outlined in this thesis, as discussed in Section 1.1. It defines the research objectives in Section 1.2. Additionally, Section 1.3 offers a concise overview of the main contributions. Lastly, in Section 1.4, the chapter outlines the structure of the document.

1.1 Motivation

The application of deep learning (DL), being rooted in data-driven methods, has raised significant concerns regarding data privacy. With the advent of regulations such as the General Data Protection Regulation (GDPR), the protection of users' privacy has gained enormous importance, and organisations have been compelled to acknowledge the ownership and security of data. From a technical perspective, the challenge of developing methods for storing, querying, and processing data in a manner that mitigates privacy concerns while building deep learning models has become increasingly difficult. The construction of privacy-preserving algorithms necessitates the integration of cryptography with deep learning, and the combination of techniques from a diverse range of fields including data analysis, distributed computing, federated learning, and differential privacy. The objective is to create models that are not only secure and efficient but also possess strong generalisability [9, 10, 11]. However, privacy-preserving approaches inherently make obtaining the necessary data for model training more difficult, adding an additional layer of complexity to the development of these techniques. While this field is still in a very early stage, [12] provided a summary of best practices in regard to privacy and security in machine learning approaches.

Given the limited data available for under-represented subjects that have been scanned with less common scanners or that have pathologies that are more scarce, clinical interest has arisen around adopting automated methods to help clinicians diagnose faster and more accurately the wide range of data variety. Research in this aspect will include the quality of all diagnostics, with more emphasis on tackling the biases that already exist in the performance of DL-based CMRI segmentation models due to the imbalance of the training set [13].

Since data is scarce and with limited annotation, approaches are needed that can benefit from the available data. Furthermore, the data can be noisy or of poor quality. In some cases, the presence of artifacts can make standard algorithms under-perform or even not work. Some subjects might suffer from other underlying conditions that will decrease the quality of the scans (e.g. not being able to hold their breath or suffering from arrhythmia [14]; see Figure 1.1). An important area of study is to make algorithms resilient to such artifacts.

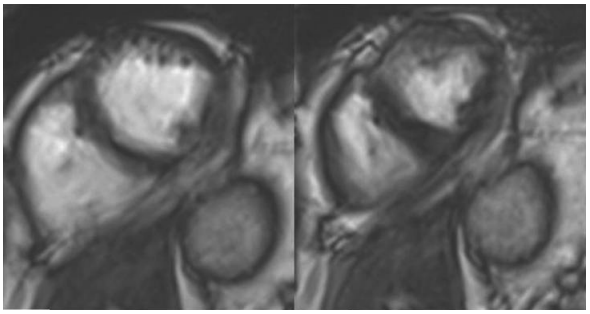


FIGURE 1.1: Cardiac MRI images with motion artifacts due to irregular intervals owing to arrhythmias. Images from [14].

Datasets in cardiovascular imaging are usually imbalanced, with uneven presence of all known diseases and characteristics, genders, ages and ethnicities. This imbalance directly affects the quality of the models developed to analyse the data. While obtaining scans is costly and time consuming, the generation of synthetic data is easier and cheaper, and can benefit from the fields of heart modeling and style transferring or image generation. Focusing on the generation of synthetic medical scans could play an important role in ensuring balanced datasets, while overcoming many challenges this field faces in the acquisition of data.

Convolutional networks are the most common type of neural networks for the task of segmentation, but their adoption in the clinical world is slower than in others and should be carefully analysed due to their very important role in the clinical workflow. Therefore, all algorithms need to be validated, by establishing clear metrics that the automated methods must satisfy in order to make it to clinical products [15]. Further efforts on interpretability would greatly benefit existing research (although this is outside the scope of this thesis).

Foundation models, also known as pre-trained models, are purposefully designed to be adaptable to various downstream cognitive tasks through techniques like fine-tuning [16]. This adaptability is achieved by initially pre-training the models on extensive and diverse datasets, enabling them to capture broad knowledge. Subsequently, these foundation models are fine-tuned for specific tasks using smaller, task-specific datasets. The core concept is to leverage the general understanding gained from pre-training on large datasets and then tailor the model's capabilities to excel in more specialised tasks.

1.1.1 Generalisable Cardiac MRI analysis

Solutions to increase the performance of cardiac MRI analysis tools on data from under-represented groups of patients aim at designing robust algorithms to train models under scarcity of annotated data where: data come from different scan vendors, data contains different pathologies, no time frames are labelled, or data contains artifacts. Solutions designed in the research reported in this thesis for the first scenario, are discussed in Chapter 3 in the context of extending labeling to all time frames. Work addressing the second scenario is presented in Chapter 4, in which the role of architectures and post-processing is studied. An approach to address the third scenario is discussed in Chapter 5, where model fine-tuning and data augmentation were explored. The work addressing the fourth scenario is presented in Chapter 6, in which regression and sequential neural network modules are studied.

A prominent amount of literature focuses on benefiting from the available data to reduce the number of cases where available neural network tools fail. Despite the significant amount of work done in that direction, the path for elaborating a foundation model for cardiac MRI is unclear due to a lack of available shareable data. Overcoming the necessity of having this data is explored in Chapter 7.

1.2 Research objectives

The field of cardiac MRI analysis is actively exploring the challenges discussed previously in this chapter, i.e. increasing the quality of the analysis with the same amount of data. However, the models used still rely on labelled varied data, and the best way to minimise it remains under discussion. Unlike the work explored in this thesis, existing methods ignore additional uses of the data present in the dataset and/or do not leverage work from related fields, such as heart modelling or image registration, to extend the generalisability of the models. Similarly, the literature that aims to minimise the dependency on varied labelled data rarely considers synthetically extending the available labelled data. The following objectives have guided and motivated the research performed and reported in this thesis:

Objective 1: To develop more generalisable models that can be used in diverse settings.

- **Objective 1.1: Enhance model robustness and generalisability to the variations presented by various imaging equipment sources.**

  Amongst the most prominent sources of poor performances in CMRI models is the lack of generalisation capacity to segment scans from vendors that are unseen or under represented to the neural network during training [17]. Prior to this research, this issue was tackled by applying intensity based data augmentations [18, 19, 20]. However, there is room for further improvement in performance. For this challenge, the objective is

to synthetically extend the labels by applying image registration between time-frames to propagate the labels across them, extending the labelled data for training.

- **Objective 1.2: To develop a model capable of effectively handling and distinguishing among various cardiomyopathies, ensuring robust performance across different pathological conditions.**

  The lack of performance on subjects with underlying pathologies is another concerning aspect of CMRI [21]. The range of diseases present in the datasets range from morphological to functional, adding complexity as models aim to process more within a single neural network. Additionally, multi-view is required to analyse some of the pathologies (e.g. tetralogy of the fallot). Within this challenging scenario, the objective is to study the performance of multiple architectures to identify the best performing one for each pathology, vendor, and view. Additionally, developing post-processing techniques to benefit from having two views from the same tissue could enhance the performance in some scenarios.

Objective 2: To automatically detect end-diastolic and end-systolic time frames for more accurate analysis and better reproducibility on benchmark datasets.

The end-diastolic and end-systolic time frames in cardiac imaging represent, respectively, the moments of maximum blood volume in the heart's ventricles at rest and the points of minimum blood volume during contraction. While segmentation performance has direct translation on better and more reliable ejection fraction calculation, the correct identification of the key time frames can reduce the accuracy of the calculation by up to 10% with just 2-3 frames of difference [22]. This is important because ECG signals do not always accompany CMRI scans. In the past, progress on solutions has remained slow and without any possibility for comparison, since proposed methods were developed with

private data [23, 24]. The objective in this setting is to provide a reliable, fast, and easy to compare approach on benchmark datasets.

Objective 3: To develop a method to mitigate against respiratory motion artifacts in cardiac MRI data.

During scanning time, patients are required to hold their breath to prevent the images from appearing blurred [25]. That is not always possible, especially for elderly people or individuals with underlying respiratory conditions. Even in normal scenarios, in case the patient was to breathe, the scan might have to be repeated, with attendant time and resources implication. Therefore, efforts to train neural networks that can overcome the extra difficulty of analysing the data can contribute to hasten, cheapen, and broaden the benefits of machine learning on CMRI.

Objective 4: To implement synthetic data balancing techniques to address class imbalance in datasets, ensuring equitable representation of different classes for more robust model training and improved performance across diverse scenarios.

Initial approaches to generate synthetic cardiac MRI images proposed methods based on GANs [26], VAEs [27], etc. However, these approaches do not consider the utilisation of currently available representation of patients [28, 29, 30], such as atlases, as a starting point for the generation of new data. A cardiac atlas is a detailed visual representation or map of the typical anatomy and structures of the heart, serving as a reference for medical professionals in the analysis of cardiac images. In fact, successfully using a source of data like a cardiac atlas, could contribute to elaborating a foundational model which could be re-trained on demand, ensuring the model meets the needs of the data available [31]. As a proof of concept, the aim is to research the synthetic extension of datasets constituted solely of healthy patients to train networks that achieve equal performance on diseased subjects.

1.3 Contributions

This section presents a description of the main contributions of this thesis as a result of the work described in subsequent chapters.

First contribution:

To enhance the applicability of the developed models across diverse settings, the research focuses on the creation of more generalisable models. By leveraging comprehensive pre-training on diverse datasets, a robust foundation is established that captures broad knowledge. This approach facilitates the development of models that can be fine-tuned for specific tasks using smaller, task-specific datasets. The contribution of this work is the resultant generalisability, allowing the models to demonstrate efficacy and relevance in a wide array of contexts and applications.

Second contribution:

To prevent a decrease in performance in the case of incorrect frame labelling, the research includes the creation of models that can detect the key-frames in parallel to the segmentation task. The result of successfully incorporating these two tasks together in the clinical setting ensures the correct calculation of the key metrics. The contribution of this work in this matter is to present accurate and fast models, on public datasets for reproducibility.

Third contribution:

To mitigate the impact of respiratory motion artifacts on the efficiency of diagnostics, the research also focuses on the development of a method for crafting models resilient to the former. The outcome of fine tuning pre-trained models in combination with augmentations that resemble the effect of the artifacts results in an increased performance. The contribution in this case is the development of a novel deep learning model that enhances the accuracy and reliability of cardiac image analyses in the presence of motion-induced variations.

Fourth contribution: The key idea behind foundation models is to leverage pre-training on large and diverse datasets, capturing broad knowledge, and then fine-tune them for specific tasks with smaller, task-specific datasets. Designing methods to generate synthetic shareable data hugely contributes to the establishment of a foundational model. The contribution of this work lies in successfully developing a proof of concept pipeline that can generate on-demand data by applying atlas deformation in accordance to known morphological features, and obtaining realistically looking MRI images from those deformations throw deep learning.

1.4 Thesis structure

The remainder of the thesis is structured as follows:

In Chapter 2, the technical background of the thesis is delved into and the related work is examined. This chapter elucidates the fundamental principles underlying cardiac MR analysis with deep learning, encompassing a comprehensive review of diverse datasets, the experimental setup, and the notation employed throughout the subsequent chapters of the thesis.

In Chapter 3, the application of image registration to synthetically label unannotated time frames is investigated, effectively enhancing performance across subjects from previously unseen vendor scans, by employing a semi-supervised approach..

In Chapter 4, a comprehensive examination is conducted on various architectures along with a post-processing technique that utilises pixel mapping to optimise performance across patients with distinct pathologies. This comparative analysis offers a detailed exploration of these components, providing valuable insights for the training of neural networks specifically tailored for cardiac MRI segmentation.

In Chapter 5, time frame detection is tackled using state-of-the-art neural network modules.. The chapter presents results obtained from public datasets and showcases a substantial acceleration in previously reported inference times.

In Chapter 6, a transfer learning method designed for respiratory motion correction is introduced to alleviate the impact of respiratory artifacts. Additionally, the chapter explores specific data augmentation techniques that emulate these artifacts for improved robustness.

In Chapter 7, the synthetic balancing of cardiac MR datasets is thoroughly investigated, with a particular focus on two pathologies featuring notable morphological implications. The methodology employs an atlas and style transfer in a proof-of-concept experiment to evaluate the feasibility of this approach and its potential implications in the development of a foundational model.

Lastly, Chapter 8 provides an overview of the research presented in this thesis. The research objectives are discussed, the contributions made are highlighted, and the limitations of the work are outlined. Additionally, suggestions for future exploration in the domain of cardiac MR analysis with deep learning are offered.

Chapter 2

Background

In this chapter, an initiation to and exposition of the fundamental elements integral to the research is provided. Commencing with an introduction to cardiac anatomy and the corresponding principles related to the cardiac cycle, the narrative underscores the importance of appreciating the foundational physiological aspects required for the generation of efficient predictive models. Following this, the discourse transitions into an in-depth explanation of the procedures utilised for quantifying cardiac activity and examining anatomical detail. This leads to the presentation of foundational principles central to this research strategy, with deep learning serving as a significant focus. This chapter concludes with a literature review of the topic and an overview of the currently available public datasets that are extensively utilised in the forthcoming chapters.

2.1 Cardiovascular System Anatomy

This research focuses on two of the most clinically relevant chambers of the heart (see Figures 2.1 and 2.2, short-axis, and long-axis slices, respectively), namely the right and left ventricles. It is important to note that the reader would see the ventricles as if they were facing the patient (the famous "left-right swap" rule in radiology).

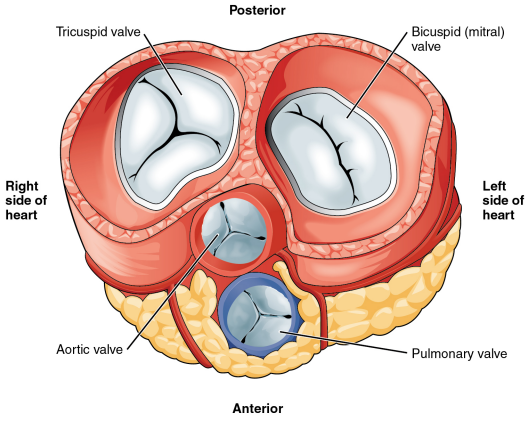


FIGURE 2.1: Cardiac short-axis cut displaying the tricuspid and mitral valves. In the images resulting from an MRI, the left part of the heart is observed on the right, and the right part is observed on the left. Image from [32] CC by 3.0 licensed imagery.

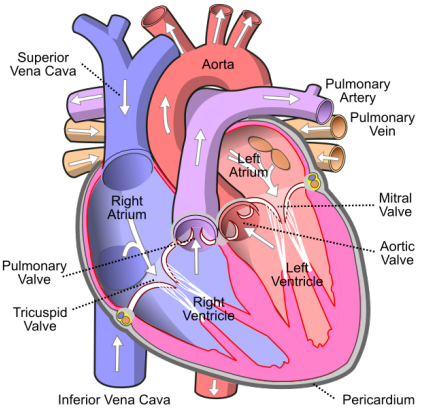


FIGURE 2.2: Cardiac long-axis cut displaying four of the most important parts of the heart: the left ventricle, the right ventricle, the right atrium and the left atrium. Image from [32] CC by 3.0 licensed imagery.

Between the epicardium (the outer layer of the heart wall) and the endocardium (the inner layer of the heart wall), there is a thick middle layer called the myocardium, and blood is supplied via coronary circulation. Myocardial cells are called cardiomyocytes (heart muscle cells) and are joined by intercalated discs. The matrix encasing them with collagen fibres and other substances is called the extracellular matrix.

The myocardium is an involuntary signal-paced striated muscle that constitutes the main tissue of the heart's walls. The myocardium is an important focus of the work conducted in this research, as it represents the boundary of the left ventricle chamber. The examination of myocardial tissue can aid in the diagnosis of cardiac disorders.

2.1.1 The Cardiac Cycle

Events occurring within a single heartbeat can be divided into four phases: isovolumetric contraction, ventricular ejection, isovolumetric relaxation, and ventricular filling [33]. These events (contraction and relaxation) can be electrical and mechanical and happens for both left and right ventricles at different timings. The atria and ventricles alternatively perform the contractions.

The first part of the cycle is the isovolumetric contraction, which causes ventricle pressures to rise above atria pressure. This contraction causes the mitral and tricuspid valves to close but forces the aortic and pulmonary valves to open. This blood ejection is called ventricular ejection. The systole term refers to those two phases of the heart cycle.

The blood ejection decreases the ventricular pressure, which falls below the atrial pressure. Then, the aortic and pulmonary valves close and the atrias start refilling. This phase is referred to as the isovolumetric relaxation, and it is followed by the ventricular filling when the mitral and tricuspid valves open. These two phases belong to the diastole. When those phases end, the cardiac cycle starts again. See Figure 2.3.

In addition to those concepts, other important hemodynamic parameters are:

- **Ventricular end-diastolic volume (EDV):** volume of blood inside the ventricle prior to contraction (end of diastolic phase).
- **Ventricular end-systolic volume (ESV):** volume of blood inside the ventricle prior to ventricle filling (end of diastolic phase).
- **Ejection fraction (EF):** measures the blood pumping quality of the ventricle for a cardiac cycle. It is calculated as the percentage of blood ejected from the ventricle.

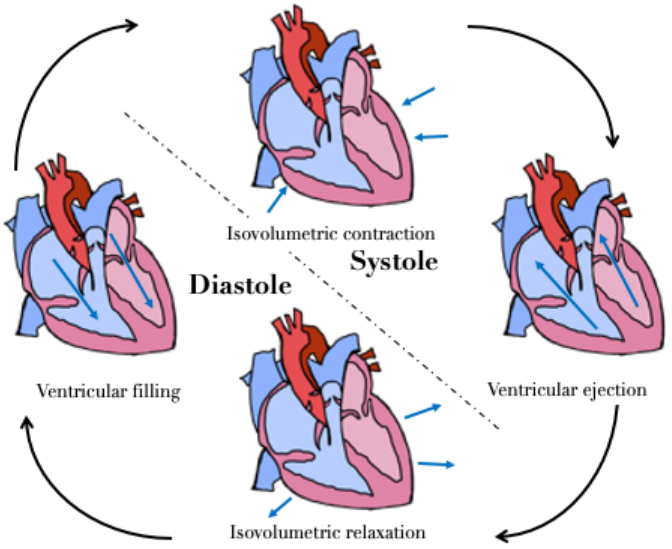


FIGURE 2.3: The cardiac cycle displaying the four phases and their belonging to diastole or systole respectively.

2.1.1.1 Electrocardiography

Electrocardiography (ECG) is a non-invasive diagnostic procedure employed to record the heart's electrical activity over time. This technique utilises electrodes (usually between 12 and 15 [34]) placed strategically on a patient's skin to detect and measure the electrical signals generated by the heart's specialised cells during each cardiac cycle. The resulting data is represented as a graphical waveform, which provides invaluable insights into the heart's overall function, rhythm, and conduction system.

The electrocardiogram offers critical information for the evaluation of various cardiac conditions, including arrhythmias, ischemic heart disease, and structural abnormalities. By analysing the amplitude, duration, and morphology of the ECG waveforms, medical professionals can identify patterns indicative of specific pathologies or dysfunctions. Furthermore, electrocardiography serves as a fundamental tool in monitoring the efficacy of pharmacological treatments and interventions targeted at alleviating cardiac issues, and also for being able to allocate imaging data in the correct frame [35].

2.2 Cardiomyopathy

In this section, various cardiac pathologies identifiable through imaging and diagnostic procedures will be explored. These conditions, which encompass congenital abnormalities present at birth and diseases that develop over time, are characterised by their impact on the structure and function of the heart [36]. Understanding the characteristics and implications of these pathologies is important for accurate diagnosis and effective treatment.

2.2.1 Dilated Left Ventricle (LV)

Left Ventricular Hypertrophy is the enlargement and thickening (hypertrophy) of the walls of the left ventricle of the heart. The thickened heart wall loses elasticity, leading to increased pressure needed to fill the heart's pumping chamber and send blood to the rest of the body. Eventually, the heart may not pump with sufficient force.

2.2.2 Hypertrophic Cardiomyopathy (HCM)

Hypertrophic Cardiomyopathy is a disease in which the heart muscle becomes abnormally thick (hypertrophied). The thickened heart muscle can make it more difficult for the heart to pump blood.

2.2.3 Congenital Arrhythmogenesis (ARR)

Congenital Arrhythmogenesis is an irregular heartbeat present at birth. Heart rhythm problems (arrhythmias) occur when the electrical signals that coordinate the heart's beats do not work properly. This failure in signalling causes the heart to beat too fast (tachycardia), too slow (bradycardia), or irregularly.

2.2.4 Tetralogy of Fallot (FALL)

Tetralogy of Fallot is a rare condition caused by a combination of four heart defects present at birth (congenital). These defects, which affect the structure of the heart, cause oxygen-poor blood to flow out of the heart and to the rest of the body.

2.2.5 Interatrial Communication (CIA)

Interatrial Communication is a congenital cardiac malformation characterised by a communication between the atrial chambers of the heart. The hole increases the amount of blood that flows through the lungs.

2.2.6 Dilated Right Ventricle (DRV)

Dilated Right Ventricle occurs when the muscle on the right side of the heart thickens and enlarges, similar to what happens in the left ventricle with dilated left ventricle pathology.

2.2.7 Tricuspidal Regurgitation (TRI)

Tricuspid Regurgitation is a type of heart valve disease in which the valve between the two right heart chambers (right ventricle and right atrium) does not close properly. As a result, blood leaks backwards into the upper right chamber (right atrium).

2.3 Magnetic Resonance Imaging (MRI)

Magnetic Resonance Imaging (MRI) is a non-invasive diagnostic modality that utilises the principles of nuclear magnetic resonance (NMR) to generate detailed images of anatomical structures and physiological processes in the human body.

2.3.1 Basic principles of MRI

This section aims to introduce the fundamental principles underlying MRI, including the concepts of magnetic fields, proton alignment, radiofrequency (RF) pulses, Larmor frequency, echo time (TE), and signal detection.

2.3.1.1 Magnetic field and proton alignment

The primary component of an MRI system is a powerful magnet, typically a superconducting magnet, which generates a strong and uniform static magnetic field (B_0) within the imaging volume. The human body is primarily composed of water molecules, each containing two hydrogen atoms (^1H), with their nuclei comprised of protons. When placed within the static magnetic field, these protons align either parallel (low-energy state) or anti-parallel (high-energy state) to the direction of the magnetic field. The summation of all the magnetic moments of the protons is named net magnetisation (M_z), and it is aligned to the direction of the main magnetic field.

The MRI system transmits RF pulses to the patient's body to excite the protons, causing them to flip from their low-energy state to their high-energy state. The frequency of these RF pulses is specifically tuned to match the resonance frequency of the protons, which is known as the Larmor frequency. According to the Larmor equation ($\omega_0 = \gamma B_0$, where ω is the Larmor frequency and γ is the gyromagnetic ratio), the Larmor frequency is proportional to the strength of the static magnetic field. The angle by which the protons are flipped is referred

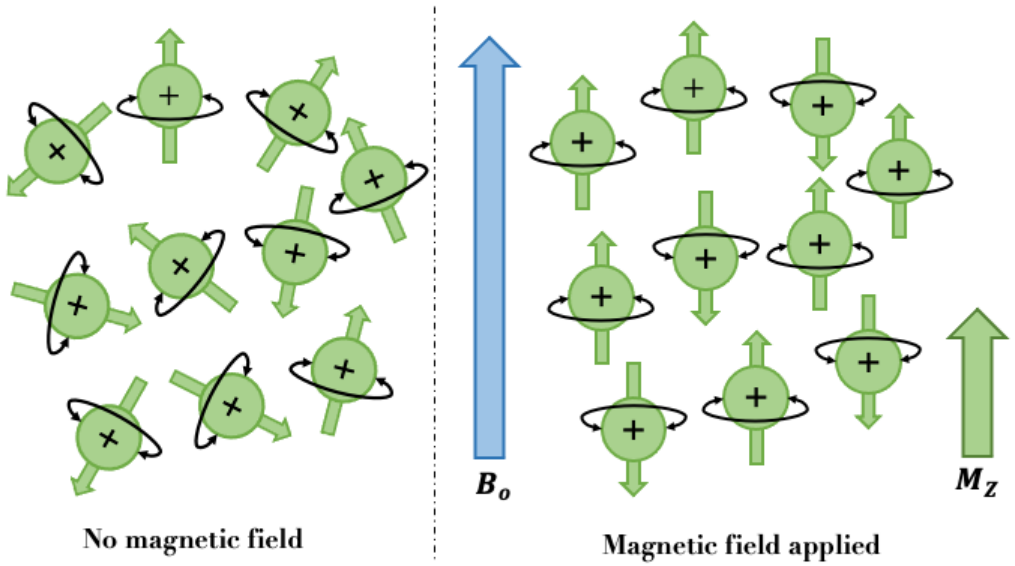


FIGURE 2.4: In the left, the protons with no external magnetic field. In the right, an external magnetic field (B_0) is applied, causing a net magnetisation (M_z) in the direction of B_0.

to as the flip angle and is determined by the duration and amplitude of the RF pulse. For ^1H, γ is equal to 42.58MHz/T. Figure 2.4 depicts the two stages, one before the application of the magnetic field and one after.

2.3.1.2 Relaxation times, echo time and signal detection

Upon cessation of the RF pulse, the protons begin to relax back to their equilibrium state, releasing energy as a weak NMR signal. Two fundamental relaxation processes occur: longitudinal relaxation (T1) and transverse relaxation (T2). Longitudinal relaxation (T1) refers to the time taken for the protons to realign with the static magnetic field, while transverse relaxation (T2) describes the time taken for the protons to lose phase coherence in the transverse plane. These relaxation times are tissue-specific and contribute to the contrast observed in the MRI images.

Echo time (TE) is a crucial parameter in MRI, representing the time interval between the application of the RF pulse and the acquisition of the NMR signal.

TE significantly influences image contrast, determining the extent of T2 relaxation occurring before signal detection. Shorter TEs emphasise T1 contrast, while longer TEs enhance T2 contrast.

To generate the final MRI image, a receiver coil detects the emitted NMR signal, which is subsequently processed through a series of mathematical transformations, such as the *Fourier Transform*. The spatial encoding of the MRI signal is achieved by applying a set of gradient magnetic fields that vary linearly across the imaging volume, enabling the localisation of the signal in three-dimensional space. The detected signal contains both magnitude and phase information.

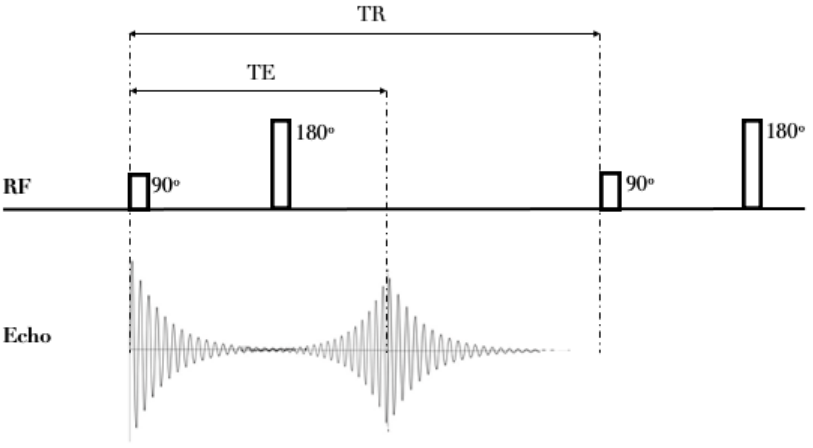


FIGURE 2.5: Basic representation of a spin-echo pulse. When an RF pulse is applied, the magnetisation is flipped into the transverse plane. Subsequently, a second RF pulse is used to refocus the spins and create an echo at the designated echo time (TE). The repetition time (TR) is the space between pulse sequences.

Numerous iterations of the pulse sequence are required to obtain adequate information for the generation of an image. The temporal interval between successive pulse sequences is denoted as the repetition time (TR). Varying TR and TE parameters can yield distinct weightings for T1 and T2 relaxation times, consequently producing divergent tissue contrasts in the resulting MRI images [37]. In Figure 2.5, a basic spin-echo pulse is represented.

2.3.2 K-space

This section presents a brief overview of MRI image generation, paying special attention to parameters related to resolution, signal-to-Noise Ratio (SNR), and field of view (FOV).

The images are obtained by applying the *Fourier Transform* to the k-space. The k-space is where the samples obtained while scanning are stored, and it has its own spatial resolution, as seen in Figure 2.6, and following: $\Delta k_x = \frac{1}{\text{FOV}_x}$ (Frequency Resolution), and $\Delta_x = \frac{1}{2K_{x,\max}}$ (Spatial Resolution). The FOV captures the dimension of the anatomic region included in the scan.

In k-space, each point is characterised by a complex number. Consequently, a data representation of k-space facilitates the visualisation of diverse representations, including real and imaginary components of complex signals or the magnitude and phase signals. It is crucial to acknowledge that Fourier transforms invariably result in complex numbers. Analogous to k-space, the transformed image can also be represented in a Cartesian coordinate system (real and imaginary components) or a polar coordinate system (magnitude and phase). Generally, only the magnitude image is retained, while the phase image is discarded.

The scanning time has a strong impact on the quality of the resulting image, as can be seen in the Scan Time and SNR Tradeoff formulas:

$$\text{Scan Time} = N_y \times \text{TR} \times \text{NEX} \qquad \text{(Scan Time)}$$

$$\text{SNR} = \frac{B_0 \times \sqrt{\text{NEX}} \times \text{Slice Thickness} \times \text{FOV}^2}{\sqrt{N_x} \times \sqrt{N_y} \times \sqrt{\text{Receiver Bandwidth}}} \qquad \text{(SNR Tradeoff)}$$

where NEX is the number of signal averages, B_0 is a constant when working with MR and TR is the repetition time used to obtain the samples. N_x and N_y refer to the number of samples in the image space in their respective dimensions. Therefore, to reduce the scanning time, images are sometimes up-sampled.

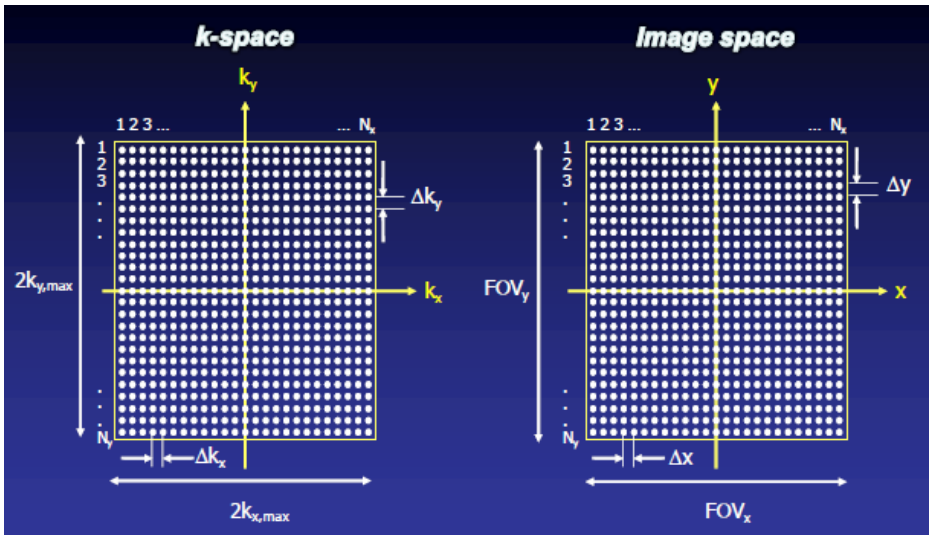


FIGURE 2.6: Association between Image Space and k-space. K-space axes correspond to frequency resolution, and Image space axes correspond to pixels. Image from [38].

2.3.3 MRI sequences

MRI scans contain high-quality information from the heart structures and functions [39]. In general, the acquisition is performed by selecting 2D slices with a given slice thickness. The assessment of the heart is usually performed using short- and long-axis slices, detailing the different cavities or chambers.

The acquired slices contain voxels (volumetric pixels), and when a single voxel contains different tissue it is averaged. This partial volume effect results in the blurring of boundaries.

2.3.3.1 Cine

Cardiac Cine MRI studies involve image acquisition, during which data from multiple heartbeats is typically required to populate the k-space matrix for each frame. Cine studies are generally conducted by repeatedly imaging the heart at a single slice location throughout the cardiac cycle, with 10 to 30 cardiac phases commonly sampled. To avoid confusion with other usages of the term "phase" in MRI, these individual cardiac images are frequently referred to as "frames."

The fundamental pulse sequence in cine studies is a bright-blood technique, typically employing a balanced steady-state free precession (SSFP) gradient echo method [40]. This method generates a high intravascular signal relative to other tissues, attributable to the intrinsically high T2/T1 ratio of blood. The structure of this sequence enables the utilisation of remarkably short TR and TE values, thereby facilitating the acquisition of multiple lines of k-space (echoes) during a single heartbeat.

A comprehensive evaluation of the entire heart necessitates obtaining separate cine image sets at various locations. For instance, a standard cine study might encompass three sets of short-axis images (base, mid-ventricle, and apex), three sets of long-axis images (2-, 3-, and 4-chamber), in addition to valvular and outflow tract images as required.

The short-axis view, often referred to as the transverse view, is oriented perpendicularly to the long-axis of the heart. This perspective is parallel to the atrioventricular (mitral and tricuspid) valves and allows for an effective evaluation of the myocardial walls and ventricular function. The view essentially "slices" the heart from base to apex, providing transverse cross-sectional images. An example of a short-axis slice is depicted in Figure 2.7, in particular the selected slice is a mid-ventricle slice.

In contrast, the long-axis view, also known as the longitudinal view, is aligned with the long-axis of the heart, which runs from the apex of the left ventricle to the center of the atrioventricular plane. This view offers essential insights into the heart valves, the chambers, and the overall structure and function of the left ventricle. An example of a long-axis view is displayed in Figure 2.8. Additionally, Figure 2.9 displays the position of both the short and long-axis views in the corresponding anatomical positions. Lastly, Figure 2.10 presents the superposition of the long-axis and short-axis images in their corresponding positions.

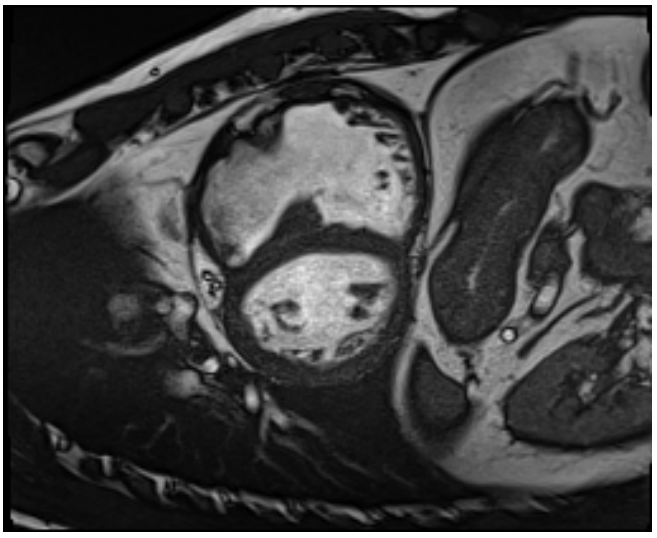


FIGURE 2.7: Example of a mid-ventricle short-axis view from a CMRI scan. Image from M&Ms2 [21].

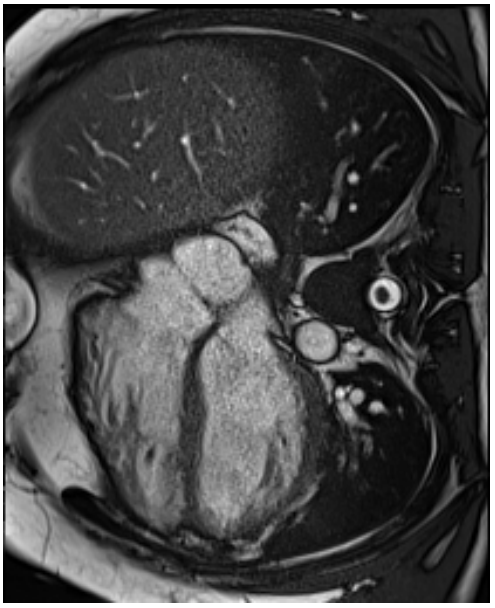


FIGURE 2.8: Example of a long-axis view from a CMRI scan. Image from M&Ms2 [21].

To maintain SSFP conditions for bright-blood imaging, radiofrequency (RF) pulses must operate continuously. Retrospective ECG-gating is performed to ensure that imaging data can be assigned to the appropriate phase of the cardiac cycle [35].

Cine sequences employ extremely short TEs (1-2 ms), allowing for the acquisition of multiple lines of k-space within each frame during a single heartbeat

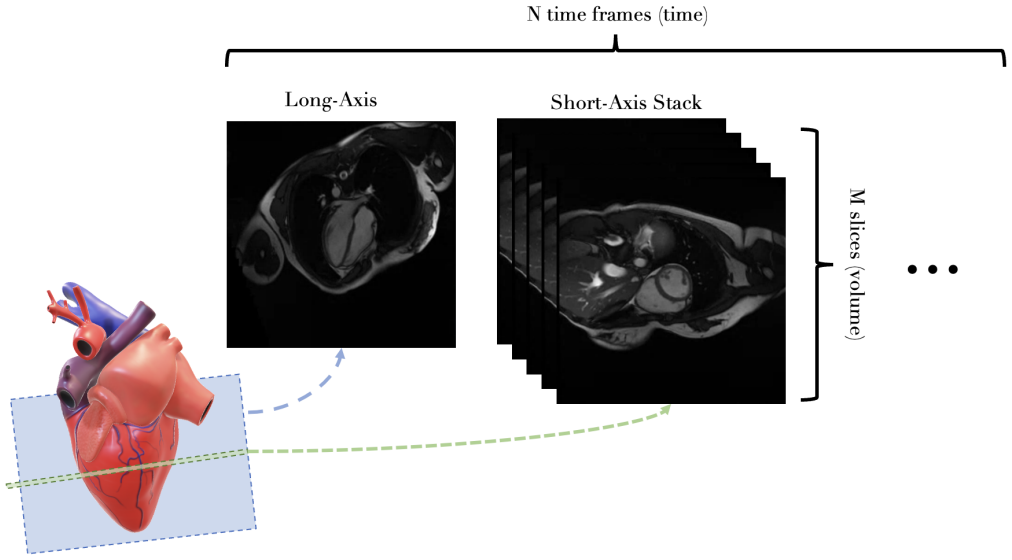


FIGURE 2.9: Example of the anatomical positions of the short-axis and long-axis views. Representative M values range between 10 and 15 slices, and N values range between 10 and 30 time frames. Scans from M&Ms2 [21].

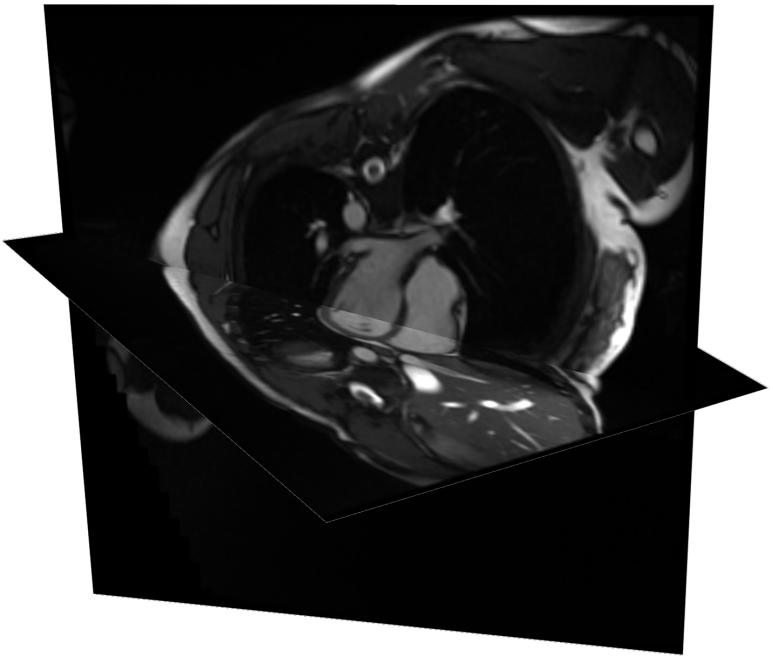


FIGURE 2.10: Long-axis and short-axis images superimposed. Scans from M&Ms2 [21].

(R-R interval). This value is an operator-selectable parameter known as views per frame (vps) or lines per segment (lps).

Given that only a limited number of views per frame are feasible, data collection must occur over multiple cardiac cycles. Nonetheless, a complete set of cine images for a given slice location can typically be obtained within 5-10 seconds, well within the single breath-hold capabilities of most patients. The breath-hold prevent any heart movement apart from the heartbeat. Not doing this hold will generate motion artifacts that disrupt the image.

2.3.3.2 Artifacts

There are three types of artifacts that can arise in MR: (1) tissue-related, (2) motion-related, and (3) technique-related; these artifacts reduce the quality of the examination since they are features that do not appear in the original imaged body. They should be considered, especially when pre-processing or adding data augmentation to the pipeline, as they can greatly improve or decrease performance [41].

2.4 Deep learning

Deep learning, a subfield of artificial intelligence and machine learning, has garnered significant attention in recent years due to its remarkable success in a wide range of tasks, including image classification, natural language processing, speech recognition, and semantic segmentation.

2.4.1 Deep learning fundamentals

The core idea of deep learning is to develop algorithms capable of learning abstract, hierarchical representations of data by training computational models known as deep neural networks (DNNs). These models are inspired by the structure and function of biological neural networks, but their inner mechanisms are different.

An artificial neuron, or perceptron, is the fundamental processing unit in a neural network. It receives input from other neurons or the input data, applies a linear transformation followed by a non-linear activation function, and produces an output value. The weights in the linear transformation are learnt during the training process by minimising a predefined loss function.

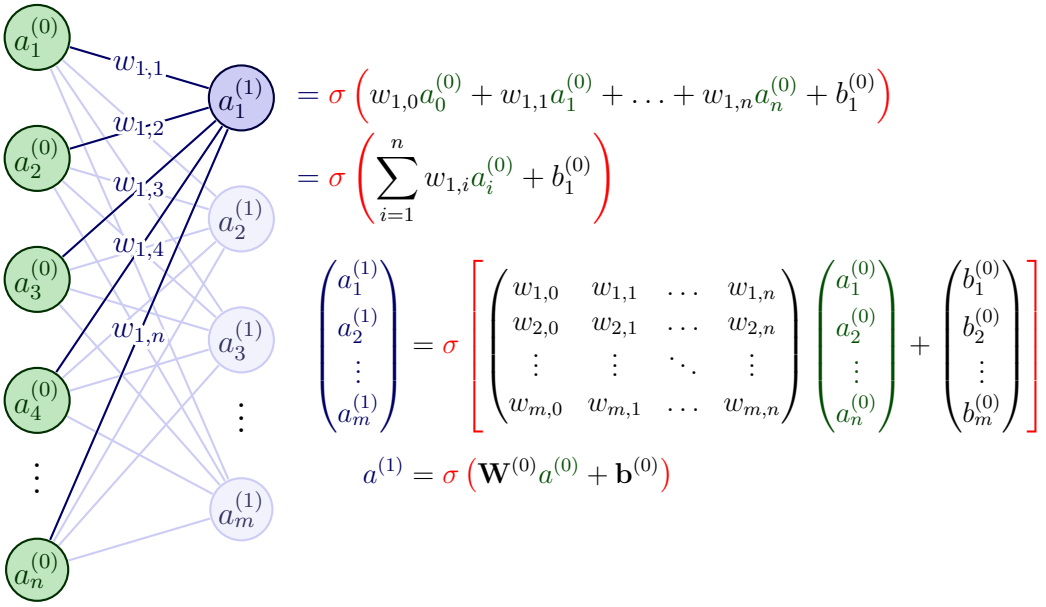


FIGURE 2.11: Multiple perceptrons in the context of a dense connection between two neural network layers. Green neurons represent the input neurons, and blue neurons represent neurons in the following layer. On the right side of the figure, there are the mathematical expressions of a single neuron and the entire layer calculations, where w are the weights, a are the input values, and b are the bias values. Finally, in red, σ is the activation function.

2.4.1.1 Multi-layer perceptron

The multi-layer perceptron (MLP) is an extension of the perceptron concept, consisting of multiple layers of interconnected artificial neurons. An MLP typically includes an input layer, one or more hidden layers, and an output layer. The input layer receives the data features, while the output layer generates the final predictions. The hidden layers in between are responsible for learning complex, non-linear representations of the input data. By stacking multiple layers, MLPs can model intricate relationships in the data, overcoming the limitations of single-layer perceptrons, which lack generalisation capabilities. The

perceptron is inherently incapable of classifying non-linearly separable patterns. Furthermore, it exhibits a predisposition towards overfitting and susceptibility to noise.

2.4.1.2 Activation functions

Activation functions are critical components in artificial neurons, introducing non-linearity into the neural network model. They enable the network to learn and approximate complex, non-linear relationships in the data. Common activation functions include the sigmoid, hyperbolic tangent (tanh), rectified linear unit (ReLU), and variants of ReLU, such as leaky ReLU and parametric ReLU, depicted in Figure 2.12.

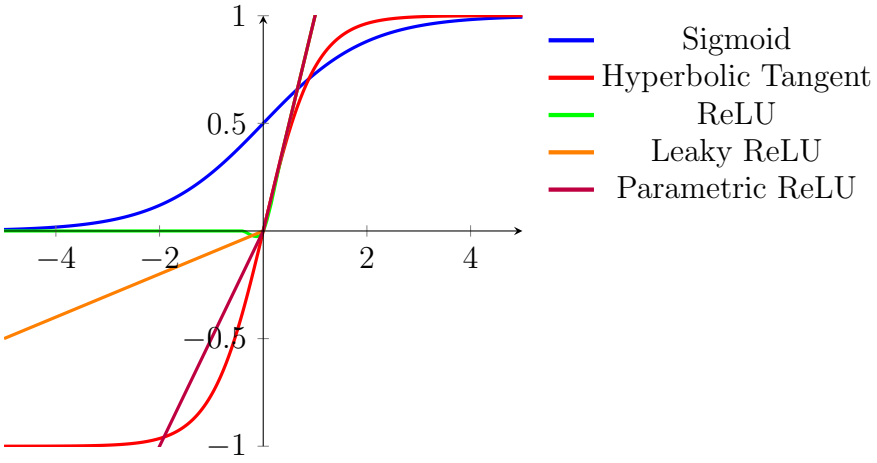


FIGURE 2.12: Graphical representations of various activation functions used in neural networks. The depicted functions include the Sigmoid function (blue), the Hyperbolic Tangent function (red), the Rectified Linear Unit (ReLU) function (green), the Leaky ReLU function (orange), and the Parametric ReLU function (purple). Each function exhibits unique characteristics, influencing the behaviour and performance of the corresponding neural network.

2.4.1.3 Fully connected layers

A fully connected layer, also known as a dense layer, is a layer in a DNN where each neuron is connected to every neuron in the adjacent layers. These layers

are responsible for integrating the learnt features and producing the network's final output, such as class probabilities in a classification task or real-valued predictions in a regression task. In deep learning architectures designed for computer vision tasks, fully connected layers are often placed after the feature extraction layers, such as convolutional or recurrent layers, to aggregate the learnt features and generate the desired output.

2.4.2 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) have emerged as a prominent class of deep learning models, particularly in the field of computer vision. They have demonstrated outstanding performance in various tasks, such as image classification, object detection, and semantic segmentation. The key innovation in CNNs is the incorporation of local spatial information and weight sharing to exploit the inherent structure in grid-like data. By having a significantly fewer parameters than fully connected layers, their adoption became rapid. A simple example of a CNN is depicted in figure 2.13.

2.4.2.1 Convolutional layers

The convolutional layer is the central component of a CNN, designed to capture local patterns and hierarchical features in the input data. It performs a convolution operation, where a set of learnable filters, or kernels, are applied to the input data in a sliding window manner. Each filter is responsible for detecting a specific feature or pattern in the input, such as edges, corners, or textures. Combining multiple filters allows a convolutional layer to learn a diverse set of features, which are then hierarchically aggregated in subsequent layers.

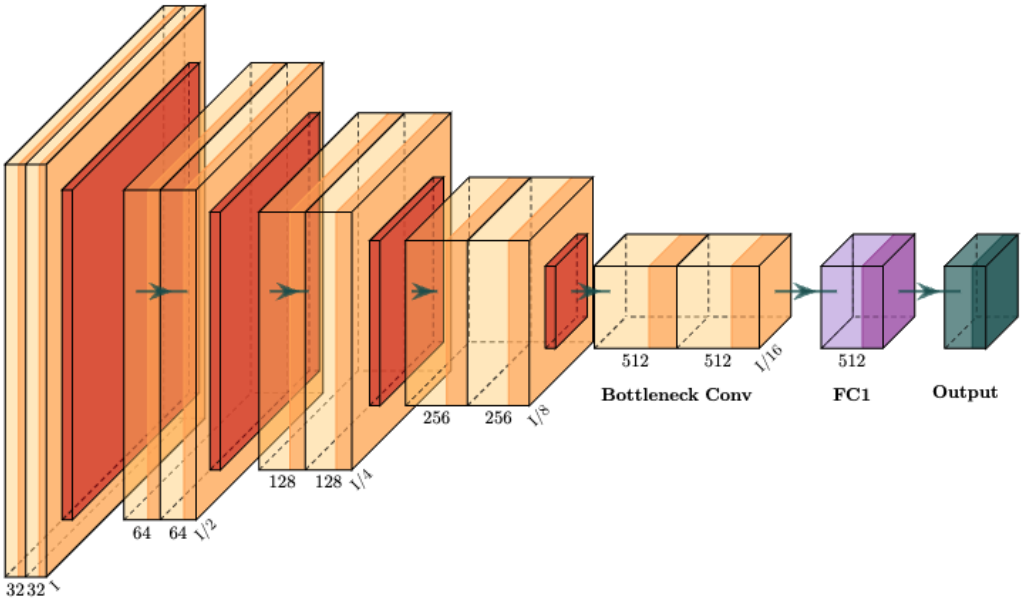


FIGURE 2.13: Illustration of a simple CNN. The convolutional layers are depicted in light orange followed by a pooling layer in stronger orange. After the bottleneck, a single fully connected layer is followed by the output layer.

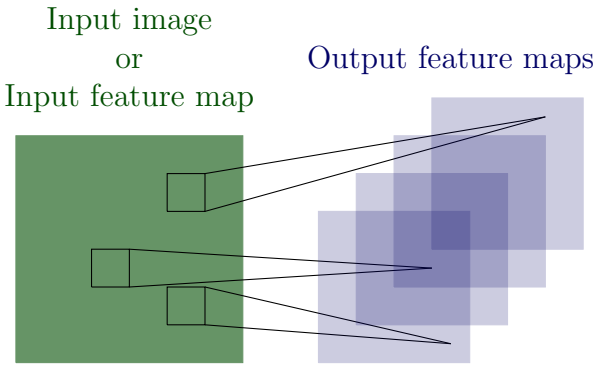


FIGURE 2.14: Illustration of a single convolutional layer. If layer l is a convolutional layer, the input image (if $l = 1$) or a feature map of the previous layer is convolved by different filters to yield the output feature maps of layer l.

2.4.2.2 Pooling layers

The pooling layers are another essential building block in CNN architectures, responsible for reducing the spatial dimensions of the input data while preserving the most critical features. By performing a downsampling operation, pooling layers provide a form of spatial variance, which means the output remains relatively invariant to small translations, rotations, or other forms of spatial

transformations. Moreover, this downsampling plays a pivotal role in mitigating the risk of overfitting. By reducing the number of parameters and thereby the complexity of the model and its computation, pooling layers help control the network's capacity, limiting its ability to memorise noise or irrelevant details from the training set. Consequently, this contributes to the model's improved performance on unseen, test data. The most common types of pooling operations are max pooling and average pooling, which compute the maximum or average value, respectively, within a local neighbourhood in the input data.

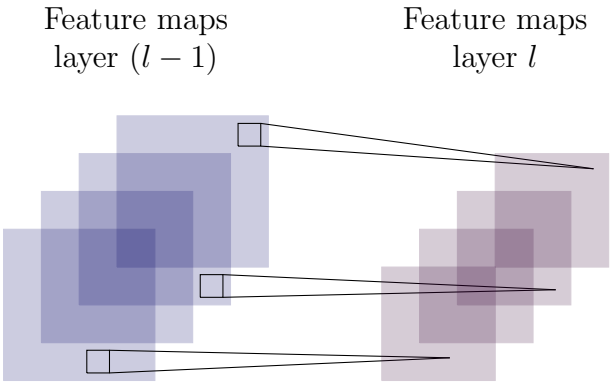


FIGURE 2.15: Illustration of a pooling and subsampling layer. If layer l is a pooling and subsampling layer and given $m_1^{(l-1)} = 4$ feature maps of the previous layer, all feature maps are pooled and subsampled individually. Each unit in one of the $m_1^{(l)} = 4$ output feature maps represents the average or the maximum within a fixed window of the corresponding feature map in layer $(l-1)$.

2.4.2.3 Notable architectures

Over the years, several influential CNN architectures have been proposed, which have advanced the state-of-the-art in various computer vision tasks. Some of the most notable architectures include LeNet-5 [42], AlexNet [43], VGGNet [44], GoogLeNet (Inception) [45], and ResNet[46]. These architectures have evolved to become deeper and more complex, incorporating novel techniques such as residual connections [46], inception modules[45], and dilated convolutions[47] to improve performance and address challenges associated with training deep networks.

2.4.3 Training a Neural Network

Training neural networks is a critical aspect of deep learning, aimed at tuning the model parameters to optimise its performance on a specific task. The process involves iteratively reducing a predefined loss function, which measures the discrepancy between the model's predictions and the ground truth.

2.4.3.1 Loss function

The loss function, also referred to as the objective function or cost function, quantifies the difference between the predicted output and the target for a given input. The goal of training a neural network is to minimise the loss function across the entire dataset. Various loss functions can be employed depending on the problem type and the desired properties of the model. Common loss functions include mean squared error (MSE) for regression tasks, cross-entropy for classification tasks, and more complex, task-specific loss functions for structured output prediction problems.

2.4.3.2 Backpropagation

Backpropagation or reverse mode automatic differentiation is one of the key algorithms used for training neural networks, including multi-layer perceptrons, convolutional neural networks, and recurrent neural networks. It is an iterative optimisation method based on the chain rule of calculus, which computes the gradient of the loss function with respect to each model parameter. The backpropagation algorithm consists of two main steps: the forward pass and the backward pass.

In the forward pass, the input data is propagated through the network to compute the output predictions and the loss function value. In the backward pass, the gradients of the loss function are calculated for each model parameter, starting from the output layer and moving towards the input layer. This is achieved

by applying the chain rule of calculus to compute the partial derivatives of the loss function with respect to each weight and bias in the network, in other words, on the input and output of each neuron.

2.4.3.3 Gradient descent

Gradient descent is a first-order optimisation algorithm used to minimise a differentiable function, such as the loss function in a neural network. The algorithm iteratively updates the model parameters in the direction of the negative gradient of the loss function, aiming to find the minimum of the loss surface. The learning rate, a hyperparameter in gradient descent, determines the step size taken in the direction of the negative gradient and significantly impacts the convergence rate and stability of the optimisation process.

Stochastic gradient descent (SGD) is a popular variant of gradient descent that updates the model parameters using the gradient computed for a single data point, chosen randomly from the dataset. In contemporary discourse, the term SGD is frequently used to denote mini-batch SGD. This signifies that the gradient is approximated utilising a modest subset of examples, as opposed to a singular instance. SGD introduces an element of randomness in the optimisation process, which can help escape local minima and converge faster than batch gradient descent. However, the random nature of SGD can also result in more oscillations and instability during training, which can be mitigated by employing techniques such as learning rate schedules, momentum, and adaptive learning rates.

2.4.3.4 Optimisers

Modern deep learning frameworks often employ more advanced optimisation algorithms, building upon the concept of gradient descent and SGD, to improve the training process's convergence rate, stability, and robustness. These

algorithms typically incorporate additional strategies, such as adaptive learning rates, momentum, and second-order information, to address challenges associated with the optimisation of deep neural networks. Some of the most popular optimisation algorithms include:

- **Momentum-based methods**: These algorithms, such as momentum SGD and Nesterov accelerated gradient, introduce a momentum term to the weight updates, which accumulates the gradients over time and helps the optimisation process to overcome local minima and oscillations.
- **Adaptive methods**: These algorithms, such as AdaGrad, RMSprop, and Adam (which also uses momentum), adjust the learning rate for each model parameter based on the historical gradients, aiming to achieve faster convergence and better performance in the presence of noisy or sparse gradients. These methods typically perform well in practice and are often recommended for training deep neural networks, predominantly Adam.

2.4.4 Deep Learning in Medical Imaging Analysis

Since segmentation is the most important part of the work conducted in this research, the particularities it has over the field of cardiac analysis are discussed. In particular, deep learning has been the gold standard for segmentation for almost a decade now.

The convolutional network is the most important type of network for this task, similar to other vision applications. In particular, U-shaped network variations have remained the go-to architectures during these past years.

To train these models, the basic and most used loss functions are the cross-entropy loss and the dice score loss for segmentation [48], and the mean squared error [49] in regression tasks such as in time frame labelling. Additionally, the Hausdorff distance serves as an essential metric for evaluating and comparing

the performance of various algorithms. This measure is notable for its tolerance to minor positional errors, yet it displays sensitivity to anomalies when objects are misclassified and located far from their correct regions [50].

The literature review will provide a more detailed overview of the area's progress (Section 2.6).

2.5 Cardiac MRI Datasets

During the last few years, a number of datasets have been made publicly available. Those datasets target different regions and contain different pathologies within them. In more recent years, datasets have focused on specific challenges, and the diversity held within the data is wider.

For the purposes of this research, the areas bearing the most significant clinical implications are the left ventricle (LV), the myocardium (Myo), and the right ventricle (RV). These critical regions are illustrated in Figures 2.16 and 2.17.

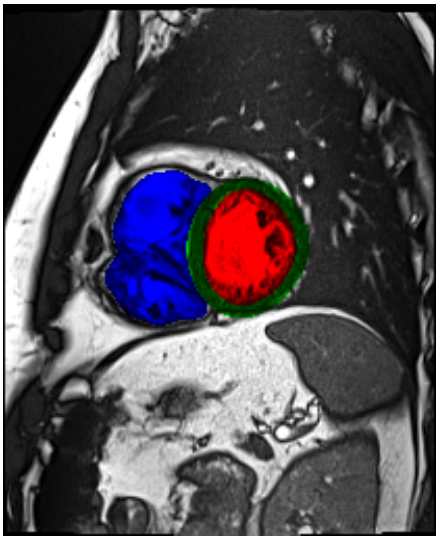


FIGURE 2.16: Example of a mid-ventricle short-axis view from a CMRI scan with annotations from experts. In red, the left ventricle; in green the myocardium; and in blue, the right ventricle. Scan from M&Ms2 [21].

The research was focused and based on publicly accessible datasets, which play an essential role in cardiac MRI research.. Publicly available datasets drive the

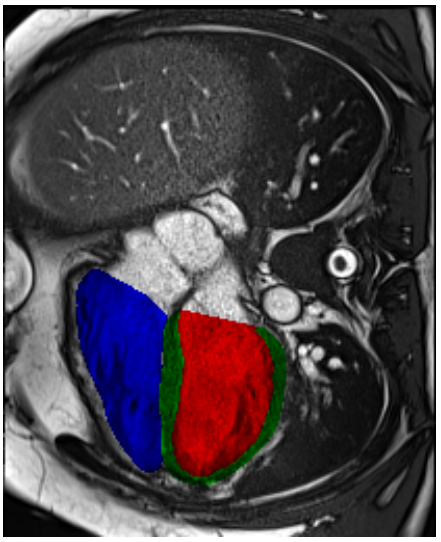


FIGURE 2.17: Example of a long-axis view from a CMRI scan with annotations from experts. In red, the left ventricle; in green the myocardium; and in blue, the right ventricle. Scan from M&Ms2 [21].

advancement of international research by offering a standardised reference point, promoting collaboration, and mitigating the barriers posed by the proprietary nature of much clinical data. Consequently, they contribute to the democratisation of data, which is particularly important in regions with limited resources for data collection.

Those datasets employ expert annotators who have the responsibility of rectifying clinical contours, thereby guaranteeing consensus among them regarding the ultimate outcome. These annotators strictly adhere to specific guidelines, which encompass comprehensive coverage of both the left ventricle and right ventricle cavities, including papillary muscles. Furthermore, the guidelines prohibit any interpolation of the LV myocardium at the base, and stipulate that the RV must exhibit a larger surface area during end-diastole compared to end-systole, while simultaneously avoiding the pulmonary artery. Additionally, the long-axis view is frequently employed as a frame of reference for precisely delineating the basal and apical regions.

Some of the most important public datasets are:

- **York** [51]: targets LV and contains cardiomyopathy, aortic regurgitation, enlarged ventricles and ischemia.
- **Sunnybrook** [52]: targets LV and contains hypertrophy and heart failure (with or without infarction).
- **Left Ventricle Segmentation Challenge (LVSC)** [53]: targets LV and contains coronary artery disease and myocardial infarction.
- **Right Ventricle Segmentation Challenge (RVSC)** [54]: targets RV and contains suspicion of arrhythmogenic, right ventricular dysplasia, dilated cardiomyopathy, hypertrophic cardiomyopathy and aortic stenosis.
- **Automated Cardiac Diagnosis Challenge (ACDC)** [55]: targets LV and RV and contains myocardial infarction, dilated hypertrophic cardiomyopathy and abnormal RV.
- **Multi-Centre, Multi-Vendor & Multi-Disease Cardiac Image Segmentation Challenge (M&Ms)** [56]: contains short-axis images and focuses on multi-vendor data (4 different scanners). Has contours for RV, LV and myocardium.
- **Multi-Disease, Multi-View & Multi-Center Right Ventricular Segmentation in Cardiac MRI (M&Ms-2)** [21]: contains short- and long-axis views. Focused on RV but having contours for LV, RV and myocardium. It is composed of healthy subjects and patients with seven different pathologies.
- **Extreme Cardiac MRI Analysis Challenge under Respiratory Motion (CMRxMotion)** [25]: contains short-axis images with different grades of motion artifacts. Provides ground truth for LV, RV, and myocardium.

In general, the patient data within a CMRI dataset comprises a stack of short-axis images or a single long-axis image for each time frame. The recordings

typically contain between 10 and 30 time frames and between 10 and 15 slices for short-axis stack, with sizes ranging from 180x180 pixels to 360x360 pixels. Additionally, depending on the dataset, further information about the characteristics of the subject may be included. Metadata containing the scanner parameters, which is essential for recovering the anatomical position of the voxels, and voxel sizes accompanies the scan. Finally, the images typically have a bit depth of 16 bits per pixel, although in some cases, they may have a bit depth of 32 bits per pixel.

Where segmentation annotations are provided, they have the same characteristics as the corresponding scan. The segmentation masks are mostly available for key time frames (ED and ES). These annotations adhere to Standard Operating Procedures (SOP). For instance, in the case of [21]:

- LV and RV cavities must be completely covered, with papillary muscles included.
- No interpolation of the LV myocardium must be performed at the base.
- RV must have a larger surface in end-diastole compared to end-systole and avoid the pulmonary artery.
- Additionally, long-axis view is used as a reference to delimit the basal and apical regions.

2.6 Literature review

This section gives a general overview of the cardiac ventricle segmentation landscape, shows the dependence on labelled data and the domain specific performance that the models have, and introduces the main work done within the topic to improve clinical tools regarding ventricle segmentation. An extensive literature review is provided for each piece of research in the corresponding chapters.

During the last decade, advances in ventricle segmentation have been consistently based on public datasets, such as LVSC [53] and RVSC [54] among others [51, 52, 57]. Those datasets consist of short-axis images annotated by expert clinicians. In this scenario, researchers have investigated different methodologies and, as in many other domains, deep neural networks [43] were adopted. While fully convolutional networks were the main architecture [58, 59, 60], specifically U-Net [61] was preferred after its inception from 2017 on [62, 63, 64]. The same year, the Automated Cardiac Diagnosis Challenge (ACDC) [55] took part, and after the results were released, the ensemble of 2D and 3D U-Net of [63] won. The researchers then started asking the question if the problem was solved [55]. The author of [65] subsequently raised several limitations with these methods, and a new paradigm of challenges arose.

More recently, other datasets have been produced to target more specific deficits in the state-of-the-art, benefiting from the progress made in the ACDC challenge. The Multi-Centre, Multi-Vendor & Multi-Disease Cardiac Image Segmentation Challenge (M&Ms) [56] focused on the analysis of the performance of models when scans from unseen vendors are present. In general, the combination of MRI-specific data augmentations and normalisations [18, 19, 20], and a U-shaped architecture produced the best results [66, 67, 68]. In Chapter 3, an unsupervised method is proposed to benefit from image registration techniques [69, 70] to extend clinical annotation to unlabelled time frames to improve the performance on unseen scanners. The results are reported in [1], where an improvement of 2.6% was achieved compared to baseline results.

After the success of the first M&Ms challenge, the second edition was held [21], this time focusing on the right ventricle in a multi-view perspective. In general, the techniques used were similar to previous challenges [71], in addition to the adoption of new modules to benefit from both views [72, 73], and the merging of the feature representations of both views in bottlenecks [74, 75]. Among all architectures, the self-configuring framework proposed by [76] scored the highest and was adopted by different methods [77, 78]. In Chapter 4, different

architectures for each vendor and pathology on each view are analysed, and pixel-mapping post-processing is proposed to benefit from multi-view data.

While improved segmentations yield better morphological and functional analysis tools, and better ejection fraction calculations, the correct time frame detection received much less attention. Since 2016, when [23] was proposed, no further investigation has been performed despite the fact that a small error in this task can greatly affect the calculation of the EF [22]. In Chapter 5, two different approaches based on public data are proposed to correctly label the different frames within the heart cycle, with the inference time improved by a factor of 500.

More recently, the artifact correction topic has received more attention, and a dataset with respiratory motion artifacts has been made publicly available, the Extreme Cardiac MRI Analysis Challenge under Respiratory Motion (CMRxMotion) [25]. Several works aim to correctly segment these scans with approaches similar to what has been mentioned above; some did multitask methods [79, 80] or used high quality pseudo-labels [81]. Chapter 6 reports the approach of using transferred learning and MRI specific data augmentation to correctly segment the scans under the presence of artifacts of different intensities. The results are reported in [2], where it was found that the above-mentioned approach is very competitive in the presence of respiratory motion artifacts..

Moreover, within the realm of generating synthetic CMRI images, recent studies have indicated numerous advantages in incorporating such data for the purpose of balancing datasets [82]. These studies predominantly relied on either Generative Adversarial Networks (GANs) [26] or Variational Autoencoders (VAEs) [27]. Noteworthy is the emphasis in previous research on direct data augmentation, rather than generating data from sources distinct from existing scans [30, 29, 83]. These approaches draw inspiration from advancements in other computer vision domains, such as Style GANs [84] or Cycle GANs [85]. In Chapter 7, a novel source of synthetic data, the atlas deformation, is introduced. Within

this chapter, style transfer is applied to slices derived from deformed versions of the atlas.

While other imaging modalities like CT have recently seen the development of foundational models that have significantly impacted the field [86], CMRI still grapples with challenges before achieving a similar feat. The proposition of foundational models often accompanies the concept of multi-modal models [87]. Developing a successful foundational model entails pre-training a model with ample data and testing it on benchmark datasets. If the performance proves sufficient, the model can then be fine-tuned with smaller, more specific datasets [16]. Although initial attempts have been made across multiple imaging modalities, including MRI, the formulation of a foundational model tailored to CMRI remains an ongoing potential.

2.7 Summary

This chapter contains an introduction to basic anatomical and functional concepts about the human heart (Section 2.1), the fundamentals of the cardiac MRI acquisition process (Section 2.3) and a brief presentation of the principles of neural networks and their applications to computer vision applications (Section 2.4), with emphasis on the segmentation task. In Section 2.5, an example of the data utilised during the research is depicted, along with a review of the available public datasets and their peculiarities. The literature review presented in Section 2.6 addresses the more relevant works in the area of cardiac MRI segmentation.

In subsequent chapters, the results achieved through the course of this research are described and discussed. Furthermore, each chapter encompasses extended literature reviews focused on specific sub-topics, accompanied by comprehensive descriptions of the employed methodologies and experiments conducted.

Chapter 3

Advancing Model Generalisation with Synthetic Label Propagation

3.1 Introduction

This chapter provides a detailed description of the contributions of this research to cardiac MRI segmentation using synthetic label propagation, along with a complete description of the tools and methods used during the experimentation. This research includes the synthetic labelling of intermediate time frames by using image registration, to increase the available data for the training. The objective is to increase the models' generalisation to segment the end-diastolic and end-systolic time frames from scans from unseen vendors. In addition, the next steps to be researched and the applications of the research outcomes are discussed. The experiments and results described in this chapter were published the Irish Machine Vision and Image Processing Conference (IMVIP), 2021 [1].

The following section describes the adoption of semi-supervised learning techniques, particularly image registration methods, while targeting the lack of capacity in computer vision models to generalise across scans obtained from different vendor machines. Section 3.2 gives an overview of previous attempts to adopt these learning techniques and the state-of-the-art tools in the image registration topic. Section 3.3 provides a detailed description of the proposed method. Section 3.4 describes the experiments, the evaluation procedure, and the results obtained. Finally, Section 3.5 presents the conclusions and suggestions for future improvements.

In Chapter 1, it was discussed how cardiac image segmentation serves as an important first step for many approaches to quantitative analysis for cardiac diagnostic evaluation. This process requires partitioning the image into a number of clinically meaningful regions such as the left ventricle, right ventricle, or myocardium. Acquiring this information allows clinicians to evaluate important metrics, such as the ejection fraction and the volume that the heart is managing at different times. These metrics are then used to determine whether there is any possible pathology and how bad it is [8].

Later, in Chapter 2, the exploration involved how data poses a key challenge when using off-the-shelf algorithms in this area, specifically due to the limited amount of annotated data available and its quality. Many researchers report that they struggle to achieve improved results with existing annotated data, especially when working with open datasets [65]. Difficulties range from low availability of data to domain shift using data from a certain scanner vendors, to images from patients with rare conditions.

There is growing interest in the community in understanding how to transfer models that work well for specific scanners to unseen ones. In this work, these challenges are addressed using the M&Ms challenge dataset, which is considered to be the representative dataset for this issue [56].

Convolutional neural networks (CNNs) are the most common type of deep neural networks for image analysis and have advanced the state-of-the-art in many object segmentation tasks, including in the medical imaging domain. In particular, U-Net [61] is the architecture with the best results in most of the challenges that cardiac MRI currently faces [65]. Therefore, it has been selected for the experimentation.

Registration is a fundamental process in the field of image analysis, encompassing a range of techniques used to align and spatially transform images acquired from different sources or at different time points. The primary objective of registration is to establish a correspondence between corresponding anatomical or functional structures in the images, facilitating comparative analysis and integration of information.

In the context of image registration, an atlas refers to a reference image or a template that serves as a common anatomical or functional space to which other images are aligned. On the other hand, a registration field, also known as a deformation field or a displacement field, characterises the spatial transformation required to align an individual image to the atlas or another reference image. The registration field represents a mapping of each point in the individual image to its corresponding location in the atlas or reference space. It captures the geometric differences and deformations between the images, enabling spatial alignment.

Recently, [66] proposed a method where labels can be propagated using image registration in an unsupervised manner. Those labels are used to enhance the process where a 3D U-Net learns features. Although the results ranked second in the M&Ms challenge [56], the study did not test these propositions in 2D networks, a key novelty of this work, and the authors also did not consider different tools for the registration part.

Contributions: this work proposes a method to enhance short-axis cardiac magnetic resonance segmentation by synthetically labelling volumes without

annotations, by propagating the existing labels with a registration field. The proposed system works by registering the labelled volumes of the end-systolic (ES) and end-diastolic (ED) time frames and using the warping field over their ground truth labels. This helps the network learn key features from those volumes, which are important towards increasing the network's performance over different scanners. This study extends recent work where the propagation of labels is used in a 3D U-Net and using a different registration tool. In contrast to that work, the investigation into whether this technique is valid for 2D U-Net applications that can be used in less computationally powerful machines is undertaken, with expectations of achieving better performance, considering that the low through-plane resolution limits the potential of 3D networks.

3.2 Related work

The clinical interest in ventricle segmentation has pushed the community toward improving performance in this task. With the advent of deep learning, a succession of different approaches was investigated. One of the first was the usage of a fully convolutional network [58]. From there, many works have improved networks, increasing the learning capacity for segmentation [62].

In this context, the effectiveness of 3D networks has been hindered by the low through-plane resolution and the presence of motion artifacts [64]. Moreover, the utilisation of data in the 3D modality results in a significant reduction in the number of available training examples, thereby posing challenges to the training process. Furthermore, the computational demands of 3D networks necessitate a larger GPU memory capacity, which may necessitate substantial downsampling of the data in situations where computational resources are limited. Consequently, this downsampling process has the potential to compromise the integrity of the information, leading to a potential loss of critical details.

Annotated data is the foundation of fully supervised approaches, but at the same time, it is scarce and costly to obtain. Moreover, annotating medical images requires significant expertise and manual effort and, even then, can lead to noisy labels and/or biases. In addition, there is an imbalance between the amount of available data on different sources of scanner, target pathology, or scanning parameters. All the above-mentioned challenges point strongly towards the need to design methods where the accuracy of the networks is maintained compared to fully supervised approach, even when there are significantly less data available.

Several works in the literature proposed unsupervised or semi-supervised techniques to overcome the scarcity of labels. One approach was to use a scribble annotation consisting of a set of quickly drawn labels, and recursively retraining the network using the output segmentation, including a conditional random field and an uncertainty estimator [88]. The advances in the calculation of optical flow inspired the usage of these algorithms for motion estimation and this has been useful in multitask approaches where a Siamese network performs motion estimation and segmentation simultaneously, exploiting the information contained in unlabelled data [89].

VoxelMorph [70] is an image registration tool for alignment and registration that can also model deformations. It has been used for atlas-based registration [90] and for probabilistic diffeomorphic registration [91]. Given its proven reliability [92, 93, 94], it was adapted for the registration part of this study.

3.3 Methodology

This section describes the proposed method, detailing the different steps and implementations of all the tools involved in the process. Distinguishing two first steps corresponding to the registration and label propagation, and two last parts

corresponding to training and testing the two different models i.e. the original dataset and the one with the addition of the synthetic labels.

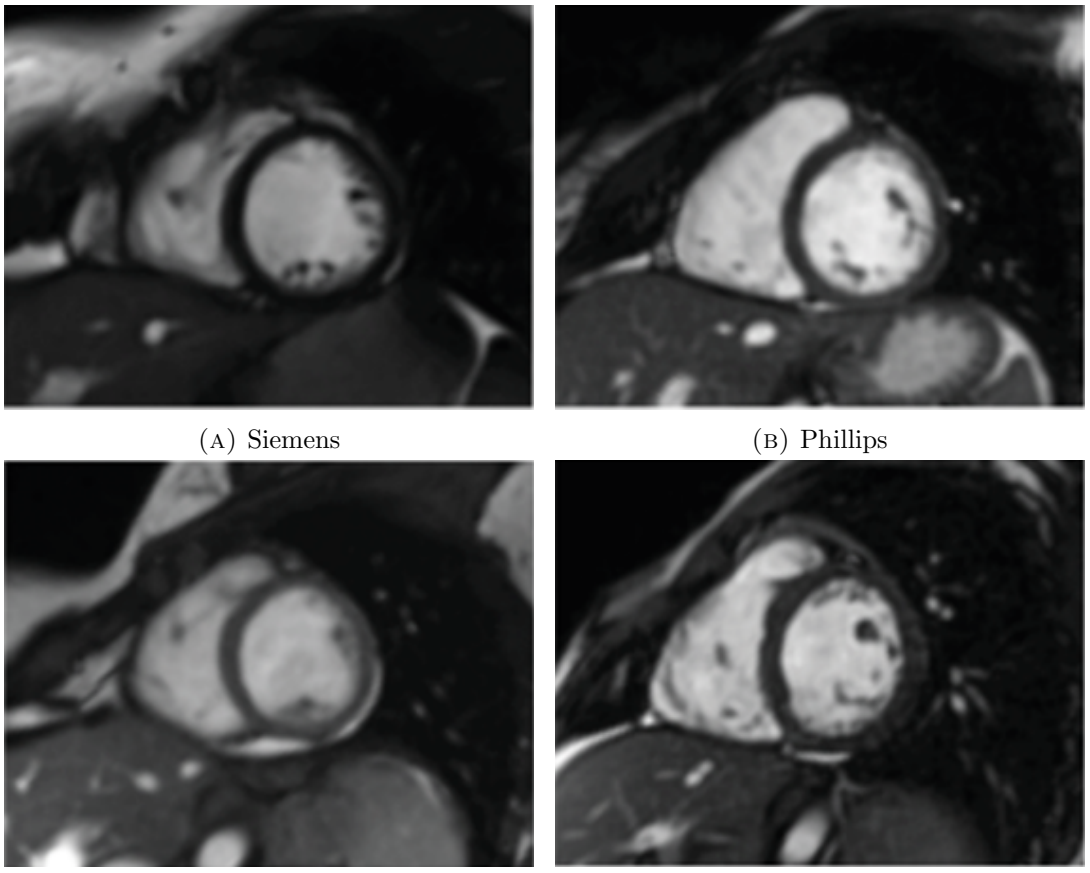
The aim of the proposed methodology is to enhance the performance of the models when applied to previously unseen datasets, specifically scans obtained from vendors that were not included in the training set. During the training phase, unlabeled time-frames belonging to the systolic phase were utilised, and the corresponding labels were subsequently obtained for these frames through the label propagation process, which is elaborated upon in Section 3.3.2. The evaluation of the results, as presented in Section 3.4, is conducted based on the labelled time-frames within the test set, as annotations are unavailable for the remaining time-frames.

3.3.1 Data

The M&Ms dataset was used for the experimentation, released during the Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Image Segmentation Challenge. This dataset consists of 345 patients with hypertrophic and dilated cardiomyopathies and healthy subjects, and it also allows the opportunity to measure performance over different scanners, and to evaluate the resilience of this approach towards different data sources. The visual appearances of CMRI scans for each vendor are presented in Figure 3.1.

In particular, the training set includes 150 cases (75 from vendor A and 75 from vendor B). The test set includes cases from two more vendors (C and D), see Figure 3.2.

The CMRI images have been segmented by experienced clinicians from the respective host institutions, including contours for the LV and RV blood pools, as well as for the left ventricular MYO. Only ED and ES time frames have been annotated, see Figure 3.3.



(A) Siemens (B) Phillips

(C) General Electric (D) Canon

FIGURE 3.1: Visual appearance of a CMRI short-axis middle slice for anatomically similar subjects in the four different vendors considered. Images reproduced from the M&Ms Dataset Challenge [56].

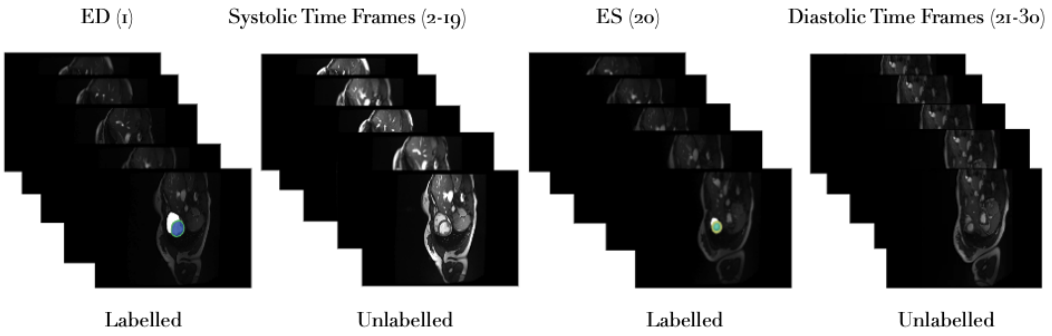


FIGURE 3.3: M&Ms [56] challenge data overview for a single patient. ED and ES frames are labelled while Systolic and Diastolic time frames are not. Time-frame ranges can vary from subject to subject.

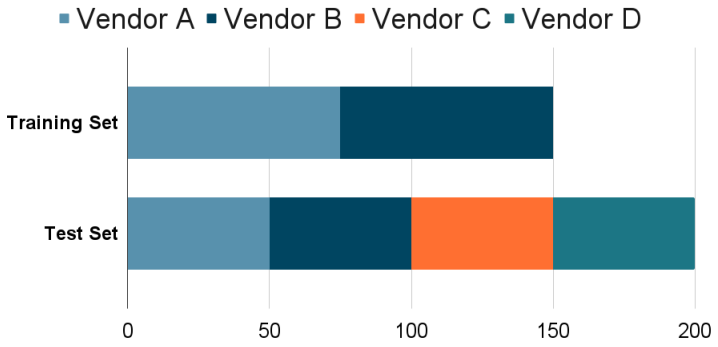


FIGURE 3.2: M&Ms [56] challenge data: number of subjects for each vendor for each set. The training set includes scan vendors A and B, while scans from vendors C and D are just present in the test set.

3.3.2 Label Propagation

VoxelMorph [69] was adapted as the selected image registration technique. Given that the pre-trained models were trained with images from other tissues, the decision was made to train one using publicly available CMRI data. The ground-truth segmentation labels were also included in the training process to produce a model that was later used for label propagation. The standard configuration of Voxelmorph was used for this part of the process(TensorFlow version)[1].

After training, all the warping fields for the time frame between end-diastole and end-systole were computed (systolic time frames), always in an intra-subject way. The propagation was carried out from the ED mask to the remaining systolic time frames. The warping fields were used to modify the ground truth for the ED time frame, obtaining the synthetic labels for the target time frames, see Figure 3.4.

[1]VoxelMorph https://github.com/voxelmorph/voxelmorph.

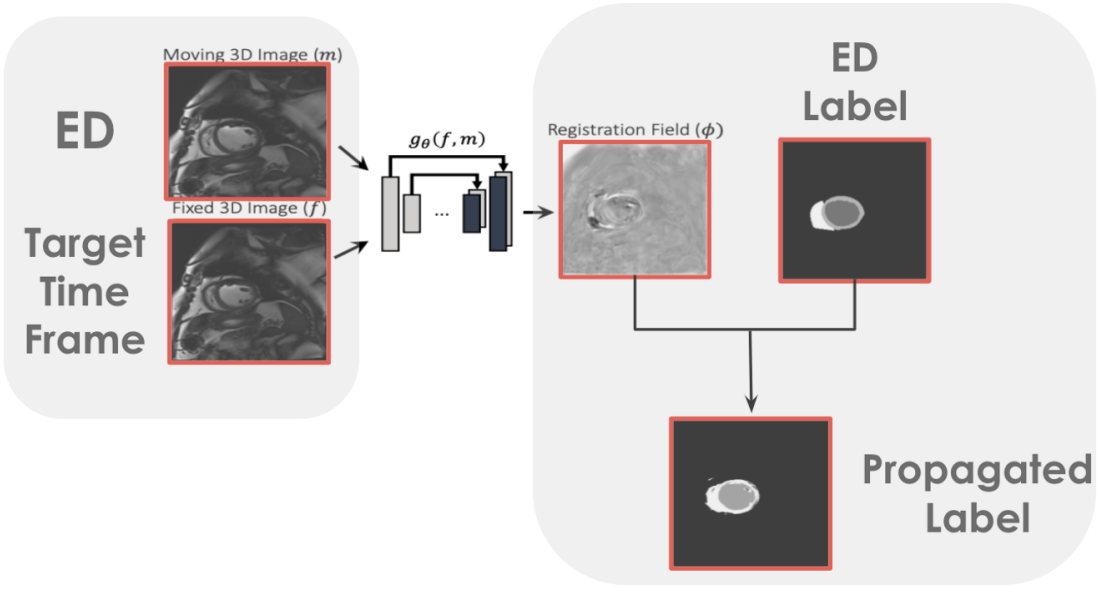


FIGURE 3.4: Semi-supervised cardiac MRI segmentation using image registration: an overview of the registration scheme. The registration network learns to output a registration field between the ED frame and the target one. Then this registration field is used to deform the ED label, obtaining a synthetic label for the target time frame, namely the propagated label. Figure adapted from [69], light grey background correspond to the changes.

Lastly, all the volumes with annotations, including the original and the synthetic ones, were combined into a single dataset.

3.3.3 Segmentation

This approach uses a 2D U-Net to segment end-diastolic and end-systolic volumes in the target dataset. In particular, the U-Net model used 32 feature channels with kernel size equal to three at the first level of the convolutional layers, where batch normalisation was also applied and the activation function was ReLU, followed by a 2×2 max-pooling layer. The final 1×1 convolution is set with four channels that match the four target regions: background, right ventricle, left ventricle, and myocardium.

For training the model, an Adam optimiser was used (learning rate = 10^{-3}) with a plateau learning rate scheduler, and the cross-entropy loss function was

employed. Additionally, the Dice score for all regions was computed in each epoch for the validation set.

3.4 Results

This section presents the results of the models obtained in Section 3.3. The results were computed using solely the original labels of the test set of the M&Ms dataset.

Table 3.1 shows the results obtained for both datasets, including regions and vendors (M&Ms). The proposed method performs 2.6% better than the baseline, where the region that benefits most from this change is the region corresponding to the myocardium. When comparing the difference in performance of both models across different scanners, it is observed how the performance of unseen scanners C and D improves in the model resulting from the proposed method, while it sees its performance almost untouched for the scanners present in the training set (A and B). It can be seen in Figure 3.5 the qualitative impact of the proposed method, which has the potential to decrease in time in clinical reporting sessions compared to the baseline.

	Regions			**Vendors**				**Total**
	LV	**Myo**	**RV**	**A**	**B**	**C**	**D**	**Dice**
Baseline (M&Ms)	0.475	0.386	0.394	0.553	0.583	0.419	0.277	0.418
Proposed (M&Ms)	0.504	0.430	0.398	0.548	0.573	0.443	0.338	0.444

TABLE 3.1: Results (dice scores) on the M&Ms datasets. Higher is better.

While [66] obtained a DICE score of 90.9, 84.53 and 87.86 for the LV, MYO and RV respectively in the test set, their improvement in comparison to the baseline without the label propagation was just of a 0.75%. This improvement is extracted from the performance in the validation set, given that no comparison to the baseline was given in the test set.

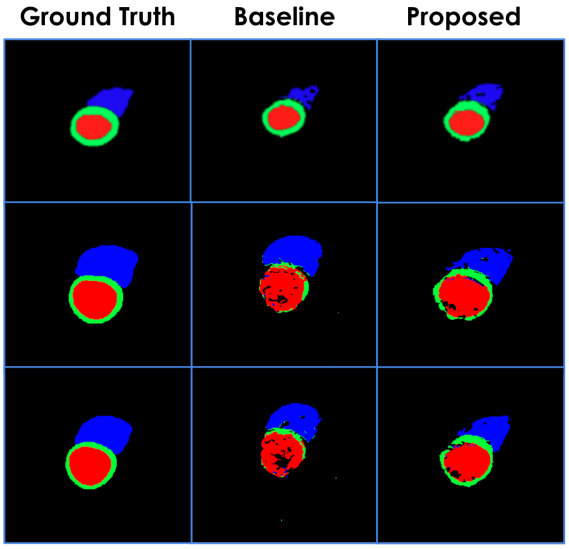


FIGURE 3.5: Qualitative Results for the semi-supervised cardiac MRI segmentation using image registration: (Red) Left Ventricle, (Green) Myocardium, and (Blue) Right Ventricle.

3.5 Discussion and Conclusions

Experimental results using synthetic labels generated from VoxelMorph improved over the baseline, demonstrating the potential to use diffeomorphic image registration as a label propagation technique. Future work should investigate the use of this technique with a more robust baseline, including the propagation of the labels from both key time frames. In particular, it would be interesting to adopt nnUNet [76], an automatic segmentation framework for medical images based on U-net architectures, which has been demonstrated to give state-of-the-art results. Finally, it is anticipated that the results could be improved by leveraging research focused on unreliable or noisy labels (e.g. [95]), weighting them differently to reflect the level of trust in the annotations, or post-processing the synthetic annotations.

The success of this work in using synthetic annotation opens the possibility for other kinds of synthetic annotation or data. Other techniques should be inveatigated that could benefit from available data to extend the capacity of the models by balancing the number of subjects from specific cohorts (different pathologies, ages or gender).

In the next chapter, an architecture study is conducted to investigate the usage of nnUNet [76] and other convolutional neural networks in a multi-view multi-pathology scenario, which provides more detail regarding the optimal setting for the segmentation of cardiac MRI. Additionally, a post-processing technique is tested to maximise the usage of resources, in this case, the metadata, to link the segmentations of the two different views.

Chapter 4

Multi-view Segmentation of Cardiac MRI

4.1 Introduction

This chapter presents a study on the segmentation of multi-view, multi-disease, and multi-vendor Cardiac MRI scans. Experimentation was performed using public data from the Multi-Disease, Multi-View and Multi-Center Right Ventricular Segmentation in Cardiac MRI Challenge (M&Ms-2) [21]. This research includes the usage of multiple architectures, frameworks and proposes a post-processing method based on pixel-mapping using anatomical metadata.

Following the motivation introduced in Chapter 1, Cardiac MRI segmentation is an important step during quantitative analysis for the diagnosis of cardiovascular diseases. During the process, the image is partitioned into meaningful regions. Acquisition of this information helps clinicians calculate and understand important features, such as the ejection fraction, which are then used to determine whether subjects have a particular pathology or if surgery is necessary [8]. Other functional analysis, such as strain, flow, or torsion, also require or benefit from the segmentation being precise. The contours extracted from the

segmentation can be used to compute the left ventricle wall thickening or strain. Those indexes are important in evaluating ischemia, myocardial infarction and ventricular dyssynchrony [96].

Given the limited data, usually biased by under-representing specific cohorts of subjects such as patients that have been scanned with less common scanners or that have pathologies that are more scarce, clinical interest has arisen towards adopting automated methods to help them diagnose faster and more accurately the wide range of data variety. In particular, improving the generalisability of models for scans from different vendors, with different protocols or subjects with different pathologies has become an important challenge.

Convolutional networks are the most common type of neural networks used for this task, but their adoption in the clinical world is slower than in others and should be carefully analysed due to their very important role in the clinical workflow. Hence, all algorithms are validated accordingly, and further efforts on interpretability could greatly benefit from research in other domains. Chapter 2 provides more information in this context.

4.1.1 Contributions

This work propose a post-processing method valid for any multi-view data that includes a header with anatomical correspondences with the scanner coordinates.

Second, a precise study over four different architectures has been made for short-axis and long-axis images, carefully analysing the performance over the different vendors and diseases subsets of data, bearing in mind if they were present or not in the training set.

Last, the prediction time was recorded; therefore, the accuracy can be analysed in parallel with its computational cost, easing the adoption of these algorithms for clinical tools and machines with low computational power.

4.2 Related work

In the following chapter, the literature related to the experimentation is discussed, situating the research performed on the topic of cardiac MRI segmentation, the available public data, and the current state-of-the-art for the tasks involved.

In particular, the focus will be on available architectures and frameworks, as well as on the data preprocessing that has demonstrated the best results.

4.2.1 Cardiac MRI segmentation

The field of cardiac MRI segmentation has received a lot of attention in the last decade and has adopted several deep learning strategies to automate daily clinical workflows. Importing those innovations implied re-designing parts of those methods, and eventually, dedicated models to medical imaging segmentation arose.

With these adoptions, new challenges and opportunities appeared; they can be summarised in: (1) lack of annotated data, (2) generalisation to unrepresented subsets, and (3) interpretability of the results [65]. In this work, the research will be focused on generalisation, particularly to subsets of vendors and diseases.

The ACDC dataset [55], has been the most widely used dataset to benchmark the performance of algorithms. The results show how 2D CNNs have achieved the best performance metrics in the proposed cohort of subjects. However, recent public data publication has been made to leverage performance in underrepresented cohorts (M&Ms [21]), promoting research towards the above-mentioned challenge 2 [1].

4.2.1.1 Architectures

While CNNs have been the best performing neural network module for segmentation, other techniques have been added to improve their performance.

In particular, U-Net [61] adopted skip connections and a U shape to extract multiscale features and preserve contextual information, where DeepLabv3 [97] approached this using the atrous spatial pyramid pooling.

More recently, the adoption of transformers in transUnet [98] was suggested to preserve global contexts. On the other hand, CE-Net [99] added atrous convolution and inception-like structures to preserve that context.

- **nnU-Net:** Aiming to tackle the non triviality of crafting solution to biomedical image segmentation, nnU-Net [100] proposed a self-configuring framework that includes pre-processing, architecture, training, and post-processing for a given task. Built-in options offer 2D, 3D in cascade and in full resolution, and ensembles of the previous U-Net based methods.

4.2.1.2 Multi-view segmentation

In the context of multi-view segmentation during the M&Ms2 Challenge [21], Tampera [101] was proposed, implementing a geometric spatial transformer to map the segmentation from one view to condition the segmentation of the other view.

4.2.1.3 MRI specific normalisation and augmentation

MRI data has many particularities due to the nature of its acquisition, and therefore it shall be managed accordingly, pre-processing it with techniques that maximise the performance yielded by each costly image. In this work, Torchio [102], a Python library for efficient loading, preprocessing, and augmentation,

was used to implement this preprocessing, which included different normalisations and augmentations that are presented in the following sections.

The following normalisations and augmentations have demonstrated to greatly benefit state-of-the-art methods in the context of the MRI segmentation task [103].

- **Z-Normalisation:** It consists of subtracting the mean and dividing by the standard deviation. Allowing for centering and scaling the input data, which will ease and improve the training process.
- **Histogram normalisation:** Perform histogram standardisation of intensity values as suggested in [104], see Figure 4.1.

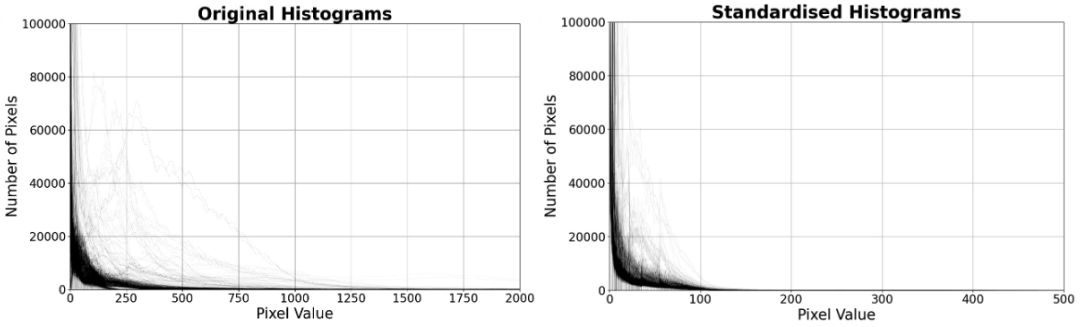


FIGURE 4.1: Histogram Standardisation of pixel values from the SA images in the M&Ms2 dataset [21].

- **Random motion:** Magnetic resonance images suffer from motion artifacts when the subject moves during image acquisition. This transform follows [105] to simulate motion artifacts for data augmentation.
- **Random ghosting:** Discrete "ghost" artifacts may occur along the phase-encode direction whenever the position or signal intensity of the imaged structures within the field of view vary or move in a regular (periodic) fashion. The flow of blood, cardiac motion, and respiratory motion are the most important patient-related causes of ghost artifacts in clinical magnetic resonance imaging [106].

- **Random bias field:** MRI magnetic field inhomogeneity creates intensity variations of very low frequency intensity variations throughout the image. The bias field is modelled as a linear combination of polynomial basis functions, as in [107]. It was implemented on NiftyNet by Carole Sudre and used by [108].

4.3 Methodology

In this section, the setup of the proposed methods is presented, detailing their components and hyperparameters.

4.3.1 Data

For this thesis, the M&Ms2 dataset [21] was used, which was recently released during the Multi-Center, Multi-View and Multi-Disease Right Ventricular Segmentation in Cardiac MRI Challenge. Although the challenge focused on the right ventricle, the performance of the various methods was evaluated in all chambers.

This dataset contains short-axis and long-axis images from each subject (for a total of 160), as well as annotations of the right ventricle for each, and an example of both views can be seen in Figure 4.2 and Figure 4.3.

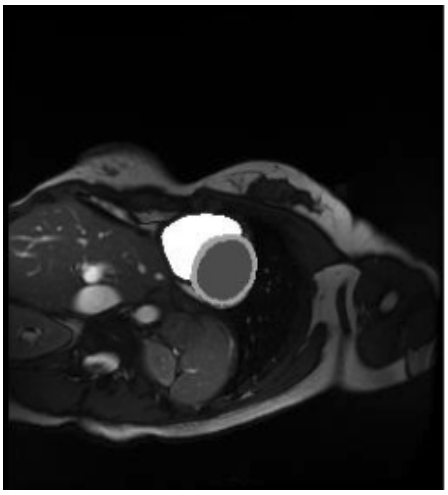


FIGURE 4.2: Example of a short-axis (SA) image. Right ventricle (white), myocardium (grey), and left ventricle (dark grey).

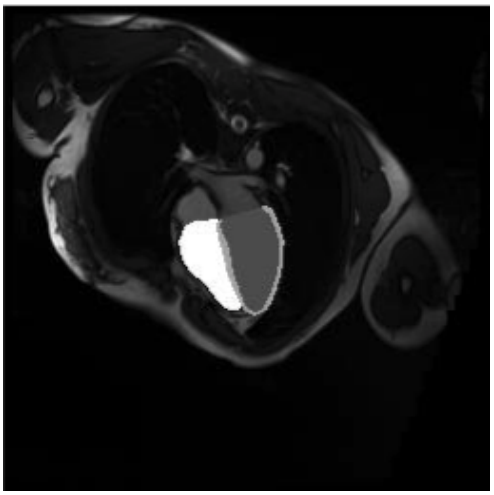


FIGURE 4.3: Example of a long-axis (LA) image. Right ventricle (white), myocardium (grey), and left ventricle (dark grey).

The challenge cohort consisted of 360 patients with different pathologies of the right and left ventricles, as well as healthy subjects. All subjects were scanned at three clinical centers in Spain using three different magnetic resonance scanner vendors (Siemens, General Electric, and Philips).

The training set contained 200 annotated subjects, and the testing set contained 160 more. Two pathologies (tricuspid regurgitation and congenital arrhythmogenesis) were not present in the training set, but were in the validation and testing sets to evaluate generalisation to unseen pathologies.

The complete list of pathologies is:

- Dilated Left Ventricle (LV).
- Hypertrophic Cardiomyopathy (HCM).
- Congenital Arrhythmogenesis (ARR).
- Tetralogy of Fallot (FALL).

- Interatrial Communication (CIA).
- Dilated Right Ventricle (DRV).
- Tricuspidal Regurgitation (TRI).

4.3.2 nnU-Net and post-processing

In the following section, the parameters chosen for nnU-Net are discussed, followed by the introduction of a post-processing technique developed to improve the results.

4.3.2.1 nnU-Net setup

For the experimentation, the regular 2D version of nnU-Net was used, without any other version or combination, and training was not conducted using cross-validation; therefore, FOLD was set to ALL. An individual model for short-axis and a model for long-axis images were trained.

The combination between Cross-Entropy loss and Dice loss was chosen as a loss function.

The rest of the parameters of the framework were left as default.

4.3.2.2 Post-processing

To boost the efficiency of this configuration, a post-processing technique that leverages predictions from both perspectives to enhance segmentation along the long-axis was implemented. This was a strategic move to capitalise on the superior performance typically observed in the short-axis view, which boasts a larger sample size. The implementation of a post-processing pipeline was

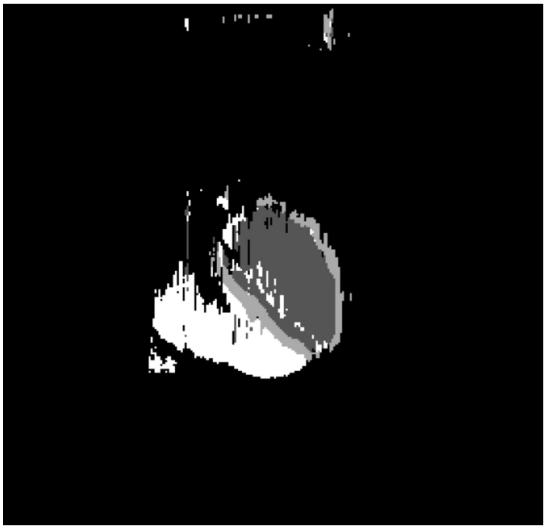


FIGURE 4.4: Example of the original predictions for LA images in the nnU-Net pipeline.

prompted by persistent spurious predictions in the LA images within the nnU-Net pipeline, despite efforts to address them. An example of these problematic predictions is shown in Figure 4.4.

The post-processing consisted on mapping the voxels from the short-axis view into the long-axis view. This operation changed depending on the scanning parameters, consequently the metadata held in the header file of the NIfTI [109] format (the source format for the chosen dataset) was extracted.

NIfTI information and methods

The NIfTI format, short for Neuroimaging Informatics Technology Initiative format [109], is a widely-used file format for storing and sharing neuroimaging data. It was introduced as a replacement for the earlier Analyse format, which was limited in its ability to represent complex data structures and lacked support for modern imaging techniques.

One of the key advantages of the NIfTI format is its flexibility and extensibility. It can represent a wide range of neuroimaging data, including structural and functional MRI, diffusion-weighted imaging, and positron emission tomography (PET) data, among others. It also supports the use of extensions, which allow

researchers to add additional information to the header of the file, such as study-specific metadata or additional image data.

Continuous coordinates can be associated with voxels utilising three distinct methods, named methods 1, 2, and 3. This discourse primarily highlights the context of three-dimensional volumes, in which these continuous coordinates are represented by the variables (x, y, z). The voxel index coordinates, correspondingly depicted as (i, j, k), are essentially the array indices.

The qform is the coordinate transformation that takes the coodinates of the image into the coordignates of the magnet.

The sform is a coordinate transformation that aligns the images to a standard reference.

The q or s form code determines the coordinate system used as the reference: 0 is arbitrary (apply Method 1), 1 is a scanner-based anatomical coordinates, 2 is aligned to another file's or to anatomical truth, 3 is aligned to Talairach-Tournoux atlas, 4 is MNI 152 normalised coordinates and 5 is aligned to normalised coordinates.

The conditions for utilising the different methods are:

- **Method 1**: when the qform code equals 0. This old method remains useful just for compatibility with the older format.
- **Method 2**: when the qform code is greater than 0. It combines an affine transformation and the voxel dimensions.
- **Method 3**: when sform code is greater than 0. It uses an affine transformation.

In this case, the sform code is equal to 2, indicating that the coordinate mapping should be performed with Method 3. This also indicates that the coordinate system is aligned with an anatomical truth.

Method 3

In Method 3 the (x, y, z) coordinates are given by a general affine transformation 4.1 of the (i, j, k) indices:

$$\begin{aligned} \mathrm{x} &= \mathrm{srow}_x[0] \cdot i + \mathrm{srow}_x[1] \cdot j + \mathrm{srow}_x[2] \cdot k + \mathrm{srow}_x[3] \\ y &= \mathrm{srow}_y[0] \cdot i + \mathrm{srow}_y[1] \cdot j + \mathrm{srow}_y[2] \cdot k + \mathrm{srow}_y[3] \\ z &= \mathrm{srow}_z[0] \cdot i + \mathrm{srow}_z[1] \cdot j + \mathrm{srow}_z[2] \cdot k + \mathrm{srow}_z[3] \end{aligned} \tag{4.1}$$

where srow items are the different values in the 4×4 affine matrix, the sub-indexes refer to the row indexes, and the numbers between brackets are the indexes of the different columns. The matrix contains the necessary information to apply a rotation and translation transform.

Consequently, the expression in 4.1 can also be expressed in the matrix form as:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \mathrm{srow}_x[0] & \mathrm{srow}_x[1] & \mathrm{srow}_x[2] \\ \mathrm{srow}_y[0] & \mathrm{srow}_y[1] & \mathrm{srow}_y[2] \\ \mathrm{srow}_z[0] & \mathrm{srow}_z[1] & \mathrm{srow}_z[2] \end{bmatrix} \cdot \begin{bmatrix} i \\ j \\ k \end{bmatrix} + \begin{bmatrix} \mathrm{srow}_x[3] \\ \mathrm{srow}_y[3] \\ \mathrm{srow}_z[3] \end{bmatrix} \tag{4.2}$$

In this method, there is no use of voxel dimensions.

In the mapping processes, the 4.2 operation is firstly performed with the srow values from one view, and then the inverse operation is computed with the srow values of the other view.

In this study, two distinct applications of the third method were executed, referred to as the forward and backward transformations. The forward transformation operation applies a comprehensive mapping of all voxel data from the collection of short-axis slices onto the long-axis image. Conversely, the backward transformation operation involves a more selective mapping, in which only the corresponding voxel from the short-axis is mapped for each specific long-axis coordinate.

The code employed to implement this method is presented in Appendix A.

Forward method 3

In the forward variant of Method 3, the short-axis to anatomical reference coordinate mapping is computed first, followed by the anatomical to long-axis mapping. This results in the overlapping of short-axis masks in the long-axis scan, see Figure 4.5. Due to the slice thickness, the mapping consists of a set of ring-like objects stacked from base to apex. Moreover, the overlapping also presents tissues that are behind or in front of the long-axis plane, in this case parts of the right ventricle and myocardium.

By performing morphological closing and opening, a mask is obtained for the long-axis just from the mapped stack of short-axis annotations; see Figure 4.6. The structural element selected for the morphological operations was an ellipse with dimensions of (20,20) for closing and (5,5) for opening. These design decisions were made based on qualitative analysis conducted on a small set of images from five randomly selected patients. The gap between the mapped stack originates from the slice thickness the scans have.

The mapped annotation and the original annotation (see Figure 4.7) were then combined to obtain a cleaner prediction; see Figure 4.8. The merging consisted on using the mapped and processed annotation from SA as structuring element for closing on the predicted LA from the nnU-Net. Before this merging, the original annotation had a morphological opening with an ellipse of size (20,20). This procedure arose from the idea of discarding spurious annotations across the image and to fill the respective holes in the mask.

Backward method 3

An alternative to the forward method 3 is the backward method 3, which computes the operation in 4.1 from the long-axis to the anatomical and then to the short-axis to allocate the corresponding values from the short-axis prediction

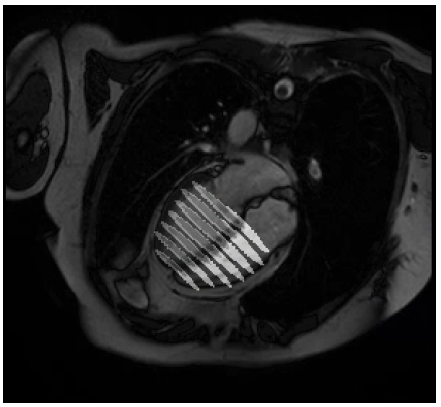


FIGURE 4.5: Method 3 (forward): overlaid short-axis predictions over the long-axis image.

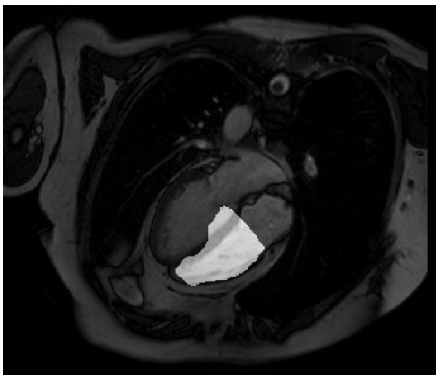


FIGURE 4.6: Method 3 (forward): stack of right ventricle mapped SA annotations after using morphological operations to fill the gaps between the different slices.

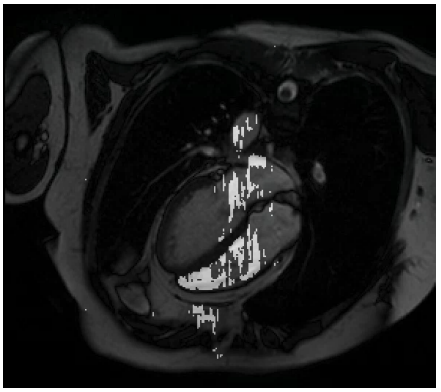


FIGURE 4.7: Method 3 (forward): original predictions for the RV on LA images in the nnU-Net pipeline before the post-processing.

to the long-axis scan. In this operation, nearest neighbour interpolation is performed to determine the value to be mapped. This version of the method 3 can be seen in Figure 4.9.

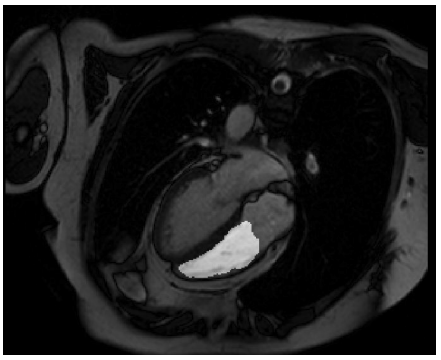


FIGURE 4.8: Method 3 (forward): processed right ventricle mapped SA annotations merged with the predicted LA mask from nnU-Net.

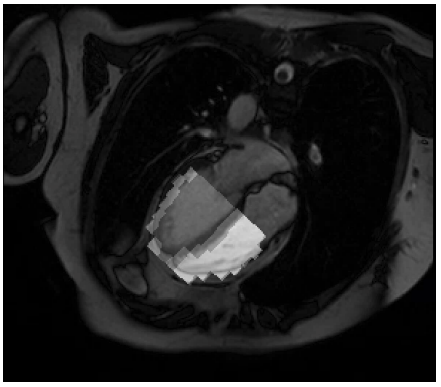


FIGURE 4.9: Method 3 (backward): short-axis annotations mapped into the long-axis view.

In contrast to the forward method 3, the backward method 3 represents the intersection of perpendicular planes, each with its corresponding slice thickness. The distinction lies in the fact that, in the backward method 3, each pixel value is mapped from the short-axis image using the corresponding interpolation. Conversely, in the forward method 3, each pixel value is mapped to the long-axis image. The computation of the exact locations of different pixels often yields decimal values; thus, the nearest integer location is selected. This process results in the observed effect illustrated in Figure 4.5.

4.3.3 Independent networks for each view

For the experimentation of the architectures, four different architectures were prepared and then two versions of each were trained, one for each of the views.

All the networks were trained in the same way. The parameters for training, data preprocessing, and network design decisions are presented below.

4.3.3.1 Hyper-parameters

Training was conducted for 100 epochs with a learning rate of 10^{-3}, using an Adam optimiser and a learning rate scheduler on plateaus with a patience of 10.

The loss function chosen was the summation of the cross-entropy loss and the Dice score loss.

4.3.3.2 Data normalisation and augmentation

The data preprocessing was performed by concatenating the following operations in this order:

- Crop or padding to 256×256.
- Z-Normalisation.
- Histogram normalisation: the landmark were extracted from the training set, but applied to all the data. There was one landmark for long-axis images and another for the short-axis images.
- Intensity normalisation between 0 and 1.

During training, the following augmentations were also added (parameters that are not mentioned below follow the default values of the Torchio library):

- Random motion with 15 degrees of rotation.
- Random ghosting with 3 ghosts.
- Random bias field.

4.3.3.3 Architectures

All the different architectures were trained using the above-mentioned training and the same data. In the following sections, the hyperparameters chosen for these architectures are presented.

For architectures involving the utilisation of pre-trained networks, the input was methodically modified. This modification was achieved by expanding the input to encompass the three channels – red (R), green (G), and blue (B). Each of these channels received identical input. In addition to this, a normalisation process was performed on the input to enhance its compatibility with the pre-existing network.

- **U-Net:**

  U-Net [61] is one of the most widely used architectures in segmentation for medical imaging. It consists of concatenations of pairs of convolution layers followed by a pooling layer when contracting and up-sampling steps in the expansion. Each contraction step has a skip connection to their respective expansive step. The final layer is composed of a 1×1 convolution layer; see Figure 4.10.

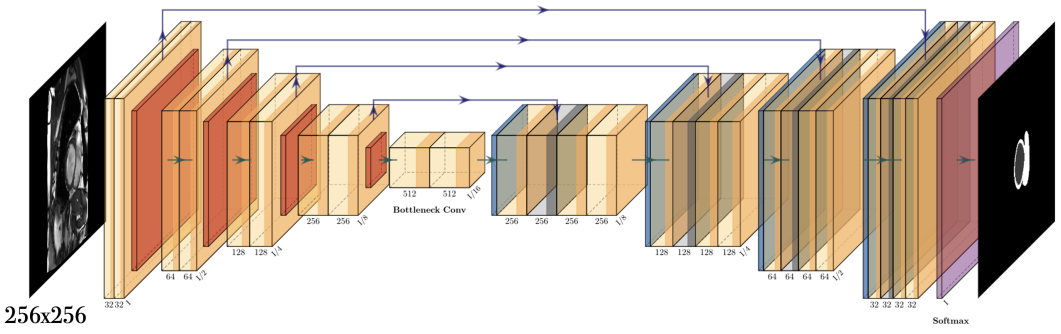


FIGURE 4.10: U-Net [61] Architecture.

  The U-Net followed a vanilla implementation with an initial feature map of 32 filters.

- **DeepLabv3:**

  DeepLabv3 [97] is a semantic segmentation architecture that proposes an atrous convolution in cascade or in parallel to capture multiscale context by adopting multiple atrous rates, these blocks are ResNet blocks.

  In particular, this version has an Atrous Spatial Pyramid Pooling module (ASSP) that differs from previous versions by applying global average pooling on the last feature map of the model, inputs the resulting features in a 1×1 convolution with 256 filters, and then there is a bilinear upsampling to the desired spatial dimension; see Figure 4.11.

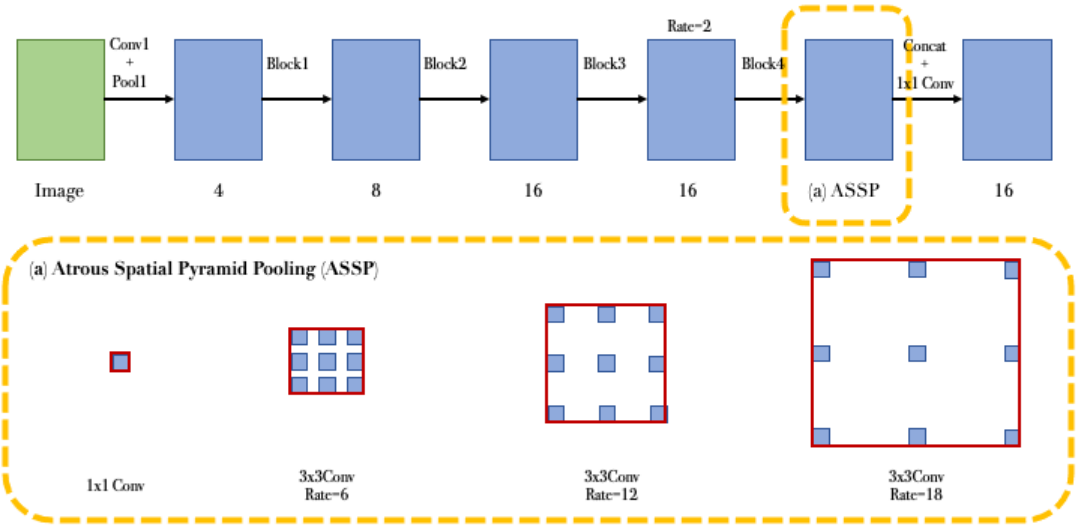


FIGURE 4.11: DeepLabv3 Architecture. Figure reproduced from [97].

The DeepLabv3 that was trained had a ResNet101 backbone that had been pre-trained from Torchvision [110]. The head had 2048 neurons.

In order to use this network, the single channel that the input images have was expanded to the 3-channel input that ResNet101 expects by duplicating the channel. Additionally, during normalisation, the rescaling of the values addressed the difference between the bits used for the encoding of the colour or greyscale intensities.

- **TransUNet:**

TransUNet [98] is a hybrid U-shaped CNN-Transformer architecture developed to perform medical image segmentation. The transformer encodes tokenised image patches from CNN feature map as the input sequence to extract global contexts. On the other hand, the decoder up-samples the encoded features, which are then combined with the high-resolution CNN feature maps to enable precise localisation, see Figure 4.12.

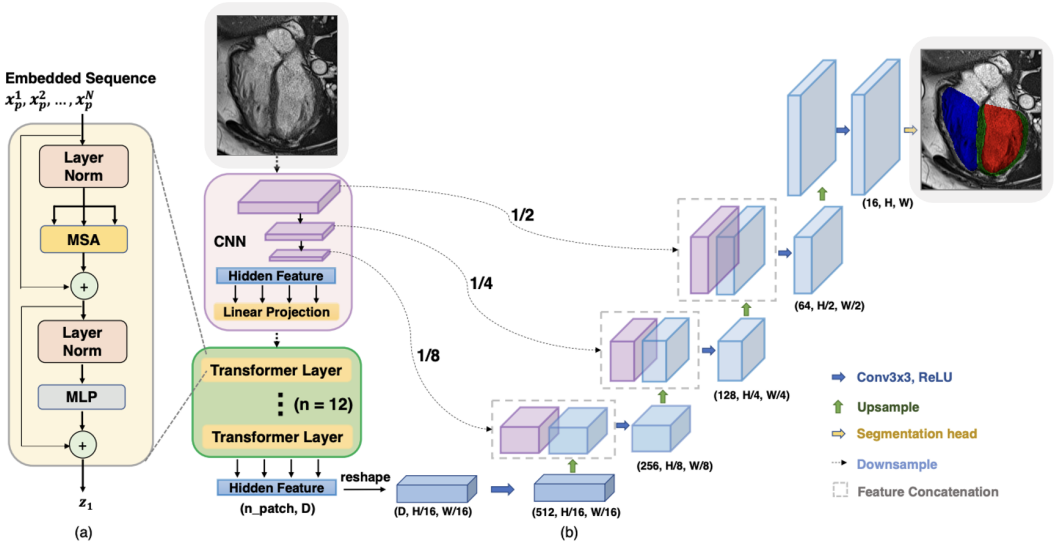


FIGURE 4.12: TransUNet Architecture. Figure adapted from [98] using M&Ms2 scans [21], changes are indicated with a light grey background.

The TransUNet was set to have 8 visual attention blocks and 512 neurons in the head of the multilayer perceptron.

- **CE-Net:**

  CE-Net [99] (Context Encoder Network), was proposed to capture more high-level information and preserve spatial information for 2D medical image segmentation, see Figure 4.13.

  The three main components are: (1) a feature encoder module, (2) a context extractor, and (3) a feature decoder module. As a feature extractor, they proposed to use a pre-trained ResNet block. The context extractor module is formed by a newly proposed dense atrous convolution (DAC) block, depicted in Figure 4.14; and a residual multi-kernel pooling (RMP) block, represented in Figure 4.15.

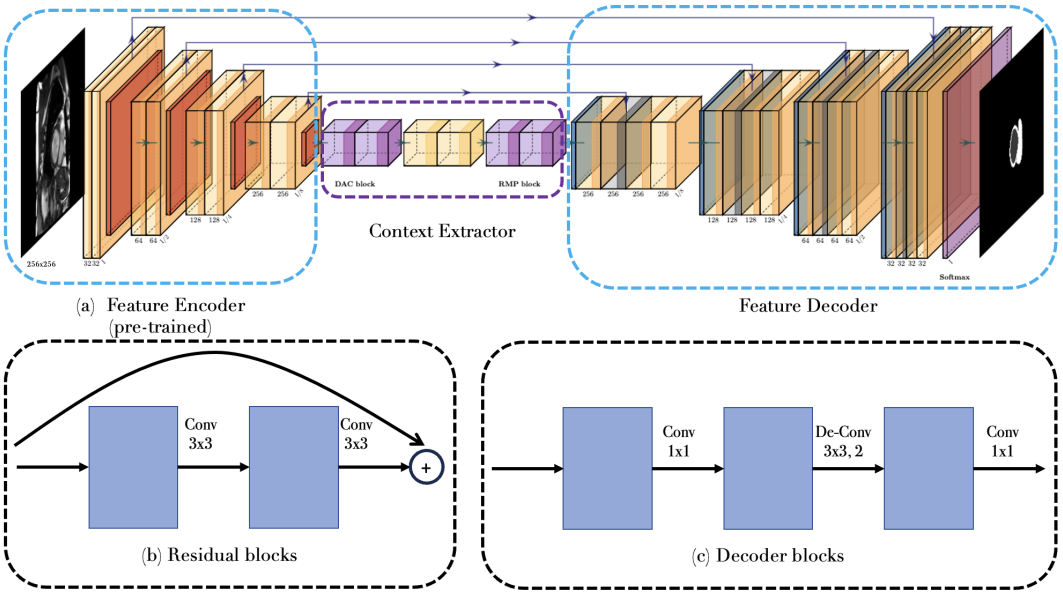


FIGURE 4.13: CE-Net Architecture. Figure reproduced from [99].

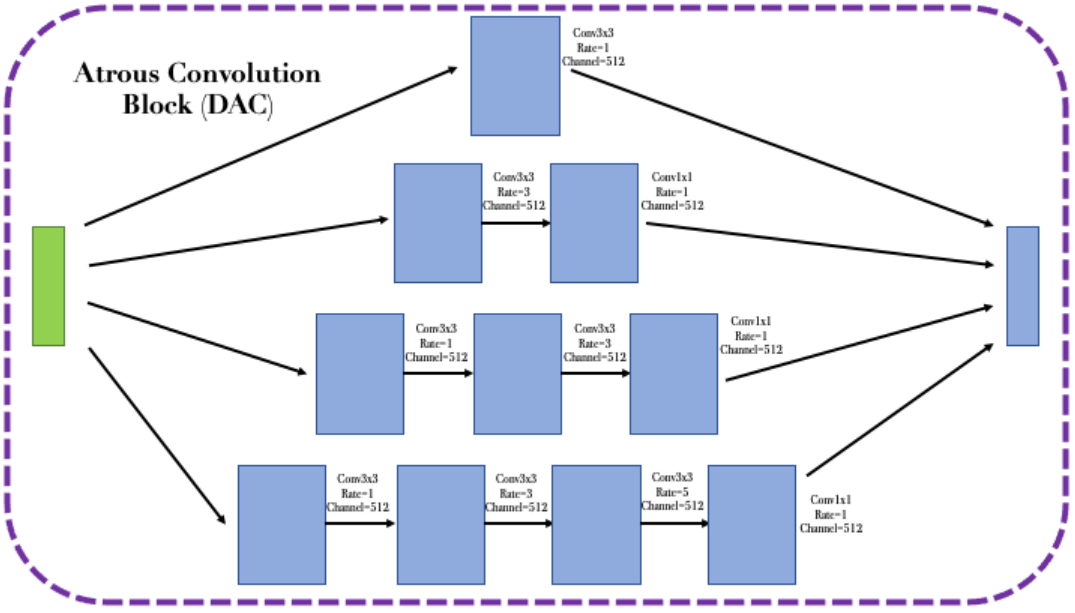


FIGURE 4.14: The dense atrous convolution (DAC) block. Figure reproduced from [99].

The DAC has four cascade branches with the gradual increase of the number of atrous convolutions, from 1 to 1, 3, and 5, then the receptive field of each branch will be 3, 7, 9, 19. It employs different receptive fields, similar to Inception structures.

The RMP encodes global context information with four different-sized

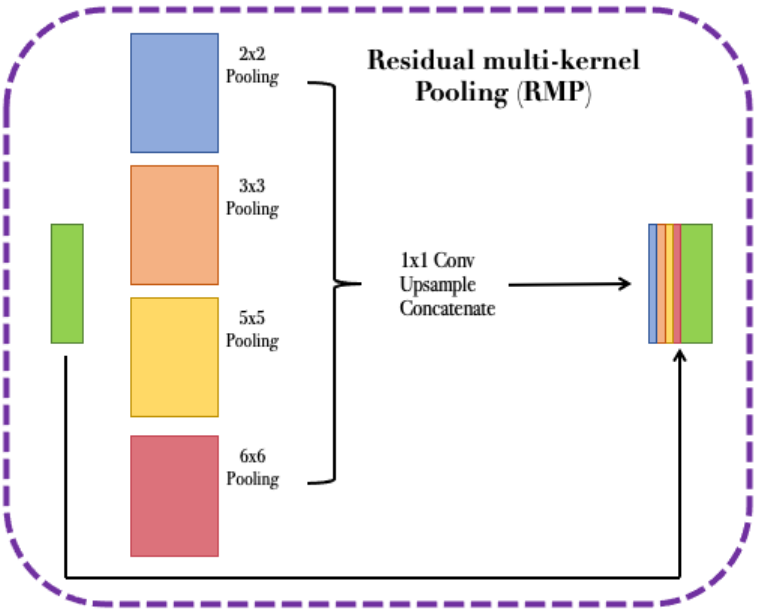


FIGURE 4.15: The residual multi-kernel pooling (RMP) block. Figure reproduced from [99].

receptive fields. The four-level outputs contain feature maps of various sizes. To reduce the dimension of weights and computational cost, it was proposed to use a 1×1 convolution after each level of pooling. The CE-Net was set to have its vanilla values: 64 filters in the initial convolution block. The backbone in this case was a ResNet34 that was pre-trained.

In the same manner as the adaptation made for the DeepLabv3 network, the channels of the input data were also expanded to accommodate the ResNet of CE-Net.

4.4 Experiments and results

The results for the baselines and the two different approaches are presented in the following sections.

4.4.1 nnU-Net and post-processing

This section provides the results for the nnU-Net baseline and for the proposed post-processing. The results are displayed in Table 4.1, note that those results have been obtained in the validation set and only for the right ventricle cavity. The averages are calculated based on the number of images rather than for individual patients.

	DICE		**Hausdorff**		**Average**	
	SA	**LA**	**SA**	**LA**	**Dice**	**Hausdorff**
Baseline	0.9118	0.3498	10.3670	145.0771	0.7713	44.0445
Proposed (Forward)	0.9118	0.6478	10.3669	28.8535	0.8458	14.9886
Proposed (Backward)	0.9118	**0.7310**	10.3669	**12.7729**	**0.8709**	**10.7296**

TABLE 4.1: Results for nnU-Net with and without the proposed post-processing in the validation set for the right ventricle. Higher is better for DICE, and lower is better for Hausdorff.

4.4.2 Independent networks

In this section, the results for the four different architectures for both views are presented. Table 4.2 shows the best architectures for each vendor and view, and Table 4.3 presents the results for each disease and view. Additionally, in the Appendix B, the results for each vendor pertaining to a specific view and models have been provided.

	SA	**Architecture**	**LA**	**Architecture**
Siemens	0.9225	CE-Net	0.8947	DeepLabv3
Philips	0.9126	CE-Net	0.9082	CE-Net
GE	0.8832	DeepLabv3	0.8434	DeepLabv3

TABLE 4.2: Multi-view segmentation: best performing architectures for each view and scanner. Results in DICE (higher is better).

Additionally, inference times per image are presented in Table 4.4. The graphical processor unit (GPU) employed for processing the data was a Nvidia GeForce RTX 2080 Ti.

	SA	Architecture	LA	Architecture
NOR	0.9500	CE-Net	0.8835	DeepLabv3
LV	0.9305	U-Net	0.9011	DeepLabv3
HCM	0.9251	CE-Net	0.9078	DeepLabv3
ARR	0.9171	CE-Net	0.8928	DeepLabv3
FALL	0.9322	DeepLabv3	0.8826	DeepLabv3
CIA	0.8999	DeepLabv3	0.8802	DeepLabv3
DRV	0.9188	CE-Net	0.8671	DeepLabv3
TRI	0.9143	U-Net	0.8921	DeepLabv3

TABLE 4.3: Multi-view segmentation: best performing architecture per view and disease. Healthy (NOR), Dilated Left Ventricle (LV), Hypertrophic Cardiomyopathy (HCM), Congenital Arrhythmogenesis (ARR), Tetralogy of Fallot (FALL), Interatrial Communication (CIA), Dilated Right Ventricle (DRV), Tricuspidal Regurgitation (TRI). Results in DICE (higher is better).

	Time (s/image)	
	SA	**LA**
U-Net	**0.027**	**0.023**
DeepLabv3	0.042	0.038
TransUNet	0.034	0.03
CE-Net	0.032	0.028

TABLE 4.4: Multi-view segmentation: inference times for the different architectures and views in seconds per image. U-Net performed faster than the other architectures.

4.5 Discussion and Conclusions

In this section, the results obtained are discussed, along with the limitations of the work. Some future work is also proposed.

Firstly, the nnU-Net framework was set up, and while the results for the short-axis were promising, tuning the parameters to the intended values took more time than expected. Furthermore, an issue with predictions arose during the training of the long-axis images. After several debugging sessions, this tool was abandoned due to its poor accuracy and inadequate usability. Moreover, nnU-Net demanded quite a lot of computational power, plus much more time to train.

To make the original idea work, a post-processing method was developed, inspired by other works that benefited from the information of the NIfTI header to map both views. This consisted of mapping predictions from short-axis images into long-axis images. This transformed the result from unusable to an effective prior during clinical reporting sessions. The applicability of this technique depends on having the requisite metadata to interpret the scanning reference to anatomical positions. In scenarios where the data format differs (e.g., DICOM), a similar procedure could be applied assuming the presence of the aforementioned metadata.

Based on the results of the architecture search conducted to replace nnU-Net, it can be concluded that DeepLabv3 would be the best architecture to choose from the range tested, but CE-Net would also be a reasonable option when computing power is limited. Pre-trained models were generally more accurate than architectures trained from scratch. Further efforts and investigation could include pre-trained encoders in the existing models, which is likely to improve their performance.

Moreover, the explanation on why the performance of all models was worse for the GE scans is due to the fact that there are fewer cases in the training set belonging to this particular scanner, and hence further work is needed to ensure the generalisation of the models where datasets are unbalanced.

Surprisingly, the performance on the pathologies that were not present in the training set, RV and TRI, was still acceptable, even better than some pathologies seen during training. Further experimentation should explore which diseases are more difficult to achieve good predictions on when they are not present in training. The intuition is that the anatomical features of dilated right ventricles and tricuspid regurgitation are similar to other diseases and easy to learn by CNNs. In contrast, pathologies such as CIA or FALL have morphologies that are more prone to miss-classification.

Additionally, the better results on SA scans are attributed to having more images in this view than in the LA view

Although the top scoring method in the M%Ms2 Challenge [111] outperformed the various models presented in this chapter, the DICE score remained superior to that of several participants in the final phase. The performance in the presented models was approximately 0.03 lower than that of the best-performing algorithm, with comparable inference time.

Work on exploring augmentations that change the morphology of healthy scans into different diseases, leveraging the scarcity of data for specific pathologies, is presented in Chapter 7.

Chapter 5

Time Frame Detection using Sequential Neural Networks

5.1 Introduction

This chapter presents a detailed analysis of the contributions of this research to cardiac MRI time frame detection. In particular, the adoption of new deep learning modules to accelerate and improve the average frame difference (aFD) of previous work was explored. Results trained and tested on public data (M&Ms) are presented to ease reproducibility and bring more attention to this topic, which has received far less attention than segmentation despite its importance in the ejection fraction calculation.

The work aims to advance the automation of ED and ES phase detection in cardiac MRI through deep learning, with a specific focus on improving precision and reducing reliance on accompanying ECG signals. The significance lies in enhancing the accuracy of clinical indicators and making the analysis process more efficient and less prone to errors.

The experiments and results described in this chapter have been accepted for publication at the Irish Conference on Artificial Intelligence and Cognitive Science (AICS) 2023.

Accurate identification of the End-Diastolic and End-Systolic phases in cardiac functional analysis is paramount, as errors during these phases can significantly impact critical clinical indicators such as ejection fraction (EF) and global longitudinal strain (GLS). Studies have indicated that discrepancies within the two and three frame difference can introduce up to a 10% error in these indicators, underscoring the importance of precision in phase detection [112, 113, 22]. These studies were developed using echocardiographic data, but [114] compared echocardiographic measurements to CMRI and concluded that LV volume is underestimated in the former. Additionally, the variation is consistent over time, suggesting that the error in selecting the incorrect frame remains similar.

The research reported in this chapter focuses on leveraging deep learning techniques to automate the detection of ED and ES phases in cardiac short-axis Magnetic Resonance Images. Notably, the work aims to eliminate the reliance on accompanying Electrocardiogram signals, thereby enhancing algorithmic robustness and reducing the likelihood of errors associated with manual phase labeling. Such advancements not only streamline the analysis process but also contribute to more accurate clinical assessments.

In this work, cardiac short-axis Magnetic Resonance Images are targeted for the automated detection of ED and ES phases through deep learning in a regression problem. In particular, by leveraging pre-trained models with sequential neural network modules. The contributions are the following:

- **Report performance on public data:** previous works on this topic reported results on private data [23], hence reproducing their results was not possible. Evidence of strong performance in the available open M&Ms dataset [56] is provided.

- **Comparison of two different architectures:** The performance of different novel models was studied, where two elements that have shown great performance in a variety of problems involving sequences were experimented with: (1) Long Short-Term Memory (LSTM) [115] and (2) Transformers [116].

- **Inference time:** the inference time of the proposed models is 500 faster than in previous work on the same problem [23]. While the results were obtained faster, the computational resources of the previous method are unknown.

5.2 Methodology

In this section, the methodology and experimentation framework employed in the study are delineated, focusing on MRI sequences from publicly available datasets. The investigation revolves around the implementation and evaluation of two deep learning modules applied after a pre-trained encoder, a combination introduced in this work. The ensuing details elucidate the intricacies of this approach and the rigorous experimentation process, providing a comprehensive understanding of the methodological choices and research design.

5.2.1 Data

In the experimentation, the data released during the Multi-Centre, Multi-Vendor, and Multi-Disease Cardiac Image Segmentation Challenge (M&Ms) was employed. In particular, the dataset consists of 150 training cases, 34 for validation, and 136 for testing. For the pre-trained part of the network, the M&Ms2 training data was employed 4.3.1.

Notably, data was acquired on scanners from four different vendors, two of them are present in the training set, and all of them are present in the testing one.

For the purposes of the experimentation, a single middle slice for each time frame was focused on, streamlining the analysis while maintaining key temporal information. This approach ensures a consistent and representative selection for the deep learning models.

5.2.2 Architecture

The model architecture comprises a Convolutional Neural Network (CNN) module,corresponding to the encoder part and the bottleneck of a U-net [61]. Subsequently, the CNN features are fed into a fully connected layer where they are flattened. Following this, the resulting features undergo sequence-wise feature extraction using either an LSTM or a Transformer Encoder in the first and second experiments, respectively. The output of the sequence module is then connected to a second fully connected layer that produces a vector corresponding to one element per frame in the sequence. The complete architecture is depicted in Figure 5.1. The parameters chosen for the networks were the following:

- The CNN encoder was pre-trained in a segmentation task using the M&Ms2 data and then frozen while training the rest of the network. In particular, the parameters of the network were: 32 filters in the first out of five pairs of convolutional layers, and a max-pooling layer after each of the four first pairs of convolutional layers.
- The fully connected layers had 512 and one neurons, respectively.
- The LSTM had the default PyTorch parameters, except that it was set to bidirectional. Both the input size and the hidden size were 512, and it had two layers.
- The Transformer Encoder was set to have 512 as the input size, with four heads and two layers.

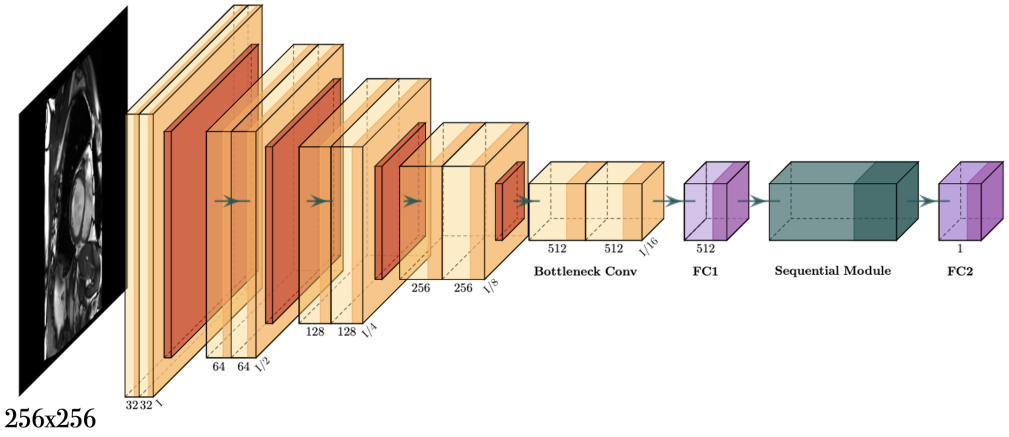


FIGURE 5.1: The proposed network with the sequential module referring to an LSTM or a Transformer encoder in each experiment.

5.2.3 Experiment

In the experiment, the performance of the two proposed networks was tested, which were trained using a loss function consisting of two components: (1) Mean Squared Error between the prediction and the synthetically generated signal (see Equation 5.1), and (2) a temporal structured loss (see Equation 5.2), both proposed by TempReg-Net [23]. The loss adds both components together.

$$y_k = \begin{cases} \left|\frac{k-N_{es}}{N_{es}-N_{ed}}\right|^{\delta}, & \text{if} \quad N_{ed} < k \leq N_{es} \\ \left|\frac{k-N_{es}}{N_{es}-N_{ed}}\right|^{\upsilon}, & \text{otherwise} \end{cases} \tag{5.1}$$

Where N is the ground truth for each phase, and k is the time frame number. δ and υ are hyperparamaters set to 3 and 1/3 respectively to mimic the behaviour of the left ventricle in the cardiac cycle.

$$\begin{aligned} L_{temp} &= \tfrac{1}{2}(L_{inc} + L_{dec}) \\ L_{inc} &= \tfrac{1}{T}\sum_{k=2}^{T} \mathbb{1}(y_k > y_{k-1}) \max(0, \eta_{k-1} - \eta_k) \\ L_{dec} &= \tfrac{1}{T}\sum_{k=2}^{T} \mathbb{1}(y_k < y_{k-1}) \max(0, \eta_k - \eta_{k-1}) \end{aligned} \quad \eta \text{ is the prediction.} \tag{5.2}$$

To label the time frames, the maximum and the minimum of the signal are set as the ED and ES time frames. An example of the resulting signal is depicted in Figure 5.2. In this signal, the time frames corresponding to ED and ES are the first and eighth frames of the scan.

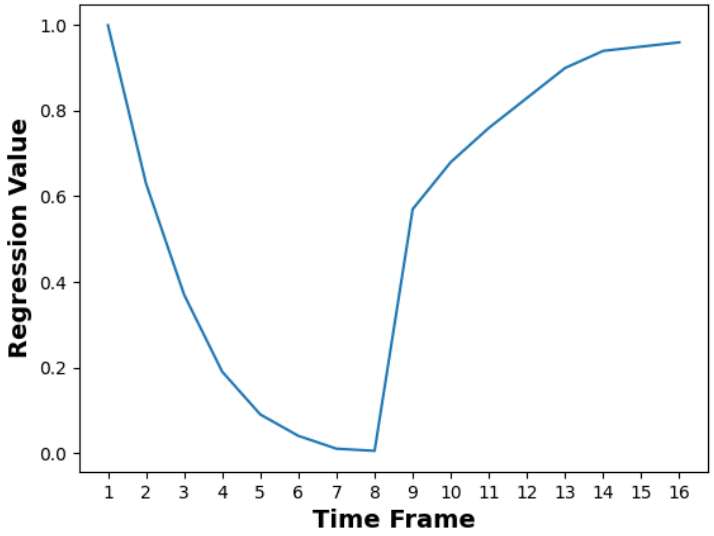


FIGURE 5.2: An example of the output regression curve. The maximum (first frame) and the minimum (eighth frame) correspond to ED and ES time frames, respectively.

To evaluate the performance, the average Frame Difference (aFD) (see Equation 5.3) was used to quantify the error.

$$aFD = \frac{1}{N}\sum_{t=1}^{N}|\hat{y}_t - y_t| \tag{5.3}$$

5.3 Results

Table 5.1 presents the performance metrics on the test set of the M&Ms dataset for both proposed architectures: Average Frame Difference (aFD), where lower values indicate better performance, and the detection time for each frame. The computations were conducted using an Nvidia GeForce RTX 2080Ti.

Model	aFD ED (ms)	aFD ES (ms)	Detection Time (s)
CNN + LSTM	1.70 (77)	1.75 (79)	0.0028
CNN + Transformer	2.03 (92)	1.84 (83)	0.00246

TABLE 5.1: Time frame detection results: average frame different and detection times. LSTM performs marginally better than the transformer while detection time remains similar.

The selection of the middle slice for the analysis constitutes a chosen hyperparameter and can affect the reproducibility. Results may vary depending on the exact frame, although the results should remain similar. Adding additional slices to the computation would reduce the variability at a computational cost.

5.4 Conclusions

The models consistently delivered accurate results, even when applied to data from previously unseen scanners, particularly excelling in scenarios with a two-frame difference. Notably, the LSTM model demonstrated superior performance compared to the model incorporating a transformer encoder. This difference in performance might stem from the need for a more refined training strategy for the transformer model. Additionally, both models exhibited remarkable speed during testing, processing data at 0.0025 seconds, a significant improvement over other methods that take more than one second.

these findings underscore the effectiveness of the proposed method, showcasing performance on par with human annotators while operating at a significantly faster speed. This suggests the potential for seamless integration into current clinical tools. For future research directions, exploring additional training strategies, such as incorporating segmentation tasks or optimising the training process, and investigating the synergy between the proposed recurrent modules present promising avenues for further enhancement.

In future work, it is speculated that the addition of optical flow [117] between frames could lead to improvements. As a direct input in a parallel encoder or indirectly by obtaining a feature from it. Moreover, incorporating more slices might provide additional benefits.

Chapter 6

Leveraging Pre-trained Models and MRI Specific Augmentations to Mitigate Respiratory Artifacts

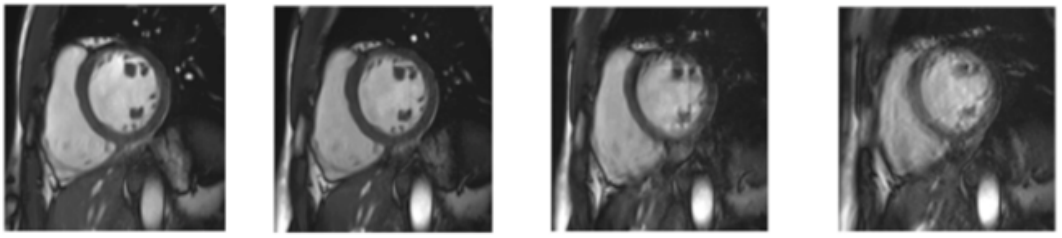
6.1 Introduction

This chapter provides a detailed description of the contributions of this research to cardiac MRI segmentation in the presence of respiratory motion artifacts, which took part during Extreme Cardiac MRI Analysis under Respiratory Motion Challenge (CMRxMotion). The experiments and results described in this chapter have been accepted for publication at the Medical Image Computing and Computer Assisted Interventions (MICCAI) 2022, in the Statistical Atlases and Computational Modeling of the Heart (STACOM) proceedings [2].

As discussed in Chapter 1, extracting morphological and functional information from CMRI data is tedious and intensive and can lead to observer bias [118]. As was presented in Chapter 2, the automation of these tasks has attracted

the attention of scientists due to its high impact on daily clinical workflows. In the last few years, the innovation of deep learning techniques, and in particular convolutional neural networks, have brought more interest in the topic and have demonstrated great potential [119, 65, 1]. While recent efforts address approaches to resolve the quality of the segmentation of all the chambers within the heart (right ventricle, left ventricle, and myocardium) and multiple views of the tissue (long-axis and short-axis) [56], sub-optimal segmentations, e.g., those affected by respiratory artifacts, are still under-explored. Examples from the mentioned artifacts are displayed in Figure 6.1.

To address this, in this chapter, a method is proposed that benefits from publicly available pre-trained models. Using pre-trained weights not only accelerates the training process but also enhances the predictions of the network. In particular, the results are substantially better for the right ventricle and the myocardium, leading to increased DICE scores and reduced Hausdorff distance. Results on the left ventricle, however, remained unchanged. This work was conducted in the context of the CMRxMotion challenge, with a focus on Task 2 (the segmentation task).



(A) Full Breath hold (B) Half Breath hold (C) Free Breath (D) Intensive Breath

FIGURE 6.1: The four different breathing intensities resulting in motion artifacts present in the CMRxMotion Challenge data (Figure reproduced from the official challenge images [25]).

6.2 Method

The proposed method consists of: (1) resampling, preprocessing, and normalising the data, (2) loading a pre-trained model and setting the appropriate training parameters, and (3) training the loaded network with data that have been augmented with a number of different deformation and intensity changes.

The 3D data is referred to as volumes and the short-axis slices as images. The data resampling, preprocessing, and normalisation were performed in 3D, while the data augmentation was carried out over the slices.

The presented method focuses on fine tuning an encoder during the segmentation task (with a segmentation head and decoder).

6.2.1 Data resampling, preprocessing and normalisation

First, the volumes were reoriented to the canonical orientation, which were then resampled. A crop was then applied to the region of interest. The data were subsequently split into training and validation subsets, where validation represented 20% of the available subjects.

Secondly, the intensity was normalised using a histogram obtained from the training samples. The histogram standardisation [120] of all sets is performed using the mentioned histogram.

6.2.2 Architecture study

In this work, two different versions of the widely used U-Net architecture [61] were tested to investigate if pre-trained weights from a different problem and data domain could lead to improved results. The two approaches are:

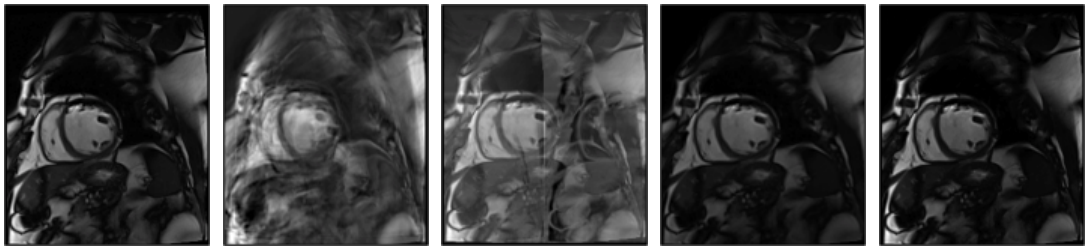
- a U-Net trained from scratch;

- a ResNet-based [46] U-Net architecture with weights pre-trained for ImageNet [121] classification.

6.2.3 Data augmentation

An important part of this study consisted of applying four different types of augmentation techniques, which have previously been shown to enhance the quality of the predictions on scans without the above-mentioned artifacts [122]. Additionally, these augmentations consist of applying intensity artifacts, which are hypothesised to improve the models' generalisation and strength against the target artifacts. These augmentations were the random motion, random ghosting and random gamma, presented in Section 4.2.1.3, plus the Random Gamma. The Random Gamma is an intensity transform that consists of a random change in the contrast of an image by raising its values. The ranges of this transformation vary according to the gamma value.

Figure 6.2 depicts an example of the application of each augmentation over the same scan.



(A) Raw Image (B) Random Motion (C) Random Ghosting (D) Random Bias Field (E) Random Gamma

FIGURE 6.2: Application of augmentations over the same scan. From right to left: 6.2a the raw image or scan, 6.2b random motion, 6.2c random ghosting, 6.2d random bias field, and 6.2e random gamma.

6.2.4 Cardiac MRI Dataset

The dataset for this study was provided by the Extreme Cardiac MRI Analysis Challenge under Respiratory Motion [25]. In particular, the data from the segmentation challenge (task 2) was utilised.

Training data represented a cohort of 20 subjects, which were scanned four different times, each under a different grade of respiratory intensity as in Figure 6.1, described as follows: (1) full breath hold, (2) half breath hold, (3) free breathing, and (4) intensive breathing. The evaluation data represented a cohort of five subjects with the same four different intensities of breathing. Lastly, the test data represented a cohort of 15 subjects, unavailable for participants, just used in the offline testing phase. An expert radiologist recorded and labelled the end-diastolic and end-systolic phases for all subjects and breathing intensities.

6.3 Experiment settings

The experiment settings set during the method's training and inference are described in the following section.

First, the architectures in Section 6.2.2 had the following details:

- U-Net trained from scratch: 32 filters in the first out of five pairs of convolutional layers and a max-pooling layer after each of the four first pairs of convolutional layers. The convolutional layers used the ReLU activation and batch normalisation.
- U-Net pre-trained: used a ResNet101 [46] backbone as an encoder, pre-trained with ImageNet [121].

The general learning rate was set at 10^{-3} except for fine tuning, where the learning rate for the trained encoder was set at 10^{-4}. The learning rate was

scheduled to change on plateau with a patience of 100 epochs, reducing the learning rate to half its previous value. The optimiser chosen was Adam.

The steps detailed in Sections 6.2.1 and 6.2.3 were done using the TorchIO library [102]. The pre-trained network was downloaded from PyTorch Segmentation Models [110].

The augmentation policy consisted of always applying one of the techniques described in Section 6.2.3. In particular, it was found that tripling the random motion augmentation compared to the other augmentations resulted in a model that was more resilient to respiratory motion artifacts, leading to better segmentation performance.

6.4 Results

The results are presented in two different sections: (1) validation results, and (2) evaluation results, which correspond to the results provided by the challenge platform.

6.4.1 Validation results

Table 6.1 shows the results on the validation split of the training set. Four different models are listed, indicating whether the training included augmentations, and if the weights were trained from scratch or pre-trained on ImageNet. The hypothesis that was initially had is validated, as all regions performed better when pre-trained weights and the augmentations were used.

	DICE			
	LV	**MYO**	**RV**	**ALL**
U-Net (scratch)	0.97	0.945	0.963	0.959
U-Net (scratch) Augs	0.974	0.949	0.965	0.963
U-Net (ImageNet)	0.97	0.948	0.966	0.962
U-Net (ImageNet) Augs	**0.976**	**0.952**	**0.97**	**0.966**

TABLE 6.1: Validation results for the segmentation of the Cardiac MRI scans with respiratory motion (DICE). Augs indicate the additional data augmentation. The best results are in bold.

6.4.2 Evaluation results

Table 6.2 shows the results of the inference of the evaluation data on the challenge platform. The four models are the same used in the previous section.

	DICE			Hausdorff (mm)		
	LV	**MYO**	**RV**	**LV**	**MYO**	**RV**
U-Net (scratch)	0.88	0.768	0.789	11.78	7.64	11.37
U-Net (scratch) Augs	0.88	0.771	0.782	**10.57**	7.74	11.87
U-Net (ImageNet)	0.879	0.796	0.826	11.4	6.2	8.71
U-Net (ImageNet) Augs	**0.883**	**0.797**	**0.851**	11.04	**5.64**	**7.77**

TABLE 6.2: Evaluation results for the segmentation of the Cardiac MRI scans with respiratory motion (DICE and Hausdorff 95). Augs indicate the additional data augmentation. The best results are in bold.

The best results were achieved with the model trained using pre-trained weights and additional data augmentation. However, in this model, the Hausdorff distance is worse than in the equivalent model trained from scratch. This discrepancy in the LV might be due to a general tendency to misclassify some pixels, placing them slightly away from their correct location.

6.4.3 Test results

Table 6.3 shows the results of the inference of the test data, where the data were processed on the challenge platform through the submission of the algorithm. The submitted model corresponded to the pre-trained version of the U-net including the augmentation strategy.

	DICE			Hausdorff (mm)		
	LV	**MYO**	**RV**	**LV**	**MYO**	**RV**
U-Net (ImageNet) Augs	0.897	0.837	0.842	6.17	5.03	7.69

TABLE 6.3: Test results (DICE and Hausdorff 95).

The test set results demonstrate that the presented method not only preserves but also improves performance. This highlights the model's generalisability and resilience to varying levels of artifacts. Notably, the Myo region showed the most significant improvement in DICE, while the LV exhibited a substantial reduction in Hausdorff distance.

6.5 Conclusions

In this chapter, it was proposed that the training be started from weights obtained in a classification problem in another data domain. In addition, an augmentation policy was proposed, consisting of four different augmentations with random motion applied three times more than the rest.

While the contribution of the random motion was positive in the experimentation, which consisted of applying this transformation to all the scans (including the data already containing artifacts), it decreased the performance if utilised for the classification task of the same challenge. It is hypothesised that this policy will directly affect the relation of the scan to its class label. The success of this method in this task (segmentation) is attributed to the contribution of having a wider range of data, triggering a greater capacity of the model to generalise.

From the quantitative analysis conducted on the validation split of the training data and the evaluation data from the official platform, it can be said that both additions corresponded to an increase in the quality of segmentation. Furthermore, the adoption of the pre-trained weights also accelerated training times.

In conclusion, the model trained using pre-trained weights and additional data augmentation demonstrated superior performance, as evidenced by its improved results on the test set. This highlights the model's strong generalisability and resilience to varying levels of artefacts. While there was a slight increase in the tendency to misclassify some pixels in the evaluation set, the overall performance remained robust. Notably, the Myo region showed the most significant improvement in DICE, and the LV region achieved a substantial decrease in Hausdorff distance, further validating the effectiveness of the proposed method in the test set.

Chapter 7

Synthetic Balancing of Cardiac MRI Datasets using Style Transfer

7.1 Introduction

This chapter presents the research performed on synthetically balancing cardiac MRI datasets using style transfer and 3D deformations. Synthetic balanced datasets are crucial in cardiac MRI data analysis because they help address the class imbalance problem. In cardiac MRI, the number of abnormal or diseased cases is often significantly lower than the number of normal cases. This imbalance can lead to biased machine learning models that perform poorly in detecting cardiac conditions.

Synthetic balancing of datasets involve creating artificial samples of the minority class (e.g., diseased hearts) to match the number of samples in the majority class (e.g., healthy hearts). This enables machine learning algorithms to learn from a more balanced dataset, improving their ability to detect and diagnose cardiac conditions accurately. By using synthetically balanced datasets, researchers

and clinicians can enhance the performance and reliability of cardiac MRI data analysis, ultimately leading to more accurate diagnoses and better patient care

The novelty in this proposed method lies in using style transfer techniques to achieve synthetic data balancing through deformations applied to an atlas. Unlike previous attempts [28, 83, 29, 84], the synthetic data was obtained from an atlas.

Once the atlas is processed, different slice cuts are obtained from it and style transfer is applied to make them appear as real short-axis MRI scans. Those synthetic scans are then added to the training set for the segmentation network. It was found that using synthetic scans to balance the dataset resulted in up to a 0.05 increase in the DICE score.

The following sections detail how synthetic scans are leveraged to augment the dataset and optimise the performance of the analysis for the two separate patient cohorts with distinct cardiac pathologies. This innovative strategy enable models to address the unique challenges posed by these cohorts and obtain more robust and precise results in the CMRI data analysis. Section 7.1.1 provides an overview of relevant literature, focusing on works related to this method and generative techniques that allow for comparisons with certain steps in this pipeline. In Section 7.2, the various components of this novel method are broken down, with detailed descriptions of each element being offered. Section 7.3 outlines the experimental setup, evaluation procedures, and showcases the results. Lastly, in Section 7.4, the conclusions are presented, and suggestions for future enhancements and research directions are provided.

The experiments and results described in this chapter have been accepted for an oral talk at the Scientific Assembly and Annual Meeting of the Radiological Society of North America (RSNA) 2023 and as a full paper at the Medical Image Understanding and Analysis (MIUA) 2024.

Segmenting cardiac MRI images poses a specific challenge related to achieving a balanced performance across various cardiac pathologies and normal cases.

Imbalanced datasets can lead to biased machine learning models that perform poorly on underrepresented classes [21]. Various approaches, such as data augmentation and domain adaptation, have been proposed [17] to address this issue.

In this chapter, the problem of dataset balancing in the context of semantic segmentation for cardiac MRI is explored. The effectiveness of synthetic dataset balancing in improving the performance of machine learning algorithms on CMRI data with pathologies that are unseen to the model is investigated. These synthetic datasets are generated by modelling the underlying physics of the imaging process and simulating variations in patient anatomy, physiology, and imaging parameters. While synthetic datasets have the advantage of being easily scalable and customisable, they often suffer from a lack of diversity and realism.

The proposed method combines modifications of an existing atlas [123] that represents a healthy patient with image-to-image style transfer [124] to produce synthetic subsets of data that constitute additional images to train segmentation networks.

7.1.1 Related Work

Three main components were integrated into this research: (1) heart modelling from 2D slices and cardiac atlases, (2) image style transfer and (3) cardiac MRI segmentation.

7.1.1.1 Heart Modelling

The field of heart modelling from slices has made significant progress recently, driven by statistical models and computational efficiency [125]. Recent works have focused on improving shape modelling by refining boundary extraction [126]. Those steps focus on the misalignment correction of the statistical shape model (SSM) extracted from the contours of the 2D images. From the SSM, a

3D mesh is generated and evaluated to fit plausible heart shapes [127], building the final reconstruction of the heart from the scanned structure.

7.1.1.2 Cardiac Atlas

In the context of heart modelling and statistical parametric mapping (SPM), a method and an atlas were proposed by [123]. The authors provided an analysis of the number of subjects and methods necessary to represent a healthy and anatomically normal heart. This representation can be seen in the end-diastolic frame in Figure 7.1.

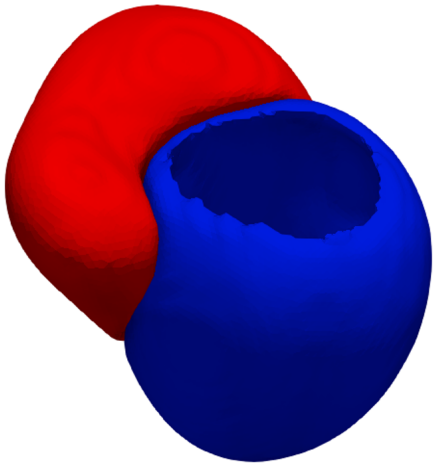


FIGURE 7.1: 3D model of the atlas [123] at ED.

7.1.1.3 Style transfer

Image-to-image translation has attracted a lot of attention during the last 5 years, and in particular, the topic of style transfer has remained an important application within the computer vision community.

In particular, in unpaired image-to-image translation, recent efforts have been made to improve the results and make them realistic [124]. CUT [124] uses a single direction mapping based on patches of images from both domains and

benefits from the contrastive representation learning recent advances. In Figure 7.2, the logic followed by the patch-wise contrastive loss can be seen.

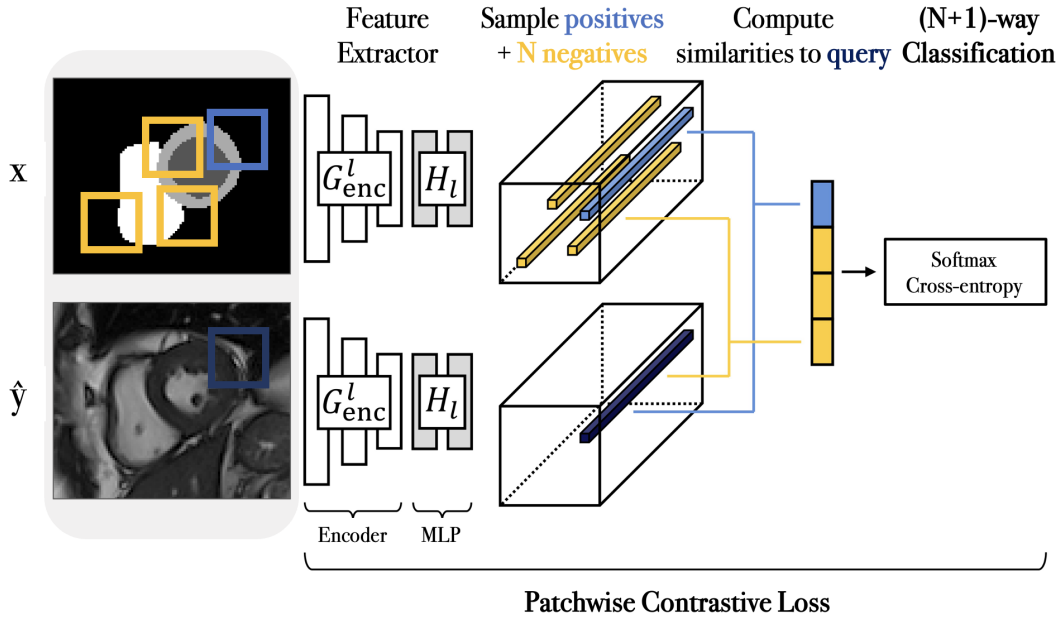


FIGURE 7.2: Patchwise contrastive loss of CUT [124]. Figure adapted from [124], changes are indicated with a light grey background.

While style transfer has found applications in regular images, only recently have a few endeavours harnessed these advances to enhance clinical applications, as evidenced by previous research such as [128] and [30]. Notably, while these works have successfully applied this technique, few have specifically explored its potential in the context of cardiac MRI applications or its role in addressing dataset imbalances [28, 83, 29, 84]. Indeed, none have tackled this imbalance without incorporating real data into their pipeline, whereas this approach relies solely on a single atlas.

In Computed Tomography (CT) imaging, there have been successful approaches using style transfer to augment available training data, as noted by [129]. While this study emphasises preserving contextual features in latent representations, it diverges from the focus of this chapter's work on model objectives. In [129], style transfer aims to diversify the training set for improved generalisability, whereas in this chapter, style transfer transforms mask-like images from a morphologically altered atlas into realistic MRI scans.

7.1.1.4 Cardiac MRI segmentation

In the last 5 years, a new paradigm has been established in the topic of cardiac magnetic resonance segmentation, working to extend the success of homogeneous datasets to more heterogeneous and diverse datasets [17, 21]. While the U-shaped architectures have remained a usual and powerful backbone for deep learning based models [61, 76], data processing has significantly seen a prolific production of new methods [2, 18, 19, 20]. Additionally, there have been efforts of extending the available data to increment the number of images for training by using image registration [1]. Despite these efforts, the exploration of how atlases representing one type of subject can be used to complement a dataset by applying deformations to their 3D model has not been conducted. A more extensive analysis is displayed in Chapter 2, in Section 2.6.

7.1.2 Contributions

The contributions of this work are threefold: (1) a method to modify a representative cardiac atlas and a sampling technique to obtain new subsets of data, specifically data from synthetic subjects with hypertrophic cardiomyopathy (HCM) and dilated right ventricle (DRV), is proposed; (2) recent advances in the image-to-image style transfer domain are built upon to obtain realistic MRI scans from its labels; (3) synthetic and real data are successfully integrated to address the imbalance in the dataset, and segmentation networks are trained to improve performance in patients with diseases not present in the training set. These three points constitute a novel pipeline that can be implemented in deep learning models to complement available data, similar to existing data augmentation techniques.

7.2 Method

The proposed method is founded on the premise that inaccuracies in model predictions stem from the unique cardiac morphology observed in patients with pathologies. The hypothesis is that the absence of patients with specific pathologies in the training data can cause the model to overfit to the pathologies that are present. To mitigate this, scans exhibiting morphology similar to the pathologies not originally represented in the training data are introduced. This augmentation increases the dataset's diversity, resulting in enhanced segmentation model performance. The overall structure of the method's pipeline is depicted in Figure 7.3.

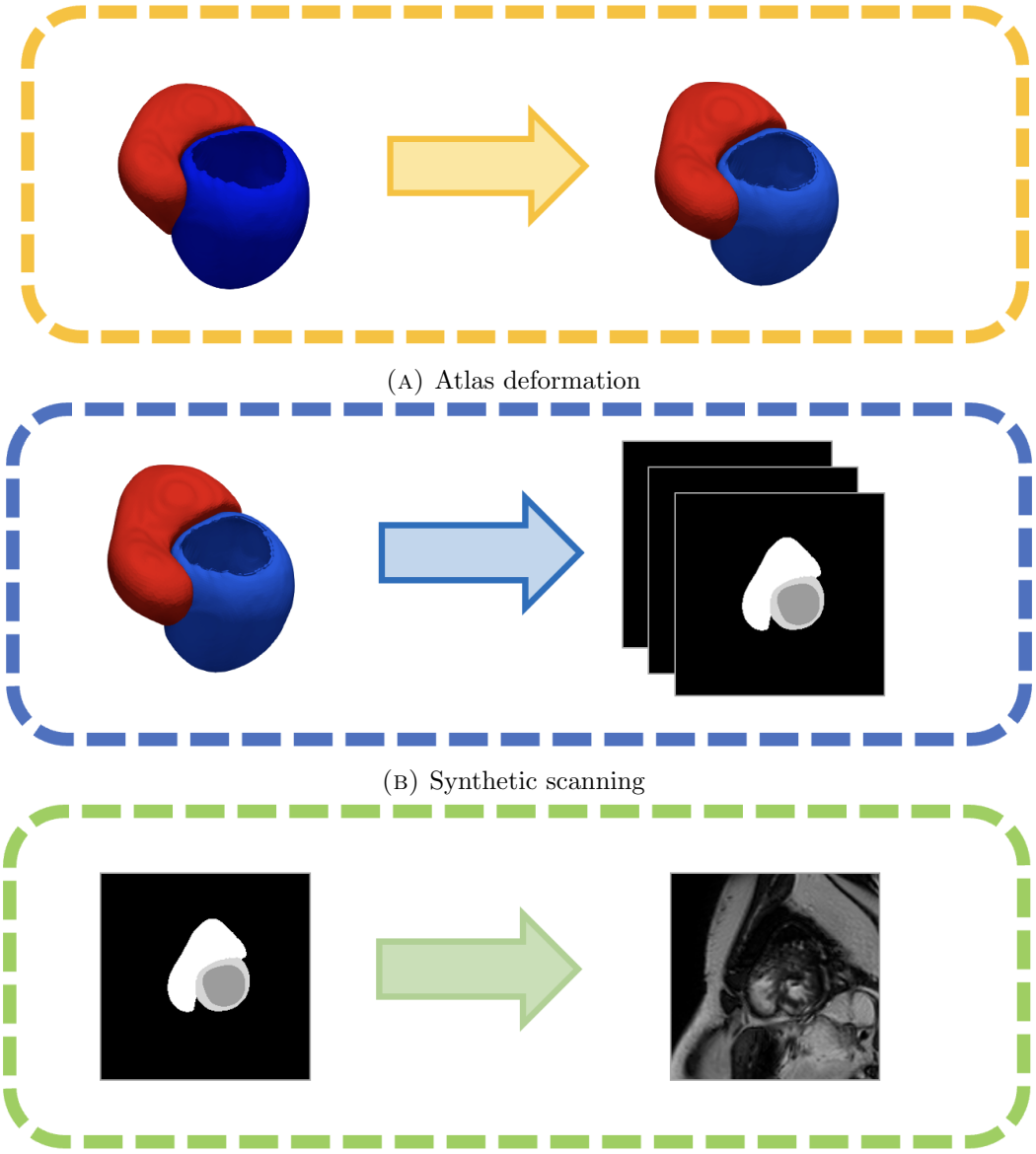
To rectify the imbalance in the training data, what is referred to as "synthetic patients" was introduced. These synthetic patients are created by applying deformations to an atlas, essentially simulating the heart structures of the two previously mentioned diseases, Hypertrophic Cardiomyopathy (HCM) and Dilated Right Ventricle (DRV). These simulations are achieved by virtually slicing the 3D model of the heart to generate these synthetic patient representations.

The incorporation of these synthetic patients into the dataset involves a crucial style transfer phase. During this step, images that resemble labels or structural representations are transformed into realistic MRI scans. This transformation was accomplished using a technique called "CUT" as described in the reference [124].

Finally, a segmentation network is trained using real and synthetic data.

7.2.1 Atlas-based data representation

Rather than having a model created based on the existing data, which would be computationally costly and could introduce errors during the 3D reconstruction steps, the atlas described in [123] was chosen for utilisation. This atlas represents



(A) Atlas deformation

(B) Synthetic scanning

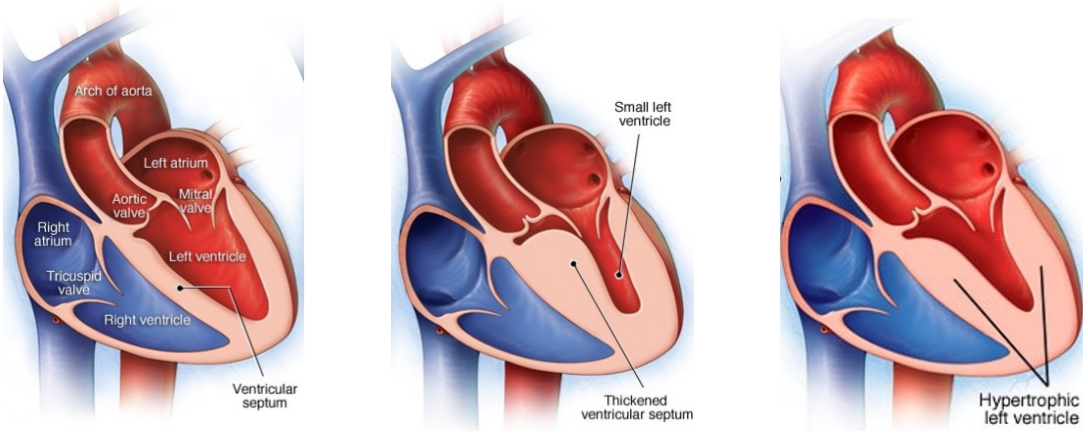
(C) Style transfer

FIGURE 7.3: The pipeline of the proposed method. The three main components correspond to: 7.3a the atlas deformation, 7.3b the synthetic scanning, and 7.3c the style transfer.

over 1000 healthy subjects, providing a comprehensive and reliable reference for this purposes.

7.2.2 3D Deformations and Virtual Scanning Techniques

For this work, three different deformations representing three different types of subjects were created: (1) a healthy subject, (2) 20 subjects with heart structures similar to patients with a dilated right ventricle, and (3) 20 subjects with heart structures similar to patients with hypertrophic cardiomyopathy. The visual representation of these diseases is presented in Figure 7.4.



(A) Healthy (B) DV (C) HCM

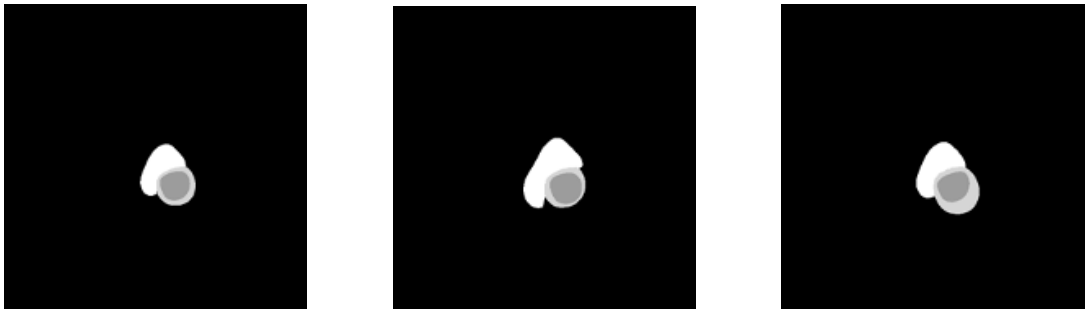
FIGURE 7.4: Anatomical structure of the heart. Figure reproduced from [36]. Where: 7.4a is a healthy heart, 7.4b is a dilated ventricle, and 7.4c ventricular hypertrophy.

The synthetic patients are obtained by performing a spherical cast operation (see Figure 7.5) to the cavities of the heart model. The spherical cast consists of a *To Sphere* transformation that will generate different results depending on the number and arrangement of the elements selected. Depending on the number of selected parts in each region (epicardium, endocardium, or right ventricle) and the range of the deformation, a more or less prominent disease in the synthetic subject was obtained. For HCM, the epicardium and endocardium were enlarged, while for DRV, the RV was enlarged. The factor of the spherical cast ranged between 0 and 0.5.

(a) **Factor = 0** (b) **Factor = 0.25** (c) **Factor = 0.5** (d) **Factor = 0.75** (e) **Factor = 1**

FIGURE 7.5: Example of spherical cast applied to a cube with different factors.

The resulting 3D models are then processed as if they were scanned by obtaining slices throughout the longitudinal axis, to sample the equivalent to short-axis labels. An example for each subset of data is presented in Figure 7.6.



(A) Healthy (B) DRV (C) HCM

FIGURE 7.6: Representations of the sampled slices from the 3D heart model. Where: 7.6a is no deformation, 7.6b deformations on the right ventricle representing a dilated right ventricle, and 7.6c deformations on the left ventricle representing hypertrophic cardiomyopathy in the left ventricle.

7.2.3 Style transfer

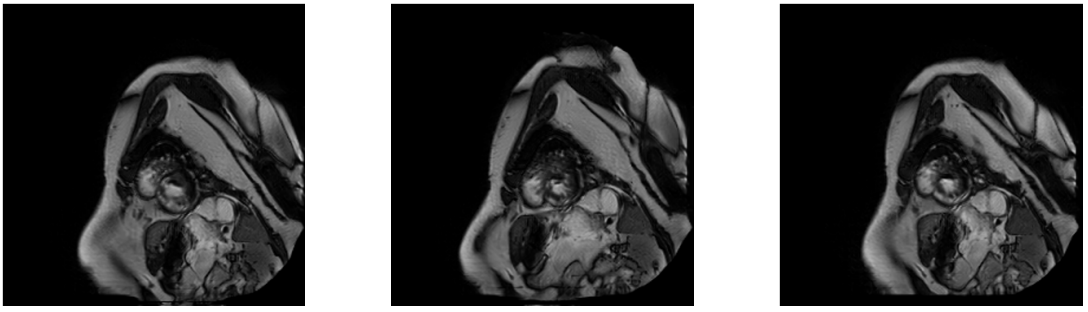
A specific CUT [124] model with the available data from M&Ms 2 [21] was trained for this purpose, where one domain is the MRI scan and the other is the label from the MRI scan.

CUT was the preferred style transfer model because, besides being based on CycleGAN [85], it has a contrastive term that encourages spatial consistency in the generated image. The CUT loss function is presented by:

$$\mathcal{L}_{\text{GAN}}(G, D, X, Y) + \lambda_X \mathcal{L}_{\text{PatchNCE}}(G, H, X) + \lambda_Y \mathcal{L}_{\text{PatchNCE}}(G, H, Y) \quad (7.1)$$

where $\lambda_Y \mathcal{L}_{\text{PatchNCE}}(G, H, Y)$ is the contrastive term that enhances spatial consistency with the source image in X (synthetic image in this case). This term promotes the proximity of input-output patches from a specific location in an image within the feature space while ensuring that such patches are distant from other patches in the image. The discriminator part of the GAN is represented by D and the generator part as a G. The weights of a two-layer perceptron that projects the patches to the feature space are denoted as H, and the hyperparameters λ_X and λ_Y regulate the influence of the respective contrastive terms.

The results of applying a CUT trained on the available data to the synthetic data obtained from the previous steps are presented in Figure 7.7. The two domains are: (1) the MRI scans, and (2) the labels of the MRI scans. In other words, the CUT model learns to transfer the style of an MRI scan to a label, resulting in a realistic MRI scan with the structure displayed in a label. The resulting data was used as annotated data for the segmentation models.



(A) Healthy (B) DRV (C) HCM

FIGURE 7.7: Slices from Figure 7.6 processed with the trained CUT [124]. Where: 7.7a is no deformation, 7.7b deformations on the right ventricle representing a dilated right ventricle, and 7.7c deformations on the left ventricle representing hypertrophic cardiomyopathy in the left ventricle

7.2.4 Segmentation Network

To test the performance of the models trained with the different datasets the same architecture and the same preprocessing and normalisation techniques were used. In particular, a standard U-Net [61] with 32 filters in the first convolutional

layer was implemented. Data normalisation was performed with z normalisation and histogram standardisation. The preprocessing consisted in centre-cropping with padding.

7.2.4.1 Short-axis Cardiac MRI Data

M&Ms2 dataset [21] was used, released as part of the Multi-Center, Multi-View & Multi-Disease Right Ventricular Segmentation in Cardiac MRI Challenge. The performance was evaluated in all chambers for the short-axis view. In particular, the performance of the different pathologies present in the subjects of the dataset was compared. The pathologies are dilated left ventricle (LV), hypertrophic cardiomyopathy (HCM), congenital arrhythmogenesis (ARR), tetralogy of fallot (FALL), interatrial communication (CIA), dilated right ventricle (DRV), and tricuspidal regurgitation (TRI).

7.3 Experimental Methodology and Results

The experiment consisted of evaluating the performance of four models trained with three different sets of data: (1) normal (healthy) cases from M&Ms2 plus the atlas sampled without deformation, (2) normal (healthy) cases from M&Ms2 plus 20 dilated right ventricle deformations applied to the atlas, (3) normal (healthy) cases from M&Ms2 plus 20 hypertrophic cardiomyopathy deformations applied to the atlas, and (4) normal (healthy) cases from M&Ms2 and classic data augmentation (affine and elastic transformation). The first model serves as a control model to compare the improvements when adding the synthetic data, and the last one is a comparison to classic data augmentation. The augmentations were applied with an overall probability of 0.5, and then each augmentation had equal probabilities of being applied.

The evaluation of the models was performed over the test split of the M&Ms2. In particular, the results for normal (healthy), hypertrophic cardiomyopathy,

	Normal	**HCM**	**DRV**	**Augs**
Real Patients	40	40	40	40
Synthetic Patients Healthy	1	1	1	1
Synthetic Patients HCM	0	20	0	0
Synthetic Patients DRV	0	0	20	0

TABLE 7.1: Number of real and synthetic patients used to train the segmentation model. Synthetic refers to the atlas or its deformations with the style-transfer performed.

and dilated right ventricle are presented; a part of overall performance. DICE scores are provided for each region. The results are presented in Table 7.2.

Data subset	**Model**	**DICE**				**Hausdorff (mm)**
		LV	**MYO**	**RV**	**ALL**	**All**
Normal (Healthy)	Healthy	0.9417	0.8252	0.8542	0.8737	6.311
	HCM	0.9400	0.8397	0.8273	0.8690	**6.2916**
	DRV	0.9333	0.8380	0.8053	0.8589	6.4883
	Augs	**0.9477**	**0.8416**	**0.8790**	**0.8894**	6.4223
HCM	Healthy	**0.9382**	**0.8583**	0.8284	**0.8750**	**6.5544**
	HCM	0.9195	0.8574	**0.8317**	0.8700	6.6080
	DRV	0.9148	0.8450	0.8239	0.8612	6.7190
	Augs	0.9089	0.8198	0.8111	0.8467	6.635
DRV	Healthy	0.7808	0.6036	**0.7742**	0.7195	7.3460
	HCM	0.7625	0.6146	0.7662	0.7149	7.1411
	DRV	**0.8049**	**0.6499**	0.7723	**0.7424**	6.9020
	Augs	0.7916	0.6243	0.7142	0.7101	**6.5030**
All	Healthy	0.8577	0.7149	0.7720	0.7815	6.8173
	HCM	0.8313	0.7137	**0.7856**	0.7769	6.770
	DRV	**0.8672**	**0.7388**	0.7729	**0.7930**	6.9027
	Augs	0.8581	0.7228	0.7667	0.7825	**6.5474**

TABLE 7.2: Segmentation performances for each subset of data. Each subset presents the DICE score (higher is better) and Hausdorff distance (lower is better) for each of the four models and each of the regions (Left Ventricle, Myocardium, and Right Ventricle). Models refer to the model trained with healthy patients of the original dataset [21] plus the atlas or its synthetic deformed aliases. The best results are in bold.

The results show how the addition of synthetic data had a positive impact on the performance, with a strong emphasis on the dilated right ventricle deformations. Overall performance increased the DICE score in that model by 0.01, and by 0.05 in the myocardium for the DRV subset of data, where the mean performance was more than 0.02 better than in the healthy subset.

Additionally, Figure 7.8 provides a visual comparison of the various models trained for the different tissues. The DRV model, represented in green, outperforms the other models with the exception of the RV, where the HCM model, displayed in yellow, achieves a better DICE score. The Augs model, depicted in red, shows an improvement over the Healthy model (in blue), as expected, but does not surpass the performance of the models utilising synthetic balancing.

In comparison to the study by [129], the results demonstrate similar enhancements despite expanding the available training data through a distinct method. This approach offers the additional benefit of generating extra data specifically in targeted directions where increased training data is needed. The evaluation of this research focuses on assessing the impact in these specific directions.

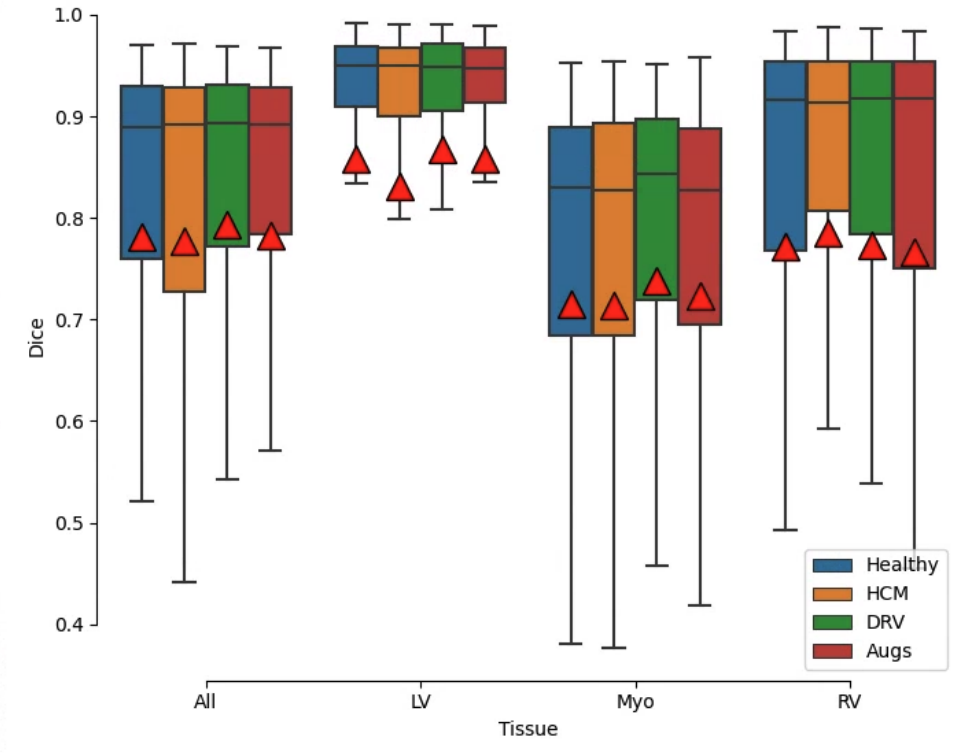


FIGURE 7.8: Segmentation performances for each subset of data. Each subset presents the DICE score (higher is better) for each of the four models (Healthy, HCM, DRV, and Augs) and each of the tissues (Left Ventricle, Myocardium, Right Ventricle and all regions). Models refer to the model trained with healthy patients of the original dataset plus the atlas or its synthetic deformed aliases. The mean is plotted with a red mark.

7.4 Conclusions

In this study, a style transfer model was successfully trained to effectively generate synthetic cardiac MRI images sampled from a deformed atlas. Those synthetic images were successfully used to balance the training dataset. The results show how the addition of this data meant an increase in the DICE score of up to 0.05 in some regions within the target pathologies of the data. Nonetheless, the method was significantly less succcesful when simulating hypertrophic cardiomyopathies than it was when synthesising dilated right ventricle patients.

The incorporation of synthetic data in the DRV model led to improvements of up to 0.01 in the overall dataset and up to 0.03 on the DRV subset compared to classical augmentations. This demonstrates a distinct advantage over conventional techniques.

Conversely, the HCM subset exhibits optimal performance when utilising the Healthy model. This unexpected outcome indicates that the model generalised from healthy data may capture essential features or patterns in the HCM subset more effectively than a condition-specific model. Additionally, this discrepancy might also be attributed to this method's lack of precision in capturing the specific morphologies of HCM data. These contrasting results underscore the importance of selecting appropriate modelling strategies based on the unique attributes of each data subset, and they open avenues for further investigation into the underlying reasons for these performance disparities.

Future endeavours will delve into incorporating additional and more accurate deformations to synthesise a more extensive range of diseases, coupled with enhancements in the style-transfer component of the pipeline. Moreover, extending the pipeline's design to generate synthetic patients with pathologies identified through functional analysis would constitute a significant advancement in the field.

In the culmination of this research journey, the preceding chapters collectively unveil a comprehensive exploration into various facets of CMRI analysis. From the meticulous investigation of diverse architectures and the development of a synthetic labelling pipeline to address dataset imbalances, to the detailed examination of time frame detection, mitigation of respiratory motion artifacts, and the pioneering use of on-demand synthetic data, each chapter contributes a unique perspective to the overarching goal of enhancing CMRI segmentation. The findings underscore the potential for foundational models in cardiac MR analysis, as demonstrated by the successful fusion of pre-trained models, innovative data augmentation strategies, and the generation of synthetic data. As the research objectives align with the ongoing advancements in deep learning methodologies, the implications extend beyond the immediate scope of this study. These insights not only bolster the understanding of CMRI analysis but also pave the way for future research directions and the continued evolution of foundational models in medical imaging. In the subsequent concluding chapter, these collective contributions are distilled into a comprehensive synthesis, with reflections on the achievements, limitations, and proposals for future research in the dynamic landscape of cardiac MR analysis.

Chapter 8

Conclusion

Neural networks have reached impressive results in many medical imaging applications, in particular, in cardiac MR. Clinical tools have integrated them since they have become the new state-of-the-art for a number of their intrinsic tasks. These results, however, come with some constraints: to achieve top scoring results neural networks require large and diverse datasets, with expensive labels, and significant computational requirements. Obtaining this data is costly and it is difficult to share. Adding to this challenge, there are cohorts of patients who can not get the scans without holding their breaths, resulting in artifacts that deform the contouring of the heart's cavities. The research presented in this thesis addresses these challenges: Chapters 3 and 4 address alternatives to obtain more benefits from the available data and labels and to enhance generalisability towards diverse data (scan vendors, pathologies and views); Chapter 5 addresses the possible performance leak due to incorrect time frame labelling; Chapter 6 focuses on mitigation against respiratory motion artifacts; and lastly, Chapter 7 focuses on developing a pipeline to generate synthetic data on demand that could be integrated in the building of a foundational model.

In particular, Chapters 3, and 4 present innovations in advancing model generalisation with synthetic label propagation and in handling multi-view data. This is necessary to tackle the challenges in developing accurate segmentation

models from cardiac MRI data that is essential for extracting specific metrics such as the Ejection Fraction for reliable clinical interpretation. The former explores the utilisation of synthetic labels obtained through image registration of intermediate time frames, bolstering the model's generalisation capabilities to analyse scans from previously unseen vendors. The latter examines various model architectures and post-processing techniques involving the anatomical position of scans, evaluating their performance on images from different views.

Chapter 5 focuses on including the time frame labelling in the context of segmentation, benefiting from the networks trained with that task. This is important to prevent the accuracy of clinical metrics from dropping when no ECG signal or time frame label accompanies the scans. This chapter combines pre-trained encoders with sequential modules to provide a fast, accurate, and reproducible model.

The focus of Chapter 6 is to mitigate the respiratory motion artifacts in cardiac MR data. In this chapter, a method for leveraging pre-trained models for cardiac segmentation and the integration of MRI-specific augmentations to mitigate the former artifacts is presented.

Finally, in Chapter 7, the idea of building a foundational model by generating synthetic data is worked on. In particular, a novel pipeline to obtain on-demand synthetic data was developed. For this proof of concept, two different pathologies with known distinctive morphologies were chosen to develop the idea. The pipeline consisted of deforming an atlas according to the morphologies, sampling the data, and finally applying style transfer to convert the images into MRI-looking scans with a proposed deep learning model.

In this chapter, Section 8.1 addresses the objectives proposed in Chapter 1, and how the research presented in Chapters 3, 4, 5, 6 and 7 addresses important aspects for the former. Section 8.2 summarises the research contributions of the thesis. Finally, Section 8.3 presents the limitations of this work and elaborates on the suggestions for future research.

8.1 Research objectives

The research objectives introduced in Chapter 1 are discussed in this section in light of the research presented in the corresponding chapters.

Objective 1: To develop more generalisable models that can be used in diverse settings.

- Objective 1.1: Enhance model robustness and generalisability to the variations presented by various imaging equipment sources.
- Objective 1.2: To develop a model capable of effectively handling and distinguishing among various cardiomyopathies, ensuring robust performance across different pathological conditions.

Chapters 3 and 4 empirically demonstrate that exploiting relationships among data within the same dataset enhances the neural network's ability to generalise across diverse settings. Specifically, the experiments detailed in Chapter 3 reveal that extending labels from annotated time frames to intermediate ones improves network performance on data from previously unseen vendors. In the experiments conducted in Chapter 4, it is established that the choice of architectures during training significantly influences performance across different views, vendors, and pathologies. Moreover, the findings indicate that pre-trained networks consistently yield superior performance in most cases. Additionally, within this chapter, a post-processing method is proposed to label predictions from one view using predictions from another view, accomplished through pixel mapping with metadata derived from the images.

Objective 2: Automatically detect end-diastolic and end-systolic time frames for more accurate analysis and better reproducibility on benchmark datasets.

The experiments carried out in Chapter 5 delved into assessing the performance, both in accuracy and speed, of a neural network incorporating a pre-trained

encoder and a sequential module. Two distinct modules of this nature were tested: an LSTM [115] and a transformer encoder [116]. Furthermore, this chapter established a benchmark using a publicly available dataset, facilitating the reproducibility of future research endeavours.

Objective 3: To develop a method to mitigate respiratory motion artifacts on cardiac MRI data.

In Chapter 6, the research centered on mitigating against respiratory motion artifacts in CMRI data. The experimentation revolved around fine-tuning pre-trained models in conjunction with data augmentation techniques that replicated the blurring effects associated with the mentioned artifacts. It is concluded that this combination yielded a more robust and resilient model, effectively mitigating the effect of artifacts.

Objective 4: To synthetically balance datasets.

Chapter 7 centered on exploring the use of on-demand synthetic data to address dataset imbalances. Specifically, the experimentation aimed to balance a dataset encompassing various pathologies, selecting two with known morphological differences compared to healthy subjects. The approach involved designing a pipeline capable of generating new CMRI data from an atlas. This pipeline included: (1) deforming the atlas to match the corresponding morphology, (2) sampling the resulting 3D model to generate images, and (3) applying style transfer to produce MRI-like images. The experiment compared the impact of adding various synthetically generated subjects to the training data on a segmentation network, contrasting the results with other classic data augmentation methods based on deformation. The conclusion highlighted the success of the proof of concept, suggesting that this approach could represent a significant advancement in foundational models.

8.2 Research contributions and proposed solutions

In the following section, a discussion related to contributions of this thesis is presented.

How can cardiac MR analysis be improved with deep learning methods?

To enhance the applicability of the developed models across diverse settings, the research focuses on the creation of more generalisable models. By leveraging comprehensive pre-training on diverse datasets, the aim is to establish a robust foundation that captures broad knowledge. This approach facilitates the development of models that can be fine-tuned for specific tasks using smaller, task-specific datasets. The contribution of this work lies in the pursuit of generalisability, allowing the models to demonstrate efficacy and relevance in a wide array of contexts and applications.

Can image registration be used to propagate labels to intermediate time frames?

In publicly available cardiac MR datasets, annotations are typically provided only for key time frames, such as End-Diastole and End-Systole. To address this limitation, in Chapter 3, annotations were expanded to intermediate frames by propagating labels from the ED frame. This extension was achieved through image registration. Initially, the registration field between two images was computed and subsequently applied to the corresponding mask.

Can the analysis be improved if intermediate unlabelled data is included?

The integration of scans enriched with synthetically propagated labels yielded notable performance improvements, particularly evident when applied to scans

originating from vendors not included in the initial training set, as seen in Chapter 3. This strategy effectively contributed to enhancing the model's generalisation capabilities, showcasing its adaptability to diverse data sources and underscoring the robustness of the trained model in handling variations across different vendors. The extended capability to generalise across a broader spectrum of datasets reinforces the practical utility and versatility of the model in real-world applications, where data heterogeneity is often encountered.

Can information from the short-axis be exploited to improve model performance in the long-axis?

CMRI data typically includes metadata containing crucial information, with pixel coordinates in the anatomical position being particularly vital. To leverage this information effectively, it is imperative to identify the mapping corresponding to the specific case. In the experiments conducted in Chapter 4, this mapping was successfully accomplished using the method denoted as method 3. By employing this post-processing step, results comparable to those obtained through neural network predictions were achieved. The introduction of this proposed post-processing technique proves advantageous in scenarios where there is limited long-axis data or when model performance is sub-optimal. This innovation addresses challenges associated with insufficient data and enhances the overall robustness of the model's predictions.

Is it important to select different architectures for different tasks?

One strategy for optimising performance on CMRI data involves training dedicated models for specific subsets of data. In Chapter 4, this approach was investigated by exploring four distinct architectures and evaluating their performances across a variety of vendor scans and pathologies. The experimentation produced intriguing findings, as the optimal architecture varied depending on the specific scanner and pathology under consideration. Nevertheless, it was consistently observed that architectures incorporating a pre-trained encoder outperformed others in the majority of cases. This underscores the effectiveness of leveraging

pre-training strategies to enhance the overall robustness and adaptability of the models across diverse datasets.

Can end-diastolic and end-systolic time frames from cardiac MR data be detected automatically?

This contribution entails the creation of an automated system designed to precisely detect end-diastole and end-systole time frames. The automated process significantly contributes to the efficiency and robustness of cardiac imaging analyses. In Chapter 5, the utilisation of pre-trained segmentation neural networks and sequential modules to expedite this task was investigated, enabling parallelised execution without compromising accuracy. This research also establishes a benchmark using publicly available datasets, facilitating reproducibility and seamless comparisons across studies. In essence, this work not only advances efficiency and accuracy in cardiac imaging analyses but also sets a standard for future research in the field.

Can a deep learning model be resilient to respiratory motion artifacts?

Addressing the challenge of preventing models from underperforming on CMRI data with artifacts is noteworthy due to its implications, though inherently difficult due to the variable intensities of motion blurring caused by different levels of patient breathing. Chapter 6 details the exploration of this challenge, where experimentation with a pre-trained model and various data augmentation techniques resulted in notable improvements and the successful mitigation of artifacts. Specifically, augmentations designed to mimic the effects of motion artifacts in the images were applied, encompassing a range of intensities to comprehensively address the variability observed in patient breathing.

Can atlas data be leveraged to produce synthetic cardiac MR data?

In Chapter 7, the exploration of utilising an atlas to generate synthetic data on demand was conducted, offering a convenient means to acquire specific data

for balancing datasets. Throughout the experimentation, a pipeline method designed to deform the atlas to emulate two diseases with distinct morphological characteristics was introduced. This method involved sampling images from the deformed atlas and applying style transfer to render images with an appearance akin to MRI scans. The outcomes of the experimentation showcased an enhancement in results, particularly for one of the pathologies. Furthermore, a comparative analysis with results obtained through traditional data augmentation deformations was conducted, revealing that the proposed pipeline outperformed in comparison. The overall success of the proof of concept was affirmed, signalling the promising potential of this approach.

8.3 Limitations, recommendations and future work

In this section, the primary research avenues outlined throughout the thesis are gathered, potential prospects for future research within these domains are explored, and the limitations of the work are elucidated:

- Generalisable models that can be used in diverse settings:
  - A limitation within the proposed work in this context is found in the fact that image registration was conducted exclusively in a 2D scenario. It remains uncertain whether incorporating a combination of 2D and 3D registration in analogous approaches could yield more significant performance improvements.
  - Continuing in the same vein, the architecture search was based on various 2D architectures that offer avenues for further exploration. Additionally, future research might consider incorporating ensembled networks or models in cascade. However, it is important to note that such approaches could introduce more complex training strategies

and potentially demand a larger computational budget. Finally, embracing holistic approaches that simultaneously consider both views could potentially lead to further improvements, as highlighted by some researchers [72].

- Automatically detect end-diastolic and end-systolic time frames for more accurate analysis and better reproducibility on benchmark datasets:
  - While this exploration resulted in strong performance, future methods could potentially concentrate on unsupervised approaches to attain comparable or superior results. Approaches like [24], which explore deriving a signal from the registration process, appear particularly interesting in this context. The hypothesis suggests that the incorporation of optical flow [117] might further enhance performance and reduce the reliance on labelled data.
- To develop a method to mitigate respiratory motion artifacts on cardiac MRI data:
  - In addressing this specific challenge, the efforts were centered on motion artifacts without considering their diversity (e.g. age or pathologies). It would be intriguing to explore how artifacts may vary in this context and to study mitigating each specific type. Similarly, the domain of MRI reconstruction stands out as a challenging yet captivating area for exploration with similar motivation. Perhaps, the advancements made in the field of artifact mitigation could offer valuable insights and improvements in the realm of MRI reconstruction.
- To synthetically balance datasets:
  - The experimentation primarily centered around two diseases with notable morphologies observable in short-axis images. As such, there is room for exploration of additional pathologies in long-axis views,

such as tetralogy of Fallot. Future research could investigate other pathologies that exhibit significant changes in heart function. While these may be more challenging to replicate, they hold greater potential for enhancing diagnostic capabilities.

– *Is it feasible to create a foundational model that harnesses pre-training on extensive and diverse datasets, capturing a wide range of knowledge, and subsequently fine-tunes this model for specific tasks using smaller, task-specific datasets, particularly with the incorporation of synthetic CMRI data?*

  Building on the success of the proof-of-concept pipeline, it is posited that the construction of a foundational model for cardiac MR is an attainable goal. Atlases serve as invaluable resources for representing large patient cohorts without compromising individual privacy. Nonetheless, a comprehensive study is essential for the style transfer component, which captures CMRI texture features, to enhance its performance and explore both its limitations and implications.

Appendix A

NIfTI Method 3 code

```
import nibabel as nib
import numpy as np
import numpy.linalg as npl

def method3_forward(i,j,k,SA,LA):
    # Loading and info bloc
    SA_pred = SA.get_fdata()
    SA_info = SA.header
    SA_s = SA_info.get_sform()
    LA_pred = LA.get_fdata()
    LA_info = LA.header

    # Matrix bloc
    affine_la = LA_info.get_qform()
    affine_sa = SA_info.get_qform()
    M1 = affine_sa[:3,:3]
    M2 = affine_la[:3,:3]
    M2_prime = npl.inv(M2)

    # Offsets bloc
    SA_off = affine_sa[:3,3][np.newaxis]
    LA_off = affine_la[:3,3][np.newaxis]

    # SA to anatomical
    SA_anatomical = M1.dot([i, j, k]) + SA_off

    # LA to anatomical
    SA_anatomical_off = SA_anatomical - LA_off
    SA_LA = M2_prime.dot(SA_anatomical_off.T)

    return int(SA_LA[0]), int(SA_LA[1])


```

Appendix B

Results for each architecture and view evaluated for the different scan vendors.

Short Axis

	DICE			Total	
	LV	**MYO**	**RV**	**DICE**	**Hausdorff**
Siemens	0.9579	0.8871	0.9187	0.9212	6.8208
Philips	0.9479	0.8630	0.9224	0.9111	7.1085
GE	0.8930	0.8330	0.8580	0.8613	8.3549

Table B.1: Vendor results for U-Net over the Short Axis images.

	DICE			Total	
	LV	**MYO**	**RV**	**DICE**	**Hausdorff**
Siemens	0.9578	0.8837	0.9161	0.9192	6.6107
Philips	0.9489	0.8600	0.9065	0.9052	6.7915
GE	0.9262	0.8427	0.8806	**0.8832**	8.2118

Table B.2: Vendor results for Deeplabv3 on the Short Axis images.

	DICE			Total	
	LV	**MYO**	**RV**	**DICE**	**Hausdorff**
Siemens	0.9399	0.8607	0.8801	0.8935	7.7813
Philips	0.9122	0.8295	0.8792	0.8736	7.6662
GE	0.8605	0.7893	0.8322	0.8274	9.8808

TABLE B.3: Vendor results for transUNet on the Short Axis images.

	DICE			Total	
	LV	**MYO**	**RV**	**DICE**	**Hausdorff**
Siemens	0.9608	0.8884	0.9183	**0.9225**	6.7653
Philips	0.9507	0.8635	0.9234	0.9126	7.0532
GE	0.9079	0.8236	0.8502	**0.8606**	8.1474

TABLE B.4: Vendor results for CE-Net over the Short Axis Images.

Long Axis

	DICE			Total	
	LV	**MYO**	**RV**	**DICE**	**Hausdorff**
Siemens	0.9402	0.8436	0.8784	0.8874	8.1299
Philips	0.9398	0.8662	0.8949	0.9003	8.0030
GE	0.8762	0.7638	0.7749	0.8050	8.5716

TABLE B.5: Vendor results for U-Net over the Long Axis images.

	DICE			Total	
	LV	**MYO**	**RV**	**DICE**	**Hausdorff**
Siemens	0.9424	0.8486	0.8931	**0.8947**	7.6309
Philips	0.9492	0.8568	0.9087	0.9049	7.1991
GE	0.9209	0.7735	0.8358	**0.8434**	8.1294

TABLE B.6: Vendor results for Deeplabv3 in Long Axis images.

	DICE			Total	
	LV	**MYO**	**RV**	**DICE**	**Hausdorff**
Siemens	0.9022	0.7931	0.8019	0.8324	8.9289
Philips	0.9220	0.8173	0.8379	0.8591	8.7059
GE	0.8347	0.7210	0.7028	0.7528	10.1022

TABLE B.7: Vendor results for transUNet on the Long Axis images.

	DICE			Total	
	LV	**MYO**	**RV**	**DICE**	**Hausdorff**
Siemens	0.9433	0.8502	0.8831	0.8922	8.2293
Philips	0.9507	0.8652	0.9087	**0.9082**	7.7915
GE	0.8895	0.7650	0.8353	0.8299	8.8623

TABLE B.8: Vendor results for CE-Net over the Long Axis images.

References

[1] Carles Garcia-Cabrera, Kathleen M. Curran, Noel E. O'Connor, and Kevin McGuinness. "Semi-supervised learning of cardiac MRI using image registration". In: *Irish Pattern Recognition & Classification Society Conference Proceedings* (Sept. 2021). URL: https://doras.dcu.ie/26161/.

[2] Carles Garcia-Cabrera, Eric Arazo, Kathleen M. Curran, Noel E. O'Connor, and Kevin McGuinness. "Cardiac Segmentation Using Transfer Learning Under Respiratory Motion Artifacts". en. In: *Statistical Atlases and Computational Models of the Heart. Regular and CMRxMotion Challenge Papers*. Ed. by Oscar Camara et al. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2022, pp. 392–398. ISBN: 978-3-031-23443-9. DOI: 10.1007/978-3-031-23443-9_36.

[3] Carles Garcia-Cabrera, Kathleen M. Curran, Noel E. O'Connor, and Kevin McGuinness. "Cardiac Magnetic Resonance Phase Detection Using Neural Networks". In: *2023 31st Irish Conference on Artificial Intelligence and Cognitive Science (AICS)*. 2023, pp. 1–4. DOI: 10.1109/AICS60730.2023.10470772.

[4] Carles Garcia-Cabrera et al. "Synthetic Balancing of Cardiac MRI Datasets". In: *Medical Image Understanding and Analysis*. Ed. by Moi Hoon Yap, Connah Kendrick, Ardhendu Behera, Timothy Cootes, and Reyer Zwiggelaar. Cham: Springer Nature Switzerland, 2024, pp. 353–365. ISBN: 978-3-031-66958-3.

[5] James Callanan, Carles Garcia-Cabrera, Niamh Belton, Gennady Roshchupkin, and Kathleen M. Curran. "Integrating feature attribution methods into the loss function of deep learning classifiers". In: *Irish Pattern Recognition & Classification Society Conference Proceedings* (Sept. 2022). URL: https://doras.dcu.ie/27762/.

[6] Julia Dietlmeier, Carles Garcia-Cabrera, Anam Hashmi, Kathleen M. Curran, and Noel E. O'Connor. "Cardiac MRI Reconstruction from Undersampled K-Space Using Double-Stream IFFT and a Denoising GNA-UNET Pipeline". In: *Statistical Atlases and Computational Models of the Heart. Regular and CMRxRecon Challenge Papers*. Ed. by Oscar Camara et al. Cham: Springer Nature Switzerland, 2024, pp. 326–338. ISBN: 978-3-031-52448-6.

[7] Jun Lyu et al. "The state-of-the-art in Cardiac MRI Reconstruction: Results of the CMRxRecon Challenge in MICCAI 2023". In: *arXiv e-prints* (2024), arXiv–2404.

[8] Nadine Kawel-Boehm et al. "Normal values for cardiovascular magnetic resonance in adults and children". In: *Journal of Cardiovascular Magnetic Resonance* 17.1 (2015), p. 29. ISSN: 1532-429X. DOI: 10.1186/s12968-015-0111-7. URL: https://doi.org/10.1186/s12968-015-0111-7.

[9] Martin Abadi et al. "Deep Learning with Differential Privacy". In: *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, Oct. 2016. DOI: 10.1145/2976749.2978318. URL: https://doi.org/10.1145%5C%2F2976749.2978318.

[10] Keith Bonawitz et al. "Practical Secure Aggregation for Privacy-Preserving Machine Learning". In: *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. CCS '17. Dallas, Texas, USA: Association for Computing Machinery, 2017, pp. 1175–1191. ISBN: 9781450349468. DOI: 10.1145/3133956.3133982. URL: https://doi.org/10.1145/3133956.3133982.

[11] Theo Ryffel et al. *A generic framework for privacy preserving deep learning*. 2018. DOI: 10.48550/ARXIV.1811.04017. URL: https://arxiv.org/abs/1811.04017.

[12] Nicolas Papernot. "A Marauder's Map of Security and Privacy in Machine Learning: An Overview of Current and Future Research Directions for Making Machine Learning Secure and Private". In: *Proceedings of the 11th ACM Workshop on Artificial Intelligence and Security*. AISec '18. Toronto, Canada: Association for Computing Machinery, 2018, p. 1. ISBN: 9781450360043. DOI: 10.1145/3270101.3270102. URL: https://doi.org/10.1145/3270101.3270102.

[13] Esther Puyol-Antón et al. "Fairness in cardiac magnetic resonance imaging: Assessing sex and racial bias in deep learning-based segmentation". en. In: *Front. Cardiovasc. Med.* 9 (Apr. 2022), p. 859310. DOI: 10.3389/fcvm.2022.859310. URL: https://www.frontiersin.org/articles/10.3389/fcvm.2022.859310/full.

[14] Khalid Alfudhili et al. "Current artefacts in cardiac and chest magnetic resonance imaging: tips and tricks". en. In: *Br. J. Radiol.* 89.1062 (June 2016), p. 20150987. DOI: 10.1259/bjr.20150987. URL: https://www.birpublications.org/doi/10.1259/bjr.20150987.

[15] Samer Alabed et al. "Validation of artificial intelligence cardiac MRI measurements: Relationship to heart catheterization and mortality prediction". en. In: *Radiology* 305.1 (Oct. 2022), pp. 68–79. DOI: 10.1148/radiol.212929. URL: https://doi.org/10.1148/radiol.212929.

[16] Nanyi Fei et al. "Towards artificial general intelligence via a multimodal foundation model". In: *Nature Communications* 13.1 (June 2022), p. 3094. ISSN: 2041-1723. DOI: 10.1038/s41467-022-30761-2.

[17] Víctor M. Campello et al. "Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Segmentation: The M&Ms Challenge". In: *IEEE Transactions on Medical Imaging* 40.12 (2021), pp. 3543–3554. DOI: 10.1109/TMI.2021.3090082.

[18] Fanwei Kong and Shawn C. Shadden. "A Generalizable Deep-Learning Approach for Cardiac Magnetic Resonance Image Segmentation Using Image Augmentation and Attention U-Net". In: *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*. Ed. by Esther Puyol Anton et al. Cham: Springer International Publishing, 2021, pp. 287–296. ISBN: 978-3-030-68107-4.

[19] Xiaoqiong Huang et al. "Style-Invariant Cardiac Image Segmentation with Test-Time Augmentation". In: *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*. Ed. by Esther Puyol Anton et al. Cham: Springer International Publishing, 2021, pp. 305–315. ISBN: 978-3-030-68107-4.

[20] Jun Ma. "Histogram Matching Augmentation for Domain Adaptation with Application to Multi-centre, Multi-vendor and Multi-disease Cardiac Image Segmentation". In: *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*. Ed. by Esther Puyol Anton et al. Cham: Springer International Publishing, 2021, pp. 177–186. ISBN: 978-3-030-68107-4.

[21] Víctor M. Campello et al. "Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Segmentation: The M&Ms Challenge". In: *IEEE Transactions on Medical Imaging* 40.12 (2021), pp. 3543–3554. DOI: 10.1109/TMI.2021.3090082.

[22] Razvan O. Mada, Peter Lysyansky, Ana M. Daraban, Jürgen Duchenne, and Jens-Uwe Voigt. "How to Define End-Diastole and End-Systole?: Impact of Timing on Strain Measurements". In: *JACC: Cardiovascular Imaging* 8.2 (2015), pp. 148–157. ISSN: 1936-878X. DOI: https://doi.org/10.1016/j.jcmg.2014.10.010. URL: https://www.sciencedirect.com/science/article/pii/S1936878X1400936X.

[23] Bin Kong and et al. "Recognizing End-Diastole and End-Systole Frames via Deep Temporal Regression Network". In: Cham: Springer International Publishing, 2016, pp. 264–272. ISBN: 978-3-319-46726-9.

[24] Sven Koehler et al. *Self-supervised motion descriptor for cardiac phase detection in 4D CMR based on discrete vector field estimations*. 2022. DOI: 10.48550/ARXIV.2209.05778. URL: https://arxiv.org/abs/2209.05778.

[25] Shuo Wang et al. *The Extreme Cardiac MRI Analysis Challenge under Respiratory Motion (CMRxMotion)*. 2022. DOI: 10.48550/ARXIV.2210.06385. URL: https://arxiv.org/abs/2210.06385.

[26] Ian J. Goodfellow et al. *Generative Adversarial Networks*. 2014. arXiv: 1406.2661 [stat.ML].

[27] Diederik P Kingma and Max Welling. *Auto-Encoding Variational Bayes*. 2022. arXiv: 1312.6114 [stat.ML].

[28] Sina Amirrajab et al. "Label-informed cardiac magnetic resonance image synthesis through conditional generative adversarial networks". eng. In: *Computerized Medical Imaging and Graphics: The Official Journal of the Computerized Medical Imaging Society* 101 (Oct. 2022), p. 102123. ISSN: 1879-0771. DOI: 10.1016/j.compmedimag.2022.102123.

[29] Youssef Skandarani, Nathan Painchaud, Pierre-Marc Jodoin, and Alain Lalande. *On the effectiveness of GAN generated cardiac MRIs for segmentation*. 2020. arXiv: 2005.09026 [eess.IV].

[30] Mengting Liu et al. "Style Transfer Using Generative Adversarial Networks for Multi-site MRI Harmonization". In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*. Ed. by Marleen de Bruijne et al. Cham: Springer International Publishing, 2021, pp. 313–322. ISBN: 978-3-030-87199-4.

[31] Peilun Shi et al. "Generalist Vision Foundation Models for Medical Imaging: A Case Study of Segment Anything Model on Zero-Shot Medical Segmentation". In: *Diagnostics* 13.11 (2023). ISSN: 2075-4418. DOI: 10.3390/diagnostics13111947. URL: https://www.mdpi.com/2075-4418/13/11/1947.

[32] Wikipedia contributors. *Heart*. Jan. 2023. URL: https://en.wikipedia.org/wiki/Heart.

[33] A.P. Farrell. "Design and physiology of the heart — Physiology of Cardiac Pumping". In: *Encyclopedia of Fish Physiology*. Ed. by Anthony P. Farrell. San Diego: Academic Press, 2011, pp. 1030–1037. ISBN: 978-0-08-092323-9. DOI: https://doi.org/10.1016/B978-0-12-374553-8.00060-5. URL: https://www.sciencedirect.com/science/article/pii/B9780123745538000605.

[34] Ioannis Vogiatzis et al. "The Importance of the 15-lead Versus 12-lead ECG Recordings in the Diagnosis and Treatment of Right Ventricle and Left Ventricle Posterior and Lateral Wall Acute Myocardial Infarctions". In: *Acta Informatica Medica* 27.1 (Mar. 2019), pp. 35–39. ISSN: 0353-8109. DOI: 10.5455/aim.2019.27.35-39.

[35] Marcelo Souto Nacif et al. "Cardiac magnetic resonance imaging and its electrocardiographs (ECG): tips and tricks". en. In: *The International Journal of Cardiovascular Imaging* 28.6 (Aug. 2012), pp. 1465–1475. ISSN: 1573-0743. DOI: 10.1007/s10554-011-9957-4.

[36] Mayo Clinic. *Diseases and conditions: Cardiomyopathies*. https://www.mayoclinic.org/diseases-conditions/chronic-fatigue-syndrome/symptoms-causes/syc-20360490. Jan. 2023.

[37] S.C. Bushong and G. Clarke. *Magnetic Resonance Imaging: Physical and Biological Principles*. Elsevier Health Sciences, 2003. ISBN: 9780323277655. URL: https://books.google.es/books?id=Xe9OAQAAQBAJ.

[38] Carles Garcia-Cabrera. *Automatic Ventricle Segmentation Using CNNs in Cardiac MRI. Bachelor's Thesis*. 2019.

[39] M Steinmetz, HC Preuss, and J Lotz. "Non-invasive imaging for congenital heart disease–Recent progress in cardiac MRI". In: *J Clin Exp Cardiolog. S* 8 (2012), p. 008.

[40] Wen-Yih Isaac Tseng, Mao-Yuan Marine Su, and Yao-Hui Elton Tseng. "Introduction to cardiovascular magnetic resonance: technical principles and clinical applications". In: *Acta Cardiologica Sinica* 32.2 (2016), p. 129.

[41] MRI Questions. *MRI Artifacts*. URL: https://mriquestions.com/hellipmr-artifacts.html.

[42] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324. DOI: 10.1109/5.726791.

[43] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Ed. by F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger. Vol. 25. Curran Associates, Inc., 2012. DOI: 10.1145/3065386. URL: https://dl.acm.org/doi/10.1145/3065386.

[44] Karen Simonyan and Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2014. DOI: arXiv:1409.1556. arXiv: 1409.1556 [cs.CV]. URL: https://doi.org/10.48550/arXiv.1409.1556.

[45] C. Szegedy et al. "Going deeper with convolutions". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, June 2015, pp. 1–9. DOI: 10.1109/CVPR.2015.7298594. URL: https://doi.ieeecomputersociety.org/10.1109/CVPR.2015.7298594.

[46] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[47] Fisher Yu and Vladlen Koltun. "Multi-Scale Context Aggregation by Dilated Convolutions". In: *International Conference on Learning Representations*. 2015. DOI: arXiv:1511.07122v3. arXiv: 1511.07122 [cs.CV].

[48] Shruti Jadon. "A survey of loss functions for semantic segmentation". In: *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)* (Oct. 2020). DOI: 10.1109/cibcb48159.2020.9277638. URL: http://dx.doi.org/10.1109/CIBCB48159.2020.9277638.

[49] "Mean Squared Error". In: *Encyclopedia of Machine Learning*. Ed. by Claude Sammut and Geoffrey I. Webb. Boston, MA: Springer US, 2010, pp. 653–653. ISBN: 978-0-387-30164-8. DOI: 10.1007/978-0-387-30164-8_528. URL: https://doi.org/10.1007/978-0-387-30164-8_528.

[50] D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge. "Comparing images using the Hausdorff distance". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15.9 (1993), pp. 850–863. DOI: 10.1109/34.232073.

[51] Alexander Andreopoulos and John K. Tsotsos. "Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI". In: *Medical Image Analysis* 12.3 (2008), pp. 335–357. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2007.12.003. URL: https://www.sciencedirect.com/science/article/pii/S1361841508000029.

[52] P. Radau et al. "Evaluation Framework for Algorithms Segmenting Short Axis Cardiac MRI." In: The MIDAS Journal, July 2009. DOI: 10.54294/g80ruo.

[53] Avan Suinesiaputra et al. "A collaborative resource to build consensus for automated left ventricular segmentation of cardiac MR images". In: *Medical Image Analysis* 18.1 (2014), pp. 50–62. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2013.09.001. URL: https://www.sciencedirect.com/science/article/pii/S1361841513001217.

[54] Caroline Petitjean et al. "Right ventricle segmentation from cardiac MRI: A collation study". In: *Medical Image Analysis* 19.1 (2015), pp. 187–202. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2014.10.

004. URL: https://www.sciencedirect.com/science/article/pii/S1361841514001509.

[55] Olivier Bernard et al. "Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved?" In: *IEEE Transactions on Medical Imaging* 37.11 (2018), pp. 2514–2525. DOI: 10.1109/TMI.2018.2837502.

[56] Víctor M. Campello et al. *Multi-Centre, Multi-Vendor & Multi-Disease Cardiac Image Segmentation Challenge*. Mar. 2020. DOI: 10.5281/zenodo.3715890. URL: https://doi.org/10.5281/zenodo.3715890.

[57] Cathie Sudlow et al. "UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age". In: *PLOS Medicine* 12.3 (Mar. 2015), pp. 1–10. DOI: 10.1371/journal.pmed.1001779. URL: https://doi.org/10.1371/journal.pmed.1001779.

[58] Phi Vu Tran. *A Fully Convolutional Neural Network for Cardiac Segmentation in Short-Axis MRI*. 2016. DOI: arXiv:1604.00494v3. arXiv: 1604.00494 [cs.CV].

[59] Jesse Lieman-Sifry, Matthieu Le, Felix Lau, Sean Sall, and Daniel Golden. *FastVentricle: Cardiac Segmentation with ENet*. 2017. DOI: 10.48550/ARXIV.1704.04296. URL: https://arxiv.org/abs/1704.04296.

[60] Ahmed S. Fahmy, Hossam El-Rewaidy, Maryam Nezafat, Shiro Nakamori, and Reza Nezafat. "Automated analysis of cardiovascular magnetic resonance myocardial native T1 mapping images using fully convolutional neural networks". In: *Journal of Cardiovascular Magnetic Resonance* 21.1 (Jan. 2019), p. 7. ISSN: 1532-429X. DOI: 10.1186/s12968-018-0516-1.

[61] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro

F. Frangi. Cham: Springer International Publishing, 2015, pp. 234–241. ISBN: 978-3-319-24574-4.

[62] Mahendra Khened, Varghese Alex Kollerathu, and Ganapathy Krishnamurthi. "Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers". In: *Medical Image Analysis* 51 (2019), pp. 21–45. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2018.10.004. URL: http://www.sciencedirect.com/science/article/pii/S136184151830848X.

[63] Mihaela Pop et al., eds. *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Springer International Publishing, 2018. DOI: 10.1007/978-3-319-75541-0. URL: https://doi.org/10.1007%2F978-3-319-75541-0.

[64] Christian F. Baumgartner, Lisa M. Koch, Marc Pollefeys, and Ender Konukoglu. *An Exploration of 2D and 3D Deep Learning Techniques for Cardiac MR Image Segmentation*. 2017. DOI: https://doi.org/10.48550/arXiv.1709.04496. arXiv: 1709.04496 [cs.CV].

[65] Chen Chen et al. "Deep Learning for Cardiac Image Segmentation: A Review". In: *Frontiers in Cardiovascular Medicine* 7 (2020). ISSN: 2297-055X. DOI: 10.3389/fcvm.2020.00025. URL: https://www.frontiersin.org/articles/10.3389/fcvm.2020.00025.

[66] Yao Zhang et al. "Semi-supervised Cardiac Image Segmentation via Label Propagation and Style Transfer". In: *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*. Springer, 2021, pp. 219–227. ISBN: 978-3-030-68107-4.

[67] Peter M. Full, Fabian Isensee, Paul F. Jäger, and Klaus Maier-Hein. "Studying Robustness of Semantic Segmentation Under Domain Shift in Cardiac MRI". In: *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*. Ed. by Esther Puyol Anton et al. Cham: Springer International Publishing, 2021, pp. 238–249. ISBN: 978-3-030-68107-4.

[68] Mario Parreño, Roberto Paredes, and Alberto Albiol. "Deidentifying MRI Data Domain by Iterative Backpropagation". In: *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*. Ed. by Esther Puyol Anton et al. Cham: Springer International Publishing, 2021, pp. 277–286. ISBN: 978-3-030-68107-4.

[69] Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca. "VoxelMorph: A Learning Framework for Deformable Medical Image Registration". In: *IEEE Transactions on Medical Imaging* 38.8 (Aug. 2019), pp. 1788–1800. ISSN: 1558-254X. DOI: 10.1109/tmi.2019.2897538. URL: http://dx.doi.org/10.1109/TMI.2019.2897538.

[70] Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, Adrian V. Dalca, and John Guttag. "An Unsupervised Learning Model for Deformable Medical Image Registration". In: *Computer Vision and Pattern Recognition (CVPR)* (June 2018). DOI: 10.1109/cvpr.2018.00964. URL: http://dx.doi.org/10.1109/CVPR.2018.00964.

[71] Tewodros Weldebirhan Arega, François Legrand, Stéphanie Bricq, and Fabrice Meriaudeau. "Using MRI-specific Data Augmentation to Enhance the Segmentation of Right Ventricle in Multi-disease, Multi-center and Multi-view Cardiac MRI". In: *Statistical Atlases and Computational Models of the Heart. Multi-Disease, Multi-View, and Multi-Center Right Ventricular Segmentation in Cardiac MRI Challenge*. Ed. by Esther Puyol Antón et al. Cham: Springer International Publishing, 2022, pp. 250–258. ISBN: 978-3-030-93722-5.

[72] Sandro Queirós. "Right Ventricular Segmentation in Multi-view Cardiac MRI Using a Unified U-net Model". In: *Statistical Atlases and Computational Models of the Heart. Multi-Disease, Multi-View, and Multi-Center Right Ventricular Segmentation in Cardiac MRI Challenge*. Ed. by Esther Puyol Antón et al. Cham: Springer International Publishing, 2022, pp. 287–295. ISBN: 978-3-030-93722-5.

[73] Mitchell J. Fulton, Christoffer R. Heckman, and Mark E. Rentschler. "Deformable Bayesian Convolutional Networks for Disease-Robust Cardiac MRI Segmentation". In: *Statistical Atlases and Computational Models of the Heart. Multi-Disease, Multi-View, and Multi-Center Right Ventricular Segmentation in Cardiac MRI Challenge*. Ed. by Esther Puyol Antón et al. Cham: Springer International Publishing, 2022, pp. 296–305. ISBN: 978-3-030-93722-5.

[74] Christoforos Galazis et al. "Tempera: Spatial Transformer Feature Pyramid Network for Cardiac MRI Segmentation". In: *Statistical Atlases and Computational Models of the Heart. Multi-Disease, Multi-View, and Multi-Center Right Ventricular Segmentation in Cardiac MRI Challenge*. Ed. by Esther Puyol Antón et al. Cham: Springer International Publishing, 2022, pp. 268–276. ISBN: 978-3-030-93722-5.

[75] Sana Jabbar, Syed Talha Bukhari, and Hassan Mohy-ud-Din. "Multi-view SA-LA Net: A Framework for Simultaneous Segmentation of RV on Multi-view Cardiac MR Images". In: *Statistical Atlases and Computational Models of the Heart. Multi-Disease, Multi-View, and Multi-Center Right Ventricular Segmentation in Cardiac MRI Challenge*. Ed. by Esther Puyol Antón et al. Cham: Springer International Publishing, 2022, pp. 277–286. ISBN: 978-3-030-93722-5.

[76] Fabian Isensee, Paul F Jaeger, Simon A A Kohl, Jens Petersen, and Klaus H Maier-Hein. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". In: *Nature Methods* 18.2 (2021), pp. 203–211. ISSN: 1548-7105. DOI: 10.1038/s41592-020-01008-z. URL: https://doi.org/10.1038/s41592-020-01008-z.

[77] Lei Li, Wangbin Ding, Liqin Huang, and Xiahai Zhuang. "Right Ventricular Segmentation from Short- and Long-Axis MRIs via Information Transition". In: *Statistical Atlases and Computational Models of the Heart. Multi-Disease, Multi-View, and Multi-Center Right Ventricular Segmentation in Cardiac MRI Challenge*. Ed. by Esther Puyol Antón

et al. Cham: Springer International Publishing, 2022, pp. 259–267. ISBN: 978-3-030-93722-5.

[78] Xiaowu Sun, Li-Hsin Cheng, and Rob J. van der Geest. "Right Ventricle Segmentation via Registration and Multi-input Modalities in Cardiac Magnetic Resonance Imaging from Multi-disease, Multi-view and Multi-center". In: *Statistical Atlases and Computational Models of the Heart. Multi-Disease, Multi-View, and Multi-Center Right Ventricular Segmentation in Cardiac MRI Challenge*. Ed. by Esther Puyol Antón et al. Cham: Springer International Publishing, 2022, pp. 241–249. ISBN: 978-3-030-93722-5.

[79] Amin Ranem, John Kalkhof, Caner Özer, Anirban Mukhopadhyay, and Ilkay Oksuz. *Detecting respiratory motion artefacts for cardiovascular MRIs to ensure high-quality segmentation*. 2022. DOI: 10.48550/ARXIV.2209.09678. URL: https://arxiv.org/abs/2209.09678.

[80] Michal K. Grzeszczyk, Szymon Płotka, and Arkadiusz Sitek. *Multi-task Swin Transformer for Motion Artifacts Classification and Cardiac Magnetic Resonance Image Segmentation*. 2022. DOI: 10.48550/ARXIV.2209.02470.

[81] Wanqin Ma, Huifeng Yao, Yiqun Lin, Jiarong Guo, and Xiaomeng Li. *Semi-Supervised Domain Generalization for Cardiac Magnetic Resonance Image Segmentation with High Quality Pseudo Labels*. 2022. DOI: 10.48550/ARXIV.2209.15451. URL: https://arxiv.org/abs/2209.15451.

[82] Yasmina Al Khalil et al. "On the usability of synthetic data for improving the robustness of deep learning-based segmentation of cardiac magnetic resonance images". In: *Medical Image Analysis* 84 (2023), p. 102688. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2022.102688.

[83] T. Karras, S. Laine, and T. Aila. "A Style-Based Generator Architecture for Generative Adversarial Networks". In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 43.12 (Dec. 2021), pp. 4217–4228. ISSN: 1939-3539. DOI: 10.1109/TPAMI.2020.2970919.

[84] Fanwei Kong and Shawn C. Shadden. "A Generalizable Deep-Learning Approach for Cardiac Magnetic Resonance Image Segmentation Using Image Augmentation and Attention U-Net". en. In: *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*. Ed. by Esther Puyol Anton et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021, pp. 287–296. ISBN: 978-3-030-68107-4. DOI: 10.1007/978-3-030-68107-4_29.

[85] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. "Unpaired image-to-image translation using cycle-consistent adversarial networks". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2223–2232.

[86] Tarun K. Mittal et al. "The Use and Efficacy of FFR-CT". In: *JACC: Cardiovascular Imaging* 16.8 (2023), pp. 1056–1065. DOI: 10.1016/j.jcmg.2023.02.005. URL: https://www.jacc.org/doi/abs/10.1016/j.jcmg.2023.02.005.

[87] Chaoyi Wu, Xiaoman Zhang, Ya Zhang, Yanfeng Wang, and Weidi Xie. *Towards Generalist Foundation Model for Radiology by Leveraging Web-scale 2D&3D Medical Data*. 2023. arXiv: 2308.02463 [cs.CV].

[88] Yigit B. Can et al. *Learning to Segment Medical Images with Scribble-Supervision Alone*. 2018. arXiv: 1807.04668 [cs.CV].

[89] Chen Qin et al. *Joint Learning of Motion Estimation and Segmentation for Cardiac MR Image Sequences*. 2018. arXiv: 1806.04066 [cs.CV].

[90] Adrian V. Dalca, Marianne Rakic, John Guttag, and Mert R. Sabuncu. *Learning Conditional Deformable Templates with Convolutional Networks*. 2019. arXiv: 1908.02738 [cs.CV].

[91] Adrian V. Dalca, Guha Balakrishnan, John Guttag, and Mert R. Sabuncu. "Unsupervised Learning for Fast Probabilistic Diffeomorphic Registration". In: *Lecture Notes in Computer Science* (2018), pp. 729–738. ISSN: 1611-3349. DOI: 10.1007/978-3-030-00928-1_82. URL: http://dx.doi.org/10.1007/978-3-030-00928-1_82.

[92] S. Kevin Zhou et al. "A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies With Progress Highlights, and Future Promises". In: *Proceedings of the IEEE* 109.5 (2021), pp. 820–838. DOI: 10.1109/JPROC.2021.3054390.

[93] Yabo Fu et al. "Deep learning in medical image registration: a review". In: *Physics in Medicine & Biology* 65.20 (Oct. 2020), 20TR01. DOI: 10.1088/1361-6560/ab843e. URL: https://dx.doi.org/10.1088/1361-6560/ab843e.

[94] Grant Haskins, Uwe Kruger, and Pingkun Yan. "Deep learning in medical image registration: a survey". In: *Machine Vision and Applications* 31.1 (Jan. 2020), p. 8. ISSN: 1432-1769. DOI: 10.1007/s00138-020-01060-x. URL: https://doi.org/10.1007/s00138-020-01060-x.

[95] Eric Arazo, Diego Ortego, Paul Albert, Noel O'Connor, and Kevin McGuinness. "Unsupervised label noise modeling and loss correction". In: *International Conference on Machine Learning*. PMLR. 2019, pp. 312–321.

[96] Peng Peng et al. "A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging". In: *Magnetic Resonance Materials in Physics, Biology and Medicine* 29 (2016), pp. 155–195.

[97] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. *Rethinking Atrous Convolution for Semantic Image Segmentation*. 2017. DOI: 10.48550/ARXIV.1706.05587. URL: https://arxiv.org/abs/1706.05587.

[98] Jieneng Chen et al. *TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation*. 2021. DOI: 10.48550/ARXIV.2102.04306. URL: https://arxiv.org/abs/2102.04306.

[99] Zaiwang Gu et al. "CE-Net: Context Encoder Network for 2D Medical Image Segmentation". In: *IEEE Transactions on Medical Imaging* 38.10 (2019), pp. 2281–2292. DOI: 10.1109/tmi.2019.2903562. URL: https://doi.org/10.1109%2Ftmi.2019.2903562.

[100] Fabian Isensee et al. *nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation*. 2018. DOI: https://doi.org/10.48550/arXiv.1809.10486. arXiv: 1809.10486 [cs.CV].

[101] Christoforos Galazis et al. "Tempera: Spatial Transformer Feature Pyramid Network for Cardiac MRI Segmentation". In: *Lecture Notes in Computer Science*. Springer International Publishing, 2022, pp. 268–276. DOI: 10.1007/978-3-030-93722-5_29. URL: https://doi.org/10.1007%5C%2F978-3-030-93722-5_29.

[102] Fernando Pérez-García, Rachel Sparks, and Sébastien Ourselin. "TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning". In: *Computer Methods and Programs in Biomedicine* 208 (2021), p. 106236. ISSN: 0169-2607. DOI: https://doi.org/10.1016/j.cmpb.2021.106236. URL: https://www.sciencedirect.com/science/article/pii/S0169260721003102.

[103] Sandro Queirós. "Right Ventricular Segmentation in Multi-view Cardiac MRI Using a Unified U-net Model". In: *Statistical Atlases and Computational Models of the Heart. Multi-Disease, Multi-View, and Multi-Center Right Ventricular Segmentation in Cardiac MRI Challenge*. Ed. by Esther Puyol Antón et al. Cham: Springer International Publishing, 2022, pp. 287–295. ISBN: 978-3-030-93722-5.

[104] L.G. Nyul, J.K. Udupa, and Xuan Zhang. "New variants of a method of MRI scale standardization". In: *IEEE Transactions on Medical Imaging* 19.2 (2000), pp. 143–150. DOI: 10.1109/42.836373.

[105] Richard Shaw, Carole Sudre, Sebastien Ourselin, and M. Jorge Cardoso. "MRI k-Space Motion Artefact Augmentation: Model Robustness and Task-Specific Uncertainty". In: *Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning*. Ed. by M. Jorge Cardoso et al. Vol. 102. Proceedings of Machine Learning Research. PMLR,

2019, pp. 427–436. URL: https://proceedings.mlr.press/v102/shaw19a.html.

[106] MRI Questions. *Random Ghosting*. URL: https://mriquestions.com/why-discrete-ghosts.html.

[107] Koen Van Leemput, Frederik Maes, Dirk Vandermeulen, and Paul Suetens. "Automated model-based tissue classification of MR images of the brain". In: *IEEE transactions on medical imaging* 18.10 (1999), pp. 897–908.

[108] Carole H. Sudre, M. Jorge Cardoso, and Sebastien Ourselin. "Longitudinal segmentation of age-related white matter hyperintensities". In: *Medical Image Analysis* 38 (2017), pp. 50–64. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2017.02.007. URL: https://www.sciencedirect.com/science/article/pii/S1361841517300257.

[109] Neuroimaging Informatics Technology Initiative. *NifTI*. URL: https://nifti.nimh.nih.gov/.

[110] Adam Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[111] Carlos Martín-Isla et al. "Deep Learning Segmentation of the Right Ventricle in Cardiac MRI: The M&Ms Challenge". In: *IEEE Journal of Biomedical and Health Informatics* 27.7 (2023), pp. 3302–3313. DOI: 10.1109/JBHI.2023.3267857.

[112] Brage Høyem Amundsen. "It Is All About Timing!" In: *JACC: Cardiovascular Imaging* 8.2 (2015), pp. 158–160. ISSN: 1936-878X. DOI: https://doi.org/10.1016/j.jcmg.2014.11.005.

[113] Elisabeth S. Lane et al. "Multibeat echocardiographic phase detection using deep neural networks". In: *Computers in Biology and Medicine* 133 (2021), p. 104373. ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2021.104373.

[114] Rupert Simpson et al. "6 Comparing echocardiography and cardiac magnetic resonance measures of ejection fraction: implications for HFMRF research". In: *Heart* 104.Suppl 5 (2018), A3–A3. ISSN: 1355-6037. DOI: 10.1136/heartjnl-2018-BCVI.6. eprint: https://heart.bmj.com/content/104/Suppl_5/A3.1.full.pdf. URL: https://heart.bmj.com/content/104/Suppl_5/A3.1.

[115] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory". In: *Neural Comput.* 9.8 (Sept. 1997), pp. 1735–1780. ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735.

[116] Ashish Vaswani et al. "Attention is All you Need". In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[117] Mingliang Zhai, Xuezhi Xiang, Ning Lv, and Xiangdong Kong. "Optical flow and scene flow estimation: A survey". In: *Pattern Recognition* 114 (2021), p. 107861. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2021.107861.

[118] C. Petitjean and J. N. Dacher. "A review of segmentation methods in short axis cardiac MR images". In: *Med. Image Anal.* 15 (2011). DOI: 10.1016/j.media.2010.12.004. URL: https://doi.org/10.1016/j.media.2010.12.004.

[119] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". In: *Nature Methods* 18 (2021). DOI: 10.1038/

s41592-020-01008-z. URL: https://doi.org/10.1038/s41592-020-01008-z.

[120] L.G. Nyul, J.K. Udupa, and Xuan Zhang. "New variants of a method of MRI scale standardization". In: *IEEE Transactions on Medical Imaging* 19.2 (2000), pp. 143–150. DOI: 10.1109/42.836373.

[121] Jia Deng et al. "ImageNet: A large-scale hierarchical image database". In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009, pp. 248–255. DOI: 10.1109/CVPR.2009.5206848.

[122] Tewodros Weldebirhan Arega, François Legrand, Stéphanie Bricq, and Fabrice Meriaudeau. "Using MRI-specific Data Augmentation to Enhance the Segmentation of Right Ventricle in Multi-disease, Multi-center and Multi-view Cardiac MRI". In: *Statistical Atlases and Computational Models of the Heart. Multi-Disease, Multi-View, and Multi-Center Right Ventricular Segmentation in Cardiac MRI Challenge*. Ed. by Esther Puyol Antón et al. Cham: Springer International Publishing, 2022, pp. 250–258. ISBN: 978-3-030-93722-5.

[123] Wenjia Bai et al. "A bi-ventricular cardiac atlas built from 1000+ high resolution MR images of healthy subjects and an analysis of shape and motion". In: *Medical Image Analysis* 26.1 (2015), pp. 133–145. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2015.08.009. URL: https://www.sciencedirect.com/science/article/pii/S1361841515001346.

[124] Taesung Park, Alexei A. Efros, Richard Zhang, and Jun-Yan Zhu. "Contrastive Learning for Unpaired Image-to-Image Translation". In: *European Conference on Computer Vision*. 2020.

[125] Abhirup Banerjee et al. "A completely automated pipeline for 3D reconstruction of human heart from 2D cine magnetic resonance slices". en. In: *Philos. Trans. A Math. Phys. Eng. Sci.* 379.2212 (Dec. 2021), p. 20200257.

[126] Krithika Iyer et al. "Statistical Shape Modeling of Biventricular Anatomy with Shared Boundaries". In: *Statistical Atlases and Computational Models of the Heart. Regular and CMRxMotion Challenge Papers*. Ed. by Oscar Camara et al. Cham: Springer Nature Switzerland, 2022, pp. 302–316. ISBN: 978-3-031-23443-9.

[127] Benjamin Villard, Vicente Grau, and Ernesto Zacur. "Surface Mesh Reconstruction from Cardiac MRI Contours". In: *Journal of Imaging* 4.1 (2018). ISSN: 2313-433X. DOI: 10.3390/jimaging4010016. URL: https://www.mdpi.com/2313-433X/4/1/16.

[128] Enric Moreu, Eric Arazo, Kevin McGuinness, and Noel E. O'Connor. "Joint one-sided synthetic unpaired image translation and segmentation for colorectal cancer prevention". In: *Expert Systems* (Sept. 2022), e13137. DOI: https://doi.org/10.1111/exsy.13137.

[129] Yin Xu, Yan Li, and Byeong-Seok Shin. "Medical image processing with contextual style transfer". In: *Human-centric Computing and Information Sciences* 10 (2020), pp. 1–16.