UNIVERSITY OF LEEDS

Deep Learning-based and Motion-Compensated Reconstruction for Undersampled Cardiac MRI

Kun Wu
University of Leeds
School of Computing

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy
January, 2025

Intellectual Property Statement

The candidate confirms that the work submitted is his own and that appropriate credit has been given within the thesis where reference has been made to the work of others. The contribution of the candidate and other authors to this work is explicitly indicated in the following.

I am the main author of all the following publications. I led the design of the study and wrote the manuscript, including analysis and discussion of the results. I also wrote the analysis code and carried out the data processing, statistical analysis, and experiments. The contribution of other authors was to help me discuss the results and review the manuscript.

1. **Chapter 3:**
   Wu, K., Xia, Y., Ravikumar, N. and Frangi, A.F., 2024. Compressed sensing using a deep adaptive perceptual generative adversarial network for MRI reconstruction from undersampled K-space data. Biomedical Signal Processing and Control, 96, p.106560.

Acknowledgements

First and foremost, I would like to extend my most sincere gratitude to my team of supervisors: Dr Nishant Ravikumar, Professor Alejandro F.Frangi and Dr Yan Xia, for their professional and invaluable guidance, responsible attitude and persistent patience during my doctoral study. Their rigorous approaches to academic research and excellent research abilities has had a profound impact on me. I have not only gained doctoral supervisors, but also long-term and solid friendships. In addition, I would like to thank Professor David Buckley, Professor Andrew Scarsbrook and Dr Sharib Ali, who provided valuable suggestions in my PhD transfer and annual progress review, and affirmed my stage works.

My heartfelt appreciation goes to my lab mates and friends Haoran Dou, Fengming Lin, Nina Cheng, Ning Bi, Cynthia Maldonado Garcia, Rodrigo Bonazzola, over the four years, they provide me with a good atmosphere and support in my study. Further, I want to thank our software engineers Alejandro Lebrero and Kattia Hoz de Vila. Special thanks to Xiang Chen, Qiongyao Liu, the common pursuit of life and hobbies has made us become close friends in scientific research and life. I also like to thank Domantas Dilys and Jiangbei Yue, even though we come from different research teams, I have gained valuable friendships and company during the last three months in our university. Although my student life is over, the support and love these friends have given me is an endless asset to me in the future.

I would also like to thank the School of Computing and the University of Leeds, for providing me with an excellent research environment and an efficient support team.

My most special thanks go to my family, especially my parents, Genlong Wu and Shiying Gao, who have given me the most selfless love and unconditional support. I am very grateful that I have a happy and harmonious family. I am not a gifted student, but they let me know that I would always be the best in their eyes, and they would even be happy for my small improvement. My family is my motivation to overcome obstacles, explore the world, and try to break through myself.

At this point in time, my four-year PhD programme is coming to an end, bringing my twenty-two years as a student to a close. The journey has been like an arduous yet rewarding marathon and my PhD degree is my final medal. As the old Chinese saying goes, there is no end to learning. I am about to start the next chapter of my life, and I believe that I can achieve a life without regrets and a better version of myself.

Abstract

Cardiovascular diseases (CVDs) remain one of the leading causes of mortality worldwide in the clinical factors, accounting for an estimated 17.9 million deaths in 2019, which represents 32% of all global deaths. Medical imaging technologies, particularly Magnetic Resonance Imaging (MRI), is a classic and reliable tool for non-invasive imaging, with the aim of diagnosing CVDs. Cardiac MR (CMR) imaging has emerged as an essential clinical tool for assessing cardiac morphology and function, due to its high contrast of anatomical structures. However, CMR is hindered by the long acquisition times required to achieve high-quality measurements in both spatial and temporal resolution. Additionally, cardiac motion during acquisition further increases scan time. These factors make MR imaging time-consuming and prone to variability. Recently, deep learning-based CMR reconstruction approaches have been proposed to address these challenges by learning the ill-posed inverse optimization process, thereby offering state-of-the-art solutions and advancing other relative technologies in medical image analysis.

This thesis encapsulates the development and validation of deep learning-based frameworks in the field of MRI reconstruction and motion modelling from undersampled cardiac MRI scans of k-space. It introduces a systematic approach for accelerating cardiac MR imaging, underpinned by spatio-temporal coherence, capable of dynamic reconstruction of high-quality CMR images from significantly fewer measurements. Three models are derived from this foundational research, each contributing to different aspects of the process.

Firstly, adaptive supervision is proposed through a feature guidance mechanism capable of operating at multiple levels during extraction. This approach addresses aliasing robustness and efficiently leverages limited reference data, ensuring the capture of fine details while emphasising features of underlying anatomical structures. Next, we introduce a pyramid network that leverages combinational-attention-based correlation mapping, enabling accurate estimation of cardiac motion with the capability to handle deformation fields from coarse to fine. Furthermore, we introduce the CS+M (compressed sensing + motion estimation) method for dynamic CMR reconstruction. This joint optimisation facilitates the connection between spatio-temporal MRI reconstruction and the estimation of inherent physical motion. This innovative approach opens new avenues by utilising motion as a compensatory factor captured during accelerated data acquisition and intermediate reconstruction processes, acting as a temporal regularisation embedding in a single optimisation model. This is critical for accomplishing reliable reconstruction even under aggressive scenarios with various undersampling strategies.

Contents

LIST OF FIGURES

List of Tables

Abbreviations

ACDC	Automated Cardiac Diagnosis Challenge	ACS	Autocalibration signal
AMI	Acute Myocardial Infarction	ARV	Abnormal Right Ventricle
BS	Brain Stem	CVD	Cardiovascular Disease
CT	Computed Tomography	CTA	Computed Tomography Angiography
CAD	Coronary Artery Disease	CS	Compressed Sensing
CNN	Convolutional Neural Network	CC 359	Calgary Campinas 359
CCM	Corpus Callosum of Middle-anterior	CMR	Cardiac Magnetic Resonance
DSC	Deep-wise Separable Convolutions	DL	Deep Learning
DCM	Dilated Cardiomyopathy	DC	Data Consistency
DS	Dice Score	ED	End Diastole
ES	End Systole	FT	Fourier Transform
HCM	Hypertrophic Cardiomyopathy	HD95	Hausdorff Distance
IFT	Inverse Fourier Transform	LA	Left Atrium
LAX	Long-Axis	LPIPS	Learned Perceptual Image Patch Similarity
LP	Left Putmamen	LT	Left Thalamus
LVEDV	Left Ventricle End Diastole Volume	LVESV	Left Ventricle End Systole Volume

Abbreviations

LVSV	Left Ventricle Stroke Volume
LVEF	Left Ventricle Ejection Fraction
LVM	Left Ventricle Myocardial mass
LVBP	Left Ventricle Blood Pool
LVM	Left Ventricle Myocardium
MRI	Magnetic Resonance Imaging
MINF	Myocardical Infarction
MYO	Myocardium
MSE	Mean Squared Error
MAE	Mean Absolute Error
NJD	Negative Jacobian Determinant
PET	Positron Emission Tomography
PSNR	Peak Signal-to-Noise Ratio
RA	Right Atrium
RP	Right Putamen
RT	Right Thalamus
RV	Right Ventricle
RVEDV	Right Ventricle End Diastole Volume
RVESV	Right Ventricle End Systole Volume
RVSV	Right Ventricle Stroke Volume
RVEF	Right Ventricle Ejection Fraction
RVM	Right Ventricle Myocardial mass
SPECT	Single Photon Emission Computed Tomography
SAX	Short-Axis
SOTA	State-Of-The-Arts
SSIM	Structure Similarity Index Measurement
TV	Total Variation
UKBB	UK Biobank
VM	Voxelmorph

CHAPTER 1

Introduction: Background, Motivation and Contribution

Cardiovascular disease (CVD) ranks as the foremost cause of mortality globally, accounting for approximately 19 million deaths in 2020. To aid doctors in clinical practices in terms of comprehending, detecting, and analyzing CVDs, the role of medical imaging technologies, along with their associated processing and analytical algorithms, has grown increasingly crucial. Particularly notable is the surge in advancements in automatic CVD analysis following the introduction of deep learning techniques. In this chapter, we first outline the essential knowledge concerning the human heart and Cardiac MR imaging, in particular, the motivations and contributions are demonstrated. Finally, an architectural overview of this thesis is presented.

1.1 Cardiac Clinical Background and MR Imaging

1.1.1 Cardiac Anatomy and cardiac cycle

The human heart is a vital muscular organ responsible for pumping blood throughout the body, supplying oxygen and nutrients while removing metabolic waste products. The heart's anatomical structure is intricately designed to support its physiological functions, encompassing several key components, including the cardiac chambers, valves, and the conduction system.

As shown in Fig. 1.1 and Fig. 1.2, anatomically, the heart is segmented into four chambers: the right atrium, right ventricle, left atrium, and left ventricle. The atria serve as the receiving chambers for blood returning to the heart, whereas the ventricles function as the primary pumping chambers. The right atrium and right ventricle constitute the right heart, which manages the pulmonary circulation by directing deoxygenated blood to the lungs. Conversely, the left atrium and left ventricle form the left heart, responsible for systemic circulation by propelling oxygenated blood throughout the body.

The cardiac valves play a crucial role in ensuring unidirectional blood flow and preventing regurgitation. These include the tricuspid valve which is situated between the right atrium (RA) and right ventricle (RV), the pulmonary valve situated between the RV and pulmonary artery, the mitral valve situated between the left atrium (LA) and left ventricle (LV), and the aortic valve situated between the LV and aorta. Each valve's anatomical configuration is integral to its function, ensuring efficient and coordinated blood flow through the heart's chambers and into the circulatory system.

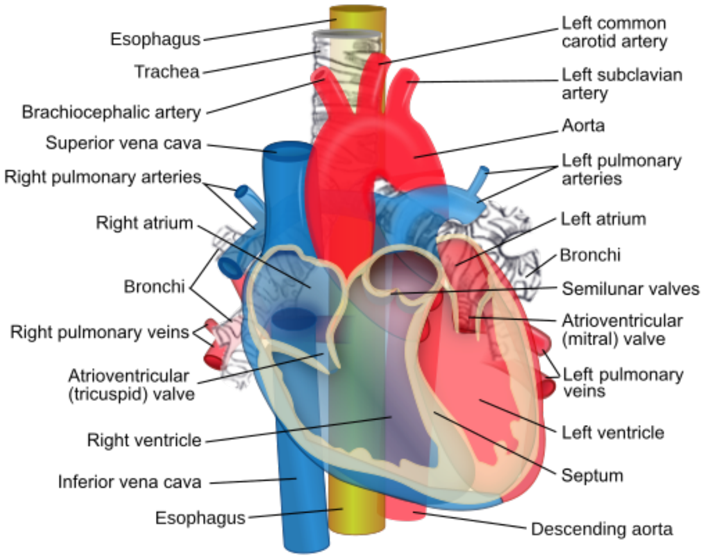


Figure 1.1: The anatomy of the heart. The internal structures of the heart from the anterior view. It shows the four chambers, the major vessels and the values. The image is from ZooFari et al.[1].

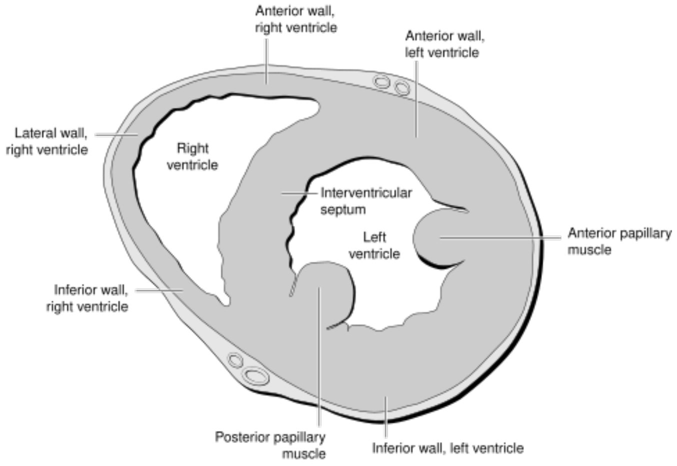


Figure 1.2: The normal short-axis echocardiography view of the heart. The structures indicate the LV, RV and myocardium. Compared with the relatively thinner RV myocardium, the LV myocardium comprises three distinct layers: the epicardium, myocardium, and endocardium, located from the internal surface to the outermost layer of the LV. The image is adapted from Patrick J. Lynch[2].

The cardiac cycle is a natural property of the human heart from the beginning of one heartbeat to the beginning of the next heartbeat in line with the fixed time and space rhythm law. A single complete beating cycle of the heart consists of two dynamic phases: diastole and systole. The start of the general cardiac cycle is marked by atrial contraction, when the heart at the critical point at the end of the previous diastolic phase pushes blood into the ventricles. The ventricular myocardium begins to accommodate the injected blood. The ventricles begin to contract with the relaxation of the atria, at which point the ventricular pressure increases. At the same time, the isovolumic ventricles are activated, characterised by atrioventricular values closure. The critical point of semilunar valve opening is the time point when the ventricular pressure exceeds the pressure values of the aorta and pulmonary arteries. This critical point is followed by the ejection phase, where blood is injected by the ventricles into the aorta and pulmonary arteries. This is also the driving mechanism of blood circulation throughout the body. This is followed by isovolumic ventricular diastole, in which the ventricular myocardium begins to relax and the semilunar valve closes. The ventricular pressure enters a decreasing phase, but its volume does not change. The atrioventricular values valve opens passively, and when the ventricular pressure is lower than the atrium, the ventricle begins to be filled again with blood injected from the atrium.

For cardiac anatomy structural processing and analysis, the end-diastole (ED) and end-systole (ES) time frames are the two most be adopted phases with their distinct anatomical motion status. CMR temporal sequences with ED frame as the first position make it easier for relevant analysis, such as group-wise motion estimation, and clinical key indices computing. From the visual inspection view, the ED frame contains the largest area for LV-related volume of a cardiac cycle, while ES is the minimal size.

1.1.2 Cardiac MR Imaging: From accelerated data acquisition to deep learning-based reconstruction

Medical imaging technologies offer different advantages in evaluating the structure, function, and pathology of the human heart. Based on its unique nature, cardiac imaging is essential in diagnosing, monitoring and managing cardiovascular diseases.

Among the mainstream imaging approaches, MR imaging, as a non-invasive technique, enables a better contrast effect on soft tissue than other medical imaging solutions, further, detailed tissue characterisation without ionising radiation is available based on this operation. In addition, MRI enables multiplane imaging and a much more large field of view.

To form image-space information of a target body area, for instance, human heart, MRI scanners apply intense magnetic fields to all the organs of the patient. Specific resonance frequency-based oscillating magnetic field is formed for the cardiac region as the target area. The fundamental theory during this process is, under the effection of external energy fields, hydrogen nuclei in patient's tissues will reflect abnormal signals. By acquisition and processing the signals through certain media, different densities of nuclei can form images corresponding to multiple organs of the specific areas.

Cardiac Magnetic Resonance Imaging (CMR) is considered the gold standard for assessing key cardiac anatomical structures especially myocardial structure and function.

There are several imaging tools adopted in clinical practice, such as echocardiography, which is conducted based on ultrasound waves. As a real-time, portable and relatively low-cost imaging, it is widely available. But the disadvantage is obvious, less detailed characterization can be presented especially compared with MRI. Another one is computed tomography (CT), especially coronary CT angiography (CTA), which is highly effective in visualising coronary arteries and detection of disease, especially because it enable shorter scan times than MRI, however, it is exposure to ionizing radiation and heavily limited in contrast tissue details. Another widely used method is single photon emission computed tomography (SPECT/PET), which is highly sensitive for detecting ischemia and assessing myocardial viability. However, the performance of these technologies is limited by the exposure of ionizing radiation, higher cost and longer procedure times, and much lower spatial resolution especially compared to CT and MRI, in particular, PET requires more specialised equipment and expertise.

This section focuses on introducing two essential knowledgements within cardiac MRI: the pipeline between k-space undersampling and deep-learning-based reconstruction networks for CMR acceleration, another aspect is Cine-MRI as the main research target of this thesis. These elements play a significant role in forming accurate and practical insights for cardiac mechanics, morphology, and functional analysis from quantitative and qualitative views, as a professional assistant approach for clinicians.

K-space undersampling and image-space learning

The concept of k-space is a generalisation of the simple relation of a time-variant signal to a spectrum as shown in Fig. 1.3 to two or more dimensions. An image is related to a two-dimensional k-space data set by two-dimensional Fourier transformation (2DFT) in the same way a spectrum is related to the signal by one-dimensional FT. 2DFT is formally achieved by successive row-by-row and column-by-column FT. FT is an information-preserving operation and therefore the k-space data contain exactly the same information as the image data. Thus, in the situation of MRI data acquisition with regular scan time, the full-sampled feature in the image domain can be obtained once the full-sampled k-space data are measured in advance. The task of acquiring a scanned image is then converted to the task of finding a way to measure the necessary corresponding frequency feature in the k-space data.

For an understanding of the physical meaning of the acquisition in the frequency domain, the k-space signal can be originally derived from the basic 1D procedure. The time-domain signal defines the 1D k-space. Its coordinates are obviously defined by the area under the gradient that the spins have experienced prior to the collection of each k-space data point. The main limitation for CMR imaging in practice is the prolonged time cost for data acquisition. It is more challenging for cardiac compared with other areas due to the requirement of an achievable spatio-temporal high resolution will lead to a slow speed for scanning. However, high resolution in both spatial and temporal dimensions is essential for appropriate and efficient diagnoses. Reducing the measurements of k-space data and conducting reconstruction in image-space becoming the most successful approach.

To simulate retrospective undersampling from single-coil frequency domain data acquisition to image domain via inverse Fourier transform, in the second row of Fig. 1.3, we present 2D undersampling masks which are generally utilised to perform data acceleration on the fully-sampled k-space data. Nevertheless, the undersampling operation introduces incoherent aliasing artifacts into the image domain, which can be removed by image reconstruction algorithms. Here, both conventional and deep learning-based methods can be the options to address the reconstruction problem. The difference is the way to perform regularisation. Based on the above theories, the undersampled images are fed into a non-linear reconstruction network for retrieving the final predictions with high-quality image. The sampling masks are binary, signifying that a sample from

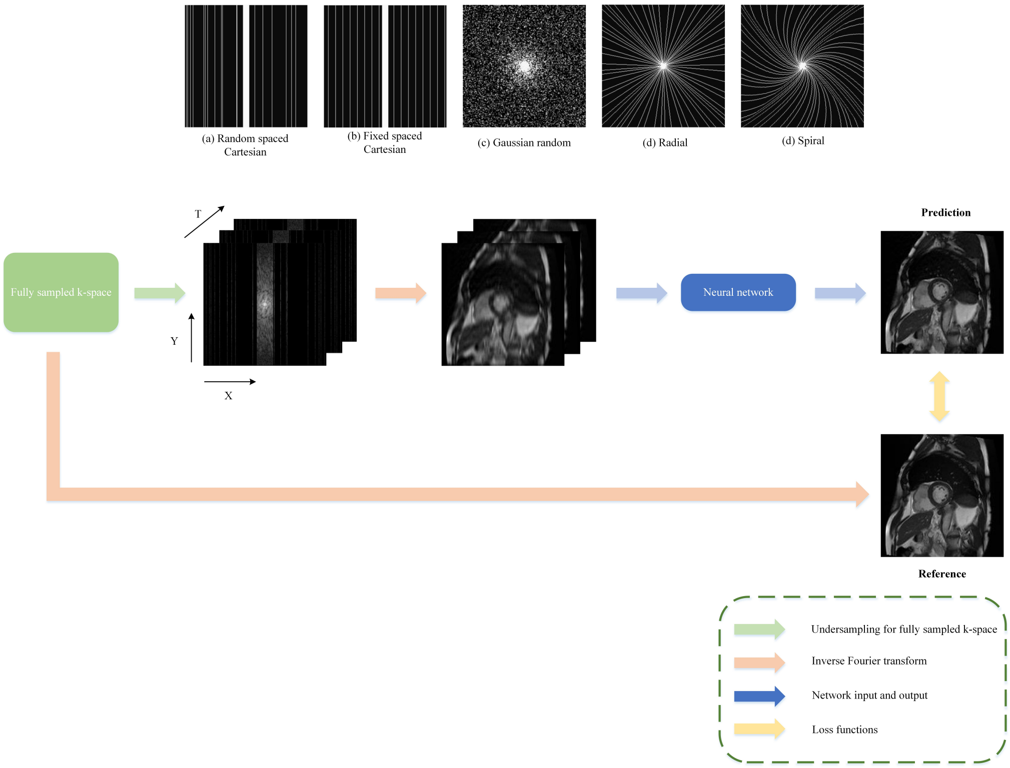


Figure 1.3: The first row: introduction of five most adopted sampling mask for acceleration of data acquisition to facilitate sparse sampling. The second row: the pipeline of supervised deep learning-based reconstruction networks for single-coil data acceleration across frequency domain and image domain

the fully sampled data was masked if and only if the corresponding mask entry was zero. Here we introduce widely-adopted undersampling strategies for the acceleration of MRI for subsequent DL-based reconstruction, which includes Cartesian sampling mask, Gaussian random mask, Pseudo-radial mask, radial mask and spiral mask. In particular, we illustrate the effect of accelerated acquisition based on partial sampling utilising Cartesian mask as an example, which is leveraged as the basic template for multi-variants of undersampling pattern in the following research.

As shown in Fig. 1.3(a), Cartesian sampling aims to collect samples on an equal or randomly spaced and rectangular grid. The most common Cartesian trajectory is the rectilinear pattern in which k-space samples are acquired in a line-by-line scheme in which the resulting samples are multi-distant in both axes, while most conventional strategies adopt an equidistant scheme. A widely-used standard for this pattern is fully sampling a small region of the centre region of the k-space, which is known as the autocalibration signal (ACS) which includes low frequencies [3]. We employed three distinct rectilinear sampling patterns in Fig. 1.3: (a) ACS only: only sampling on the ACS region without any additional sampling lines. (b) Equispaced: Sampling lines are included with a fixed distancing that satisfies the desired accelerated quality. (c) Random-spaced: Sampling lines are adopted from the uniform distribution with possible overlap with the ACS lines. As illustrated in Fig. 1.3 (b), in Gaussian random sampling, the probability of sampling a point in k-space follows a Gaussian distribution, a pattern that samples more points near the centre of k-space (ACS region). Based on this, the global structure of the anatomy is crucial for capturing, while in contrast, it is not afforded by fine details and edges of the structures, as a sampling result of the outer k-space regions corresponding to higher frequencies. Compared with the structured aliasing after accelerated sampling of Cartesian pattern, the aliasing of random results in more diffuse and incoherent artefacts, which are spread throughout the image. In contrast, the Uniform Cartesian aliasing is structured and coherent, the undersampling introduces anisotropy in the image resolution, often leading to streaks or lines aligned with the undersampling direction.

Radial and spiral sampling masks are another two categories of Non-Cartesian trajectories. As shown in Fig. 1.3(d), in the radial mask, k-space signal samples are acquired along several radial spokes radiating its centre, resulting in the centre being sampled multiple times. Meanwhile, the spiral mask includes acquiring data on helical curves starting from the centre of the k-space, as illustrated in Fig. 1.3(e). In non-Cartesian trajectories, k-space measurements are acquired on a non-Cartesian grid and are therefore not equidistant with each other. For instance, in the radial filling, samples closer to the centre are dense compared to samples further on the radial spokes.

Cardiac CINE MRI

Based on the superior ability to provide high contrast on tissues, in the area of CVD image analysis, MR images are most adopted as the gold standard. To leverage the advantage, this thesis paper presents the research works and improvements in the analysis and processing of cardiac CINE MR images. In which, acquired data can be categorised into two types: short-axis images (SAX) and long-axis images (LAX). There are

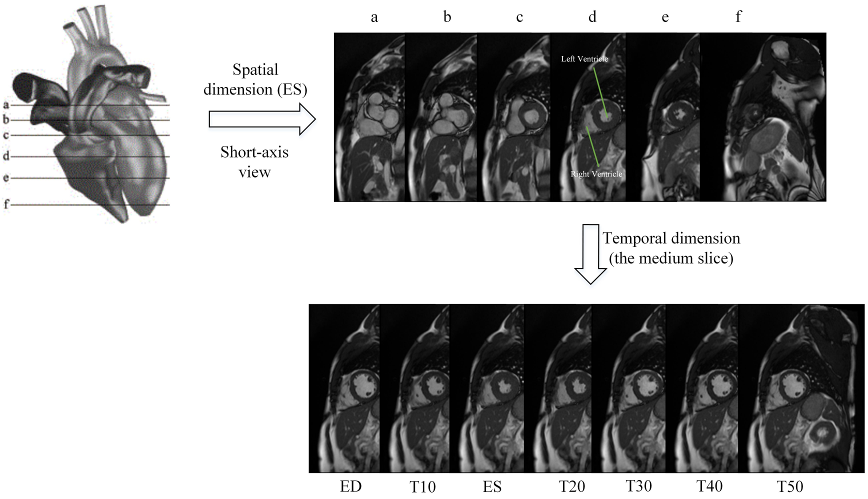


Figure 1.4: Introduction of cardiac cine MR images along spatial dimension and temporal dimension. The general format of the UKBB CMR dataset is a group of separate 2D slice images with a fixed slice thickness of 8 mm for each case. The stacked volume data is available for each time frame of a complete cardiac cycle.

several most adopted cardiac image datasets for the research in the area of CMR image analysis, regarding to the different acquisition operations during scan. For instance, UK Biobank (UKBB).

CMR images in SAX facilitate an ideal visual inspection for anatomical structures of LV and RV. Meanwhile, limited by the inherent characteristics of CMR, high in-plane resolution which causes huge sampling distances between obtained slices generally exists. As illustrated in Fig. 1.4, SAX volume data conducts a stack of a sequence of 2D images (usually containing less than 20 slices), in contrast with other data in 3D form. In addition, each case contains multiple temporal volume frames normally. The average amount of a cardiac cycle can be roughly categorised into three groups: 20, 30, and 50 time frames. While LAX is available to provide supplementary information for SAX images. In addition, LAX enables indicating the structure of RA and LA, including three slices of 2D images from multiple visual directions, which are 2-chamber (2CH), 3-chamber(3CH) and 4-chamber (4CH), respectively. In certain cases, usually, only no more than 2 slices of LAX are available. An example from UKBB is shown in Fig.5, where SAX images and LAX images are presented.

1.1.3 Overview of Cardiovascular Diseases and the Current Analysis

Cardiovascular disease is a general category of diseases that affect the heart or circulation. Arterial damage to organs such as the brain, heart, kidneys and eyes is also a potential contributor to cardiovascular disease. The most common cardiovascular diseases include coronary artery disease (CAD), heart failure, hypertension, stroke, and arrhythmia. In recent years, CVD has become one of the primary causes of mortality worldwide. It accounts for about 30 % of human deaths worldwide and about 10 % of the disease burden. Given the complex nature and widespread effects on the human body, CVDs have been categorised in various ways. Despite this diversity, the major types of CVDs are consistently recognised across these classification systems, with ischemic (coronary) heart disease and hypertensive disorders being among the most prominent. Main CVD normally can be divided into four categories: coronary heart disease (including heart attack, heart failure, and angina), aortic disease, transient ischaemic attack and stroke. It is worth noting that CVDs and heart disease are two definitions with overlapping areas. However, heart disease, specifically, includes more detailed and multiple classifications, which include coronary heart disease, heart failure, acute rheumatic fever, chronic rheumatic heart disease, pulmonary heart disease, hypertensive heart disease, pulmonary circulation and hypertensive renal disease[4]. Among the several major datasets that are available for research to the public in the field of cardiac medical image analysis(e.g.UK BioBank, Automated Cardiac Diagnosis Challenge-ACDC), the following main categories of CVDs are covered: acute myocardial infarction (AMI), myocardial infarction (MINF), dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM) and abnormal right ventricle (ARV).

In order to assist clinicians in more comprehensively understanding the complex disease situation and provide a scientific and accurate preoperative or postoperative auxiliary diagnosis for clinical practice, artificial intelligence-based medical image analysis and processing technology has been widely used. Considering multiple influence factors including various modalities of different patients, different types of scanners, and acquisition times accordingly, relying on manual analysis and processing by clinicians only can no longer be regarded as the unique gold standard due to the limitations of non-efficient and potential risk of error. Based on the assistance from computer vision-aided technologies as automatic tools, complex tasks including reconstruction, registration, detection, super-resolution, recognition and classification of regions of interest (ROIs) can be individually or jointly addressed accurately. A prediction with higher-quality images, efficient and accurate analysis with sufficient information for clinical decision-making of CVDs analysis can be achieved.

1.2 Thesis Motivation, Contributions and Overview

In this section, as shown in Fig. 1.5, the research background and motivations, the main contributions of the three works in the thesis, and the relationships and outlines of each chapter in the thesis, are presented respectively.

1.2.1 Background and motivation

As mentioned above, CMR CINE imaging offers the gold standard for measurements of cardiac morphology and function, enabling the fine visual inspection quality of anatomical structures with high contrast. In particular, CMR imaging is beneficial for downstream evaluations with multiple quantitative indicators, including key volumetric indices and functional analysis for evidence-based CVDs. A reduction with much fewer measurements and scanning time accordingly would be desirable to accelerate acquisition time, minimize potential image artefacts due to motion, meanwhile, increase patient comfort. However, this violates the Nyquist-Shannon sampling theorem based on the limitation of the reduced k-space data that can be collected for each temporal frame, resulting in highly undersampled k-space data. In particular, it is a linear balance problem between the amount of sparse measurements to be transferred into the image domain finally and the scan time. The constraints include the sensitivity to physical motion-based image degradation, the reduction of clinical throughput, etc.

In the past few decades, with the help of the theory of Compressed Sensing MRI (CS-MRI) and the fast development of deep learning-based methods, it showed a great improvement and potential in solving CS-MRI with much faster reconstruction. Despite this, the CMR reconstruction remains challenging due to the complexities associated with detailed anatomical variations and cardiac inherent motion, and potential improvement can be observed in terms of the following aspects.

The high-quality CMR reconstruction is primarily focused on three major branches:

1. The first is improving the generalisation ability of image reconstruction along spatial resolution. Handling the denoising problem for separate frames from a

complete sequence is the main motivation.

2. The second is accurate motion estimation capable of learning the potential physical information of cardiac in temporal dimension.
3. The last step is a hybrid model that integrates data-driven and physics-based prior knowledge to achieve more accurate and motion-robust reconstructions of dynamic CMR, which allows for improved handling of complex motion patterns and variations in cardiac dynamics.

1.2.2 Thesis Contributions

This thesis offers substantial improvements in the efficiency, accuracy, and reliability of high-quality MRI reconstruction, cardiac motion estimation and subsequent analysis for clinical relevance, contributing to diagnosis and related technologies for medical image analysis, which is crucial for patient care. The key contributions are the following:

1. **Exploration of supervision-efficient reconstruction (Chapter 3):** This thesis explores an adaptive supervision mechanism, capable of providing multi-level feature learning guidance in terms of perceptual details, enhancing the generalisation of the network. The results presented in my published paper [5] demonstrate the de-aliasing robustness of the proposed method, particularly when high acceleration rates and aggressive sampling strategies are applied.
2. **Innovations in cardiac motion estimation (Chapter 4):** The development of an attention-aware cardiac motion estimation network, integrated with the pyramid correlation mapping with multi-range, systematically improves the multi-range motion extraction capability of global cardiac region and inter-anatomical sub-regions (i.e. LVendo, LVmyo, RV.) in a parallel fashion. Considering how to well aligned with clinical interest, the proposed approach enables a well-balance between smoothness of deformation field and structural consistency of the target anatomical details. This is a benefit for downstream clinical applications based on strain analysis along radial and circumferential directions.
3. **Achieving group-wise dynamic CMR reconstruction with spatio-temporal coherence (Chapter 5):** A unified $3D + time$ optimisation between MR reconstruction and motion estimation, with the latter driving in the iterative updating

in an unrolled framework. In particular, based on the previous study on fully-sampled and pair-wise cardiac motion estimation, a new variant that can conduct group-wise attention-based residual learning motion estimation with accelerated data specifically was proposed. For clinical relevance, the proposed method allows reconstructing high-quality images from a complete cardiac cycle to facilitate improving spatial and temporal resolutions by incorporating physical information into the compressed-sensing theory. Higher acceleration factors can be performed to reduce the required measurements in practice, while retaining no influence on image quality closer to fully sampled data.

1.2.3 Thesis Overview

The overview of the chapters in this thesis is summarised as follows:

Chapter 2 presents a thorough literature review of deep learning-based MRI reconstruction from theoretical and methodological perspectives. Section 2.1 provides a brief introduction to the reason we chose MR imaging as the imaging technology for the research target - cardiac medical images, based on the unmatched advantages. Also, this section gives a special emphasis on introducing the structure of the following detailed literature review from the view of accepting compressed sensing as the theoretical backbone and related DL-based approaches. Section 2.2 outlines the problem formulation and development from the general (discretised) principles of MR signal generation and image formation to CS-MRI imaging theory. Section 2.3 reviews the mainly adopted network architectures and the classic combined with more recent approaches which achieved higher performance during the development of MRI reconstruction. The review focuses on four main categories including deep CNN trained in end-to-end, generative networks, self-supervised methods and dual-domain frameworks, elucidating their theoretical-based and foundational components, with a special emphasis on approaches for CMR. Section 2.4 discusses and reviews the advanced DL algorithms which focus more on being translated into clinical practice applications, primarily exploring DL-based methods in terms of quantitative and dynamic MRI, laying the foundation of the main motivation for subsequent chapters. Then, Section 2.5 introduces two cardiac cine-MRI datasets utilised in this thesis. Section 2.6 discusses the challenges for the general MR reconstruction and CMR-specific perspectives.

Chapter 3 delves into the improvement of the robustness of aliasing and blur when

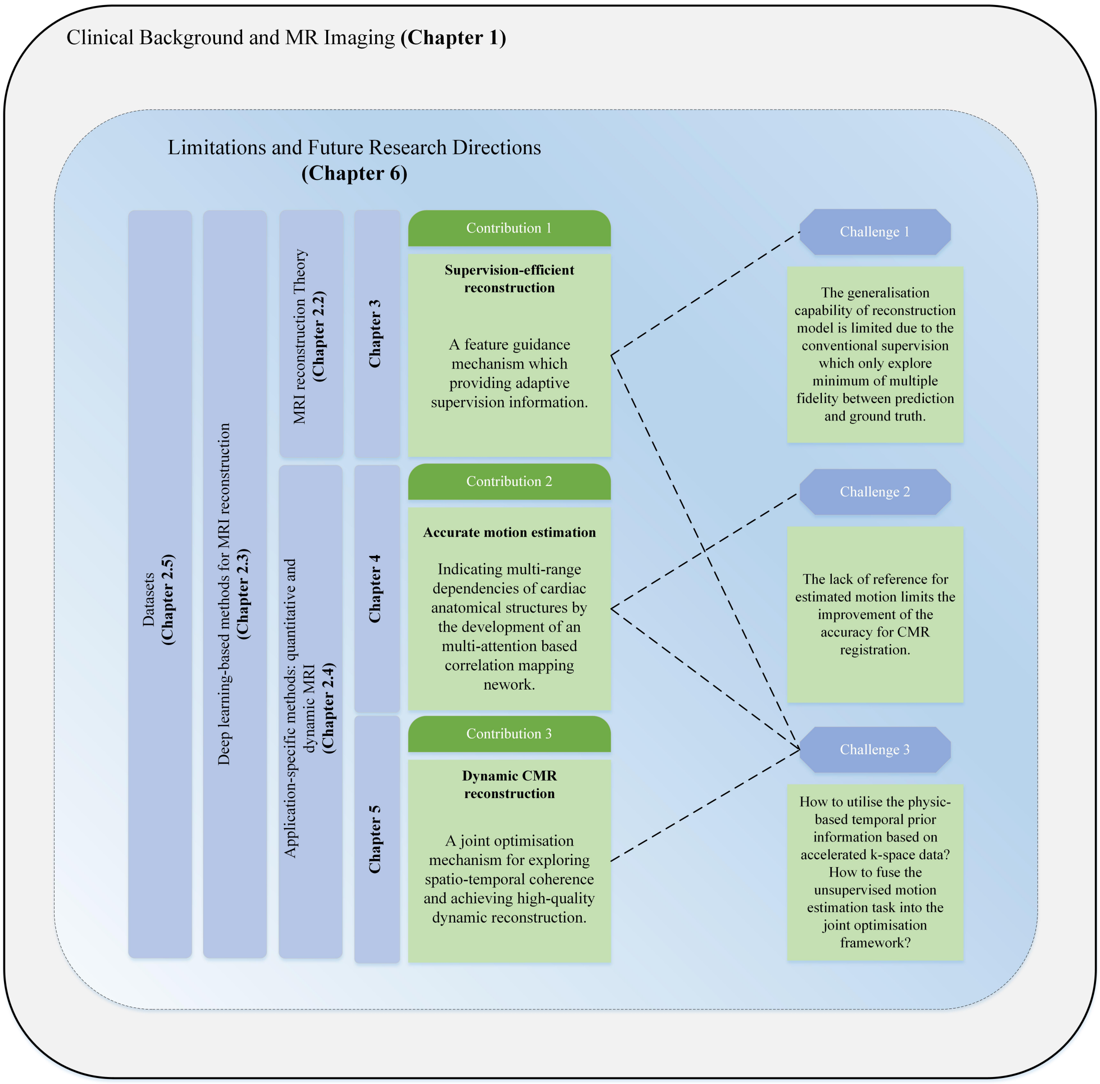


Figure 1.5: Diagram of thesis structure.

extremely high acceleration rates are applied in the simulation of clinical data acquisitions, in addition, this work focuses on addressing the nature-looking structures that are hallucinated in the prediction. It proposes an MRI reconstruction model, DAPGAN, which incorporates adaptive feature guidance as a supervision-efficient mechanism to the generative model in multi-level, instead of introducing extra loss functions.

Chapter 4 explores the global and local contextual information along channel and inter-spatial dimensions, focusing on improving the feature representation for dense estimation of cardiac motion in multi-scale. Especially for medical anatomical structures, emphasising meaningful features along different axes is a significant standard. The proposed model, MAPC-Net, rooted in a pyramid network with correlation mapping operation applied from coarse to fine, integrated with multi-attention-based cost volume calculating and spatially-adaptive instance normalisation (ASAIN), is capable of the refinement of intrinsic deformation fields. It shows great potential in multi-range dependencies and uses the relationship between local objects and global view embedded into correlation mapping, which is necessary for capturing motion fields of different scales.

Chapter 5 extends the foundational static MRI reconstruction and unsupervised motion estimation models by simultaneously computing the dual tasks while in a joint optimisation problem, in which iterative optimisation of reconstruction and motion estimation are carried out alternatively. In particular, a special motion estimation network is designed for addressing the inaccurate motion estimation as a result of undersampled or intermediate prediction during reconstruction. This can perform precise and rapid dynamic CMR reconstruction on full time frames of SAX CINE CMR, capturing high-quality spatio-temporal features.

Chapter 6 concludes the thesis, additionally, it discusses the limitations of this research and the potential directions of future works.

CHAPTER 2

Literature Review on Deep Learning-Based MR Image Reconstruction

2.1 Introduction

Magnetic Resonance Imaging (MRI) has been a foundational tool in clinical medical imaging due to its unparalleled capability for non-invasive assessment of anatomical structures, characterised by its exceptional soft tissue contrast. Its unmatched versatility in detecting subtle pathological changes and guiding therapeutic interventions, in comparison to CT, PET, and other imaging modalities, has solidified MRI as an indispensable instrument in contemporary diagnostic medicine. Specifically, in the context of cardiovascular diseases, Cardiac MRI (CMR) is recognised as a clinically pivotal technique for the comprehensive evaluation of various cardiac parameters, including cardiac morphology, function, perfusion, viability, and tissue characterisation, exemplified by myocardial strain analysis.

Nevertheless, a significant limitation of CMR lies in its inherently long acquisition time, driven by the necessary measurement of image features. This prolonged duration inevitably reduces patient comfort and imposes constraints on the achievable spatial and temporal resolution, as well as the range of contrast settings. Moreover, in the context of dynamic imaging, the requirement to account for cardiac and respiratory-induced motion of the heart further increases the scan time.

To overcome the obstacles of CMR imaging, various technical advancements have been proposed over the past decades. These include research on acquiring less data than traditionally required, in the Nyquist sense, to facilitate undersampled reconstruction techniques, compensating for information loss due to cardiac motion, and leveraging multiple receiver coils in the frequency domain for parallel imaging (PI). Compressed sensing (CS) has also emerged as a key technique for undersampled reconstruction, significantly accelerating CMR imaging. However, both PI and CS introduce active aliasing artefacts, which must be addressed through reconstruction algorithms to recover images that closely approximate fully sampled ones.

Recent efforts to benefit from machine learning have significantly expanded the regularisation options for image reconstruction, achieving performance that rivals SOTA approaches. Deep learning (DL)-based pipelines are now enabling clinical simulations with more aggressive acceleration rates while producing reconstructed images with higher accuracy compared with fully sampled data and more perceptual details in clinical practice.

In this chapter, we introduce the foundational theories, conventional methods, and state-of-the-art approaches related to CS-MRI and deep learning-based CMR reconstruction. The first section provides an overview of the basic theory behind MRI reconstruction. We then proceed with a detailed review of developments in the field, beginning with an introduction to the theory of deep learning-based reconstruction and its application as an alternative regularisation term in the compressed sensing (CS) optimisation problem. Our focus is on summarising the key methods and related theories for addressing MRI reconstruction challenges using deep learning. In the third section, we specifically review the application-specific approaches from the aspects of quantitative analysis and dynamic reconstruction for CMR. Then we review the most-utilised public dataset for CMR analysis and processing. In the last section, we discuss the current and consistent challenges related to general MR reconstruction and CMR-specific areas. Finally, we explore the future outlook of the field and present our conclusions.

2.2 Basic Deep Learning Networks

Due to the increasing requirements of accelerating imaging time and high accuracy of predictions, significant improvements have been achieved by DL algorithms and advanced computational hardware, for instance, GPUs. In particular, according to various medical imaging analysis tasks (e.g. classification, detection, reconstruction, segmentation, registration), multiple architectures of DL networks have been developed. In this section, we briefly introduce the three most widely explored basic structures for MRI reconstruction, which are fundamental networks adopted by conventional and state-of-the-art approaches in the area.

2.2.1 U-Net

U-Net[6] is the most widely adopted fundamental network for multi-tasks, as the superior ability for feature extraction and learning, despite the original task belonging to image segmentation in their proposed work. As depicted in Fig. 2.1, an encoder-decoder structure conducted the basic U-shaped framework with downsampling and upsampling operations, respectively. Each downsampling block includes several convolution layers and pooling layers, the process is termed as ’feature extraction’; the upsampling performs deconvolution layers as a mirrored structure compared with the encoder. A feature with different perception fields from coarse to fine can be reconstructed with different resolutions according to the design. In addition, skip connections are performed to facilitate fusion learning between local subjects from lower levels with fine details and contextual information from higher and coarse-level features. The main improved design between U-Net and autoencoder is skip-connection. U-Net has become the mainstream architecture for generation-based tasks. Approaches utilising U-Net as the backbone structure or a branch network have been increasing recently [7–11]. These works commonly conducted deep CNNs which leverage U-Net variant with different layers and conditions to achieve reconstruction from undersampled k-space data. In the work [11], a residual U-Net architecture enables interpolation of k-space information into spatial multi-channel features for learning is introduced.

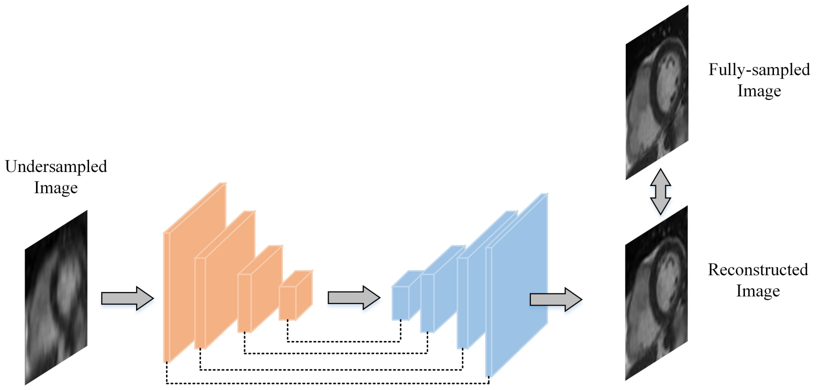


Figure 2.1: A basic framework for MRI reconstruction utilising U-Net architecture. The network takes undersampled image as input and conducts loss functions between reconstructed image and fully-sampled image.

2.2.2 Resdual Net (ResNet)

ResNet[12] has superior ability to solve the degradation problem of DL-based models in the training process. In the work of ResNet (see Fig. 2.2), a short-cut connection is proposed and conducted as a mechanism every two to three layers in a normal convolutional network design. The structure is the proposed residual block. This mechanism allows the ResNet-based network to enhance feature learning ability through a deeper structure while avoiding gradient vanish or explosion by directly increasing the amount of convolutional layers or pooling layers. In the work[13], multiple residual blocks were proposed and a feature-sharing block for temporal information was utilised to achieve dual-domain learning in the reconstruction network. Xue et al.[14] proposed a residual learning-based network for MR super-resolution task. The architecture enables extracting potential high-frequency residual features with sub-band residual module and reconstruction in a parallel style.

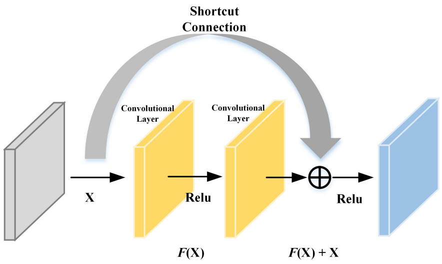


Figure 2.2: A demonstration of Residual-Net architecture

2.2.3 Generative Adversarial Network (GAN)

GAN[15] has achieved outstanding performance in medical image generation-related tasks, for instance, reconstruction or synthesis-related studies in the areas of MRI. It is proposed for improving the generalisation capability of natural images originally. As shown in Fig. 2.3, the design of the architecture follows the principle of zero-sum contrast problem, two sub-branches can facilitate improving the corresponding ability by confronting and constraining each other. The structure of GAN is indicated in Fig. 2.3, in which a generator aims to generate predictions that are close to the ground-truth value by training, and a discriminator to identify the generated images with the real data as accurately as possible. For MRI reconstruction, advanced development has been achieved by various GAN-based approaches[16], represented by CGAN, DCGAN, DAGAN, WGAN, LAPGAN, etc. GAN-based reconstruction network has received increasing attention, the related works are discussed in more detail in Section 2.4.2.

2.3 Image Reconstruction Theory

As the foundational theory that supports CS-based reconstruction solutions throughout the rest of the review, the general theory of MRI reconstruction serves as a key concept underlying MR image reconstruction.

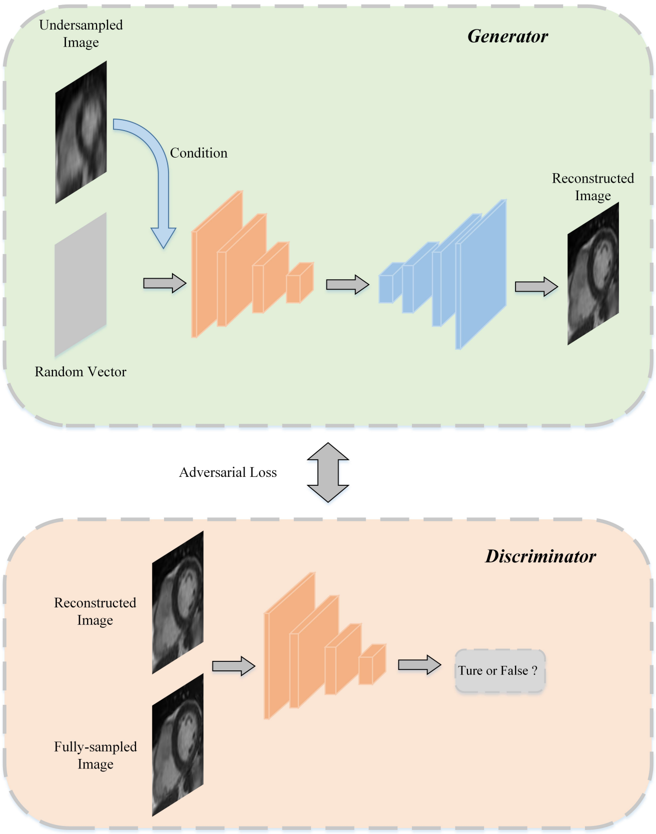


Figure 2.3: An illustration of GAN-based architecture for CMR reconstruction. The generator is a U-Net-based network, and the undersampled image is concatenated with a random vector as conditional input. The discriminator is conducted to discriminate the difference between the reconstructed image and the fully sampled image. The improvement of the generalisation ability for the generator is driven by the discriminator.

2.3.1 MR Reconstruction: An Inverse Problem

Given that general MR processing can be represented as a linear function for signal acquisition and feature domain transformation, the corresponding equation [17] is as follows:

$$s = E\rho, \tag{2.1}$$

where E represents the MR encoding operator, a complex entity composed of several components: the undersampling matrix, the Fourier transform operation and the coil sensitivity matrix. The variable ρ represents the image to be reconstructed, expressed in the spatial domain. Specifically, ρ typically refers to the distribution of proton densities and their relaxation properties within the tissue. In practical terms, ρ is the anatomical or functional image that we seek to reconstruct from the k-space data. While s stands for the acquired k-space data, which is a frequency domain representation of the MRI signals. The data in k-space is ultimately what the MRI scanner measures.

Thus, this process is formulated as an inverse problem, where the goal is to reconstruct ρ as closely as possible to the fully sampled signal, based on the observed encoding matrix E and the obtained signal s. Due to the ill-posed nature of the problem, MR image reconstruction does not satisfy the criteria of a well-posed problem. For the contextual research of MRI reconstruction, the conditions of existence, uniqueness, and stability of a solution are often not fully met. The key factors contributing to the ill-posed nature of MRI reconstruction include (1) the computational intensity of the problem, particularly when employing iterative and advanced reconstruction algorithms; (2) signal model simplifications, which may involve assumptions about the linearity, stationarity, or homogeneity of the magnetic field and tissue properties, leading to discrepancies between the model and actual physical phenomena, thus affecting the fidelity of the reconstructed images; (3) a reduction in the number of phase encoding steps acquired; and (4) the presence of multiple sources of noise during data acquisition, such as thermal noise, physiological noise, and system noise.

A reformulated form [17] which introduced the regularised optimisation additionally for MRI reconstruction problem is expressed as follows:

$$\hat{\rho} = \text{argmin}_{\rho} \, \|E\rho - s\|_2^2 + \lambda R(\rho), \tag{2.2}$$

where it is a trade-off optimisation problem to balance between the data consistency term $\|E\rho - s\|_2^2$ and the additional added regularisation term $R(\rho)$. The regularisation parameter λ indicates the weight of regularisation applied and should be selected based on the noise level in the sampled data. Particularly, the integration of sparsity priors and the statistical characteristics inherent in MRI to regularise the reconstruction problem have demonstrated considerable potential. In the following subsections, we will focus on reviewing compressed-sensing-based optimisation for the acceleration of MR imaging.

2.3.2 Compressed sensing-based MRI

As a mainstream approach to MRI acceleration, the methods that employ compressed sensing (CS) for MRI reconstruction are collectively referred to as CS-MRI[17].

Compressed sensing (CS)[18] is based on the principle of pseudo-random undersampling for acquiring k-space data in the frequency domain. During data acquisition, k-space data signals are sampled using multiple patterns with non-uniform strategies. Pseudo-random undersampling in MRI is strategically employed to optimise compressed sensing reconstruction by maximising incoherence, ensuring uniform coverage of k-space, and minimising predictable artefacts. This approach allows for balanced sampling of both central and peripheral regions of k-space, which is crucial for maintaining image resolution and contrast. The stochastic nature of pseudo-random patterns enhances the performance of optimisation algorithms used in CS, enabling them to exploit image sparsity more effectively. Moreover, the flexibility of pseudo-random sampling accommodates the specific constraints and requirements of MRI scans, facilitating faster acquisition and producing higher-quality images that meet clinical needs.

In the transformed domain, converted from the originally sampled frequency domain, the image is represented with properties that exhibit sparsity in the corresponding real space. To enforce the combined properties of data consistency and sparsity for MR features in this transformed domain, a non-linear reconstruction approach is employed. To achieve sparse sampling, the regularisation term $R(\rho)$ in Equation (2.2) can be replaced by an alternative option l_0 which conducts the specific coefficients for achieving sparsity. Based on the fact that utilising l_0 will lead to a non-convex normalisation, particularly when expansive datasets are applied. A prevalent substitution [18] employs l_1 to conduct convexity of normalisation as follows:

$$\hat{\rho} = \operatorname{argmin}_{\rho} \|E\rho - s\|_2^2 + \lambda\|\Phi\rho\|_1. \tag{2.3}$$

where $\Phi\rho$ in the formulation acts as the role of the regularisation term, a certain transform which conducted the image to be reconstructed to the sparsity assumption. The parameter λ controls the trade-off with the fidelity term of the acquired raw data, which is convex. From the view of conventional and iterative learning-based solutions [19], typical transforms utilised as the regularisation term here can be chosen from the temporal Fourier transform (FT) [20], spatio-temporal total variation (STTV) [21], and spatio-temporal wavelets [22, 23], and low-rank[24–26]. Building on the foundations of initial studies, data-driven methods became another mainstream of CS-based approaches. For instance, dictionary learning [27–29] proposed to leverage sparse transform combined with fixed dictionaries as priority information, compared with the regularisation role in a conventional way. An optimal dictionary is extracted directly from undersampled data to fill the missing lines in k-space.

Although conventional CS-based methods have achieved notable success, several critical drawbacks still hinder their full integration into routine clinical practice for CS-MRI. These limitations include: (1) the iterative process under non-linear conditions is time-consuming; (2) the optimisation of regularisation through parametric tuning is highly sensitive to various factors, such as undersampling trajectories, acceleration rates, anatomical complexity, and the type of sparse transform; and (3) the absence of a transformation bias guided by prior information makes it challenging to achieve efficient sparse representations. In particular, conventional CS-based methods rely on established mathematical transformations to enforce sparsity, using explicit, analytically defined transforms (e.g., Fourier, wavelet) or properties (e.g., total variation). The selection of these methods often depends on the specific characteristics of the data and the desired output properties. The limitations of traditional methods be leveraged as regularisation terms highlight the need for more robust, adaptive, and efficient approaches, particularly as the scale and complexity of data in compressed sensing applications have increased more recently. These challenges pave the way for the adoption of advanced techniques, such as deep learning-based regularisation.

2.4 Deep Learning-Based MRI Reconstruction: Architectures and Approaches

Based on the significant dependency of the specific sparse transform, especially in a situation which may be incapable of extracting and learning the complex anatomical structure of multiple organs, especially Cardiac MRI, the demand for efficient and robust reconstruction algorithms gives rise to the rapid adoption of deep learning-based approaches as an alternative protocol in CS-based optimisation problem.

In this section, we review the deep learning-based algorithms for MRI reconstruction. The approaches are categorised into the following four aspects: (1) end-to-end training; (2) generative networks; (3) self-supervised methods; and (4) dual-domain frameworks.

2.4.1 End-to-End Training-Based DL Approaches

Mapping Network: De-noising

Contrast with conventional CS-based reconstruction which leverages alternative transform in iteration methods, DL-based approaches for reconstruction explore optimising relative parameters for solving the task in advance, so it can be utilised as an analytical tool to solve Equation 2.3 to solve CS-based MRI problem. After altering the conventional one with explicitly described regularisation term with a mapping cost function $\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \operatorname{loss}\left(f_\theta\left(\mathbf{y}_i\right), \mathbf{x}_i\right)$ which is expressed as a neural network with parameters θ correspondingly, the optimisation problem is carefully defined. Thus, end-to-end training is commonly adopted for mapping the accelerated raw data, which is often affected by multi-degree noise and artefacts, to the corresponding fully-sampled data, as the ground truth.

Early architectures were developed to map image degradation with aliasing artefacts to finely sampled images. One notable approach, AUTOMAP (Automated Transform by Manifold Approximation) [30], investigates the forward model for transforming the data from the frequency domain to the image domain. Another commonly applied architecture is the convolutional neural network (CNN)[31], a widely adopted method that was first introduced to address the challenge of mapping zero-filled images to fully sampled ones. This approach demonstrated preliminary success, producing convincing results with detailed reconstructions.

For CNN-based approaches, in particular, depending on the specific tasks, various downstream architectures have evolved based on the foundational theory of mapping noisy measurements to high-quality results. These include: (1) UNet-based methods[32, 33]: these methods capture spatial hierarchies and details across multiple scales, making them particularly effective for static image reconstruction. The work [33] introduced a U-Net-based method with an improved module named Faster Fourier to accelerate the reconstruction speed for multi-dimensional MRI reconstruction. However, purely U-Net-based methods face challenges in extremely undersampled k-space data, where the global receptive field may fail to capture significant features. (2) ResNet-based methods[34]: By introducing skip connections (residual connections), ResNet helps mitigate the vanishing gradient problem, making it well-suited for very deep architectures. This approach is robust and capable of learning complex transformations, making it useful for both static and dynamic MRI, especially in preserving fine details. (3) RNN (Recurrent neural networks)[35, 36]: specialized for tasks involving sequences or temporal data. While less common for static MRI, they are valuable in dynamic MRI reconstruction. In the work[35], they established a framework that integrates extracted regularisation as prior information into a recurrent network for T1 mapping. Nevertheless, the complexity of the RNN limits the transfer learning applicability to other types of MRI scans or datasets from other resources. (4) Another noteworthy category is transformer-based methods for MRI reconstruction[37] more recently. Originally developed for natural language processing tasks, transformers have shown exceptional promise due to their ability to model long-range dependencies and integrate contextual information effectively. By operating on the principle of attention, transformers can identify crucial spatial and temporal patterns in complex imaging data.

Unrolled Networks

Unrolled networks indicate a noteworthy development in the area of MRI reconstruction. It is a novel type of DL-based method that unfolds the iterative optimisation utilised in conventional MRI reconstruction into a series of CNN layers. It should be noted that this mechanism allows for integration with end-to-end learning, where the algorithm is trained to learn the mapping relationship from undersampled raw data to the reconstructed results. This approach conducts the fusion between the advantages of model-based optimisation methods with the flexibility of data-driven learning, contributing to the improvement of image quality and saving reconstruction times.

The work[38] first proposes the concept of an unrolled mechanism, which is a CNN-based algorithm with initialization of a specific iterative-based method, while efficiently estimating the corresponding sparse code for obtaining good quality visual inspections. For the context of CS-based MRI reconstruction[39, 40], they all derived from the iterative procedures optimizing a CS-based MRI model, and adopted deep architecture consists of multiple stages, or unrolled cascaded sub-networks, each of which corresponds to an iteration in the algorithm. Fabian *et al.* [41] introduced a hybrid framework that fuses CNN with transformer blocks within an unrolled network from coarse to fine.

A classic and pioneering work [42] firstly proposed a generalisation-based compressed sensing reconstruction algorithm which is embedded in an unrolled gradient descent scheme, which is an alternative term for the regularisation $R(\rho)$ in the Equation (2.2) integrated with a variation model. The iterations of gradient descent can be conceptualized by conducting a neural network framework, where each iteration involves two key elements: enforcing data consistency (based on the gradient of the least-squares loss function) and applying a modified term through a convolutional network (derived from the gradient of the regularisation factor).

$$\mathbf{x}_{t+1} = \mathbf{x}_t - \eta_t \mathbf{A}^T \left(\mathbf{A}\mathbf{x}_t - \mathbf{y} \right) + \mathrm{CNN}_t \left(\mathbf{x}_t \right) \tag{2.4}$$

where the weight value η_t and the network parameter $\mathbf{x}_t$ can be updated during optimisation.

For CMR, a significant difference is capturing the heart motion in a complete cycle, leading to heavily affected image quality and more complex imaging dimensions, such as 2D with a temporal component (2D + time) or 3D volumetric data combined with time (3D + time). These additional dimensions significantly increase the computational demands of the reconstruction process. Here we review some representative approaches.

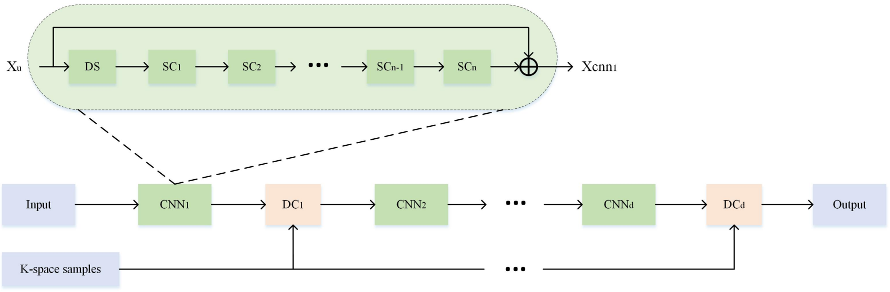


Figure 2.4: An illustration of CascadeNet architecture. DS denotes data sharing block, SC is sub-cascade network, DC represents data consistency block.

Schlemper *et al.*[40] introduced a novel framework which is formed in an unrolled global structure, termed CascadeNet (see Fig. 2.4). This algorithm cascades multiple sub-networks of identical structures, demonstrating that repeatedly applying Convolutional Neural Networks (CNNs) for regularisation learning can significantly enhance the reconstruction of highly under-sampled features through iterative learning. The cascade approach addresses the limitations of single CNN structures and single encoder-decoder models in conventional style, which typically require a large number of parameters and exhibit high computational complexity. One of the key innovations is the work introduced data consistency block for employing a mapping supplementation with original sampled k-space data, a trade-off linear combination to merge the Fourier-transformed estimates from the previous module with the original measurements is conducted depending on if the k-space data leveraged to reconstruct the regularisation term is sampled. Additionally, the work explored temporal redundancy along the complete time frames from spatial deformations between adjacent frames. Thus, data consistency in k-space can be further optimised by not only leveraging data from the previous cascaded sub-network but also by concatenating it with transformed k-space data from neighbouring time frames.

However, during the training procedure, facilitating the unrolling architecture requires multiple consecutive CNN blocks for feature extraction, which improves the complexity of signal processing from k-space, and their performance will be limited in high-dimensional (e.g., 3D, 4D) application scenarios. The works [43, 44] introduced a concept similar to CS processing, which is applied to improve memory utilisation efficiency during training procedures. It is used to achieve data consistency by alternating iterations and is based on gradient descent algorithms and conjugate gradient algorithms based on fractional functions. Zimmermann *et al.*[45] explored the inter-coil relationship from the acquired data by an unrolled network which aims to achieve an estimate of the coil sensitivity map implicitly. Another work[46] proposed a density-compensated method based on unrolled architecture for non-Cartesian sampled MRI reconstruction, which shows the advantage of improving the reconstruction quality across spiral and radial sampling, offering high speed and better aliasing reduction in such strategies.

2.4.2 Image prior-based Generative networks

GAN-based methods

While general CNN-based approaches in previous have shown significant progress in addressing the inverse problem of CS-MRI than conventional CS methods, their primary benefit over traditional approaches remains the aspect of faster reconstruction times. However, to fulfil the practical demands of clinical MRI, deep learning must deliver markedly higher imaging quality than conventional CS-MRI techniques, along with enhanced robustness to various imaging factors. A significant challenge, as in other deep learning applications, is determining the optimal network architecture. Additionally, achieving convergence during deep network training is often challenging and necessitates extensive parameter adjustments.

To address these challenges, a fine-grained and computationally efficient method is urgently needed as the next improvement direction. The emergence of Generative Adversarial Networks (GANs) offers a compelling solution. The superior ability of GANs to balance the accelerating convergence and reduce the need for exhaustive parameter tuning makes the framework an ideal solution to enhance CS-MRI capabilities.

In particular, GANs provided significant improvement in enhancing the generalisation ability of the network through the adversarial loss function. Adversarial loss contributes to ensuring the reconstructed images are not only accurate but also realistic, thus effectively reducing common artefacts and improving overall image quality. Regularisation is essential to maintain the fidelity of the reconstructed images while ensuring that the optimisation process is more robust and efficient.

Generative Adversarial Networks (GANs)[15] can utilise prior information based on a known probability distribution. The primary goal of adversarial learning is to adjust this probability distribution by minimizing the distance between the generated results and the set of labels. This approach eliminates the need for explicit input-related assumptions. However, the accuracy of the final output from the generator can be significantly influenced by the actual probability distribution between the real data utilised in the discriminator and the generated output. Specifically, when the generator and discriminator reach a state of equilibrium, they can effectively approximate each other.

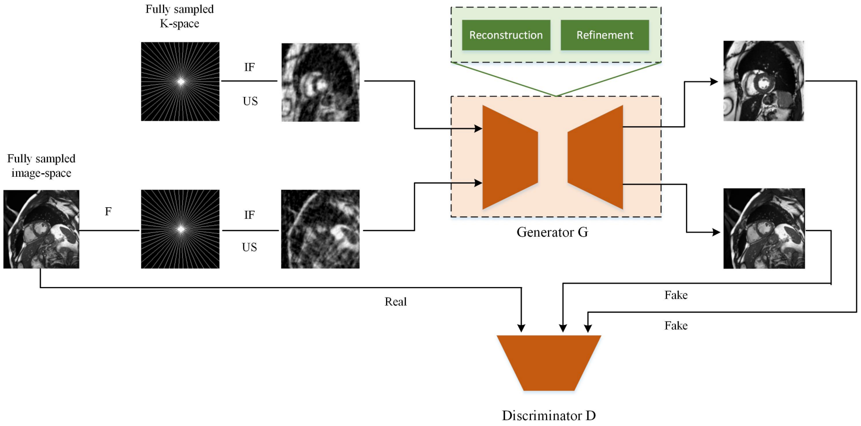


Figure 2.5: The framework of RefineGAN.

In recent years, Generative Adversarial Networks (GANs) have gained significant interest in the field of MRI reconstruction. A notable example is the pioneering work of DAGAN (Deep De-Aliasing Generative Adversarial Networks)[47]. This innovative approach introduced a conditional GAN architecture, incorporating a refinement-learning phase. It also employed a composite loss function that integrated both adversarial and perceptual components. In particular, DAGAN optimized MSE loss in both image domain and frequency domain. This optimisation concept inspired subsequent developments in MRI reconstruction, for instance, another pioneering work, RefineGAN[48], which proposed cyclic loss for k-space data consistency with the observed matrix based on accelerated signal (see Fig. 2.5). Further, the network is performed based on a chained architecture to achieve better performance on reconstruction quality. The architecture is inspiring for the related tasks by proposing a variant that achieves the combination of residual auto-encoder and GANs. By fusing the attention module with cyclic loss in the GAN framework, Li *et al.*[49, 50] introduced a framework, utilising residual embedded U-net which conducted multi-wise attention that enhances the feature extraction on specific image regions, further improved the learning ability for relevant features in high-frequency level. Shaul *et al.*[51] proposed a GAN-based network to directly predict the k-space data as close as to the full sampled data, to be used as priority information for image domain reconstruction and refinement. The work [52] described an outline of most adopted GANs architectures and applications, however for brain MRI, it was also demonstrated that on a different dataset, the potential to generate synthetic MRI data that enable extract and predict real distribution was confirmed.

Despite their potential, GAN-based approaches for MRI reconstruction face significant limitations, primarily due to the need for extensive and diverse training datasets. These datasets are crucial for ensuring the generalisation of the models across various anatomical regions, imaging sequences, and pathological conditions. Moreover, the effectiveness of these methods hinges on the availability of substantial ground-truth data to train the generator adequately. A critical challenge is the tendency of these approaches to introduce visual hallucination or propagate the artefacts, particularly when dealing with highly under-sampled data.

2.4.3 Self-supervised Methods

CNN algorithms, typically trained in an end-to-end format, generally require ground truth as a reference value to guide supervised training. However, in clinical practice, access to this gold standard may not always be feasible due to several factors, such as limited scan time, signal decay effects along echo trains, or physiological motion. As a result, there has been growing interest in weakly supervised or self-supervised algorithms. Recent strategies have focused on reducing or eliminating the need for reference values by simulating the generation of measurements or partial measurements, or by utilising conditional noise information to train the network. These approaches have shown promising results in recent years.

DL Methods Based on noise measurement assumption

Assumptions about noise distribution have long been a foundational approach as an unsupervised solution. A theory known as Stein Unbiased Risk Estimation (SURE)[53] originally introduced this concept to address such problems.

Building on this theoretical estimation, some studies[54, 55] have employed SURE as an unsupervised penalty term for image noise processing. These approaches have demonstrated quantitative performance comparable to those achieved by methods trained on labelled data. To enhance the network's ability to fit Gaussian random noise, the work[54] trained each network layer by incorporating a constraint layer based on the generalised SURE within each individual layer. Despite the assumption that the residual aliasing error after each iteration approximates Gaussian random noise, this is not crucial for methods based on the optimisation of inverse problems. Consequently, this method falls short of the performance achieved by supervised approaches.

Numerous SURE-based methods require estimation of the statistical properties of the noise and have high computational resource requirements as the computation involves second-order derivatives. Overall, such methods rely on accurate knowledge of the noise model, and their effectiveness can be compromised if the noise distribution assumptions are inaccurate.

More recently, increasing works[56–59] leverage advanced deep learning architectures, such as diffusion models, score-based models, and generative models. These architectures are often unrolled or iteratively applied, with an emphasis on maintaining data consistency and incorporating domain-specific knowledge. They often incorporate statistical or probabilistic models to estimate and mitigate the noise in the reconstruction process, thereby improving image quality. Nevertheless, these methods rely heavily on accurate assumptions about noise distribution, which may not always hold in practice. Incorrect noise modelling can lead to sub-optimal reconstruction results. In particular, some of these methods, especially those involving diffusion models, can be computationally intensive, potentially limiting their applicability in real-time or resource-constrained environments.

DL Methods based on Noise2noise

In contrast, Noise2Noise-based methods do not require prior knowledge of the statistical distribution of the noise or reliance on a specific noise model. Instead, they operate using pairs of noisy images, which broadens the applicability of these methods. However, this approach necessitates the consistency of noise characteristics across the image pairs. Noise2Noise [60] is a widely used framework that is well-suited for constructing self-supervised losses. It operates on pairwise noise measurements that are consistent with the noise characteristics.

To formulate the procedure, by conducting an inverse problem from compressed sensing, but assuming two accelerated acquisitions from different random samplings, while $\mathbf{F}$ is the Fourier transform. The self-supervised loss then can be expressed from the pair-wise measurements of noise:

$$\ell_{\text{self}}\left(f_{\boldsymbol{\theta}}(\mathbf{y}), \mathbf{y}'\right) = \left\|\mathbf{R}'\mathbf{F} f_{\boldsymbol{\theta}}(\mathbf{y}) - \mathbf{y}'\right\|_2^2 . \tag{2.5}$$

Huang *et al.* [61] proposed a self-supervised denoising network with a plug-and-play approach, named DURED-Net (see Fig. 2.6), which established a fusion between a novel regularization utilising imaging physics with the reconstruction process, thus improved the performance of the model.

Yaman *et al.* [62] is one of the significant works to achieve high accuracy based on mapping pairwise noise properties and implementing unsupervised learning. This algorithm proposes to use the observed k-space sampled values for two mutually independent subtasks. The first is the implementation of data consistency at each stage in the unrolled network, a module that no longer requires fully sampled k-space data, but observed data. The second is that this observed data can be used if supervised training. The experimental part of this method also demonstrates that its reconstruction capab-

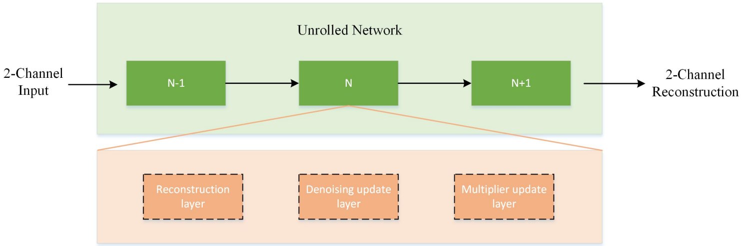


Figure 2.6: The architecture of DURED-Net.

ility can be of the calibre of fully supervised learning methods. However, variations in noise characteristics can affect the reconstruction quality, furthermore, they still need large datasets of noisy images to achieve high performance. In addition, may struggle to generalize across different types of noise or MRI acquisition protocols. This can limit their applicability in diverse clinical settings, subtle artefacts may still be introduced during the reconstruction process, especially when the underlying assumptions about noise do not fully hold.

DL framework based on k-space bands

Recent advancements in self-supervised MRI reconstruction have largely concentrated on sparse sampling techniques such as compressed sensing. These methods aim to reduce the effects of noise and artefacts resulting from limited sampling times, through the performing of under-sampling masks for accelerated scans. Additionally, there is growing interest in high-resolution reconstruction within the field of MRI reconstruction. Unlike the aforementioned accelerated sampling strategies, high-resolution reconstruction specifically tackles the challenges arising from incomplete K-space coverage by the sampling mask. This incomplete coverage limits the original resolution, which poses significant challenges for subsequent reconstruction processes.

For the simulation with only local k-space undersampling, the self-supervised learning methods[63–65] can effectively solve the prior knowledge based on low resolution, so as to realise the high-resolution reconstruction of the complete frequency domain data. The essence of the method based on the K-space band is to reconstruct the frequency domain data in different bands. These methods leverage the inherent structure of the MRI frequency domain by learning the potential connections between various k-space bands, for instance, low-frequency and high-frequency. This mechanism allows models to be trained without relying entirely on paired undersampled and fully sampled data. Commonly employed architectures include stochastic gradient descent over k-space subsets, dual-domain learning, and matrix completion theories, the common target is to leverage partial k-space data effectively. Moreover, these methods enhance generalisation across different types of MRI scans by enabling models to understand the underlying physics and structure of k-space data, rather than relying solely on image-space features. This is effective even when dealing with varying anatomical structures or diverse acquisition protocols. However, the success of these approaches depends on the careful design of loss functions and strategies to ensure the model accurately captures the relationships between different k-space components. Additionally, while these methods emphasize local or band-specific information within k-space, they may sometimes result in sub-optimal reconstructions due to limited consideration of the global context and an over-reliance on specific sampling patterns.

2.4.4 Dual-domain and Transformer-based Frameworks for MRI Reconstruction

Most existing research focuses on enhancing reconstruction quality under accelerated sampling by improving network structures, supervision techniques, or iterative methods within a single image domain. However, a significant limitation of these approaches is the lack of consideration for signal learning in k-space as a parallel branch. While some studies[66–68] do incorporate k-space learning, it is often relegated to a secondary role, primarily used to compensate for image domain feature learning and to ensure data consistency.

Currently, some researchers explore achieving a higher performance by learning features and reconstruction in both the original sampling domain - k-space, and the image domain after the Inverse Fourier transform, known as dual-domain networks, which has suggested a new perspective for MRI reconstruction. Approaches of this type can be roughly divided into two groups according to the implicit and explicit roles. First, some works [66–68] learn the k-space and spatial domain data separately, and learn the feature information in their respective domains alternately through different network structures and orders. However, ignoring the data sharing between the information of the two domains limited the ability of internal connection across dual domains, and sub-networks all performed in the form that learns separately and sequentially, even though different methods have their own characteristics in arranging the learning network order in the spatial domain and frequency domain. At the same time, these methods only implicitly add certain frequency domain data consistency features to the reconstruction results as a priority prior, which may ignore the internal interactions between the two domains. Different from the previous methods, MD-Recon-Net[69], as the pioneering approach of integrating parallel learning with parameters sharing and cross-fusion learning in this field, by using the method of alternately cascading parallel learning and fusion learning of dual-domain networks, and arranging data consistency as a prior during fusion learning, the efficient use of dual-domain latent features can significantly enhance the quality of reconstruction. Souza *et al.* [70] demonstrated that multi-channel configurations are also beneficial with hybrid domain, which emphasises the importance of domain-specific processing in MR reconstruction. However, the multi-coil approaches require training separate models for each coil-channel configuration, which leads to reducing its flexibility for network optimisation, and improving the complexity of computing. Furthermore, complex-valued data are not supported with corresponding defined complex networks, which limits the potential performance improvement. Wang *et al.* [71] proposed a dual domain network for MRI reconstruction with the guidance from motion corruption through the uncertainty of k-space, similar dual domain reconstruction block and k-space uncertainty block with [69] are applied in this work. The regularization term conducted by the motion correction embedding operation offers a new solution for parallel learning.

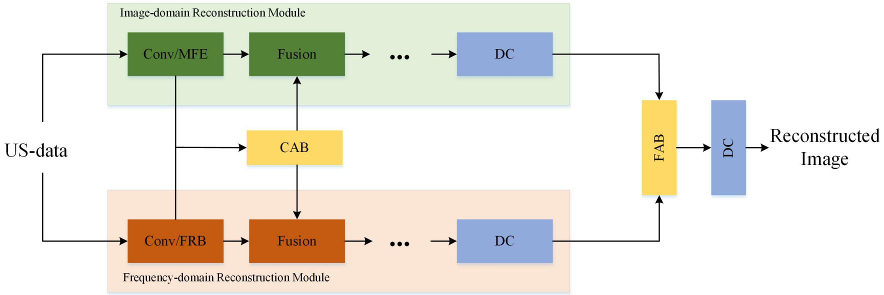


Figure 2.7: The framework of DCT-Net. The elements included in the architecture: MFE is Multi-scale Feature Extractor, FRB is Frequency Reconstruction Block, CAB is Cross Attention Block, DC is Data Consistency block, FAB is Fusion Attention Block, US denotes undersampling.

Recently, Transformers[72, 73] have introduced attention-based adaptive learning to the reconstruction problem, allowing networks to assign different weights to input data based on their varying importance. Originally proposed in the field of natural language processing, Transformers have extracted increasing attention and demonstrated essential improvement of performance in computer vision tasks, such as motion analysis and image recognition, due to their ability to capture long-range dependencies. The concept of integrating Transformer with dual-domain MRI reconstruction is represented by the work named SwinGAN [74], which allows capturing long-range dependencies, thus enhancing the learning of complex anatomical structures within MRI images. Transformers, in particular, can be adapted to various data modalities and reconstruction tasks, offering flexibility within reconstruction pipelines. Additionally, they help reduce the impact of noise and artefacts as a constraint combined in the regularisation conducted in dual-domain. Wang *et al.*[75] introduced another transformer-based network for MRI reconstruction, known as DCT-Net (see Fig. 2.7), which operates in a dual-domain framework. To enhance the efficiency of dual-domain information fusion and improve the capture of organ-specific characteristics, DCT-Net introduces two key modules: the Cross Attention Block (CAB) and the Fusion Attention Block (FAB). These modules facilitate efficient information sharing between domains, ensuring data consistency in the frequency domain while compensating for the neural network's limitations in extracting fine structural details at small scales. Although this network design shares similarities with our work discussed in Chapters 3 to 5, the dual-domain adaptive reconstruction method proposed in this paper advanced DCT-Net in several aspects. Our approach integrates an attention mechanism with motion estimation compensation, offering greater efficiency, reduced computational complexity, and suitability for multitasking scenarios involving both reconstruction and registration.

However, the transformer-based frameworks may not be the optimal choice for the task of CMR reconstruction. Specifically, transformer-based algorithms demand large-scale datasets for effective training, which results in substantial computational complexity and memory usage. This can be a significant limitation for practical deployment. Additionally, the complexity of the transformer architecture significantly increases training time, which is particularly not afforded when the goal is to save scanning time by using accelerated data in clinical practice.

2.5 Application-Specific Methods

2.5.1 DL methods for quantitative MRI

In recent years, cardiac MRI has been recognized as the gold standard for various quantitative analyses of the heart, including tissue characterization, volumetric assessment, functional evaluation, and mass measurement. The introduction of novel approaches, such as myocardial mechanics and 4D flow imaging, has further enhanced the diagnostic capabilities of these technologies.

Among various quantitative analyses, T1/T2 contrast mapping, which is conducted to assess tissue parameters, is an essential task of quantitative MRI.

Hwan *et al.*[76] validated a deep learning approach that automates T2 mapping by accurately segmenting the left ventricular myocardium and detecting elevated T2 values, indicative of myocardial inflammation. Their study demonstrated that this automated method performs comparably to expert radiologists' manual assessments. The software exhibited high sensitivity and specificity in identifying these abnormal T2 values. Wanni *et al.* [77] introduced an optimised encoder-decoder network for precise segmentation of cardiac anatomical structures. Enhancing the Dice coefficient and Hausdorff Distance (HD) index for the left ventricle (LV) and right ventricle (RV) is a crucial improvement for calculating physiological parameters and diagnosing cardiac diseases. This improvement contributes to more accurate cardiac function measurements and supports real-time practical diagnostics. Yan *et al.*[78] developed an advanced algorithm that enhances segmentation accuracy through multi-channel, fully convolutional feature extraction using annular shape level-set methods. This innovative approach demonstrated robust segmentation, which is essential for precise evaluation of clinical indices such as myocardial wall thickness and ventricular volume. To achieve more efficient reconstruction and accelerate the speed of mapping between T1/T2, Yohan *et al.*[79] established a zero-shot and low-rank-based DL method. In this work, the scan-specific model is integrated with subspace modelling technology to improve data fidelity, facilitating better monitoring capabilities for quantitative MRI. Chang *et al.* [80] introduced a framework utilising physics guidance. Accurate patient mass density maps from MRI data can be predicted with improved quantitative performance. In particular, utilising the achieved results, the work further obtained an optimal performance on proton therapy planning, which also reduced the range uncertainty.

2.5.2 DL methods for dynamic MRI

To address the fundamental challenge in high-quality MRI acquisition, a time-consuming problem, there arises a significant and complex issue: improving the reconstruction quality of MRI while extremely decreasing the scan time, which is an aggressive balance, especially when aiming to capture the necessary frequency measurements. This balance becomes even more challenging when considering the theoretical and practical aspects of performing joint optimisation tasks[81, 82] with accelerated data, which is a highly corrupted measurement. Image reconstruction and registration are two basic tasks in computer vision, serving as essential tools for the processing and analysis of MRI. They play a pivotal role in improving our understanding across various clinical applications, such as diagnostics, surgical planning, and radiotherapy.

Given the motivation to avoid image degradation and improve both spatial and temporal resolutions in cardiac MRI, particularly when addressing the challenges of limited acquisitions and motion correction, motion-corrected algorithms have recently gained popularity. These algorithms are effective in learning spatio-temporal correlations and reconstructing images from highly undersampled sequential data.

The motion guidance strategy for dynamic MRI reconstruction can be roughly categorised into two branches: the implicit mechanism for dynamic reconstruction and the explicit mechanism for motion-compensated reconstruction.

Implicit Motion Correction for Reconstruction

Dynamic reconstruction approaches in an implicit scheme normally leverage the temporal redundancies as separate prior knowledge without explicitly calculating deformation fields or addressing motion as an explicitly regularisation term. K-t domain frameworks[83–87] serve as an efficient reconstruction mechanism without attention to the non-rigid cardiac optical flow. These methods perform Fourier transform along the temporal axis, utilising the k-t domain particularly to achieve dynamic reconstruction by conducting the corresponding redundancies. Based on the research of previous stages, to further obstacle the limitation that most compressed sensing-based approaches focused on single-channel only, also motivated by parallel imaging reconstruction, Qin *et al.*[85] established a model utilising deep recurrent architecture to formulate the multi-variable minimisation problem. In addition, a dynamic parallel MR reconstruction process is performed, combined with the regularisation which is conducted in both spatio-temporal and frequency-temporal domains. Zou *et al.*[88] introduced a method that jointly learns the parameters of the deep network and the latent vectors for each slice, capturing motion-induced dynamic variations from the k-t space data.

A new direction based on utilising low-rank(L) and sparse (S) matrix completion as the components to form the reconstruction problem, gained increased attention as an optimised and extended theory based on CS. To conduct the reconstruction by leveraging the sampled sparsity from the transform domain and furthermore, the rank deficiency, some works[86, 89, 90] explored the concept by a combinational minimisation with L1 transform coefficients normalization and nuclear normalization. In addition, [91–95] investigated solutions based on another concept by an alternative algorithm involves disaggregating the sampled signal with a fusion of low-rank and sparse parameters linearly. The category of these approaches is defined as L plus S (L + S) decomposition. In particular, extended from the work [93], Otaza *et al.*[94] presented the new L+S model which enables facilitating the reconstruction from the aggressive acceleration and background separation for improving the resolution in spatio-temporal domain. In this model, the L+S reconstruction is formulated based on multi-coli sampling and optimized by a convex design.

Explicit Motion-compensation for Dynamic MRI reconstruction

Accurately estimating the deformation fields of the heart is a significant step in forming the motion-compensated MRI reconstruction. In particular, embedding the temporal dimension into the procedure of reconstruction of spatial dimension is essential for achieving dynamic optimisation. Nevertheless, this task is particularly challenging for certain CMR techniques, as in some cases, obtaining fully sampled data as ground truth for motion estimation is impractical, which is essential for robust training. The complexity of calculating cardiac motion fields makes it difficult to obtain precise, fully annotated datasets in clinical diagnoses, further complicating the deep learning (DL) reconstruction process and presenting significant challenges for CMR reconstruction.

As the most adopted and researched pipeline, dynamic measurements based on the observation of motion fields describe the deformation vector with optical flow theory and can be utilised as an explicit motion compensation branch of the global joint optimisation for reconstructing MRI. Batchelor *et al.* [96], proposed the pioneering work which established a framework that formulated the mechanism of embedding general nonrigid motion estimation to the general reconstruction with the conjugate gradient method. Thus, a basic motion-compensated MR reconstruction pipeline is explored.

There have been different attempts to perform motion estimation integrated into the process of MR image reconstruction in the community, the explicit mechanism included two sub-tasks that are facilitated either separately or most recently jointly.

Most MR image reconstruction frameworks performed with motion compensation by a solution that includes the estimated deformation fields in a sequential fashion[21, 97–105] which are established on the foundation of [96]. For example, the work [100] established an MRI reconstruction algorithm based on 4D total variation with estimated motion, furthermore, they proposed a CS-based model which is employed in an iterative-based optimisation with distinctive architecture, the motion information is embedded in a framework of joint architecture for reconstruction. The framework proposed by Qi *et al.*[102] disentangling the joint task into two stages, the motion will be estimated in the first stage for the inter-frame motion vectors and then completing the following refinement for the subsequent reconstruction, as a separate matrix to achieve constraint. However, since the deformation field is estimated from artefact degradation with blurring affected quality, the temporal information is corrupted with artefact heavily, and individual optimisation and processing for the sub-task is time-consuming and increases the cost as a result of computational complexity.

By performing the reconstruction and motion estimation sub-tasks into a joint optimized mechanism that is based on combining the intensity-based motion constraint with the CS scheme, the reconstruction is refined and driven by the motion estimation from undersampled data. The two tasks, image reconstruction and motion estimation, are executed alternately under the global framework to achieve complementary optimization in iterations. Odille *et al.*[106] introduced a comprehensive reconstruction framework known as GRICS, designed to correct various types of motion, including non-rigid and affine motion. This framework facilitates fully auto-calibrated reconstructions by deriving coil sensitivities and motion model coefficients directly from corrupted raw data. By predicting motion through external sensors and optimizing the motion model, GRICS effectively integrates motion correction into the reconstruction process. While maintaining focus on tackling the challenge of imaging during free-breathing, and addressing issues related to motion artefacts for improving the reconstruction quality, Odillel *et al.*[107] extended their work by facilitating more motivations including a joint optimization framework for the simultaneous reconstruction of multiple MR images and a non-rigid motion model, contrast with the primary work, the significant difference is this model accounts for both intra-image and inter-image motion, effectively correcting ghosting/blurring artefacts and misregistration. In addition, the progress represents a shift from a generalized framework for motion correction to a more specialized and clinically validated approach. The latter leverages joint optimization of multiple images and a sophisticated motion model to achieve higher image quality and alignment, specifically validated in myocardial T2 quantification for heart transplant patients. Most of the works for CMR consider pair-wise registration for the deformation field estimated in parallel for compensation, in contrast, Royuela-del Val *et al.*[108] proposed a robust group-wise nonrigid motion estimation technique specifically designed for dynamic cardiac cine MRI sequences, which facilitate the potential of improving the accuracy of motion estimation by considering the motion of multiple frames simultaneously rather than pairwise. Zhao *et al.*[109] introduced a framework that facilitates joint MR image reconstruction by proposing a general pipeline supporting various sub-tasks. In addition, they designed an affine model for motion estimation from coarse to fine. Research of joint MRI reconstruction based on dictionary learning[110] applied another solution, the optical flow approximation is compensated in the alternative problem by extracting the sparse dynamic representation over a learned dictionary. Inspired by the work[111], which introduced a generally formulated principle of combining compressed sensing with optical flow components, Aviles-Rivero *et al.*[112] introduced a new framework based on computationally tractable variational mechanism. It employs a multi-scale network for motion prediction in particular, further, the algorithm performed a model of CS+M as a simultaneously addressed framework which enables employing primal-dual splitting, variable splitting and regularization for CS-based reconstruction and motion estimation, incorporating the data fidelity term. However, the minimisation theories for alternating optimization of the reviewed works above are heavily affected by the cost of prediction time, as a result of the nature of iterative-based approaches.

Developed as the alternative time-efficient term, deep learning (DL) methods have been devoted to addressing the previous limitations of dynamic motion-compensated MRI reconstruction. Recently, by leveraging the outstanding feature extraction ability

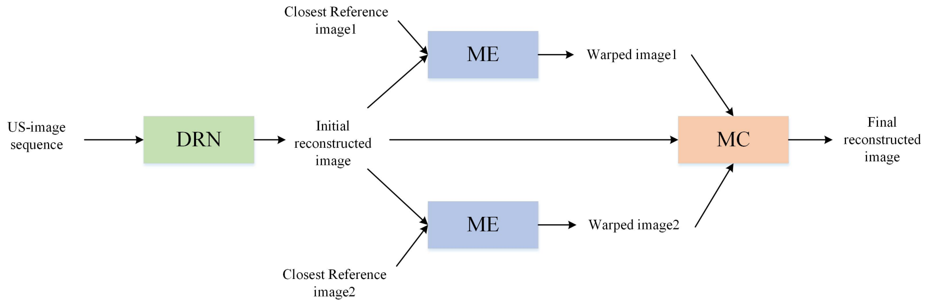


Figure 2.8: The framework of DRN-Net.

with multi-scale of U-Net as the backbone, a number of representative approaches[113–117] that explicitly injection for deformation field into the reconstruction procedure via compensation has emerged. Among these methods, Huang *et al.*[117] proposed a recurrent CNN network driven by a novel loss for estimating the motion field in terms of long-distance, thus a motion-guided framework including dynamic reconstruction, and motion processing is formulated (see Fig. 2.8). The limitations of specific architecture for estimating the aggressive noise-affected deformation vectors and the universality for both single and multi-coil channels have not been verified by these methods. Based on two fundamental studies[118, 119] for non-rigid cardiac motion estimation and group-wise motion-guided learning for CMR reconstruction, Pan *et al.*[120–122] developed their work to a systematic mechanism, which combined the elements of deep learning-based architecture, unrolled optimization with an embedded group-wise motion estimation network. In particular, most of them enable multi-coli parallel imaging. The second stage of each unrolled block is the reconstruction procedure with implementing CG-SENSE[123] as the course of the motion-compensated CMR reconstruction iteration.

2.6 Datasets and Software

2.6.1 UK Biobank (UKBB)

The UK Biobank Dataset (UKBB) is a widely utilised dataset resource in biomedical-related projects, offering a comprehensive data platform with broad coverage of categories and detailed classification across various research domains. Adhering to strict data collection protocols[124], the UKBB gathered and curated medical data from 500000 participants across the UK over a period of five years from 2006, covering different population categories, with the range of age defined from 40 to 69. This dataset enhances the representativeness of middle-aged and elderly populations by encompassing a wide range of causative factors, including health-related questionnaires, physical measurements, biological samples, and a substantial collection of standardised medical imaging data[125]. The platform includes imaging data of the basic vital organs and tissues, such as the brain, heart, bones, and abdomen, and supports multiple common clinical imaging modalities, including MRI, X-ray, and CT scans.

For CMR, Littlejohns *et al.*[126] provided a comprehensive review of the imaging enhancements within the UK Biobank (UKBB). The UKBB contains an extensive repository of cardiovascular imaging data, including detailed cardiac imaging, offering a unique opportunity to investigate sub-clinical cardiovascular mechanisms linked to a broad category of cardiovascular and non-cardiovascular diseases. This includes research aimed at identifying pathology markers from an early stage, along with their genetic and lifestyle determinants. Cardiac MRI scans are performed using a Siemens 1.5 Tesla MAGNETOM Aera scanner (Siemens Healthineers, Erlangen, Germany), equipped with VD13A software and a spine and body flex matrix coil. The protocol, which lasts approximately 20 minutes, is conducted individually without the leveraging of pharmacological stressors or contrast agents. This thorough protocol enables assessments with detailed structure and function of the left and right ventricles, left and right atria, and the aorta. The measurements obtained include volumes and their variations across the cardiac cycle, cardiac wall thickness and mass, tissue motion via tagging, and thoracic aorta size and distensibility. However, further exploration is needed for the automatic extraction of cardiac features. Research teams based in London and Oxford established a ground-truth dataset[127, 128] utilised as a reference for structural segmentation, derived from the manual annotation of 5000 cardiac scans. An automated system for large-scale image quality control, analysis and image-based phenotype extraction has also been developed in collaboration with the University of Leeds, under UKBB Application 1135. This dataset includes approximately 5,000 CMR cases with manually labelled contours for the anatomical structures of the heart, including the left ventricle (LV) and right ventricle (RV). As a standard, a complete temporal sequence includes 50 time points from SAX stacks under the b-Steady-State Free Precession (SSFP). Each volume of data contains around 10 slices in depth dimension. The corresponding scan slices are acquired with a matrix size of 208 × 187, further, the scale of 1.8×1.8 mm^2 is obtained as a fixed resolution for the view of in-plane. The reference annotation labels for the LVendo, LVmyo and RV are also offered as the additional attachment data with ED and ES frames, most ED frames are located at the first position, and ES is normally located at different positions of a cycle, depending on the clinical scan.

2.6.2 Automated Cardiac Diagnosis Challenge dataset (ACDC)

The ACDC dataset, introduced by the Automated Cardiac Diagnosis Challenge at the Medical Image Computing and Computer Assisted Intervention (MICCAI) 2017, is another widely leveraged data resource for cardiac magnetic resonance (CMR) analysis. Originally designed to provide a robust benchmark for validating segmentation algorithms[129, 130], the dataset has since extended its impact to other areas of medical image analysis. It now plays a critical role in tasks such as image reconstruction and cardiac functional assessment, facilitating standardised evaluations and enabling high-quality comparisons of different approaches particularly.

The dataset comprises MR images from 150 patients attached with annotations, representing five distinct pathological conditions: normal, myocardial infarction, dilated cardiomyopathy, hypertrophic cardiomyopathy, and abnormal right ventricle. Each sample includes multiple image sequences captured at different cardiac phases, accompanied by ground truth labels for various anatomical structures, including the left ventricle, right ventricle, and myocardium.

The data collected in the ACDC achieves the acquisition via two types of MRI scanners utilising different magnetic strengths: 1.5 Tesla (Siemens Area, Siemens Medical Solutions, Germany) and 3.0 Tesla (Siemens Trio Tim, Siemens Medical Solutions, Germany). The protocol that ACDC follows is [131]. The dataset includes short-axis cine MRI scans acquired using both retrospective and prospective gating during breath-hold under the SSFP sequence. The scan direction to capture the anatomical structure is from base to apex, producing short-axis slices with the thicknesses of either 5 mm or 8 mm to cover across LV and RV regions. A complete or partial cycle covers the necessary cycle and concludes 28 - 40 frames with a 1.37 to 1.68 mm^2 per pixel for the spatial resolution.

The ACDC test set is applied with 100 SAX cases, each covering a complete cardiac cycle. These cases are divided into two categories: 20 samples are classified as normal, while the remaining 80 samples represent various pathological conditions, including previous myocardial infarction (MINF), dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), and abnormal right ventricle (ARV). Corresponding segmented labels are provided for reference at the end-diastolic (ED) and en-systolic (ES) phases.

2.7 Discussion

Previous sections have reviewed and summarised the publications which are the most relevant work in the field of MRI reconstruction, especially the approaches for cardiac CINE MRI. In this section, we introduce the main challenges for MRI and CMR specifically, which conducted the potential directions for future research in the field.

2.7.1 CMR-Specific Challenges for CMR Reconstruction

This section reviews the key challenges unique to deep learning (DL)-based reconstruction in CMR reconstruction, driven by the dynamic nature of the heart, especially when compared to static organs. Secondly, we review the mainstream solutions with corresponding representative works specifically developed to overcome these specific challenges.

Challenge 1: Increased Dimensionality

End-to-end training strategy based on backpropagation is widely utilised in deep data-driven or model-driven algorithms, especially for unrolling framework, more recently. Nevertheless, for obtaining a competitive performance, the large memory requirement for a huge amount of parameters of the network is the main obstacle for clinically relevant practice. Furthermore, the potential layers for normalisation and activation enhanced this challenge. In particular, for MRI reconstruction to address the cardiac CINE MR images, an additional dimension is required for temporal redundancy, which further increases the challenge.

In the case of clinical practice, the main focus of the clinicians is the robustness of the analysis efficiency utilising the model. The multiple qualifications, accordingly, and the complexity of the proposed frameworks are increasing along with the development of different methods. Nevertheless, a balance problem has emerged between the memory requirements and the training settings including batch size, and iterations, to improve the efficiency of the pipeline methods.

In order to strike an optimal balance between memory usage and performance, it is beneficial to explore innovatively structured network frameworks or training strategies for efficient memory utilisation, which can directly reduce computational complexity. Chen *et al.*[132] introduced a framework which enables fusion learning to overcome the GPU memory limitation through the improved network design for reconstruction in both low-frequency and high-frequency, by conducting the multi-layer perceptron and convolutional fitters with smaller sizes. The algorithm offers a computationally efficient solution in terms of GPU memory usage, which allows the model to be deployed on a broader range of hardware, including systems with limited computational resources. Ozturkler *et al.*[133] proposed a training mechanism which is established to address the constraint on computational challenges in terms of high-dimensional data - 3D MRI. By performing an efficient design for the architecture of end-to-end training through the reduction of module size, each block is optimised independently. In addition, to reduce the memory footprint further, a gradient updates strategy by decoupling sub-modules in a parallel style based on different GPUs is presented.

Challenge 2: Limited Training Data

To achieve the high-quality reconstruction of undersampled MRI, the mechanisms employed in a supervised solution require a large amount of fully sampled data as a reference. However, over-reliance on reference data for training the model as a gold standard is not always practical for the clinical situation. Especially considering cardiac MRI, the scan time constraints usually preclude the availability of fully sampled data. Reducing the supervised dependence from fully sampled data or eliminating such supervised models is a research direction of interest for the field.

More recently, the self-supervised mechanism is the significant and primary solution that solves the limitation of the scale of training resources. A detailed review of self-supervised approaches for MRI reconstruction is presented in Section 2.3.3. In addition, Deep Image Prior (DIP), as another potential approach which obtains increasing attention, offers a novel training strategy which is free of pre-training. The first work[134] introduced the concept, a mapping framework that established the connection between an unchanged latent matrix to the undersampled measurements. Based on this, some works[135–137] further developed the theory from static to dynamic reconstruction. In the work[135], a framework without pre-training is proposed which establishes a hybrid learning based on the integration between the temporal variations and spatial redundancies of the latent domain. In addition, to capture the dynamic measurements which reflect the embedding of a fixed manifold into the latent domain, a Time-Dependent DIP is conducted. Elmas *et al.*[137] introduced a fusion learning of cross-site feature extraction of generative images prior to information adaptation to local imaging conditions. In particular, this framework enables collaborative training without the training based on data sharing, further, it can produce more high-quality reconstructions across different sites with consistent and accurate performance.

2.7.2 Robustness challenges for MRI reconstruction

In this section, we present robustness-relevant challenges of general MRI reconstruction.

Perceptual feature or realistic-looking texture

One significant challenge encountered by models based on the theoretical architecture of feature generation is the emergence of false yet realistic features, despite outstanding performance on quantitative metrics such as SSIM and PSNR. This issue is often temporarily resolved through the applications of manual annotation. A primary reason that leads to this phenomenon[138] is inaccurate prior assumptions, particularly the insufficient robustness of models to withstand distribution shifts when the training data differs from the test data, such as different sample resources, and data domains. The term ”hallucinations”, introduced in the work of FastMRI challenge[139] as a conclusion role for the limitation, highlights this issue. It was observed that many methods, despite their excellent performance on quantitative indices, generate abnormal anatomical structures during visual inspection. These artefacts typically arise during the encoding process for reconstruction, leading to unrealistic global features.

In the context of MRI reconstruction as an inverse problem, there is increasing researches[140–145] analysis how DL-based approaches suffer from hallucinations and offer their solutions to address this issue. Among these studies, Mao *et al.*[142] introduced an efficient regularisation term to address the issue of mode collapse in conditional generative models. This framework explicitly maximizes the rates that reflect the difference between generated images with the latent encoding procedure outputs, thus encouraging the generators to exhibit less focus on common modes. Genzel *et al.*[144] conducted an extensive study of the robustness of DL-based approaches for solving underdetermined inverse problems, particularly in real-world MRI scenarios, by focusing on adversarial perturbations of the measurements that maximize reconstruction errors. More currently, pipeline works for reducing hallucinative features in image quality are primarily categorized into two types. The first involves evaluating reconstructed images based on manual annotations from clinicians such as MR imaging experts. The second solution is introducing enhanced metrics, such as Learned Perceptual Image Patch Similarity (LPIPS), which is a more effective evaluation tool in terms of human perception compared to SSIM.

Adversarial balance and robustness

Image analysis tasks, especially classification and generation-based processing, are particularly vulnerable to perturbations from small-scale features and adversarial learning. The issue has been extensively highlighted in recent studies[146–148], to explore the relevant impact on the stability of the reconstruction. In the work presented by Huang *et al.*[146], an algorithm is established for addressing the challenges of boundary distortion, edge blurring, and intensity biases suffered from the missing data. The research also simultaneously exhibited increasing concern about the robustness of neural networks in clinical settings, where the model is particularly susceptible to adversarial examples. They mainly focus on investigating the impact factors that contribute to perturbations or noise, which can cause neural networks to overlook existing lesions.

Based on the present research, there is no definitive literature indicating that the reconstruction methods based on deep neural networks, most are performed in black box learning mode, are more susceptible to the superposition of various disturbance factors than reconstruction methods conducted in a conventional way based on sparse sampling. The stability and robustness of adversarial learning for weight coefficient still require deeper and more extensive investigation. Morshuis *et al.*[149] presented the adversarial attacks on reconstruction based on k-space and evaluated the performance across multiple CNN-based pipelines. Additionally, their findings suggest that adversarial attacks offer a valuable tool for understanding potential failure modes, also their focus was on specific regions of interest (ROI) rather than diagnostically relevant areas. A similar theoretical description can be found in the research presented in [150].

To enhance the robustness against adversarial attacks, the work introduced by Pramanik *et al.*[44] demonstrates that imposing a monotone constraint is both necessary and sufficient to ensure uniqueness, along with convergence and stability properties, for the fixed point in arbitrary inverse problems. However, the performance is often limited by the global convexity or monotone constraint[151], leading to instability. In the work of John *et al.*[152], it is noted that recent works based on monotone operator learning typically conduct the generation procedure based on Lipschitz constraint, which can reduce performance on quantitative view compared to the approaches that leverage unrolling mechanism. Furthermore, they proposed applying the monotone constraint to the sum of the gradient of the data term and the CNN, rather than constraining the CNN itself with performing further operations on monotone. Some work[153, 154] has suggested embedding adversarial attacks into the encoding process of the learning procedure to enhance the fitting abilities of the model.

2.8 Conclusions

In this chapter, we introduced a comprehensive review of the evolution of MRI reconstruction with a focus on relevant CMR reconstruction. Additionally, based on the previous review of theories and networks, we summarised the application-specific methods in terms of two aspects: quantitative MRI and dynamic MRI, which is the core motivation of this thesis, in particular. Then, we present a thorough summary of two widely adopted and publicly available datasets of cardiac MRI for high-quality evaluations. Last but not least, we further discussed the existing challenges from the views of general MRI and CMR-specific reconstruction.

The improvement of MRI reconstruction approaches benefits significantly from the development of deep learning and the compressed sensing theory which takes sparsely sampling as backbone. The initial attempt that utilise DL for MRI reconstruction began with CNNs, among various architectures, based on the superior ability to reconstruct feature detail from coarse to fine, U-Net gained the most adoption as the basic architecture to handle the specific requirements of MRI data. The primary stage focuses on denoising and deblurring. Conventional methods demonstrated the potential of DL to accelerate MRI scans. As deep learning techniques advanced, specialised models tailored specifically for MRI reconstruction emerged. Generative Adversarial Networks

(GANs) and Variational Autoencoders (VAEs) became prominent, addressing the limitations of earlier models in handling the unique artefacts and resolution requirements of MRI. These models offered improved reconstruction quality, capable of generating high-frequency images from under-sampled data. The current works in the third stage explore the integration of physical MRI acquisition with deep learning techniques, leading to hybrid models that combine data-driven and physics-based approaches. This integration enabled the development of models that could leverage both the learned representations from deep learning and the known physical constraints of MRI, indicating the measurements from clinical practice including multiple resolutions in spatial and temporal resolutions. While for this stage, more accurate and robust reconstructions were achieved. More recently, attention mechanisms and Transformer-based models gained increasing attention, providing enhanced capabilities in handling complex imaging tasks and improving the generalisation ability of the reconstruction process. The most recent developments focus on translating these advanced deep-learning models into clinical practice. This stage involves creating large, publicly available datasets to train and validate models, ensuring their reliability and robustness across diverse clinical settings. Efforts are also being made to standardize evaluation criteria and foster multi-disciplinary collaboration to refine these models further. Additionally, there is a growing emphasis on explainability and model interpretability to gain clinician trust and facilitate integration into routine diagnostic workflows.

In the near future, related works will keep focus on aspects such as generalisation ability, perceptual realism, and physics-based compensation for dynamic resolution improving CMR reconstruction. Image reconstruction is a fundamental task for collaborating with other medical image processing tasks and enhancing multi-dimensional feature learning for potential downstream analysis.

CHAPTER 3

Compressed Sensing using a Deep Adaptive Perceptual Generative Adversarial Network for MRI Reconstruction from Undersampled K-space Data

Objective: Magnetic resonance imaging (MRI) reconstruction from undersampled k-space data has received great interest due to its capability to reduce physical scan time. Meanwhile, the reconstruction problem is challenging because of its ill-posed and inverse nature. Nevertheless, existing compressed sensing (CS) and deep learning-based methods still need improvement as they suffer from limited generalisation ability, especially when higher undersampling factors are applied.
Methods: To recover high-quality images with reliable fine anatomical structures, we propose DAPGAN - a Deep Adaptive Perceptual Generative Adversarial Network that reconstructs high-quality MR images from undersampled k-space data. In particular, a novel Perceptual Feature Guidance (PFG) mechanism is proposed which has the capability to retrieve effective features from each level that are useful in emphasising underlying anatomical structures. In addition, the model explores information in a dual domain style.
Results: Experimental results show that the proposed method outperforms state-of-the-art baselines in terms of quantitative and qualitative evaluations. Our method improves the average SSIM (structure similarity index measurement) from 0.81 to 0.93 at a low CS ratio of 10%, compared to the average performance of competing methods on cardiac datasets using Cartesian sampling.
Conclusion: An innovative mechanism was proposed for accurate and perceptual feature guidance. It's an adaptive error-correction-based mechanism during multi-level feature reconstruction. The effectiveness was proved by the superior performance in extracting reliable anatomical details.
Significance: The architecture of our proposed model offers a new solution for accurate feature guidance, considering enhancing conventional optimization-based problems. In particular, it's a robust mechanism in aggressive undersampling scenarios.

3.1 Introduction

Magnetic resonance imaging (MRI) is a widely used noninvasive medical imaging modality for clinical diagnosis and treatment planning. MRI has excellent strengths to provide outstanding contrast, especially in the field of assessing brain diseases, cardiovascular diseases, and corresponding treatment monitoring. Nevertheless, the long acquisition time caused by the hardware or patients nature always lead to discomfort to patients, which may hinders the related time-critical issues. Additionally, due to the limitations of MRI scanning, the image quality may be influenced heavily by aliasing like motion artefacts. As a result, an efficient MR image reconstruction approach that recovers high-quality images with faithful anatomical structures from undersampled data is a challenging problem.

To accelerate MR imaging, the efforts can be categorised into two directions: hardware-based parallel imaging techniques[155] and compressed sensing-based MR image reconstruction algorithms using accelerated k-space signal. As a mainstream direction, most researches focus on reconstructing MR images with undersampled data in k-space. The violation of the Nyquist sampling theorem results in aliasing artefacts when undersampled images are acquired. To solve this, the MR image reconstruction process can be formulated as an optimization problem composed of two elements: data fidelity which ensures the consistency between reconstructed and original items in k-space, and prior regularization, which usually includes sparsity, smoothness or spatio-temporal redundancy. This technique is well-known as compressed sensing MRI [17, 18](CS-MRI). Conventional CS methods are essentially regularised convex optimization algorithms, most commonly based on Total Variation (TV) regularisation [156–158] and low-rank constraints [95, 158]. Great progress has been made since CS-MRI-based conventional methods were proposed, nevertheless, these methods are iterative and they are heavily limited by the reconstruction speed which is time-consuming.

Recently, deep learning (DL) approaches have gained popularity for CS-MRI, as they are more efficient (i.e. faster execution/reconstruction times) and yield higher-quality for image reconstructions than their conventional CS-MRI counterparts [47]. There have been quite a few newly proposed methods, which can be roughly categorized into two types, (a) single domain-based methods and (b) dual-domain-based methods. The models performed in a single domain usually leverage the inverse Fourier transform (IFT) to obtain the original input from initial frequency domain data, then feed it to an image domain network or transfer the outputs from a network in k-space to obtain the final image domain results. The former one is the most adopted workflow for most current single-domain convolutional neural network(CNN)-based methods. For instance, Liu *et al.* [159] exploited the shareable features from potentially misaligned target images of an arbitrary contrast to reconstruct accelerated images based on an end-to-end CNN. Dar *et al.* [160] addressed the problem of data scarcity in network training for accelerated MRI by utilising transfer-learning based approach.The methods above addressed high computational cost relatively compared with conventional CS-based algorithms. Further, generative adversarial network (GAN)[15] has been proven to be effective in MRI reconstruction of image space [47, 48, 161–165]. De-aliasing generative adversarial network (DAGAN) [47] was the first study to explore using GANs for MRI reconstruction. To improve the recovery of high-frequency data, the adversarial loss was combined with a mean squared error (MSE) loss, mean absolute error (MAE) loss, perceptual loss and other non-adversarial losses. Zhou *et al.* [161] proposed an approach which introduced structurally strengthened connections for enhancing feature propagation, further more the in-between concatenated autoencoders were utilised except residual layers. Considering widening the channels of feature propagation in image space, Liu *et al.* [162] used a cross-stage skip connection between two end-to-end cascaded U-Net to form the generator. For learning the long-range dependency of feature maps, Xu *et al.* [164] proposed a dilated depthwise separable convolution dense block and a self-attention module which is squeeze-and-excitation lightweight. In order to consider the data consistency (DC) with k-space information for the correction of reconstructed results, several works [40, 68, 70, 166–170] proposed to incorporate the data fidelity in the network architecture by applying linear combination between the network prediction and the known undersampled k-space measurements for the known samples, and replacing the unknown samples by the prediction of the CNN. For example, Wang *et al.* [67] proposed a deep cascade-based convolutional neural network to simulate the iterative reconstruction of dictionary learning-based approaches by embedding the data consistency layer into the unrolling iteration. By adding a data-sharing layer, the temporal correlation for dynamic reconstruction can be learned as an enhanced regularisation for better performance. The dual-domain approaches [70, 171, 172] adopt learning feature information in both k-space and image domain. To simultaneously reconstruct information in k-space and image space, Zhou *et al.* [171] proposed a recurrent network with deep T1 prior embedded for accelerating the acquisition of MRI.

We compared the proposed method with the state-of-the-art in Table 5.1 from the following aspects: Firstly, most comparison methods consider the reconstruction of the k-space data to aid the spatial domain through data consistency. For example, DIMENSION designs k-space networks to learn the domain information in particular. Both of our methods and MD-Recon Net addressed the sequential learning of splitting the k-space and image domains, fully exploring the internal and intersection information

Table 3.1: The configuration of investigated methods in terms of the algorithms and experiments.

	Ours	Ran et al.[69]	Wang et al.[67]	Sun et al.[169]	Yang et al.[47]	Ravishankar et al.[27]
K-space loss	✓	✓	✓	✗	✓	✗
Image-space loss	✓	✓	✓	✓	✓	✓
Parallel domain learning	✓	✓	✗	✗	✗	✗
Active adaptation-based guidance	✓	✗	✗	✗	✗	✗
Multi-medical datasets	✓	✓	✗	✗	✗	✓
Irregular/aggressive undersampling patterns	✓	✗	✗	✗	✗	✗
Ablation study	✓	✗	✓	✓	✓	✗
3D downstream analysis (clinical relevance)	✓	✗	✗	✗	✗	✗

of the dual domains. Our method facilitates an adaptive error-correction-based feature guidance mechanism. From the aspect of the configuration for experiments, several methods utilise one type of dataset. Our method verified the superior performance on both brain and cardiac datasets. Further, we demonstrated the robustness of our model with simulated irregular undersampling patterns based on compressed sensing. Considering the clinical relevance of investigated methods, we conducted 3D downstream analysis with volumetric quantifications of key indices for both brain and cardiac MRI.

Previous works on dual-domain MRI reconstruction have proved the superiority compared with single-domain approaches. Nevertheless, the reconstruction quality is still limited, and the performance can be improved based on two observations:

1. From the view of network design, the conventional optimization strategies adopted the combination of different loss functions which only consider the fidelity between the final output and the ground truth. Therefore, the performance is easy to suffer from blurry results or natural-looking structures which are hallucinated in the reconstructed results, because the prediction is essentially an average of all likely solutions to satisfy the multiple criteria. There are still more valuable intermediate predictions regarding different levels of reconstruction to be utilized for much more accurate MR image reconstruction.

2. From the perspective of data to be reconstructed, the increased acceleration factors of CS, and changing the target organs for reconstruction are two main terms that affect the robustness of the network when facilitating reasonable structural anatomy recovery.

To essentially address the above limitations, we propose the deep adaptive perceptual generative adversarial network (DAPGAN), a novel DL-based method for CS-MRI reconstruction from under-sampled k-space data, it performs in an adaptive supervision style to provide aliasing-robust and supervision-efficient reconstruction. Our network simultaneously leverages feature learning in both frequency and spatial domains. The main contributions of our approach are listed below:

1. A perceptual feature guidance at multi-level is proposed in this work. It can be embedded in the original output of the previous multi-level CNN layers to build the adaptive mechanism which iterates between intermediate reconstructions. Such a mechanism can adaptively constrain the spatial domain information as an additional and efficient supervision which enables emphasizing underlying and missed anatomical structures of multi-levels based on the previous iterations.

2. De-aliasing robustness with multi-datasets and multi-undersampling conditions. We perform extensive experiments on brain and cardiac MRI datasets. Results show the DAPGAN outperforms several state-of-the-art methods, particularly under high undersampling acceleration factors and irregular undersampling masks.

3. Clinical relevance. We address the clinical relevance of this approach by demonstrating that subsequent image analysis steps –viz. segmentation– can be performed using the reconstructed images from undersampled data, without significant influence on the biomarkers/anatomical measurements derived thereof, relative to those extracted from the original images (i.e. reconstructed from fully sampled data).

3.2 Problem Formulation for CS–MRI

The observation and data acquisition model for image reconstruction can be regarded as a discrete linear model:

$$\boldsymbol{F}\boldsymbol{X} + \varepsilon = \boldsymbol{Y}, \tag{3.1}$$

where $\boldsymbol{X} \in \mathbb{C}^{\mathrm{N}}$ represents the complex-valued MR image which is desired, $\boldsymbol{Y} \in \mathbb{C}^{M}$ denotes an observation matrix that represents undersampled k-space measurements, ε is acquisition noise and $\boldsymbol{F}$ represents an undersampled Fourier coding matrix. From the CS perspective, the CS-MRI reconstruction task can be expressed as an optimization problem:

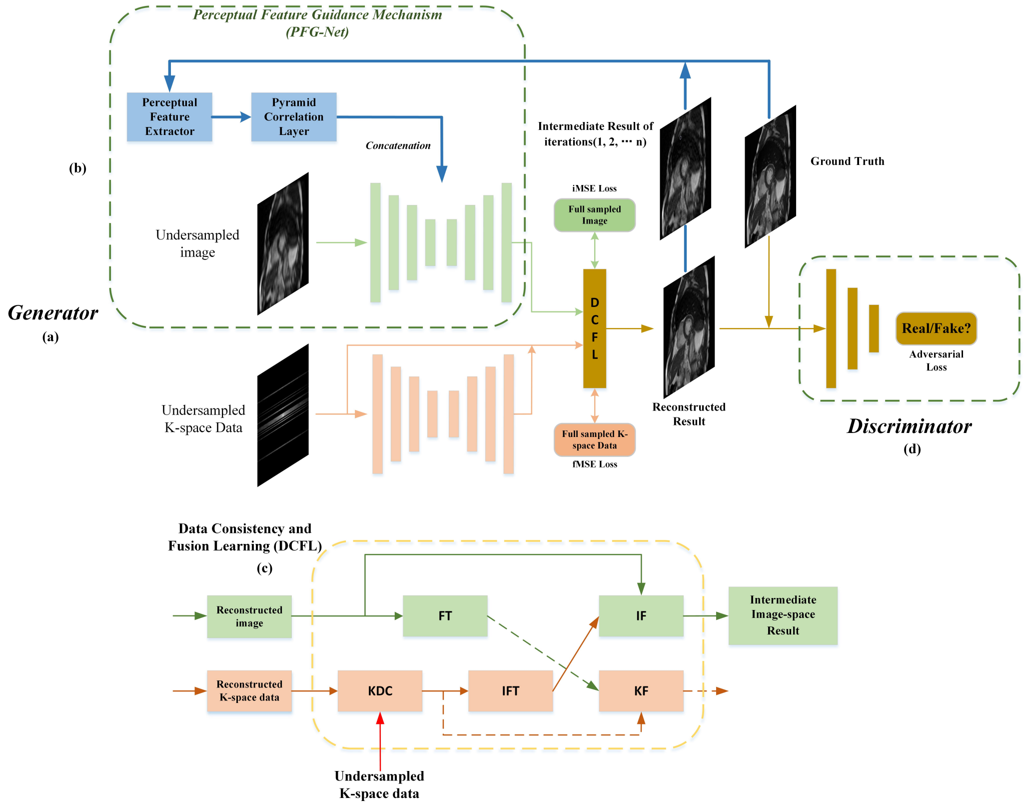


Figure 3.1: Network architecture of DAPGAN for high-quality MRI reconstruction. (a) The generative network takes the undersampled k-space data and the zero-filling images (with inverse Fourier transform of undersampled k-space data) as inputs in a parallel learning fashion, and outputs reconstructed images. (b) For the image domain, the proposed PFG-Net is a U-Net-based network driven by the PFG mechanism; for the k-space domain, we utilise a basic U-Net to reconstruct the frequency information. (c) Data consistency module and fusion learning layers are utilised in the cross-domain for information sharing. (d) The discriminator distinguishes the reconstructed images from the fully sampled images. (KDC: K-space Data Consistency, FT: Fourier Transform, IFT: Inverse Fourier Transform, IF: Image-space Fusion, KF: K-space Fusion)

$$\min_{x} \zeta R(\boldsymbol{x}) + \lambda \left\| \boldsymbol{F}_u \boldsymbol{x} - \boldsymbol{y} \right\|_2^2 , \tag{3.2}$$

where $\boldsymbol{F}_u \in \mathbb{C}^{M \times N}$ denotes the undersampled Fourier encoding matrix, R denotes a regularisation term based on input vector $\boldsymbol{x}$, λ and ζ are regularisation parameters.

The deep learning-based CS-MRI reconstruction model can be formulated as follows:

$$\min_{x} \zeta \left\| \boldsymbol{x} - f_{cnn}\left(\boldsymbol{x}_u \mid \boldsymbol{\vartheta}\right) \right\|_2^2 + \lambda \left\| \boldsymbol{F}_u \boldsymbol{x} - \boldsymbol{y} \right\|_2^2, \tag{3.3}$$

where f_{cnn} represents a CNN parameterised by $\boldsymbol{\vartheta}$, and ζ is a regularisation parameter, $\boldsymbol{x}_u$ denotes the zero-filled reconstruction utilised as the initial input of the network. Here, the reconstruction process applied by CNN methods (i.e. $f_{cnn}\left(\boldsymbol{x}_u \mid \boldsymbol{\vartheta}\right)$) are used as alternatives for the regularisation term $R(x)$ by optimizing the parameters of the model.

Inspired by the conventional regularisation term in CNN-based methods (Equation. 3.3), the proposed perceptual feature guidance can be leveraged as an additional constrain to formulate the new optimization problem:

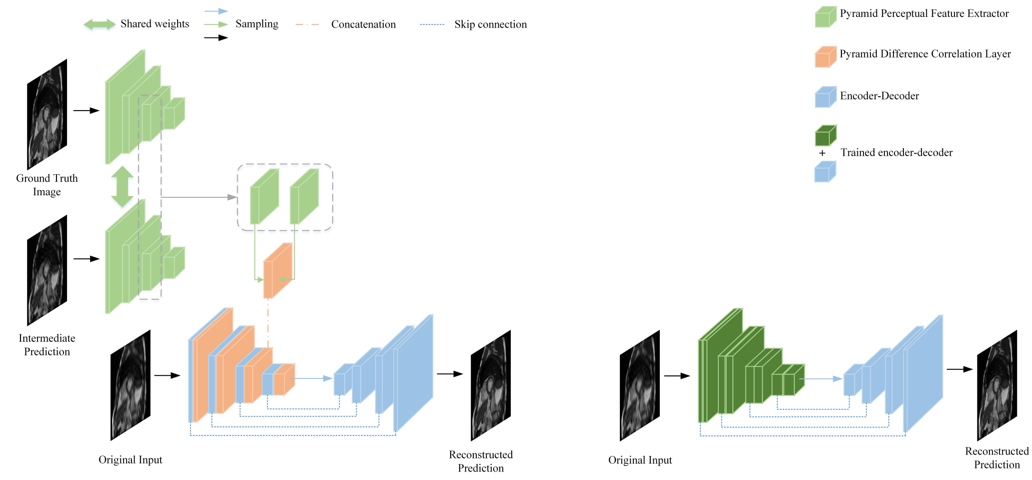
$$\min_{x} \zeta \left\| \boldsymbol{x} - f_{cnn}\left(c\left(f_{corr}\left(\left(\boldsymbol{x} - \boldsymbol{x_i}\right) \mid \hat{\boldsymbol{\vartheta}}\right), \boldsymbol{x_u}\right) \mid \boldsymbol{\vartheta}\right) \right\|_2^2 + \lambda \left\| \boldsymbol{F_u} \boldsymbol{x} - \boldsymbol{y} \right\|_2^2, \tag{3.4}$$

where $\boldsymbol{x}_i$ denotes the intermediate prediction of each iteration, f_{corr} represent the correlation layers parameterised by $\hat{\boldsymbol{\vartheta}}$, which can acquire and learning the difference feature in each training iteration when we force $\boldsymbol{x}$ to be well-approximated by the reconstructed results. Then the correlation features are concatenated with the original input as a constrain for next iteration of $f_{cnn}\left(\boldsymbol{x}_u \mid \boldsymbol{\vartheta}\right)$. How to provide accurate guidance for training process to find the optimal parameters is what we focused on in this paper.

3.3 Methodology

In this section, we introduce the details of our proposed method – DAPGAN. Inspired by the concept of dual-domain learning (MRI Dual-domain Reconstruction Network: MD-Recon-Net) [69], we developed a novel deep learning-based framework DAPGAN as illustrated in Fig. 3.1. Most approaches so far have only focused on fidelity between the output and the referenced ground truth based on minimising a cost function. This mechanism conventionally derived a tendency for model training to fall into the local minima. In contrast, our approach handles image reconstruction by proposing a self-adaptive guidance mechanism to address the problem relatively. The generator comprises two parallel convolutional neural networks: a multi-level perceptual feature guidance network (PFG-Net) based on U-Net architecture for real-space (image-domain) data and a standard U-Net for processing k-space data. The structure of PFG-Net is shown in Fig. 3.2, and PFG-Net is driven by the mechanism of perceptual feature guidance (PFG), which is presented in the following subsections.

3.3.1 Perceptual Feature Guidance mechanism of PFG-Net



(A) Pyramid Perceptual Feature Guidance for training phase

(B) Trained model for testing and validation phase

Figure 3.2: The overall schematic architecture of proposed pyramid Perceptual Feature Guidance (PFG) mechanism for deep conditioning reconstruction and PFG-Net. (A) The pipeline includes PFE, Pyramid multi-level correlation layers and concatenation-generation layers for the encoder at different scales. (B) The testing and validation phase of the generator.

We construct the pyramid Perceptual Feature Guidance (PFG) as an active-tuning-based deep conditioning mechanism, which consists of three steps (as Fig. 3.2.A shows): PFE (a VGG-based extension) for multi-scale feature extraction with shared weights, in which two parallel branches take reference signals (fully sampled or multi-accelerated data) and intermediate prediction as input respectively, instead of using the VGG-Net as the loss network only [173], we employ the proposed PFE as a pre-trained and pure multi-level feature extractor as the initial step of PFG; a pyramidal correlation layer between the corresponding two branches; a pyramidal generation with correlation features from pairs of frames. Thus, the adaptive tuning function of our generator is facilitated by learning the dynamic correlation guidance alongside training iterations. For model testing and validation, we employ the architecture with trained parameters in the training phase and leave out PFG (see Fig. 3.2.B).

Perceptual Feature Extractor

As the first stage of the PFG mechanism, PFE includes the following process: the reconstruction results of the last iteration are processed as the intermediate prediction IP and then fed with the ground truth GT jointly to the PFE to obtain the paired output $\boldsymbol{PFE_k^{GT}}$and $\boldsymbol{PFE_k^{IP}}$ as the inputs of the next stage (see Fig. 3.2.A). PFE is an extension of the VGG-Net [174] (Fig. 3.3). Previous approaches [175] have shown that preserving the negative features with reference to the inputs is beneficial for MRI reconstruction. Motivated by this, we employed recursive residual learning to enhance the capacity of the VGG in the multi-level feature extraction task. Each layer of PFE contains three pre-activation residual blocks (PRB), and each PRB comprises batch normalisation (BN), a pre-activation unit (ReLU) and convolutional layers (in sequence) as sub-blocks with skip connections.

Considering $\boldsymbol{R_{i-1}}$ and $\boldsymbol{R_i}$ as the input and output of the ith PRB module, respectively, the PRB can be expressed as

$$\boldsymbol{R_i} = f_{PRB}\left(\boldsymbol{R_{i-1}}\right) + \boldsymbol{R_{i-1}} = \boldsymbol{W_{RB}}\left(\boldsymbol{BN}\left(r_a\left(\boldsymbol{R_{i-1}}\right)\right)\right) + \boldsymbol{R_{i-1}}. \tag{3.5}$$

where $f_{PRB}(\bullet)$ is a function of the residual network, $\boldsymbol{W_{RB}}$ represents the weights of the convolutional network, r_a and $\boldsymbol{BN}$ denote ReLU and batch normalisation which included in the pre-activation function.

We extract the convolution and pooling layers from the standard VGG-Net and concatenate them with the PRB modules. We can represent the expression as follows:

$$SP_k = \boldsymbol{W_k} \otimes \left[f_c\left(\boldsymbol{R_k^i} + \boldsymbol{X_k}\right)\right] + b_k, \tag{3.6}$$

where SP_k represents a sub-unit of PFE, k denotes the sequential level of feature extraction that corresponds to multi-scale features in the encoder, $\boldsymbol{W_k}$ represents the convolution network weight of each block in standard VGG, b_k is a bias, f_c denotes the concatenation and fusion of the PRB $\boldsymbol{R_k^i}$ and initial input $\boldsymbol{X_k}$, $\otimes$ denotes the convolution operation.

The input for the PFE is the pairs of ground truth $\boldsymbol{X_{GT}}$ and the intermediate predicted $\boldsymbol{X_{IP}}$ over the iterations during training; we instantiate Eqn. (3.6) as

$$
\begin{aligned}
PFE_k &= SP_k\left(\boldsymbol{X_{GT}}, \boldsymbol{X_{IP}}\right) \\
&= \boldsymbol{W_k} \otimes \left[f_c\left(\boldsymbol{R_k^i} + \boldsymbol{X_k}\right)\right] + b_k \mid \boldsymbol{X_{GT}}, \boldsymbol{X_{IP}}.
\end{aligned}
\tag{3.7}
$$

Pyramidal Correlation and Generation(PCG)

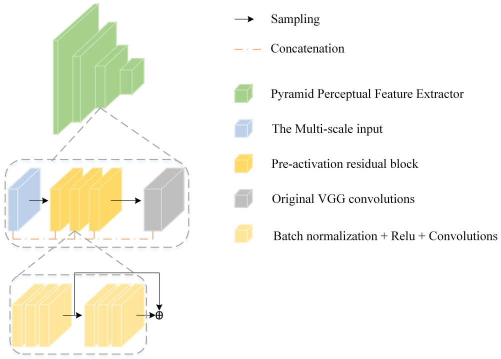


Figure 3.3: Overview of the proposed pyramid Perceptual Feature Extractor, for exploring multi-level features without information redundancy, a pre-activation module with residual option is employed, inheriting the VGG-16 architecture before fully connected layer.

For PCG, as Fig. 3.2.A shows, there are two stream branches utilised as inputs: $\boldsymbol{PFE_k^{GT}}$and $\boldsymbol{PFE_k^{IP}}$ represent the pyramidal outputs from the PFE, the pairwise features are inputs of pyramidal correlation layer, the original features denoted by $\boldsymbol{X_u}$ from the convolutional block as the backbone of the generator. Along the iteration axis, we match the correlation features to the primary encoder at four distinct scales to form the generation layers, which are 256×256, 128×128, 64×64, 32×32, respectively. The outputs from convolution and residual blocks are concatenated with the correlation feature maps in the form of $\left[f_{\text{Conv }+RB}\left(\boldsymbol{X_u}\right), f_{\text{Corr}}\ \right]$. The correlation layer comprises convolutional layers and activation options (ReLU). We formulate the PFG mechanism as follows:

$$
\begin{aligned}
F_{PFG} = {} & \boldsymbol{W_k^{DAP}} \otimes \left(f_c\left(\boldsymbol{W_k^{Corr}} \otimes \left(\boldsymbol{PFE_k^{GT}} - \boldsymbol{PFE_k^{IP}}\right), f_{\text{Conv }+RB}^{k}\left(\boldsymbol{X_u}\right)\right)\right) \\
& + b_k^{DAP},
\end{aligned}
\tag{3.8}
$$

where, $\boldsymbol{W}_{\boldsymbol{k}}^{\boldsymbol{DAP}}$ represents the weight and activation function of the spatial-domain reconstruction network PFG-Net at the convolutional scale k, $\boldsymbol{W}_{\boldsymbol{k}}^{\boldsymbol{Corr}}$ represents the weights of the convolutional block and activation function in correlation layers, from where we obtain the correlation features f_{Corr} of level k in the mechanism, f_c represents the concatenation function, b_k^{DAP} denotes the bias of layer k and $\otimes$ represents a convolution operation.

3.3.2 Data consistency and dual-domain fusion

Motivated by the concepts of CDNN [40] and MD-Recon Net [69], we use the data consistency of the k-space to incorporate data fidelity into the network architecture and employ a fusion block to enhance dual-domain learning.

A closed-form solution of Equation. 3.3 is applied for k-space data consistency (KDC module in Fig. 3.1)

$$DC(j) = \begin{cases} \hat{\boldsymbol{Y}} & \text{if} \quad \boldsymbol{j} \notin \Omega \\ \frac{\hat{\boldsymbol{Y}}+\lambda_0 \boldsymbol{U}(\boldsymbol{j})}{1+\lambda_0} & \text{if} \quad \boldsymbol{j} \in \Omega \end{cases}, \tag{3.9}$$

where, $\hat{\boldsymbol{Y}}$ denotes the predicted value learned from the k-space network. $\boldsymbol{j}$ denotes the index of k-space data, $\boldsymbol{U}(\boldsymbol{j})$ represents the undersampling value in the frequency domain. The DC block is used to perform linear fitting between the network prediction and original measurements. The hyper-parameter λ_0 is related to the imaging noise level, which is determined during training and it controls the constraints of the original k-space measurements, Ω represents the index set of the obtained k-space samples. If the k-space coefficient is initially not sampled ($\boldsymbol{j} \notin \Omega$), implying that we use the network prediction results as the final results. For the sampled k-space entries, we calculate it linearly and weighted by λ_0 thus we combine the original sampled data with the unconditionally embedded network prediction. The forward and backward passes of the k-space data consistency can be derived easily.

As the merit of dual-domain reconstruction is taking advantage of the reconstructed details that only k-space domain network can extract, to further improve the ability of PFG-Net in image-domain, the iterative outputs of k-space network and KDC module are interactive via the image domain fusion (IF) layers, thus, it is utilised as one of the input for reconstructed intermediate results (see Fig. 3.1). We employ fusion blocks as the last step of each DC block. For k-space fusion (KF), the prediction from the branch network of the k-space is fed into the k-space data consistency (KDC) module with the original undersampled k-space data. Then, KF takes the output of KDC and the output of FT which is the transformed prediction of the image-domain network, as inputs.

3.3.3 Optimisation

The proposed approach is trained as a conditional GAN, where the perceptual feature guidance maps are used as the conditioning variables to drive image reconstruction. The network is trained end-to-end by minimising the adversarial loss and normalising MSE between the reconstructed data and fully sampled data. In particular, the MSE loss is defined in both the k-space and image domains to enforce data consistency across both domains.

Multi-scale perceptual Loss

Let Θ be the set of all the learnable parameters in the model, which includes the perceptual feature pyramid extractor and the encoder at different pyramid levels. Let n denote the iteration number, and $\boldsymbol{W_{n-1}^{l}}$ denote the intermediate prediction at the lth pyramid level reconstructed by our model, and $\boldsymbol{W_{GT}^{l}}$ the corresponding reference signal. The basic training process can be formulated as:

$$\mathcal{L}(\Theta_n) = \sum_{l=l_0}^{L} \alpha_l \sum_{\boldsymbol{X}} \left| \boldsymbol{W}_{n-1}^{l}(\boldsymbol{X}) - \boldsymbol{W}_{\boldsymbol{GT}}^{l}(\boldsymbol{X}) \right|_2 + \gamma|\Theta|_2, \tag{3.10}$$

where we compute the $\boldsymbol{L_2}$ norm $|\cdot|_2$ between the intermediate results and the ground truth in a multi-scale style. The second term denotes the regularisation of the model. We designed robust training loss for fine-tuning:

$$\mathcal{L}(\Theta_n) = \sum_{l=l_0}^{L} \alpha_l \sum_{\boldsymbol{X}} \left(\left| \boldsymbol{W}_{n-1}^{l}(\boldsymbol{X}) - \boldsymbol{W}_{\mathrm{GT}}^{l}(\boldsymbol{X}) \right| + \epsilon \right)^q + \gamma|\Theta|_2, \tag{3.11}$$

where $|\cdot|$ denotes the MAE norm, and we set further $q < 1$ to give less penalty to the short-range difference of perceptual fields, further, ϵ denote a constant.

Conditional Adversarial Loss

We formulate the conditional GAN as follows:

$$\begin{aligned}\min_{\theta_G}\max_{\theta_D}\mathcal{L}_{cGAN}\left(\theta_D,\theta_G\right) &= \mathrm{E}_{X_g\sim p_{\text{train }(X_g)}}\left[\log D_{\theta_D}\left(\boldsymbol{X_g}\mid\boldsymbol{Y}\right)\right] \\ &+\mathrm{E}_{X_u\sim p_{G(X_u)}}\left[\log\left(1-\left(D_{\theta_D}\left(G_{\theta_G}\left(\boldsymbol{X_u}\mid\boldsymbol{Y}\right)\right)\right)\right)\right],\end{aligned} \tag{3.12}$$

where, $\boldsymbol{X_u}$ denotes the undersampled data and $\boldsymbol{X_g}$ represents the fully-sampled ground-truth data; Y denotes the guidance information that acts as conditioning variables, with G_{θ_G} and D_{θ_D} representing the generator and discriminator networks parameterised by θ_G and θ_D respectively; and $p_{\text{train }(X_g)}$ and $p_{G(X_u)}$ denote the data distribution induced by the discriminator and generator respectively.

Dual-domain MSE Loss

As the aforementioned PFG is additional supervision when considering the MSE loss as the basic content loss to improve the quality of our reconstruction results, here we define MSE loss across the dual domain. This helps overcome the non-smooth inspection, which lacks coherent image details, based solely on the optimization of normalised MSE which is defined by pixel-wise difference.

The MSE loss in the image-space domain is formulated as follows:

$$\min_{\theta_G}\mathcal{L}_{iMSE}\left(\theta_G\right)=\frac{1}{2}\left\|\boldsymbol{X_g}-\hat{\boldsymbol{X}}_{\boldsymbol{u}}\right\|_2^2, \tag{3.13}$$

where X_g represents the ground truth, and $\hat{\boldsymbol{X}}_{\boldsymbol{u}}$ represents the reconstructed image.

Alternatively, the MSE formulation for the frequency domain is expressed as:

$$\min_{\theta_G}\mathcal{L}_{fMSE}\left(\theta_G\right)=\sum_{i=1}^{m}\frac{1}{2}\left\|\boldsymbol{K_{fs}}-\boldsymbol{K_{DC}^{i}}\right\|_2^2, \tag{3.14}$$

where $\mathcal{L}_{fMSE}$ denotes the MSE loss in the frequency domain, $\boldsymbol{K_{fs}}$ represents the full-sampled k-space data and $\boldsymbol{K_{DC}^{i}}$ denotes the output of the i-th generation sub-net with the data-consistency branch.

Consider that K^i denotes the generation networks with convolution blocks in the k-space domain. The forward-pass process can be expressed as

$$\begin{cases} \boldsymbol{K^1} = \sigma\left(\boldsymbol{W^1} \otimes \boldsymbol{K_u} + \boldsymbol{b^1}\right) \\ \quad\quad\vdots \\ \boldsymbol{K^i} = \sigma\left(\boldsymbol{W^i} \otimes \boldsymbol{K^{i-1}} + \boldsymbol{b^i}\right) \\ \boldsymbol{K^i_{DC}} = KDC\left(\boldsymbol{K^i}\right) \end{cases} \tag{3.15}$$

where $\boldsymbol{K^i_{DC}}$ represents the results from the KDC (k-space-domain data consistency). $\boldsymbol{W^i}$ and $\boldsymbol{b^i}$ denote the weight and bias of the i-th filter, respectively, and $\otimes$ denotes a convolution operation.

We express the total loss function as

$$\mathcal{L}_{\text{TOTAL}} = a\mathcal{L}(\Theta_n) + b\mathcal{L}_{cGAN} + c\mathcal{L}_{iMSE} + d\mathcal{L}_{fMSE}. \tag{3.16}$$

where a, b, c and d are weighting factors that control the trade-off between the different terms in the overall loss function.

3.4 Experimental Setup

3.4.1 Datasets

We trained and evaluated the proposed DAPGAN using a publicly available brain MRI dataset, namely, Calgary Campinas 359[1] (CC-359), and cardiac MRI data from the UK Biobank[176] (UKBB). We applied different undersampling schemes/patterns to the datasets and reconstruct the undersampled data using the proposed method.

In the CC-359 brain dataset, images were acquired from a clinically used MR scanner (Discovery MR750, GE Healthcare, Waukesha, WI). 35 subjects were used, and the data were divided into training (70%), validation (15%) and test (15%) sets. The dataset contains sequential two-dimensional slices stacked as volumetric images, with a size of $170 \times 256 \times 256$.

In the UKBB dataset, cardiac MRI data were acquired using a clinical wide bore 1.5T MR system (MAGNETOM Aera, Syngo Platform VD13A, Siemens Healthcare, Erlangen, Germany) equipped with an 18-channel anterior body surface coil (45 mT/m and 200 T/m/s gradient system). Further, fully sampled 2D cine b-SSFP SAX (short axis) image stacks were acquired with the following acquisition protocol: in-plane spatial resolution of 1.8×1.8 mm, slice thickness of 8 mm, slice gap of 2 mm and image

---

[1] https://sites.google.com/view/calgary-campinas-dataset/home/mr-reconstruction-challenge

size of 198×208. We used 250 subjects to train the model. For each subject's cine-MR sequence, we selected two time points, namely, at end diastolic (ED) and end systolic (ES) frames respectively, resulting in 5000 slices (10 slices per time point); 1000 slices at ED and ES of another 50 subjects were used as validation and test data respectively.

3.4.2 Implementation Details

We propose to use a U-Net-based architecture to construct the generators for image-space and k-space. For image space, the PFG-Net consists of 5 convolutional layers as encoding layers with 32 filters in each layer, followed by leaky ReLU activation functions with negative slope 1×10^{-2}, and 7 decoding layers with 32,32,32,32,32, 2 and 2 filters followed by leaky ReLU layers, respectively. Skip connections are introduced to connect mirrored layers between the encoder and decoder of matched scales. All kernels of the encoder are set to 3×3, the stride is set to 2 except for the first layer (set to 1), and padding is set to 1. Pyramidal correlation features are concatenated between convolutional layers and residual blocks in each encoder block to form the final generation layer. For the decoder of PFG-Net, all the kernels size are set to 4×4, and we set stride to 2, while padding is set to 1. As for the generator of k-space, it consists of 5 convolutional layers for encoder and decoder respectively, followed by batch normalisation and leaky ReLU layers. On the other hand, the discriminator D is formed using a standard CNN with 9 convolutional layers, followed by batch normalization and leaky ReLU layers.

The proposed approach was trained using the Adam optimiser, with an empirical learning rate of 5×10^{-5}. The specific settings of the optimiser are $\alpha = 5 \times 10^{-5}$, $\beta_1 = 0.9$ and $\beta_2 = 0.999$. For the training procedure, we performed 500 iterations. In Equation. 3.2, Equation. 3.3 and Equation. 3.4, λ and ζ are determined as learning parameters during network training. The hyper-parameters a, b, c and d (weighting factors in the total loss function, refer to Equation. 3.16) were set to 0.5, 0.5, 0.05 and 0.05 respectively. It is worth noting that, from the perspective of practical experience, these hyperparameters were determined empirically with the average value between the brain and cardiac datasets during fine-tuned training. From the view of theoretical design of feature guidance, the proposed method is robust with minimal parameter tuning when different anatomical structures are applied.

Our training strategy is to pre-train the PFE using brain and cardiac MRI datasets from different groups in UKBB as the first step, then utilise the pre-trained PFE as an additional feature extractor during PFG to leverage correlation feature guidance in the training and fine-tuning processes of our model with CC-359 and UKBB. The PFG plays a semi-supervision role in the training process and we used a trained model to infer the independent testing and validation dataset.

3.4.3 Competing Methods

Competing methods that the proposed approach, DAPGAN, is compared against in this study can be divided into three categories: i) DLMRI [27] and DAGAN [47] representing conventional methods without k-space data consistency; ii) DIMENSION [67] and ADMM-CS Net [169] involving reconstruction that leverages data consistency in the frequency and spatial domains and iii) MD-Recon Net[69] that is most related to the proposed method, and was the first study to employ dual-domain learning. The proposed DAPGAN inherits the concept of dual-domain learning with data consistency, and implements a mechanism that leverages multi-level PFG in the spatial domain. This is demonstrated later in the ablation study, in which we evaluate the influence of different network components at retaining key structural and morphological characteristics of the anatomical regions of interest in the case of the cardiac MRI data.

3.4.4 Evaluation Design

We evaluated the performance of the proposed DAPGAN and investigated competing methods to reconstruct high-resolution imaging data, from various versions of undersampled imaging data, generated using different types of undersampling patterns/schemes. Two different imaging data sets (i.e. brain and cardiac MRI) were used across all experiments conducted in this study. The performance of the proposed approach was evaluated and compared against competing methods in terms of image reconstruction quality, and the quality of downstream analyses (e.g. image segmentation, biomarker quantification) afforded by images reconstructed from undersampled data.

To demonstrate the flexibility of the proposed approach, we simulate two types of k-space masks for generating undersampled data: variable-density Cartesian and random at sampling rates of 10%, 20%, equal to acceleration factors of ×10 and ×5. To evaluate the robustness of the proposed model under aggressive undersampling conditions, we design two challenging sampling patterns with sampling rates of 10% and 20 % based on the Cartesian sampling rules.

Two quantitative evaluation metrics are considered to assess the reconstructed results in terms of structure similarity index measurement (SSIM) and peak signal-to-noise ratio (PSNR); these metrics are used to evaluate the structural similarity and information-to-noise ratio, respectively. The fully sampled data was used for the ground-truth in training and the reference in evaluation. Higher SSIM and PSNR scores mean better reconstruction, which can be verified by the retention of perceptual features and artefact removal indicated by visual inspection simultaneously.

We address several aspects of clinical relevance by demonstrating that subsequent image analysis steps, such as segmentation, can be performed using images reconstructed by the proposed method, producing results comparable to those derived using the original/fully sampled data on both brain and cardiac MRI datasets. Then we compared volumetric indices/biomarkers extracted from segmented images, in addition to comparing the quality of segmentations achieved by 3 state-of-the-art methods.

3.5 Results

3.5.1 Brain MRI Reconstruction

Table 3.2: Quantitative comparison of CS-MRI reconstruction on brain data using Cartesian and random sampling at different rates in terms of SSIM and PSNR(dB). To further verify the robustness of investigated methods, Type I and Type II of Cartesian sampling are conducted as challenging masks, with multi-irregular phase-encoding gradients.

	Cartesian (Mean Std.)								Random(Mean Std.)			
	10%		Type I of Irregular 10%		Type II of Irregular 10%		20%		10%		20%	
Methods	SSIM	PSNR(dB)	SSIM	PSNR(dB)	SSIM	PSNR(dB)	SSIM	PSNR(dB)	SSIM	PSNR(dB)	SSIM	PSNR(dB)
DLMRI	0.737(0.087)	24.994(2.718)	0.701(0.084)	24.260(2.681)	0.673(0.082)	24.288(2.443)	0.763(0.079)	25.309(2.995)	0.663(0.071)	24.906(3.104)	0.757(0.049)	28.150(2.988)
DAGAN	0.774(0.076)	25.784(2.927)	0.765(0.060)	26.174(2.774)	0.719(0.063)	25.649(2.528)	0.791(0.067)	26.947(2.952)	0.737(0.049)	27.364(2.863)	0.785(0.042)	29.132(2.726)
ADMM-CS Net	0.856(0.034)	29.449(2.370)	0.805(0.047)	27.265(2.651)	0.818(0.040)	28.947(2.564)	0.895(0.026)	32.161(2.411)	0.824(0.030)	30.098(2.348)	0.853(0.027)	31.683(2.301)
DIMENSION	0.874(0.037)	29.573(2.870)	0.835(0.038)	28.594(2.466)	0.841(0.038)	29.598(2.579)	0.903(0.024)	32.654(2.344)	0.844 (0.027)	30.865(2.237)	0.864(0.025)	32.145(2.258)
MD-Recon	0.898(0.031)	31.179(2.554)	0.855(0.040)	29.259(2.758)	0.868(0.032)	30.839(2.438)	0.923(0.020)	33.865(2.236)	0.869(0.024)	31.859(2.188)	0.880(0.023)	32.838(2.228)
DAPGAN	**0.937(0.017)**	**34.808(2.188)**	**0.923(0.023)**	**32.861(2.126)**	**0.925(0.015)**	**34.092(2.033)**	**0.951(0.018)**	**35.261(2.192)**	**0.911(0.019)**	**33.368(2.375)**	**0.915(0.025)**	**34.222(2.180)**

We trained and evaluated DAPGAN across a series of experiments using brain MRI data, where images were undersampled to different degrees using both regular and irregular undersampling strategies, i.e. based on Cartesian grid and random undersampling patterns. DAPGAN was trained to reconstruct high-resolution images from their undersampled counterparts, and was compared against several state-of-the-art image reconstruction methods.

Regular undersampling reconstruction

As shown in Table 3.2, when using the Cartesian mask with a sampling rate of 20 %, DLMRI and DAGAN achieved SSIM values in the range of 0.750–0.800. ADMM-CS Net and DIMENSION achieved a better SSIM value of 0.900, while MD-Recon Net yielded an SSIM value of 0.923 and the highest PSNR value of 33.865 dB among all competing methods. Our DAPGAN achieved the best SSIM value of 0.951 and reached a PSNR value of 35.261 dB. At a sampling rate of 10% that results in a more aggressive reconstruction problem, the performance of all state-of-the-art methods and that of our proposed method decreased, as expected. Whilst DIMENSION, ADMM-CS Net and MD-Recon Net achieved SSIM values in the range of 0.850–0.900, our DAPGAN still achieved an SSIM value of >0.900, reaching 0.937. The decrease in the SSIM values of these three competing methods was ∼3%, whilst that of our method was <2%. The PSNR values of the state-of-the-art methods reduced by ∼3 dB, whereas that of our DAPGAN was <1 dB.

When a random mask was used, our DAPGAN achieved the highest SSIM values of 0.911 and 0.915 at sampling rates of 10% and 20%, respectively. Only our method could reach SSIM values of >0.900. Moreover, our method achieved the highest PSNR value at the aggressive sampling rate of 10% and maintained competitive performance at 20%.

The reconstruction results of five competing methods and our proposed method using different sampling strategies are shown in Fig. 3.4 and Fig. 3.5, respectively. Zoomed-in images indicated by red boxes and reconstruction error maps presented below the reconstructed images clearly highlight the improvements in image quality achieved by DAPGAN relative to all other competing methods.

When the Cartesian sampling rate is modified to 10%, our approach reconstructed the most accurate perceptual textures of underlying images with the smallest differences observed from the error maps (see Fig. 3.4). Images reconstructed using DLMRI and DAGAN contained significant artefacts, with poor resolution and blurring of structural and textural details throughout the brain. While ADMM-CS suppressed a large propor-

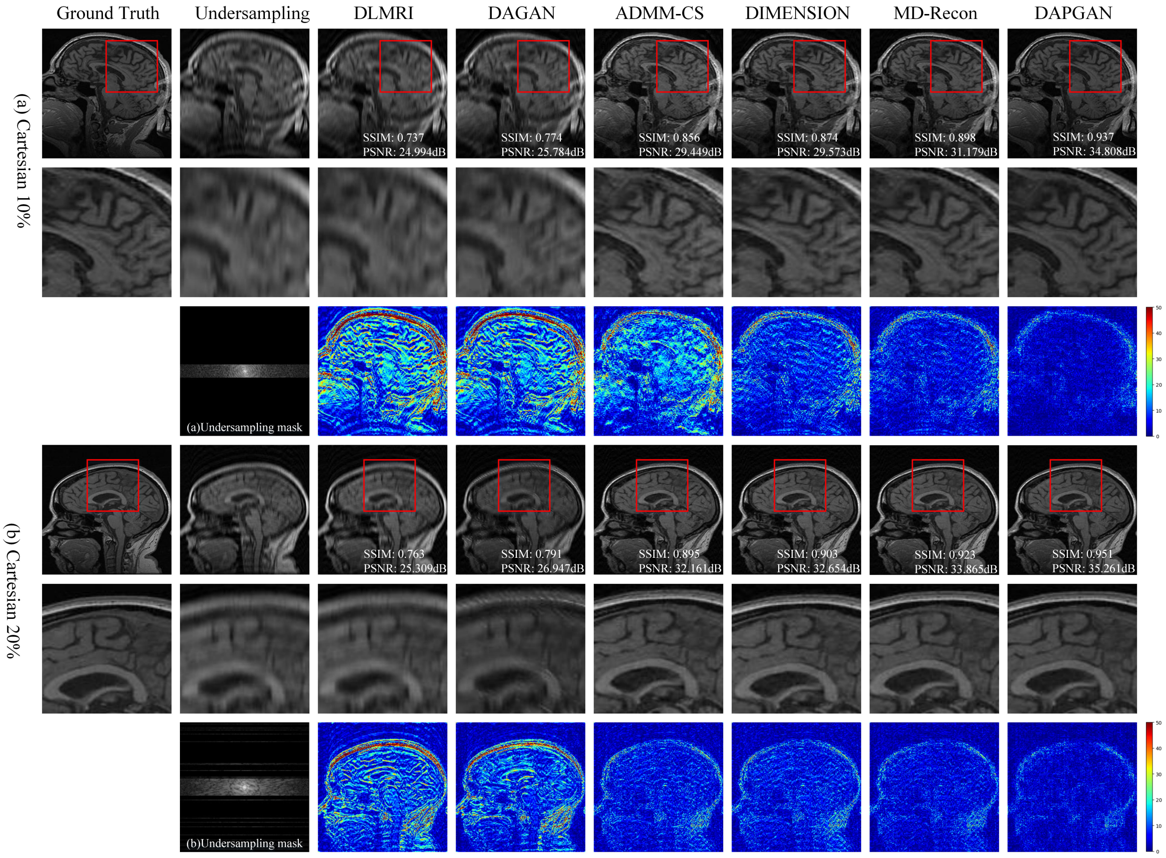


Figure 3.4: Visual inspection results of CS-MRI reconstruction on brain images using regular Cartesian sampling at rates 10% (from first row to third row) and 20% (from fourth row to last row). Here we show the comparison between the ground truth, undersampled results, and the investigated methods. The third column to last column are the results of different methods. The second and fifth rows represent the magnifying views of the red boxes in the original images. The third row and last row show the undersampling masks and the absolute difference maps (error maps compared to the full reconstruction) indicating the intensity of differences that range from 0 to 50 in the colorbar, which are visual inspected regions with multi-degrees of distinctions. The quantitative evaluation metrics (SSIM/PSNR) of the reconstructed images are indicated in the first and fourth rows.

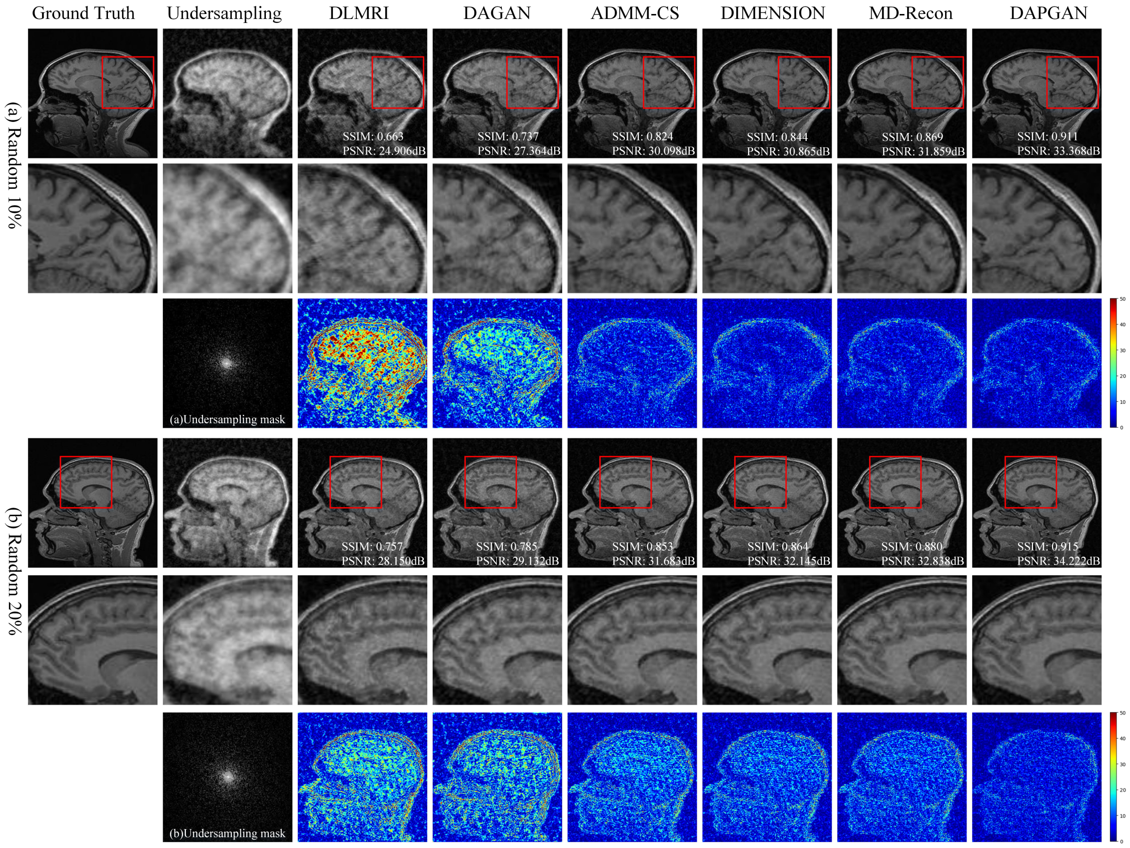


Figure 3.5: Visual inspection results of CS-MRI reconstruction on brain images using regular random sampling at rates 10% (from first row to third row)and 20% (from fourth row to last row), here we show the comparison between the ground truth, undersampled results and investigated methods. The second column (Undersampling) of the first and fourth rows represent the random sampling pattern in image space. The third column to the last column are the results of different methods. The second and fifth rows represent the magnifying views of the red boxes in the original images. The third row and last row show the undersampling masks and the absolute difference maps(error maps compared to the full reconstruction) indicating the intensity of differences that range from 0 to 50 in the colorbar, which are visual inspected regions with multi-degrees of distinctions. The quantitative evaluation metrics (SSIM/PSNR) of the reconstructed images are indicated in the first and fourth rows.

tion of such artefacts, recovering structural information more accurately than DLMRI and DAGAN throughout the brain, the error map indicates it still failed to preserve realistic and detailed textural information. DIMENSION and MD-Recon Net were the competitive approaches in terms of visual inspection, which reduced most of the artefacts, but still suffered defects of unrealistic textures and loss of structural details in some specific regions, compared with our method. When we set the sampling rate to 20%, our method again achieved the best reconstruction results based on visual inspection of the reconstructed images and corresponding error maps (shown in Fig. 3.4), with all other approaches achieving higher quality image reconstructions than when a sampling rate of 10% was used.

With regards to random undersampling at different sampling rates (see Fig. 3.5), again, the proposed method recovered high-frequency content better than the competing methods and captured detailed structural information more accurately than the rest (as evidenced by the error maps presented in Fig. 3.5).As seen before in the Cartesian undersampling scenario, images reconstructed using DLMRI and DAGAN contained severe structural and textural artefacts. The other approaches, namely ADMM-CS Net, DIMENSION and MD-Recon, were more successful in suppressing artefacts, but still exhibited higher errors than DAPGAN.

Irregular Cartesian undersampling reconstruction

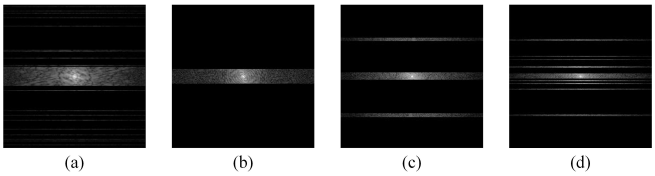


Figure 3.6: Examples of regular and irregular Cartesian undersampling masks. (a) Regular Cartesian mask for 20% sampling, (b) Regular Cartesian mask for 10% sampling, (c) Type I of irregular Cartesian mask for 10% sampling, (d) Type II of irregular Cartesian mask for 10% sampling

Among types of k-space imaging trajectories, Cartesian sampling is the most common type, where each line represents a phase-encoding gradient of frequency-encoding readout. All lines in the raster are provided in parallel as a sampling mask in k-space. Typical Cartesian sampling employs equally spaced lines in the raster, which can be reconstructed consistently after IFT. Usually, partial or parallel acquisition is adopted to save scan time. As illustrated in Fig. 3.6, to further verify the robustness of investigated methods, two types of Cartesian sampling are investigated (see Fig. 3.6 c, d), namely, Type I and Type II, to generate undersampling masks under a fixed sampling rate of 10%, with multi-irregular phase-encoding directions. The lines in the raster of Type I reduce the density of the central region compared with regular masks (instead of only acquiring data at the central region, we also acquire the k-space data towards the top and bottom directions), while the space between lines in the raster of Type II is provided with random values, and the lines are randomly located.

Quantitative image reconstruction results for the undersampling with the Type I and Type II irregular Cartesian undersampling masks are presented in Table 3.2. The SSIM yielded by DLMRI were both close to 0.700, DAGAN achieved SSIM values of 0.700–0.800 for both types, outperforming DLMRI marginally. DIMENSION and ADMM-CS Net meanwhile, improved image reconstruction quality for both Type I and II undersampled data, relative to DAGAN and DLMRI, achieving SSIM values in the range of 0.800–0.850. Similarly, MD-Recon Net afforded further improvement, achieving SSIM values in the range of 0.850–0.900. The proposed approach outperformed all competing methods for both types of irregular undersampled data, consistently achieving SSIM values > 0.900. The performance of all competing methods in terms of PSNR, for Type I undersampled data, was $< 30dB$. DAPGAN consistently outperformed all methods in this regard, achieving PSNR of $> 30dB$. For the Type II undersampled data, both MD-Recon Net and DAPGAN achieved PSNR values $> 30dB$, while the PSNR values for all other methods were $< 30dB$.

Fig. 3.7 provides a visual comparison of image reconstruction results achieved using several state-of-the-art approaches and the proposed approach, for two types (i.e. Type I and Type II) of Cartesian undersampling patterns. DLMRI and DAGAN failed to preserve detailed structural and textural information, and the reconstructed images contain significant visible artefacts. ADMM-CS and DIMENSION also performed poorly in this regard, although their reconstructions were of higher quality than DLMRI and DAGAN, as evidenced by the error maps depicted in Fig. 3.7. Images reconstructed by ADMM-CS and DIMENSION appear to be affected by streak-artefacts, as illustrated in the enlarged regions in Fig. 3.7 (refer to red boxes in Fig. 3.7). Visually, images reconstructed by MD-Recon Net appeared similar to those reconstructed using our approach. The corresponding error maps, however, suggest that our approach, DAPGAN, achieved lower magnitudes of reconstruction errors throughout the brain than MD-Recon Net, as illustrated in Fig. 3.7.

Clinical Brain Quantification

To demonstrate the clinical feasibility of the proposed image reconstruction approach, we also conducted a volumetric analysis in terms of the selected key brain indices based on segmentations derived from images reconstructed using undersampled data. Volumetric measurements of key anatomical structures are often used as biomarkers for disease diagnosis and monitoring. The rationale here is that for the proposed approach to be clinically useful, volumetric biomarkers derived from reconstructed images (from undersampled data) should show no/limited statistically significant differences to values derived from the original images. To this end, volumetric analysis was conducted using reconstructions of structural brain MR images, by quantifying volumes for the following structures using FreeSurfer [177] — lateral ventricles, white matter, thalamus, and putamen for both left and right brain hemispheres, and the brain stem and middle-anterior corpus callosum.

Table 3.3: Quantification of brain voxel volume (mm^3) of multiple volumetric indices extracted from segmented 3D volumes with 2D reconstructed results derived from Ground Truth, our method and competing methods. (LV/RV: left/right ventricle, LWM/RWM: left/right white matter, LT/RT: left/right thalamus, LP/RP: left/right putamen, BS: brain stem, CCM: corpus callosum of middle-anterior)

	GT	**DAPGAN**		MD-Recon		DIMENSION		ADMM		DAGAN	
Parameters	Mean(Std)	**Mean(Std)**	**p-value**	Mean(Std)	p-value	Mean(Std)	p-value	Mean(Std)	p-value	Mean(Std)	p-value
LV	3302.67(188.23)	**3519(224.55)**	**0.0002**	4798.17(749.63)	<0.0002	4712.17(1210.88)	<0.0002	5034.83(671.01)	<0.0002	5716.33(837.57)	<0.0002
LWM	8722.33(478.60)	**8406.17(509.41)**	**0.0179**	7176.33(1106.57)	<0.0002	5600.67(579.03)	<0.0002	6597.33(689.54)	<0.0002	4792.17(967.35)	<0.0002
LT	5873.5(487.71)	**5619(455.72)**	**0.0446**	5218.83(1433.46)	0.0234	5539.33(1379.54)	0.0237	4840.67(1375.83)	0.0003	4527.24(1484.65)	0.0256
LP	1669.67(67.89)	**1735.83(67.91)**	**0.0005**	2860.83(1012.60)	<0.0002	4433.83(645.89)	<0.0002	4193.67(826.52)	<0.0002	4819.43(891.38)	<0.0002
BS	14415.83(1232.73)	**13109.5(1469.73)**	**0.0005**	7909.33(948.93)	<0.0002	8441.67(364.38)	<0.0002	8447.67(1059.51)	<0.0002	7233.17(918.24)	<0.0002
RP	2747.83(117.60)	**2647.5(148.03)**	**0.0059**	2130.33(160.67)	<0.0002	1834.83(226.38)	<0.0002	1998.17(89.010)	<0.0002	1761.63(288.29)	<0.0002
RT	5050.83(459.48)	**4700(461.32)**	**0.0052**	4194(750.15)	<0.0002	3913.17(641.05)	<0.0002	4210.33(375.90)	<0.0002	3792.19(259.26)	<0.0002
RWM	10335.17(817.08)	**9856(802.90)**	**0.0281**	7349.83(704.03)	<0.0002	7941.17(1049.23)	<0.0002	6425(759.40)	<0.0002	6193.21(785.73)	<0.0002
RV	2731.33(422.33)	**2869.5(397.27)**	**0.1523**	3005.67(402.98)	0.0091	1751.83(351.69)	<0.0002	1830.33(490.11)	<0.0002	1585.45(493.17)	<0.0002
CCM	1562.67(127.64)	**1413.17(218.09)**	**0.0023**	934(53.53)	<0.0002	785.17(120.88)	<0.0002	959.67(279.73)	<0.0002	717.06(138.83)	<0.0002

Brain volumetric analysis following image reconstruction using DAPGAN and all other state-of-the-art methods of interest in this study is summarised in Fig. 3.8, alongside the ground truth structural brain volumes (i.e. volumes derived from the original images). For fair comparison, all evaluations were conducted identically, via automatic brain segmentation and volumetric quantification using FreeSurfer. Our method, DAPGAN, consistently outperformed all competing methods and showed no statistically significant differences to the ground truth volumes for most brain structures analysed (see Table 3.3).

3.5.2 Cardiac MRI Reconstruction

For the cardiac dataset, we evaluated the performance of our method on the reconstruction of undersampled data generated using Cartesian and Gaussian random undersampling patterns, using different sampling ratios (regular and irregular masks with 10 % and 20%). Furthermore, we conducted a set of ablation experiments to evaluate the key components of the proposed approach.

Regular Cartesian undersampling

Table 3.4: Quantitative comparison of cardiac image reconstruction performance, using Cartesian and random undersampling patterns at different rates in terms of SSIM and PSNR(dB). To further verify the robustness of investigated methods, Type II of Cartesian sampling is conducted as a challenging mask, with irregular phase-encoding gradient.

	Cartesian Mean(Std.)				Random Mean(Std.)			
	Type II of Irregular 10%		20%		10%		20%	
Methods	SSIM	PSNR(dB)	SSIM	PSNR(dB)	SSIM	PSNR(dB)	SSIM	PSNR(dB)
DLMRI	0.710(0.040)	26.400(1.369)	0.803(0.036)	25.758(2.338)	0.689(0.031)	29.146(1.604)	0.767(0.030)	32.843(1.437)
DAGAN	0.801(0.027)	29.375(1.332)	0.849(0.024)	30.152(1.359)	0.743(0.033)	29.892(1.604)	0.815(0.030)	33.762(1.502)
ADMM-CS Net	0.826(0.024)	29.913(1.343)	0.894(0.018)	31.859(1.481)	0.832(0.024)	31.823(1.687)	0.873(0.019)	34.960(1.558)
DIMENSION	0.855(0.021)	31.376(1.412)	0.907(0.016)	32.383(1.545)	0.869(0.018)	32.941(1.705)	0.919(0.016)	36.681(1.477)
MD-Recon	0.859(0.021)	31.193(1.441)	0.925(0.013)	33.291(1.576)	0.887(0.017)	33.639(1.669)	0.938(0.017)	37.347(1.446)
DAPGAN	**0.930(0.011)**	**34.512(1.472)**	**0.959(0.007)**	**35.657(1.579)**	**0.918(0.018)**	**35.081(1.546)**	**0.953(0.015)**	**37.877(1.413)**

Table 3.4 shows the results of an experiment with an undersampling rate of 20% in the regular Cartesian pattern. DLMRI and DAGAN achieve SSIM values of $0.800 - 0.850$, with the latter providing marginal improvements over the former. DIMENSION, ADMM-CS Net and MD-Recon Net all outperform DLMRI and DAGAN, achieving SSIM values of 0.894, 0.907 and 0.925, respectively. As with the brain MR data set, our approach significantly outperformed all competing methods, at reconstructing images undersampled using a regular Cartesian pattern at a rate of 20%, achieving an average

SSIM value of 0.959. Our method achieved the best performance with an SSIM value of 0.959. Moreover, our method reached the highest PSNR value of 35.657 dB, more than 2 dB higher than that achieved using MD-Recon Net. In the random sampling mask with a sampling rate of 10%, DLMRI yielded an SSIM lower than 0.700, and DAGAN achieved an SSIM value of only 0.743. Both ADMM-CS Net and DIMENSION achieved scores close to 0.850. MD-Recon Net achieved an SSIM value of almost 0.900, while our method achieved an SSIM value of >0.900. When the sampling rate was increased to 20%, the SSIM value increased by 4% and 5% in the cases of ADMM-CS Net and DIMENSION, respectively. Our method and MD-Recon Net, DIMENSION, achieved SSIM values of >0.900.

Figure 3.9 provides results of visual inspection between the proposed approach and other approaches on the cardiac dataset. When we set the 20% Cartesian strategy, as shown in the enlarged images and error maps, DLMRI and DAGAN show severe blurriness, and the remaining three approaches produced the images with artefacts corrupted near the myocardium (MYO) wall (as indicated more clearly in error maps compared with magnifying views). The DAPGAN consistently outperformed the other methods in terms of the removal of the artefacts and recovering high-resolution details.

For Gaussian random results with a sampling rate of 10% in Fig. 3.10, the five compared methods failed to remove the severe noise signals and did not preserve realistic textures in the ventricle regions. However, the proposed method yielded the most comparable results to the GT with fine details.

Irregular Cartesian undersampling

Table 3.4 evaluates the de-aliasing robustness of the model, regarding Type II of the irregular mask (see Fig. 3.6) performing qualitative and quantitative experiments under aggressive sampling conditions. We alter the regular mask with a challenging pattern designed based on the Cartesian sampling rule. The SSIM and PSNR values obtained using DLMRI were 0.700 and 26 dB, respectively. DAGAN achieved an SSIM value of 0.800. The performance of ADMM-CS Net was most affected under the aforementioned sampling conditions. Compared with the 20% sampling rate from a longitudinal perspective, the SSIM value of ADMM-CS Net decreased by 7% to 0.830 under this pattern, while the SSIM values obtained using DIMENSION and MD-Recon Net decreased by 5% (0.855) and ∼7% (0.859), respectively. Our method maintained an SSIM value of >0.900, and the decrease was only <3%.

Figure 3.11 shows that DAPGAN yields superior and consistent performance in preserving accurate perceptual details with better contrast and higher resolution, as depicted by the magnified cardiac regions, which reflected more underlying anatomy. We faithfully reconstructed the images with the smallest difference illustrated by the error maps, especially the improved performances in the region of the MYO wall. In contrast, the comparison methods failed to recover the high-frequency details, and produced relatively large differences, as shown in the error maps and enlarged images.

Clinical Cardiac Quantification

In this section, we demonstrate the impact of our method on various cardiac functional indexes based on the two most representative time points: ED and ES. Table 3.5 and Fig. 3.12 demonstrate the performance of the competing methods in terms of several key cardiac clinical analysis metrics, including the LV ED volume (LVEDV), ES volume (LVESV), LV stroke volume (LVSV), LV ejection fraction (LVEF), LV myocardial mass (LVM), RV ED volume (RVEDV), ES volume (RVESV), RV stroke volume (RVSV), and RV ejection fraction (RVEF). Our proposed DAPGAN outperformed all the competing methods in all clinical parameters and yielded the most comparable results to the reference ranges derived from the fully-sampled data with no statistically significant differences.

Table 3.5: 3D cardiac segmentation indices with 2D reconstruction results

	GT	**DAPGAN**		MD-Recon Net		DIMENSION		ADMM-CS Net		DAGAN	
Parameters	Mean(Std.)	**Mean(Std.)**	**p-value**	Mean(Std.)	p-value	Mean(Std.)	p-value	Mean(Std.)	p-value	Mean(Std.)	p-value
LVEDV (mL)	154.31(32.17)	**153.11(31.87)**	**0.676**	149.79(32.79)	0.121	152.18(31.87)	0.461	150.57(31.58)	0.192	142.17(31.37)	0.116
LVESV (mL)	69.93(22.18)	**67.38(22.28)**	**0.202**	62.75(22.81)	<0.005	65.66(22.18)	0.034	62.64(22.44)	<0.005	56.52(22.57)	<0.005
LVSV (mL)	84.38(16.74)	**85.74(16.01)**	**0.369**	87.05(17.88)	0.090	86.53(16.24)	0.161	87.92(16.71)	0.020	89.81(16.93)	0.036
LVEF (%)	55.09(6.46)	**56.51(6.34)**	**0.015**	58.64(7.43)	<0.005	57.39(6.41)	<0.005	58.97(6.84)	<0.005	60.89(7.83)	<0.005
LVM (g)	87.97(16.93)	**84.94(16.79)**	**0.049**	75.10(16.53)	<0.005	80.92(16.38)	<0.005	78.25(15.97)	<0.005	67.36(16.19)	<0.005
RVEDV (mL)	165.48(31.11)	**164.78(31.22)**	**0.813**	160.50(28.99)	0.067	162.02(30.41)	0.221	159.62(29.82)	0.032	156.24(30.65)	0.022
RVESV (mL)	76.82(19.63)	**76.34(19.15)**	**0.756**	73.28(19.88)	0.044	74.98(18.79)	0.286	73.47(18.95)	0.053	70.37(18.68)	0.043
RVSV (mL)	88.65(15.91)	**88.44(15.93)**	**0.878**	87.22(14.34)	0.265	87.04(15.46)	0.251	86.14(15.50)	0.076	84.58(15.69)	0.032
RVEF (%)	53.92(5.32)	**53.99(5.15)**	**0.799**	54.84(6.26)	0.067	54.06(5.19)	0.610	54.33(5.78)	0.335	55.62(5.76)	0.117

Ablation Study

To fully evaluate the benefit of our algorithm, we perform an ablation experiment. The results were analysed under the challenging 10 % Cartesian undersampling conditions.

Table 3.6: Quantitative comparison of CS-MRI reconstruction on cardiac data using irregular Cartesian sampling at rates 10% for ablation study in terms of SSIM and PSNR (dB). IS: image-space Net only, KS + IS: dual-domain learning with k-space, KS + IS + PL: adding perceptual loss into the loss functions, and KS + IS + PFG: a complete structure for DAPGAN with PFG.

	Type II of Irregular 10% Cartesian Mean(Std.)	
Methods	SSIM	PSNR
IS	0.816(0.030)	29.780(1.469)
KS + IS	0.846(0.027)	31.644(1.457)
KS + IS + PFL	0.875(0.023)	32.427(1.491)
KS + IS + PFG (DAPGAN)	**0.930(0.011)**	**34.512(1.472)**

We use three structural variants of our approach: image-space only (IS), dual-domain learning, which is image-space joint with k-space (KS + IS), adding perceptual feature loss into the loss functions (KS + IS + PFL) and a complete structure for DAPGAN with PFG (KS + IS + PFG). We first evaluate the importance of incorporating the k-space domain into the reconstruction process. In Table 3.6, compared with single-domain learning, better performance can be achieved using dual-domain learning, with SSIM and PSNR value improvements of 3% and 2 dB, respectively. In Fig 3.13, when we performed the dual-domain network, an improved performance with smaller errors in the region of MYO wall could be observed. Furthermore, we evaluated the variant with perceptual loss added to the linear loss function. The model achieved a SSIM value with an improvement of 3% compared with 'KS+IS'. However, from the view of error maps, 'KS + IS + PFL' appeared smaller differences slightly than the former two variants, which indicated the limitation of utilising the linear combination of multi-loss functions as the only supervision. DAPGAN considerably outperformed the other variants in terms of all metrics and visual inspection based on the challenge mask, significantly decreasing differences in the regions of MYO wall and ventricle. This consistently confirmed the merits of exploiting multi-level perceptual feature guidance (PFG) as an effective supervision mechanism, instead of optimising or simply combining loss functions linearly in a conventional way.

3.6 Discussion

Accurate assessment and high acceleration of the MR image reconstruction from undersampled k-space data is essential. The relationship between convolutional networks and conventional methods has already been justified by the theoretical analysis, which solidly supports the development of the CNN-based method. Herein, we focus on exploring an innovative approach for MRI reconstruction suitable for practical undersampling conditions, saving the physical sampling time, and enabling the reconstruction of undersampled data with high quality and comparable to reconstruction results of fully-sampled data.

Exploring the information in image-space and k-space in parallel improved the average performance relatively compared with single domain-based approaches. Nevertheless, we should pay more attention to the control of the optimisation solution of a CNN-based method during training, in other words, the 'black box', if we desire an effective mechanism for avoiding overfitting and achieving superior performance. In summary, recent methods are inadequate for two main terms: clinical practice and improvements for regularisation constraints. In addition, an efficient architecture enables the network to have a good potential for application extension.

Firstly, when the amount of training data based on fullysampling as references is not sufficient for the task, the supervised learning paradigms may not be able to train a sufficiently robust model. Moreover, for accelerated MR image acquisition, higher undersamping factors add severe aliasing to the initial zero-filled inputs, making the reconstruction more challenging. In addition, the generalization ability across multiple conditions is significantly reduced when training the model with certain datasets. For instance, the reconstruction tasks for medical imaging with multiple organs. Compensating the improvement of network generalization capability with existing transfer learning offers a better solution.

For another aspect, it is defined as the inverse problem which learns the end-to-end mapping from aliased undersampled data to reconstructed high-quality prediction. Nevertheless, as the regularization term utilised in the optimization problem, recent deep learning-based methods have adopted optimization strategies based on the fusion of multiple typical loss functions. Therefore, it will make them still suffer from blurry results, because the output is essentially an averaged optimal solution to satisfy the multiple criteria based on corresponding weights. Meanwhile, the intermediate information of multi-level features which perceive more detailed structural information is less considered for optimizing specially. The effectiveness is also confirmed by our simulation experiments, especially under aggressive sampling strategies. Inspired by the VGG network for deep feature extraction, we proposed multi-level feature extraction and difference correlation layer as an iterate-specific error-correction unit, the outputs are inserted into embedding and reconstruction nets. By retaining the useful information, this unit can adaptively correct the intermediate prediction of each iteration. Overall, for reflecting more accurate tissue structures, the feature guidance adaptively constrains the feature learning of multi-level in the solution space. For relative tasks like segmentation or synthesis, it can be a flexible mechanism that enables the network to have a good potential for application extension. Furthermore, our work proved another option for controlling the ’black box’ of deep learning-based optimization, which is the effective feature guidance that enables avoiding gradient vanishing or overfitting.

Considering that the performances of most deep learning-based CS approaches are limited by higher undersampling factors, in addition, it would also be influenced when exploring the application on multiple organs with different structural complexity. We evaluate our method on brain data and cardiac data. For imaging acceleration simulation, we considered CS-MRI reconstruction at sample rates lower than 20% which are relatively hard to deal with utilising optimization-based approaches. In view of simulating more undersampling patterns for improving the robustness of our model, we conducted irregular masks at a 10% rate for simulating aggressive CS-MRI situations with significantly increased aliasing. Our experimental results show that the proposed DAPGAN can achieve better promising results both in terms of visual images and quantitative indicators compared with conventional and competing approaches.

3.7 Limitations

Currently, we only focus on the reconstruction in static 2D cardiac cine MR imaging, while clinical indices such as myocardium strain are in demand to be computed to assess the dynamic deformation functionality of the heart, which indicates the potential correlation and redundancy exist along the spatio-temporal dimension. Additionally, as estimating the deformation field of multiple cardiac frames in a complete cardiac cycle is a crucial step for dynamic reconstruction with cine imaging, it is still a challenge to estimate accurate respiratory motion fields directly from undersampled images. It would be interesting to investigate how motion compensation improves dynamic CMR reconstruction and research its generalizability for different downstream tasks.

3.8 Conclusion

We proposed a novel deep learning-based approach, namely DAPGAN, for the reconstruction of undersampled k-space data with high quality. In particular, an active adaptation-based guidance mechanism for multi-level feature reconstruction is proposed. Also, the dual-domain learning with k-space further improved the generalisation ability. DAPGAN is robust to de-aliasing and other artefacts removal in accelerated k-space datasets. It produces perceptual features-focused results, which is confirmed by superior performance on evaluated metrics.

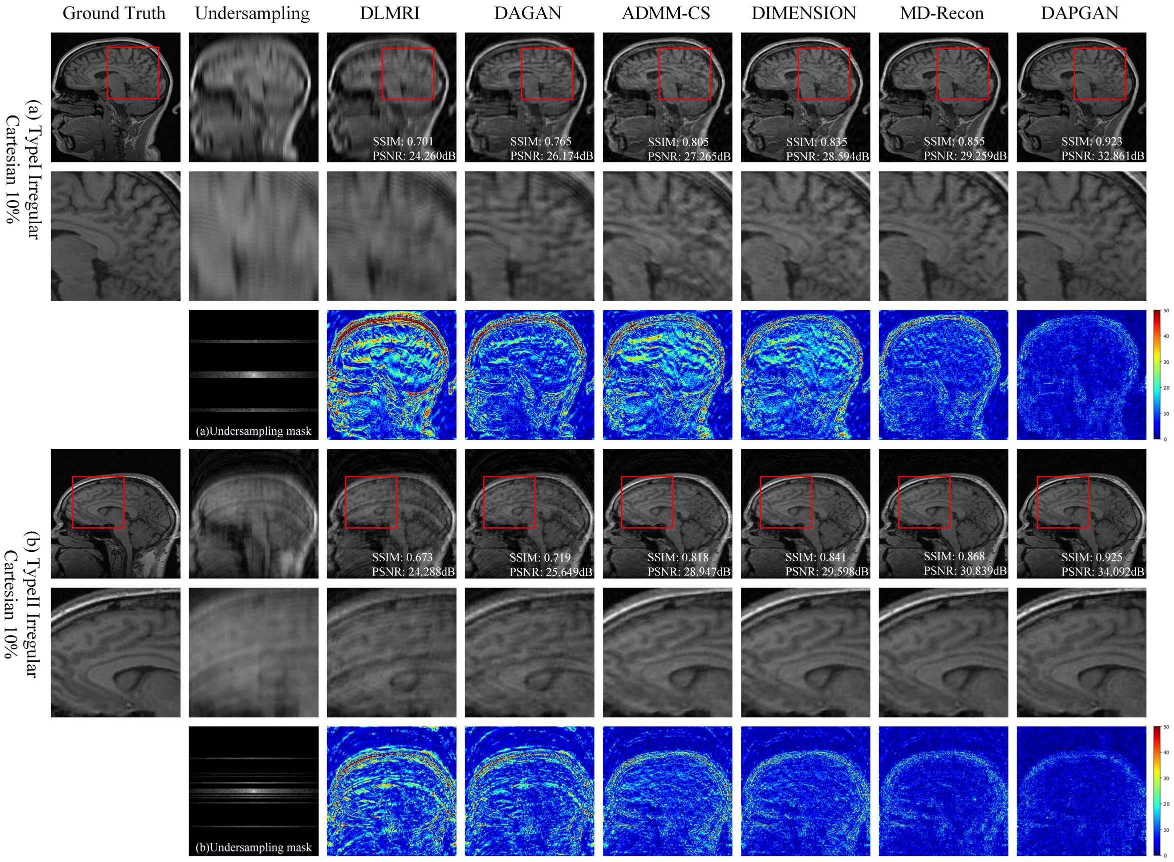


Figure 3.7: Visual inspection results of CS-MRI reconstruction on brain images using two types of irregular Cartesian sampling at rates 10%, here we show the comparison between the ground truth, undersampled results and investigated methods. The third column to the last column are the results of different methods. The second and fifth rows represent the magnifying views of the red boxes in the original images. The third row and last row show the irregular undersampling masks and the absolute difference maps (error maps compared to the full reconstruction) indicating the intensity of differences that range from 0 to 50 in the colorbar, which are visual inspected regions with multi-degrees of distinctions. To further verify the robustness of investigated methods, Type I and Type II of Cartesian sampling are conducted as challenging masks, with multi-irregular phase-encoding gradients. The lines in the raster of Type I reduce the density of central region compared with regular masks, while the space between lines in the raster of Type II is provided with random values, and the lines are randomly located. The quantitative evaluation metrics (SSIM/PSNR) of the reconstructed images are indicated in the first and fourth rows.

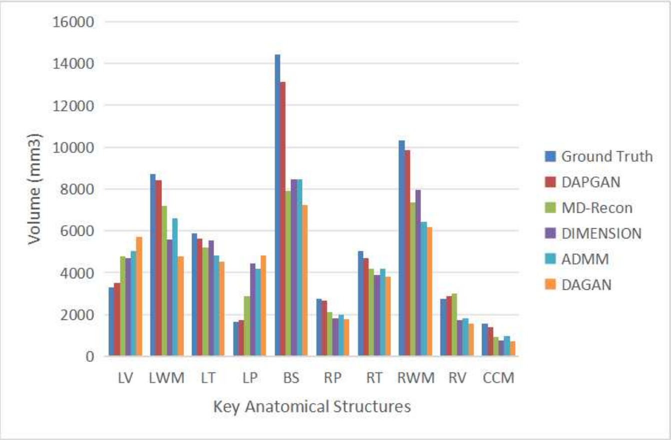


Figure 3.8: Quantification of brain volumetric indices based on segmentations derived from the original images (Ground Truth) and from images reconstructed from undersampled data using DAPGAN and each competing method of interest. (LV/RV: left/right ventricle, LWM/RWM: left/right white matter, LT/RT: left/right thalamus, LP/RP: left/right putamen, BS: brain stem, CCM: corpus callosum of middle-anterior)

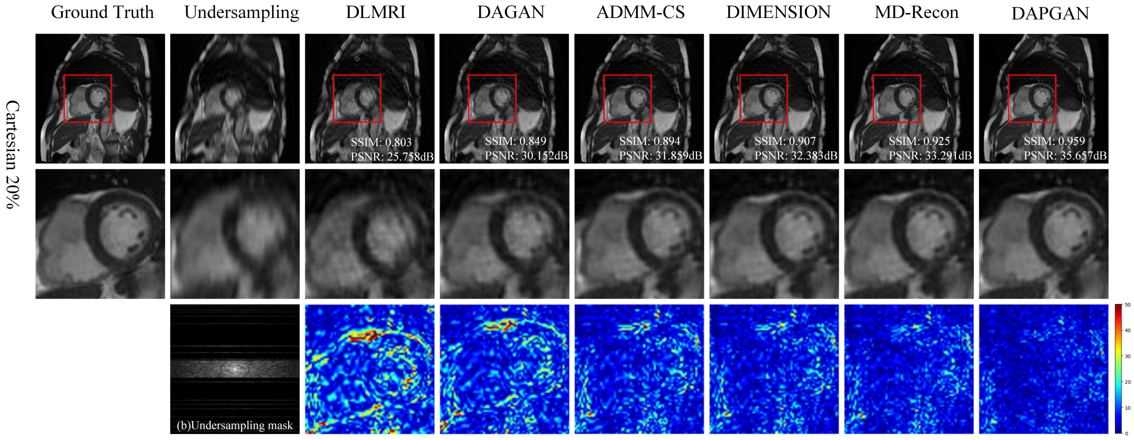


Figure 3.9: Visual inspection results of CS-MRI reconstruction on cardiac images using regular Cartesian sampling at rates 20%, here we show the comparison between the ground truth, undersampled results and investigated methods. The third column to the last column are the results of different methods. The second row represents the magnifying views of the red boxes in the original images. The third row shows the undersampling mask and the absolute difference maps (error maps compared to the full reconstruction) indicating the intensity of differences that range from 0 to 50 in the colorbar, which are visual inspected regions with multi-degrees of distinctions. The quantitative evaluation metrics (SSIM/PSNR) of the reconstructed images are indicated in the first row.

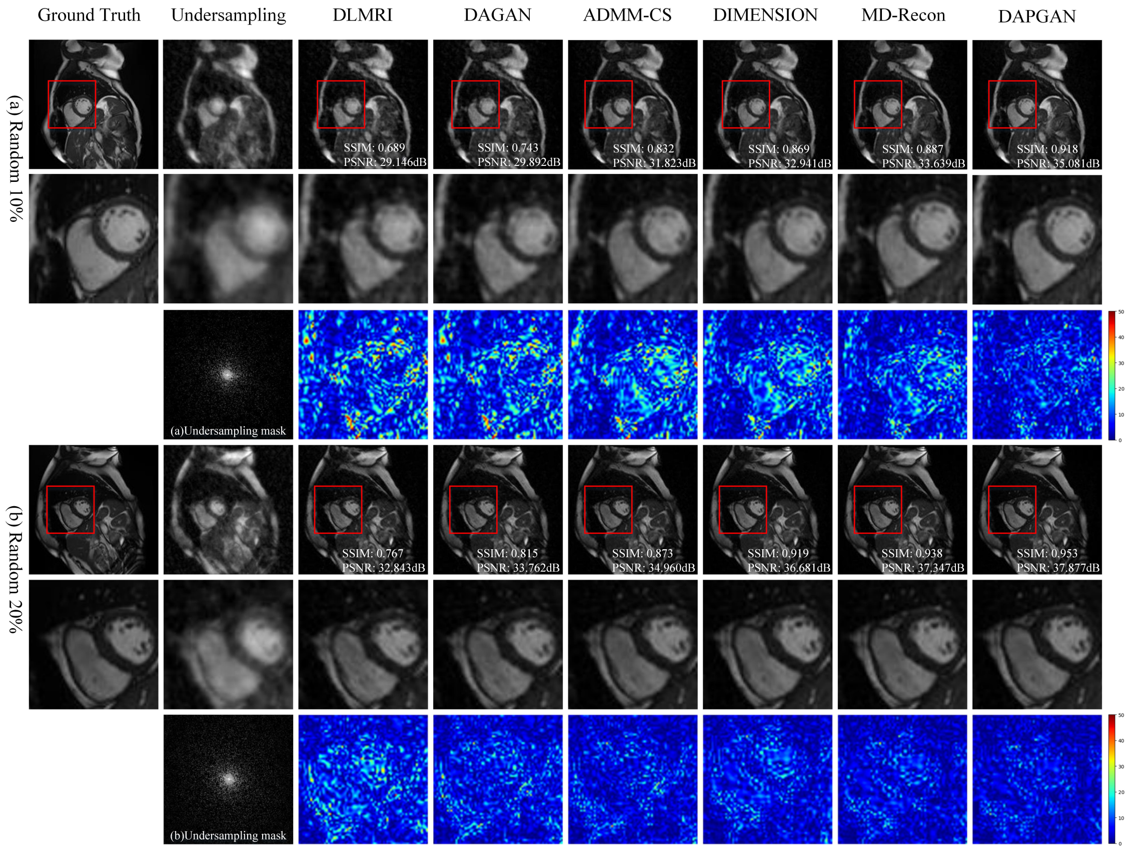


Figure 3.10: Visual inspection results of CS-MRI reconstruction on cardiac images using regular random sampling at rates 10% (from first row to third row) and 20% (from fourth row to last row), here we show the comparison between the ground truth, undersampled results and investigated methods. The second column (Undersampling) of the first and fourth rows represent the random sampling pattern in image space. The third column to the last column are the results of different methods. The second and fifth rows represent the magnifying views of the red boxes in the original images. The third row and last row show the undersampling masks and the absolute difference maps (error maps compared to the full reconstruction) indicating the intensity of differences that range from 0 to 50 in the colorbar, which are visual inspected regions with multi-degrees of distinctions. The quantitative evaluation metrics (SSIM/PSNR) of the reconstructed images are indicated in the first and fourth rows.

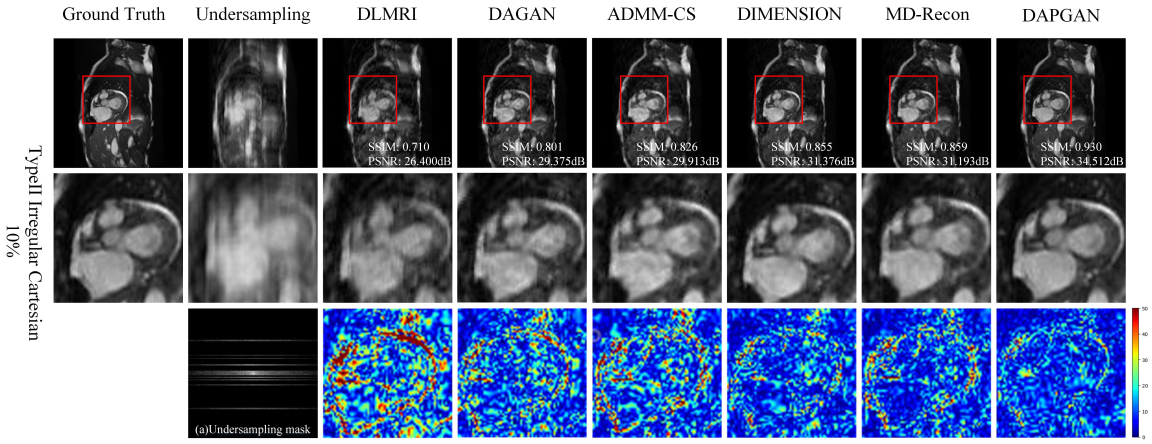


Figure 3.11: Visual inspection results of CS-MRI reconstruction on cardiac images using irregular Cartesian sampling at rates 10%, here we show the comparison between the ground truth, undersampled results and investigated methods. The third column to the last column are the results of different methods. The second row represents the magnifying views of the red boxes in the original images. The third row shows the irregular undersampling mask and the absolute difference maps (error maps compared to the full reconstruction) indicating the intensity of differences that range from 0 to 50 in the colorbar, which are visual inspected regions with multi-degrees of distinctions. To further verify the robustness of investigated methods, Type II of Cartesian sampling is conducted as a challenging mask, with irregular phase-encoding gradient. The space between lines in the raster of Type II is provided with random values and the lines are randomly located. The quantitative evaluation metrics (SSIM/PSNR) of the reconstructed images are indicated in the first row.

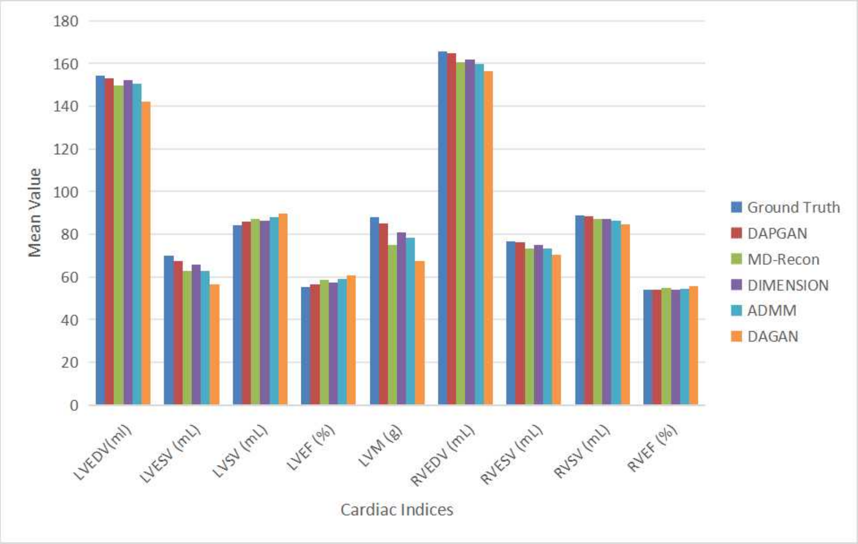


Figure 3.12: Mean value of 3D cardiac functional indices for the ground truth and investigated methods.

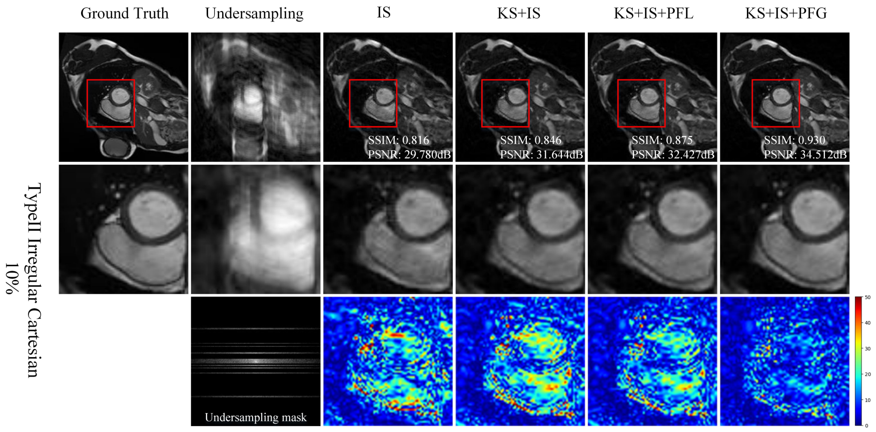


Figure 3.13: Visual inspection results of CS-MRI reconstruction on cardiac images using irregular Cartesian sampling at rates 10% for ablation study, here we show the comparison between the ground truth, undersampled results and investigated variants. IS: image-space Net only, KS + IS: dual-domain learning with k-space, KS + IS + PL: adding perceptual loss into the loss functions, KS + IS + PFG: a complete structure for DAPGAN with PFG. The third column to the last column are the results of different variants of our method. The second represents the magnifying views of the red boxes in the original images. The third row shows the undersampling mask and the absolute difference maps (error maps compared to the full reconstruction) indicating the intensity of differences that range from 0 to 50 in the colorbar, which are visual inspected regions with multi-degrees of distinctions. The quantitative evaluation metrics (SSIM/PSNR) of the reconstructed images are indicated in the first row.

CHAPTER 4

Multi-Attention-Based Feature Pyramidal Correlation Network For Cardiac Motion Estimation

Cardiac motion estimation plays a significant role in Cardiac MRI (CMRI) analysis and processing, including various primary downstream tasks, such as feature tracking, quantitative assessments of cardiac myocardium functional strain, and aiding in the diagnosis of cardiovascular diseases. Nevertheless, most existing approaches ignore accounting for the multi-range dependence within the motion estimation process. Additionally, efficiently leveraging features at different semantics levels to capture information at varying granularity remains a challenge. In this chapter, we propose a Multiple Attention-based Pyramidal Correlation Network (MAPC-Net) for accurate cardiac motion estimation. This method introduces a hybrid mechanism which integrates spatial and channel attention, allowing for exploring global and inter-spatial information in parallel. In particular, to demonstrate the significant deformation fields, we propose a multi-attention-aware pyramid correlation mechanism. The conventional correlation operation is optimised by embedding the integration of local textures with the corresponding global dependencies. Moreover, a multi-range mapping is introduced to enhance the capability of correlation computing. The proposed method was evaluated on UKBB and ACDC. The comparison results of experiments demonstrated the proposed method can achieve superior and consistent performance over the baseline approaches quantitatively and qualitatively.

4.1 Introduction

Cardiac magnetic resonance imaging (CMR) is a fundamental technology for evaluating cardiovascular morphology and function. Cardiac motion estimation (CME), developed from optical flow estimation for natural image registration, provides deformation information with meaningful insights, offering dynamic features that benefit various primary analyses for CMR, involving dynamic cardiac image reconstruction and unsupervised segmentation.

To estimate the optical flow field, conventional methods are conducted in a knowledge-driven mode and generally employ an energy function that includes a smoothness term and a data term. Horn and Schunck [178] pioneered the use of assumptions regarding smoothness and brightness constancy to address the constraints in generating deformation fields. However, these conventional approaches involve energy optimization processes that are computationally intensive and time-consuming.

As deep learning achieved increasing theoretical development, DL-based algorithms received significant attention due to the breakthrough application of CNNs. CNN-based networks offer the advantages of their hierarchical structure and the superior ability to abstract features across multiple scales. Recently, most CNN-based optical flow estimation approaches adopted the encoder-decoder architecture, termed U-Net [6]. Existing U-Net-based algorithms can be divided into the following stages during development, based on how they address the challenge of model size limitations, which arise from the simultaneous extraction of both large and small displacements. The initial concept of utilising U-Net for optical flow estimation was introduced by Dosovitskiy *et al.*[179], two fundamental applications, named FlowNetS and FlowNetC, are trained on the Flying Chairs dataset. Nevertheless, limited by the complexity of redundant design in the initial version of the correlation parameters, the estimated deformation fields can not afford superior performance, especially when compared with the conventional approaches. In the work of FlowNet2.0 [180], by establishing the warping layers which can be embedded into the cascaded FlowNet[179], an enhanced variant is conducted, and allows for more efficient utilisation of estimated flow fields across multi-scales. The approaches of the second stage focus on better balance accuracy and computational cost. As previously mentioned, although the encoder-decoder architecture improved the capability of motion estimation by conducting an extended receptive field, it still suffers from significant drawbacks, including large model sizes, an extensive number of network parameters, and high computational complexity for feature extraction, all of which reduce efficiency. The work proposed by Ranjan *et al.*[181] addressed this issue by indicating an algorithm based on pyramidal structure, named SpyNet, which estimated optical flow from coarse to fine at different convolutional levels. Despite SpyNet achieving performance similar to FlowNet, it remains less competitive compared to FlowNet2.0. PWC-Net[182] is inspired by the basic architecture of SpyNet. In addition, this work is the first framework to introduce the concept of utilising warping operations with upsampled flow fields fused with activated feature maps from the previous scale. As the next stage, it integrates the above feature maps with moving and fixed frames to achieve sequential correlation mapping from local to global in a pyramid structure. PWC-Net achieved milestone performance with promising results, and it kept the constancy on light-weight model size. However, how to adaptively refine the feature maps and facilitate the normalisation of the cost volume calculation has not been addressed by these methods.

For cardiac motion estimation, due to the lack of ground truth motion fields, constraining the learning with unsupervised loss functions is adapted by most deep learning-based approaches. In the works [183, 184], losses penalize differences in appearance and local spatial variations were utilised. For evaluations, Dice coefficients and Hausdorff Distance based on the respective segmentation are generally used metrics.

However, the performance of previous methods can be further improved from the following limitations: (1) existing DL-based methods still need to improve the ability to refine the process of motion estimation in an adaptive way, due to the lack of ground-truth for motion, (2) most methods ignore to balance the relationship between local small objects and the corresponding long-range spatial dependencies (3) most approaches only consider the pyramidal learning, estimating the motion field at the last stage by fusing the multi-scale features directly, the redundancy of stacking operation leads to less efficiency and increases the computation burden. While existing approaches ignore computing the correlation features more efficiently with contextual information. It is a better solution to enhance the discriminative ability of the correlation layer to improve the motion estimation accuracy and robustness with an adaptive guidance mechanism.

In this paper, we introduce a deep learning network for accurate cardiac motion estimation, termed a Multi-Attention-based Pyramidal Correlation Network (MAPC-Net). In this work, we mainly focus on improving the capability of correlation mapping at each scale of the pyramid network, in order to refine the motion fields from coarse to fine. The main contributions of our work can be summarized as follows:

1. We introduce a multi-attention mechanism into cardiac motion estimation task, to facilitate the ability to fine-learning local features of cardiac anatomical structures while emphasising global contextual information through modelling long-range dependencies.

2. We propose a multi-attention guided warping-correlation module, which can enhance the discriminative ability of feature representations, by exploring contextual information of moved frames and fixed frames individually, then conducting global dependencies embedded in the correlation procedure along channel-wise and spatial-wise combinationally.

3. We propose a pyramid correlation mapping mechanism based on different patch scales to match the specific feature level. This operation can improve the correlation efficiency while reducing the computational complexity compared with utilising a fixed patch size.

4.2 Problem Formulation

The moving frame and fixed frame are fed to the encoding network for feature extraction, in order to form the feature pyramid with the output from the parallel frames based on weight sharing. At each level of the feature pyramid, the motion estimated by the previous scale is upsampled to the current scale, and then the frame from the moving image is warped with the upsampled motion to the fixed image.

The warping operation can be formulated as:

$$\boldsymbol{F_w} = \boldsymbol{F_2}\left(x + \boldsymbol{W_s}\right), \tag{4.1}$$

where the warped image is presented by $\boldsymbol{F_w}$, x denotes the spatial position of pixel point for moving image $\boldsymbol{F_2}$ and $\boldsymbol{W_s}$ denotes the upsampled motion field from the previous level.

Conventionally, the correlation layer is conducted following the feature warping, facilitating the cost volume calculating through matching moved frame $\boldsymbol{F_w}$ to fixed frame$\boldsymbol{F_1}$, in addition, a patch with fixed correlation range is indicated here. The processing can be expressed as follows:

$$\boldsymbol{C}\left(x_1, x_w\right) = \sum_{\mathbf{o}\in\Omega} \left\langle \mathbf{F}_1\left(\mathbf{x}_1 + \mathbf{o}\right), \mathbf{F}_w\left(\mathbf{x}_w + \mathbf{o}\right)\right\rangle . \tag{4.2}$$

where x_1 and x_w indicate the point positions at $\boldsymbol{F_1}$ and $\boldsymbol{F_w}$ respectively. $o \in \Omega = [-k, k] \times [-k, k]$ is the correlation range (displacement size).

4.3 Methodology

The proposed method, Multi-Attention-based Pyramidal Correlation Network (MAPC-Net), is presented from three aspects in this section. First, we introduce the correlation mapping pyramid conducted with multiple displacement ranges (Section 4.3.1). Then, the combinational attention module is introduced in Section 4.3.2. The core innovation of the proposed method, the multi-scale attention-guided warping-correlation mechanism, is presented in Section 4.3.3. At the last, a fusion and normalisation block is proposed in Section 4.3.4.

4.3.1 Multi-range Correlation Mapping Pyramid

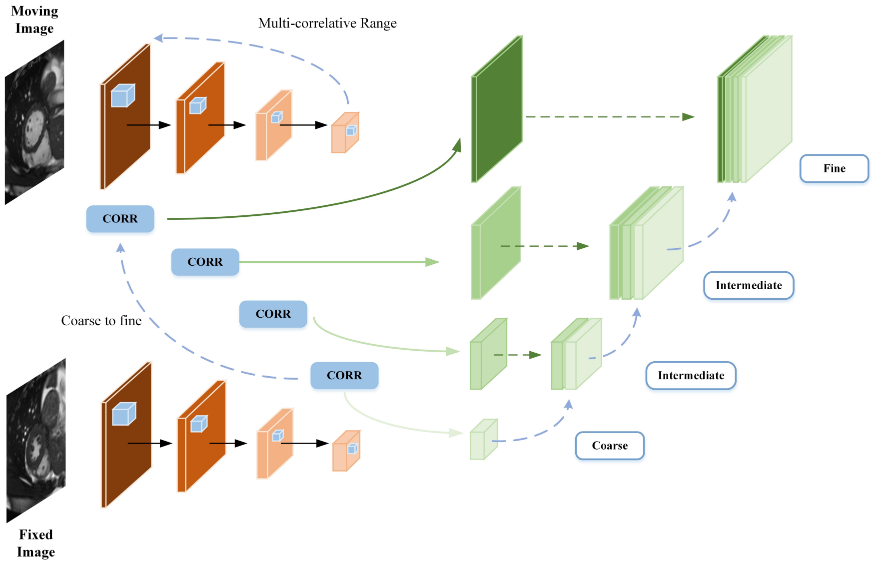


Figure 4.1: The architecture of the proposed multi-range correlation mapping pyramid. The displacement patch size is flexible regarding different feature levels. In addition, the calculated correlation maps are mapped and concatenated from coarse to fine.

Inspired by FlowNetC[179] and PWC-Net[182], it is the most adopted framework to extract pyramidal features from weights sharing and pair-wise frames of different levels. Feature correlation between warped and fixed images at each pyramidal level is an essential stage of the refinement for estimated motion. Nevertheless, a unified value for patch size performed at different feature levels limited the computational efficiency due to the receptive field of correlation not matching the scale of the feature map, as one of the reasons leading to failure to preserve the deformation details. In order to overcome this obstacle, we conduct a variety of displacement sizes for different pyramidal levels, as shown in Fig. 4.1. In particular, a correlation mapping mechanism is applied as the fundamental architecture of the proposed network. By concatenation operation between the upsampled correlation feature from all the previous levels and the feature from the current stage, potential displacement information is extracted and propagated from coarse to fine. The displacement sizes from high-level to low-level are 2, 2, 4, and 8, respectively.

$$\boldsymbol{C}_l = \begin{cases} \boldsymbol{C}_1^c, & l = 1 \\ \boldsymbol{C}_l^c \odot \left(\boldsymbol{C}_{l-1}^u\right), & l > 1 \end{cases} \tag{4.3}$$

where $\boldsymbol{C}_{l-1}^u$ denotes the correlation feature from the previous level $l-1$ through upsampling, with the sampling size of 2, and average pooling additionally. $\odot$ represents the concatenation operator along the channel dimension.

4.3.2 Combinational Self-Attention module

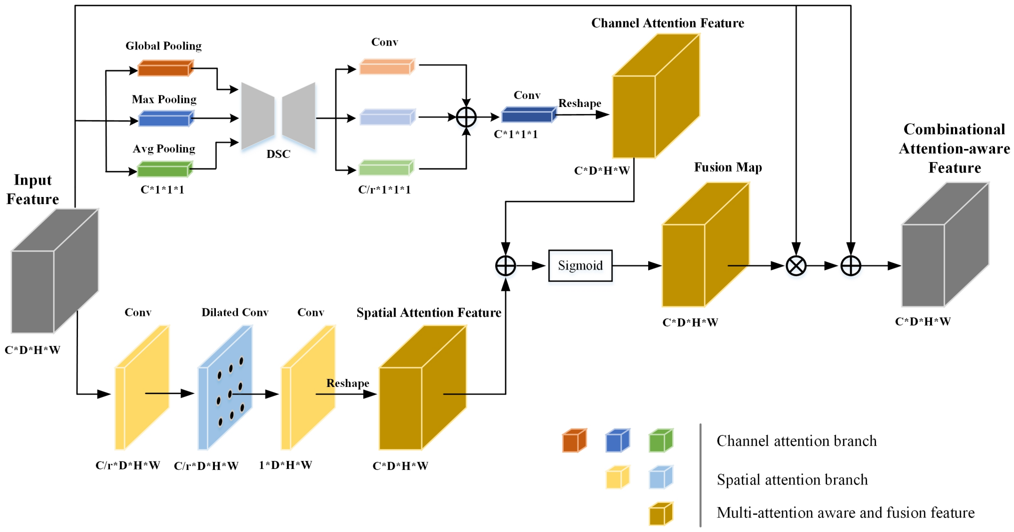


Figure 4.2: Illustration of the Combinational Attention. Two sub-branches including attention mechanism along channel and spatial dimensions are applied, respectively. The channel attention conducts three parallel pooling blocks and general convolutional layers. The spatial attention branch is formed through both general and dilated convolution operations. A fusion module is introduced as the final stage.

As shown in Fig. 4.2, the combinational self-attention module is conducted with two sub-pipeline: channel-wise attention $\boldsymbol{CA}$ and spatial-wise attention $\boldsymbol{SA}$, respectively.

Channel Attention branch

The channel attention branch is conducted with three stages, first, to yield spatial context descriptors, three parallel pooling operators $\mathbf{F^c_{avg}}$, $\mathbf{F^c_{max}}$, $\mathbf{F^c_{global}}$ are performed on the input feature $\mathbf{F} \in \mathbb{R}^{C \times D \times H \times W}$. At the second stage, a weights-sharing network is applied based on multi-layer deepwise separable convolutions (DSC), for receiving the three descriptors. In particular, an activated hidden operator with the size of $\mathbb{R}^{C/r \times 1 \times 1 \times 1}$ to achieve a reshaped feature size. At the last stage, the three features output from DSC are concated and fed to a convolutional layer to form the feature $\mathbf{CA_1}$ with the output size of $C \times 1 \times 1 \times 1$, then the feature map is reshaped to $C \times D \times H \times W$ again to produce the final channel attention map $\mathbf{CA_2}$.

$$\begin{aligned}\mathbf{CA_1} &= \mathrm{DSC}(\mathrm{AvgPool}(\mathbf{F})) + \mathrm{DSC}(\mathrm{MaxPool}(\mathbf{F})) + \mathrm{DSC}(\mathrm{GlobalPool}(\mathbf{F})) \\ &= \mathbf{W_1}\left(\mathbf{W_0}\left(\mathbf{F^c_{avg}}\right)\right) + \mathbf{W_1}\left(\mathbf{W_0}\left(\mathbf{F^c_{max}}\right)\right) + \mathbf{W_1}\left(\mathbf{W_0}\left(\mathbf{F^c_{global}}\right)\right),\end{aligned} \tag{4.4}$$

$$\mathbf{CA_2} = \mathrm{Conv}\left(\mathrm{Conv}\left(\mathbf{CA_1}\right)\right), \tag{4.5}$$

Spatial Attention branch

According to the different stages of the proposed multi-attention guided correlation in section 4.3.3, the spatial attention utilised in the module can be categorised into two types: we perform the attention block based on criss-cross [185] for the pair-wise input including warped frame and fixed frame before fed into correlation stage, a normal spatial attention block is applied as a branch in the combinational self-attention module as the third stage.

To achieve encoding the contextual feature across cardiac anatomical structures into local objects with more details, a spatial attention branch is introduced. Based on this, the long-range dependencies can be further explored in an adaptive style.

As Fig. 4.2 shows, the attention feature map is obtained through a hybrid convolution block, including standard and dilated convolution operations. Assume $\mathbf{F} \in \mathbb{R}^{C \times D \times H \times W}$ as the input feature, the shape of the feature after the first convolution layer Conv_1 (kernel size is 1) is $\mathbb{R}^{C/r \times D \times H \times W}$, with a reduced channel dimension C/r by the rate r of 16. A dilated convolutional layer is conducted to capture the contextual relationship across long-range features. Further, a standard 3D convolution layer Conv_2 with kernel size of 1 and reshape operation are applied sequentially, the shape final attention map $\boldsymbol{SA}$ is $C \times D \times H \times W$.

$$\boldsymbol{SA} = \mathrm{Conv}_2(D(\mathrm{Conv}_1(\boldsymbol{F_c}))), \tag{4.6}$$

where Conv(·) denotes the standard 3D convolution, D(·) indicates the dilated convolutional layer, presents the Sigmoid function.

$$\boldsymbol{A_{sum}} = \mathrm{Sig}\left(\boldsymbol{SA} + \boldsymbol{CA_2}\right). \tag{4.7}$$

By conducting a Sigmoid function Sig(·), the fusion map between across channel and spatial wise is gated to the hybrid attention map $\boldsymbol{A_{sum}}$ as a weight vector.

4.3.3 Multi-scale Attention guided Warping - Correlation

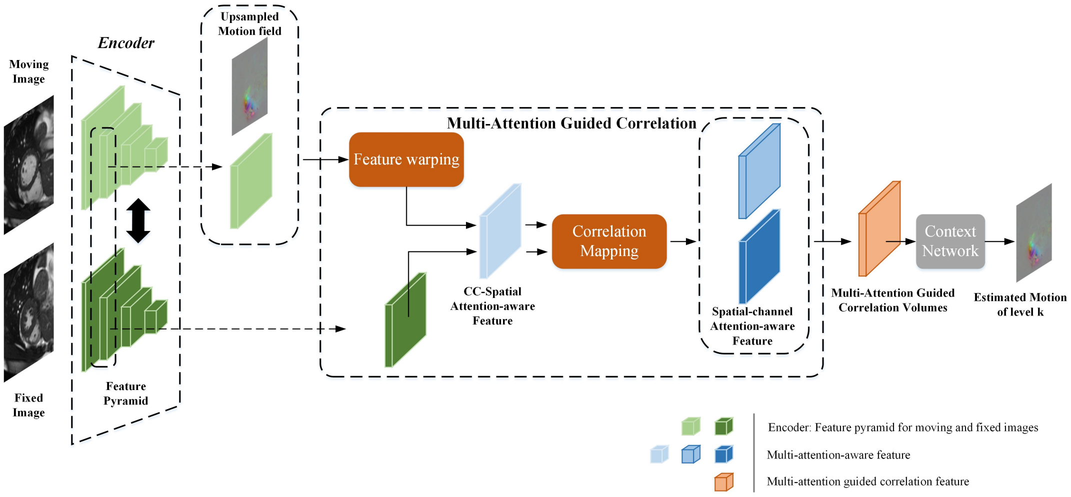


Figure 4.3: The proposed Multi-scale Attention guided Warping - Correlation. In each pyramidal level, Criss-Cross attention is performed for warped images and fixed images, the combinational attention is performed after the correlation layer.

With the intention to refine the motion estimation process with a more precise prediction with global and local features from coarse to fine, the multi-attention guided warping-correlation mechanism is performed at each pyramidal level. Given the pair-wise images F_1 and F_2, each scale of the proposed network performs an individual estimation of the motion field to access the corresponding warped image F_w. As shown in Fig. 4.3, the proposed multi-scale attention-guided warping-correlation mechanism is the core step of the global framework. The complete process includes feature pyramid encoding to extract pair-wise image features from low-level to high-level with stride size of 2 and output channel of 16, 32, 64, 96 and 128; criss-cross spatial attention (CC-attention) for warped and fixed images; multi-attention correlation mapping (the displacement ranges are 8, 4, 2, 2, respectively); correlation fusion block and motion estimator.

Moving image $F_{k,2}$ and fixed image $F_{k,1}$ at each level are first fed in the feature warping module, $F_{k,2}$ is warped to the view of $F_{k,1}$ via the upsampled flow M_{k-1} from stage $k-1$. The warped feature is denoted as $\hat{F}_{k,2}$. Then, considering enhancing the generalisation on the dependencies of both large and small displacements, for a spatial refinement, the pair-wised features of warped image and fixed image are fed into the criss-cross attention module in parallel, the process can be defined as

$$\begin{aligned} \boldsymbol{F_{ccAW}} &= \boldsymbol{SA_{cc}} \otimes \boldsymbol{\hat{F}_{k,2}} + \boldsymbol{\hat{F}_{k,2}}, \\ \boldsymbol{F_{ccAF}} &= \boldsymbol{SA_{cc}} \otimes \boldsymbol{F_{k,1}} + \boldsymbol{F_{k,1}}, \end{aligned} \tag{4.8}$$

where $\otimes$ represents element-wise multiplication.

Following the warping layer, we utilise the correlation layer to calculate the matching cost based on the cost volumes, in addition, we performed specific correlation patch sizes for different layers. The correlation operation can be formulated as

$$\boldsymbol{C}\left(x_1, x_w\right) = \sum_{\mathbf{o} \in \Omega} \left\langle \mathbf{F}_1\left(\mathbf{x}_1 + \mathbf{o}\right), \mathbf{F}_w\left(\mathbf{x}_w + \mathbf{o}\right)\right\rangle, \tag{4.9}$$

where x_1 and x_w present the point positions at $\boldsymbol{F_{k,1}}$ and $\boldsymbol{F_{k,w}}$ respectively. $o \in \Omega = [-k, k] \times [-k, k]$ is the mapping patch size of the current level.

To boost the ability of emphasising reliable details in the correlation maps, a combinational attention module $\boldsymbol{A_{sum}}$ is embedded into each pyramid level to refine the correlation features from spatial and channel aspects. The module can be defined as

$$\boldsymbol{F_{ACorr}} = \boldsymbol{A_{sum}} \otimes \boldsymbol{C} + \boldsymbol{C}, \tag{4.10}$$

where the $\boldsymbol{A_{sum}}$ denotes the combinational attention map which is fused by two separated attention maps, channel-wise attention map and spatial attention map.

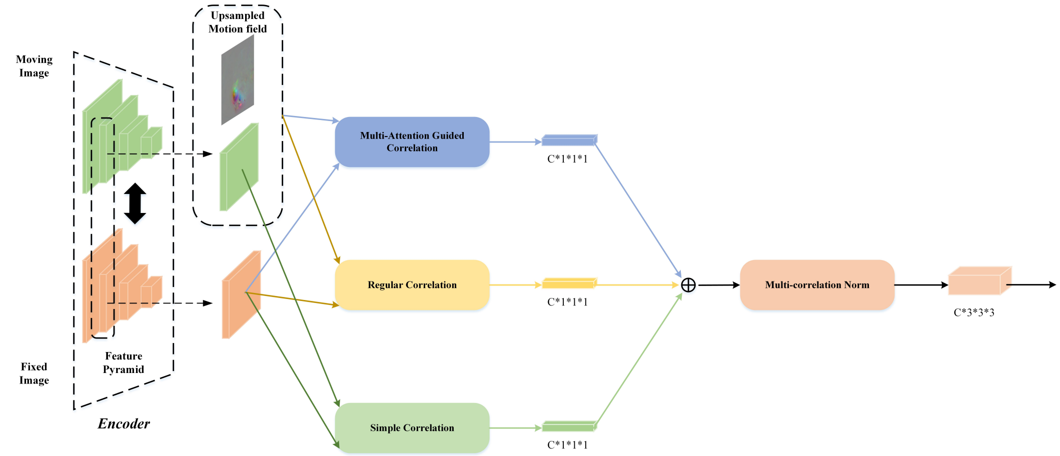


Figure 4.4: Illustration of the proposed multi-correlation fusion and normalization module. An attention-based mechanism is performed for an adaptive fusion of multiple correlation maps, and a channel normalisation is conducted following the fusion operation to enhance the feature generalisation.

4.3.4 Multi-correlation Fusion and Normalisation

Based on our knowledge, current methods adopt two types of correlation operations to calculate the matching cost with cost volumes, correlation between moving image and fixed image, and correlation between warped image and fixed image. Considering performing an effective trade-off between the existing correlations and ours, we propose a fusion and normalization module for adaptive refinement of multi-correlations. Fig. 4.4 shows the details of the fusion module, each type of correlation followed by a $1*1*1$ convolutional layer. After the fusion operation, we utilise a channel normalisation module followed by a standard convolutional layer with a kernel size of 3.

$\boldsymbol{I}$ presents the input feature, we express the normalisation as:

$$\begin{aligned} \boldsymbol{O_{fusion}} &= Conv_1\left(\boldsymbol{F_{A-Corr}}\right) + Conv_1\left(\boldsymbol{F_{Corr1}}\right) \\ &\quad + Conv_1\left(\boldsymbol{F_{Corr2}}\right), \\ \boldsymbol{O_{norm}}(\mathbf{x}, c) &= Conv_2\left(I(\mathbf{x}, c) / \left(\alpha \sum_{\mathbf{x}} (O_{fusion}(\mathbf{x}, c))^2 + \epsilon\right)^{\beta}\right), \end{aligned} \tag{4.11}$$

where $\boldsymbol{F_{A-Corr}}$ denotes the attention-guided correlation, $\boldsymbol{F_{Corr1}}$ and $\boldsymbol{F_{Corr2}}$ are the two types of conventional correlation mappings. c presents the channel of $\boldsymbol{I}$. The weight of multiplication, the exponent and the respective additive constant utilised in normalisation, are presented by α, β, ϵ, respectively.

We applied motion estimator layers after the normalisation block to infer better representations for the final estimated motion, scale by scale, until reaching the full resolution.

4.3.5 Training Loss Function

Unsupervised loss for consistency

As the data fidelity term in the summary loss function, an unsupervised loss function is performed for evaluating the registration consistency between the moved image utilising estimated motion field $W_k(\mathbf{x})$ to warp the moving image $F_{k,2}$ to achieve registration with fixed image $F_{k,1}$ in level k. The process is formulated as:

$$\mathcal{L}^{C}_{\phi,k} = \sum_{\mathbf{x}} \left| F_{k,1}(\mathbf{x}) - F_{k,2}\left(\omega\left(\mathbf{x}, W_k(\mathbf{x})\right)\right) \right|, \tag{4.12}$$

where $\omega(\cdot)$ denotes the warping function conducted by linear interpolation, ϕ represents the parameters of the proposed network.

Smoothness loss for regularisation

To achieve balanced performance between the spatial smoothness of the estimated motion and preserving the discontinuity deformation at the edge area of the cardiac region, a smoothness loss as the regularisation term for the deformation field is formulated as follows:

$$\mathcal{L}^{R}_{\phi,k} = \sum_{d\in\{u,v,w\}} \sum_{\mathbf{x}} \left| \nabla W_k(\mathbf{x}, d) \right| \circ \exp\left(-\left| \nabla F_{k,1}(\mathbf{x}) \right|\right), \tag{4.13}$$

where d represents the 3D direction u, v, w of motion field.

Summary Loss

$$\mathcal{L}_{\mathrm{SUM}} = a\mathcal{L}^{C}_{\phi,k} + b\mathcal{L}^{R}_{\phi,k}. \tag{4.14}$$

where a, and b are weighting factors that control the trade-off between the different terms in the overall loss function.

4.4 Experiments and Results

4.4.1 Dataset and Implementation

The proposed method, MAPC-Net, is trained on the UKBB [176] dataset and evaluated on UKBB and another publicly available dataset ACDC. We chose 1000 short-axis (SAX) subjects and stacked each slice from the same case to a volume 3D data. In addition, we select time frames of end diastolic (ED) and end systolic (ES) from each case for the experiments (ED to ES motion estimation). Each subject includes pairwise images and respective manually annotated segmentation masks by experts. The images were acquired with the following acquisition protocol: spatial resolution of $1.8 \times 1.8 \times 10mm^3$. We utilised cropping and padding to pre-process our dataset to a fixed size of $128 \times 128 \times 16$. The dataset is split into a training set (600), a validation set (200), and a test set (200). The proposed model was implemented in PyTorch, on an NVIDIA A100 GPU machine. The learning rate of our model is 1e-4 and the batch-size is 3. The hyper-parameters a, b in the total loss $\mathcal{L}_{\text{SUM}}$ (refer to Equation. 4.14) were set to 1, 0.01 respectively, which were tuned using the validation set. α, β, ϵ in Equation. 4.11 were set to 0.99, 0.5 and 0.01.

4.4.2 Competing Methods and Comparison settings

We compare the estimated motion of the proposed MAPC-Net to conventional registration methods Symmetric Normalisation (SyN) registration in ANTS[186] and Demons (with 800 iterations and standard deviations 1.0)[187], in addition, DL-based approaches Voxelmorph[188], and DDIR[189]. In particular, MAPC-Net utilises segmentation masks as an evaluation for proving the accuracy of the estimated optical flow only in an unsupervised style. For a fair comparison, we build an unsupervised version for the pipeline approach DDIR which leverages the segmentation as the condition for training.

To demonstrate the advantage of the proposed modules including multi-range correlation mapping and multi-attention-based warping-correlation, ablation study including different variants is applied.

To quantitatively compare MAPC-Net and the investigated approaches, we utilised several metrics for evaluation including Dice Score (DS - %) and the 95% - Hausdorff Distance (HD95 - mm) across the three anatomical cardiac structures LV blood-pool

(LVBP), LV myocardium (LVM) and RV blood-pool (RV), by warping the segmentation mask from ED frame to ES frame with the estimated cardiac motion fields. Then we computed the two metrics between the ground truth segmentation and warped segmentation. To evaluate the smoothness of the investigated approaches, the average number of positions with a negative Jacobian determinant value (NJD - %) in the deformation fields is performed, which measures the diffeomorphism of the estimated deformation fields. Besides, to demonstrate the superior capability in clinical scenarios of MAPC-Net, we conducted strain analysis for peak signal at the ES frame.

4.4.3 Results tested on UKBB data

Quantitative Results

Table 4.1: Quantitative comparison between MAPC-Net and investigated methods on testing datasets over UKBB.

Methods	Avg.DS(%)	DS-LVBP(%)	DS-LVM(%)	DS-RV(%)	Avg.HD95(mm)	HD95-LVBP(mm)	HD95-LVM(mm)	HD95-RV(mm)	NJD(%)
befor Reg	45.29(7.49)	56.83(6.19)	43.29(8.86)	35.76(7.42)	12.15(4.31)	10.23(2.83)	12.47(3.29)	13.75(6.82)	-
Demons	63.79(15.84)	69.29(12.57)	58.91(15.93)	63.16(19.02)	9.16(4.79)	7.17(3.59)	8.92(3.42)	11.38(7.35)	0.36(0.23)
SyN	65.14(15.36)	70.37(13.38)	63.26(17.25)	61.78(15.45)	6.91(4.20)	6.47(2.31)	4.59(2.33)	9.66(7.96)	1.60(0.75)
VM	69.66(15.72)	78.57(10.62)	66.82(17.66)	63.58(18.89)	5.98(4.26)	4.53(2.48)	5.30(3.52)	8.12(6.79)	0.34(0.25)
DDIR	75.20(7.74)	81.25(9.25)	70.12(5.61)	**74.22(8.36)**	5.31(1.57)	3.81(1.38)	**3.94(1.03)**	8.19(2.31)	0.30(0.26)
MAPC	**77.20(6.18)**	**84.37(6.35)**	**75.02(5.47)**	72.20(6.73)	**4.44(2.23)**	**3.27(1.06)**	4.72(2.93)	**5.32(2.71)**	**0.20(0.15)**

As shown in Table 4.1, DL-based approaches outperform conventional registration methods in terms of both DS and HD95. The average Dice score achieved by conventional methods SyN and Demons are around 60.00%-65.00% based on the three anatomical structures. Voxelmorph improves the performance to higher than 65.00%, and further improves LVBP from 70.00% to 78.00%. While DDIR and our method yield better results than VoxelMorph, DDIR achieves 5.54% higher average DSC than VM, while our method further improves the average value by 7.54% than VM. In addition, DDIR and MAPC improve the Dice score on LVBP and LVM to over 80.00% and 70.00%, respectively. In particular, for the three anatomical sub-regions of interest in the heart, our method achieved the best performance consequently on LVBP and LVM. For RV, our method also achieved very close performance to the SOTA. Our method outperforms other investigated methods in terms of HD95 on both the global area and the three sub-regions. Demons and SyN achieved an average HD95 higher than 6.00mm. VM achieved an average value slightly lower than 6.00mm, reaching

5.98mm. Our method yielded an outstanding average value of 4.00mm and the lowest value of 3.27mm and 5.32mm on LVBP and RV, respectively. We found MAPC further improved the HD registration performance 15.00% better than DDIR compared with VM (25.80% vs 11.20%).

Qualitative Results

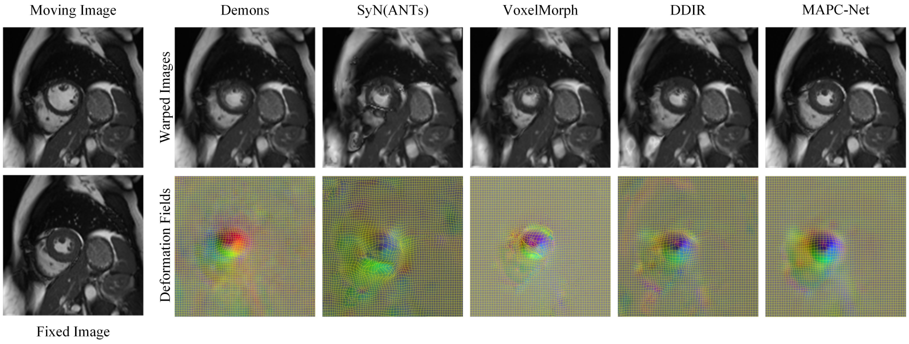


Figure 4.5: Visual comparison of results on UKBB motion estimation using MAPC-Net and state of the art methods. The corresponding warped moving images, and deformation fields with grids in rows 1-2 based on the estimation methods including Demons, SyN, Voxelmorph, DDIR and MAPC (columns 2-6).

Fig. 4.5 presents the visual inspections for the comparison between the investigated methods and the proposed method in terms of the consistency between the warped images and the fixed image, the estimated motion is presented which demonstrates the spatial smoothness of the deformation fields additionally. In order to achieve the most obvious contrast of deformation estimated based on different approaches, we selected the ES time point as the fixed time frame and, the image at the ED time point as the moving frame, as shown in the first column. The warped images obtained by the conventional registration method Demons and SyN (ANTs) perform distinctly different structures to the fixed image on LV and RV, respectively. In addition, partial deformation information is missed in the estimated motion from the two methods. While DL-based methods can achieve more accurate anatomical features, in addition, improved accuracy of deformation within terms of the inside region and edges can be observed. Despite DDIR achieving significant improvement on LVBP and RV than VM, some detailed textures are still distinctly different to fixed images on LVBP and

RV. While MAPC-Net achieved a better quality with more precise and reasonable anatomical features in terms of texture and edge across all the anatomical regions. The deformation fields with grids also suggested that our method achieves a better balance performance between improving the similarity with the fixed image and preserving the smoothness of the estimated motion.

4.4.4 Results tested on ACDC data

Table 4.2: Quantitative comparison between MAPC-Net and investigated methods on testing datasets over ACDC.

Methods	Avg.DS(%)	DS-LVBP(%)	DS-LVM(%)	DS-RV(%)	Avg.HD95(mm)	HD95-LVBP(mm)	HD95-LVM(mm)	HD95-RV(mm)	NJD(%)
befor Reg	59.11(7.68)	63.39(8.78)	53.61(6.67)	60.33(7.58)	6.95(1.41)	5.37(1.59)	7.21(1.26)	8.28(1.38)	-
Demons	75.79(3.27)	82.97(2.42)	73.18(4.57)	71.23(2.81)	4.56(1.27)	3.06(1.21)	5.97(1.57)	4.66(1.03)	0.31(0.26)
SyN	73.52(7.99)	80.45(9.24)	70.62(8.36)	69.85(6.38)	4.99(1.32)	3.43(1.29)	6.24(1.33)	5.31(1.35)	0.46(0.21)
VM	77.49(4.19)	83.78(3.61)	75.42(5.29)	73.28(3.68)	3.41(1.27)	2.43(1.15)	3.94(1.50)	3.85(1.17)	0.27(0.22)
DDIR	80.59(4.18)	85.13(2.35)	80.17(6.74)	76.48(3.46)	2.98(1.16)	2.29(1.26)	3.18(1.03)	3.47(1.18)	0.18(0.10)
MAPC	**86.92(3.06)**	**92.19(2.62)**	**85.33(4.51)**	**83.24(2.04)**	**1.62(0.58)**	**1.16(0.63)**	**1.52(0.32)**	**2.18(0.79)**	**0.16(0.11)**

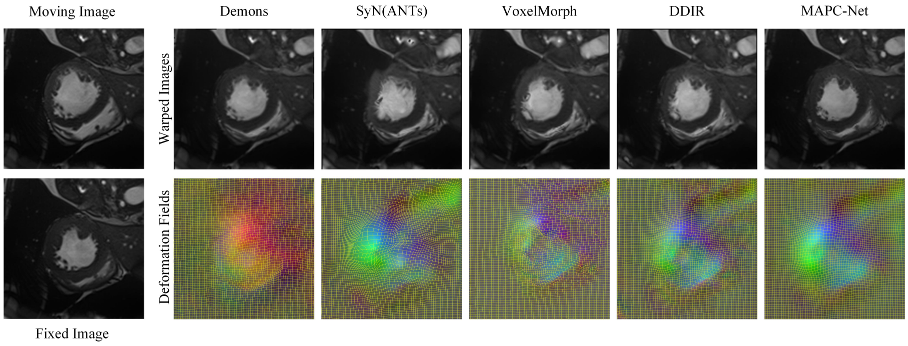


Figure 4.6: Visual comparison of results on ACDC dataset motion estimation using MAPC-Net and state of the art methods.

To highlight the benefits of the proposed MAPC-Net in terms of the robustness of samples from different datasets, especially generalising to multiple real-world data acquired routinely in clinical examinations, we apply the pre-trained methods leveraging UKBB CMR cases, to ACDC particularly. For most cases in ACDC, less significant deformation can be observed from ED to ES time frames, which leading the difference in the estimated motion between investigated methods are not pronounced. The quantitative and qualitative results obtained from investigated methods are shown in Table 4.2

and Fig. 4.6, respectively. In Table 4.2, for conventional methods, conventional registration approaches are still worse than deep learning-based methods across all metrics. For the Dice score, SyN only achieved an average value with 73.00%, while Demons achieved better performance than SyN (2.27% higher) on average performance. VM achieve only marginally better performance than Demons. While MAPC outperforms other methods significantly. In particular, improved performances on average DS with an increase of 11.13% and HD95 with a decrease of 2.94mm than Demons are achieved by MAPC-Net, compared with 4.69% and 1.58mm achieved by DDIR. MAPC yielded the best performance across all the three sub-regions of heart correspondingly, and we found that it is also the only method that achieved performance of over 80.00% on each individual anatomical region, with a significant improvement on LVBP (over 90.00%, 9.22% higher than Demons, 7.06% higher than DDIR). This highlights the merit of a multi-attention-aware correlation block. The proposed method further achieved superior registration accuracy across all anatomical regions with the lowest NJD and HD95. As shown in Fig. 4.6, the prediction of SyN and Demons are heavily limited by the over-smoothed texture and object boundaries in terms of the optical flow of motion fields. While for the warped moving images, unreliable structures are generated particularly in some specific regions including LVBP and LVmyo. Despite VM and DDIR fixed this relatively, the consistency of small objects is especially evident across RV and LVBP, which suggests that MAPC-Net can achieve the best similarity to the fixed image. In addition, the motion fields with grids indicate that the estimated motion by MAPC-Net presents a better capability to reduce the size of the negative Jacobian determinant. This visual qualification is consistent with the corresponding quantitative results. The deformation fields further indicate that the motion fields estimated by the proposed method preserve accurate smoothness while maintaining the accuracy of the warped image similarity.

4.4.5 Ablation Study

To analyse the contribution of each module in MAPC-Net, we conducted an ablation study on the proposed network evaluated by UKBB data. The results are shown in Table 4.3, where RP denotes performing motion estimation task with a regular pyramid network purely, $CM + FP$ represents correlation-mapping operation based on fixed patch size, $CM + MP$ indicates correlation mapping based on multi-patch size,

Table 4.3: Quantitative comparison between MAPC-Net and investigated variants on testing datasets over UKBB.

Methods	Avg.DS(%)	DS-LVBP(%)	DS-LVM(%)	DS-RV(%)	Avg.HD95(mm)	HD95-LVBP(mm)	HD95-LVM(mm)	HD95-RV(mm)	NJD(%)
RP	66.01(13.34)	75.42(10.53)	62.07(15.61)	60.53(13.88)	8.81(6.27)	6.24(3.37)	9.83(7.42)	10.36(8.01)	2.38(0.91)
CM+FP	69.24(12.22)	77.84(8.34)	67.39(15.84)	62.49(12.47)	6.81(2.12)	5.29(1.17)	5.47(2.54)	9.67(2.65)	1.53(0.61)
CM+MP	72.97(11.48)	81.79(9.65)	71.51(12.97)	65.61(11.82)	5.41(3.07)	4.03(2.38)	5.29(3.21)	6.91(3.63)	0.70(0.38)
CM+MP+AC	**77.20(6.18)**	**84.37(6.35)**	**75.02(5.47)**	**72.20(6.73)**	**4.44(2.23)**	**3.27(1.06)**	**4.72(2.93)**	**5.32(2.71)**	**0.20(0.15)**

$CM + MP + AC$ denotes the complete structure of MAPC-Net, including correlation mapping operation based on multi-patch size at each pyramidal level, and attention-aware correlations.

According to Table 4.3, based on the conventional pyramidal structure as an alternative network, the average performance on Dice score is 66.00%, which is marginally better than the performance of SyN in Table 4.1 (66.01% vs 65.14%), nevertheless, the performance of HD95 utilising RP is 27.00% higher than SyN. A fixed patch size for deformation is a widely adopted mechanism by most pyramidal correlation-based approaches. The variant with multi-scale correlation mapping marginally improved the average DS and HD95 by 3.23% and 2mm, respectively. However, the performance of DS leveraging $CM + FP$ still lower than 70.00%, and the performance be achieved is roughly at the same level compared with VM (69.66% vs 69.24% Dice score) is evident to verify the limited capability by performing conventional correlation mapping additionally. When we alternate the fixed size with multi-patch size which is adjusted to the pyramidal level, the variant achieved a value above 70.00% in terms of average Dice score, with 7.00% improvement than RP, while $CM + FP$ achieved 3.00%. A significant improvement is achieved when completing the network with multi-attention aware correlation, achieving 12.00% higher global Dice score than RP, compared with the improvement achieved by $CM + MP$. In addition, it is the only version among the variants which achieved performance over 70.00% in terms of DS across both the three anatomical regions, and a performance of HD95 lower than $5mm$. Furthermore, from the aspects of HD95 and NJD, the complete version outperforms other variants with significantly decreased value on each sub-region, which indicates that the attention-based module and multi-range patch size strategy are beneficial for enhancing the generalisation ability of the correlation operation.

4.4.6 Quantitative Results on Cardiac Strain Analysis

Table 4.4: Comparison of strain results derived from reference instance and the investigated methods in terms of LV SA peak radial (Peak RS) and circumferential strains (Peak CS) on testing datasets over UKBB. Asterisk indicate statistically significant differences.

	Ref	MAPC	DDIR	VM	SyN	Demons
Peak RS(%)	45.38(6.85)	43.28(7.03)	39.93(8.15)	37.16(12.33)*	33.59(11.69)	36.08(10.76)
Peak CS(%)	-24.06(2.52)	-22.46(2.47)	-18.37(2.92)	-19.07(5.07)	-15.53(6.22)*	-16.29(5.35)*

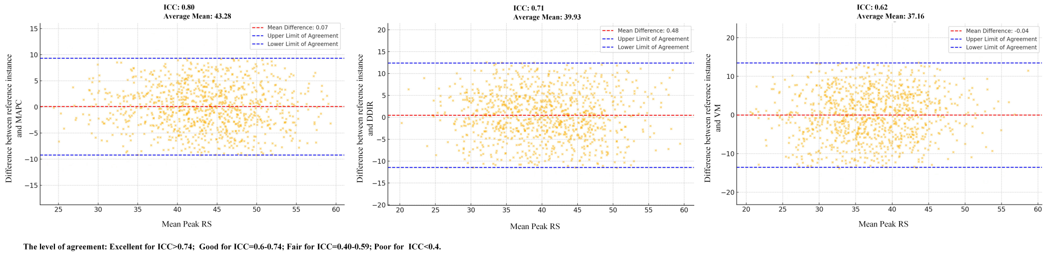


Figure 4.7: Bland Altman plots between reference instance and the investigated competitive methods for peak global radial strain (%). The red dashed line denotes the mean difference between methods, and the blue dashed lines present the limits of agreement(95% confidence).

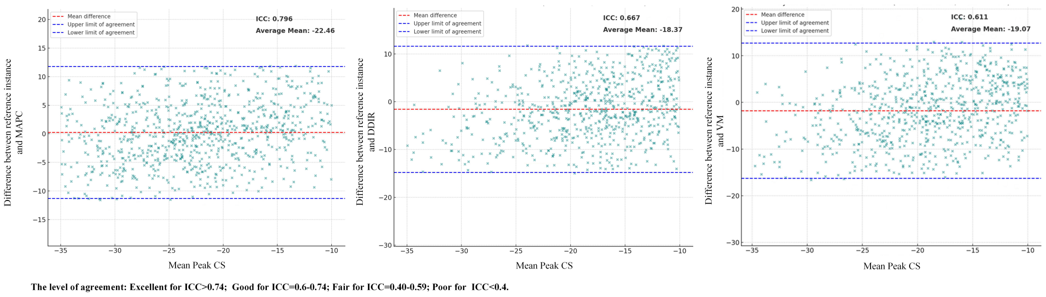


Figure 4.8: Bland Altman plots between reference instance and the investigated competitive methods for peak global circumferential strain (%). The red dashed line denotes the mean difference between methods, and the blue dashed lines present the limits of agreement(95% confidence).

In this section, strain analysis is conducted utilising the proposed method and investigated methods in terms of peak short-axis LV strain values based on the UKBB data set. Leveraging the global instance references as the ground truth. Both the global instance references and the investigated methods calculated the strain value measure the registration from ED time frame to ES utilising the segmentation of ES and estimated motion.

LV SA circumferential (Ecc, SA) and radial(Err, SA) strains were calculated from three mid-ventricular SA slices, which are determined by clinical practices. For the quantification of LV strain on ES, Table 4.4 demonstrates the mean value, and standard deviation of peak LV strain values in terms of the two metrics using the investigated methods, and the results were compared using a Welch's t-test(significant differences reported for p-value<0.05, with Bonferroni's correction).

Table 4.4 indicates the average peak strain values of 1000 patients using investigated methods. The mean value and an unpaired sample t-test showed that the proposed method was more successful than other methods in detecting a statistically significant in terms of both the two different strain metrics, compared to the reference instance. As shown in Figure. 4.7 and Figure. 4.8, compared with other investigated methods, the results demonstrate that our method achieved excellent agreement, compared with reference instance, with the lowest variation of agreement limit in terms of radial and circumferential, than the other investigated methods with larger range of agreements. Also, the Intraclass correlation coefficients (ICC) value also suggested the outperforming which is an additional evidence to prove that MAPC-Net can benefit the downstream analysis with the estimated motion field.

4.5 Conclusion

In this paper, we proposed MAPC-Net for cardiac motion estimation, which is an attention-based pyramid network which conducts attention-guided correlation mechanism and mapping operation for different scales from coarse to fine. Fine details of deformation are further extracted through the capturing of contextual information embedded in both local subjects and the corresponding long-distance dependencies. We compensated the attention-based correlation with several conventional correlation layers in a normalization module to form a fusion learning. Experiments show that the proposed method achieves the state-of-art performance in terms of investigated evaluation metrics and downstream analysis, for CMR motion estimation without utilising segmentation as a condition.

Chapter 5

Dynamic Accelerated Cardiac CINE MRI Reconstruction Based on Motion-Compensation

In recent years, compressed sensing has developed as a promising solution for accelerating MRI data acquisition. Nevertheless, balancing between reconstructing high-quality images from undersampled MRI data and saving scan time is still a challenge due to violation of the Nyquist-Shannon sampling principle, further, it is a highly ill-posed problem. In particular, considering the motion information between temporal frames can be utilised to explore spatial-temporal resolution during cardiac CINE MR reconstruction, an effective approach for highly accelerated MRI based on motion compensation has emerged as a popular and promising research direction.

5.1 Introduction

Magnetic resonance (MR) CINE imaging has long been at the forefront of medical imaging. Cardiac MR (CMR), in particular, offers outstanding potential to visualise the human heart's anatomical structures and function. In addition, for the diagnosis of cardiovascular disease, CMR is beneficial for downstream quantitative analysis. To facilitate precise CMR acquisition, establishing a high-quality reconstruction along both spatial and temporal dimensions is the core problem. For general reconstruction in the spatial domain as the primary stage, the essential problem is the balance between reducing the number of necessary measurements to save acquisition time and the increasing reconstruction difficulty as a result of the sensitivity to active noise responsible for the aggressive image degradation. The robustness of the de-aliasing (blurring effects or geometric distortions) ability of the algorithm and the acceleration rate for undersampling should be positively correlated. Despite the work of Chapter 3, the problem is addressed considering a single time frame, however, clinical practice normally requires a high spatio-temporal resolution across the complete cardiac cycle. The dynamic reconstruction of CMR measurements is crucial for reducing the effect on patient non-compliance due to the requirement of breath holds, and it is achieved by leveraging the motion information between different temporal frames. Based on the above observations, a main motivation in this work is to address the question - How to facilitate efficient embedding of physic-based information during image acquisition into recovering high-quality reconstructions whilst decreasing the long data measurement time?

As the footstone of reconstruction theory, the solutions to improve static reconstruction quality can be summarised into two categories: the first is hardware-based, and another, software-based attempts. At the present stage, the limited optimisation through enhancing or alternating partial mechanical conduction of the scanner can not match the needs of clinical practice, furthermore, there is no significant progress purely relying on optimising hardware can be the reference to indicate the solution for addressing the balance problem [190, 191]. In contrast, Compressed Sensing MRI (CS-MRI)[17, 192] is introduced as a solution with significant potential to effectively reduce the scanning time during signal measurement. Considering processing the reconstruction procedure as an inverse problem, to minimise the distance between the originally acquired k-space data and the target fully sampled data in image domain by transform with Fourier transform, under a sparse representation priority.

The second stage is the integration of spatio-temporal reconstruction in a pipeline of dynamic process, this research direction has received increasing attention recently. Notwithstanding that the literature of Chapter 3 has solid evidence that the approaches based on CS for CMR reconstruction have achieved convincing performance. Based on the progress, the motion patterns in accelerated acquisitions for dynamic spatio-temporal CMR reconstruction[96, 105] becomes a mainstream research direction and has been explored as potential and valuable terms for performance-enhancing in terms of reconstruction quality, especially utilising the correction of estimated motion fields from accelerated data in parallel. Contained temporal resolution, motion information can be integrated into dynamic CMR reconstruction[193–195]. Implicit motion embedding is leveraged for exploring the improvement of reconstruction quality with temporal information in the works of [193, 195]. As a contrast, in [194], motion information is performed as an augmentation module combined with a data-consistency block, while it is applied based on a cascade workflow. However, artefacts-affected errors could be propagated through reconstruction, thus image degradation would limit the accuracy of estimated motion fields. The impact and solutions of motion estimation under aggressive accelerated priority have not been fully explored yet. How to fuse the unsupervised motion estimation task into the joint optimisation problem from undersampled k-space data is still a challenge [118, 119]. In addition, previous studies focus on reconstructing based on pair-wise frames, while contextual motion information with spatial-temporal redundancy is ignored, utilising deformation fields from a time frame to the remaining time frames of a complete cardiac cycle can introduce a promising improvement to the performance, which is valuable for attention.

In this work, to achieve high-quality CMR reconstruction, based on extending the static reconstruction to dynamic reconstruction leveraging adequate spatio-temporal information, we address the previous limitations by proposing an explicit motion compensation mechanism which enables joint and alternating optimisation between reconstruction and motion estimation in a dynamic and unrolled framework.

Inspired by the conventional concept of explicit motion compensation reconstruction (EMCR) for CMR, a novel multi-attention-based motion residual learning network can conduct rapid and accurate registration along the unrolled reconstruction process, thus an attention learning and reconstruction-based and unrolled motion-compensation CMR reconstruction framework for high-quality dynamic reconstruction (DRMC). Our contributions can be summarized as follows:

1. A joint optimisation framework: we propose a DL-based framework for dynamic CMR reconstruction with motion compensation by conducting alternative optimisation between the joint tasks. In particular, by establishing the integration of the motion estimation from undersampled k-space data with motion-guided compressed sensing problem, the spatio-temporal redundancy can be leveraged in the dynamic reconstruction process, contrast with most of the previous methods require an additional branch for pre-estimation for motion field from fully sampled images, which is performed separately.

2. An accurate and efficient motion estimation network for achieving registration from one moving image to all target frames of the entire cardiac cycle. In particular, the residual learning-based attention mechanism and the temporal contextual module are proposed regarding to improve the accuracy and enhance the robustness of motion estimation from highly accelerated data.

3. The experiment results indicate that fusing physical motion fields into the CS-based reconstruction to form the dynamic learning procedure improves the performance in terms of accelerated k-space data, and the reconstructed quality can persist in high-level performance compared to fully sampled data especially higher and aggressive acceleration factors are applied only if motion is effectively embedded in the iterative model.

4. We address the multi-purpose properties and generalisation capabilities of the proposed method in terms of clinical relevance. The downstream applications

include cardiac key indices volumetric analysis, and cardiac strain curve analysis.

5.2 Problem formulation

5.2.1 General Compressed Sensing-based MR Reconstruction

Assuming $\mathbf{x} = \left[x^{(1)}, \ldots, x^{(N)}\right]$ denotes the input CMR sequence, which consists of N temporal frames. $x^{(n)} \in \mathbb{C}^M$ represents the frames of the sequence contains M pixels in its 3 dimensions. Let $\mathbf{y} = \left[y^{(1)}, \ldots, y^{(N)}\right]$ represents the undersampled k-space data which utilising Fourier Transform with the matched image domain. Although, $y^{(n)} \in \mathbb{C}^M$. The compressed sensing-based CMR reconstruction task can be expressed as the following inverse problem:

$$\min_{x} \sum_{n=1}^{N} \left\| \mathbf{A}^{(n)} x^{(n)} - y^{(n)} \right\|_2^2 \tag{5.1}$$

where $\mathbf{A}^{(n)}$ denotes the encoding procedure, which can be formulated as $\mathbf{A}^{(n)} = \mathbf{D}^{(n)}\mathbf{F}$, it concludes F, the Fourier transform matrix for transferring to the corresponding k-space, and D denotes the undersampling operator. The Equation. 5.1 can be solved by the reconstruction algorithm proposed by our work in Chapter 3. However, this approach is designed for pure reconstruction instead of considering the integration of spatio-temporal information.

5.2.2 Motion-compensated MR Reconstruction

From the view of joint optimisation for multi-tasks, which is based on utilising temporal redundancy with cardiac deformation fields between different time frames to form a resolution improvement mechanism with spatial domain, in this study, for the primary stage, the unrolled mechanism requires initial reconstructed results as the input for the following iterative joint tasks, here we perform DAPGAN from Chapter 3. For the following reconstruction, complementary information from temporal frames is embedded into the reconstruction forward process, here we follow the principle proposed in the work[96], known as conjugate-gradient SENSE (CG-SENSE).

For a conventional format, the optimisation procedure is conducted by applying the temporal redundancy on static CMR reconstruction algorithm, Equation. 5.1 can be reformulated as an explicit motion compensated reconstruction (EMCR) problem.

However, it should be noticed that only the regularisation term of reconstruction matrix $\mathbf{x^{(n_2)}}$ optimised in the iterations:

$$\min_{x} \sum_{n_1=1}^{1} \sum_{n_2=2}^{N} \left\| \mathbf{A}^{(n_2)} \mathbf{W} \left(u^{(n_1 \to n_2)}, x^{(n_1)} \right) - y^{(n_2)} \right\|_2^2 , \tag{5.2}$$

The estimated motion embedding learning of spatial-temporal redundancies is achieved by alternating the reconstructed image with the warped time frame image, denoted by $\mathbf{x^{(n_2)}}$, which is moved by the optical flow field $\mathbf{u^{(n_1 \to n_2)}}$ from the moving frame $\mathbf{x^{(n_1)}}$. In which, the W denotes the warping operation which takes ED as the moving frame, the remaining time frames of a complete cardiac cycle as fixed frame to predict the corresponding motion fields, which will be elaborated in Section 5.3.2.

The mechanism adopted by conventional EMCR theory can be categorised as conditional guided reconstruction. In this workflow, the motion field is estimated to take fully sampled data as a reference to complete the registration in advance of reconstruction, performed as a fixed matrix excluded from optimisation terms.

5.3 Methods

In this work, we unroll the conventional EMCR problem in Equation. 5.2 with a semi-supervised motion estimation network and perform the joint optimisation between the reconstruction task and motion estimation task.

For the EMCR problem, the temporal frames utilised as input to the motion estimation network are either raw accelerated data or intermediate predictions during reconstruction which suffer from severe aliasing artefacts. The noise-affected deformation fields can heavily limit the reconstructed results for the downstream dynamic reconstruction. Considering the above factors, the proposed MCMR performing a joint optimisation is specifically designed for solving the EMCR problem:

$$\min_{x,u} \sum_{n_1=1}^{1} \sum_{n_2=2}^{N} \left\| \mathbf{A}^{(n_2)} \mathbf{W} \left(u^{(n_1 \to n_2)}, x^{(n_1)} \right) - y^{(n_2)} \right\|_2^2 , \tag{5.3}$$

Here, the temporal resolution from moving image $\mathbf{x^{(n_1)}}$ to the fixed image is indicated by the motion $\mathbf{u^{(n_1 \to n_2)}}$, which will be optimised combined with fixed image $\mathbf{x^{(n_2)}}$ as two joint regularisation terms.

5.3.1 Dynamic CMR Reconstruction with motion compensation

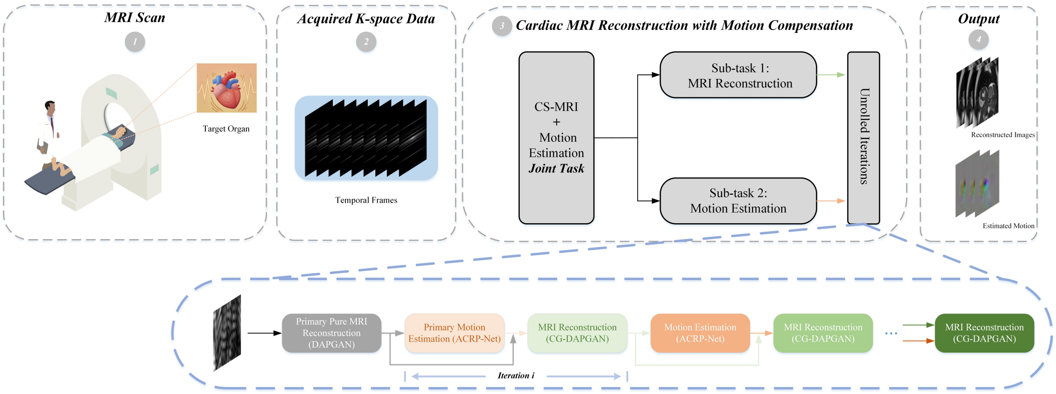


Figure 5.1: The framework of the proposed dynamic Cardiac MRI reconstruction with motion compensation. Two sub-tasks are applied and driven by a dynamic joint optimisation between reconstruction and motion estimation. The primary reconstruction is performed without motion compensation, and the following image reconstruction using the CG for embedding the prediction of registration. The framework is applied for a complete cardiac cycle along the temporal resolution.

The workflow of the proposed dynamic motion compensated reconstruction for CMR acceleration is illustrated in Fig. 5.1. At the initial stage, it allows for faster acquisition of the raw data with extreme undersampling rates, while the images conclude the complete time frames along the temporal resolution of k-space (Stage 1,2). Via Inverse Fourier Transform (IFT), the zero-filled undersampled data in the image domain is obtained, as the input for the framework of the EMCR joint model, the architecture of the joint model follows an unrolled mechanism (stage 3). At the output stage, both reconstructed images with accurate anatomical structures and the corresponding motion fields simulated based on optical flows are obtained (stage 4).

At stage 3, as the core step of the proposed workflow (see Fig. 5.1(3)), the undersampled image sequence is first fed to the proposed DAPGAN which conducts a pure reconstruction without motion embedding. The initial stage is utilised for suppressing the artefacts primarily. The following steps perform the iterative learning, which enables an alternative performing between motion estimation using the intermediate reconstructed frames as input to comprise the estimated motion fields to apply temporal redundancy during reconstruction, and dynamic reconstruction using the conjugate-gradient-based DAPGAN in each iteration. Skip connection is leveraged for concatenation sptio-temporal resolution for CG reconstruction.

To solve the dynamic compressed sensing problem with motion compensation depicted in Eq. 5.3, the proposed DRMC-Net is conducted with two intertwined terms including an unrolled motion estimation network and motion compensated reconstruction network. Based on this, an alternating optimisation is performed to address Eq. 5.3.

Before the unrolled iteration procedure, an initial step is conducted by utilising the original undersampled image x_0 concatenated with the random noise as a conditional input to perform pure reconstruction. This step aims to obtain the de-aliased result primarily thus facilitating the following unrolled motion estimation and reconstruction procedure.

Eq. 5.4 describes the optimisation for motion estimation sub-network in the unrolling mechanism, including a similarity consistency term between the warped frame and fixed frame, and motion smoothness term in spatial and temporal dimensions, respectively. Here, the similarity consistency is treated by minimising the supervised training loss Eq. 5.7 and the unsupervised training loss Eq. 5.9.

$$\mathbf{u}_i = \arg\min_u \sum_{n_1=1}^{T=1} \sum_{n_2=2}^{N} \left\| \rho \left(x_{i-1}^{(n_2)} - \mathbf{W} \left(u^{(n_1 \to n_2)}, x_{i-1}^{(n_1)} \right) \right) \right\|_1 + \alpha \mathcal{R}_s(\mathbf{u}) + \beta \mathcal{R}_t(\mathbf{u}), \quad (5.4)$$

where $\mathbf{u} = \left[u^{(1 \to 2)}, \ldots, u^{(1 \to N)} \right]$ represents the estimated motion sequence from group-wise registration, between ED frame to each frame of the remaining time points in a complete cardiac cycle. $\mathcal{R}_s$ and $\mathcal{R}_t$ denote the motion smoothness terms in spatial and temporal resolution, respectively. ρ represents the Charbonnier penalty function.

As indicated in the Eq. 5.5, for an intermediate iteration i in dynamic motion compensated MRI reconstruction which is conjugate gradient-based, an estimated motion $\mathbf{u_{i-1}^{(n_1 \to n_2)}}$ from last iteration leveraging the motion estimation network is essential, then warp the reconstructed moving frame $\mathbf{x^{(n_1)}}$ from previous iteration to fixed frame $\mathbf{x^{(n_2)}}$, based on the temporarily freezing u_{i-1}, a motion compensated data consistency term utilising $\mathbf{A}^{(n_2)}$ and $y^{(n_2)}$ indicated in Eq. 5.1 is facilitated. The ℓ_2 regularisation term in Eq. 5.5 is conducted with the reconstruction network DAPGAN parameterised with ϑ.

$$x_i = \arg\min_{x} \sum_{n_1=1}^{T=1} \sum_{n_2=2}^{N} \left\| \mathbf{A}^{(n_2)} \mathbf{W} \left(u_{i-1}^{(n_1 \rightarrow n_2)}, x^{(n_1)} \right) - y^{(n_2)} \right\|_2^2 + \gamma \left\| x - f_{cnn} \left(\boldsymbol{x}_{i-1} \mid \boldsymbol{\vartheta} \right) \right\|_2^2 . \tag{5.5}$$

In summary, the difficulty of the two joint tasks under aggressive acceleration rates is alleviated alternatively along the unrolling framework.

5.3.2 Accurate Motion Estimation from Undersampled Data

For the motion estimation based on accelerated time frames, the core challenge is how to establish an aliasing-robust network for obtaining accurate temporal information without unrealistic deformation fields. To address this, we propose an attention and contextual residual correlation-based pyramid network. It takes the artefacts-affected image sequence $x^{(n)}$ as input, achieving $1 \rightarrow N$ group-wise motion estimation from ED to the remaining frames of a cardiac cycle. In particular, to overcome the obstacles of aggressive artefacts left during the unrolled iterations of reconstruction, attention-aware correlation mapping of multi-level and residual learning blocks for contextual features of fixed frames are introduced as a complementary operation which offers enhanced robustness for registration. The proposed architecture is depicted in Fig. 5.2.

Motion estimation for complete cardiac cycle

To facilitate the group-wise registration, we select a frame at ED time point as moving image $\mathbf{x}^{(\mathbf{m})}$ and register ED to all remaining frames $\mathbf{x}^{(\mathbf{f})}$ of the entire cardiac cycle. Let $f(\cdot)$ denote the proposed motion estimation network, and Φ represents its parameters. In order to estimate the motion fields $\mathbf{u}^{(\mathbf{m} \rightarrow \mathbf{f})}$, with $m = 1$, and $1 < f < N$ where N denotes the number of frames in the cycle. The group-wise registration estimates the motion fields from frame $\mathbf{x}^{(\mathbf{m})}$ to $\mathbf{x}^{(\mathbf{f})}$ can be formulated as:

$$\mathbf{u}^{(m \rightarrow f)} = f\left(x^{(m)}, x^{(f-1)}, x^{(f)}, x^{(f+1)}; \Phi \right), \tag{5.6}$$

The temporal contextual encoder in Fig. 5.2 takes the fixed image together with its adjacent time frames as input, offering additional structural references along the temporal resolution.

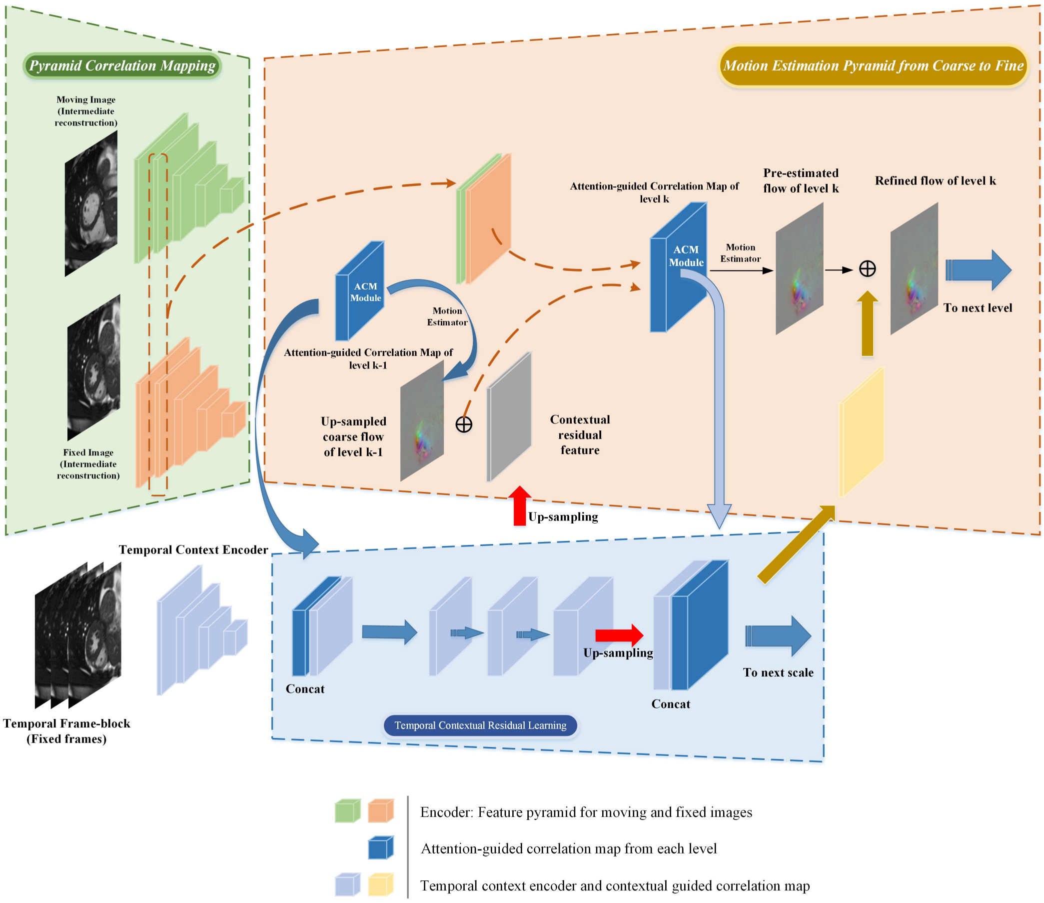


Figure 5.2: Details of residual learning-based contextual correlation framework for the refinement of motion estimation. (a) contextual residual sub-network for temporal feature learning. (b) refinement block for attention-guided correlation utilising the contextual temporal block. (c) feature pyramid.

Residual Learning-based Contextual Correlation for Motion Estimation

The proposed residual learning-based contextual correlation pyramid network is illustrated in Fig. 5.2, as the motion estimation branch in the alternating optimisation framework. As the outputs of undersampling acquisition or intermediate reconstruction, the pair-wise inputs including moving and fixed frames are fed into the motion estimation encoder with shared weights to form a feature pyramid. To achieve detailed refinement from coarse to fine, we conduct the network by two core modules: (a) Residual learning branch: which explores contextual information between adjacent temporal frames. (b) Motion estimation branch: which facilitates fusion learning by integrating residual learning branch into the attention-aware correlation pyramid for estimating an refined motion field.

As shown in the light blue block of Fig. 5.2, the residual learning sub-network is independently structured by encoding the temporal features to yield contextual features in the corresponding scale of the temporal features pyramid. In the initial stage, the temporal context encoder utilises the fixed image from level $k-1$ with the adjacent time frames as input, the encoder has five convolutional layers, with a kernel size of 3, and an encoding scale stride of 2. The target is to obtain the corresponding pyramidal contextual features which match the specific level of the pyramid encoder for moving and fixed images in level k. The remaining stages consist of (a) cascades of residual convolution stack module, as the learning procedure, and (b) concatenation operation between the output of convolution unit, multi-attention-based correlation, and the up-sampled flow field, as the fusion and refinement procedure.

The framework is a variant of the originally proposed motion estimation architecture in Chapter 4. The original input is the concatenation between the attention-guided correlation map (ACM) from the previous level and extracted contextual features from the temporal context encoder in the same scale. The second stage consists of a cascades structure by stacking three convolution blocks, in each convolution unit, we performed one layer with a kernel size of 1, one medium layer with a kernel size of 3 based on dilated convolution, the last layer with kernel size of 1, and one layer with kernel size of 1 as skip connection layer, all followed by a Batch Normalisation, and Relu activation layer except the last layer of the unit. In particular, the residual learning result from the medium layer will concatenate with the estimated motion from level $k-1$, embedded into the pair-wised moving and fixed images to generate the ACM in level k, then a transposed convolution is applied to upsample the output of the cascade residual block with the scale of 2 and fuse it with ACM feature. The last stage is a concatenation operation between the motion fields estimated from the main branch (in light orange) and residual learning branch, and then the final prediction of the motion field in level k is formed.

5.3.3 Loss Functions and Training

Global Training Loss function

The proposed network is driven by Equation 5.4, which is the core optimisation term for Equation 5.5. Due to the limitation that there is no reference CMR motion in clinical situations, cardiac motion estimation is regularly trained in an unsupervised mechanism. Nevertheless, the motion estimation task is more challenging by utilising accelerated measurements and artifact-corrupted reconstruction. To address this in the proposed unrolled joint optimisation framework, we proposed a combinational data consistency term for brightness including semi-supervised similarity penalty and unsupervised similarity penalty functions. Hence we introduce a combinational loss function as follows, which is elaborated with three parts, semi-supervised part, unsupervised part, and unsupervised smoothness loss:

$$\begin{aligned}\mathcal{L}_{\phi,k}\left(F_{k,1},F_{k,2},\hat{\mathbf{F}}_{\mathbf{k},\mathbf{1}},\hat{\mathbf{F}}_{\mathbf{k},\mathbf{2}},\hat{\mathbf{W}}_{\mathbf{k}}\right) &= \mathcal{L}^{S}_{\phi,k}\left(F_{k,1},F_{k,2},\hat{\mathbf{W}}_{\mathbf{k}}\right)\\ &+\alpha\mathcal{L}^{U}_{\phi,k}\left(\hat{\mathbf{F}}_{\mathbf{k},\mathbf{1}},\hat{\mathbf{F}}_{\mathbf{k},\mathbf{2}},\hat{\mathbf{W}}_{\mathbf{k}}\right)+\beta\mathcal{L}^{G}_{\phi,k}\left(\hat{\mathbf{W}}_{\mathbf{k}}\right)\end{aligned} \tag{5.7}$$

where ϕ represents the optimised parameters of the motion estimation network, $\mathcal{L}^{S}_{\phi,k}$, $\mathcal{L}^{U}_{\phi,k}$ and $\mathcal{L}^{G}_{\phi,k}$ denote the semi-supervised loss, unsupervised spatial loss and smoothness loss, respectively. $\hat{\mathbf{W}}_{\mathbf{k}}$ denotes the estimated motion from undersampled data in the k th level. The loss is applied for motion estimation based on multi-level with the trade-off weights.

Semi-supervised brightness consistency term

We facilitate the semi-supervision by leveraging fully-sampled fixed frame $\mathbf{x}^{(\mathbf{n_2})}_{\mathbf{gt}}$ and moving frame $\mathbf{x}^{(\mathbf{n_1})}_{\mathbf{gt}}$ combined with the estimated motion field from undersampled or artefact-corrupted reconstruction to form the warping operation and data consistency penalty.

$$\mathcal{L}_i = \sum_{n_1=1}^{T=1}\sum_{n_2=2}^{N}\left\|\rho\left(\mathbf{x}^{(\mathbf{n_2})}_{\mathbf{gt}} - \mathbf{W}\left(\mathbf{u}^{(\mathbf{n_1}\rightarrow\mathbf{n_2})}_{\mathbf{i}},\mathbf{x}^{(\mathbf{n_1})}_{\mathbf{gt}}\right)\right)\right\|_1 \tag{5.8}$$

where it should be noticed the motion u_i for warping is still estimated from the undersampled or reconstructed frames of intermediate status during the training procedure. Through this, any deformation falsely originating from aliasing-affected images is not reserved.

Unsupervised brightness consistency term

We perform the unsupervised loss by minimising the pixel-wise mean squared error between the warped (moved) frame and the fixed frames.

$$\mathcal{L}^{U}_{\phi,k,t} = \frac{1}{T}\sum_{t=2}^{T}\sum_{\mathbf{x}} \left\| F_{k,t}(\mathbf{x}) - F_{k,1}\left(\omega\left(\mathbf{x}, W_k(\mathbf{x})\right)\right)\right\|^2 , \tag{5.9}$$

where the fixed image is randomly selected from the entire cardiac cycle, utilising T represents the total amount. The registration is conducted from ED (the moving image) to the view of the selected fixed image. The time position for ED is equal to $t = 1$. While $\omega\left(\mathbf{x}, W_k(\mathbf{x})\right)$ denotes the warping function to move the frame from $F_{k,1}$ to the view of $F_{k,t}$ by predicted motion field $W_k(\mathbf{x})$. We evaluate the multi-level similarity at the values of fixed image and warped image from a time sequence.

Unsupervised Temporal Smoothness Loss

To ensure local smoothness, we penalise the gradients of the motion field by utilising an approximation of Huber loss [196], in addition, we performed a corresponding regularisation term to constrain the deformation field to behave smoothly in the temporal dimension, where $\epsilon = 0.01$. The loss functions can be expressed as follows:

$$\mathcal{L}^{G}_{\phi,k} = \alpha \underbrace{\sqrt{\epsilon + \sum_{i=u,v,q}\left(\delta_u \Delta i^2 + \delta_v \Delta i^2 + \delta_q \Delta i^2\right)}}_{\mathcal{R}_s} + \beta \underbrace{\sqrt{\epsilon + \sum_{i=u,v,q,t} \delta_t \Delta i^2}}_{\mathcal{R}_t}, \tag{5.10}$$

where $\mathcal{R}_s$ and $\mathcal{R}_t$ denote the similarity loss functions for spatial and temporal dimensions, respectively.

Total Loss

The total loss for motion estimation can be described as follows:

$$\mathcal{L}_{\text{total}} = \sum_{i=1}^{I}\eta^{I-i}\mathcal{L}_i = \sum_{i=1}^{I}\eta^{I-i}\left(\mathcal{L}^{S}_{\phi,k} + \lambda\mathcal{L}^{U}_{\phi,k,t} + \alpha\frac{1}{T}\sum_{l=2}^{T}\mathcal{H}\left(\delta_{x,y,z}\Delta_{t+l}\right) + \beta\mathcal{H}\left(\delta_t\Delta\right)\right). \tag{5.11}$$

where α and β are regularization parameters to trade off between semi-supervised motion loss, unsupervised image similarity loss, and local and temporal smoothness. η is the loss weighting factor at different iterations.

Reconstruction Loss

For the reconstruction task, we perform the training with Equation 5.5, which is established based on the original loss functions represented in the work of Chapter 3.

5.4 Experiments

5.4.1 Dataset

The training and evaluation of the proposed approach and the other investigated methods for comparison are performed on SAX-CMR images (spatial resolution at 1.8×1.8 mm^2) of UK Biobank[128] datasets. We chose images from 3000 subjects at random, and test on 25 time frames of a complete cardiac cycle. The spatial distance of pix for each image in the UKBB is $1.8 \times 1.8 \times 10mm^3$. Among these, the end-diastole (ED) time point is utilised as a moving frame, and all remaining frames include the end-systole (ES) as fixed images. Among these, 1600 subjects were selected randomly for training the proposed method, equating to 40000 image pairs (ED-to-25 time frames from the cardiac cycle). Group image pairs from the remaining 1000 subjects were applied for testing. All CMR images were cropped to a fixed size of $128 \times 128 \times 32$.

5.4.2 Implementation Details

Our approach was implemented in Pytorch (v1.11.0) and trained on an NVIDIA A100 GPU. The Adam optimiser 1e-4 was used to optimise Eq.5.5. We set the batch size to 5, and the number of unrolled iterations to $I = 8$. The iteration number value is determined based on three factors, firstly the data consistency of the sampled data in the frequency domain is pursued in the model training to achieve the goal that the image domain reconstruction is infinitely close to the ground truth. Secondly, the estimation accuracy of the deformation field, as a complementary term in the regularisation, also images the choice of the number of iterations. Finally, our model maintains high average accuracy at around $I = 8$ for various sample acceleration rate scenarios and achieves the highest performance for the joint optimisation task. The hyperparameters α, β, λ, γ for trade-off the different terms in Eq.5.11 and Eq.5.5 were set to 10, 10, 0.0001 and 0.01, respectively. For verifying the robustness of investigated methods, we designed a strategy of mixed acceleration factors with $R = 5$, $R = 10$ and $R = 15$ for undersampling the k-space data, while we test the investigated approaches on each rate individually. For the temporal axis, we utilised all 25 time frames of a complete cardiac cycle when training, and tested ED to ES for performance verifying, while for strain analysis, we tested on all frames of 25 time positions. Furthermore, the training and testing data were performed with raw k-space data without any prior reconstruction or motion correction.

5.4.3 Baseline Comparisons

We compare the proposed method to four other methods that can be used as baseline comparison from the following two perspectives: First, non-motion compensation MR reconstruction approaches in k-t domain: (1) L+S[94], based on acquired data from k-t space, utilising the incoherence between low-rank (L) and sparse (S) as dynamic components. (2) CTF-Net[85], in which the temporal information is extracted by RNN (recurrent neural network), and further utilised as a supplement to the spatial information for dynamic reconstruction. The other aspect is considering explicit motion learning embedded in their network, as the state-of-the-art: (3) GRAFT[119] apply the network without unrolling iterations, but a 4D all pairs correlation for group registration and motion refinement is valuable to compare the performance between ours. It should be noticed that GRAFT performs the task of motion estimation as a separate task in advance, and it is optimisation-free as a fixed condition. (4) The last one is the most recent approach[120] which adopted unrolled operation for optimisation to form joint optimisation of reconstruction and motion estimation (UR-MCMR).

5.4.4 Evaluation

For evaluation of the quality of the dynamic reconstruction performance quantitatively from the proposed method and the investigated baseline approaches, here we applied multiple metrics from the view of regular reconstruction and perceptual reconstruction: Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR), besides, Learned Perceptual Image Patch Similarity (LPIPS), a metric for evaluating the similarity compared to human perceptual standard. In particular, as a joint optimisation for reconstruction and motion estimation, it is necessary to evaluate the performance on motion estimation, leveraging the evaluation metrics - Dice Scores (denoted as DS) across the three anatomical cardiac regions: LVBP, LVM and RV, and Hausdorff distance (95%) (HD95), and average number of positions in the deformation fields with a negative Jacobian determinant (NJD) indicates the performance of the diffeomorphism of the motion. Higher DS values and lower HD95 and NJD values denote better motion estimation performance.

In addition, the downstream analysis includes clinical cardiac volumetric indices and strain analysis is also calculated as metrics to demonstrate the performance of the proposed method preserves clinically relevant volumetric indices post-image reconstruction, and it can also benefit the prediction of strain indices (radial and circumferential signal) as close as to those predicted from fully-sampled data, but based on moved frames after reconstruction. Different ranges of time frames are applied depending on different experiments. For the evaluation of reconstruction, we focus on moving the ED frame to the ES frame for comparison, while for strain analysis, a full cycle is demonstrated.

5.5 Results

5.5.1 Baseline comparisons

Quantitative analysis in terms of reconstruction:

Table 5.1: Quantitative comparison for reconstruction between the proposed method and investigated methods on cardiac testing datasets over UKBB in terms of Cartesian and random sampling mask with different acceleration rates.

Sampling Mask	Acceleration Rate	Metrics	L+S	CTF	GRAFT	MCMR	Proposed
Cartesian	R=5	SSIM	0.863(0.05)	0.891(0.05)	0.896(0.05)	0.915(0.04)	0.942(0.05)
		PSNR(dB)	25.196(2.21)	26.281(2.78)	30.362(3.52)	30.728(5.33)	33.539(5.17)
		LPIPS	0.07(0.05)	0.04(0.02)	0.03(0.02)	0.03(0.02)	0.02(0.01)
	R=15	SSIM	0.547(0.15)	0.791(0.07)	0.724(0.07)	0.873(0.07)	0.906(0.07)
		PSNR(dB)	15.733(2.93)	22.483(2.83)	21.619(2.57)	29.771(4.38)	29.572(4.07)
		LPIPS	0.17(0.06)	0.09(0.07)	0.10(0.06)	0.05(0.03)	0.04(0.02)
Random	R=5	SSIM	0.866(0.04)	0.879(0.04)	0.904(0.05)	0.952(0.04)	0.956(0.03)
		PSNR(dB)	25.091(2.33)	25.317(2.65)	30.478(3.37)	33.778(5.82)	35.247(4.83)
		LPIPS	0.05(0.03)	0.03(0.02)	0.04(0.02)	0.03(0.02)	0.02(0.01)
	R=10	SSIM	0.734(0.07)	0.826(0.07)	0.784(0.05)	0.896(0.06)	0.925(0.05)
		PSNR(dB)	20.635(2.79)	22.923(2.57)	25.593(2.63)	30.827(4.54)	33.427(4.53)
		LPIPS	0.11(0.05)	0.05(0.03)	0.06(0.04)	0.04(0.03)	0.04(0.02)

As shown in Table 5.1, the performance comparison of investigated methods is evaluated across two sampling masks, Cartesian and Random, respectively and two different acceleration rates for each of the masks, in terms of three quantitative metrics including SSIM, PSNR and LPIPS. The reconstruction challenges are reflected in two aspects, regular and irregular (less ACS region, introduced in Chapter 1) sampling strategies, and undersampling rates (aggressive rate = 15 was applied). When we focus on the Cartesian sampling strategy, all the methods achieve SSIM over 0.850, PSNR over 25.000, and LPIPS lower than 0.10. Among them, the proposed method achieves the best performance across all metrics. When the undersampling rate is accelerated to 15, an irregular mask is applied, where severe performance drops can be observed, compared with the performance when R =5, only the proposed method achieves SSIM higher than 0.900, only MCMR and the proposed method maintain PSNR higher than 25.000, and LPIPS lower than 0.05. Increasing from R =5 to R = 15, a reduction of ∼32.00%, ∼10.00%, ∼17.00%, ∼4.20% are demonstrated by L+S, CTF, GRAFT and MCMR in terms of SSIM, respectively. While in contrast, the proposed method, shows the lowest reduction of ∼3.6% and exhibits superior and consistent performance. The outstanding performance is also suggested by PSNR and LPIPS regardless of the sampling strategy including Random and acceleration rates.

Quantitative analysis in terms of motion estimation:

Table 5.2: Quantitative comparison for motion estimation between the proposed method and investigated methods on cardiac testing datasets over UKBB in terms of Cartesian mask with an acceleration rate of 16.

Methods	Avg.DS(%)	DS-LVBP(%)	DS-LVM(%)	DS-RV(%)	Avg.HD95(mm)	HD95-LVBP(mm)	HD95-LVM(mm)	HD95-RV(mm)	NJD(%)
Before Reg	4.55(3.79)	6.24(5.28)	3.83(2.92)	3.59(3.16)	170.61(13.21)	152.36(9.57)	177.16(12.38)	182.32(17.67)	-
L+S	50.27(15.59)	52.16(15.67)	50.93(17.83)	47.72(13.26)	25.66(10.62)	22.54(12.93)	26.74(10.46)	27.69(8.47)	5.36(0.83)
CTF-Net	62.12(17.56)	68.36(18.49)	56.18(16.38)	61.82(17.82)	9.60(6.08)	7.23(6.24)	10.36(5.38)	11.21(6.62)	3.75(0.89)
GRAFT	60.55(15.13)	68.22(16.21)	53.59(13.35)	59.83(15.83)	10.52(6.45)	8.02(4.18)	10.71(6.83)	12.82(8.35)	2.24(0.37)
MCMR	70.16(8.82)	77.42(9.27)	67.79(8.51)	65.27(8.68)	8.52(4.90)	6.83(3.28)	9.21(3.74)	9.50(7.67)	1.76(0.25)
Proposed	74.35(6.91)	81.17(7.73)	69.37(6.17)	72.52(6.83)	7.02(4.49)	5.66(2.18)	7.25(3.73)	8.16(7.57)	1.62(0.17)

In order to verify that the reconstruction results with motion correction can be beneficial for the estimation of the cardiac motion especially when it is based on accelerated data acquisition, leveraged as a self-supervision, is also practically relevant in clinical practices, we also evaluated the performance of motion estimation by comparing the results using the same methods investigated in the reconstruction comparisons in terms of DSC, HD95 and NJD between the moved mask volume of ED frame reconstructed from undersampled data and the fullsampled data as the ground truth at ES time frame (see Table 5.2).

The Baseline methods which leverage motion information as correction information achieve higher Dice scores on the registration results than those solely designed for image reconstruction(e.g. L+S, CTF-Net). The conventional method L+S shows no significant improvements over the performance before registration, which achieved ∼50.00% for average Dice. While CTF-Net significantly outperforms L+S across all metrics, it improved the performance on average Dice from 50.00% to 60.00% and represents a reduction from 26.00mm to 10.00mm in terms of HD95. As a motion compensation method, GRAFT only outperforms L+S, and it performs marginally worse than CTF-Net based on the three anatomical structures (60.55% vs. 62.12% on Dice, and 10.52mm vs. 9.60mm on HD95), based on the conducting of the reconstruction as a follow-up task separately. While MCMR and the proposed method yield performance over 70.00%, and HD95 lower than 10.00mm. Correspondingly, it is worth noting that only the proposed method achieved a Dice score over 80.00% similarity on LVBP while producing displacement fields with the lowest average NJD and significantly outperforms other methods in terms of DSC on LVBP and RV, and HD95 on LVmyo. This indicates that our proposed method produces more anatomically plausible motion fields and consistently highlights superior joint optimisation performance.

Qualitative analysis:

The comparisons between the proposed method and other investigated methods in terms of visual inspections for qualitative analysis are illustrated in Fig. 5.3 to Fig. 5.6. The reconstructed results are presented simultaneously with the moved image by leveraging the corresponding de-aliased images, the last item showcased is the estimated motion fields which demonstrate the deformation information from the moving image to the fixed image. In particular, key metrics and enlarged cardiac regions are indicated and attached to the predictions. Cartesian undersampling with two acceleration rates R = 15 and R = 5 and random undersampling with acceleration rate R = 10 and 5 are applied for each group, respectively.

Fig. 5.3 to Fig. 5.6 indicate that the proposed method achieved superior and con-

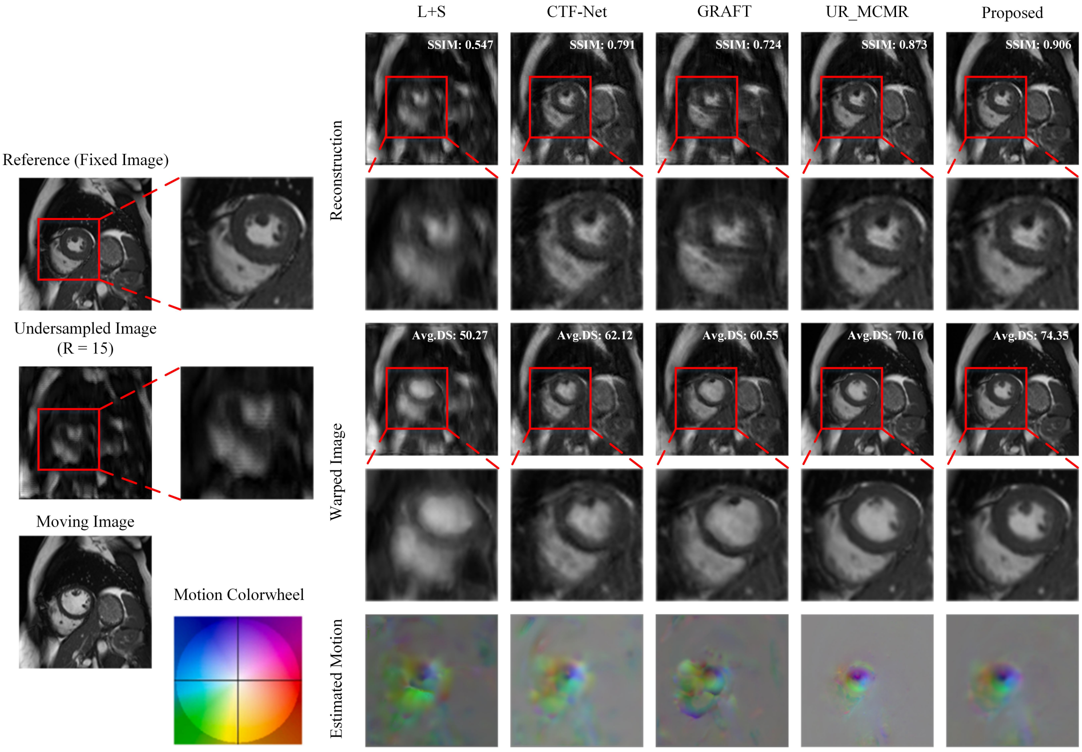


Figure 5.3: The visual inspection-based qualitative results of CMR reconstruction and motion estimation from undersampled temporal frames. The comparison is presented including the proposed method and investigated methods on the UKBB Cardiac dataset in terms of Cartesian sampling of R = 15. In the left part, the fully sampled moving image, the fully sampled fixed image (as the reference for reconstruction) and the undersampled fixed image are presented. In the right part, the first row and the second row are the reconstructed results. The third row shows the warped frames moved by the estimated motion in the last row. Motion fields are coded with colour-wheel, hue encodes orientation and saturation magnitude. The angle with the x-axis denotes the estimated direction of the motion, and the colour intensity indicates the magnitude of the deformation.

sistent performance compared with investigated baseline methods. Our method outperforms all baseline methods in both preserving higher resolution and better contrast, reflecting the high-quality reconstruction with accurate anatomical structure at all acceleration rates. Additionally, the improvement in temporal resolution consistently suggested that the proposed method performed better in agreement with the fully

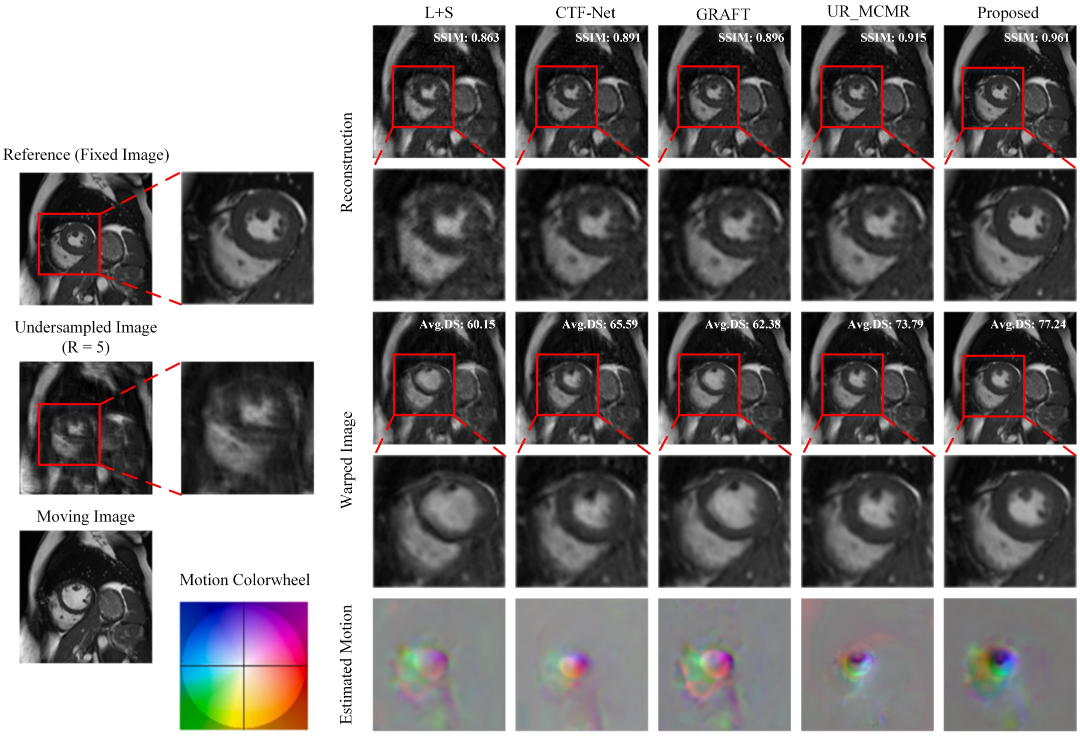


Figure 5.4: The visual inspection-based qualitative results of CMR reconstruction and motion estimation from undersampled temporal frames. The comparison is presented including the proposed method and investigated methods on the UKBB Cardiac dataset in terms of Cartesian sampling of R = 5. In the left part, the fully sampled moving image, the fully sampled fixed image (as the reference for reconstruction) and the undersampled fixed image are presented. In the right part, the first row and the second row are the reconstructed results. The third row shows the warped frames moved by the estimated motion in the last row. Motion fields are coded with colour-wheel, hue encodes orientation and saturation magnitude. The angle with the x-axis denotes the estimated direction of the motion, and the colour intensity indicates the magnitude of the deformation.

sampled reference in terms of warped images and deformation field.

In Fig. 5.3 and Fig. 5.4, the undersampled and zero-filled image with Cartesian sampling at rate R=15 represents severe blurring and streaking artefacts, whereas moderate at rate R = 5. When the sampling rate R = 15 was applied to verify the proposed method exhibits outstanding robustness in an aggressive accelerated sampling strategy, the conventional method L+S presents severe blurring and fails to obtain a reasonable anatomical texture after reconstruction. CTF-Net, as a DL-based method, while performing data acquisition in the k-t domain, appeared to slightly better improvement in artefacts removal, significantly increased quality of warped image yields a more accurate prediction over L+S, however, it still suffers from noticeably missed structures and blurring on edges for both reconstructions of fixed images and moved images. In contrast, the results of motion-compensated methods are improved with more detailed structural features. However, the reconstructed image utilising GRAFT remains low spatial resolution and it is still heavily limited by noise, inferior to the later two methods, even worse than CTF-Net. UR-MCMR recovered better anatomical structures than GRAFT and CTF-Net, but some fine features in ROI are lost, which is especially reflected in the moved frame, indicating potential perceptual texture or edge details were lost. In contrast, the proposed method can reconstruct the finest and perceptual reliable anatomical details with a high quality in similarity accuracy of moved frame compared with fixed image. Besides, despite UR-MCMR output comparable results in terms of reconstruction results for the R = 5 and the samples in the random strategy, it suffers from severe performance drops across reconstruction and motion estimation when acceleration factors increase, in addition, the sampling strategy is an irregular one. This demonstrates a more robust performance and superior generalisation of our proposed method in contrast to the baseline works.

For the comparisons of the estimated motion, as shown in Fig. 5.3 to Fig. 5.6, the proposed method outperforms the two conventional methods and the two motion-compensated methods across all undersampling rates in terms of different sampling masks. In terms of the commonality across different comparative groups, the performance of the two methods without motion field supplementation is limited to the reconstruction results, and their predicted deformation fields are not able to effectively differentiate between background noise and the heart region. As for the UR-MCMR and GRAFT, regarding the joint prediction network architecture and the specific loss terms designed for the motion estimation, they can remove most of the background noise compared to the previous two methods, but the main limiting factor exists in the predicted deformation fields are sparse due to the lack of details of the deformation properties in the different regions of the cardiac and is not smooth on the boundaries. While the usage of attention-based correlation mapping modules benefits the proposed

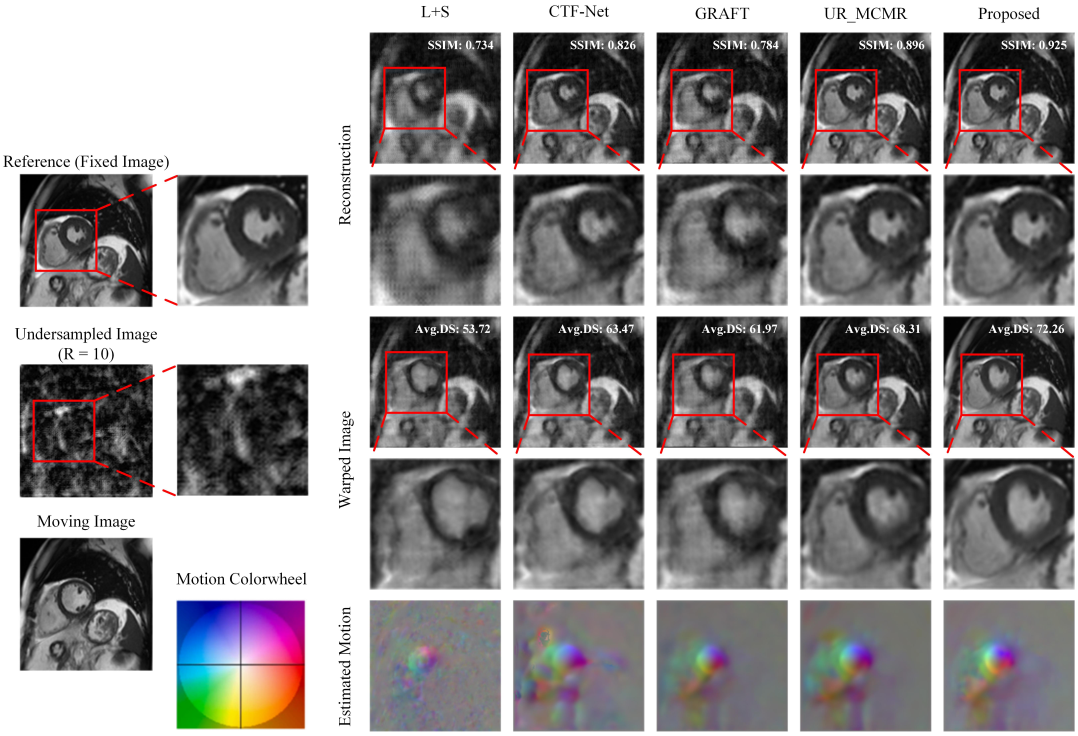


Figure 5.5: The visual inspection-based qualitative results of CMR reconstruction and motion estimation from undersampled temporal frames. The comparison is presented including the proposed method and investigated methods on the UKBB Cardiac dataset in terms of random sampling of R = 10. In the left part, the fully sampled moving image, the fully sampled fixed image (as the reference for reconstruction) and the undersampled fixed image are presented. In the right part, the first row and the second row are the reconstructed results. The third row shows the warped frames moved by the estimated motion in the last row. Motion fields are coded with colour-wheel, hue encodes orientation and saturation magnitude. The angle with the x-axis denotes the estimated direction of the motion, and the colour intensity indicates the magnitude of the deformation.

approach minimises the effect of redundant deformation fields generated by background noise and predicts the most meaningful and dense motion flow. The balanced performance between the warped image and estimated motion also verifies that our method can maintain the accuracy of the deformation field prediction in the undersampling premise.

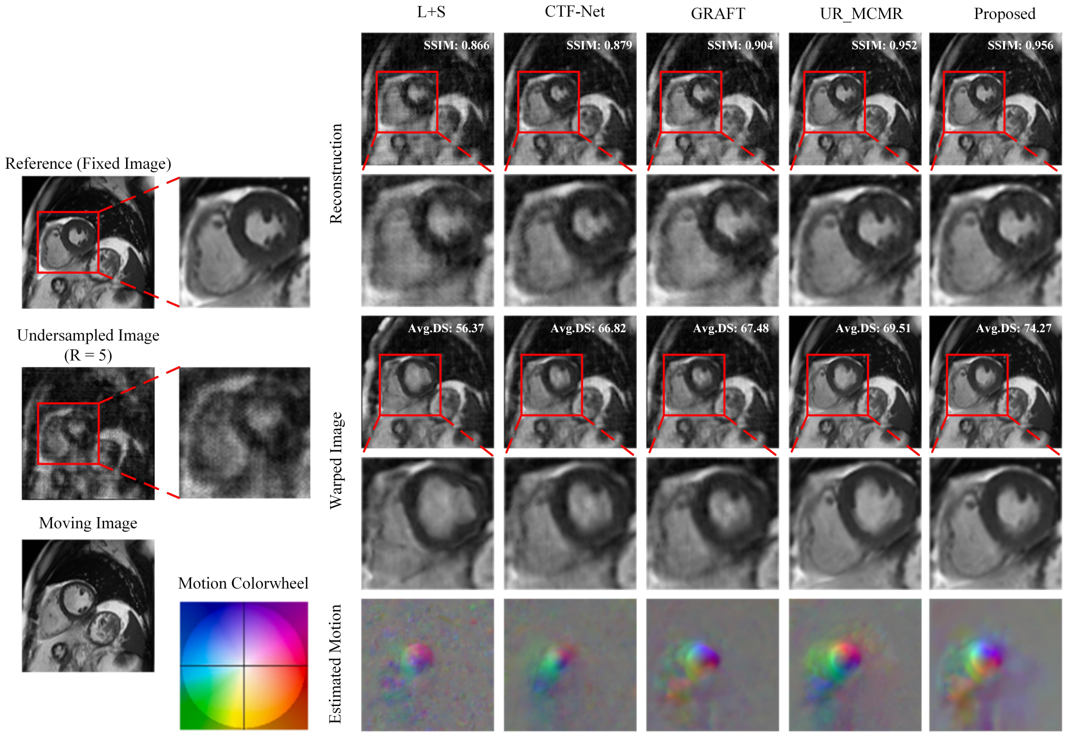


Figure 5.6: The visual inspection-based qualitative results of CMR reconstruction and motion estimation from undersampled temporal frames. The comparison is presented including the proposed method and investigated methods on the UKBB Cardiac dataset in terms of random sampling of R = 5. In the left part, the fully sampled moving image, the fully sampled fixed image (as the reference for reconstruction) and the undersampled fixed image are presented. In the right part, the first row and the second row are the reconstructed results. The third row shows the warped frames moved by the estimated motion in the last row. Motion fields are coded with colour-wheel, hue encodes orientation and saturation magnitude. The angle with the x-axis denotes the estimated direction of the motion, and the colour intensity indicates the magnitude of the deformation.

Clinical indices for cardiac quantification

Table 5.3 demonstrates the benefits of the relevant clinical functional indexes on the cardiac dataset from UKBB based on segmentations of ED and ES time frames which leverage the reconstructed results. The clinical quantified indices include the LV ED volume (LVEDV), ES volume (LVESV), LV myocardial mass (LVMM), RV ED volume

Table 5.3: Cardiac downstream quantification in terms of key clinical indices derived from reconstruction results of investigated methods

Parameters	GT	Proposed		UR-MCMR		GRAFT		CTF-Net		L+S	
	Mean (Std)	Mean (Std)	p-value	Mean (Std)	p-value	Mean (Std)	p-value	Mean (Std)	p-value	Mean (Std)	p-value
LVEDV (mL)	157.62(30.16)	155.37(31.29)	0.637	154.66(32.18)	0.136	148.25(31.85)	0.183	150.73(32.61)	0.245	145.89(31.42)	0.112
LVESV (mL)	71.39(22.62)	70.75(22.16)	0.317	66.38(22.51)	0.122	63.59(23.57)	<0.005	63.96(22.36)	0.024	53.43(22.47)	<0.005
LVMM (g)	88.79(16.81)	86.62(16.37)	0.037	81.47(16.45)	<0.005	76.33(15.79)	<0.005	75.27(16.95)	<0.005	63.55(16.18)	<0.005
RVEDV (mL)	166.57(30.16)	165.78(31.24)	0.724	161.37(32.52)	0.055	163.61(28.36)	0.232	158.51(30.67)	0.023	152.68(29.56)	0.018
RVESV (mL)	78.37(18.27)	78.12(19.61)	0.702	76.82(19.76)	0.056	74.21(17.27)	0.022	75.69(18.26)	0.053	71.28(18.56)	0.033

(RVEDV), and ES volume (RVESV). The proposed method outperformed all the investigated methods across all the clinical indices with the most comparable results yielding no significant differences, compared to the reference volumetric values derived from the fully-sampled frames, which is not achieved by any of the other methods investigated. This demonstrates the potential for applying the proposed method in real clinical downstream scenarios.

Strain analysis:

The strain values estimated with the fixed time frames selected at fixed intervals using the proposed method and the investigated methods were compared with strain analysis based on the corresponding moved time frames estimated from full-sampled images. Similar to the comparisons of warped images from ED to ES, we employ the motion estimated from fully-sampled images as our reference motions and leverage it on the downstream strain analysis, and we assume this is the most precise one that can be the role of the gold standard.

Table 5.4: Comparison of peak strain analysis derived from estimated time frames of fully-sampled and investigated methods in terms of LV short axis (SA) radial and circumferential strains. Asterisk indicate statistically significant differences.

Strains	Fully-sampled	Proposed	UR-MCMR	GRAFT	CTF-Net	L+S
LV Peak Radial, SA (%)	42.36(7.19)	40.81(7.33)	37.92(7.29)	34.20(12.37)	35.54(10.57)	24.62(11.93)*
LV Peak Circumferential, SA (%)	-21.67(2.47)	-20.36(2.83)	-17.72(5.27)	-15.17(6.18)*	-15.51(6.61)	-12.38(6.34)*

Fig. 5.7 shows the strain curve of SA circumferential and radial strain values calculated from investigated methods compared with strain values estimated from full-sampled data, which is assumed as the ground truth curve. The results show the curve

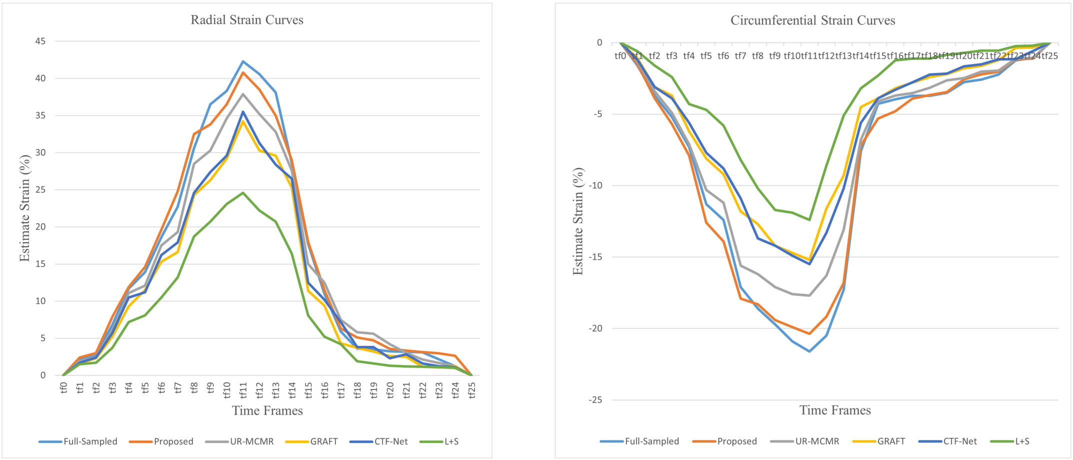


Figure 5.7: Illustration of the strain curves throughout the cardiac cycle to estimate the radial and circumferential strains

achieved by the proposed method which correlates better with the ground-truth, compared to the other investigated methods. Further, the proposed method outperforms the other methods substantially at the ES time frame, which is the peak strain value. The outstanding performance for the downstream task of the proposed method is also suggested by the results represented in Table 5.4, we evaluate the statistical significance of the investigated methods using the average peak strain values and the corresponding paired *t*-tests and found that our method showed no significant difference with the ground-truth in terms of both two strain analysis, not afforded by other approaches.

5.5.2 Ablation Study

Table 5.5: Quantitative comparison for reconstruction and motion estimation between the proposed method and investigated variants on cardiac testing datasets over UKBB in terms of Cartesian mask with an acceleration rate of 16.

Variants	SSIM	PSNR(dB)	LPIPS	Avg.DS(%)	Avg.HD95(mm)
MCMR-G	0.627(0.12)	17.382(3.17)	0.16(0.07)	52.51(15.57)	21.28(10.18)
MCMR-M	0.732(0.06)	20.523(2.65)	0.10(0.07)	60.37(16.27)	10.19(5.27)
MCMR-MT	0.815(0.07)	24.796(4.28)	0.06(0.03)	68.92(9.26)	9.57(5.24)
DRMC	0.906(0.07)	29.572(4.07)	0.04(0.02)	74.35(6.91)	7.02(4.49)

In this subsection, we will analyse quantitatively the effectiveness of the proposed architecture design of correlation mapping based on residual refinement, and the temporal context encoder.

Since the network specifically designed for motion estimation and the network employed as a sub-branch which is designed for motion compensated reconstruction are different in original inputs and the role of the global workflow, embedding the network designed especially for motion estimation directly is not available for an ideal prediction, in order to prove this and verifying the efficiency of the proposed constructing temporal context encoder and correlation residual refinement block, we performed the ablation study in this section.

The variants performed in this study for quantitative comparison can be roughly divided into two categories, the first one includes MCMR-G (reconstruction using motion estimated from general pyramidal network for compensation), MCMR-M (MCMR with MAPC-Net for motion estimation), but they are optimized by same loss function with ours (Equation. 5.5). The second category is variants MCMR-MT (MCMR with MAPC-Net and temporal context encoder), and DRMC (the proposed network for MCMR) include specific blocks for enhancing feature extraction or correlation mapping abilities by considering artefacts and aliasing affection when data were accelerated as original inputs.

We have designed a multi-level residual learning branch for finer deformation estimation. The purpose of this branch consists of residual convolution units as illustrated in Fig. 5.2 to combinational explore motion details and temporal redundancy between frames to learn fine-grained residual representation. Aggregation of an attention-based temporal context encoder provides the opportunity to capture the missing anatomical structure when the corresponding detail is aliased during reconstruction by taking the adjacent time frames as references, also it can facilitate the estimation of the deformation field when occlusion is present.

As illustrated in Table. 5.5, it presents the comparisons in terms of metrics on both reconstruction and motion estimation using a Cartesian sampling mask with an acceleration rate of 15. It can be noticed that improvements of $\sim$11% and 3dB for SSIM and PSNR can be achieved respectively when altering the general pyramidal network (MCMR-G) with MAPC-Net (MCMR-M) for motion estimation sub-task. In contrast, when the temporal context encoder was applied, MCMR-MT achieved an increase to over 0.800 and 20dB in terms of SSIM and PSNR, respectively, which indicated the necessity to conduct specific module for motion estimation with artefacts-affected data during the joint optimization. While the proposed method ACRR yields superior improvements across all the metrics, which achieved SSIM to 0.906, ∼9% higher than MCMR-MT compared with MCMR-M (∼17% vs ∼8%), and it further improved the PSNR to $\sim 30db$. It is worth noting that the proposed method is the only architecture with an average Dice score over 70, additionally, the framework significantly outperforms other investigated variants in terms of LPIPS and HD95. This indicates the merit of attention-based correlation and residual refinement modules for more accurate estimation of cardiac motion.

5.6 Conclusion

In this chapter, to obtain efficient and high-quality CMR reconstruction, we designed a novel dynamic reconstruction network with motion-compensation, named DRMC, which enables attention-aware motion estimation and is embedded into the framework of dynamic CMR reconstruction to achieve a joint optimisation between multi-tasks. We applied the unrolled optimisation combined with an end-to-end deep learning-based method. We accomplished a residual learning block by fusion the feature from attention-based correlation mapping and temporal context to refine the motion from coarse to fine, which achieved enhanced robustness when highly accelerated data was acquired. The proposed method outperforms other investigated methods in terms of quantitative metrics and visual inspections qualitatively. Furthermore, the DRMC exceeds baseline methods across all subsequent clinically relevant indices. In addition, based on the presented performance, we believe that the proposed DRMC lead to better generalisation potential in the direction of dynamic CMR reconstruction, as an efficient supplement of CVDs in future developments.

CHAPTER 6

Conclusions

6.1 Summary and Achievements

Based on the motivation of decreasing the long data-acquisition time whilst obtaining high-quality reconstructions in both spatial and temporal axes, this research sought to explore the potential of improving reconstruction quality from two different aspects: efficient supervision based on data-driven and accurate motion compensation which is physics-based. Therefore, the internal relationship of spatio-temporal resolution for cardiac cine-MRI via deep learning methods is achieved.

Rooted in sparse undersampling of k-space data, this thesis framed a unified minimisation problem which is efficiently solved by an alternating scheme between dual-domain reconstruction and accelerated motion estimation in the context of dynamic reconstruction. Three follow-up works are proposed to demonstrate the above global contribution.

Firstly, an active adaptation-guided generation mechanism is proposed for enhancing the generalisation ability for fidelity consistency, which enables emphasising small size texture that may missed in the learning process potentially and fully utilising of reference data. The proposed model, DAPGAN, demonstrates its superior performance compared to conventional and state-of-the-art methods in terms of qualitative and quantitative aspects. Moreover, the model can potentially be used to analyse the downstream task, such as volumetric indices evaluations.

Furthermore, the accurate deformation fields can be estimated through the proposed pyramid CNN, MAPC-Net, which predicts and fuses a pyramid of motion fields from multi-scales of feature presentations, in particular, attention-aware correlation mapping is integrated in the pyramidal architecture, with multi-attention normalisation applied, a refined motion field can be formed by feeding pair-wise time frames. This approach offers supreme motion estimation accuracy in an unsupervised style without any segmentation as a reference label, which demonstrates the potential to contribute to subsequent tasks for CMR such as strain assessment.

Finally, in the last work, a central question in MRI is - How to get high-quality MRI reconstructed images as close as the fully-sampled data, especially under highly undersampled factors or complex and irregular sampling strategies? Chapter 5, which established a novelty mechanism that relies on incorporating, explicitly and simultaneously in an unrolled scheme, a model for estimation of the cardiac motion from a complete cycle, compensated to the MRI reconstruction algorithm, which provides higher quality images with fewer motion artefacts and additionally, joint predicted results for multi-tasks. The potential of the proposed method has been demonstrated by exhaustive qualitative and quantitative analysis, in which we show the proposed approach outperforms the investigated conventional and state-of-the-art methods in terms of preservation of fine details, accurate edge and superior performance on reduction of blurring artefacts, in particular, the outstanding robustness of multi-tasks has also been verified by achieving the best quality of warped images and the ability of balancing the relationship between dense and smooth motion fields and most accurate moved frames.

The objective of this thesis is to establish a solid proof of concept that jointly addressing undersampled MRI reconstruction under highly accelerated scenarios and motion estimated dynamically during reconstruction intermediate status can benefit the final reconstructions. Moreover, we aim to open a new view of leveraging the reconstructed results for downstream multiple clinically relevant investigations. Thus, these contributions collectively have improved the efficiency, accuracy, and reliability of CMR dynamic reconstruction and relative analysis, directly impacting patient care and enhancing positive clinical decision-making.

6.2 Limitations and Future Research Directions

In this thesis, we focus on the intrinsic relevance of the contribution by spatio-temporal resolutions, to the learning-based dynamic reconstruction of CMR. Three connected tasks which are investigated to form the global research motivation: static MRI reconstruction, CMR motion estimation, and CMR dynamic reconstruction. While the advancements presented in this thesis are noteworthy, and promising outcomes have been achieved in the related areas, which outperforms previous studies, there still remain certain challenges as the limitations of the work in this thesis, which require further attention to facilitate more complete and superior performance for CMR reconstruction and CVDs analysis.

Opportunity always lies in challenges, the limitations can be roughly concluded in the following categories: the trade-off between data-driven and model-based learning, the dependency on the fully sampled dataset, better learning and understanding of k-space data.

In the problem of MRI reconstruction based on sparse sampling of k-space data, there are clear differences in emphasis between the two categories of methods: data-driven end-to-end methods with model-based unrolling iterative methods. Also, it is clear that deep learning has introduced new advantages for MRI reconstruction, the model-based unrolling iterative methods are based on the traditional CS model, which has advantages in terms of reconstruction speed, robustness, generalisability and interpretability. While the data-driven methods are much better at the accuracy of predictions and capturing complex non-linear relationships, as the network design can be flexibly adjusted according to the characteristics of the task. How to hybrid the two methods inferior their significant advantages is a huge challenge for the trade-off between the robustness in terms of adversarial attack with better ability for removing noise and artifacts, and the interpretability of the learning process. Despite the option of trainig, the entire system has emerged which is conducted by data-driven for individual steps followed by model-based unrolled iterations, yet it still needs further improvement, in the aspects of the way incorporating and the iteration numbers, especially in the area of dynamic reconstruction for CMR. The work in Chapter 5 is an exploration following the combinational style to address the MRI reconstruction problem.

Secondly, most of the current state-of-the-art for MRI reconstruction requires a fully sampled dataset. Some works of semi-supervised learning generally refer to the form of training that contains only part of the label data. Although we improved the generalisation ability with a more efficient way of supervision through adaptive feature guidance, however, it still relies on fully sampled data. Self-supervised learning has great potential since it does not require reference data, which is more applicable for clinical practice. It is one of the important future research directions.

Moreover, we rethink the role of k-space in the MRI workflow. In this thesis, we offer dual-domain learning for both the image domain and k-space domain, as image domain learning is better at de-aliasing, while k-space domain learning is better at keeping high-frequency information with fine structures. Future research should, therefore, include exploration of the optimal configurations for the specific frequency domain, for instance, what is the most efficient network architecture for k-space learning, and the optimal conduction for fusion learning based on data-consistency.

Another notable limitation is the gap between the ground truth of the cardiac motion field in the heart muscle between the estimated pixel-based displacement in Cine-MRI, which means the alignment might not be perfect in the preciseness. This discrepancy is an inherent limitation of pixel-based image motion estimation. Since the target of relevant models is to minimize the pixel difference between the fixed frame and the warped frame, it may fail to capture the actual or potential mechanism of cardiac motion. To address this issue, there are two potential directions for future work. The first is that mesh-based reconstruction can be established as an alternative method with more detailed cardiac anatomy, based on which the corresponding motion field can be estimated, and the second is to use semantic segmentation to enhance the supervision of motion estimation to produce more clinically relevant atlas-based motion modelling. This method bridges the gap between computational predictions and actual physiological movements, enhancing the clinical applicability of the study.

Finally, integrating or simulating the clinical feedback from the practice scenarios of the real world with the refinement of the DL-based method is essential. The feedback from clinicians can provide valuable insights that are often overlooked in data-driven situations. This can lead to the optimisation of models being developed not only utilising feature information as a condition but also based on clinically relevant facts and the trust of clinicians.

APPENDIX A

List of Publications

A.1 Journal Papers

1. Wu, K., Xia, Y., Ravikumar, N. and Frangi, A.F., 2024. Compressed sensing using a deep adaptive perceptual generative adversarial network for MRI reconstruction from undersampled K-space data. Biomedical Signal Processing and Control, 96, p.106560.

REFERENCES

[1] M. ZooFari, “Relations of the aorta, trachea, esophagus and other structures of and around the heart,” 2011. Available: https://creativecommons.org/licenses/by-sa/3.0/deed.en.

[2] P. J.Lynch, “Heart normal short axis echocardiography view,” 2006. Available: https://creativecommons.org/licenses/by/2.5/deed.en.

[3] G. Yiasemis, J.-J. Sonke, C. Sánchez, and J. Teuwen, “Recurrent variational network: a deep learning inverse problem solver applied to the task of accelerated mri reconstruction,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 732–741, 2022.

[4] E. J. Benjamin, P. Muntner, A. Alonso, M. S. Bittencourt, C. W. Callaway, A. P. Carson, A. M. Chamberlain, A. R. Chang, S. Cheng, S. R. Das, *et al.*, “Heart disease and stroke statisticsâ€”2019 update: a report from the american heart association,” *Circulation*, vol. 139, no. 10, pp. e56–e528, 2019.

[5] K. Wu, Y. Xia, N. Ravikumar, and A. F. Frangi, “Compressed sensing using a deep adaptive perceptual generative adversarial network for mri reconstruction from undersampled k-space data,” *Biomedical Signal Processing and Control*, vol. 96, p. 106560, 2024.

[6] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[7] D. Lee, J. Yoo, S. Tak, and J. C. Ye, “Deep residual learning for accelerated

mri using magnitude and phase networks," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1985–1995, 2018.

[8] Y. Han, J. Yoo, H. H. Kim, H. J. Shin, K. Sung, and J. C. Ye, "Deep learning with domain adaptation for accelerated projection-reconstruction mr," *Magnetic resonance in medicine*, vol. 80, no. 3, pp. 1189–1205, 2018.

[9] C. M. Hyun, H. P. Kim, S. M. Lee, S. Lee, and J. K. Seo, "Deep learning for undersampled mri reconstruction," *Physics in Medicine & Biology*, vol. 63, no. 13, p. 135007, 2018.

[10] D. Liang, J. Cheng, Z. Ke, and L. Ying, "Deep magnetic resonance image reconstruction: Inverse problems meet neural networks," *IEEE Signal Processing Magazine*, vol. 37, no. 1, pp. 141–151, 2020.

[11] T. Du, H. Zhang, Y. Li, S. Pickup, M. Rosen, R. Zhou, H. K. Song, and Y. Fan, "Adaptive convolutional neural networks for accelerating magnetic resonance imaging via k-space data interpolation," *Medical image analysis*, vol. 72, p. 102098, 2021.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[13] X. Li, T. Cao, Y. Tong, X. Ma, Z. Niu, and H. Guo, "Deep residual network for highly accelerated fmri reconstruction using variable density spiral trajectory," *Neurocomputing*, vol. 398, pp. 338–346, 2020.

[14] X. Xue, Y. Wang, J. Li, Z. Jiao, Z. Ren, and X. Gao, "Progressive sub-band residual-learning network for mr image super resolution," *IEEE journal of biomedical and health informatics*, vol. 24, no. 2, pp. 377–386, 2019.

[15] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[16] P. Shende, M. Pawar, and S. Kakde, "A brief review on: Mri images reconstruction using gan," in *2019 International Conference on Communication and Signal Processing (ICCSP)*, pp. 0139–0142, IEEE, 2019.

[17] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[18] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on information theory*, vol. 52, no. 2, pp. 489–509, 2006.

[19] D. Liang, B. Liu, J. Wang, and L. Ying, “Accelerating sense using compressed sensing,” *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 62, no. 6, pp. 1574–1584, 2009.

[20] S. G. Kim, L. Feng, R. Grimm, M. Freed, K. T. Block, D. K. Sodickson, L. Moy, and R. Otazo, “Influence of temporal regularization and radial undersampling factor on compressed sensing reconstruction in dynamic contrast enhanced mri of the breast,” *Journal of Magnetic Resonance Imaging*, vol. 43, no. 1, pp. 261–269, 2016.

[21] J. Royuela-del Val, L. Cordero-Grande, F. Simmross-Wattenberg, M. Martín-Fernández, and C. Alberola-López, “Jacobian weighted temporal total variation for motion compensated compressed sensing reconstruction of dynamic mri,” *Magnetic resonance in medicine*, vol. 77, no. 3, pp. 1208–1215, 2017.

[22] G. Varela-Mattatall, C. A. Baron, and R. S. Menon, “Automatic determination of the regularization weighting for wavelet-based compressed sensing mri reconstructions,” *Magnetic Resonance in Medicine*, vol. 86, no. 3, pp. 1403–1419, 2021.

[23] W. He and L. Zhao, “Mri reconstruction with separate magnitude and phase priors based on dual-tree complex wavelet transform,” *International Journal of Biomedical Imaging*, vol. 2022, no. 1, p. 7251674, 2022.

[24] M. Lustig, D. Donoho, and J. M. Pauly, “Sparse mri: The application of compressed sensing for rapid mr imaging,” *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[25] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, “Compressed sensing mri,” *IEEE signal processing magazine*, vol. 25, no. 2, pp. 72–82, 2008.

[26] D. L. Donoho and J. Tanner, “Precise undersampling theorems,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 913–924, 2010.

[27] S. Ravishankar and Y. Bresler, “Mr image reconstruction from highly undersampled k-space data by dictionary learning,” *IEEE transactions on medical imaging*, vol. 30, no. 5, pp. 1028–1041, 2010.

[28] X. Chen, Z. Du, J. Li, X. Li, and H. Zhang, “Compressed sensing based on dictionary learning for extracting impulse components,” *Signal Processing*, vol. 96, pp. 94–109, 2014.

[29] Y. Huang, J. Paisley, Q. Lin, X. Ding, X. Fu, and X.-P. Zhang, “Bayesian nonparametric dictionary learning for compressed sensing mri,” *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5007–5019, 2014.

[30] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, “Image reconstruction by domain-transform manifold learning,” *Nature*, vol. 555, no. 7697, pp. 487–492, 2018.

[31] S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng, and D. Liang, “Accelerating magnetic resonance imaging via deep learning,” in *2016 IEEE 13th international symposium on biomedical imaging (ISBI)*, pp. 514–517, IEEE, 2016.

[32] A. Kofler, M. Haltmeier, T. Schaeffter, and C. Kolbitsch, “An end-to-end-trainable iterative network architecture for accelerated radial multi-coil 2d cine mr image reconstruction,” *Medical Physics*, vol. 48, no. 5, pp. 2412–2425, 2021.

[33] X. Liu, Y. Pang, X. Sun, Y. Liu, Y. Hou, Z. Wang, and X. Li, “Image reconstruction for accelerated mr scan with faster fourier convolutional neural networks,” *IEEE Transactions on Image Processing*, 2024.

[34] A. Pal and Y. Rathi, “A review and experimental evaluation of deep learning methods for mri reconstruction,” *The journal of machine learning for biomedical imaging*, vol. 1, 2022.

[35] F. F. Zimmermann, C. Kolbitsch, P. Schuenke, and A. Kofler, “Pinqi: an end-to-end physics-informed approach to learned quantitative mri reconstruction,” *IEEE Transactions on Computational Imaging*, 2024.

[36] C. Oh, J.-Y. Chung, and Y. Han, “An end-to-end recurrent neural network for radial mr image reconstruction,” *Sensors*, vol. 22, no. 19, p. 7277, 2022.

[37] X. Zhang, L. Zhou, Y. Wan, B. W.-K. Ling, and D. Xiong, “Ttmri: Multislice texture transformer network for undersampled mri reconstruction,” *IET Image Processing*, 2024.

[38] K. Gregor and Y. LeCun, “Learning fast approximations of sparse coding,” in *Proceedings of the 27th international conference on international conference on machine learning*, pp. 399–406, 2010.

[39] H. K. Aggarwal, M. P. Mani, and M. Jacob, “Modl: Model-based deep learning architecture for inverse problems,” *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 394–405, 2018.

[40] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, “A deep cascade of convolutional neural networks for dynamic mr image reconstruction,” *IEEE transactions on medical imaging*, vol. 37, no. 2, pp. 491–503, 2017.

[41] Z. Fabian, B. Tinaz, and M. Soltanolkotabi, “Humus-net: Hybrid unrolled multiscale network architecture for accelerated mri reconstruction,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 25306–25319, 2022.

[42] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, “Learning a variational network for reconstruction of accelerated mri data,” *Magnetic resonance in medicine*, vol. 79, no. 6, pp. 3055–3071, 2018.

[43] D. Gilton, G. Ongie, and R. Willett, “Deep equilibrium architectures for inverse problems in imaging,” *IEEE Transactions on Computational Imaging*, vol. 7, pp. 1123–1133, 2021.

[44] A. Pramanik, M. B. Zimmerman, and M. Jacob, “Memory-efficient model-based deep learning with convergence and robustness guarantees,” *IEEE transactions on computational imaging*, vol. 9, pp. 260–275, 2023.

[45] F. F. Zimmermann and A. Kofler, “Nosense: Learned unrolled cardiac mri reconstruction without explicit sensitivity maps,” in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 454–466, Springer, 2023.

[46] Z. Ramzi, G. Chaithya, J.-L. Starck, and P. Ciuciu, "Nc-pdnet: A density-compensated unrolled network for 2d and 3d non-cartesian mri reconstruction," *IEEE Transactions on Medical Imaging*, vol. 41, no. 7, pp. 1625–1638, 2022.

[47] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, *et al.*, "Dagan: deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1310–1321, 2017.

[48] T. M. Quan, T. Nguyen-Duc, and W.-K. Jeong, "Compressed sensing mri reconstruction using a generative adversarial network with a cyclic loss," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1488–1497, 2018.

[49] G. Li, J. Lv, X. Tong, C. Wang, and G. Yang, "High-resolution pelvic mri reconstruction using a generative adversarial network with attention and cyclic loss," *IEEE Access*, vol. 9, pp. 105951–105964, 2021.

[50] G. Li, J. Lv, and C. Wang, "A modified generative adversarial network using spatial and channel-wise attention for cs-mri reconstruction," *IEEE Access*, vol. 9, pp. 83185–83198, 2021.

[51] R. Shaul, I. David, O. Shitrit, and T. R. Raviv, "Subsampled brain mri reconstruction by generative adversarial neural networks," *Medical Image Analysis*, vol. 65, p. 101747, 2020.

[52] H. Ali, M. R. Biswas, F. Mohsen, U. Shah, A. Alamgir, O. Mousa, and Z. Shah, "The role of generative adversarial networks in brain mri: a scoping review," *Insights into imaging*, vol. 13, no. 1, p. 98, 2022.

[53] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *The annals of Statistics*, pp. 1135–1151, 1981.

[54] C. A. Metzler, A. Mousavi, R. Heckel, and R. G. Baraniuk, "Unsupervised learning with stein's unbiased risk estimator," *arXiv preprint arXiv:1805.10531*, 2018.

[55] M. Zhussip, S. Soltanayev, and S. Y. Chun, "Training deep learning based image denoisers from undersampled measurements without ground truth and without image prior," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10255–10264, 2019.

[56] Y. Korkmaz, T. Cukur, and V. M. Patel, “Self-supervised mri reconstruction with unrolled diffusion models,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 491–501, Springer, 2023.

[57] Z.-X. Cui, C. Cao, S. Liu, Q. Zhu, J. Cheng, H. Wang, Y. Zhu, and D. Liang, “Self-score: Self-supervised learning on score-based models for mri reconstruction,” *arXiv preprint arXiv:2209.00835*, 2022.

[58] T. Yu, T. Hilbert, G. F. Piredda, A. Joseph, G. Bonanno, S. Zenkhri, P. Omoumi, M. B. Cuadra, E. J. Canales-Rodríguez, T. Kober, *et al.*, “Validation and generalizability of self-supervised image reconstruction methods for undersampled mri,” *arXiv preprint arXiv:2201.12535*, 2022.

[59] T. Xiang, M. Yurt, A. B. Syed, K. Setsompop, and A. Chaudhari, “Ddm ˆ2: Self-supervised diffusion mri denoising with generative diffusion models,” *arXiv preprint arXiv:2302.03018*, 2023.

[60] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, “Noise2noise: Learning image restoration without clean data,” *arXiv preprint arXiv:1803.04189*, 2018.

[61] P. Huang, C. Zhang, X. Zhang, X. Li, L. Dong, and L. Ying, “Self-supervised deep unrolled reconstruction using regularization by denoising,” *IEEE Transactions on Medical Imaging*, 2023.

[62] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya, “Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data,” *Magnetic resonance in medicine*, vol. 84, no. 6, pp. 3172–3191, 2020.

[63] F. Wang, H. Qi, A. De Goyeneche, R. Heckel, M. Lustig, and E. Shimron, “K-band: Self-supervised mri reconstruction via stochastic gradient descent over k-space subsets,” *arXiv preprint arXiv:2308.02958*, 2023.

[64] H. Qi, F. Wang, A. De Goyeneche, M. Lustig, and E. Shimron, “K-band: A self-supervised strategy for training deep-learning mri reconstruction networks using only limited-resolution data,” in *Proc. Intl. Soc. Mag. Reson. Med*, vol. 31, p. 1, 2023.

[65] S. Wang, R. Wu, S. Jia, A. Diakite, C. Li, Q. Liu, H. Zheng, and L. Ying, "Knowledge-driven deep learning for fast mr imaging: Undersampled mr image reconstruction from supervised to un-supervised learning," *Magnetic Resonance in Medicine*, vol. 92, no. 2, pp. 496–518, 2024.

[66] R. Souza, R. M. Lebel, and R. Frayne, "A hybrid, dual domain, cascade of convolutional neural networks for magnetic resonance image reconstruction," in *International Conference on Medical Imaging with Deep Learning*, pp. 437–446, PMLR, 2019.

[67] S. Wang, Z. Ke, H. Cheng, S. Jia, L. Ying, H. Zheng, and D. Liang, "Dimension: dynamic mr imaging with both k-space and spatial prior knowledge obtained via multi-supervised network training," *NMR in Biomedicine*, vol. 35, no. 4, p. e4131, 2022.

[68] T. Eo, Y. Jun, T. Kim, J. Jang, H.-J. Lee, and D. Hwang, "Kiki-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images," *Magnetic resonance in medicine*, vol. 80, no. 5, pp. 2188–2201, 2018.

[69] M. Ran, W. Xia, Y. Huang, Z. Lu, P. Bao, Y. Liu, H. Sun, J. Zhou, and Y. Zhang, "Md-recon-net: a parallel dual-domain convolutional neural network for compressed sensing mri," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 1, pp. 120–135, 2020.

[70] R. Souza, M. Bento, N. Nogovitsyn, K. J. Chung, W. Loos, R. M. Lebel, and R. Frayne, "Dual-domain cascade of u-nets for multi-channel magnetic resonance image reconstruction," *Magnetic resonance imaging*, vol. 71, pp. 140–153, 2020.

[71] J. Wang, Y. Yang, Y. Yang, and J. Sun, "Dual domain motion artifacts correction for mr imaging under guidance of k-space uncertainty," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 293–302, Springer, 2023.

[72] K. Han, A. Xiao, E. Wu, J. Guo, C. Xu, and Y. Wang, "Transformer in transformer," *Advances in neural information processing systems*, vol. 34, pp. 15908–15919, 2021.

[73] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, *et al.*, “A survey on vision transformer,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 1, pp. 87–110, 2022.

[74] X. Zhao, T. Yang, B. Li, and X. Zhang, “Swingan: A dual-domain swin transformer-based generative adversarial network for mri reconstruction,” *Computers in Biology and Medicine*, vol. 153, p. 106513, 2023.

[75] B. Wang, Y. Lian, X. Xiong, H. Zhou, Z. Liu, and X. Zhou, “Dct-net: Dual-domain cross-fusion transformer network for mri reconstruction,” *Magnetic Resonance Imaging*, vol. 107, pp. 69–79, 2024.

[76] H. Kim, Y. J. Yang, K. Han, P. K. Kim, B. W. Choi, J. Y. Kim, and Y. J. Suh, “Validation of a deep learning-based software for automated analysis of t2 mapping in cardiac magnetic resonance imaging,” *Quantitative Imaging in Medicine and Surgery*, vol. 13, no. 10, p. 6750, 2023.

[77] W. Xu, J. Shi, Y. Lin, C. Liu, W. Xie, H. Liu, S. Huang, D. Zhu, L. Su, Y. Huang, *et al.*, “Deep learning-based image segmentation model using an mri-based convolutional neural network for physiological evaluation of the heart,” *Frontiers in Physiology*, vol. 14, p. 1148717, 2023.

[78] Y. Wang, Y. Zhang, Z. Wen, B. Tian, E. Kao, X. Liu, W. Xuan, K. Ordovas, D. Saloner, and J. Liu, “Deep learning based fully automatic segmentation of the left ventricular endocardium and epicardium from cardiac cine mri,” *Quantitative Imaging in Medicine and Surgery*, vol. 11, no. 4, p. 1600, 2021.

[79] Y. Jun, Y. Arefeen, J. Cho, S. Fujita, X. Wang, P. E. Grant, B. Gagoski, C. Jaimes, M. S. Gee, and B. Bilgic, “Zero-deepsub: Zero-shot deep subspace reconstruction for rapid multiparametric quantitative mri using 3d-qalas,” *Magnetic Resonance in Medicine*, vol. 91, no. 6, pp. 2459–2482, 2024.

[80] C.-W. Chang, R. Marants, Y. Gao, M. Goette, J. E. Scholey, J. D. Bradley, T. Liu, J. Zhou, A. Sudhyadhom, and X. Yang, “Multimodal imaging-based material mass density estimation for proton therapy using supervised deep learning,” *The British Journal of Radiology*, vol. 96, no. 1152, p. 20220907, 2023.

[81] P. M. Johnson and M. Drangova, “Conditional generative adversarial network for 3d rigid-body motion correction in mri,” *Magnetic resonance in medicine*, vol. 82, no. 3, pp. 901–910, 2019.

[82] T. Küstner, N. Fuin, K. Hammernik, A. Bustin, H. Qi, R. Hajhosseiny, P. G. Masci, R. Neji, D. Rueckert, R. M. Botnar, *et al.*, “Cinenet: deep learning-based 3d cardiac cine mri reconstruction with multi-coil complex-valued 4d spatio-temporal convolutions,” *Scientific reports*, vol. 10, no. 1, p. 13710, 2020.

[83] C. Qin, J. Schlemper, J. Duan, G. Seegoolam, A. Price, J. Hajnal, and D. Rueckert, “k-t next: dynamic mr image reconstruction exploiting spatio-temporal correlations,” in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*, pp. 505–513, Springer, 2019.

[84] Z.-P. Liang, “Spatiotemporal imagingwith partially separable functions,” in *2007 4th IEEE international symposium on biomedical imaging: from nano to macro*, pp. 988–991, IEEE, 2007.

[85] C. Qin, J. Duan, K. Hammernik, J. Schlemper, T. Küstner, R. Botnar, C. Prieto, A. N. Price, J. V. Hajnal, and D. Rueckert, “Complementary time-frequency domain networks for dynamic parallel mr image reconstruction,” *Magnetic Resonance in Medicine*, vol. 86, no. 6, pp. 3274–3291, 2021.

[86] S. G. Lingala, Y. Hu, E. DiBella, and M. Jacob, “Accelerated dynamic mri exploiting sparsity and low-rank structure: kt slr,” *IEEE transactions on medical imaging*, vol. 30, no. 5, pp. 1042–1054, 2011.

[87] H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye, “k-t focuss: a general compressed sensing framework for high resolution dynamic mri,” *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 61, no. 1, pp. 103–116, 2009.

[88] Q. Zou, A. H. Ahmed, P. Nagpal, S. Priya, R. F. Schulte, and M. Jacob, “Joint alignment and reconstruction of multislice dynamic mri using variational manifold learning,” in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pp. 1–4, IEEE, 2022.

[89] A. Majumdar and R. K. Ward, “Exploiting rank deficiency and transform domain sparsity for mr image reconstruction,” *Magnetic resonance imaging*, vol. 30, no. 1, pp. 9–18, 2012.

[90] B. Zhao, J. P. Haldar, A. G. Christodoulou, and Z.-P. Liang, “Image reconstruction from highly undersampled (k, t)-space data with joint partial separability and sparsity constraints,” *IEEE transactions on medical imaging*, vol. 31, no. 9, pp. 1809–1820, 2012.

[91] H. Gao, S. Rapacchi, D. Wang, J. Moriarty, C. Meehan, J. Sayre, G. Laub, P. Finn, and P. Hu, “Compressed sensing using prior rank, intensity and sparsity model (prism): applications in cardiac cine mri,” in *Proceedings of the 20th Annual Meeting of ISMRM, Melbourne, Australia*, vol. 2242, 2012.

[92] V. Singh, A. H. Tewfik, and D. B. Ress, “Under-sampled functional mri using low-rank plus sparse matrix decomposition,” in *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 897–901, IEEE, 2015.

[93] R. Otazo, E. Candes, and D. K. Sodickson, “Low-rank and sparse matrix decomposition for accelerated dce-mri with background and contrast separation,” in *Proceedings of the ISMRM Workshop on Data Sampling and Image Reconstruction, Sedona*, p. 7, 2013.

[94] R. Otazo, E. Candes, and D. K. Sodickson, “Low-rank plus sparse matrix decomposition for accelerated dynamic mri with separation of background and dynamic components,” *Magnetic resonance in medicine*, vol. 73, no. 3, pp. 1125–1136, 2015.

[95] B. Trémoulhéac, N. Dikaios, D. Atkinson, and S. R. Arridge, “Dynamic mr image reconstruction–separation from undersampled (

{

\bf k}, t)−*spacevialow−rankplussparseprior,″ IEEEtransactionsonmedicalimaging, vol.* 33
−1701, 2014.

[96] P. Batchelor, D. Atkinson, P. Irarrazaval, D. Hill, J. Hajnal, and D. Larkman, “Matrix description of general motion correction applied to multishot images,” *Magnetic*

Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 54, no. 5, pp. 1273–1280, 2005.

[97] G. Cruz, D. Atkinson, M. Henningsson, R. M. Botnar, and C. Prieto, "Highly efficient nonrigid motion-corrected 3d whole-heart coronary vessel wall imaging," *Magnetic resonance in medicine*, vol. 77, no. 5, pp. 1894–1908, 2017.

[98] X. Chen, M. Salerno, Y. Yang, and F. H. Epstein, "Motion-compensated compressed sensing for dynamic contrast-enhanced mri using regional spatiotemporal sparsity and region tracking: Block low-rank sparsity with motion-guidance (blosm)," *Magnetic resonance in medicine*, vol. 72, no. 4, pp. 1028–1038, 2014.

[99] A. Bustin, I. Rashid, G. Cruz, R. Hajhosseiny, T. Correia, R. Neji, R. Rajani, T. F. Ismail, R. M. Botnar, and C. Prieto, "3d whole-heart isotropic sub-millimeter resolution coronary magnetic resonance angiography with non-rigid motion-compensated prost," *Journal of Cardiovascular Magnetic Resonance*, vol. 22, no. 1, p. 24, 2020.

[100] C. M. Rank, T. Heußer, M. T. Buzan, A. Wetscherek, M. T. Freitag, J. Dinkel, and M. Kachelrieß, "4d respiratory motion-compensated image reconstruction of free-breathing radial mr data with very high undersampling," *Magnetic resonance in medicine*, vol. 77, no. 3, pp. 1170–1183, 2017.

[101] E. Martín-González, T. Sevilla, A. Revilla-Orodea, P. Casaseca-de-la Higuera, and C. Alberola-López, "Groupwise non-rigid registration with deep learning: an affordable solution applied to 2d cardiac cine mri reconstruction," *Entropy*, vol. 22, no. 6, p. 687, 2020.

[102] H. Qi, R. Hajhosseiny, G. Cruz, T. Kuestner, K. Kunze, R. Neji, R. Botnar, and C. Prieto, "End-to-end deep learning nonrigid motion-corrected reconstruction for highly accelerated free-breathing coronary mra," *Magnetic Resonance in Medicine*, vol. 86, no. 4, pp. 1983–1996, 2021.

[103] M. Usman, D. Atkinson, F. Odille, C. Kolbitsch, G. Vaillant, T. Schaeffter, P. G. Batchelor, and C. Prieto, "Motion corrected compressed sensing for free-breathing dynamic cardiac mri," *Magnetic resonance in medicine*, vol. 70, no. 2, pp. 504–516, 2013.

[104] G. Cruz, K. Hammernik, T. Kuestner, C. Velasco, A. Hua, T. F. Ismail, D. Rueckert, R. M. Botnar, and C. Prieto, "Single-heartbeat cardiac cine imaging via jointly regu-

larized nonrigid motion-corrected reconstruction," *NMR in Biomedicine*, vol. 36, no. 9, p. e4942, 2023.

[105] Ç. Bilen, I. Selesnick, Y. Wang, R. Otazo, and D. K. Sodickson, "A motion compensating prior for dynamic mri reconstruction using combination of compressed sensing and parallel imaging," in *2011 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pp. 1–6, IEEE, 2011.

[106] F. Odille, P.-A. Vuissoz, P.-Y. Marie, and J. Felblinger, "Generalized reconstruction by inversion of coupled systems (grics) applied to free-breathing mri," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 60, no. 1, pp. 146–157, 2008.

[107] F. Odille, A. Menini, J.-M. Escanyé, P.-A. Vuissoz, P.-Y. Marie, M. Beaumont, and J. Felblinger, "Joint reconstruction of multiple images and motion in mri: Application to free-breathing myocardial

{

\rm T} _ {2} *quantification,'' IEEEtransactionsonmedicalimaging, vol.* 35, *no.* 1, *pp.* 197−
−207, 2015.

[108] J. Royuela-del Val, L. Cordero-Grande, F. Simmross-Wattenberg, M. Martín-Fernández, and C. Alberola-López, "Nonrigid groupwise registration for motion estimation and compensation in compressed sensing reconstruction of breath-hold cardiac cine mri," *Magnetic resonance in medicine*, vol. 75, no. 4, pp. 1525–1536, 2016.

[109] N. Zhao, D. Oâ€™Connor, A. Basarab, D. Ruan, and K. Sheng, "Motion compensated dynamic mri reconstruction with local affine optical flow estimation," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 11, pp. 3050–3059, 2019.

[110] T. Schmoderer, A. I. Aviles-Rivero, V. Corona, N. Debroux, and C.-B. Schönlieb, "Learning optical flow for fast mri reconstruction," *Inverse Problems*, vol. 37, no. 9, p. 095007, 2021.

[111] M. Burger, H. Dirks, and C.-B. Schonlieb, "A variational model for joint motion estimation and image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 11, no. 1, pp. 94–128, 2018.

[112] A. I. Aviles-Rivero, N. Debroux, G. Williams, M. J. Graves, and C.-B. Schönlieb, "Compressed sensing plus motion (cs+ m): a new perspective for improving undersampled mr image reconstruction," *Medical Image Analysis*, vol. 68, p. 101933, 2021.

[113] V. Corona, A. I. Aviles-Rivero, N. Debroux, M. Graves, C. Le Guyader, C.-B. Schönlieb, and G. Williams, "Multi-tasking to correct: Motion-compensated mri via joint reconstruction and registration," in *Scale Space and Variational Methods in Computer Vision: 7th International Conference, SSVM 2019, Hofgeismar, Germany, June 30–July 4, 2019, Proceedings 7*, pp. 263–274, Springer, 2019.

[114] J. Yang, T. Küstner, P. Hu, P. Liò, and H. Qi, "End-to-end deep learning of non-rigid groupwise registration and reconstruction of dynamic mri," *Frontiers in cardiovascular medicine*, vol. 9, p. 880186, 2022.

[115] T. Küstner, J. Pan, H. Qi, G. Cruz, C. Gilliam, T. Blu, B. Yang, S. Gatidis, R. Botnar, and C. Prieto, "Lapnet: Non-rigid registration derived in k-space for magnetic resonance imaging," *IEEE transactions on medical imaging*, vol. 40, no. 12, pp. 3686–3697, 2021.

[116] J. Liu, A. I. Aviles-Rivero, H. Ji, and C.-B. Schönlieb, "Rethinking medical image reconstruction via shape prior, going deeper and faster: Deep joint indirect registration and reconstruction," *Medical Image Analysis*, vol. 68, p. 101930, 2021.

[117] Q. Huang, Y. Xian, D. Yang, H. Qu, J. Yi, P. Wu, and D. N. Metaxas, "Dynamic mri reconstruction with end-to-end motion-guided network," *Medical Image Analysis*, vol. 68, p. 101901, 2021.

[118] J. Pan, D. Rueckert, T. Küstner, and K. Hammernik, "Efficient image registration network for non-rigid cardiac motion estimation," in *Machine Learning for Medical Image Reconstruction: 4th International Workshop, MLMIR 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings 4*, pp. 14–24, Springer, 2021.

[119] K. Hammernik, J. Pan, D. Rueckert, and T. Küstner, "Motion-guided physics-based learning for cardiac mri reconstruction," in *2021 55th Asilomar Conference on Signals, Systems, and Computers*, pp. 900–907, IEEE, 2021.

[120] J. Pan, D. Rueckert, T. Küstner, and K. Hammernik, “Learning-based and unrolled motion-compensated reconstruction for cardiac mr cine imaging,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 686–696, Springer, 2022.

[121] J. Pan, W. Huang, D. Rueckert, T. Küstner, and K. Hammernik, “Reconstruction-driven motion estimation for motion-compensated mr cine imaging,” *IEEE Transactions on Medical Imaging*, 2024.

[122] J. Pan, M. Hamdi, W. Huang, K. Hammernik, T. Kuestner, and D. Rueckert, “Unrolled and rapid motion-compensated reconstruction for cardiac cine mri,” *Medical Image Analysis*, vol. 91, p. 103017, 2024.

[123] K. P. Pruessmann, M. Weiger, P. Börnert, and P. Boesiger, “Advances in sensitivity encoding with arbitrary k-space trajectories,” *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 46, no. 4, pp. 638–651, 2001.

[124] S. E. Petersen, P. M. Matthews, F. Bamberg, D. A. Bluemke, J. M. Francis, M. G. Friedrich, P. Leeson, E. Nagel, S. Plein, F. E. Rademakers, *et al.*, “Imaging in population science: cardiovascular magnetic resonance in 100,000 participants of uk biobank-rationale, challenges and approaches,” *Journal of Cardiovascular Magnetic Resonance*, vol. 15, pp. 1–10, 2013.

[125] A. Fry, T. J. Littlejohns, C. Sudlow, N. Doherty, L. Adamska, T. Sprosen, R. Collins, and N. E. Allen, “Comparison of sociodemographic and health-related characteristics of uk biobank participants with those of the general population,” *American journal of epidemiology*, vol. 186, no. 9, pp. 1026–1034, 2017.

[126] T. J. Littlejohns, J. Holliday, L. M. Gibson, S. Garratt, N. Oesingmann, F. Alfaro-Almagro, J. D. Bell, C. Boultwood, R. Collins, M. C. Conroy, *et al.*, “The uk biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions,” *Nature communications*, vol. 11, no. 1, p. 2624, 2020.

[127] S. E. Petersen, N. Aung, M. M. Sanghvi, F. Zemrak, K. Fung, J. M. Paiva, J. M. Francis, M. Y. Khanji, E. Lukaschuk, A. M. Lee, *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*, vol. 19, no. 1, p. 18, 2016.

[128] S. E. Petersen, P. M. Matthews, J. M. Francis, M. D. Robson, F. Zemrak, R. Boubertakh, A. A. Young, S. Hudson, P. Weale, S. Garratt, *et al.*, "Uk biobank's cardiovascular magnetic resonance protocol," *Journal of cardiovascular magnetic resonance*, vol. 18, no. 1, p. 8, 2016.

[129] Y. Chen, W. Xie, J. Zhang, H. Qiu, D. Zeng, Y. Shi, H. Yuan, J. Zhuang, Q. Jia, Y. Zhang, *et al.*, "Myocardial segmentation of cardiac mri sequences with temporal consistency for coronary artery disease diagnosis," *Frontiers in Cardiovascular Medicine*, vol. 9, p. 804442, 2022.

[130] G. Snaauw, D. Gong, G. Maicas, A. Van Den Hengel, W. J. Niessen, J. Verjans, and G. Carneiro, "End-to-end diagnosis and segmentation learning from cardiac magnetic resonance imaging," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 802–805, IEEE, 2019.

[131] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. G. Ballester, *et al.*, "Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved?," *IEEE transactions on medical imaging*, vol. 37, no. 11, pp. 2514–2525, 2018.

[132] E. Z. Chen, C. Zhang, X. Chen, Y. Liu, T. Chen, and S. Sun, "Computationally efficient 3d mri reconstruction with adaptive mlp," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 195–205, Springer, 2023.

[133] B. Ozturkler, A. Sahiner, T. Ergen, A. D. Desai, C. M. Sandino, S. Vasanawala, J. M. Pauly, M. Mardani, and M. Pilanci, "Gleam: greedy learning for large-scale accelerated mri reconstruction," *arXiv preprint arXiv:2207.08393*, 2022.

[134] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 9446–9454, 2018.

[135] J. Yoo, K. H. Jin, H. Gupta, J. Yerly, M. Stuber, and M. Unser, "Time-dependent deep image prior for dynamic mri," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3337–3348, 2021.

[136] H. Wang, T. Li, Z. Zhuang, T. Chen, H. Liang, and J. Sun, "Early stopping for deep image prior," *arXiv preprint arXiv:2112.06074*, 2021.

[137] G. Elmas, S. U. Dar, Y. Korkmaz, E. Ceyani, B. Susam, M. Ozbey, S. Avestimehr, and T. Çukur, "Federated learning of generative image priors for mri reconstruction," *IEEE Transactions on Medical Imaging*, vol. 42, no. 7, pp. 1996–2009, 2022.

[138] S. Bhadra, V. A. Kelkar, F. J. Brooks, and M. A. Anastasio, "On hallucinations in tomographic image reconstruction," *IEEE transactions on medical imaging*, vol. 40, no. 11, pp. 3249–3260, 2021.

[139] M. J. Muckley, B. Riemenschneider, A. Radmanesh, S. Kim, G. Jeong, J. Ko, Y. Jun, H. Shin, D. Hwang, M. Mostapha, *et al.*, "Results of the 2020 fastmri challenge for machine learning mr image reconstruction," *IEEE transactions on medical imaging*, vol. 40, no. 9, pp. 2306–2317, 2021.

[140] J. P. Cohen, M. Luck, and S. Honari, "Distribution matching losses can hallucinate features in medical image translation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*, pp. 529–536, Springer, 2018.

[141] Y. Chi, L. Bi, J. Kim, D. Feng, and A. Kumar, "Controlled synthesis of dermoscopic images via a new color labeled generative style transfer network to enhance melanoma segmentation," in *2018 40Th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pp. 2591–2594, IEEE, 2018.

[142] Q. Mao, H.-Y. Lee, H.-Y. Tseng, S. Ma, and M.-H. Yang, "Mode seeking generative adversarial networks for diverse image synthesis," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1429–1437, 2019.

[143] N. M. Gottschling, V. Antun, A. C. Hansen, and B. Adcock, "The troublesome kernel–on hallucinations, no free lunches and the accuracy-stability trade-off in inverse problems," *arXiv preprint arXiv:2001.01258*, 2020.

[144] M. Genzel, J. Macdonald, and M. März, "Solving inverse problems with deep neural networks–robustness included?," *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 1, pp. 1119–1134, 2022.

[145] J. Amjad, Z. Lyu, and M. R. Rodrigues, “Deep learning model-aware regulatization with applications to inverse problems,” *IEEE Transactions on Signal Processing*, vol. 69, pp. 6371–6385, 2021.

[146] Y. Huang, T. Würfl, K. Breininger, L. Liu, G. Lauritsch, and A. Maier, “Some investigations on robustness of deep learning in limited angle tomography,” in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*, pp. 145–153, Springer, 2018.

[147] V. Antun, F. Renna, C. Poon, B. Adcock, and A. C. Hansen, “On instabilities of deep learning in image reconstruction and the potential costs of ai,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30088–30095, 2020.

[148] A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. Tamir, “Robust compressed sensing mri with deep generative priors,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 14938–14954, 2021.

[149] J. N. Morshuis, S. Gatidis, M. Hein, and C. F. Baumgartner, “Adversarial robustness of mr image reconstruction under realistic perturbations,” in *International Workshop on Machine Learning for Medical Image Reconstruction*, pp. 24–33, Springer, 2022.

[150] A. Krainovic, M. Soltanolkotabi, and R. Heckel, “Learning provably robust estimators for inverse problems via jittering,” *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[151] A. Goujon, S. Neumayer, P. Bohra, S. Ducotterd, and M. Unser, “A neural-network-based convex regularizer for inverse problems,” *IEEE Transactions on Computational Imaging*, 2023.

[152] M. John, J. R. Chand, and M. Jacob, “Local monotone operator learning using non-monotone operators: Mnm-mol,” *IEEE Transactions on Computational Imaging*, 2024.

[153] F. Calivá, K. Cheng, R. Shah, and V. Pedoia, “Adversarial robust training of deep learning mri reconstruction models,” *arXiv preprint arXiv:2011.00070*, 2020.

[154] K. Cheng, F. Calivá, R. Shah, M. Han, S. Majumdar, and V. Pedoia, “Addressing the false negative problem of deep learning mri reconstruction models by adversarial

attacks and robust training," in *Medical Imaging with Deep Learning*, pp. 121–135, PMLR, 2020.

[155] D. K. Sodickson and W. J. Manning, "Simultaneous acquisition of spatial harmonics (smash): fast imaging with radiofrequency coil arrays," *Magnetic resonance in medicine*, vol. 38, no. 4, pp. 591–603, 1997.

[156] K. T. Block, M. Uecker, and J. Frahm, "Undersampled radial mri with multiple coils. iterative image reconstruction using a total variation constraint," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 57, no. 6, pp. 1086–1098, 2007.

[157] X. Fan, Q. Lian, and B. Shi, "Compressed sensing mri with phase noise disturbance based on adaptive tight frame and total variation," *IEEE Access*, vol. 5, pp. 19311–19321, 2017.

[158] F. Shi, J. Cheng, L. Wang, P.-T. Yap, and D. Shen, "Lrtv: Mr image super-resolution with low-rank and total variation regularizations," *IEEE transactions on medical imaging*, vol. 34, no. 12, pp. 2459–2466, 2015.

[159] X. Liu, J. Wang, J. Jin, M. Li, F. Tang, S. Crozier, and F. Liu, "Deep unregistered multi-contrast mri reconstruction," *Magnetic Resonance Imaging*, vol. 81, pp. 33–41, 2021.

[160] S. U. H. Dar, M. Özbey, A. B. Çatlı, and T. Çukur, "A transfer-learning approach for accelerated mri using deep neural networks," *Magnetic resonance in medicine*, vol. 84, no. 2, pp. 663–685, 2020.

[161] W. Zhou, H. Du, W. Mei, and L. Fang, "Efficient structurally-strengthened generative adversarial network for mri reconstruction," *Neurocomputing*, vol. 422, pp. 51–61, 2021.

[162] X. Liu, H. Du, J. Xu, and B. Qiu, "Dbgan: A dual-branch generative adversarial network for undersampled mri reconstruction," *Magnetic Resonance Imaging*, vol. 89, pp. 77–91, 2022.

[163] B. Vasudeva, P. Deora, S. Bhattacharya, and P. M. Pradhan, "Compressed sensing mri reconstruction with co-vegan: Complex-valued generative adversarial network," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 672–681, 2022.

[164] J. Xu, W. Bi, L. Yan, H. Du, and B. Qiu, “An efficient lightweight generative adversarial network for compressed sensing magnetic resonance imaging reconstruction,” *IEEE Access*, vol. 11, pp. 24604–24614, 2023.

[165] X. Li, H. Zhang, H. Yang, and T.-Q. Li, “Cs-mri reconstruction using an improved gan with dilated residual networks and channel attention mechanism,” *Sensors*, vol. 23, no. 18, p. 7685, 2023.

[166] L. Sun, Z. Fan, Y. Huang, X. Ding, and J. Paisley, “Compressed sensing mri using a recursive dilated network,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 2018.

[167] Z. Fan, L. Sun, X. Ding, Y. Huang, C. Cai, and J. Paisley, “A segmentation-aware deep fusion network for compressed sensing mri,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 55–70, 2018.

[168] Q. Huang, D. Yang, P. Wu, H. Qu, J. Yi, and D. Metaxas, “Mri reconstruction via cascaded channel-wise attention network,” in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 1622–1626, IEEE, 2019.

[169] J. Sun, H. Li, Z. Xu, *et al.*, “Deep admm-net for compressive sensing mri,” *Advances in neural information processing systems*, vol. 29, 2016.

[170] W. Qiu, D. Li, X. Jin, F. Liu, and B. Sun, “Deep neural network inspired by iterative shrinkage-thresholding algorithm with data consistency (nistad) for fast undersampled mri reconstruction,” *Magnetic resonance imaging*, vol. 70, pp. 134–144, 2020.

[171] B. Zhou and S. K. Zhou, “Dudornet: learning a dual-domain recurrent network for fast mri reconstruction with deep t1 prior,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4273–4282, 2020.

[172] B. Zhou, J. Schlemper, N. Dey, S. S. M. Salehi, K. Sheth, C. Liu, J. S. Duncan, and M. Sofka, “Dual-domain self-supervised learning for accelerated non-cartesian mri reconstruction,” *Medical Image Analysis*, vol. 81, p. 102538, 2022.

[173] J. Justin, A. Alexandre, and F.-F. Li, “Perceptual losses for real-time style transfer and super-resolution,” in *European conference on computer vision*, pp. 694–711, Springer, 2016.

[174] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[175] Y. Tai, J. Yang, and X. Liu, “Image super-resolution via deep recursive residual network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3147–3155, 2017.

[176] S. E. Petersen, P. M. Matthews, J. M. Francis, M. D. Robson, F. Zemrak, R. Boubertakh, A. A. Young, S. Hudson, P. Weale, S. Garratt, *et al.*, “UK Biobanks Cardiovascular Magnetic Resonance Protocol,” *Journal of Cardiovascular Magnetic Resonance*, vol. 18, no. 1, p. 8, 2015.

[177] B. Fischl, “Freesurfer,” *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012.

[178] B. K. Horn and B. G. Schunck, “Determining optical flow,” *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.

[179] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. Van Der Smagt, D. Cremers, and T. Brox, “Flownet: Learning optical flow with convolutional networks,” in *Proceedings of the IEEE international conference on computer vision*, pp. 2758–2766, 2015.

[180] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, “Flownet 2.0: Evolution of optical flow estimation with deep networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2462–2470, 2017.

[181] A. Ranjan and M. J. Black, “Optical flow estimation using a spatial pyramid network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4161–4170, 2017.

[182] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, “Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8934–8943, 2018.

[183] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, “Voxelmorph: a learning framework for deformable medical image registration,” *IEEE transactions on medical imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.

[184] C. Qin, W. Bai, J. Schlemper, S. E. Petersen, S. K. Piechnik, S. Neubauer, and D. Rueckert, "Joint learning of motion estimation and segmentation for cardiac mr image sequences," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11*, pp. 472–480, Springer, 2018.

[185] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "Ccnet: Criss-cross attention for semantic segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 603–612, 2019.

[186] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, "A reproducible evaluation of ants similarity metric performance in brain image registration," *Neuroimage*, vol. 54, no. 3, pp. 2033–2044, 2011.

[187] T. Vercauteren, X. Pennec, A. Perchant, N. Ayache, *et al.*, "Diffeomorphic demons using itkâ€™s finite difference solver hierarchy," *The Insight Journal*, vol. 1, pp. 1–8, 2007.

[188] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu, "Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces," *Medical image analysis*, vol. 57, pp. 226–236, 2019.

[189] X. Chen, Y. Xia, N. Ravikumar, and A. F. Frangi, "A deep discontinuity-preserving image registration network," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24*, pp. 46–55, Springer, 2021.

[190] A. Majumdar, *Compressed sensing for magnetic resonance image reconstruction.* Cambridge University Press, 2015.

[191] L. N. Tanenbaum, A. J. Tsiouris, A. N. Johnson, T. P. Naidich, M. C. DeLano, E. R. Melhem, P. Quarterman, S. Parameswaran, A. Shankaranarayanan, M. Goyen, *et al.*, "Synthetic mri for clinical neuroimaging: results of the magnetic resonance image compilation (magic) prospective, multicenter, multireader trial," *American journal of neuroradiology*, vol. 38, no. 6, pp. 1103–1110, 2017.

[192] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics:*

A Journal Issued by the Courant Institute of Mathematical Sciences, vol. 59, no. 8, pp. 1207–1223, 2006.

[193] H. Qi, N. Fuin, G. Cruz, J. Pan, T. Kuestner, A. Bustin, R. M. Botnar, and C. Prieto, “Non-rigid respiratory motion estimation of whole-heart coronary mr images using unsupervised deep learning,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 444–454, 2020.

[194] G. Seegoolam, J. Schlemper, C. Qin, A. Price, J. Hajnal, and D. Rueckert, “Exploiting motion for deep learning reconstruction of extremely-undersampled dynamic mri,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 704–712, Springer, 2019.

[195] J. Feng, R. Feng, Q. Wu, Z. Zhang, Y. Zhang, and H. Wei, “Spatiotemporal implicit neural representation for unsupervised dynamic mri reconstruction,” *arXiv preprint arXiv:2301.00127*, 2022.

[196] J. Caballero, C. Ledig, A. Aitken, A. Acosta, J. Totz, Z. Wang, and W. Shi, “Real-time video super-resolution with spatio-temporal networks and motion compensation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4778–4787, 2017.