On Deep Learning in Cardiovascular Magnetic Resonance Imaging

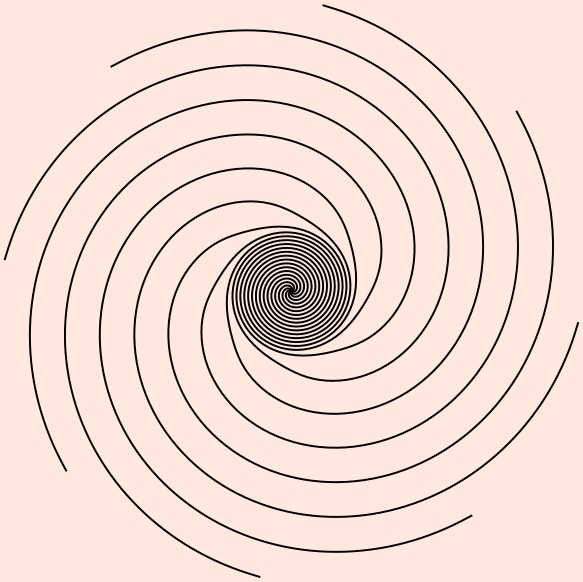
Åkesson, Julius

2025

Document Version:
Publisher's PDF, also known as Version of record

On Deep Learning in Cardiovascular Magnetic Resonance Imaging

JULIUS ÅKESSON
Department of Biomedical Engineering
Faculty of Engineering | LTH | Lund University



On Deep Learning in Cardiovascular Magnetic Resonance Imaging

On Deep Learning in Cardiovascular Magnetic Resonance Imaging

Julius Åkesson

LUND
UNIVERSITY

Thesis for the degree of Doctor of Philosophy
Thesis advisors: Dr. Einar Heiberg, Dr. Ellen Ostenfeld
Faculty opponent: Prof. Daniel B. Ennis

To be presented, with the permission of the Faculty of Engineering at Lund University, for public criticism in the Segerfalk lecture hall (Segerfalksalen) on Friday, 17th of October 2025 at 13:00.

Organization
LUND UNIVERSITY
Department of Biomedical Engineering
Faculty of Engineering

Document name
DOCTORAL DISSERTATION

Date of dissertation
Friday, 17th of October 2025 at 13:00

Sponsoring organization

Author(s)
Julius Åkesson

Title
On Deep Learning in Cardiovascular Magnetic Resonance Imaging

Abstract
Cardiovascular disease is the leading global cause of death. Cardiovascular magnetic resonance (CMR) imaging is the gold standard method for evaluating cardiac function and structure and is thus an important tool for diagnostics in cardiovascular disease. Recently, deep learning (DL) methods have achieved notable developments in various aspects of CMR. The aim of this thesis was to further assess the benefits of incorporating DL-based methods in CMR, specifically in the process of evaluating cardiac function. This was done in four studies that explored different applications and aspects of DL in CMR. In these, DL was applied alongside other new developments in image acquisition, reconstruction, and analysis. Assessments were conducted on the usefulness of the specific DL applications, and on how randomness in DL can affect the reporting of results.

In Study I, a DL-based right ventricular segmentation pipeline was developed, and the aspect of clinical time reduction was assessed, indicating that it could substantially reduce the time required to produce segmentations of clinically useful quality. Study II assessed how randomness affects the reliability of standard methods for comparing the performance of DL-based image segmentation models. Randomness caused statistically significant differences between models from the same learning algorithm, indicating that the standard methods cannot be used to reliably compare different learning algorithms. In Study III, a method was developed for synchronizing time-resolved ventricular cine images from 2D real-time exercise CMR. The results indicated that the method is feasible for deriving time-resolved measures of cardiac function and showed that DL could help accelerate the method. In Study IV, a rapid simultaneous multi-slice (SMS) bSSFP sequence and a DL-based image artifact suppression reconstruction method were developed for real-time CMR, achieving a reasonable agreement with standard CMR in volume-based clinical measures, but needing further improvements in image quality.

To conclude, DL is a powerful tool for accelerating existing workflows and facilitating new developments in CMR, but the process of developing DL methods involves randomness that can affect the reliability of comparisons. The usefulness of DL methods needs to be assessed on a case-by-case basis, but DL methods will likely continue to be a big part of future developments in CMR.

Key words
deep learning, cardiovascular magnetic resonance imaging, MRI, segmentation, reconstruction

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language
English

ISSN and key title

ISBN
978-91-8104-683-0 (print)
978-91-8104-684-7 (pdf)

Recipient's notes

Number of pages 171

Price

Security classification

Signature _____ Date _____

On Deep Learning in Cardiovascular Magnetic Resonance Imaging

Julius Åkesson

LUND
UNIVERSITY

Faculty Opponent

Prof. Daniel B. Ennis
Stanford University
Stanford, United States of America

Evaluation Committee

Prof. Peter Lundberg
Linköping University
Linköping, Sweden

Prof. Robin Strand
Uppsala University
Uppsala, Sweden

Dr. Anna Carlén
Linköping University
Linköping, Sweden

Cover: ”Vägs ände” by Julius Åkesson

Faculty of Engineering, Department of Biomedical Engineering
ISBN: 978-91-8104-683-0 (print)
ISBN: 978-91-8104-684-7 (pdf)
ISRN: LUTEDX/TEEM-1145-SE
Report number: 6/25

Written in LaTeX using a template by Berry Holl and Daniel Carrera.

Printed in Sweden by Tryckeriet i E-Huset, Lund University, Lund, Sweden, 2025

Dedicated to Helena and Mats

Contents

List of publications and manuscripts

This thesis is based on the following publications and manuscripts:

I **Deep learning can yield clinically useful right ventricular segmentations faster than fully manual analysis**

J. Åkesson, E. Ostenfeld, M. Carlsson, H. Arheden, E. Heiberg (2023)
Scientific Reports, vol. 13(1), 1216.
DOI: 10.1038/s41598-023-28348-y

II **Random effects during training: Implications for deep learning-based medical image segmentation**

J. Åkesson, J. Töger, E. Heiberg (2024)
Computers in Biology and Medicine, vol. 180, 108944.
DOI: 10.1016/j.compbiomed.2024.108944

III **Retrospectively synchronized time-resolved ventricular cine images from 2D real-time exercise cardiac magnetic resonance imaging**

J. Åkesson, J. Edlund, K. Steding-Ehrenborg, E. Heiberg (2025)
Clinical Physiology and Functional Imaging, vol. 45(5), e70027.
DOI: 10.1111/cpf.70027

IV **Accelerated real-time CMR at 1.5 T using blipped-CAIPI simultaneous multi-slice spiral bSSFP imaging and online deep artifact suppression reconstruction**

J. Åkesson, I. Dragonu, E. Heiberg, V. Muthurangu, J. Steeden
Manuscript.

Popular summary

Cardiovascular disease is the number one cause of death at a global scale. Magnetic resonance imaging (MRI) is the most accurate method for measuring the function and structure of the heart, and is thus important for diagnosing cardiovascular disease. Imaging of the heart with MRI can be done by recording multiple two-dimensional (2D) videos and piecing them together into a three-dimensional (3D) video of the heart, as if the heart was a loaf of bread and each 2D video was a slice of that loaf. From these 3D videos, the amount of blood in the heart can be determined by drawing lines around the edges of each chamber and measuring the volumes enclosed by these lines. This can tell us how much blood the heart is pumping at every heartbeat, indicating its function.

However, there are some persisting challenges with this method. Both the manual drawings and the MRI can take a long time. The MRI also usually requires patients to hold their breath and be completely still. This can be difficult for children that are not able to follow instructions. It can also be difficult for patients that are doing exercise during the MRI, in order to show problems with the heart that cannot be seen during rest. This thesis investigated the use of deep learning, a form of artificial intelligence, to tackle these issues, by using it for tasks that humans can and cannot do, in four different studies.

In the first study, a deep learning model was developed for automatically drawing the contours of one of the heart's chambers, the right ventricle. This method was able to speed up the process of getting drawings that could be used in a clinic by 87 % compared to manual drawings.

In the second study, we investigated the randomness that is used when deep learning models learn to draw contours. The learning is based on showing the model many examples of human-made drawings in a random order. We showed that by changing a single parameter that affects this order (and other random processes), the resulting models ended up with differences in performance, according to standard methods for comparing drawing performance. This showed that if we want to compare two *different* ways of learning to draw, we should not use these standard methods for comparing drawings, since they showed differences between models that used the same learning method, entirely based on randomness.

In the third study, a method was developed for creating 3D heart videos from 2D heart videos of patients that are doing exercise during the MRI, while breathing freely. In this

method, deep learning was used to detect specific time points in a heartbeat, to allow piecing together 2D videos that were taken at the same stage of breathing. Using this method, we could see that the anatomy was comparable to standard MRI during breath-holding and rest, and that deep learning could make the method faster.

In the fourth study, a method was developed for speeding up the MRI method for collecting 2D videos when patients are breathing freely. This was done by collecting two 2D videos *simultaneously* instead of just one at a time, which resulted in one video where two different videos are added on top of each other. A deep learning method was then used to separate this into two separate 2D videos. Using this method, the MRI scanning time could be reduced from minutes to seconds, but the quality of the videos was reduced compared to standard MRI.

To conclude, this thesis shows that the development of deep learning methods involves impactful randomness, but that the resulting methods can nonetheless be used to improve various aspects of heart MRI by speeding up tasks that humans can do, and by performing tasks that humans cannot do.

Preface

Studies I-III in this thesis were carried out at Clinical Physiology, Department of Clinical Sciences Lund, Lund University, Skåne University Hospital, Lund, Sweden and at the Department of Biomedical Engineering, Faculty of Engineering, Lund University, Lund, Sweden under the supervision of **Dr. Einar Heiberg** (main supervisor) and **Dr. Ellen Ostenfeld** (co-supervisor). Study IV was mostly carried out during a research visit at the Centre for Translational Cardiovascular Imaging, Institute of Cardiovascular Science, University College London, London, United Kingdom under the supervision of **Dr. Jennifer Steeden** and **Prof. Vivek Muthurangu**.

The work was supported by funding from the **Heart and Lung Foundation**, the **Medical Faculty at Lund University**, the **Knut and Alice Wallenberg Foundation**, **Region Skåne**, **Wallenberg Centre for Molecular Medicine**, the **Swedish Research Council**, the **Scandinavian Society of Clinical Physiology and Nuclear Medicine**, and **Svenska Sällskapet för Medicinsk Forskning**.

Acknowledgements

I'm endlessly grateful for all the help that I have received during my time as a PhD student in the **Lund Cardiac MR Group**. It has been a journey with many ups and downs, but at some point the ups started outweighing all those downs, and they have taken me down routes I could not foresee.

Thank you **Einar** for all your fantastic support, your guidance, and your ability to find solutions to all problems. Thank you **Ellen** for your great support, endless positivity, and for sharing your deep knowledge about writing. Thank you **Håkan** for creating and maintaining a great research environment, and for the guidance in personal development. Thank you **Jennifer** and **Vivek** for teaching me about sequence programming and image reconstruction, and for making my time in London a very developing experience. Thank you **Johannes** and **Anthony** for inspiring me to venture deeper into the technical aspects of MRI. Thank you **Erik** for being a great source of scientific inspiration and a networking mentor. Thank you **Katarina** for great help in physiology and exercise CMR. Thank you **Henrik** for being a great role model in leadership.

Thank you **Barbro**, **Shahnaz**, **Mariam**, **Per**, **Fredrik**, **Robert**, **Jonas**, **Morten**, **Henrik**, **David**, **Pia**, **Jane**, **Ashwin**, **Alessandro**, **Mattias**, **Helen**, **Karolina**, **Jelena**, **Anette**, **Ann-Helen**, **Christel**, **Reza**, **Johanna**, **Anna**, **Annmarie**, **Charlotte**, **Tony** and **Frederik** for contributing to a great research and work environment. Thank you to my fellow PhD students and research colleagues **Elsa**, **Axel**, **Kristian**, **Jonathan E**, **Martin**, **Tania**, **Theodor**, **Anders**, **Marjolein**, **Karin**, **Anna**, **Björn**, **Fanny**, **Anthony L**, **Daniel**, **Petter**, **Jonathan B**, **Jonas**, and **Romeissa** for all your help and the sharing of knowledge and struggle in our common journey into research. A special thanks to my fellow engineers for all our great times and travels. Thank you to everyone in the UCL lab for your help and companionship during my time at ZCR. Thank you to my parents **Helena** and **Mats** for always giving me unconditional support in everything I do, and to my sister **Karolina** for always being a great inspiration. Thank you **Sorrel** for everything.

List of abbreviations

2D Two-Dimensional

3D Three-Dimensional

ADC Analog-to-Digital Converter

AI Artificial Intelligence

BH Breath-Hold

BPM Beats Per Minute

bSSFP Balanced Steady-State Free Precession

CAIPI Controlled Aliasing in Parallel Imaging

CAIPIRINHA CAIPI Results In Higher Acceleration

CMR Cardiovascular Magnetic Resonance

CNN Convolutional Neural Network

CO Cardiac Output

CS Compressed Sensing

DFT Discrete Fourier Transform

DL Deep Learning

ECG Echocardiogram

ED End Diastole / End-Diastolic

EDV End-Diastolic Volume

EF Ejection Fraction

ERF Early Rapid Filling

ES End Systole / End-Systolic

ESV End-Systolic Volume

FA Flip Angle

FFT Fast Fourier Transform

FNN Feedforward Neural Network

FOV Field Of View

GB Gigabyte

GPU Graphics Processing Unit

GRAPPA Generalized Auto-calibrating Partially Parallel Acquisitions

HR Heart Rate

LA Left Atrium

LV Left Ventricle

LVM Left-Ventricular Mass

MDD Mechanical Diastolic Duration

MRI Magnetic Resonance Imaging

MS Mid Systole

MSE Mean Squared Error

NUFFT Non-Uniform Fast Fourier Transform

O1 Observer 1

O2 Observer 2

RA Right Atrium

ReLU Rectified Linear Unit

RF Radiofrequency

RT Real-Time

RV Right Ventricle

RVEDV Right Ventricular End-Diastolic Volume

RVESV Right Ventricular End-Systolic Volume

SMS Simultaneous Multi-Slice

SSIM Structural Similarity Index Measure

SV Stroke Volume

TE Echo Time

TR Repetition Time

Part I

Research context

Chapter 1

Background

1.1 Introduction

Cardiovascular disease is the leading global cause of death [1]. Thus, accurate diagnostics of cardiovascular disease are of great importance. Cardiovascular magnetic resonance (CMR) imaging is considered to be the gold standard method for evaluating cardiac function and structure [2], and is thus an important tool for diagnostics. It is also an active field of research in which new developments and improvements are constantly being published. Since the popularization of artificial intelligence (AI) methods, developments in the field of CMR have accelerated, and a lot of recent work has been focused on applications of these methods. One of the most widely applied types of AI for recent developments in the field of CMR is *deep learning*, which is the subcategory of AI that focuses on the use of neural networks. Deep learning has permeated many aspects of CMR, from reconstruction to image analysis, and new applications continue to be published.

This thesis presents a few applications of deep learning in CMR for measuring cardiac function. However, it focuses on the wider perspective of the use of deep learning, through applying it in a context together with other new developments in CMR, where it is used as a tool to reach certain goals that cannot be reached with other methods. The work in this thesis touches upon many of the different aspects of the CMR process, including image acquisition, image reconstruction, image processing, and image analysis. In addition to presenting the applicability of deep learning methods within this context, the studies focus on the aspect of *usefulness* of deep learning, in terms of its ability to speed up clinical practice or provide some added value in the given context. In the studies, new developments are presented in image segmentation, real-time exercise CMR, and real-time simultaneous multi-slice CMR. In one of the studies, a more thorough examination is also done on the effects of randomness when training deep learning methods for image segmentation, and how this can affect the reporting of results. This initial chapter provides an overview of the most important theoretical concepts that were touched upon in this doctoral thesis.

1.2 Magnetic resonance imaging (MRI)

1.2.1 Fundamentals of MRI

Magnetic resonance imaging (MRI) is a medical imaging modality based on physical phenomena occurring when atomic nuclei with a non-zero *spin* and *magnetic moment* interact with magnetic fields [3]. In the human body, the most commonly occurring atom with these properties is hydrogen (^1H), and in this thesis, only hydrogen imaging was performed. Nuclei with these properties are commonly referred to as *spins* in MRI literature [3], and this notation will be used throughout this chapter.

Spins can be thought of as very small bar magnets with their own small magnetic fields, meaning that they will also interact with other, external magnetic fields. If one observes a volumetric sample of spins without any external magnetic field, the net magnetization $\mathbf{M}$ (magnetic moment per unit volume) will be zero due to the randomly directed magnetic moments cancelling each other out (Fig. 1.1a). When interacting with an external magnetic field B_0, they tend to align with the direction of this, leading to a net magnetization vector $\mathbf{M}$ in the direction of B_0 (Fig. 1.1b) [3].

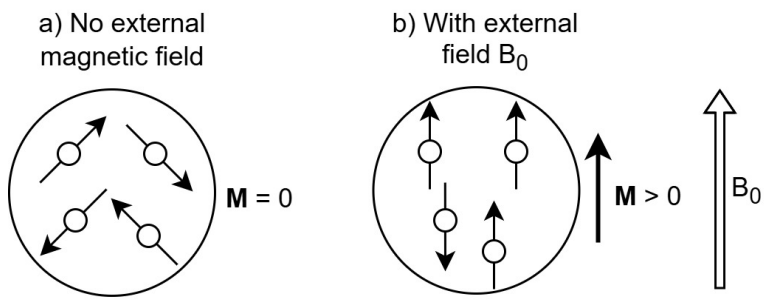


Figure 1.1: Simplified visualization of the behaviour of a volumetric sample of spins in the absence and presence of an external magnetic field B_0. Without the external field, the magnetic moments of the individual spins cancel out. Inside B_0, the spins align parallel and antiparallel to the direction of B_0, and there is a non-zero net magnetization $\mathbf{M}$.

$\mathbf{M}$ is aligned with B_0 when at thermal equilibrium. However, when $\mathbf{M}$ is not fully aligned with B_0, the torque imposed from B_0 gives rise to a precessional behaviour of $\mathbf{M}$ about B_0 [3], with a frequency known as the *Larmor frequency*, given as

$$\omega = \gamma B \text{ rad/s} \tag{1.1}$$

where γ is the *gyromagnetic ratio* (a constant specific to the nucleus) and B is the magnetic field strength. This is also the *resonant frequency* of $\mathbf{M}$. This means that if we apply a second magnetic field B_1 perpendicular to B_0 that oscillates at this particular resonant frequency, energy will be efficiently transmitted from B_1 to $\mathbf{M}$ resulting in a torque per-

pendicular to B_1 and $\mathbf{M}$ that rotates $\mathbf{M}$ away from the B_0 direction. Figure 1.2 illustrates this precessional behaviour at the presence of B_1. The B_1 field is usually applied via short *radiofrequency* (RF) pulses. The process of tipping the magnetization vector away from the B_0 field is known as *excitation*, and the angle to which the magnetization is flipped away from the direction of B_0 right after excitation is called the *flip angle*.

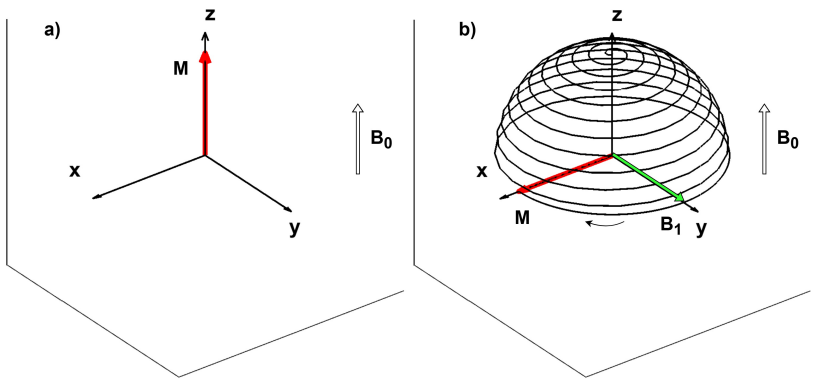


Figure 1.2: Illustration of the behaviour of the magnetization $\mathbf{M}$ at (a) thermal equilibrium, and (b) after applying a second magnetic field B_1, oscillating at the Larmor frequency. This shows that when B_1 is applied, the magnetization vector $\mathbf{M}$ moves into the xy-plane in an oscillatory manner.

After excitation, $\mathbf{M}$ will start to return to its thermal equilibrium via a process known as *relaxation*. This process is best described separately for the longitudinal ($\mathbf{M}_z$) and transversal ($\mathbf{M}_{xy}$) components of the magnetization. For the longitudinal magnetization component, the return to equilibrium can be described by

$$\mathbf{M}_z = \mathbf{M}_o + (\mathbf{M}_z(0) - \mathbf{M}_o)e^{-t/T_1} \tag{1.2}$$

where $\mathbf{M}_o$ is the equilibrium magnetization, $\mathbf{M}_z(0)$ is the longitudinal magnetization directly following the excitation (which would be zero if the flip angle was 90 degrees), t is time, and T_1 is a tissue-specific constant called the *the spin-lattice relaxation constant* that describes $\mathbf{M}_z$ during the return to equilibrium. For the transversal magnetization component, the return to equilibrium can be described by

$$\mathbf{M}_{xy} = \mathbf{M}_{xy}(0)e^{-t/T_2} \tag{1.3}$$

where $\mathbf{M}_{xy}(0)$ is the transversal magnetization directly following excitation (which will be equal to the equilibrium magnetization $\mathbf{M}_o$ if the flip angle is 90 degrees), t is time, and T_2 is a tissue-specific constant called the *the spin-spin relaxation constant* that describes $\mathbf{M}_{xy}$ during the return to equilibrium [3]. One reason for this decay is that the individual spins interact with each other, which leads to slight changes in their individual precessional

frequencies and causing them to *de-phase* [4]. The de-phasing of the individual oscillations of the spins means that the magnitude of the transversal net magnetization component will decay.

1.2.2 The MRI signal

After excitation, the net magnetization $\mathbf{M}$ will precess around B_0 at the Larmor frequency. In MRI, the source of the imaging signal is the time-variation or precession of the transverse magnetization component during relaxation. The signal measurement is carried out using receiver coils in which the time-variation of the transverse magnetization, and thus the magnetic flux through the receiver coil, induces an electromotive force through Faraday's law of induction [3]. The measured MRI signal is a voltage in the receiver coil that is sampled and digitized by using an analog-to-digital converter (ADC) [4].

Right after excitation, the transverse magnetization component precesses in the xy-plane and the de-phasing of spins causes it to decay according to Equation 1.3. The signal that can be measured during this process is called *free induction decay* (FID). However, this signal is rarely measured in itself in cardiovascular MRI [4]. Instead, the signal that is measured in most MRI sequences comes from different types of *echoes*. One way to form an echo is through using linear gradient magnetic fields (or *gradients*), that create linear spatial variations of the effective magnetic field, making the Larmor frequency of the spins vary linearly with the position along the direction of the applied gradient field according to Equation 1.1. By turning on a gradient right after RF excitation, the spins are de-phased by precessing at different frequencies based on their spatial position (if stationary). If a gradient of opposite sign to the first gradient is then played, the de-phasing can be reversed, bringing the spins back into phase [4], and this is when the echo signal occurs [5]. Echoes can also be formed in with RF pulses, but this will not be described in this thesis.

1.2.3 Spatial encoding and RF pulses

In order to make the MR signal useful for imaging, spatial dependency needs to be encoded in the measured signal. For this, gradients are employed. Figure 1.3 illustrates the use of gradients for spatial encoding. Gradients can be turned on during a signal readout (when the ADC is on) to create spatially dependent frequency variations in the measured signal. This is known as *frequency encoding*. Additionally, gradients can be turned on *before* a readout to create spatially dependent phase variations in the measured signal. This is known as *phase encoding*. Gradients can also be applied during excitation to make the RF pulse (the oscillating B_1 field) excite only a specific slice of spins. This is known as *slice selection*. When played, the slice selection gradient makes the resonant frequencies of the spins vary linearly along the z-axis. RF pulses usually excite only a specific band of frequencies, called the transmit bandwidth [4]. A common shape of the RF pulse is variants of the *sinc* $(\sin(x)/x)$ function that have been truncated to be made finite and further processed to create a better

shape of the excited slice, or *slice profile* [5]. The sinc pulse produces a slice profile that is approximately rectangular, hence its widespread use [4].

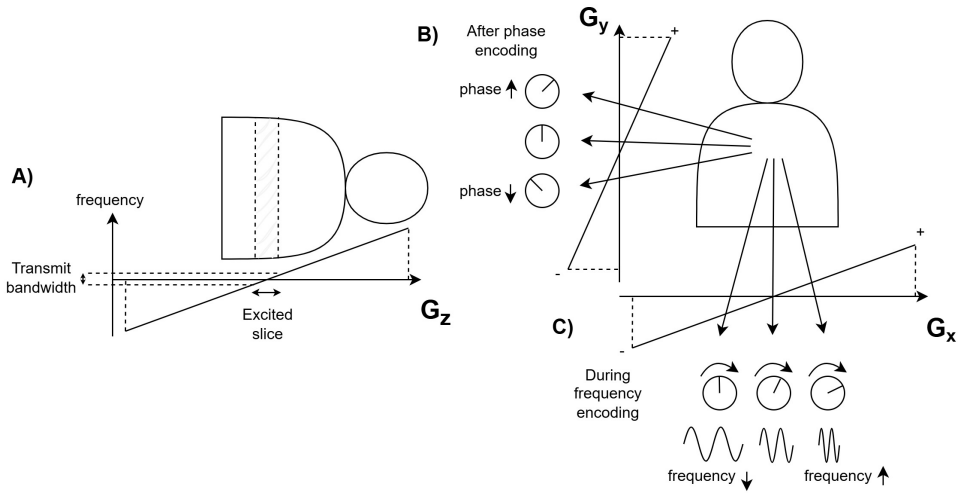


Figure 1.3: Illustration of the use of linear gradient fields for A) slice selection (G_z-axis), B) phase encoding (G_y-axis), and C) frequency encoding (G_x-axis). The slice selection gradient makes the RF pulse excite spins within the transmit bandwidth of frequencies, corresponding to a section along the G_z axis. The phase encoding gradient, after being turned on and off, imparts different phases to spins at different spatial positions along G_y. The frequency encoding gradient imparts different frequencies to spins at different spatial positions along G_x when turned on.

1.2.4 K-space

The MR signal sampled using the ADC does not provide an image directly. The signal is obtained in the spatial frequency domain, which is often referred to as *k-space*. Signals in k-space can be transformed to the image domain via the inverse *Fourier transform*. The Fourier transform is a mathematical operation that can be applied to a signal to retrieve its spectral content, i.e. what frequencies it contains. For example, given an audio signal of a chord from a musical instrument, the Fourier transform can be used to find out what musical notes were played in the chord, and how loudly each note was played [5]. In an MR signal measured into 2D k-space, each data point describes the influence of a particular *spatial* frequency, which in image space corresponds a two-dimensional sinusoidal wave at a particular frequency, direction, and amplitude. This is illustrated in Figure 1.4, where a single non-zero point in k-space corresponds to a two-dimensional sinusoidal wave in image space. Any object in image space can be described as a superposition of such two-dimensional sinusoidal wave components, and thus by spatial frequencies in k-space.

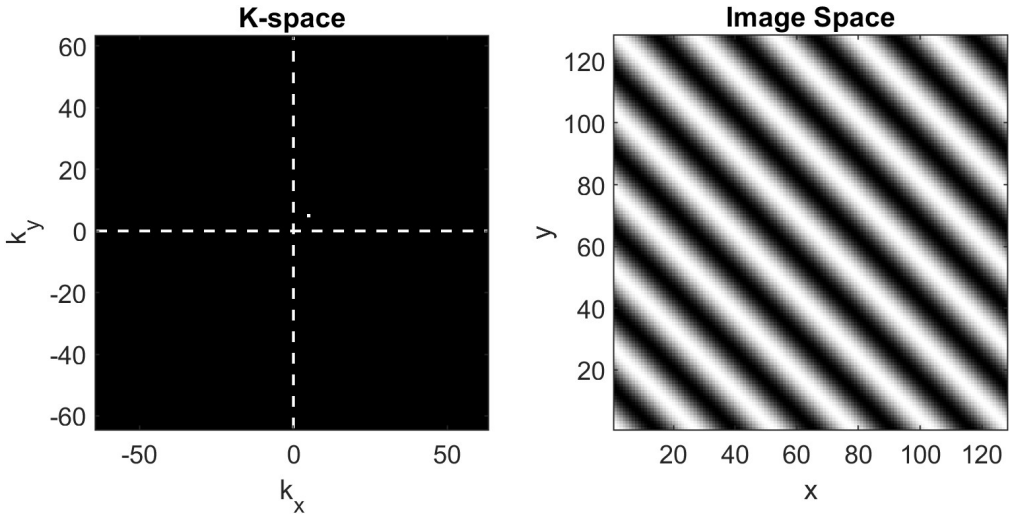


Figure 1.4: Left: K-space (the spatial frequency domain) with a single nonzero point at $(k_x, k_y) = (5, 5)$. Right: The real component of the corresponding signal in the image domain, displaying a two dimensional sinusoidal wave.

Because the sampled MR signal is discrete, the Fourier transform that needs to be applied to transform the MR signal is the discrete two-dimensional Fourier transform (DFT). More specifically, an efficient algorithm for computing the DFT, called the two-dimensional *fast Fourier transform* (FFT), is often used [5]. This method can be used in the case when k-space is uniformly fully sampled, i.e. when the values in k-space are evenly distributed across the k-space grid. This can be achieved through filling k-space rectilinearly, via one horizontal line for each phase encoding step, also known as *Cartesian* imaging. However, the collected MR signal is often deliberately under-sampled at readout to reduce acquisition time. In this case, not all points in the k-space grid are filled. In many cases of under-sampled imaging, k-space is also traversed using a trajectory with which the measured k-space points are not evenly or uniformly sampled in k-space, also known as *non-Cartesian imaging*. An example of a Cartesian and a non-Cartesian trajectory can be seen in Figure 1.5. To allow the computation of the inverse FFT on k-space data filled with a non-Cartesian trajectory, it requires for the FFT to be generalized to a non-uniform sampling grid. This is called a non-uniform fast Fourier transform (NUFFT), and is often based on algorithms for gridding the data onto a uniform Cartesian grid via interpolation to allow performing a conventional FFT [6, 7].

Spiral readouts, such as seen in Figure 1.5, can be used to cover a large part of k-space at each excitation, thereby being more efficient in their coverage of k-space than with a rectilinear filling [4]. To achieve a spiral readout in 2D imaging using gradients, the gradients are played out *simultaneously* during the readout along the x-and y directions,

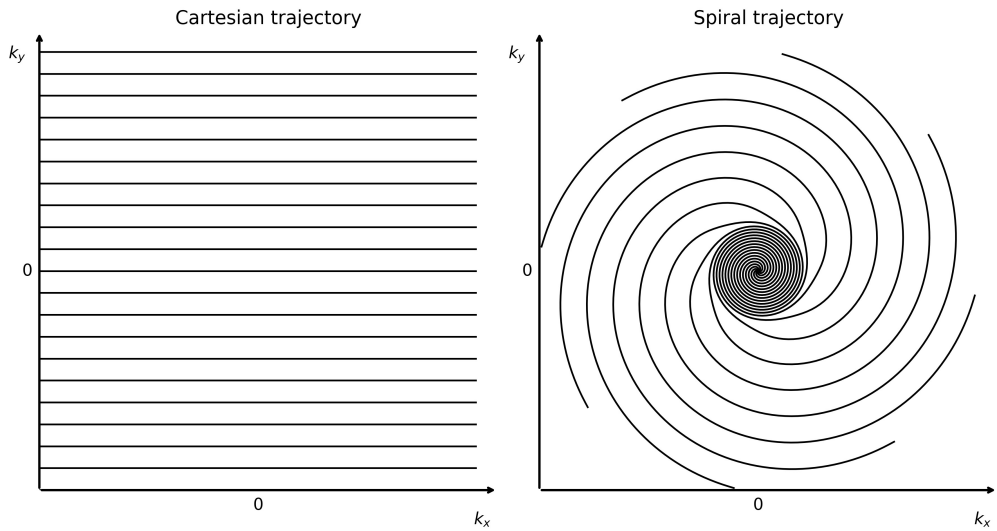


Figure 1.5: Left: A Cartesian k-space trajectory, covering k-space in a rectilinear manner. Right: A non-Cartesian, variable density interleaved multi-shot spiral k-space trajectory, where the density of the spirals are significantly higher in the center of k-space. The trajectory has 8 interleaves, meaning that it consists of 8 separate spiral arms acquired at separate excitations.

varying sinusoidally [5] (see Fig. 1.7). Usually, spiral trajectories begin at the center of k-space, and after the full length of each spiral *interleave* (or spiral arm) has been reached, a gradient is run to rewind the k-space trajectory back to the origin of k-space, to allow the next interleave to start from the origin [5].

1.2.5 Advanced image reconstruction techniques

When k-space is under-sampled, directly applying the FFT or NUFFT algorithms leads to increased amounts of *aliasing* artifacts in the resulting images. These are artifacts that occur when signals above the *Nyquist* frequency ($f_N = \frac{1}{2} f_s$, where f_s is the sampling frequency) are misrepresented as low frequency signals. This misrepresentation increases with higher under-sampling. To avoid under-sampling artifacts, these methods need to be extended with more advanced techniques in order to handle the missing k-space data. There is a wide range of techniques for reconstructing under-sampled MR images, and the central aim of most new reconstruction methods is to facilitate faster MR acquisition by allowing a heavier under-sampling of k-space without a substantial reduction in the spatial resolution. The level of under-sampling is denoted by the reduction factor R. For example, $R = 2$ if every other line is skipped in Cartesian sampling.

A central category of reconstruction techniques for under-sampled imaging is *parallel imaging* techniques. Parallel imaging utilizes receiver coil arrays with multiple coil elements (also known as *phased arrays*), where each element is used to measure the MR signal separately. Each coil element has a different sensitivity to the signal from the imaged object due to their different spatial positions. Maps of these sensitivities are leveraged to aid the reconstruction, and the separate coil measurements are combined in various ways (depending on the specific method) to fill in the gaps in k-space and to create the final image [4]. One widely used parallel imaging technique is *GRAPPA* (Generalized Auto-calibrating Partially Parallel Acquisitions). In this method, lines of k-space are skipped according to some R, while some additional k-space lines close to the center of k-space are also collected. These are called *auto-calibration lines*. The purpose of the auto-calibration lines is to calculate weights for determining how nearby k-space lines from all coil elements can be combined together to estimate each missing k-space line in each coil [8]. When the missing lines have been filled in, coil images are reconstructed separately and then combined together in image space [8].

A category of reconstruction methods that is widely used in accelerated CMR (and can be combined with parallel imaging techniques) is *iterative* reconstruction methods. In iterative reconstruction methods, the MR image reconstruction problem is generally posed as the problem of finding an image series $\mathbf{x}$ that aligns with the acquired k-space signal $\mathbf{s}$ given some prior information or assumptions. To solve this problem, iterative reconstruction methods typically seek to minimize a regularized least squares cost function that can be denoted as

$$\hat{\mathbf{x}} = \text{argmin}_x ||\mathbf{s} - \mathbf{E}\mathbf{x}||_2^2 + \lambda R(\mathbf{x}) \tag{1.4}$$

where $\mathbf{E}$ is the MRI *encoding operator* or *encoding matrix* that includes the Fourier transform and other properties of the acquisition process (e.g. coil sensitivity information), $R(\mathbf{x})$ is the *regularization* term that enforces prior information or assumptions of the image data, and λ is a tunable parameter that determines the weighting between the first and second term [9]. The term $||\mathbf{s}-\mathbf{E}\mathbf{x}||_2^2$ is commonly called the *data consistency term*, since it enforces consistency between $\mathbf{s}$ and $\mathbf{x}$ [9].

One widely used iterative reconstruction method in contemporary CMR is *compressed sensing* (CS). In CS, the regularization term $R(\mathbf{x})$ includes a *sparsity* transform of $\mathbf{x}$ (e.g. total variation [10]) which enforces the optimization to find solutions that have a sparse representation in some domain [10]. Usually, this term also employs an L_1-norm. The method works under the assumption that k-space is randomly under-sampled [11]. To solve the optimization problem in Equation 1.4, iterative optimization methods are commonly used, such as the nonlinear conjugate gradient method [10]. Another method that can be used is the L-BFGS algorithm [12].

Iterative methods, in particular CS, have achieved high-quality reconstructions at high under-sampling factors in multiple areas of CMR imaging [11, 13]. However, drawbacks of these methods include long reconstruction times and the need to tune the regularization parameters (λ) differently for different applications [11]. The long reconstruction times limit their use for *online* reconstruction. This is when images are directly reconstructed during a scan and visualized at the scanner.

Recently, deep learning has been widely applied for image reconstruction in under-sampled MR imaging [13]. Image-based *de-aliasing*, or *deep artifact suppression* is one type of deep learning-based reconstruction. In this, an under-sampled k-space is transformed to image space using e.g a NUFFT, producing an aliased image. Then, a neural network is applied to suppress aliasing artifacts through an image-to-image mapping. Thus, in this case, the deep learning problem is a mapping of an aliased image to a non-aliased image, or an artifact-free image. Training data for training this type of mapping can be created through retrospectively under-sampling a fully sampled k-space to synthesize the under-sampling process during acquisition. This type of method was employed in Study IV of this thesis, and was implemented for online reconstruction using the *Gadgetron* framework [14].

A wide variety of other deep learning-based reconstruction methods exist, and the field is under constant development. Some methods are iterative, and employ cost functions similar to Equation 1.4. For example, *unrolled networks* apply neural networks to learn the regularization $R(\mathbf{x})$ instead of using an explicit definition of it, as for example in CS [11]. However, these methods were not employed in this thesis.

1.2.6 Cardiovascular MR (CMR) imaging

The specific type of MR imaging that this thesis focuses on is *cardiovascular* MR (CMR) imaging. CMR can be used to acquire time-resolved image volumes of the heart, allowing one to make volumetric measurements of the various anatomical structures of the heart throughout the cardiac cycle. To do this, it is necessary to employ techniques that produce a spatial resolution that allows resolving the anatomical structures of interest and a temporal resolution that allows resolving the temporal events of interest.

The anatomical and temporal information that one wishes to resolve varies depending on the specific application of CMR, and the techniques used to achieve these also vary. Importantly, there is a trade off between the time that is spent on collecting the MR signal (or filling up k-space) and the temporal and spatial resolution that can be achieved. To achieve a high spatial and temporal resolution, long acquisition times are required. Longer acquisition times make clinical CMR examinations longer and more tedious for patients, and can even make it unfeasible to carry out the acquisition for some patients. Below are descriptions of the various CMR techniques used in this thesis.

1.2.7 Cine CMR imaging

Methods for collecting k-space data to construct CMR timeframes can be divided in to two main categories. The first category utilizes the periodicity of cardiac motion and fills up k-space with signal acquired over multiple heartbeats. This is known as *gated* cine imaging, and it is the method that provides the highest spatial and temporal resolution. This is also the method most commonly employed clinically. The second category does no sharing of k-space data across heartbeats, and instead uses continuously collected MR signal to reconstruct each timeframe. This is known as *real-time* imaging.

In gated cine imaging, image acquisition is usually carried out simultaneously as an echocardiogram (ECG) signal is being measured. The ECG signal measures the electrical activity of the heart and can indicate the different phases of the cardiac cycle (see section 1.4.3). This can be used to sort the acquired k-space data into a discrete set of cardiac phases [5]. The ECG can also be used for *triggering* the acquisition of data, meaning that the start of certain events of the MR pulse sequence is connected to events in the ECG signal [5]. In *retrospective* gating, the acquisition runs continuously while the ECG triggers the updating of gradients, and the sorting of data into phases is done after the acquisition. In *prospective* gating, the acquisition itself only happens during a time window triggered by events in the ECG. In gated cine imaging using Cartesian sampling of k-space, k-space lines are acquired for each cardiac phase during each heartbeat, and imaging is continued until enough k-space lines have been collected in order to reconstruct all timeframes [5]. When more than one k-space line (a *segment* of k-space) is collected for each cardiac phase for each heartbeat, it is called *segmented* gated cine imaging. This technique is illustrated in Figure 1.6a.

Gated cine imaging is based one the assumption that anatomical information and mo-

tion will be similar across all imaged heartbeats. Because of this, cine imaging is usually carried out during breath-holding. Without breath-holding, there is risk for breathing-induced displacement of the heart between heartbeats, which can lead to artifacts. Furthermore, the ECG signal is not always reliable enough to carry out gated cine imaging. This can for example be the case when a patient suffers from *arrhythmia* [15] or when a patient is carrying out exercise [16]. One way to carry out exercise during CMR is to use a supine *bicycle ergometer*, with which the patient is pedalling with their feet while laying in a supine position. In these cases where breath-holding is not feasible and the ECG is unreliable, real-time imaging can be used.

Figure 1.6b shows a schematic overview of real-time imaging. In contrast to segmented cine imaging, real-time imaging continuously fills up k-space enough to reconstruct each timeframe, without sharing k-space data across heartbeats. In practice, this makes it more difficult to achieve high spatial and temporal resolutions, and increases the need for more efficient k-space sampling. Higher spatial and temporal resolutions in real-time CMR can be attained through efficient non-Cartesian to sampling of k-space, such as spiral or radial trajectories [15].

In real-time CMR, the ECG signal is not necessary for image formation. However, ECG signals or similar signals such as *pulse oximeter* signals can still be used for triggering the acquisition. In contrast to the ECG that measures the electrical activity of the heart using electrodes that are attached to the skin of the patient, a pulse oximeter is instead attached to a peripheral body part, such as a finger. To get a signal that indicates the cardiac periodicity, it transmits light into the body part and then determines the attenuation of this light [5] as it is transmitted through the body part [17]. Continuous measurement of this property provides a signal that can indicate cardiac phase, similarly to an ECG.

1.2.8 Balanced steady-state free precession (bSSFP) sequences

In both gated cine and real-time CMR, one of the most widely used types of MRI pulse sequences are *balanced steady-state free precession* (bSSFP) sequences. In this thesis, segmented cine bSFFP imaging was used for all CMR imaging in Study I, II and III and the real-time CMR pulse sequence implemented in Study IV was also based on a bSSFP sequence.

The most fundamental aspect of bSSFP sequences is the application of a train of RF pulses that, after a certain number of repetitions, leads to a steady state of the magnetization [18]. The steady state entails that the magnetization $\mathbf{M}$ is in a dynamic equilibrium, meaning that it has the same magnitude at corresponding temporal positions across consecutive repetition times (TR) measured between the centres of two consecutive RF pulses [5]. The conditions required to reach a steady state magnetization include a constant TR, a constant amount of de-phasing in each TR, and a constant flip angle α [18], and it usually takes around 5 periods of T_1 to reach this steady state [19, 5].

The *balanced* aspect of bSSFP sequences comes from the fact that the time integrals of the gradients over every TR, also known as the *zeroth order gradient moments*, are zero. This

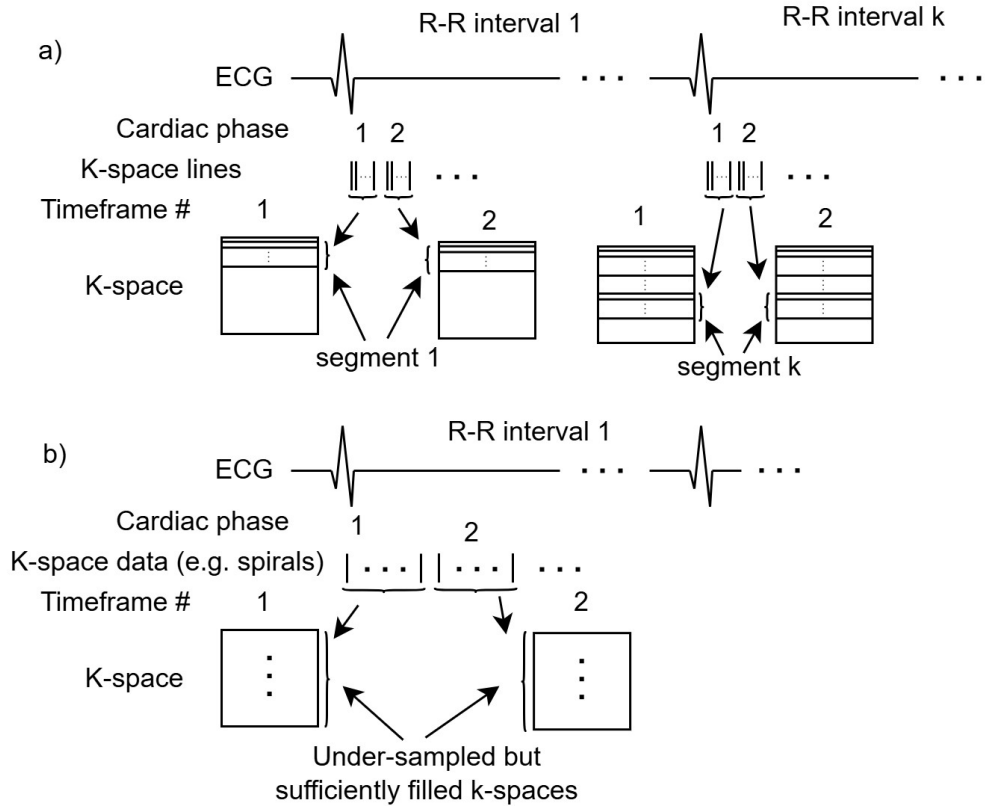


Figure 1.6: Schematic overview of how a) segmented gated cine imaging and b) real-time imaging can be carried out to fill k-space. In segmented gated cine imaging, a segment consisting of a few unique k-space lines is collected for each timeframe during each R-R interval, filling up k-space step-by-step. In real-time imaging, the k-space for every timeframe is filled up continuously within a single R-R interval, without the need for an ECG signal for gating.

also means that the net de-phasing of spins caused by gradients is zero over each TR [18]. For example, this is given along the gradient axis G_x as:

$$\int_0^{TR} G_x(t)\,dt = 0 \tag{1.5}$$

This property can be achieved by applying compensatory gradients (*rewinder* gradients) for each gradient that is played during a TR. Figure 1.7 shows a single TR of a bSSFP sequence with spiral readout gradients, and with the zeroth order gradient moments M_{0x}, M_{0y}, and M_{0z} plotted for their respective axes. Here, it can be seen that the zeroth order gradient moment is nulled across all axes at the end of the TR, as a result of the compensatory gradients applied to all gradient axes. The rewinder gradients along the G_x and G_y axes make the k-space trajectory return to the origin of k-space. In bSSFP imaging, the imaging signal in the steady state is not formed by just a single echo or a single FID signal, but rather a mix of multiple types of signals from multiple magnetization states as a result of the train of RF-pulses [19]. For CMR, bSSFP sequences are suitable because they can achieve high temporal resolution and be robust to motion [19]. However, the robustness to flow may depend on the used k-space trajectory, and some trajectories may require for additional compensatory lobes to also null the first order gradient moment [20].

1.2.9 Simultaneous multi-slice (SMS) imaging

Simultaneous multi-slice (SMS) imaging is an MR imaging technique that allows acquiring multiple ($N > 1$) image slices simultaneously, and thereby accelerating the image acquisition by a factor N [21]. An important aspect of this technique is the use of *SMS RF pulses*, that allow simultaneous excitation of slices at multiple slice positions. A single-band RF pulse can be described as

$$RF(t) = A(t) \cdot P(t) \tag{1.6}$$

where $A(t)$ is the waveform of the RF pulse (e.g. a *sinc* function) and $P(t) = e^{i(\gamma G z t + \phi)}$, where γ is the gyromagnetic ratio, G is the amplitude of the slice select gradient, z is the slice position, and ϕ is the phase directly following excitation [21]. Given this, an SMS RF pulse exciting N different slices (at N different slice positions z_n) using the same slice select gradient can be constructed as the following sum of N different RF pulses [21, 22]

$$RF_{SMS}(t) = A(t) \cdot \sum_N e^{i(\gamma G z_n t + \phi_n)}. \tag{1.7}$$

In SMS imaging, the acquired imaging signal is a superposition of the imaging signals from the N excited slices, and the k-space is thus a superposition of the N different k-spaces. Another key aspect of SMS imaging is the separation of the N superimposed slices into N different slices. To facilitate this, additional spatial encoding of the N SMS slices is often used.

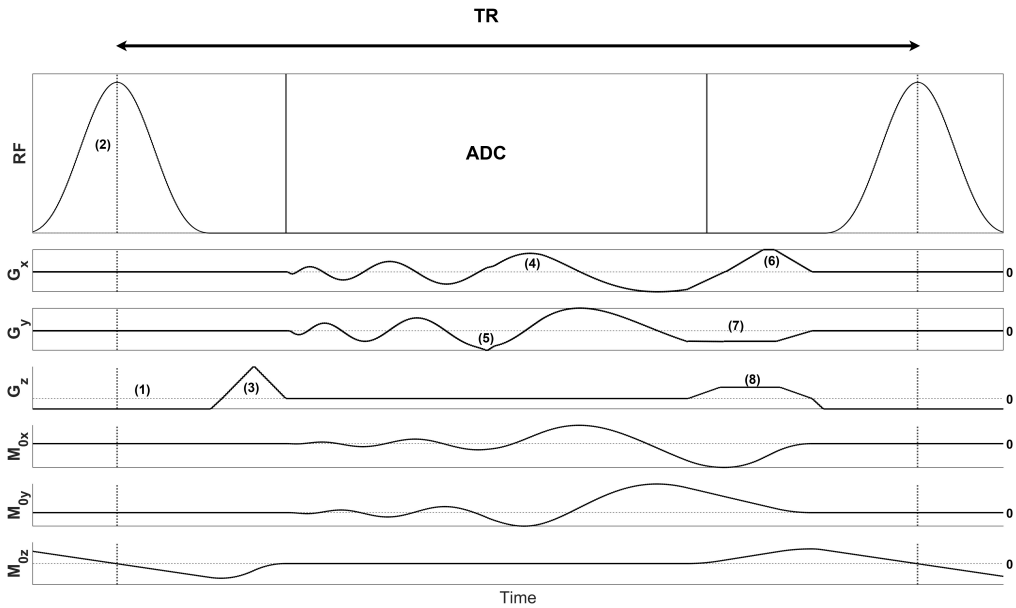


Figure 1.7: Sequence diagram showing a bSSFP sequence with spiral readout gradients. The diagram also shows the start and stop time of the analog-to-digital converter (ADC) and the zeroth order moments calculated across time for the three gradient axes M_{0x}, M_{0y}, and M_{0z}. The numbered events are (1) slice selection gradient; (2) radiofrequency (RF) pulse; (3) rewinder for the slice selection gradient; (4) spiral readout gradient (G_x); (5) spiral readout gradient (G_y); (6) rewinder gradient for G_x; (7) rewinder gradient for G_y; (8) prewinder for the slice selection gradient. Note that the zeroth order gradient moment is zero for all axes at the end of the TR.

Similarly to how in-plane phase encoding is carried out using gradients, gradients can impart relative phase differences between the SMS slices along the slice direction. However, in SMS imaging, the aim of the SMS phase encoding is generally to achieve an in-plane shift between the acquired slices, to decrease the spatial overlap [21]. A linear phase shift in k-space corresponds to a shift in the spatial position in image space according to the Fourier *shift theorem* [5]. Through this, it can be shown that specific shifts in the field of view (FOV) can be achieved through the application of specific phase modulation patterns [23]. This is known as CAIPIRINHA (Controlled Aliasing In Parallel Imaging Results In Higher Acceleration) [23], and can be used in order to achieve better separation of slices during reconstruction.

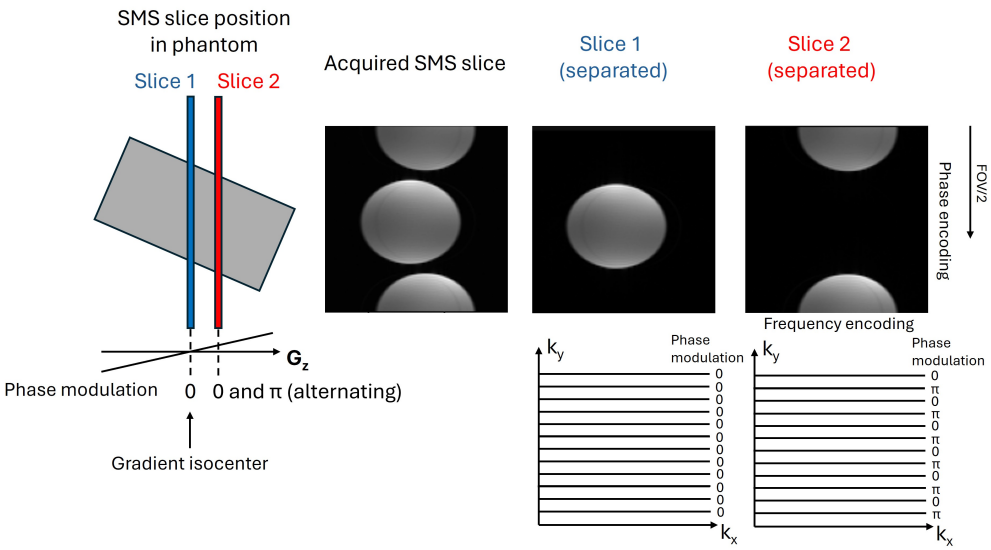


Figure 1.8: SMS acquisition of a cylindrical bottle phantom with a blipped-CAIPI phase pattern alternating between 0 and π for the phase encoding steps. The acquired SMS slice contains the superposition of the signals from the two slices. Due to the applied gradient phase modulation (gradient blips), slice 2 is shifted by FOV/2 in the phase encoding direction at acquisition. Slice 1 experiences no phase modulation due to being positioned in the gradient isocenter.

In the original CAIPIRINHA publication, phase modulation patterns were achieved using RF pulse phase cycling [23]. Another way of introducing these phase modulation patterns is through the application of short gradient pulses known as *gradient blips*, and this method is known as *blipped-CAIPI* [24]. Gradient blips applied along the G_z-direction (the SMS slice direction) introduce a phase difference between spins along this direction. When integrated into bSSFP sequences, the blips are played out before the readout in each TR, with an additional gradient blip of opposite polarity being played out after the readout

for balancing purposes (nulling the zeroth order moment) [25]. An implementation of this, where the blips are integrated into the rewinder and prewinder gradients of a bSSFP sequence, can be seen in Figure 3.2 in the *Materials and methods* section.

Figure 1.8 shows an example of an SMS acquisition where a shift of FOV/2 in the phase encoding direction is achieved through the application of a phase modulation pattern alternating between 0 and π for each phase encoding step using gradient blips. Here, the first slice is located in the gradient isocenter and the second slice is located such that it accrues a phase of π when the gradient is turned on. A specific phase shift of ϕ can be achieved through setting the properties of the blip gradient lobe based on the separation distance z_{gap} between the two slices according to

$$A_{blip} = \phi/(\gamma z_{gap}) \tag{1.8}$$

where $A_{blip} = \int_{blip} G_z dt$ is the area of the blip (the zeroth order gradient moment) and γ is the gyromagnetic ratio [24].

In the case of non-Cartesian (e.g. spiral) imaging, the same type of phase modulation pattern as in Figure 1.8 can be applied to the individual spiral interleaves in an alternating fashion. However, in non-Cartesian imaging, with trajectories that cross the centre of k-space multiple times (e.g. radial or spiral), this does not produce the effect of a shift in the FOV. Instead, the resulting effect is a constructive or destructive interference of the signal coming from the different excited slices [26, 27].

Figure 1.9 shows an example of this constructive and destructive interference using phase modulation for a spiral trajectory where the 16 spiral interleaves are phase modulated with an alternating phase modulation pattern of $-\pi/2$ and $\pi/2$. Here, the trajectory can be seen divided into its two components with different phase modulation, namely the even- and odd-numbered spiral interleaves. When these individual trajectory components are reconstructed separately using a NUFFT, the resulting images contain signal as well as aliasing artifacts from under-sampling. When instead reconstructing the full trajectory, the resulting image only consists of residual aliasing artifacts, with no visible signal. This is due to destructive signal interference. Note that the phase modulation pattern of $-\pi/2$ and $\pi/2$ provides the same effect as a phase modulation pattern of 0 and π, since the phase difference is π between interleaves. This type of signal cancellation can, similarly to the linear shifting, improve the quality of images reconstructed using parallel imaging [26, 27].

The effects of phase modulation from gradient blips can be reversed by applying the complex conjugate of the phase modulation pattern to the k-space. Through doing this, the phase can be de-modulated such that all spiral interleaves are given the same phase, and the signal interference between interleaves is then made *constructive* instead of destructive. In the case of SMS imaging, the phase modulation pattern is typically configured such that the phase modulation that is needed to achieve constructive interference for one slice corresponds to the phase modulation needed to achieve destructive interference in the other

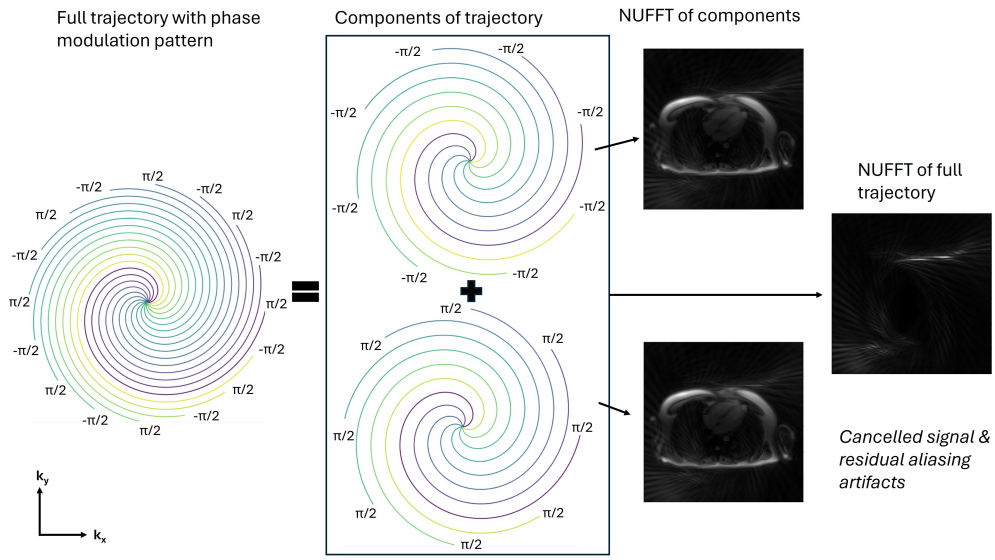


Figure 1.9: Example of signal cancellation by applying a phase modulation pattern of $-\pi/2$ and $\pi/2$ to a spiral trajectory. This shows the two individual trajectory components (even and odd spiral interleaves) between which there is a phase difference of π. The two components are reconstructed separately using a NUFFT, and the resulting images show both signal and aliasing artifacts from under-sampling. When the full trajectory is reconstructed using a NUFFT, the resulting image has no visible signal, but carries residual aliasing artifacts.

slice. Figure 1.10 shows an example of how this can be achieved when two simultaneously excited slices are evenly spaced around the G_z gradient isocenter and a phase modulation pattern of $-\pi/2$ and $\pi/2$ is applied, imparting an absolute phase difference of $\pi/2$ between slices that are z_{gap} apart. In this case, even spiral interleaves accrue the phase $\pi/4$ for slice 1 and $-\pi/4$ for slice 2, whereas odd spiral interleaves accrue the phase $-\pi/4$ for slice 1 and $\pi/4$ for slice 2. It can be seen that in this case, the phase de-modulation pattern required to achieve constructive signal interference for slice 1 (the same phase for odd and even spiral interleaves) produces destructive signal interference (a phase difference of π between odd and even spiral interleaves) for slice 2, and vice versa. This method was used in Study IV.

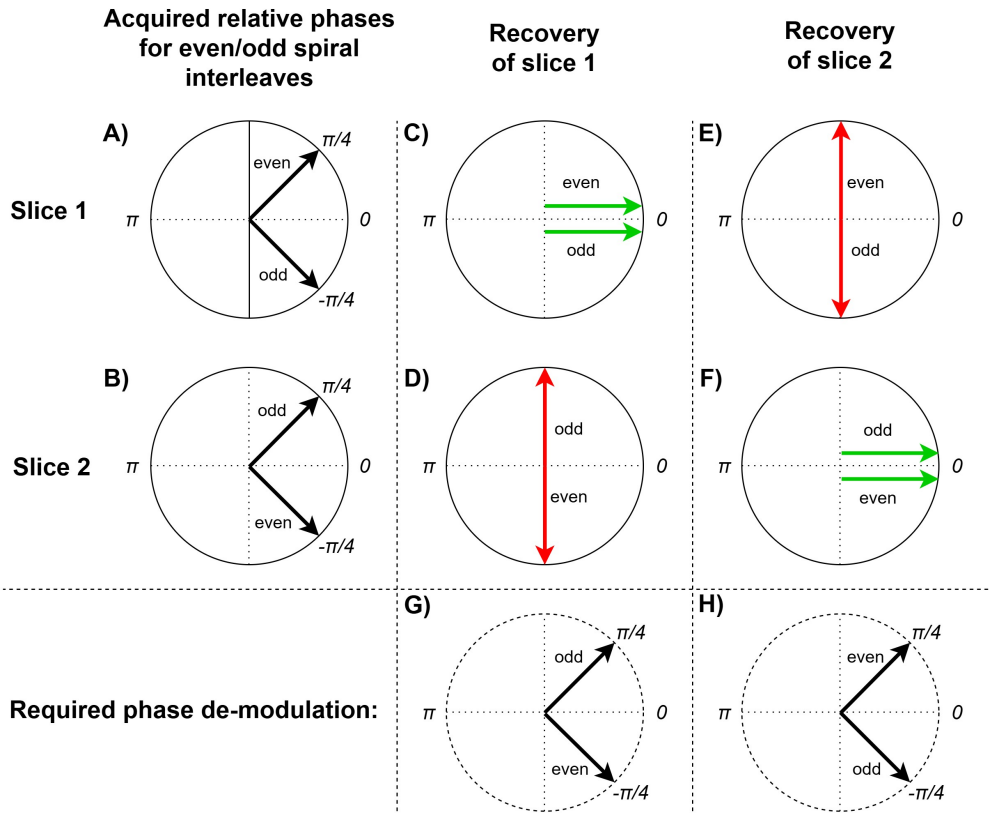


Figure 1.10: Schematic diagram of the phase de-modulation for even and odd spiral interleaves employed to separately recover two simultaneously acquired slices, where the G_z isocenter is in the center between the two slices. A) the accrued phase for slice 1. B) the accrued phase for slice 2. C) the constructive signal interference for slice 1 after recovering slice 1. D) the destructive signal interference for slice 2 after recovering slice 1. E) the destructive signal interference for slice 1 after recovering slice 2. F) the constructive signal interference for slice 2 after recovering slice 2. G) the phase de-modulation required to recover slice 1. H) the phase de-modulation required to recover slice 2. This diagram is reproduced from the manuscript of Study IV.

1.3 Deep learning

Recently, *deep learning* (DL) has had a great impact on the fields of image analysis and image processing by outperforming traditional image analysis methods on problems such as image classification, image segmentation and image de-noising. DL is a sub-category of artificial intelligence, and it focuses specifically on *neural networks*. These originated as models that imitated the function of the human brain, but have since been massively improved. Today, many different types of neural network architectures exist for solving a wide range of problems. An important factor in the rise of DL are developments in computational hardware, specifically Graphics Processing Units (GPUs), that make it possible in practice to train and employ large neural networks on large amounts of data. This chapter provides a theoretical background to the DL concepts that were touched upon in this thesis.

1.3.1 Fundamentals of neural networks

A widely used type of neural network is the *feedforward* neural network (FNN). FNNs are characterized by providing a mapping between some input $\mathbf{x}$ and some output $\mathbf{y}$ by approximating some function f^*. This mapping can be described as a parametrized function $\mathbf{y} = f(\mathbf{x}; \theta)$, where θ is a set of parameters that are *learned* [28].

To learn the parameters θ, neural networks are trained on training data using some *learning algorithm*. The learning algorithm determines all aspects of the training process, for example how training data are used in order to update θ. A learning algorithm can be *supervised*, meaning that each element in the *training dataset* $\mathbf{x}_i \in \{\mathbf{x}_1, \mathbf{x}_2 ..., \mathbf{x}_n\}$ has a corresponding labelled *ground truth* value $\mathbf{y}_i$, which is the ideal output from the neural network given the input $\mathbf{x}_i$. In this thesis, this was the only type of learning algorithm that was used.

The general structure of a neural network

Neural networks are typically constructed as computational graphs of several *layers*, where each layer performs some mathematical operation [28]. An example of the structure of a simple type of FNN can be seen in Figure 1.11a. Here, the network consists of three layers, an input layer, a *hidden* layer, and an output layer, between which information flows in a single direction. Each node in each layer is connected to each node in the next layer, making it a *fully connected* FNN. In each node of this type of network, the weighted sum of all inputs is computed, some *bias* is added, and some non-linear *activation function* is applied to this weighted sum before the output is given to the nodes of the next layer (Figure 1.11b). In a fully connected FNN, the learnable parameters θ that are updated during the execution of the learning algorithm are the weights and biases. Most DL architectures use multiple hidden layers, making them *deeper* than the architecture in Figure 1.11a. This is the origin of the term *deep learning*.

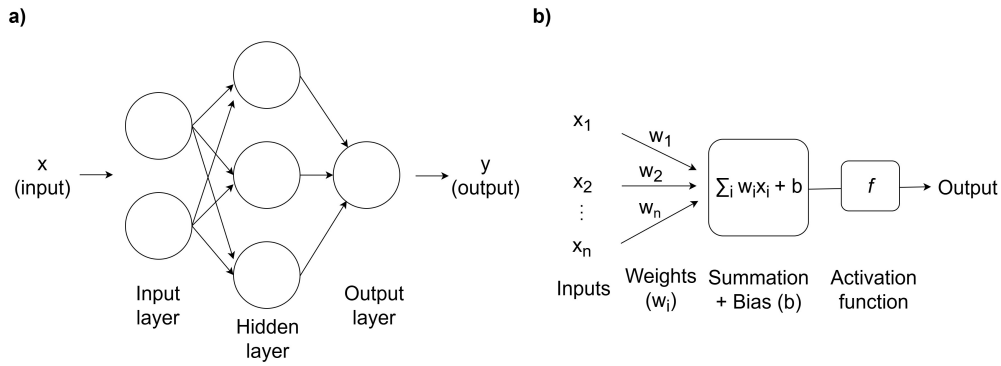


Figure 1.11: a) A fully connected feedforward neural network (FNN) with one input layer, one hidden layer, and one output layer. b) The internal structure of a node in the fully connected network. In this, inputs x_i are multiplied by weights w_i and summed, some bias b is added to the weighted sum, and a nonlinear activation function is applied before the output is given to the nodes in the next layer.

The use of nonlinear activation functions allows networks to learn nonlinear mappings. Two common nonlinear activation functions are the *rectified linear unit* (ReLU), and the leaky ReLU (LeakyReLU) functions

$$\text{ReLU}(z) = \begin{cases} 0, & z \leq 0 \\ z, & z > 0 \end{cases} \tag{1.9}$$

$$\text{LeakyReLU}(z) = \begin{cases} \alpha z, & z \leq 0 \\ z, & z > 0 \end{cases} \tag{1.10}$$

where $\alpha > 0$ is some small slope value [28]. The choice of activation function for the output layer will depend on the task of the network. For example, for a *classification* problem with n classes, the *Softmax* activation function produces a normalized probability distribution across all classes [28]. For a classification problem with n classes, the Softmax output for an input $\mathbf{z} = (z_1, z_2, ..., z_n)$ is

$$\text{Softmax}(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}. \tag{1.11}$$

The Softmax output function is also commonly applied for *segmentation* problems, since this involves the classification of each pixel in an image. For some problems, including image de-noising, linear output activation functions can be used.

Training neural networks

When an FNN produces an output $\hat{\mathbf{y}}$ from some input $\mathbf{x}$, the quality of the produced output can be determined through some *loss function*. In the case of supervised learning, the loss function typically contains a measure of the similarity between the predicted output $\hat{\mathbf{y}}$ and the ground truth $\mathbf{y}$ (see section 1.3.3). The learning algorithm then uses an algorithm known as *backpropagation* to employ the loss value for learning. Backpropagation utilizes the chain rule of calculus to propagate the computed loss value backwards trough the network to determine the gradient of the loss with respect to each learnable parameter [29]. The computed gradients can then be utilized by some gradient-based optimization method in order to determine how the parameters should be updated. In this thesis, the optimization methods *Adam* [30] and stochastic gradient descent with momentum were employed. An important parameter of these optimizers is the *learning rate*, which determines the step size for each step of the optimization.

Constructing a learning algorithm

To construct a learning algorithm for training a neural network, a series of design choices have to be made. These include problem-specific design choices, such as the choice of loss function, network architecture, and hyperparameters, but also choices such as the hardware [31]. An important aspect of a learning algorithm is also the preparation of the available training data, in terms of how it should be pre-processed and whether any *data augmentation* should be carried out before or during training. Data augmentation entails the synthetization of additional training dataset elements through slight modifications of existing elements in the training dataset, which can increase the generalizability of a model [28]. Another design choice related to the training data is the choice of *batch size*, which determines how many training data elements should be used (in each *mini-batch* of data) before each parameter update during training. Frameworks for automatically configuring these types of design choices exist. One example is the *nnU-Net* for medical image segmentation, that configures choices based on the training data [32]. However, manual design choice configuration is still commonplace.

1.3.2 Convolutional neural networks (CNNs)

Convolutional neural networks (CNNs) are a type of FNN that perform a *convolutional operation* in at least one layer [28]. In a discrete convolutional operation, a convolutional kernel (or *filter*) is applied to an input matrix in multiple steps, and a summation of the elementwise multiplication between the kernel elements and the corresponding input elements gives the elements of the resulting feature map. How the kernel is applied to the input matrix in each step is determined by the *stride*. The stride is the step size of the convolutional kernel in each dimension. For example, for a 3D convolution, a stride of [1,1,1] means that step sizes of 1 pixel are used in the x-, y-, and z-directions. Figure 1.12a and

1.12b shows examples of 2D and 3D discrete convolutional operations with strides $[1, 1]$ and $[1, 1, 1]$, respectively.

Convolutions allow the detection of local features in images by application of filters that are smaller than the image, and they also allow a more sparse connectivity than fully connected FNNs [28], making them useful for high-dimensional inputs such as images or videos. In CNNs, the filters of the convolutional layers are part of the learnable parameters θ, and these are updated during training to learn to extract useful features. Each convolutional layer in a CNN typically contains multiple filters in order to be able to extract multiple different types of features from the input. In addition to filters, convolutional layers also include a learnable bias, and some nonlinear activation function is typically applied to the resulting feature maps.

Another common operation in CNNs is *max-pooling*. In this operation, the maximum value of the elements within a specified area of the input matrix is selected. Figures 1.12c and 1.12d shows 2D and 3D max-pooling operations with strides $[2, 2]$ and $[2, 2, 2]$, respectively. Max-pooling is commonly used as a down-sampling method for reducing the dimensionality of intermediate feature maps in CNNs [29].

In CNNs, convolutions can also be *transposed*. This means that they can be made to produce the opposite effect of a standard discrete convolution and instead function as an up-sampling operation that produces output feature maps of a higher dimensionality than the input [33]. This is commonly applied in CNNs after some initial downsampling steps. In some CNN architectures, up-sampling operations are instead done without convolutional operations, for example using nearest neighbour interpolation.

CNN architectures

One of the most widely used CNN architectures in medical image *segmentation* for the past decade is the *U-Net* [34]. U-Net-based architectures were employed for image segmentation in Study I and II and for de-noising in Study IV of this thesis. The U-Net consists of an *encoding* path followed by a *decoding* path. The purpose of the encoding path is to downsample the data in multiple steps to extract features and their context in the input image [34]. The purpose of the decoding path is to do a stepwise upsampling of the extracted features to the desired output size (commonly the same as the input size) of the network. It also locates objects in the image by combining feature maps extracted at different levels of downsampling [34]. *Skip connections* between each downsampling step and the corresponding upsampling step allows combination of image features from different levels.

In the original 2D U-Net architecture by Ronneberger et al. (2015), the encoding path consisted of multiple steps of two convolutional layers with 3×3 filters and a ReLU activation followed by a 2×2 max-pooling operation, where the number of filters are doubled at each such step [34]. The decoding path consisted of multiple steps of 2×2 transposed convolutions followed by a concatenation with filters from the skip connections

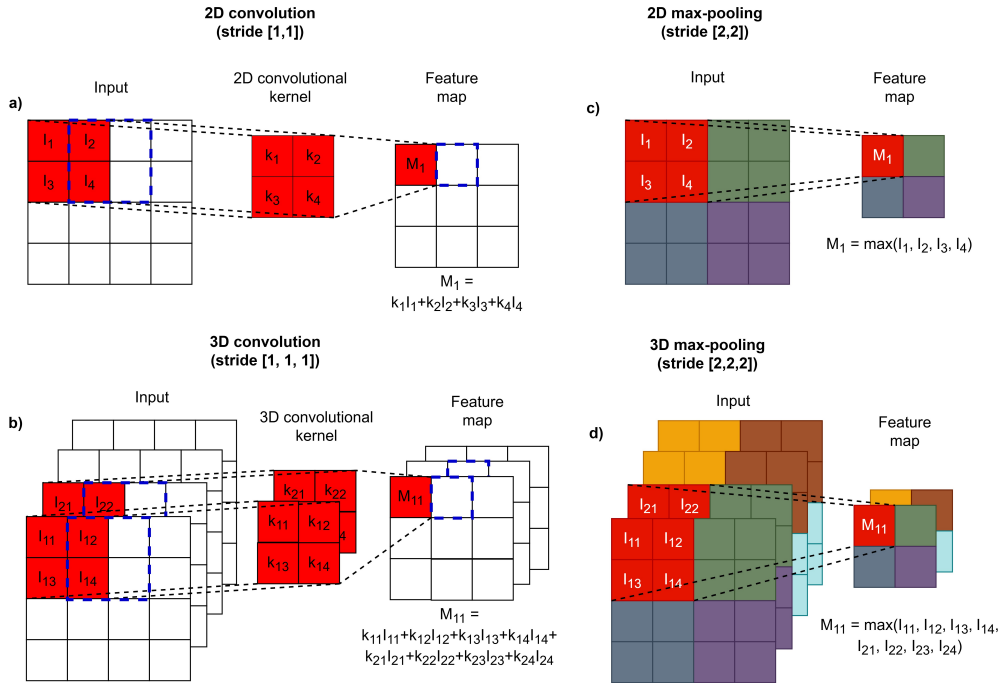


Figure 1.12: a) A 2D discrete convolutional operation where a 2×2 convolutional kernel is applied to a 4×4 input using a stride of $[1, 1]$, producing a 3×3 feature map. The blue dotted squares indicate the next kernel position and resulting feature map position according to the stride. b) A 3D discrete convolutional operation where a $2 \times 2 \times 2$ convolutional kernel is applied to a $4 \times 4 \times 4$ input using a stride of $[1, 1, 1]$, producing a $3 \times 3 \times 3$ feature map. The blue dotted squares indicate the next kernel position and resulting feature map position according to the stride. c) A 2D max-pooling operation where a 4×4 input is assessed within a 2×2 kernel using a stride of $[2, 2]$, producing a 2×2 feature map. The colours indicate the stride. d) A 2D max-pooling operation where a $4 \times 4 \times 4$ input is assessed within a $2 \times 2 \times 2$ kernel using a stride of $[2, 2, 2]$, producing a $2 \times 2 \times 2$ feature map. The colours indicate the stride.

and two convolutional layers with 3×3 filters. Softmax was used as the output activation function. An implementation of this architecture was employed in Study I.

The U-Net has also been expanded to handle 3D input images [35]. Figure 1.13 shows the version of the 3D U-Net that was implemented for Study IV of this thesis. In this implementation, the encoding path had two downsampling steps containing two 3D convolutions with $3 \times 3 \times 3$ filters followed by $1 \times 2 \times 2$ max-pooling operations. Max-pooling across time (the first dimension) was avoided to avoid blurring of motion. The decoding path used 3D up-sampling layers followed by a concatenation with filters from skip connections and two $3 \times 3 \times 3$ convolutions with LeakyReLU activations. The final operation was a $3 \times 3 \times 3$ convolution with a linear activation function. This architecture was chosen experimentally in Study IV.

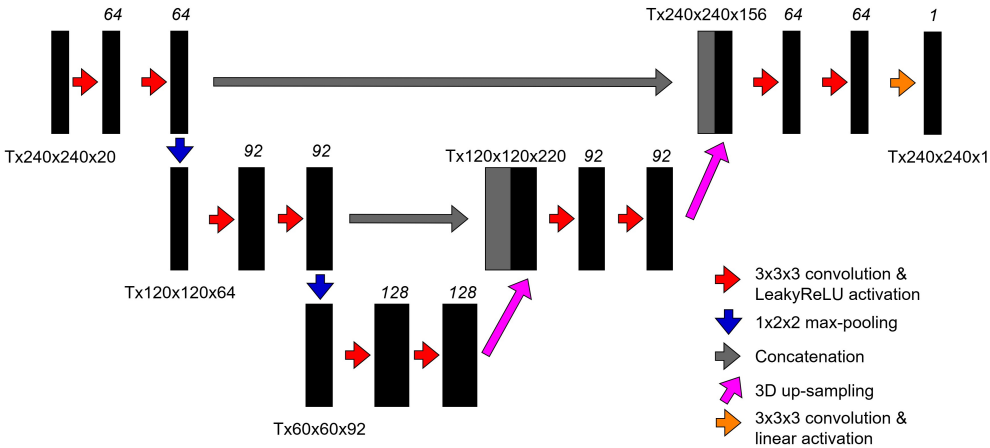


Figure 1.13: The 3D U-Net architecture implemented for Study IV of this thesis.

In Study I of this thesis, modified versions of the *Darknet-19* architecture [36] were employed for one binary classification problem and one regression problem. The original architecture consists of 19 convolutional layers with increasing numbers of 3×3 filters, 5 layers of max-pooling, and a global average pooling layer that computes the global average of each feature map, followed by a Softmax activation for classification. The modification done for the binary classification problem was to change the number of filters of the final convolutional layer to 2, such that the Softmax output activation only returned probabilities for two classes. The modification for the regression task was to additionally change the output Softmax activation to a ReLU activation function.

1.3.3 Loss functions

The image analysis problems that were addressed in this thesis were *image segmentation*, *image classification*, *image-based regression*, and *image de-aliasing*. Problem-specific loss func-

tions were used for each task. Image segmentation entails assigning a class label to each pixel of an image. For image segmentation, the *Dice loss* function can be used. The Dice loss function [37] is based on the *Sørensen-Dice coefficient*, or the *Dice score*, which is a measure of spatial overlap between two binary images. The Dice loss can be given between a predicted binary image $\hat{Y}$ and a ground truth binary image Y with N pixels or voxels $\hat{y}_i \in \hat{Y}$ and $y_i \in Y$ as $L_{Dice} = 1 - D$ where

$$D = \frac{2\sum_i^N \hat{y}_i y_i}{\sum_i^N \hat{y}_i^2 + \sum_i^N y_i^2} \tag{1.12}$$

is the Dice score [37].

Another loss function that can be used for both image segmentation and image classification is the *cross-entropy* loss function. This is useful when the output of the network consists of probability distributions, such as when using a Softmax output activation function [29]. In the binary case, for an image of N pixels, cross-entropy loss can be defined as

$$L_{CE} = -\sum_i^N (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)) \tag{1.13}$$

between a ground truth label y and the corresponding predicted probability $\hat{y}$ [38].

For image-based regression problems, the *mean squared error* (MSE) loss function is often used. The mean squared error loss function between a predicted output array $\hat{\mathbf{y}}$ of N values and the ground truth array $\mathbf{y}$ can be given as

$$L_{MSE}(\hat{\mathbf{y}}, \mathbf{y}) = \frac{1}{N} \sum_i^N (\hat{y}_i - y_i)^2 \tag{1.14}$$

For image restoration tasks, such as image de-aliasing and de-noising, a common loss function is the *structural similarity index measure* (SSIM). The SSIM is a measure of the structural similarity between images that is based on combining measures of *luminance*, *contrast* and *structure* of the images [39]. For the mathematical definition of the SSIM, the reader is referred to the original publication [39].

1.3.4 Performance estimation & performance measures

A training dataset can be split up in multiple ways to allow both training and performance estimation to be done using the dataset. A common way to split a dataset is via *hold-out validation*. This means that a part of the dataset (often 20 %) is held out from training and used for testing the performance of the trained model. Often, a dataset is split into three parts, one for training, one for validation (i.e. performance evaluation during training), and one for testing. Another way to split a dataset is through *k-fold cross-validation*. This means

that the dataset is split into k (often $k = 5$) equally sized parts (*folds*) without overlap. Performance is then estimated on one fold at a time, while using the remaining folds for training. The average performance across the k folds provides a performance measure that is less biased with respect to the chosen dataset split than a performance measure from a single hold-out validation [29]. This method is often also employed when only a limited amount of data is available.

For the deep learning models trained in this thesis, problem-specific performance measures were employed. Image segmentation performance was evaluated using the 2D and 3D Dice score (as given in Equation 1.12). It was also measured using the *Hausdorff distance*, which is a measure of the distance between two point sets, calculated as the furthest of all the closest distances between points in the two sets. When measuring segmentation performance, the point sets correspond to non-zero objects in binary segmentation masks.

To measure image classification performance, *accuracy* was used, which is the number of correct predictions divided by the total number of predictions. To measure image-based regression performance, the mean squared error (Equation 1.14) was used. To measure de-aliasing performance, the average SSIM was used. In all studies, performance measures specific to the medical applications were also used (see section 1.4.2).

1.4 Cardiovascular physiology

In this thesis, all studied methods were in various ways connected to the goal of producing measures of cardiac function from cardiovascular MRI images. This chapter gives an overview of the most important aspects of cardiovascular anatomy and physiology that were touched upon in the studies of this thesis.

1.4.1 Cardiac anatomy

The heart is an anatomical structure with the primary task of driving the blood flow in the cardiovascular system. A normal heart has four *chambers*, known as the right and left *atria* and the right and left *ventricles*. The main function of the ventricles is to pump blood, while the atria also pump blood but mainly receive and hold blood that is sent on to the ventricles. The left ventricle (LV) pumps oxygenated blood through the aorta to the *systemic circulation*, which is the circuit of blood vessels that provides oxygenated blood to the organs and tissues of the body. The left atrium and ventricle are separated by the *mitral valve*, and the left ventricle and the aorta are separated by the *aortic valve*. These valves open at different points during the cardiac cycle to let blood pass. The right ventricle (RV) pumps blood to the *pulmonary circulation*, which is the circuit of blood vessels that transports blood through the lungs in order to oxygenate it and remove carbon dioxide. The left atrium receives the blood that has passed through the pulmonary circulation, and the right atrium receives the blood that has passed through the systemic circulation.

Using CMR, the heart can be imaged in multiple different *imaging planes* to show

different aspects of its anatomy. Figure 1.14 (left) shows a CMR image of the heart in the *four chamber* view, which is an imaging plane that shows all four heart chambers. Figure 1.14 (right) shows the heart in the *short-axis* view plane. This is the imaging plane of the heart that is the most common for performing ventricular volumetric measurements, and it is the view plane that was used in all four studies of this thesis.

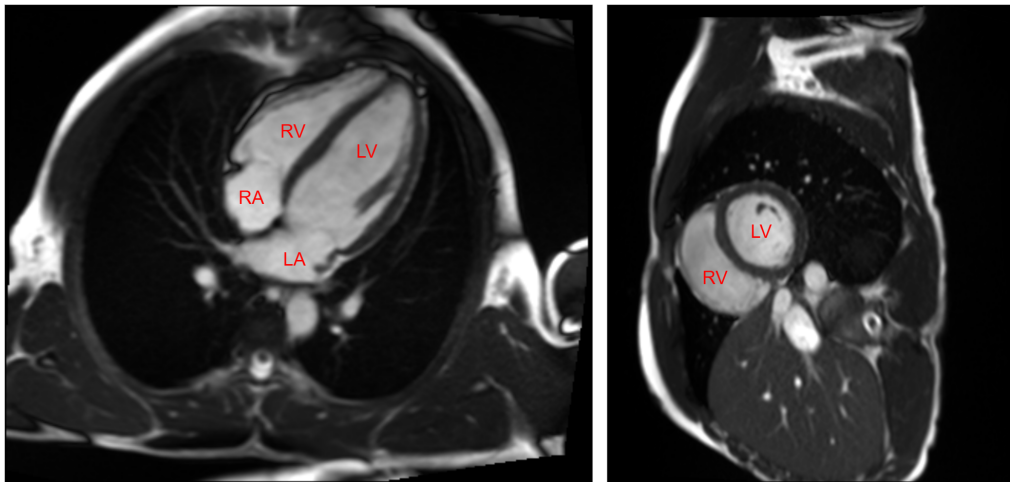


Figure 1.14: Left: A CMR image of the heart in the four chamber view, where all the four chambers of the heart are visible, i.e. the right atrium (RA), the left atrium (LA), the left ventricle (LV), and the right ventricle (RV). Right: A CMR image of the heart in the short-axis view, where the two ventricles (RV and LV) are visible.

1.4.2 Measures of cardiac function from MRI

From CMR, ventricular volumes can be measured if the images provide sufficient coverage of the ventricles. Volumes are measured by performing image segmentation, which provides a classification of each voxel as either being background or part of the volume. The volume of each voxel is known, and the full volume of the segmented object can be given as a sum of all voxel volumes. In short-axis CMR images, the LV is commonly segmented in both the *endocardium*, i.e. the innermost layer of the ventricle, and the *epicardium*, which is the outermost layer. The volume enclosed by the endocardium is also known as the *blood pool*. Extending into the blood pool, there are small pieces of muscle that are known as *trabeculations*. For the RV, it is commonly only the endocardium that is segmented. Segmentations of these structures in a short-axis CMR image can be seen in Figure 1.15.

From the ventricular volumetric measurements, measures of cardiac function can be derived separately for the two ventricles. For many measures of cardiac function, the volumes of the ventricles are measured at the two extremes of the cardiac cycle in terms of

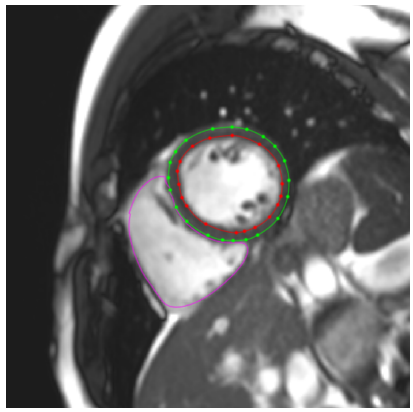


Figure 1.15: A short-axis CMR image with segmentations of the endocardium (red) and epicardium (green) of the left ventricle as well as the endocardium of the right ventricle (magenta).

blood volume, the *end-systolic* (ES) and *end-diastolic* (ED) phases (see section 1.4.3), giving the ED volume (EDV) and ES volume (ESV). The ED phase is when the ventricles are the most filled with blood, and it marks the end of ventricular filling. The ES phase is when the ventricles contain the least amount of blood, and it marks the end of ventricular contraction. From the EDV and ESV, the *stroke volume* (SV), which is the volume of blood that is pumped from each ventricle at each heartbeat, can be given as

$$\mathrm{SV} = \mathrm{EDV} - \mathrm{ESV}. \tag{1.15}$$

The *ejection fraction* (EF) can be given as

$$\mathrm{EF} = \frac{\mathrm{SV}}{\mathrm{EDV}} \times 100 \tag{1.16}$$

and indicates the percentage of the EDV that is ejected from the respective ventricles at each heartbeat. It can be calculated both for the LV (LVEF) and the RV (RVEF).

From the SV and a measured *heart rate* (HR), commonly given in beats per minute, the *cardiac output* (CO), which is a measure of the ejected blood volume per minute, can be calculated as

$$\mathrm{CO} = \mathrm{SV} \times \mathrm{HR}. \tag{1.17}$$

Given volumetric segmentations of both the endocardium and epicardium of the LV, the *left ventricular mass* (LVM) can be calculated as the difference between the epicardial and endocardial volumes multiplied by the myocardial density, which is assumed to be 1.05 g/ml [40]. The LVM does not provide a direct measure of cardiac function, but is a useful measure for validating that imaging methods provide accurate measurements. In addition,

LVM can be an important indicator of pathology, especially when measured in relation to blood volumes. The LVM should be constant across all time points of the cardiac cycle, and can thus be used to make sure that different imaging methods, e.g. different MRI pulse sequences, agree in their depiction of anatomy.

Among many other measures that can be derived from CMR, the measures of cardiac function given above can be used clinically to aid diagnosis of cardiovascular pathology. One cardiovascular disease where e.g. the EF measurement is of use for diagnosis and subtype classification is *heart failure*. Heart failure is when the heart is unable to pump the amount of blood required to fulfill the metabolic needs of the body, and it is a complex disease with a range of etiologies and subtypes [41].

1.4.3 Cardiac mechanics

The motion of the heart is periodical, and consists of a series of mechanical events. The cardiac cycle can be divided into two separate sections, *diastole* and *systole*. During diastole, the ventricles are relaxed, and during systole, the ventricles contract. These two sections in turn consist of several mechanical sub-events. Below follows an overview of these mechanical events and how they can change between rest and exercise.

Cardiac mechanics at rest

As mentioned in section 1.2.7, the triggering of the data acquisition in CMR can be done with an electrocardiogram (ECG) signal. The ECG signal is used to make sure the acquisition coincides with specific mechanical events of the cardiac cycle. Figure 1.16 shows a *Wiggers diagram* that illustrates how the timings of the events of an ECG signal relate to the mechanical events of the left side of the normal heart. The diagram starts with the final event of diastole, namely *atrial systole* [42]. Atrial systole is initiated by an electrical activation of the atria, indicated by the *P*-wave of the ECG, leading to atrial contraction. The contraction slightly increases the blood volume of the ventricle, giving the final diastolic filling before the mitral valve closes.

Directly following atrial systole, ventricular systole (or simply systole) starts. The start of systole coincides with the *QRS-complex* of the ECG signal. This series of electrical events indicates the electrical activation of the ventricles, leading to ventricular contraction [43]. The initial event of systole is *isovolumetric contraction*, where the ventricular walls develop tension while the mitral and aortic valve are closed, which increases the pressure of the blood within the ventricles without affecting the blood volume [42], producing a change in ventricular geometry. When the aortic valve opens due to the increased blood pressure in the LV exceeding the pressure in the aorta, the event of *ejection* begins. The ejection lasts until the aortic valve closes again due to a reversed pressure difference between the LV and the aorta [43], marking the end of systole and the beginning of diastole. During the ejection phase, the T-wave of the ECG occurs, indicating the return of the ventricles to their electrical resting state.

The first event of diastole is *isovolumetric relaxation*, during which the ventricles relax while the blood volume stays constant due to the mitral and aortic valves being closed. When the mitral valve opens due to the pressure difference between the left atria and the LV, *early rapid filling* starts, where blood collected in the atria flows rapidly and passively into the ventricles. When the passive filling of the ventricles starts to decrease, this marks the start of the *diastasis*, which is the longest mechanical event of the cardiac cycle for lower HR. This occurs until *atrial systole* happens again, and the cardiac cycle starts over. At rest, most of the ventricular filling occurs during early diastole [42].

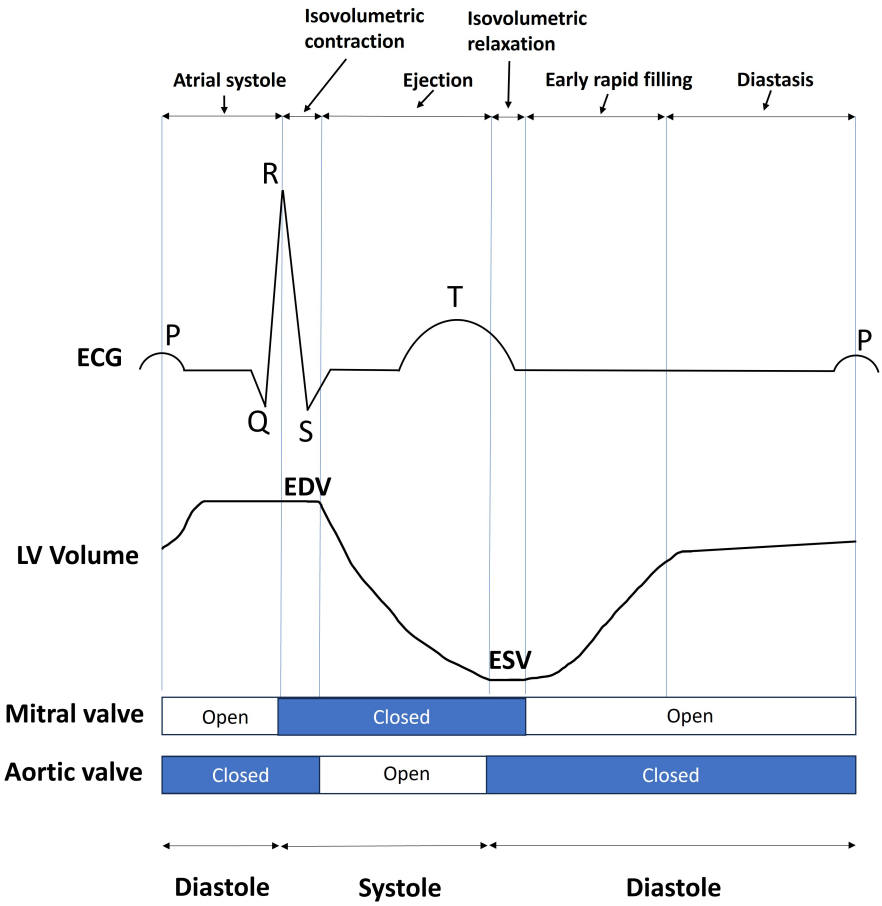


Figure 1.16: A Wiggers diagram showing the timings of the mechanical events of the cardiac cycle, an ECG signal, a typical left ventricular blood pool volume curve, and the timings of the openings and closings of the mitral and aortic valves. This diagram is inspired by the Wiggers diagram in [42].

Cardiac mechanics during exercise

During exercise, there is an increased demand for transporting oxygen and carbon dioxide to and from the active skeletal muscle, requiring an increase the cardiac output (CO) [44]. Looking at Equation 1.17, this can be done by increasing the HR, by increasing the SV, or both. For individuals with normal cardiac function, the increase in CO during exercise is mainly given by an increased HR, with minor increases in the SV [42]. The compensatory increase in HR commonly has a linear relationship with the exercise intensity, but this depends on the type of exercise performed [45]. The increase in heart rate leads to different changes in duration for the different mechanical events of the cardiac cycle.

As described above, diastole consists of isovolumetric relaxation, early rapid filling (ERF), diastasis, and atrial systole. The *mechanical diastolic duration* (MDD) can be defined as diastole without the isovolumetric relaxation. When HR increases, a nonlinear decrease in the duration of the diastasis is the primary factor in the decrease of the MDD [46]. For HR elevations at lower HRs, the elimination of the diastasis is what causes the main reduction of an R-R interval [47]. As the HR increases further, the duration of the diastasis is reduced to zero, causing a merge of ERF and atrial systole [46, 47]. Figure 1.17 shows LV volume curves across one cardiac cycle from CMR images acquired during three different exercise intensities, constructed using the method developed in Study III. This clearly shows the reduction of the diastasis between rest and exercise.

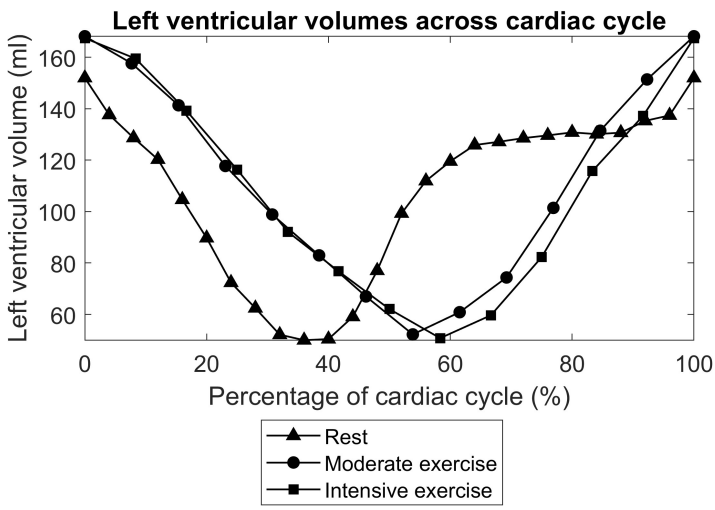


Figure 1.17: Left ventricular volume curves from cine CMR images of one healthy subject acquired at different exercise intensities. The heart rate at rest was estimated to be 59 BPM, the heart rate during moderate exercise was 143 BPM, and the heart rate during intensive exercise was 155 BPM. For exercise, the cine CMR images were constructed using the method developed in Study III.

Ventricular volumes during exercise

In addition to an increased HR, exercise can lead to changes in the level of ventricular filling compared to the resting state. However, how the changes appear depends on the intensity of exercise and the position it is performed in (e.g. upright or supine). For exercise in the supine position, which is the position used for exercise during CMR in Study III, the normal response to sub-maximal exercise for healthy individuals is for the left ventricular end-diastolic volume to stay unchanged, while the end-systolic volume decreases [48]. This means that any increase in SV occurs due to an decrease in end-systolic volume rather than increased filling [48]. Ventricular volumes are also affected by respiration, and it has been observed that left ventricular volumes are larger during expiration and the right ventricular volumes are larger during inspiration, both during rest and during exercise [49].

1.5 Study motivations

The four studies included in this thesis were all connected to the use of deep learning in CMR for assessing measures of cardiac function. Below follows the motivations behind each study.

Study I

The right ventricle (RV) can be a challenging structure to segment in short-axis images [50]. It has a complex variability in shape [51], ambiguous borders, and trabeculations [52]. These aspects make it time-consuming to segment manually, and may cause substantial inter-observer variability, which shows the need for automated segmentation methods. At the time of starting this study, many automated RV segmentation methods based on CNNs had already been presented in literature [53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63]. However, it had also been reported that CNN-based RV segmentations were still in need of expert verification and corrections [50]. Meanwhile, no studies had reported on if CNNs can actually reduce clinical RV segmentation time compared to fully manual segmentations, given this persisting need for corrections. This motivated Study I, in which a clinically applicable CNN-based segmentation pipeline was developed and the aspect of clinical time reduction with regards to the need for corrections was assessed.

Study II

Learning algorithms for deep learning-based medical image segmentation, such as the one developed in Study I, utilize randomness in various forms before and during the training process to aid the production of segmentation models. The randomness arises from the initialization of parameters (e.g. weights) [64, 65], the sampling of mini-batches [30], *drop-out* [30], data augmentation [66], non-deterministic GPU operations [65], and non-deterministic operations in deep learning software libraries [67, 68]. The randomness can

lead to performance differences between segmentation models that are generated using the very same learning algorithm. However, learning algorithms in literature are often compared based on performance comparisons between segmentation models. This motivated Study II, in which it was assessed how randomness affects the reliability of standard methods for comparing the performance of segmentation models.

Study III

Conducting CMR during ongoing exercise (exercise CMR) can help reveal pathophysiology that is not visible during rest [16]. However, exercise CMR introduces the inability to employ ECG-gating due to motion and the magneto-hydrodynamic effect [69, 16], and makes breath-holding difficult. Because of this, free-breathing real-time imaging is widely used [70], but breathing affects ventricular filling [49], and motion from breathing and from exercise leads to the displacement of the heart. To minimize the influence of breathing on volumes, retrospective respiratory and cardiac gating methods can synchronize timeframes from matching breathing and cardiac states across slice positions. Previous studies have presented methods for respiratory and cardiac gating of real-time images in image space [71, 72, 73, 74]. However, no existing method preserved the temporal order of consecutively acquired timeframes when constructing time-resolved CMR images, which is of relevance for accurate assessments of time-resolved cardiac mechanics during exercise, for example for constructing non-invasive pressure-volume loops [75]. This motivated Study III, in which a retrospective gating method for producing time-resolved ventricular CMR images from real-time exercise CMR was developed.

Study IV

In addition to exercise CMR, free-breathing real-time CMR is also of use in other cases where breath-holding is challenging, including pediatric CMR [15]. In both of these cases, fast imaging is of relevance to reduce the time that the patient needs to be in the scanner and to reduce the influence of bulk motion. Although real-time bSSFP CMR with non-Cartesian sampling of k-space allows rapid imaging [76], the acceleration is still limited by the need to collect 10-16 slices in order to fully cover the ventricles, thus requiring imaging during at least 10-16 R-R intervals. Simultaneous multi-slice (SMS) imaging can accelerate acquisitions by a factor equal to the number of simultaneously acquired slices [21], and has been combined with bSSFP imaging in several previous studies [25, 77, 78]. Furthermore, real-time bSSFP SMS CMR has been conducted at 0.55 T in a recent study [79]. However, in this, iterative reconstruction was employed [79], which requires extensive reconstruction times and limits the possibility to do online reconstruction at the scanner [11]. Rapid online reconstruction can be achieved by employing deep learning [80, 81, 82], which has also previously been employed in SMS CMR [83, 84]. This motivated Study IV, in which a bSSFP SMS sequence was implemented at 1.5 T and a deep learning-based reconstruction

method was developed to achieve fast online reconstruction of highly accelerated real-time SMS CMR.

Chapter 2

Aims

The overarching aim of this thesis was to assess the benefits of incorporating deep learning-based methods in the process of evaluating cardiac function using cardiovascular magnetic resonance (CMR) imaging. This included applications in both reconstruction and image analysis. The individual studies had specific aims, and these were:

Study I: To develop a deep learning-based pipeline for right ventricular segmentation in CMR and to assess its clinical usefulness in terms of reducing the time to obtain right ventricular segmentations of clinically sufficient quality.

Study II: To assess how randomness in neural network training affects the reliability of conventional techniques for comparing deep learning-based medical image segmentation models.

Study III: To develop a semi-automated method for retrospectively constructing time-resolved real-time exercise CMR cines that are synchronized in terms of respiratory and cardiac phase and preserves the temporal order of consecutive timeframes, and to investigate the possibility to use this method for time-resolved measures of cardiac function in heart failure patients and healthy volunteers.

Study IV: To: (1) develop a balanced steady state free precession (bSSFP) simultaneous multi-slice (SMS) sequence on a 1.5 T system for rapid free-breathing image acquisition; (2) develop a deep learning-based image artifact suppression reconstruction method for rapid online reconstruction of real-time bSSFP SMS CMR data; and (3) validate the feasibility of this method by comparing it with standard methods.

These four studies approached the overarching aim from different angles. Study I assessed the benefits of deep learning-based segmentations in clinical CMR practice. Study II contributed to the understanding of how to interpret and conduct reporting in stud-

ies on deep learning-based medical image segmentation in CMR. Study III assessed the benefits of deep learning as part of a new method for conducting measurements in exercise CMR. Study IV assessed the utility of a deep learning-based method for rapid online reconstruction of real-time imaging data from a SMS bSSFP CMR sequence.

Chapter 3

Materials and methods

The process of deriving measures of cardiac function from cardiovascular magnetic resonance (CMR) imaging involves multiple steps, including image acquisition, image reconstruction, image processing, and image analysis. This thesis touched upon all of these aspects. Figure 3.1 shows a flowchart of the general steps involved in obtaining measures of cardiac function from CMR imaging. This figure also specifies where the four different studies in this thesis fit in to this process, and how deep learning (DL) was used in each study. This chapter provides an overview of the materials and methods used to conduct the studies included in this thesis. Figures in this thesis were prepared using draw.io (diagrams.net: JGraph Ltd), MATLAB (Natick, Massachusetts: The MathWorks Inc.), and various Python libraries.

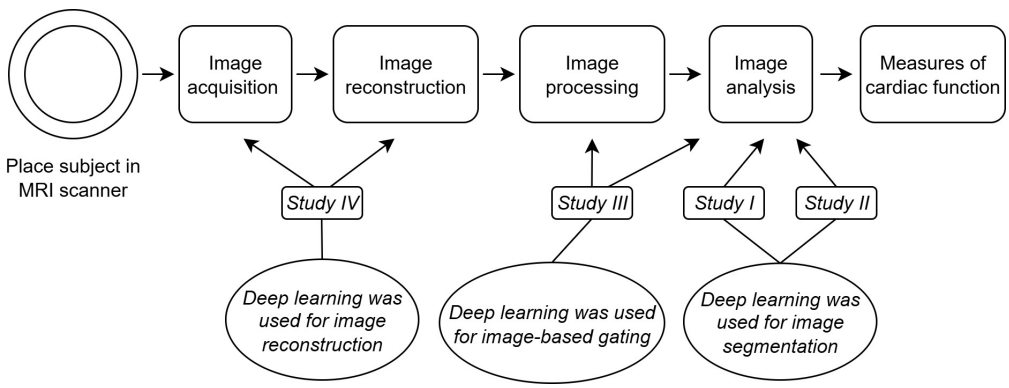


Figure 3.1: A flowchart of the general steps involved in CMR imaging for acquiring measurements of cardiac function. The flowchart indicates what aspects of CMR imaging are addressed in the four studies in this thesis, and how deep learning was employed in each study.

3.1 Imaging data and study populations

This thesis employed MR imaging data from human subjects in all studies. The imaging data was either used for training neural networks, for testing the performance of neural networks, or for validating the performance of the developed or existing methods that were employed in each study. It came from several different populations of human subjects, collected at multiple sites. In all cases, data collection was or had previously been conducted in accordance with the Declaration of Helsinki, and the use of data adhered to guidelines and regulations outlined by ethical review authorities.

Study I

In this study, a dataset of 1434 subjects was assembled for training, validation, and testing neural networks for the right ventricular segmentation pipeline. This consisted of short-axis bSSFP CMR images that were collected during clinical practice at Skåne University Hospital in Lund between 2019 and 2020, as well as during various research projects between 2004 and 2020. An additional dataset of short-axis CMR images from 50 clinical scans at Skåne University Hospital in Lund was used for validating performance. Additionally, a scan-rescan assessment was conducted using a dataset of short-axis CMR images from 10 healthy subjects acquired at Skåne University Hospital for a previous research project. Finally, additional performance assessments were done on a publicly available testing dataset of short-axis bSSFP CMR images of 50 subjects from the 2017 MICCAI Automatic Cardiac Diagnosis Challenge (ACDC) challenge [85]. Approval to use image data from research CMR scans was given by the Regional Ethical Committee in Lund (EPN Dnr 621/2004, 2010/114, 2010/248, 2011/777, 2010/55, 741/2004, 269/2005), and approval to use image data from clinical CMR scans was given by the Swedish Ethical Review Authority (Dnr 2021-03583).

Study II

In this study, only publicly available datasets were used for training, validation, and testing of the neural networks. This included the training set (100 subjects) of short-axis bSSFP CMR images from the 2017 MICCAI Automated Cardiac Diagnosis Challenge (ACDC) [85]. This also included the training set (484 subjects) from the Brain Tumour dataset and the training set (263 subjects) from the Hippocampus dataset of the Medical Segmentation Decathlon [86]. Note that this data from non-CMR MRI was used to make the assessment apply outside of the field of CMR and to DL-based medical image segmentation in general.

Study III

In this study, short-axis CMR data from 20 subjects were used for assessing the performance of the developed method. Of the 20 subjects, 10 were healthy volunteers and 10 were heart failure patients, included using the inclusion criteria of having heart failure as a clinical diagnosis. The images had been collected for previous and ongoing research projects, using both standard cine imaging during rest and real-time imaging during exercise. Exercise was carried out using a supine bicycle ergometer (Lode, Groningen, Netherlands) during CMR imaging. Ethical approval was given by the Regional Ethical Review Board in Lund (Dnr 948/2018) and the Swedish Ethical Review Authority (Dnr 2021-05044).

Study IV

In this study, a dataset of short-axis CMR k-space data from 64 subjects collected at the Royal Free Hospital in London, United Kingdom was used for training, validation, and testing of the DL-based image reconstruction method. The collection and use of this data was approved by the National Research Ethics Committee in the United Kingdom (ref. 21/EE/0037, 17/LO/1499). Furthermore, to assess the performance of the developed MRI pulse sequence and reconstruction method, short-axis CMR data was collected in 10 healthy volunteers at Skåne University Hospital in Lund using both standard cine imaging and the developed method. This data collection was approved by the Swedish Ethical Review Authority (Dnr 2025-00334-02).

3.2 Magnetic resonance imaging (MRI)

In all studies except **Study IV**, magnetic resonance imaging (MRI) data had been collected prior to the studies and made available either for in-house or public use. The in-house datasets collected for **Study I** had been acquired with multiple MRI scanners from different vendors, including Siemens, Philips and General Electric, and at field strengths of 1.5 T and 3 T. The in-house data for **Study III** had been collected using a single Siemens MAGNETOM Aera 1.5 T system. In **Study IV**, MRI scanning was carried out on Siemens MAGNETOM Aera 1.5 T systems (Siemens Healthineers, Forchheim, Germany) at the Royal Free Hospital in London, United Kingdom and Skåne University Hospital in Lund, Sweden.

3.2.1 MRI pulse sequences

For the in-house datasets in **Study I**, short-axis CMR images had been acquired using different implementations of bSSFP sequences for breath-held ECG-gated Cartesian segmented cine imaging (see section 1.2.7), that are commonly applied for clinical practice at Skåne

University Hospital in Lund. This type of pulse sequence was also used for acquiring the reference short-axis CMR data in **Studies III** and **IV**. Typical imaging parameters for this type of sequence were TE/TR = 1.07/2.14 ms, temporal resolution = 40.96 ms, voxel size = $1.0 \times 1.0 \times 8$ mm^3 without a slice gap, and FA = 69°. However, these imaging parameters varied widely between the individual datasets collected from clinical practice.

For **Study III**, the subjects had also been imaged with non-Cartesian real-time imaging during exercise, carried out using a Siemens product bSSFP sequence (TrueFISP), with a radial read-out trajectory. Typical imaging parameters were TE/TR = 1.10/2.20 ms, temporal resolution = 35-37 ms, voxel size = $1.9 \times 2.8 \times 10.0$ mm^3 without a slice gap, and FA = 60°.

MRI pulse sequence development

In **Study IV**, a spiral SMS bSSFP sequence was developed for real-time CMR on a 1.5 T Siemens MAGNETOM Aera system. The development was largely inspired by existing sequences from literature [77, 25]. Development was carried out in the IDEA development environment (Siemens Healthineers, Forchheim, Germany). The sequence was implemented by modifying an existing bSSFP sequence that had been previously modified to permit variable density spiral k-space trajectories (see Figure 1.7). The two primary modifications that were performed were: (1) replacing the original single-band RF pulse with a SMS RF pulse; and (2) incorporating *blipped-CAIPI* gradient blips into the sequence [24, 77].

Figure 3.2 shows the pulse sequence as implemented in **Study IV**. The SMS RF pulse was based on a sinc RF pulse that underwent modification to be replicated at multiple slice positions [22]. A slice acceleration factor of 2 (SMS-2) was used. The blipped-CAIPI gradient blips were integrated into the slice selection re-winder and pre-winder of the bSSFP sequence [77]. A phase modulation pattern of $\pm\pi/2$ that alternated between TR:s was achieved through calculating the zeroth order gradient moment of the blips based on the distance between SMS slices according to Equation 1.8. Within each TR, the gradient blips integrated into the re-winder and pre-winder were of opposite polarity to balance the sequence, i.e. nulling the zeroth order gradient moment along the G_z axis at excitation.

Figure 3.3 shows the trajectory that was used. This was a variable density trajectory consisting of 8 spiral interleaves that were linearly spaced. An under-sampling factor of 1.6 was employed within the inner 20 % of the k-space radius and an under-sampling factor of 14.8 was employed for the area corresponding to the outer 70 % of the k-space radius. Between timeframes, all interleaves of the trajectory were rotated by a 5th tiny golden angle ($\approx 32.04^\circ$) [87]. The trajectory and rotational angle were decided experimentally.

The imaging parameters were TE/TR: 1.15 ms/4.5 ms, temporal resolution = 36.32 ms, voxel size = $1.7 \times 1.7 \times 8.0$ mm, FA = 100°, matrix size = 240×240 mm^2, FOV = 400×400 mm. 12-16 slices were collected to cover the left ventricle.

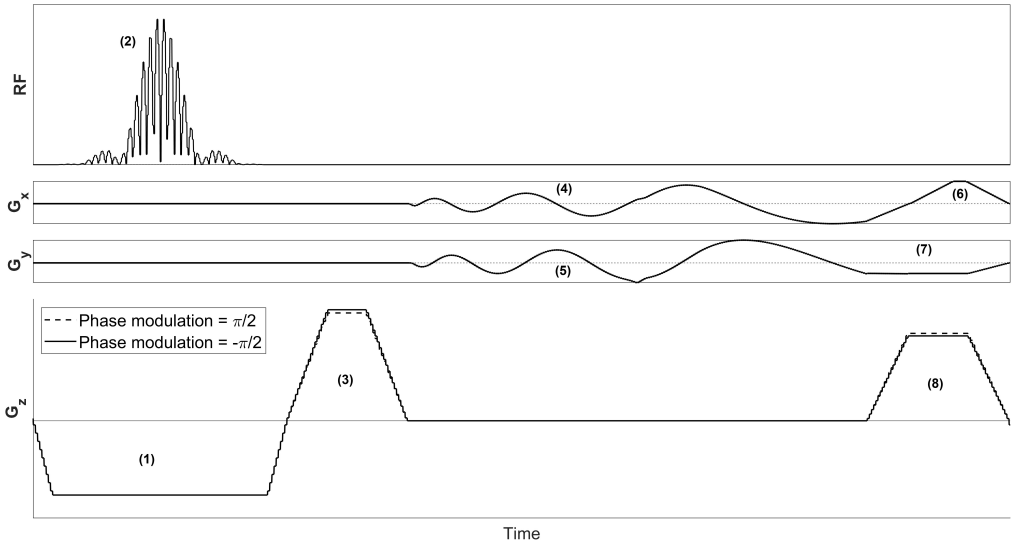


Figure 3.2: The simultaneous multi-slice (SMS) balanced steady state free precession (bSSFP) sequence developed in Study IV. The events are (1): gradient for slice selection; (2): radio-frequency (RF) pulse for SMS-2 imaging; (3): re-winder for the slice selection with integrated gradient blips (for phase modulations $\pi/2$ and $-\pi/2$); (4): spiral readout gradient along G_x; (5): spiral readout gradient along G_y; (6): re-winder for the spiral gradient along G_x (7): re-winder for the spiral gradient along G_y; (8): pre-winder for the slice selection gradient with integrated gradient blips (at phase modulations $-\pi/2$ and $\pi/2$). This figure was reproduced from the manuscript of Study IV.

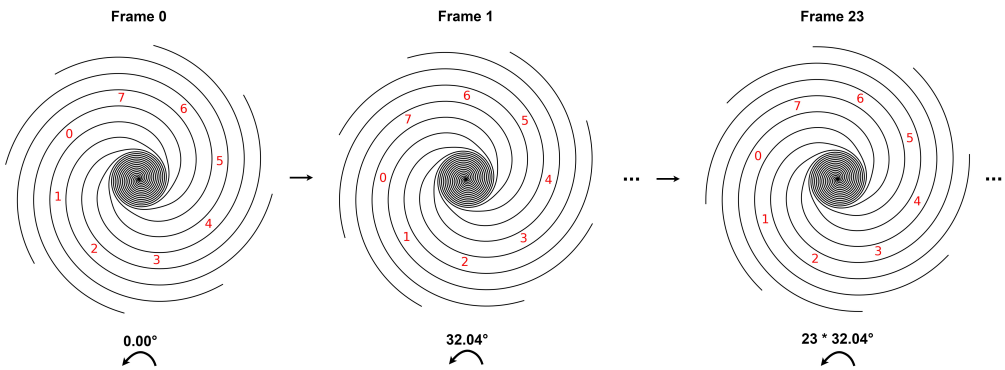


Figure 3.3: The variable density trajectory that was utilized for the developed pulse sequence in Study IV. The trajectory consists of 8 linearly spaced spiral interleaves that are rotated by a 5th tiny golden angle ($\approx 32.04°$) between each timeframe. This figure is reproduced from the manuscript of Study IV.

3.3 Image reconstruction

In all studies except **Study IV**, image reconstruction had been carried out prior to the studies. For standard CMR, the default reconstruction methods implemented clinically on the different scanners (often GRAPPA) had been used. For the real-time exercise CMR data in **Study III**, GRAPPA reconstruction with a reduction factor of 3 and a partial Fourier factor of 5/8 was used.

3.3.1 Image reconstruction development

In **Study IV**, a deep learning-based reconstruction method was developed for dynamic multi-coil real-time SMS-2 CMR data acquired using the SMS bSSFP sequence. A schematic diagram of this can be seen in Figure 3.4. During the SMS-2 acquisition, the gradient blips of the sequence introduced different phase modulation patterns to the k-spaces from the two different slices, and the SMS-2 k-space was the superposition of the k-space signals from these two slices. As a first step of the reconstruction, the SMS-2 k-space was de-modulated by applying the conjugate of the phase values that had been accrued during acquisition. This was done according to the process shown in Figure 1.10, where the de-modulation is carried out separately for the two different slices, producing constructive interference of signal from one slice and simultaneous destructive interference of signal from the other slice [27]. After phase de-modulation, singular value decomposition-based coil-compression was applied to compress the dynamic k-spaces to 10 coils, followed by density compensation [88]. Then, a NUFFT was applied to convert the two k-spaces to image space, producing dynamic multi-coil images with residual aliasing artifacts. A 3D U-Net for de-aliasing was then applied, producing dynamic cine images with suppressed aliasing artifacts as outputs. This reconstruction method was implemented as an online reconstruction at the scanner, using the Gadgetron framework [14]. This allowed image reconstruction to be done by sending k-space data to an external machine that runs the reconstruction and then sends the image data back to the scanner. A compressed sensing (CS) reconstruction method for SMS imaging was implemented based on existing code for CS in the TensorFlow MRI library, to allow performance comparisons with the developed reconstruction method. A cost function that employed an encoding matrix with blipped-CAIPI phase de-modulation was used, inspired by Tian et al. (2023) [77]. Separate regularization terms for spatial and temporal regularization using total variation were employed. The L-BFGS algorithm was used as an optimizer [12] because it was the only optimizer in the used library that was compatible with having two regularization terms.

3.4 Image analysis

Studies I, **II**, and **III** investigated automated image analysis tasks. **Study I** focused on image segmentation, while also using methods for image classification and image-based re-

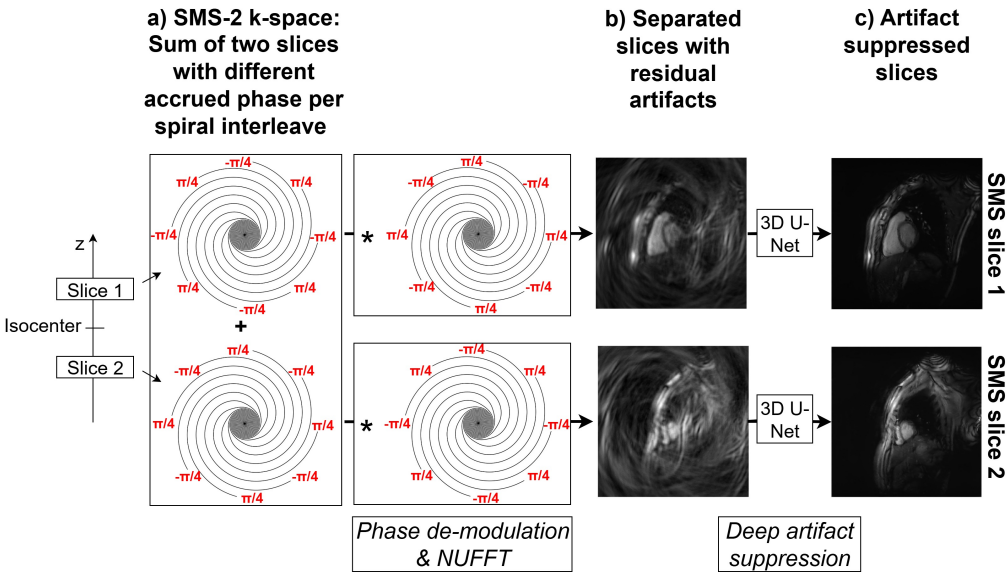


Figure 3.4: The image reconstruction method developed in Study IV. A) a SMS-2 k-space is acquired, corresponding to the superposition of signal from two different slices with different phase modulation patterns; B) phase de-modulation is carried out to separate the two slices in k-space by separately applying the conjugates of the phase values accrued during acquisition, and the slices are then transformed into image space using a non-uniform fast Fourier transform (NUFFT), resulting in dynamic multi-coil images with aliasing artifacts; C) a 3D U-Net is applied to the dynamic multi-coil images to suppress artifacts, producing coil-combined dynamic images with suppressed artifacts. This figure is reproduced from the manuscript of Study IV.

gression. **Study II** focused on image segmentation. **Study III** focused on the problem of retrospective image-based *gating* based on a method for image segmentation, and used the gating results for image processing. In **Studies I** and **III**, new methods were developed for automated and semi-automated image analysis. In **Study II**, no new image analysis methods were developed. Instead, the *nnU-Net* [32] was used, which is a deep learning-based open source framework for medical image segmentation that automatically sets parameters of the training process based on properties of the training data [32].

3.4.1 Image analysis development

Study I

In **Study I**, an automated pipeline for performing segmentation of the RV was developed. This can be seen in Figure 3.5. The pipeline took as input short-axis CMR images across all slice positions for a single cardiac phase. As an initial step, the input was re-sampled to a spatial resolution of 1.07×1.07 mm^2 using bi-linear interpolation. Then, three different CNN models were applied to the pre-processed input. The first CNN utilized a Darknet-19 architecture [36] to conduct binary classification for selecting the slices that contained an RV cross-section, as done in [89]. Only the slices containing a cross-section of the RV were used in the remaining steps of the pipeline. The second CNN used the same architecture to instead conduct regression on each slice, predicting the coordinates of the central point of the RV to allow cropping or padding of all slices to the same size (256×256 pixels). The final CNN used a 2D U-Net [34] architecture to perform segmentation of the RV in the cropped images, and the output segmentation masks were then post-processed to match the size of the input images. This pipeline was implemented into *Segment* [90], which is a freely available software for research. It was also implemented into *Segment CMR* [90], which is a commercial product for clinical use.

Study III

In **Study III**, a semi-automated pipeline was developed for constructing time-resolved cine images from 2D real-time exercise CMR images, synchronized in terms of respiratory and cardiac phases across slice positions. The pipeline can be seen in Figure 3.6. For each slice position, the initial step was to place a region of interest over the diaphragm, within which respiratory motion was estimated using a manifold learning-based method [71, 72, 73]. For this, the method implemented in Edlund et al. (2022) was used [74]. This produced a respiratory curve that allowed manually selecting a set of timeframes around a single end-expiratory state. Generally, the state containing the maximum amount of timeframes near end-expiration was selected. From these sampled timeframes, automated cardiac gating could optionally be carried out. This was based on: (1) utilizing a CNN [89] for segmentation of the LV blood pool [73]; (2) determining the cross-sectional area of the blood pool in all timeframes; and (3) identifying ED timeframes as peaks in the

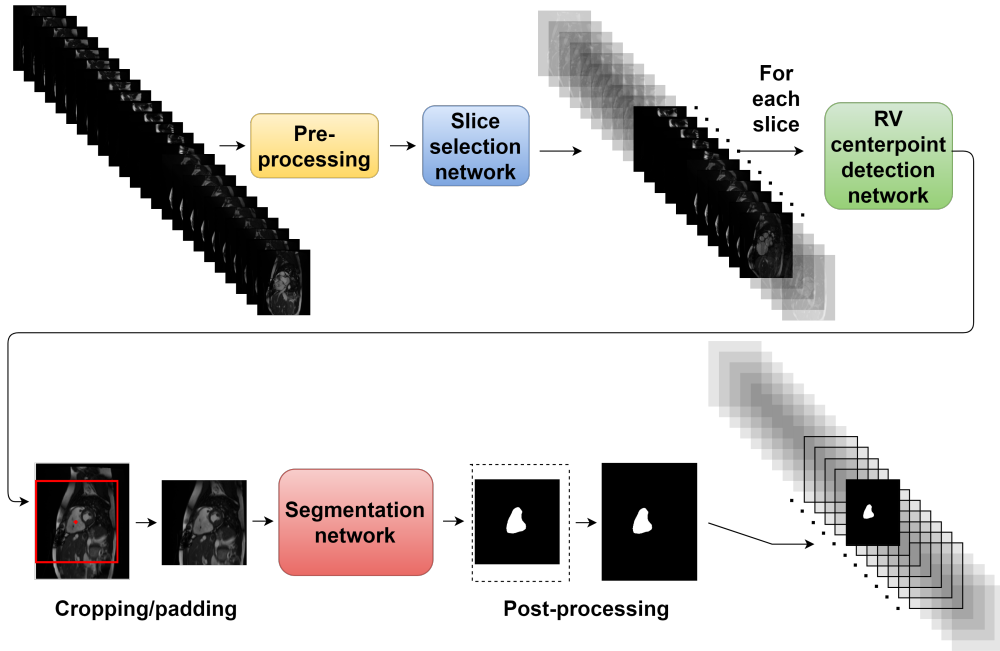


Figure 3.5: The pipeline for RV segmentation developed in Study I. The pipeline consisted of an initial pre-processing step for the input mutli-slice timeframe, followed by a network employed to select slices containing cross-sections of the RV (*slice selection network*). The selected slices were then given as input to a network for detecting the central point of the RV to allow cropping or padding, centered around the RV (*RV centerpoint detection network*). Finally, a network for segmenting the RV in the cropped images was employed (*segmentation network*), followed by post-processing. This figure is reproduced from Åkesson et al., Scientific Reports (2023) [91], licensed under a Creative Commons Attribution 4.0 International License.

area over time larger than their three surrounding timeframes in both directions. An ES timeframe was defined as the minimum area between two detected ED timeframes. From this automated suggestion of ED and ES timeframes, two consecutive ED timeframes and a corresponding ES timeframe were manually selected, and the interval between the ED timeframes was determined as the R-R interval. These steps were then repeated across all slice positions. Once R-R intervals had been defined for all slice positions, synchronization was done by re-sampling systole and diastole to the lowest number of timeframes within these intervals found across slice positions, through bi-linear interpolation. This allowed synchronizing the ED and ES timeframes, and thereby the R-R intervals, across all slice positions.

3.4.2 Manual image analysis

Manual image analysis was carried out by expert observers in **Studies I**, **III**, and **IV**. In **Studies I** and **III**, *Segment* [90] was used to conduct manual segmentation of the RV and LV, respectively. In **Study IV**, *OsiriX* [93] was used for manual segmentation of the LV.

3.5 Deep learning

Deep learning methods were employed in all four studies of this thesis. In **Study I**, all neural network training was carried out using the Deep Learning Toolbox in MATLAB R2019a and R2021a (Natick, Massachusetts: The MathWorks Inc.; 2019 and 2021). In **Study II**, all neural network training was carried out using PyTorch 1.12.0 + cu113 (Meta AI, New York City, NY, United States) and the nnU-Net framework [32]. In **Study III**, the previously trained neural network that was employed for image segmentation had been trained using the Deep Learning Toolbox in MATLAB R2019a [89]. In **Study IV**, all neural network training was carried out using Keras and TensorFlow 2.10.0 (TensorFlow, Google, Mountain View, CA, USA).

3.5.1 Network architectures

In **Study I**, a *Darknet-19* architecture [36] with an output layer modified for binary classification was used for slice selection. The same architecture but with an output layer modified for regression was used for RV centerpoint detection. A modified 2D U-Net architecture [34] was used for RV segmentation. In **Study II**, a 2D U-Net architecture [34] was used for the various image segmentation tasks, as implemented in the nnU-Net framework [32]. In **Study III**, the architecture from Bai et al. (2018) [63] was employed for LV segmentation [89], which is similar to a U-Net but with only a single up-sampling step. In **Study IV**, a modified implementation of the 3D U-Net architecture [35] was used for the suppression of aliasing artifacts.

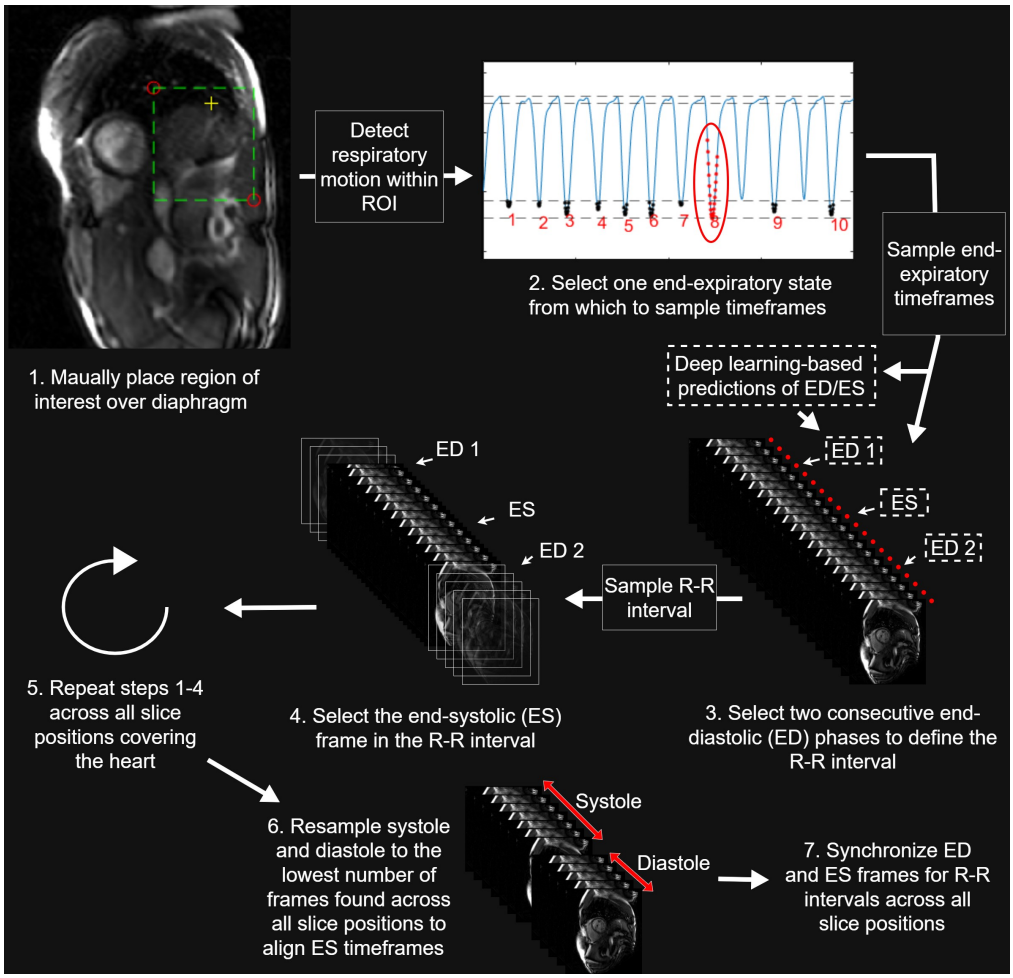


Figure 3.6: Overview of the semi-automatic pipeline for constructing synchronized time-resolved cine images from real-time exercise CMR images. In this: (1) a region of interest was placed manually across the diaphragm and respiratory motion was detected through manifold learning, producing a respiratory curve; (2) from the respiratory curve, timeframes near a single end-expiratory state (red dots) were selected, and deep learning-based detection of ED and ES timeframes could optionally be carried out; (3) two adjacent ED timeframes were identified; (4) the ES timeframe between the ED timeframes was identified; (5) all previous steps were repeated for all slice positions; (6) the R-R intervals across all slice positions were resampled to the lowest number of timeframes within the two respective intervals across all slice positions; and (7) ED and ES were aligned across slice positions. This figure is reproduced from Åkesson et al., Clinical Physiology and Functional Imaging (2025) [92], licensed under a Creative Commons Attribution 4.0 International License.

3.5.2 Training of neural networks

Training of neural networks was done in **Studies I**, **II**, and **IV**. The following sections describe the methods that were used for training.

Loss functions

In **Study I**, a cross-entropy loss function was used for the slice selection network, a mean squared error loss function was used for the RV centerpoint detection network, and a Dice loss function was used for the RV segmentation network. In **Study II**, the sum of a Dice loss function and a cross-entropy loss function was used for the various image segmentation tasks. In **Study IV**, a structural similarity index measure (SSIM) loss function was used for the network employed for the suppression of aliasing artifacts.

Optimizers

In **Study I**, a stochastic gradient descent with momentum optimizer was used for the slice selection network, and an Adam [30] optimizer was used for the RV centerpoint detection network and the RV segmentation network. In **Study II**, a stochastic gradient descent with momentum optimizer was used for the various segmentation tasks. In **Study IV**, an Adam [30] optimizer was used for the network employed for the suppression of aliasing artifacts.

Hardware

In **Studies I** and **II**, all training was performed on either an NVIDIA GeForce RTX 3090 (24 GB memory) GPU or an NVIDIA TITAN RTX GPU (24 GB memory) GPU. In **Study IV**, training was performed on a single NVIDIA RTX A6000 (48 GB memory) GPU (NVIDIA, Santa Clara, California, United States).

3.6 Evaluation and statistical analysis

Study I

In this study, analyses were conducted to evaluate (1): the ability of the developed pipeline to reduce the manual time for producing segmentations of clinically useful quality; and (2): the performance of the pipeline for segmentation and quantification of volumes and ejection fraction. All statistical analyses were performed using MATLAB R2019a or R2021a.

The assessment of time reduction was carried out by two expert observers (O1 and O2). First, the pipeline was used to automatically generate RV segmentations on the clinical validation set comprised of 50 subjects, and the time for each segmentation was measured. Then, O1 provided subjective ratings of the quality of automated segmentations according to a modified Likert scale as either (A) *sufficient for clinical use*; (B) *needing minor corrections*; or (C) *needing major corrections*. Then, O1 and O2 carried out corrections of the

segmentations that had been rated as category B or C, and the time for each correction was measured. Finally, fully manual segmentations were also carried out by O1 and O2 and times were measured. To assess statistical significance in time reduction, the two-sided *Wilcoxon signed-rank test* was used.

To evaluate segmentation performance, Dice scores and Hausdorff distances were calculated between automated and manual segmentations. To evaluate agreements in the quantification of volumes and ejection fraction, *Bland-Altman* analysis [94] and *Spearman's rank correlation coefficients* (r) were employed between automated and manual segmentations. These methods were also used to assess inter-observer variability between O1 and O2.

Study II

In this study, analyses were conducted to evaluate how randomness affects the reliability of standard methods for comparing the performance of deep learning-based medical image segmentation models. All statistical analyses were carried out using MATLAB R2021a.

As an initial step, recent literature was sampled to provide an overview of standard comparison methods. It was found that *hold-out validation* and *5-fold cross-validation* were the most common performance estimation methods, and the *paired t-test* and *Wilcoxon signed-rank test* were the most common statistical methods. For details regarding this literature analysis, the reader is referred to the publication [31].

To evaluate the effects of randomness on the reliability of these methods, three different multiclass 3D segmentation problems were addressed. Specifically, the problems included brain tumour segmentation, hippocampus segmentation, and short-axis CMR segmentation. For each segmentation problem, the same nnU-Net [32] based learning algorithm was run 50 times for 100 epochs using different random seeds. This was done for both hold-out validation and 5-fold cross-validation. For hold-out validation, this resulted in 50 different segmentation models, whose performance was estimated on the hold-out test set. For 5-fold cross-validation, this resulted in 50 different sets, each containing 5 segmentation models, whose performance was estimated across all five folds and concatenated into one array of performance measurements.

To measure segmentation performance, the 3D Dice score was used. The *best-performing seed* for each segmentation problem could be decided as the seed that provided the highest mean Dice score across images. Then, using both the paired t-test and the Wilcoxon signed-rank test, statistically significant differences in performance were assessed between the best-performing seed and the 49 remaining seeds. Statistical significance was defined as a two-tailed p value < 0.05. Results from this were evaluated by computing the percentage of *outperformances* between seeds, defined as a statistically significant difference in performance given by at least one of the two statistical methods.

Study III

In this study, analyses were conducted to (1): investigate the possibility to use the developed retrospective synchronization method for deriving time-resolved measures of cardiac function in heart failure patients and healthy volunteers; and (2) investigate the performance of the deep learning-based method for automated detection of ED and ES. All statistical analyses were performed using MATLAB R2022a.

The developed method was used by an expert observer to construct retrospectively synchronized real-time exercise CMR cines for 10 heart failure patients and 10 healthy volunteers. Then, manual segmentation of the LV was carried out by the expert observer in both exercise CMR cines and standard CMR cines at rest. LV segmentation was conducted at end diastole (ED), mid systole (MS), end systole (ES), and early rapid filling (ERF) in images from exercise, and at ED and ES for images at rest. The MS and ERF phases were chosen due to being the most challenging phases to segment in the cardiac cycle, thereby serving as strong indicators of the feasibility of time-resolved measurements. From the segmentations, the LV mass (LVM) was calculated, and agreements in LVM between rest and exercise were assessed for all cardiac phases using Bland-Altman analysis [94].

DL-based automated selections of ED and ES timeframes were compared to manual selections carried out by the expert observer during the construction of cines. Comparisons were done for the most central midventricular slice in each subject, and automated selections were made within the samples of timeframes that had been manually selected near a single end-expiratory state by the expert observer. Automated and manual selections were compared by calculating the mean and standard deviation of the difference in temporal position between corresponding timeframes (T), $\Delta T = T_{automatic} - T_{manual}$.

Study IV

In this study, analyses were conducted to evaluate the performance of the developed methods by: (1) comparing the clinical accuracy and image quality of the developed method for DL-reconstructed real-time bSSFP SMS CMR (RT-SMS) with that of standard breath-hold ECG-gated CMR (standard CMR); and (2) comparing reconstruction performance and runtime between the DL-based reconstruction method (DL) and compressed sensing (CS).

Image quality was compared between RT-SMS and standard CMR at ED and ES through a plot for qualitative visual comparison. Clinical accuracy was assessed using Bland-Altman analysis [94] for evaluating agreements in manually derived volume-based measures between RT-SMS and standard CMR. The manually derived measures were LV endocardial volumes at ED (LVEDV) and at ES (LVESV) as well as stroke volume (SV) and ejection fraction (EF). These measures were obtained through manual segmentation performed by an expert observer with 25 years of experience.

Reconstruction performance was qualitatively compared between DL and CS through visual assessments of sharpness and motion. This was done by plotting the intensity over

time for a vertical line of pixels in dynamic images of 4 different subjects reconstructed with DL and CS. The runtimes of the two reconstruction methods were also evaluated.

Chapter 4

Results and comments

4.1 Study I

The results from this study indicated that the developed deep learning-based pipeline for RV segmentation could substantially reduce the time for producing segmentations of clinically useful quality. Its performance was on par with inter-observer variability for volume-based measures and Dice scores, and it further indicated the ability to decrease inter-observer variability.

4.1.1 Time reduction

On the clinical validation set of 50 subjects, the average time for a fully manual segmentation of the RV in ED and ES was 6 min and 46 s for O1 and 5 min and 19 s for O2. The corresponding average automated segmentation time for the pipeline was 2 s. The subjective ratings of the automated segmentations showed that 58 % of automated segmentations were of category A (*sufficient for clinical use*) and the remaining 42 % were of category B (*needing minor corrections*), while none were rated to be of category C (*needing major corrections*). The average time for a manual correction was 1 min and 38 s for O1 and 2 min and 38 s for O2, both significantly shorter than the corresponding manual delineation times ($p < 0.05$). The average reduction in time given by corrected automated segmentations compared to fully manual segmentations was 5 min and 17 s, which was 87 % of the average manual segmentation time between the two observers.

This indicated that the pipeline could provide a substantial time reduction even though manual corrections were needed. Nonetheless, since the pipeline was trained on data from a single centre and evaluated within the same centre by observers that use the same segmentation guidelines [40], it is possible that less of a time reduction would be given when used in centres that apply different guidelines. This was also indicated by the somewhat reduced performance during the external validation presented in the publication [91].

4.1.2 Segmentation and quantification of volumes

On the clinical validation set of 50 subjects, the average Dice scores ($\pm$ SD) were 0.91 ± 0.02 at ED and 0.87 ± 0.04 at ES between O1 and the pipeline, 0.91 ± 0.02 at ED and 0.88 ± 0.03 at ES between O2 and the pipeline, and 0.91 ± 0.03 at ED and 0.86 ± 0.04 at ES between O1 and O2. This indicated that the inter-method variability was on par with the inter-observer variability in terms of Dice score.

Figure 4.1 shows Bland-Altman and correlation analysis for ED and ES volumes between the pipeline and the two observers. The limits of agreement between the pipeline and the two observers were more narrow than the limits of agreement between observers. A comparison of the inter-observer variability between O1 and O2 for fully manual segmentations and corrected automated segmentations can be seen in Figure 4.2. The narrower limits of agreement for corrected automated segmentations indicates that the use of the automated segmentation pipeline as a starting point for segmentation can decrease inter-observer variability.

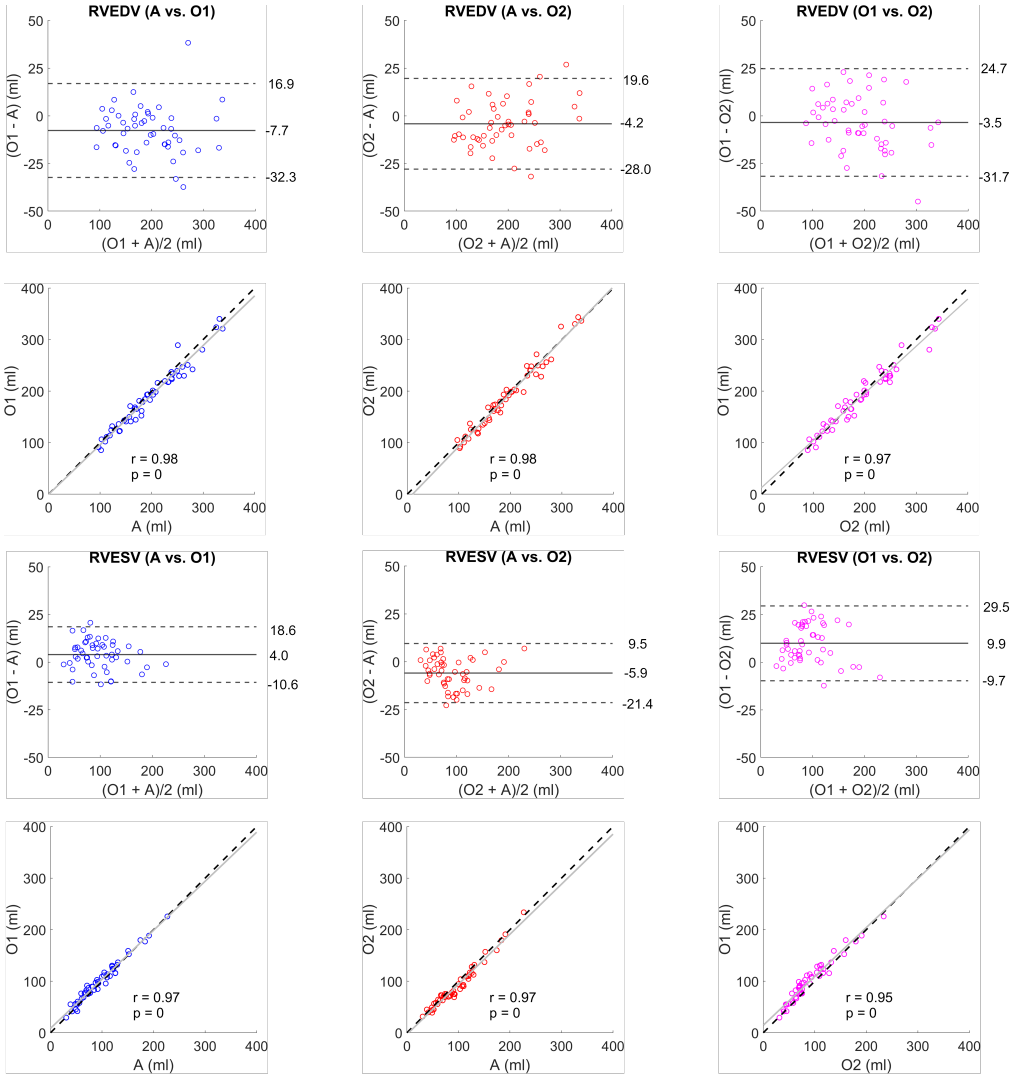


Figure 4.1: Bland-Altman and correlation plots showing agreements in RV volumes at ED (RVEDV) and ES (RVESV) between the automated pipeline (A) and Observer 1 (O1) (left column), between A and Observer 2 (O2) (middle column), and between O1 and O2 (right column). The Bland-Altman plots show bias and limits of agreement (± 1.96 SD). The correlation plots show least squares lines (grey) and Spearman's rank correlation coefficients (with p values). This figure is reproduced from Åkesson et al., Scientific Reports (2023) [91], licensed under a Creative Commons Attribution 4.0 International License.

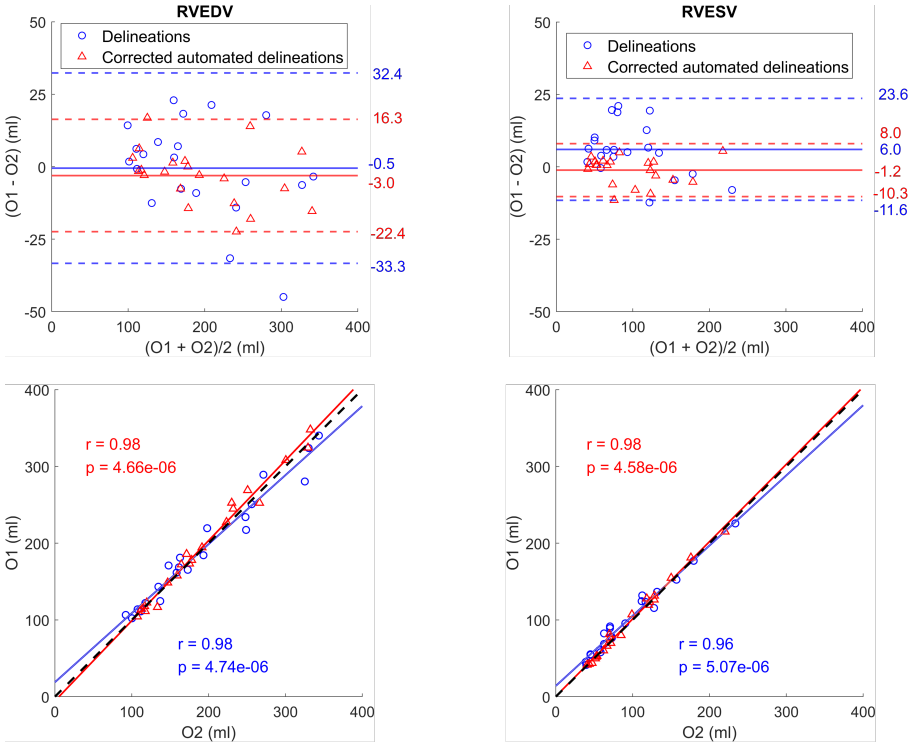


Figure 4.2: Bland-Altman and correlation plots showing agreements in RV volumes at ED (RVEDV, left column) and ES (RVESV, right column) between the two observers (O1 and O2). This indicates the inter-observer variability for fully manual delineations (blue) and corrected automated delineations (red). The Bland-Altman plots show bias and limits of agreement (± 1.96 SD). The correlation plots show least squares lines (grey) and Spearman's rank correlation coefficients (with p values). This figure is reproduced from Åkesson et al., Scientific Reports (2023) [91], licensed under a Creative Commons Attribution 4.0 International License.

4.2 Study II

The results from this study indicated that randomness, induced by simply varying the random seed, can cause statistically significant differences in performance between deep learning-based medical image segmentation models from the same learning algorithm when using standard methods for performance comparison.

4.2.1 Literature overview

Figure 4.3 shows the results from the literature overview. 460 studies were assessed, and among these studies, 399 (87 %) had presented performance comparisons between segmentation models. Out of these, 92 (23 %) had used statistical hypothesis testing for comparisons. The two most common methods for this were the paired t-test and the Wilcoxon signed-rank test. It was also noted that hold-out validation was the most common performance estimation method, followed by cross-validation. Within this category, 5-fold cross-validation was the most common.

4.2.2 Experiments

Figure 4.4 shows 3D Dice scores across images of the hold-out test set from the 50 repeated hold-out validation procedures using 50 different seeds for all three segmentation problems. It also highlights the best and worst performing seeds. This plot indicates that when varying the random seed, the different resulting models show large variations in performance on a select few images, while remaining stable in performance across most other images. The best-performing seed generally performed well across the few images with a high variance in performance across seeds, and the best-performing seed varied between classes for the same segmentation problem.

Figure 4.5 shows histogram plots of the distributions of Dice scores and the results from the pairwise statistical comparisons between the best-performing seed and the 49 remaining seeds when using hold-out validation. This shows that best-performing seed statistically significantly ($p < 0.05$) outperformed 16 %, 16 %, and 23 % of models for the three Brain Tumour classes, 23 %, and 23 % for the two Hippocampus classes, and 75 %, 76 %, and 0 % for the three CMR classes. Figure 4.6 shows histogram plots of the distributions of Dice scores and the results from the pairwise statistical comparisons between the best-performing seed and the 49 remaining seeds when using 5-fold cross-validation. This shows that the best-performing seed statistically significantly ($p < 0.05$) outperformed 35 %, 20 %, and 27 % of models for the three Brain Tumour classes, 23 %, and 38 % for the two Hippocampus classes, and 18 %, 20 %, and 10 % for the three CMR classes.

These results indicate that the choice of random seed impacts the training of a single learning algorithm enough to produce models that can outperform up to 76 % of other seeds with statistical significance, using the most widely used statistical methods in liter-

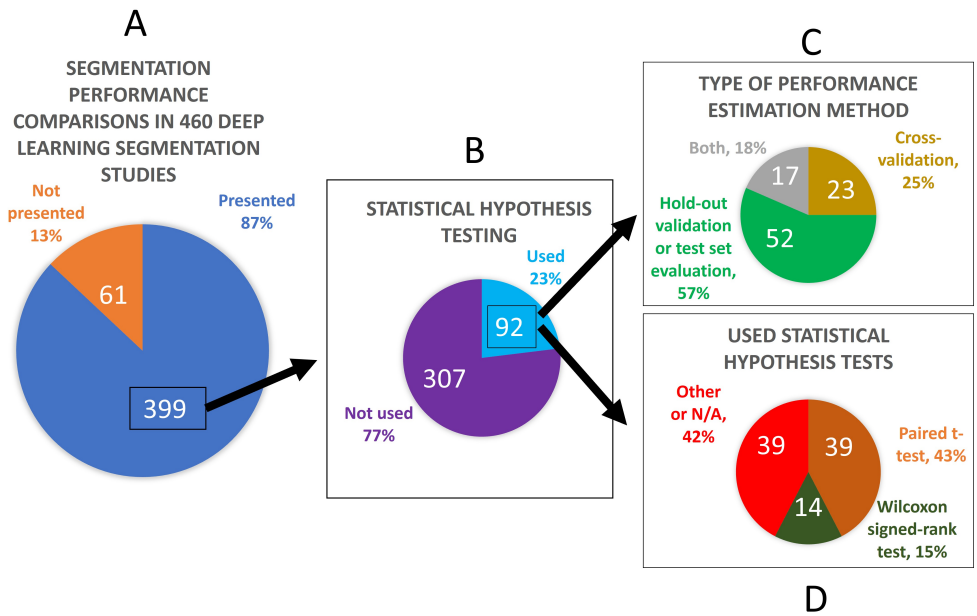


Figure 4.3: The results from the literature overview conducted in Study II. Of the 460 assessed studies, 399 (87 %) had presented performance comparisons. Out of these, 92 (23 %) had used statistical hypothesis testing for comparisons. The two most common methods for this were the paired t-test and the Wilcoxon signed-rank test. Hold-out validation and cross-validation were the most common performance estimation methods. This figure is reproduced from Åkesson et al., Computers in Biology and Medicine (2024) [31], licensed under a Creative Commons Attribution 4.0 International License.

ature. This affects the reliability of these methods in the context of measuring true performance differences between different learning algorithms, since performance differences of statistical significance can arise from changing the random seed within a single learning algorithm. Furthermore, the results indicate that using 5-fold cross-validation for performance estimation does not eliminate this effect.

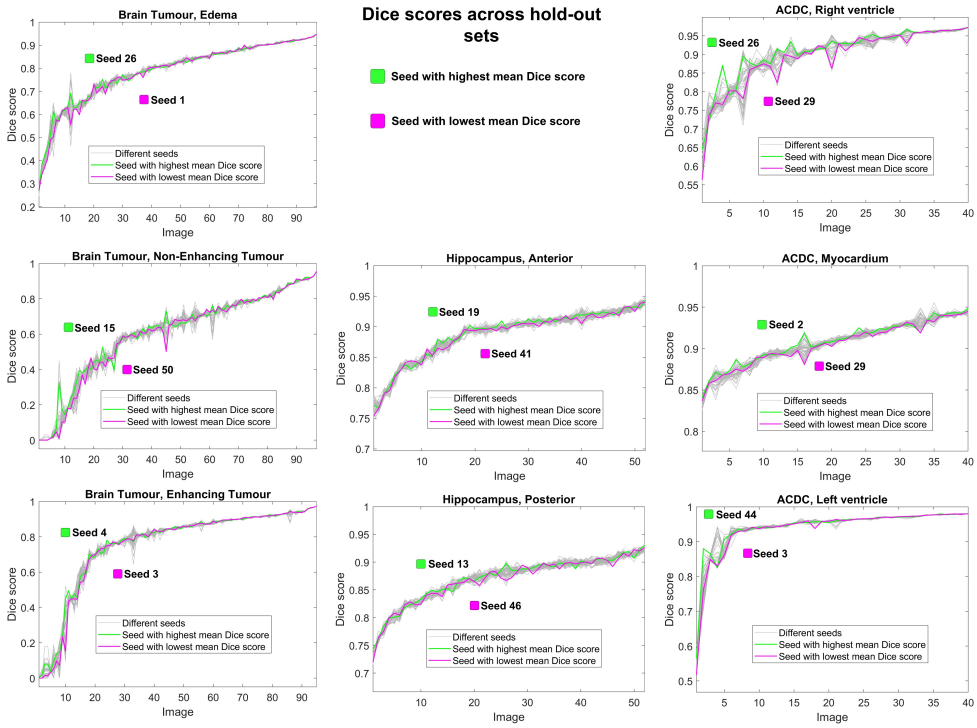


Figure 4.4: This shows the Dice scores across images in the hold-out validation test sets, obtained for all classes of the three different segmentation problems, when repeatedly running learning algorithms for 50 different seeds in Study II. Each subplot shows the results from the 50 different seeds in gray, and highlights the seed with the highest mean Dice score in green and the seed with the lowest mean Dice score in magenta. This figure is reproduced from Åkesson et al., Computers in Biology and Medicine (2024) [31], licensed under a Creative Commons Attribution 4.0 International License.

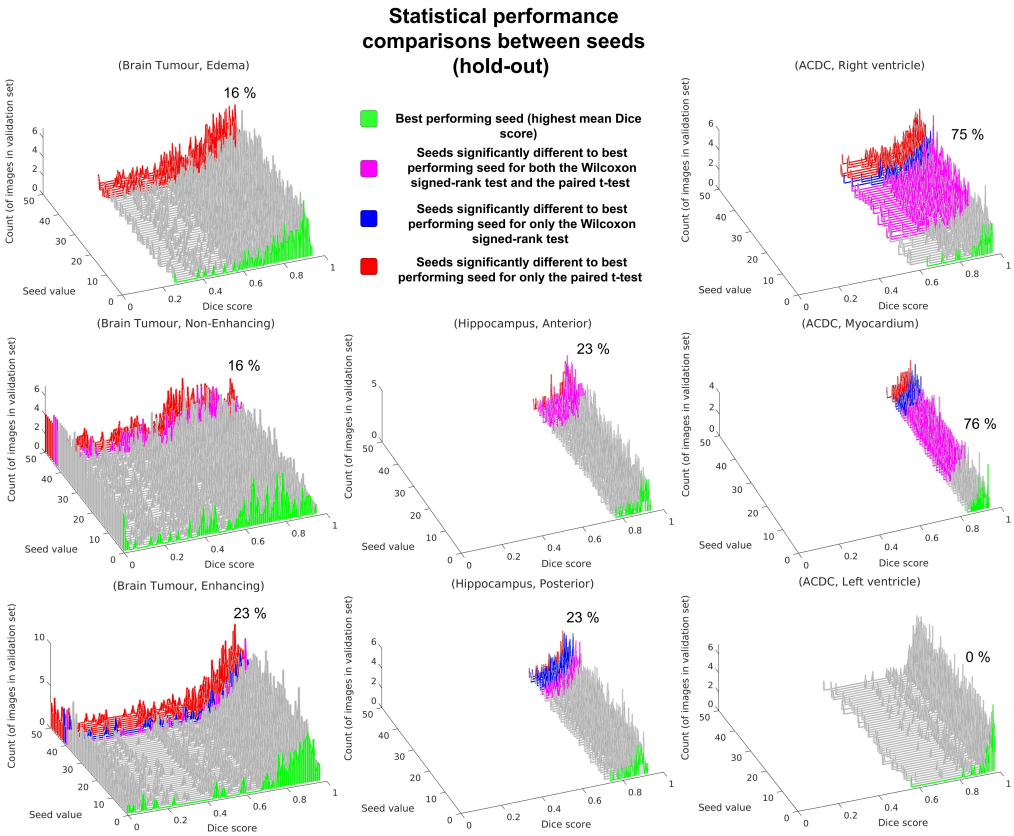


Figure 4.5: Histogram plots showing the distributions of Dice scores for the 50 different seeds for all classes of all segmentation problems assessed in Study II, when using hold-out validation. The best-performing seed is highlighted in green. The seeds that did not show a statistically significant difference to the best-performing seed from a pairwise statistical comparison are shown in gray. The seeds that did show a statistically significant difference according to one or both of the two employed statistical hypothesis tests are shown in purple, blue or red, and the percentages of outperformances are indicated above the histogram plots. This figure is reproduced from Åkesson et al., Computers in Biology and Medicine (2024) [31], licensed under a Creative Commons Attribution 4.0 International License.

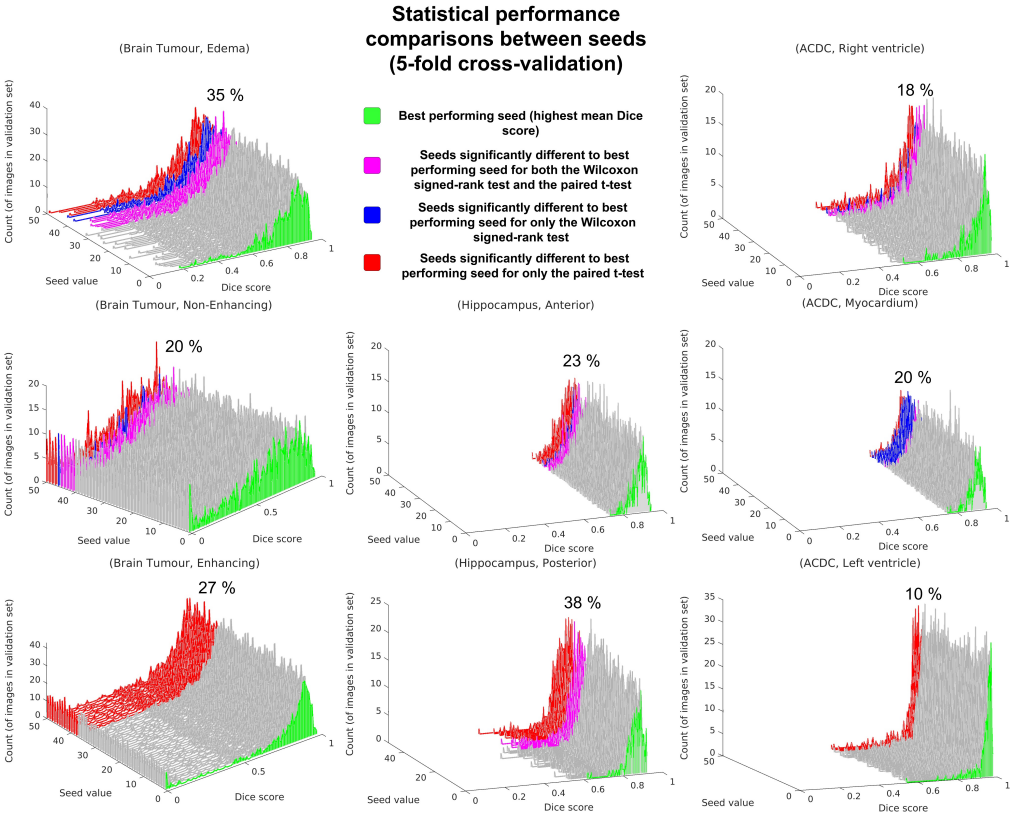


Figure 4.6: Histogram plots showing the distributions of Dice scores for the 50 different seeds for all classes of all segmentation problems assessed in Study II, when using 5-fold cross validation. The best-performing seed is highlighted in green. The seeds that did not show a statistically significant difference to the best-performing seed from a pairwise statistical comparison are shown in gray. The seeds that did show a statistically significant difference according to one or both of the two employed statistical hypothesis tests are shown in purple, blue or red, and the percentages of outperformances are indicated above the histogram plots. This figure is reproduced from Åkesson et al., Computers in Biology and Medicine (2024) [31], licensed under a Creative Commons Attribution 4.0 International License.

4.3 Study III

The results from this study indicated that the developed method for constructing synchronized time-resolved cine images from 2D real-time exercise CMR could be used for deriving time-resolved measures of cardiac function in heart failure patients and healthy volunteers. This was indicated through reasonable agreements in LVM with ECG-gated rest CMR. Assessments of the method for automated detection of ED and ES showed the ability to rapidly produce predictions of ED and ES, but indicated that this needs further improvements apically and basally.

4.3.1 Utility for time-resolved measures

Figure 4.7 shows Bland-Altman plots indicating agreements in LVM between ECG-gated rest CMR and the proposed method at ED, MS, ES, and ERF. Figure 4.8 shows Bland-Altman plots indicating the agreement in LVM for ECG-gated rest CMR between ED and ES. These plots show that the limits of agreement between ECG-gated rest CMR and the proposed exercise CMR method are within an order of magnitude of the limits of agreement between ED and ES for ECG-gated rest CMR, although the latter are visibly more narrow. This was the case for both heart failure patients and healthy volunteers. This indicates that, despite some deviation, the anatomical information is preserved when using the proposed method for all subjects. Furthermore, the agreement in LVM at the MS and ERF phases was on par with the agreement in LVM at the ED and ES phases, indicating the feasibility to conduct time-resolved volumetric measurements across all phases in cines constructed with the proposed method.

4.3.2 Automated detection of ED and ES

When evaluating the DL-based method for automated detection of ED and ES, the method failed at finding two consecutive ED phases within the selected sample of timeframes in 4 of 20 subjects. For the remaining 16 subjects, the mean $\pm$ SD of the difference in temporal position between corresponding timeframes were 0.13 ± 0.91 timeframes for ED and 1.19 ± 1.60 timeframes for ES. The average time for automatically detecting two consecutive ED phases and one ES phase within an average of 32 timeframes was 4 seconds using a laptop with a GPU (NVIDIA Quadro T1000). These results indicate that when the method does not fail, it can rapidly provide reasonably accurate predictions of ED and ES frames in midventricular slices, which could speed up the process of synchronization for the user. The method failed due to an insufficient number of selected timeframes (less than 20 timeframes in 3 out of 4 failed cases), which increased the difficulty of finding two consecutive ED timeframes for the method. Furthermore, the method performed poorly in apical and basal slices. Thus, further improvements are needed.

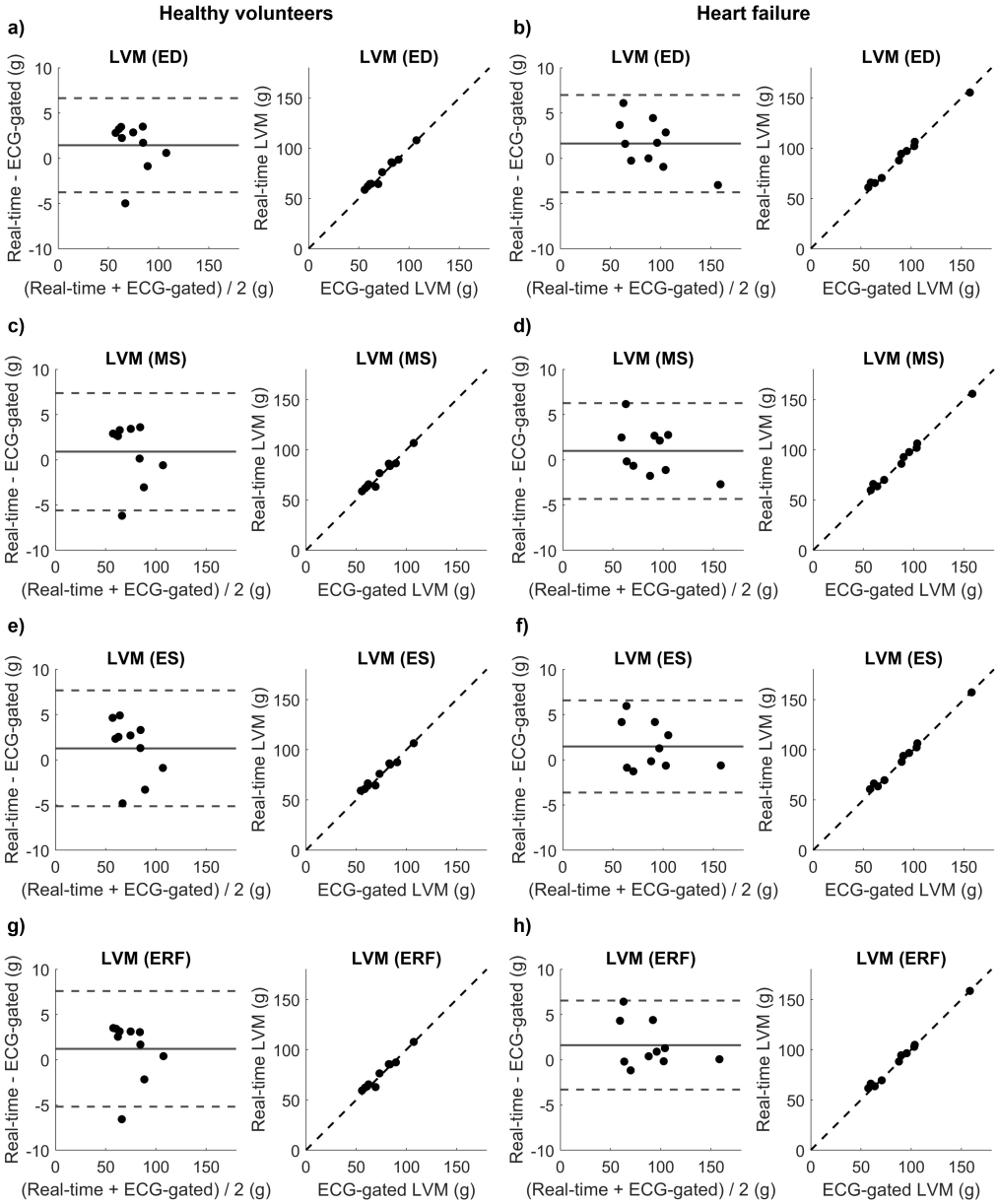


Figure 4.7: Bland-Altman and correlation plots between real-time CMR during exercise (*Real-time*) and ECG-gated CMR during rest (*ECG-gated*) in terms of left ventricular mass (LVM) at end diastole (ED), mid systole (MS), end systole (ES), and early rapid filling (ERF). This is shown for healthy volunteers (left column) and heart failure patients (right column). This figure is reproduced from Åkesson et al., Clinical Physiology and Functional Imaging (2025) [92], licensed under a Creative Commons Attribution 4.0 International License.

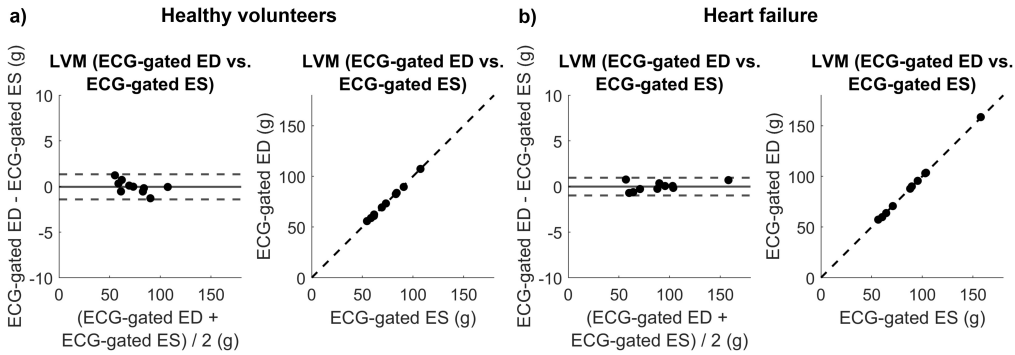


Figure 4.8: Bland-Altman and correlation plots between ECG-gated CMR during rest at end diastole (ED) and end systole (ES) in terms of left ventricular mass (LVM) for healthy volunteers (left) and heart failure patients (right). This figure is reproduced from Åkesson et al., Clinical Physiology and Functional Imaging (2025) [92], licensed under a Creative Commons Attribution 4.0 International License.

4.4 Study IV

The results from this study indicated that the developed real-time SMS-2 bSSFP sequence (RT-SMS) with online DL-based reconstruction for image artifact suppression could achieve a reasonable agreement with standard CMR in terms of volume-based clinical measures, and that the online DL-based reconstruction could substantially reduce reconstruction time compared to CS. However, the results also indicated that further improvements in image quality are required before the method can be applied in clinical practice.

4.4.1 Comparisons with standard CMR

Figure 4.9 shows CMR images at ED and ES collected using both RT-SMS and standard CMR at 16 slice corresponding positions. This shows that compared to standard CMR, the blood-myocardium contrast is reduced when using RT-SMS. Also, there is visible blurring of the borders of the blood pool that is more substantial during systole than diastole, and also a blurring and in-plane shift of fat signal visible in all RT-SMS slices.

Figure 4.10 shows Bland-Altman and correlation plots between RT-SMS and standard CMR for LV ED volume (LVEDV), LV ES volume (LSESV), stroke volume (SV) and ejection fraction (EF). The bias $\pm$ SD (in ml and in percentages of standard CMR volumes) were -13.2 ± 8.0 ml (-7 ± 4 %) for LVEDV, 2.1 ± 5.9 ml (5 ± 12 %) for LVESV, and -15.1 ± 8.1 ml (-14 ± 6 %) for SV. The bias $\pm$ SD of EF were -3.9 ± 3.3 %. This shows that when using RT-SMS, the LVEDV is largely underestimated while the LVESV is overestimated. This could be due to the visible blurring and reduced contrast. Despite the visibly higher amount of blurring at ES, LVESV showed a lower bias and standard deviation than LVEDV. This indicates that further improvements of the sequence or reconstruction

are needed to reduce blurring and improve contrast. However, the acquisition time of the standard CMR was several minutes using multiple breath-holds, while the acquisition time of the RT-SMS was around 12-16 seconds during free-breathing.

4.4.2 Comparisons with compressed sensing (CS)

Online DL reconstruction required 31 seconds of GPU reconstruction time for a 12-slice dataset (6 SMS-2 slices) with 34 coil elements and 25 timeframes. The offline CS reconstruction required 16 min and 9 seconds of GPU reconstruction time for the same dataset. Thus, the DL accelerated the reconstruction by a factor of 31.26 compared to CS. However, it is possible that a more efficient implementation using another optimizer or a more optimal set of parameters could speed up the CS reconstruction.

Figure 4.11 shows RT-SMS data acquired in four different subjects reconstructed using both DL and CS, shown as diastolic timeframes as well as plots of intensity over time for one R-R interval. This indicates an agreement in motion and anatomy between CS and DL. However, the CS reconstruction appears less sharp, both in the diastolic timeframes and the motion plots. This could be because the used k-space trajectory was optimized for the DL reconstruction, with high levels of under-sampling in the outer parts of k-space, possibly causing the blurring of CS. It also indicates the sharpening properties of the DL method. It is possible that for a trajectory more optimized for CS, the CS reconstruction could provide sharper results.

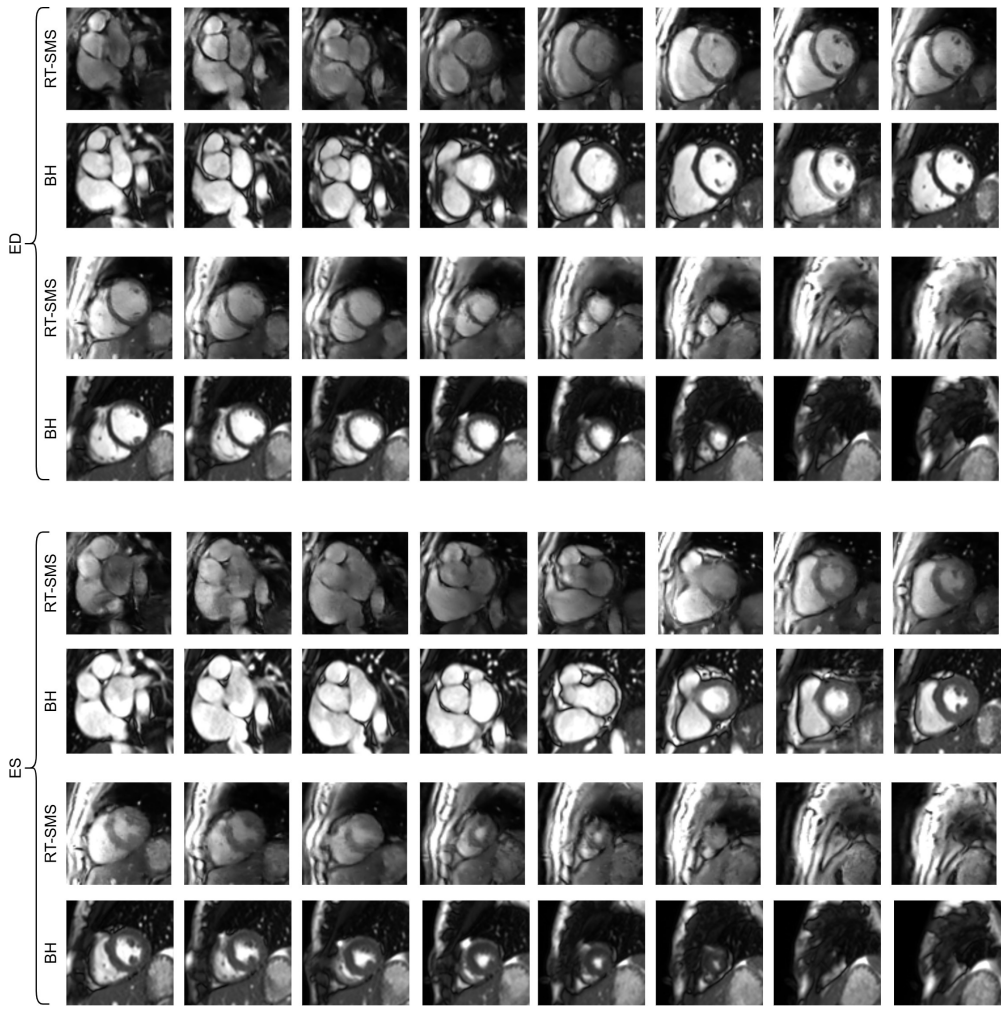


Figure 4.9: CMR images in one healthy volunteer at end diastole (ED, first 4 rows) and end systole (ES, last four rows) collected using both the real-time SMS-2 bSSFP sequence with DL reconstruction (RT-SMS) and standard breath-hold ECG-gated CMR (BH) at 16 corresponding slice positions. This figure is reproduced from the manuscript of Study IV.

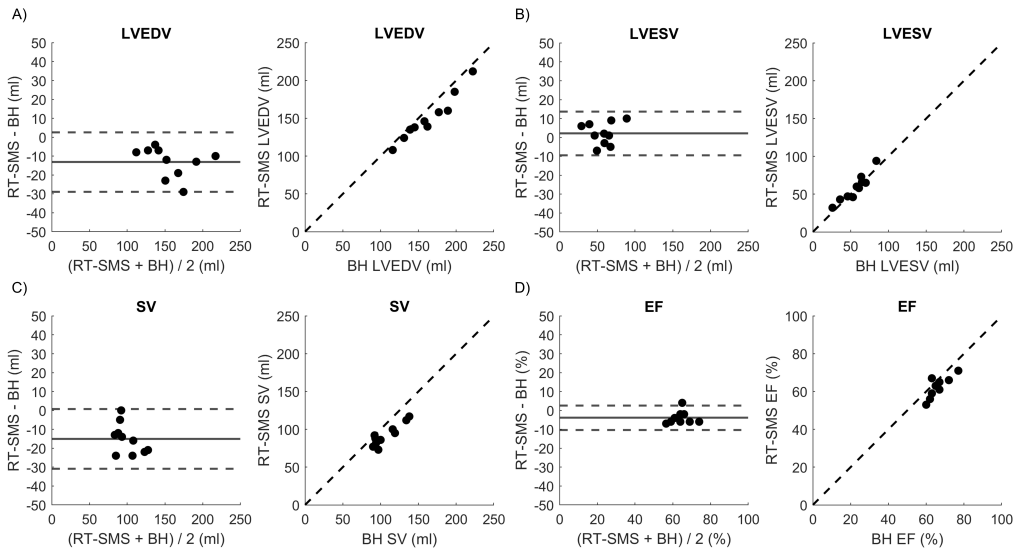


Figure 4.10: Bland-Altman and correlation plots between the real-time SMS-2 bSSFP sequence with DL reconstruction (RT-SMS) and standard breath-hold ECG-gated CMR (BH) for A) LV ED volume (LVEDV), B) LV ES volume (LSESV), C) stroke volume (SV), and D) ejection fraction (EF). This figure is reproduced from the manuscript of Study IV.

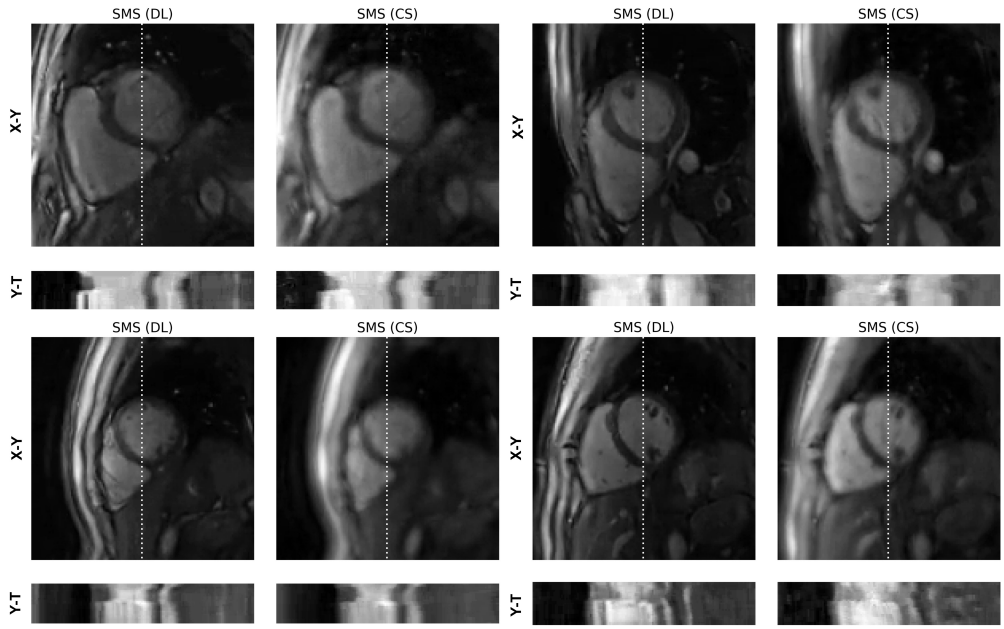


Figure 4.11: Real-time SMS-2 bSSFP data in four different subjects reconstructed using deep learning-based image artifact suppression (DL) and compressed sensing (CS). This is shown as X-Y plots of diastolic timeframes as well as Y-T plots of intensity over time in the positions indicated by the dotted lines for one R-R interval. This figure is reproduced from the manuscript of Study IV.

Chapter 5

Conclusions and future directions

In the studies of this thesis, several different aspects of the application of deep learning (DL) in the process of evaluating cardiac function using CMR were assessed. The benefits of DL were evaluated in several parts of this process together with other new developments in CMR. Specifically, it was investigated whether DL could speed up the clinical segmentation process for the right ventricle (RV) (**Study I**), help in synchronizing time-resolved cine images from exercise CMR (**Study III**), and reconstruct real-time SMS images through image artifact suppression (**Study IV**). Furthermore, the impact of randomness when developing DL methods for medical image segmentation was assessed (**Study II**). Below follows the conclusions of the four studies and discussions on future directions.

Study I

Conclusion: The clinical applicability of the developed deep learning-based pipeline for RV segmentation was indicated through its ability to reduce the time needed to obtain segmentations of clinically useful quality compared to segmentations performed manually. The average time reduction was 87 %, but the actual time reduction may vary between users that employ different segmentation guidelines.

Future directions: To further study the general clinical utility of this specific pipeline, it would be of relevance to assess the aspect of time reduction at multiple different centres, between which segmentation guidelines may differ. The aspect of pathology-specific segmentation performance and time reduction would also be of interest to study.

Since this method has been implemented in software available for clinical use [90], it would be of interest to investigate the actual clinical impact of this or similar methods, and how widespread the actual use of DL-based methods for RV segmentation is compared to manual segmentation.

Since this study was conducted, there has been substantial technical developments in the field of DL, and newer methods, such as those based on *transformers*, could potentially

lead to improved RV segmentation performance. However, it would also be of interest to study how a larger and better curated training dataset that includes an even wider range of pathologies and more refined manual segmentations could improve segmentation performance and decrease the need for manual corrections.

Study II

Conclusion: Changing the random seed can lead to statistically significant differences in performance between DL-based medical image segmentation models from a single learning algorithm, using standard methods for performance estimation and statistical testing. Thus, statistically significant differences in model performance do not reliably indicate differences in performance between different learning algorithms.

Future directions: To build upon the results of this study, it would be relevant to investigate how performance differences between models change with the number of epochs. To move away from unreliable performance comparisons, there are suggested methods that use repeated model training [95]. However, methods that require multiple repeated training procedures have a high computational demand and can be time consuming, and are therefore not likely to be widely applied. Therefore, methods for reliable performance estimation of learning algorithms that require minimal training are of great relevance for future use in the field of deep learning-based medical image segmentation. For example, methods from neural architecture search [96] may be of relevance.

Study III

Conclusion: The developed semi-automated method for constructing time-resolved ventricular cine images from exercise CMR data demonstrated feasibility for providing time-resolved volume-based measures of cardiac function in both heart failure patients and healthy volunteers. The DL-based method for rapid automated detection of end-diastolic and end-systolic timeframes provided an adequate prediction performance at mid-ventricular slice positions, but needs improvements apically and basally.

Future directions: Future applications of the presented method include producing non-invasive pressure-volume loops [75] during exercise. The method could also potentially be of use for 4D flow MRI during exercise. To improve the presented method for ED and ES detection, it would be of relevance to conduct training with a dataset consisting of actual exercise CMR data, while either formulating the ED and ES detection as a segmentation problem or a classification problem. It could also be of relevance to move towards k-space-based gating.

The use of SMS imaging, as presented in Study IV, could aid the retrospective synchronization of timeframes, since the simultaneous acquisition entails that multiple timeframes

are already synchronized across slice positions. This would substantially accelerate the retrospective synchronization.

Image-based methods for retrospective gating have the drawback of requiring assumptions regarding the appearance of certain image features (e.g. the heart or the diaphragm) to function. For more robustness to image artifacts and a more generalizable synchronization of respiratory and cardiac phases in real-time exercise CMR, it would therefore be relevant to look towards alternative physiological monitoring methods, such as *pilot tone* [97] or improved methods for pulse oximeter or ECG measurements.

Study IV

Conclusion: The developed real-time bSSFP SMS-2 sequence and DL-based online image artifact suppression reconstruction provided substantial acceleration of acquisition time compared to standard CMR and substantial acceleration of reconstruction time compared to compressed sensing. The agreement in volume-based clinical measures with standard CMR was reasonable, but further improvements in image quality are required before the method can be applied clinically.

Future directions: Potential improvements to the DL-based image artifact suppression method include reducing the amount of blurring in the reconstructed data. This could potentially be achieved by increasing the amount of training data, changing the network architecture, or using another loss-function. To reduce artifacts, it could be relevant to further optimize the trajectory for the DL-based reconstruction, as in [98]. Future improvements also include increasing the SMS acceleration factor to further decrease the acquisition time.

Exercise CMR is an important potential future application of this method. Partially because the SMS acquisition could aid the retrospective synchronization of timeframes, but also because the accelerated acquisition could shorten the exercise time for patients with reduced exercise capacity. Pediatric imaging could also benefit from the accelerated acquisition.

Concluding remarks

DL provides a powerful set of methods for solving various problems in CMR, but it is not a global solution. The time reduction that DL enables through speeding up certain tasks can be useful in a clinical context and can facilitate new developments in CMR. The development of a clinically applicable DL method involves multiple sources of randomness, which can make it a long process of trial and error, and can make it difficult to do reliable comparisons. The usefulness of DL methods needs to be assessed on a case-by-case basis as a trade-off between performance, runtime, and the time it would take to carry out the task manually or using other methods. However, DL methods will likely continue to be a big part of future developments in CMR.

Bibliography

[1] World Health Organization. Cardiovascular diseases (CVDs) (2025). Geneva, Switzerland; 2025. Available from: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[2] Petersen SE, Aung N, Sanghvi MM, Zemrak F, Fung K, Paiva JM, et al. Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in Caucasians from the UK Biobank population cohort. Journal of Cardiovascular Magnetic Resonance. 2017;19(1):1-19.

[3] Nishimura DG. Principles of Magnetic Resonance Imaging. 1st ed. Stanford University; 2010.

[4] McRobbie DW, Moore EA, Graves MJ, Prince MR. MRI from Picture to Proton. 2nd ed. Cambridge University Press; 2006.

[5] Bernstein MA, King KF, Zhou XJ. Handbook of MRI Pulse Sequences. 1st ed. Elsevier; 2004.

[6] Hansen MS, Kellman P. Image reconstruction: An overview for clinicians. Journal of Magnetic Resonance Imaging. 2015;41(3):573-85.

[7] Jackson JI, Meyer CH, Nishimura DG, Macovski A. Selection of a convolution function for Fourier inversion using gridding (computerised tomography application). IEEE Transactions on Medical Imaging. 1991;10(3):473-8.

[8] Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magnetic Resonance in Medicine. 2002;47(6):1202-10.

[9] Akcakaya M, Doneva MI, Prieto C. Magnetic Resonance Image Reconstruction: Theory, Methods, and Applications. vol. 7. Academic Press; 2022.

[10] Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. Magnetic Resonance in Medicine. 2007;58(6):1182-95.

[11] Bustin A, Fuin N, Botnar RM, Prieto C. From Compressed-Sensing to Artificial Intelligence-Based Cardiac MRI Reconstruction. Frontiers in Cardiovascular Medicine. 2020;7:17.

[12] Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. Mathematical Programming. 1989;45(1-3):503-28.

[13] Oscanoa JA, Middione MJ, Alkan C, Yurt M, Loecher M, Vasanawala SS, et al. Deep Learning-Based Reconstruction for Cardiac MRI: A Review. Bioengineering. 2023;10(3):334.

[14] Hansen MS, Sørensen TS. Gadgetron: An open source framework for medical image reconstruction. Magnetic Resonance in Medicine. 2013;69(6):1768-76.

[15] Nayak KS, Lim Y, Campbell-Washburn AE, Steeden J. Real-Time Magnetic Resonance Imaging. Journal of Magnetic Resonance Imaging. 2022;55(1):81-99.

[16] Craven TP, Tsao CW, La Gerche A, Simonetti OP, Greenwood JP. Exercise cardiovascular magnetic resonance: development, current utility and future applications. Journal of Cardiovascular Magnetic Resonance. 2020;22(1):65.

[17] Spicher N, Kukuk M, Maderwald S, Ladd ME. Initial evaluation of prospective cardiac triggering using photoplethysmography signals recorded with a video camera compared to pulse oximetry and electrocardiography at 7T MRI. Biomedical engineering online. 2016;15(1):1-28.

[18] Scheffler K, Lehnhardt S. Principles and applications of balanced SSFP techniques. European Radiology. 2003;13(11):2409-18.

[19] Bieri O, Scheffler K. Fundamentals of balanced steady state free precession MRI: Fundamentals of Balanced SSFP MRI. Journal of Magnetic Resonance Imaging. 2013;38(1):2-11.

[20] Bieri O, Scheffler K. Flow compensation in balanced SSFP sequences. Magnetic Resonance in Medicine. 2005;54(4):901-7.

[21] Barth M, Breuer F, Koopmans PJ, Norris DG, Poser BA. Simultaneous multislice (SMS) imaging techniques: SMS Imaging. Magnetic Resonance in Medicine. 2016;75(1):63-81.

[22] Wong E. Optimized phase schedules for minimizing peak RF power in simultaneous multi-slice RF excitation pulses. In: Proceedings of the 20th Annual Meeting of the International Society of Magnetic Resonance Imaging in Medicine, Melbourne, Australia; 2012. p. 2209.

[23] Breuer FA, Blaimer M, Heidemann RM, Mueller MF, Griswold MA, Jakob PM. Controlled aliasing in parallel imaging results in higher acceleration (CAIPIRINHA) for multi-slice imaging. Magnetic Resonance in Medicine. 2005;53(3):684-91.

[24] Setsompop K, Gagoski BA, Polimeni JR, Witzel T, Wedeen VJ, Wald LL. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced *g*-factor penalty. Magnetic Resonance in Medicine. 2012;67(5):1210-24.

[25] Price AN, Cordero-Grande L, Malik SJ, Hajnal JV. Simultaneous multislice imaging of the heart using multiband balanced SSFP with blipped-CAIPI. Magnetic Resonance in Medicine. 2020;83(6):2185-96.

[26] Yutzy SR, Seiberlich N, Duerk JL, Griswold MA. Improvements in multislice parallel imaging using radial CAIPIRINHA. Magnetic Resonance in Medicine. 2011;65(6):1630-7.

[27] Yang Y, Meyer CH, Epstein FH, Kramer CM, Salerno M. Whole-heart spiral simultaneous multi-slice first-pass myocardial perfusion imaging. Magnetic Resonance in Medicine. 2019;81(2):852-62.

[28] Goodfellow IJ, Bengio Y, Courville A. Deep Learning. Cambridge, MA, USA: MIT Press; 2016.

[29] Chollet F. Deep learning with Python. 2nd ed. Manning Publications; 2021.

[30] Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint. 2014. arXiv:1412.6980.

[31] Åkesson J, Töger J, Heiberg E. Random effects during training: Implications for deep learning-based medical image segmentation. Computers in Biology and Medicine. 2024;180:108944.

[32] Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods. 2021;18(2):203-11.

[33] Dumoulin V, Visin F. A guide to convolution arithmetic for deep learning. arXiv preprint. 2018. arXiv:1603.07285.

[34] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer; 2015. p. 234-41.

[35] Özgün Çiçek, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. arXiv preprint. 2016. arXiv:1606.06650.

[36] Redmon J, Farhadi A. YOLO9000: better, faster, stronger. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 7263-71.

[37] Milletari F, Navab N, Ahmadi SA. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). IEEE; 2016. p. 565-71.

[38] Jadon S. A survey of loss functions for semantic segmentation. In: 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB); 2020. p. 1-7.

[39] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing. 2004;13(4):600-12.

[40] Schulz-Menger J, Bluemke DA, Bremerich J, Flamm SD, Fogel MA, Friedrich MG, et al. Standardized image interpretation and post processing in cardiovascular magnetic resonance: Society for Cardiovascular Magnetic Resonance (SCMR) board of trustees task force on standardized post processing. Journal of Cardiovascular Magnetic Resonance. 2013;15(1):1-19.

[41] Bozkurt B, Coats AJ, Tsutsui H, Abdelhamid M, Adamopoulos S, Albert N, et al. Universal Definition and Classification of Heart Failure. Journal of Cardiac Failure. 2021;27(4):387-413.

[42] Widmaier E, Raff H, Strang KT. Vander's Human Physiology. 15th ed. McGraw-Hill Education; 2018.

[43] Haug E, Sand O, Sjaastad ØV. Människans fysiologi. 1st ed. Stockholm: Liber; 1998.

[44] Patel PN, Zwibel H. Physiology, Exercise. In: StatPearls [Internet]. StatPearls Publishing; 2022. Available from: https://www.ncbi.nlm.nih.gov/books/NBK482280/.

[45] Åstrand PO, Rodahl K, Dahl HA, Strømme SB. Textbook of work physiology: physiological bases of exercise. 4th ed. Human kinetics; 2003.

[46] Chung CS, Karamanoglu M, Kovács SJ. Duration of diastole and its phases as a function of heart rate during supine bicycle exercise. American Journal of Physiology-Heart and Circulatory Physiology. 2004;287(5):H2003-8.

[47] Stoylen A. Left Ventricular Mechanics During Exercise: A Doppler and Tissue Doppler Study. European Journal of Echocardiography. 2003;4(4):286-91.

[48] Beaudry RI, Samuel TJ, Wang J, Tucker WJ, Haykowsky MJ, Nelson MD. Exercise cardiac magnetic resonance imaging: a feasibility study and meta-analysis. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology. 2018;315(4):R638-45.

[49] Claessen G, Claus P, Delcroix M, Bogaert J, Gerche AL, Heidbuchel H. Interaction between respiration and right versus left ventricular volumes at rest and during exercise: a real-time cardiac magnetic resonance study. American Journal of Physiology-Heart and Circulatory Physiology. 2014;306(6):H816-24.

[50] Wang S, Chauhan D, Patel H, amir Khalili A, da Silva IF, Sojoudi A, et al. Assessment of right ventricular size and function from cardiovascular magnetic resonance images using artificial intelligence. Journal of Cardiovascular Magnetic Resonance. 2022;24(1):27.

[51] Peng P, Lekadir K, Gooya A, Shao L, Petersen SE, Frangi AF. A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. Magnetic Resonance Materials in Physics, Biology and Medicine. 2016;29(2):155-95.

[52] Petitjean C, Zuluaga MA, Bai W, Dacher JN, Grosgeorge D, Caudron J, et al. Right ventricle segmentation from cardiac MRI: a collation study. Medical image analysis. 2015;19(1):187-202.

[53] Miao Y, Shen CX, Shi WL, Zhang K, Jiang ZG, Yang HM. A Right Ventricle Segmentation Method based on U-Net with Weighted Convolution and Dense Connection. In: Proceedings of the 2020 2nd International Conference on Intelligent Medicine and Image Processing. Tianjin China: ACM; 2020. p. 40-6.

[54] Liu Z, Feng Y, Yang X. Right Ventricle Segmentation of Cine MRI Using Residual U-net Convolutinal Networks. In: 2019 20th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT). Gold Coast, Australia: IEEE; 2019. p. 364-8.

[55] Liu Z, Yang X. A Squeeze Convolutional Network For MRI Right Ventricle Segmentation. In: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). San Diego, CA, USA: IEEE; 2019. p. 697-700.

[56] Borodin G, Senyukova O. Right ventricle segmentation in cardiac mr images using u-net with partly dilated convolution. In: International Conference on Artificial Neural Networks. Springer; 2018. p. 179-85.

[57] Zhang L, Karanikolas GV, Akcakaya M, Giannakis GB. Fully Automatic Segmentation of the Right Ventricle Via Multi-Task Deep Neural Networks. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2018. p. 5.

[58] Isensee F, Jaeger PF, Full PM, Wolf I, Engelhardt S, Maier-Hein KH. Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. In: International Workshop on Statistical Atlases and Computational Models of the Heart. Springer; 2017. p. 120-9.

[59] Calisto MB, Lai-Yuen SK. AdaEn-Net: An ensemble of adaptive 2D–3D Fully Convolutional Networks for medical image segmentation. Neural Networks. 2020;126:76-94.

[60] Simantiris G, Tziritas G. Cardiac MRI Segmentation With a Dilated CNN Incorporating Domain-Specific Constraints. IEEE Journal of Selected Topics in Signal Processing. 2020;14(6):1235-43.

[61] Dangi S, Linte CA, Yaniv Z. A distance map regularized CNN for cardiac cine MR image segmentation. Medical physics. 2019;46(12):5637-51.

[62] Painchaud N, Skandarani Y, Judge T, Bernard O, Lalande A, Jodoin PM. Cardiac MRI segmentation with strong anatomical guarantees. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2019. p. 632-40.

[63] Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. Journal of Cardiovascular Magnetic Resonance. 2018;20(1):65.

[64] Hertel L. On Hyperparameter Optimization for Deep Learning. University of California, Irvine; 2020.

[65] Bhojanapalli S, Wilber K, Veit A, Rawat AS, Kim S, Menon A, et al. On the Reproducibility of Neural Network Predictions. arXiv preprint. 2021. arXiv:2102.03349.

[66] Renard F, Guedria S, Palma ND, Vuillerme N. Variability and reproducibility in deep learning for medical image segmentation. Scientific Reports. 2020;10(1):13724.

[67] Piantadosi G, Marrone S, Sansone C. On Reproducibility of Deep Convolutional Neural Networks Approaches. In: Kerautret B, Colom M, Lopresti D, Monasse P, Talbot H, editors. Reproducible Research in Pattern Recognition. vol. 11455. Cham: Springer International Publishing; 2019. p. 104-9.

[68] Menart C. Evaluating the variance in convolutional neural network behavior stemming from randomness. In: Overman TL, Hammoud RI, Mahalanobis A, editors. Automatic Target Recognition XXX. Online Only, United States: SPIE; 2020. p. 11.

[69] Gerche AL, Claessen G, Van de Bruaene A, Pattyn N, Van Cleemput J, Gewillig M, et al. Cardiac MRI: A New Gold Standard for Ventricular Volume Quantification During High-Intensity Exercise. Circulation: Cardiovascular Imaging. 2013;6(2):329-38.

[70] Trankle CR, Canada JM, Jordan JH, Truong U, Hundley WG. Exercise Cardiovascular Magnetic Resonance: A Review. Journal of Magnetic Resonance Imaging. 2022;55(3):720-54.

[71] Usman M, Atkinson D, Kolbitsch C, Schaeffter T, Prieto C. Manifold learning based ECG-free free-breathing cardiac CINE MRI. Journal of Magnetic Resonance Imaging. 2015;41(6):1521-7.

[72] Ruijsink B, Puyol-Antón E, Usman M, van Amerom J, Duong P, Forte MNV, et al. Semi-automatic Cardiac and Respiratory Gated MRI for Cardiac Assessment During Exercise. In: Cardoso MJ, Arbel T, Gao F, Kainz B, van Walsum T, Shi K, et al., editors. Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment. vol. 10555. Cham: Springer International Publishing; 2017. p. 86-95.

[73] Kerfoot E, Puyol Anton E, Ruijsink B, Clough J, King AP, Schnabel JA. Automated CNN-Based Reconstruction of Short-Axis Cardiac MR Sequence from Real-Time Image Data. In: Stoyanov D, Taylor Z, Kainz B, Maicas G, Beichel RR, Martel A, et al., editors. Image Analysis for Moving Organ, Breast, and Thoracic Images. vol. 11040. Cham: Springer International Publishing; 2018. p. 32-41.

[74] Edlund J, Haris K, Ostenfeld E, Carlsson M, Heiberg E, Johansson S, et al. Validation and quantification of left ventricular function during exercise and free breathing from real-time cardiac magnetic resonance images. Scientific Reports. 2022;12(1):5611.

[75] Seemann F, Arvidsson P, Nordlund D, Kopic S, Carlsson M, Arheden H, et al. Noninvasive Quantification of Pressure-Volume Loops From Brachial Pressure and Cardiovascular Magnetic Resonance. Circulation: Cardiovascular Imaging. 2019;12(1).

[76] Wang X, Uecker M, Feng L. Fast Real-Time Cardiac MRI: a Review of Current Techniques and Future Directions. Investigative Magnetic Resonance Imaging. 2021;25(4):252.

[77] Tian Y, Cui SX, Lim Y, Lee NG, Zhao Z, Nayak KS. Contrast-optimal simultaneous multi-slice bSSFP cine cardiac imaging at 0.55 T. Magnetic Resonance in Medicine. 2023;89(2):746-55.

[78] McElroy S, Kunze KP, Nazir MS, Speier P, Stäb D, Villa ADM, et al. Simultaneous multi-slice steady-state free precession myocardial perfusion with iterative reconstruction and integrated motion compensation. European Journal of Radiology. 2022;151:110286.

[79] Yagiz E, Garg P, Cen SY, Nayak KS, Tian Y. Simultaneous multi-slice cardiac real-time MRI at 0. 55T. Magnetic Resonance in Medicine. 2025;93(4):1723-32.

[80] Hauptmann A, Arridge S, Lucka F, Muthurangu V, Steeden JA. Real-time cardiovascular MR with spatio-temporal artifact suppression using deep learning–proof of concept in congenital heart disease. Magnetic Resonance in Medicine. 2019;81(2):1143-56.

[81] Jaubert O, Steeden J, Montalt-Tordera J, Arridge S, Kowalik GT, Muthurangu V. Deep artifact suppression for spiral real-time phase contrast cardiac magnetic resonance imaging in congenital heart disease. Magnetic Resonance Imaging. 2021;83:125-32.

[82] Jaubert O, Montalt-Tordera J, Knight D, Coghlan GJ, Arridge S, Steeden JA, et al. Real-time deep artifact suppression using recurrent U-Nets for low-latency cardiac MRI. Magnetic Resonance in Medicine. 2021;86(4):1904-16.

[83] Le J, Tian Y, Mendes J, Wilson B, Ibrahim M, DiBella E, et al. Deep learning for radial SMS myocardial perfusion reconstruction using the 3D residual booster U-net. Magnetic Resonance Imaging. 2021;83:178-88.

[84] Adam NL, Kowalik G, Tyler A, Mooiweer R, Neofytou AP, McElroy S, et al. Fast reconstruction of SMS bSSFP myocardial perfusion images using noise map estimation network (NoiseMapNet): a head-to-head comparison with parallel imaging and iterative reconstruction. Frontiers in Cardiovascular Medicine. 2024;11:1350345.

[85] Bernard O, Lalande A, Zotti C, Cervenansky F, Yang X, Heng PA, et al. Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? IEEE Transactions on Medical Imaging. 2018;37(11):2514-25.

[86] Antonelli M, Reinke A, Bakas S, Farahani K, Kopp-Schneider A, Landman BA, et al. The medical segmentation decathlon. Nature communications. 2022;13(1):4128.

[87] Feng L. Golden-Angle Radial MRI: Basics, Advances, and Applications. Journal of Magnetic Resonance Imaging. 2022;56(1):45-62.

[88] Pipe JG, Menon P. Sampling density compensation in MRI: Rationale and an iterative numerical solution. Magnetic Resonance in Medicine. 1999;41(1):179-86.

[89] Berggren K, Hedstrom E, Steding Ehrenborg K, Carlsson M, Engblom H, Ostenfeld E, et al. Multiple Convolutional Neural Networks for Robust Myocardial Segmentation. In: In proceedings of SSBA 2020; 2020. .

[90] Heiberg E, Sjögren J, Ugander M, Carlsson M, Engblom H, Arheden H. Design and validation of Segment - freely available software for cardiovascular image analysis. BMC Medical Imaging. 2010;10(1):1.

[91] Åkesson J, Ostenfeld E, Carlsson M, Arheden H, Heiberg E. Deep learning can yield clinically useful right ventricular segmentations faster than fully manual analysis. Scientific Reports. 2023;13(1):1216.

[92] Åkesson J, Edlund J, Steding-Ehrenborg K, Heiberg E. Retrospectively synchronized time-resolved ventricular cine images from 2D real-time exercise cardiac magnetic resonance imaging. Clinical Physiology and Functional Imaging. 2025;45(5):e70027.

[93] Rosset A, Spadola L, Ratib O. OsiriX: An Open-Source Software for Navigating in Multidimensional DICOM Images. Journal of Digital Imaging. 2004;17(3):205-16.

[94] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet. 1986;327(8476):307-10.

[95] Bajgar O, Kadlec R, Kleindienst J. A boo (n) for evaluating architecture performance. In: International Conference on Machine Learning. PMLR; 2018. p. 334-43.

[96] Chen Y, Guo Y, Chen Q, Li M, Zeng W, Wang Y, et al. Contrastive Neural Architecture Search with Neural Architecture Comparators. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE; 2021. p. 9497-506.

[97] Chen C, Liu Y, Simonetti OP, Tong M, Jin N, Bacher M, et al. Cardiac and respiratory motion extraction for MRI using pilot tone—a patient study. The International Journal of Cardiovascular Imaging. 2024;40(1):93-105.

[98] Jaubert O, Montalt-Tordera J, Knight D, Arridge S, Steeden J, Muthurangu V. HyperSLICE : HyperBand optimized spiral for low-latency interactive cardiac examination. Magnetic Resonance in Medicine. 2024;91(1):266-79.

Part II

Research papers

Author contributions

Table 5.1 and the next pages summarize my contributions to each study included in this thesis.

	Study design	Method development	Numerical data collection	Data analysis	Statistics	Figures and tables	Interpretation of results	Conceptualization	Manuscript preparation	Manuscript revision	MRI data collection	Manual delineations
Study I	2	3	3	3	3	3	3	2	3	3	0	0
Study II	3	3	3	3	3	3	3	3	3	3	-	-
Study III	3	3	3	3	3	3	3	1	3	3	0	0
Study IV	2	3	3	3	3	3	3	0	3	-	2	0

Not applicable	-
No contribution	0
Minor contribution	1
Moderate contribution	2
Significant contribution	3

Figure 5.1: Table summarizing my contributions to the four studies in this thesis.

Study I:

Performed data curation and processing of all the training and testing data. Developed all the code in the deep learning-based segmentation pipeline, inspired by previous work. Conducted training and optimization of all neural networks. Collected numerical data from delineations. Analysed all data on neural network segmentation performance and time measurements. Carried out all statistics. Created all figures and tables. Did the initial interpretation of results for the manuscript. Drafted the manuscript. Revised the manuscript.

Study II:

Carried out the literature analysis. Coded all needed modifications to the used publicly available code for neural network training. Performed the training of all neural networks. Analysed the data on neural network performance. Carried out all statistics. Created all figures and tables. Did the initial interpretation of results for the manuscript. Drafted the manuscript. Revised the manuscript.

Study III:

Developed and coded the semi-automated pipeline, inspired by previous work. Collected numerical data from delineations. Analysed the data. Carried out all statistics. Created all figures and tables. Did the initial interpretation of results for the manuscript. Drafted the manuscript. Revised the manuscript.

Study IV:

Implemented the needed modifications to the existing MRI pulse sequence. Developed the DL-based reconstruction pipeline, inspired by previous work. Curated and processed training data for neural network training. Performed training and optimization of the neural network for image reconstruction. Collected in-silico and in-vivo data to debug and optimize the MRI pulse sequence and reconstruction method during development. Implemented the reconstruction method for online reconstruction using the Gadgetron framework [14]. Set up the imaging protocol and reconstruction method for MRI data collection of in-vivo data. Collected the MRI data in healthy volunteers but did not do the anatomical planning. Did all data analysis. Carried out all statistics. Created all figures and tables. Did the initial interpretation of results for the manuscript. Drafted the manuscript. The manuscript is not yet submitted.

Printed by Tryckeriet i E-huset, Lund 2025

LUND
UNIVERSITY

ISBN: 978-91-8104-683-0 (print)
ISBN: 978-91-8104-684-7 (electronic)
ISRN: LUTEDX/TEEM–1145–SE
Report-nr: 6/25