Real time on-the-fly dynamic MRI for radiotherapy

by

Bryson Dietz

A Thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

MEDICAL PHYSICS

Department of Oncology

University of Alberta

Abstract

External beam radiotherapy (EBRT) utilizes high energy radiation (primarily photons or electrons) generated from a linear accelerator (linac) for the treatment of cancer. Current linac technologies lack the ability to continuously monitor the tumour and surround area during treatment. Real-time imaging in combination with EBRT is ideal, as it can allow for adaptive radiotherapy, or advanced gating to reduce the radiation damage to healthy surrounding tissues and organs at risk. To achieve this, several groups around the world have developed systems that combine a linac with a magnetic resonance imaging (MRI) system (LMR). In fact, many LMRs have already been installed around the world and have already started clinical trials. With the increasing popularity of LMR systems, real-time MR imaging is becoming an important aspect for achieving adaptive radiotherapy or advanced gating techniques. Real-time on-the-fly MRI presents a challenge, given the length of time it takes to acquire fully sampled MR data. Speeding up the acquisition can be achieved via the undersampling of k-space; however, this violates the Nyquist criterion resulting in aliasing artefacts occurring through the reconstructed image. Therefore, there exists a trade-off between image quality and image acquisition speed. The work presented within this thesis investigates MR image reconstruction techniques that are able to speed up the data acquisition (via k-space undersampling), while preserving the image fidelity; allowing for real-time MRI. In particular, this work focuses on image reconstruction for patients with non-small cell cancer (lung tumours), as this provides a challenging scenario of tracking a tumour during involuntary breathing motion.

Two novel techniques were developed for real-time MR imaging. Compressed sensing principal component analysis (CS-PCA), which uses an incoherent phase encoding

scheme to undersample k-space. Principal component analysis is conducted on previously acquired data, to generate principal components that are used to fill the missing k-space in order to rapidly reconstruct undersampled data in under 10 ms per image. Secondly, 3D convolutional neural networks (CNNs) were used to reconstruct undersampled data. Two phase encoding schemes were investigated, incoherent undersampling and coherent low-resolution undersampling. A CNN was trained for each patient individually, making it—what we have termed—a patient-specific CNN. The CNN utilized a cascading network of convolutional layers and data consistency layers for reconstruction. Training the CNN took approximately 6 hours to using our hardware, which consisted of an Intel Xeon E5-2650 CPU and NVIDIA GTX 1080Ti GPU. The image reconstruction time using the trained CNN parameters was less than 65 ms per image.

Preface

The work presented within this thesis is the original work of Bryson Dietz. Dr. Eugene Yip previously acquired the patient data used to evaluate my image reconstruction techniques. Dr. Jihyun Yun developed an auto-contouring software that was used to quantitatively evaluate the image reconstruction segmentability. The manuscript co-authors provided guidance and comments for the published manuscripts.

Chapter 1 provides introductory background for radiotherapy and the various image guided technologies currently available and the benefits for the use of linac-MR systems. Chapter 2 outlines the pertinent theory including the basic physics of magnetic resonance imaging, as well as the current real-time reconstruction techniques such as principal component analysis, compressed sensing, and convolutional neural networks. Chapter 3 describes a real-time MR reconstruction technique that utilizes incoherent sampling and principal component analysis (PCA), which has been published as a first authored paper in the journal *Medical Physics*. Chapter 4 details a real-time MR reconstruction technique that utilizes 3D convolutional neural networks (CNNs). This work has been published as a first authored paper in the journal *Physics in Medicine and Biology*. Chapter 5 investigates the use of a coherent low-resolution phase encoding scheme for CNN reconstruction. The coherent-LR phase encoding scheme is compared to the incoherent scheme presented in the CNN reconstruction

in Chapter 4. This work has been published as a first authored paper in the journal *Physics in Medicine and Biology*. Chapter 6 investigates the effects of translational and rotational shifts on the CNN reconstruction image quality. We intend to publish this work in the future. Chapter 7 presents the conclusion and possible future work.

Furthermore, during my literature review of real-time image reconstruction it became evident that there was some confusion surrounding the nomenclature of the term "real-time". This resulted in a published first authored letter discussing this issue in the journal *Magnetic Resonance in Medicine*, and is included in Appendix A.

Appendix B contains a step-by-step example of the patient-specific convolutional neural network investigated in Chapter 4.

To my wife, Kayla.

Acknowledgments

Throughout the duration of my PhD I have accomplished many milestones in my life. The most important of which being, getting married to my wife Kayla and having my son Milo.

Earning a PhD, as I have learnt, demonstrates that one has the ability to persevere. Getting papers published in reputable journals is not an easy feat, and it could not have been possible without the guidance and help from my exemplary supervisors Dr. Keith Wachowicz and Dr. Gino Fallone. Their keen eye to quality research are what propelled my academic career swiftly. A big thank you to my committee/candidacy members: Dr. Satyapal Rathee, Dr. Jonathan Sharp, Dr. Nicola De Zanche, Dr. Boguslaw Tomanek, who always asked insightful questions. I would also like to thank my external examiner Dr. Ian Cameron for critiquing my thesis.

I would like to acknowledge everyone who has helped me along the way.

Contents

List of Tables

List of Figures

List of Abbreviations

AAPM American Association of Physics in Medicine

ACS Auto Calibration Signal

bSSFP Balanced Steady State Free Precession

BW Bandwidth

CBCT Cone Beam Computed Tomography

CCI Cross Cancer Institute

CNN Convolutional Neural Network

CNR Contrast to Noise Ratio

CPU Computer Processing Unit

CS Compressed Sensing

CS-PCA Compressed Sensing Principal Component Analysis

CTV Clinical Treatment Volume

DC Dice Coefficient

DCT Discrete Cosine Transform

DNA Deoxyribonucleic Acid

EBRT External Beam Radiotherapy

EPID Electronic Portal Imaging Device

FE Frequency Encode

FLASH Fast Low Angle SHot

FOV Field Of View

FPS Frames Per Second

GPU Graphical Processing Unit

GRAPPA GeneRalized Autocalibrating Partial Parallel Acquisition

GRE GRadient Echo

GTV Gross Tumour Volume

ITV Internal Target Volume

JPEG Joint Photographic Experts Group

LHS Left Hand Side

MAPE Mean Absolute Percentage Error

MCP McCulloch–Pitts Neuron

MRI Magnetic Resonance Imaging

MLC Multi-leaf Collimator

NMR Nuclear Magnetic Resonance

NMSE Normalized Mean Square Error

NN Neural Network

NSCLC Non-small Cell Lung Carcinoma

OAR Organ at Risk

PCA Principal Component Analysis

PD Proton Density

PDACS Prior-Data Assisted Compressed Sensing

PE Phase Encode

PDF Probability Density Function

PSNR Peak Signal to Noise Ratio

PTV Planning Treatment Volume

RF Radio Frequency

RHS Right Hand Side

RMSE Root Mean Square Error

RT-MRI Real-Time Magnetic Resonance Imaging

SENSE SENSitivity Encoding

SBRT Stereotactic Body Radiotherapy

SNR Signal to Noise Ratio

SRS Sterotactic Radiosurgery

SE Spin Echo

SS Slice Selection

SSIM Structured SIMilarity Index

TE Echo Time

TLU Threshold Logic Unit

TPS Treatment Planning System

TR Repetition Time

VMAT Volumetric Modulated Arc Therapy

Chapter 1

Introduction

1.1 Thesis Organization

The work presented in this thesis investigated the use of real-time magnetic resonance imaging (MRI) for adaptive radiotherapy. Chapter 1 introduces the advancements of external beam radiotherapy, including the Linac-MR system and the necessity for real-time imaging. Chapter 2 details the theory of MRI as well as two techniques used for accelerated MRI; principal component analysis (PCA) and convolutional neural networks (CNNs). Chapter 3 demonstrates the feasibility of using compressed sensing (CS) and principal component analysis (PCA) for accelerated MR image reconstruction. Chapter 4 investigated the use of 3D convolutional neural networks (CNNs) for real-time image reconstruction. Chapter 5 investigated the use of a coherent low resolution phase-encoding scheme for CNN reconstruction. Chapter 6 investigated the robustness of CNN to translations and rotations. Chapter 7 presents the conclusions and future research endeavors.

1.2 Background

Nearly 50% of Canadians will be diagnosed with cancer in their lifetime [1]. The most recent available statistics state that one in four Canadians are expected to die from cancer. Lung cancer remains the leader in those diagnosed at 14% of all cancers [1]. Unfortunately, lung cancer has a very low survival rate, which is likely linked to the detection of the cancer being diagnosed at a late stage. Cancer is the seventh most costly illness in Canada accounting for for $4.4 billion dollars in economic costs; these costs include those associated with health care ($3.8 billion) and lost productivity ($586 million) [2]. It is evident that we must advance in cancer detection and treatment options in order to reduce the financial burden associated with cancer.

There are several techniques that can be used to treat cancer, which are dependent on the location, aggressiveness, and stage of the cancer. Surgery involves removing the diseased pathology but is invasive and may not remove all the cancer cells. Surgery is most often used when cancer has been detected early and there is a low risk of metastasis [1]. Chemotherapy uses drugs that are often injected into the bloodstream, in order to systematically attack cancer cells throughout the body [1]. Chemotherapy drugs generally attack cells in the body, often resulting in undesirable side effects such as hair loss, vomiting, and loss of appetite. Radiotherapy uses high energy radiation to destroy the cellular DNA of the cancerous region. Photons, electrons, protons, are the most commonly used particles; however, photons are the most common clinically. Radiation can be delivered externally via external beam radiotherapy (EBRT) with the use of a linear accelerator (linac), or internally via brachytherapy. The research described within this work focuses on the use of EBRT.

1.3 External Beam Radiotherapy

External beam radiotherapy (EBRT) is a cancer treatment technique that uses high energy radiation, often in the form of photons or electrons, to destroy cancerous tumour cells. Photons are the primary workhorse of radiotherapy as they are able to penetrate through the body to treat deep tumours [3]. Electrons are generally only used to treat superficial cancers, such as basal and squamous cell carcinomas or melanoma [4]. These particles are often delivered using a medical linear accelerator (linac) which uses a klystron (or magnetron) to accelerate electrons through a waveguide until the specified energy is achieved. The electrons travel past the primary collimator, which is a large tungsten conical component that is designed to define the maximum radiation field size, and shield against leakage radiation. The electrons then either strike a scattering foil or high-Z target (usually tungsten), depending on which is to be delivered for treatment. The electron scattering foil is required to broaden the beam of highly focused electrons, as the beam is initially too focused [3]. The high-Z target creates a forward peaked dose distribution, which may be flattened using a flattening filter; however, there has been a recent emergence of using flattening filter free (FFF), as the dose rate can be much higher (resulting in a reduced treatment time) [5]. This is important for hypo-fractionated deliveries such as stereotactic body radiotherapy (SBRT) and stereotactic radiosurgery (SRS). The delivered radiation is conformed to a particular rectangular field using large jaws located within the linac. There will be two sets of jaws, one the x-axis and one for the y-axis. The radiation field is further conformed to the cancerous region using a device such as a multi-leaf collimator (MLC), which consists of several thin tungsten leaves used to shape the radiation field. Another integral technology required for EBRT is the treatment couch. The treatment couch must be precisely calibrated, as various shifts may be

incorporated to align the patient prior to treatment. Furthermore, modern treatment couches have six degrees of freedom (6-DOF) that can be modified along the three cardinal axes, rotated about the isocenter, and adjusted along the pitch and roll. The use of 6-DOF couches for non-SRS treatments remains debatable [6].

Patients will often receive radiation several times over the course of their treatment. The amount of radiation and number of fractions is very much dependent on the type of cancer, the organ in which it is situated, and what treatment is delivered [3]. The clinical implementation for treatment using EBRT is as follows. Once the patient has been diagnosed with cancer by their oncologist, the patient will undergo a computed tomography (CT) scan of the region. The CT is important for three reasons: firstly, it is used to localize the cancerous region (contoured by the oncologist) as well as surrounding structures, which (for our site) are contoured by the dosimetrist; secondly, the CT is used to calculate the electron densities of the tissues, which is needed for the treatment planning system (TPS); thirdly, the CT data will also be used to ensure the patient is accurately set-up during their treatment. To position the patient the linac is equipped with a cone beam CT (CBCT), which will image the patient each fraction (using either kV/kV orthogonal images, or CBCT), in order to ensure the patient is positioned accurately.

1.3.1 A Brief History of Linear Accelerators

Over one century ago in 1895, William Rontgen discovered x-rays while studying the range of cathode rays produced by a Crookes tube. This discovery led to Rontgen being awarded the Nobel prize in Physics in 1901 [7]. The use of x-rays for imaging and treatment of pathologies was quickly investigated and implemented (without knowing the consequences of occupational exposure). In fact, the first clinical use

of x-rays was only 60 days after the discovery, where an ulcerated breast cancer was treated in Chicago [8]. Prior to the use of linear accelerators, there were other devices that were developed for the treatment of cancer using external radiation. The first devices operated in the kV range, and therefore were only suitable for superficial tumours. One such device is the Coolidge tube developed in 1913, which is essentially a precursor to the modern x-ray tube [9]. The next major breakthrough was in 1948, when the first MV teletherapy unit was installed in Hamilton Canada [9]. The teletherapy unit used a cobalt 60 source to deliver photons with energies of 1.17 MeV and 1.33 MeV. This was a large technological advancement as these systems allowed for the treatment of deep tumours, while sparing dose to the skin. Furthermore, deliveries of up to 45-60 Gy were possible. Over the course of 20-30 years, 1,000's of teletherapy systems were sold worldwide and they became the workhorse for external radiotherapy treatment [9]. During this time, radar research led to the development of magnetron and klystron technology. These developments would result in the first MV linac, which was installed in London England in 1953. Linacs are ideal clinically, as they are not affected by radioactive decay (Co-60 has a half-life of 5.27 years) and could produce variable energies between 6-20 MV for the delivery of photons or electrons.

With the advent of the linac for cancer treatment, there have been many milestones in improving these systems. The multi-leaf collimator (MLC) was developed in the 1990's which allowed for 3D conformal treatment. Today, the MLC consists of 80-120 tungsten alloy leaves that can be used along with the primary and secondary collimators, or as a replacement for one of the collimators, to shape the radiation field to avoid healthy structures and/or conform to the tumour. The placement of the MLC is dependent on the linac manufacturer. Further progress was made in the early 2000's when intensity-modulated radiotherapy (IMRT) was introduced along

with inverse treatment planning using a dedicated treatment planning system (TPS). By modulating the intensity of the radiation beam, sensitive organs could be better spared while increasing the conformity to the tumorous region [10]. In 2007 further progress was made with the use of volumetric modulated arc therapy(VMAT) [11]. This technique allows for three parameters to be simultaneously adjusted, MLC aperture, gantry speed, and dose rate, to ensure optimal dose conformity.

1.3.2 Prevalence of EBRT in Cancer Treatment

External beam radiotherapy (EBRT) is the most common type of radiation therapy for the treatment of cancer. Of the 14 million cases of cancer that are diagnosed every year globally, 50% of these cases will use EBRT for either curative treatment or palliative care (approximately a 25% / 25% split) [12]. Furthermore, EBRT is used to treat a wide variety of sites including head and neck, lung, breast, brain, and prostate. Many of these sites are in regions where motion may be prevalent and involuntary, such as breathing or abdominal gas. A tumour in the lung, for example, may experience motion along the superior-inferior direction of over 3 cm during a respiratory cycle [13]. A common way to mitigate the motion caused by involuntary breathing is with the use of respiratory gating. This may involve using a chest strap or bellows to track the inspiration and expiration, which can be used to ensure radiation is delivered only during the expiration phase, for example. In order to increase the accuracy of patient set-up and reproducibility for every treatment fraction, masks (for head and neck) can be use to secure the patients head to the treatment couch. Vacuum immobilization bags can be used to minimize unwanted body motion and to increase patient stability and reproducibility. However, masks may still present some variation between treatments and gating results in a longer treatment. Furthermore, respiratory gating cannot be used efficiently for patients

that have an irregular breathing cycle. One possible way to solve these issues is by having real-time imaging feedback to ensure accurate localization of the radiation.

1.3.3 Radiotherapy Treatment Margins

Before discussing the current and future linac technologies, it is important to understand why we strive for more precise localization of the radiation. In general, there are several terms that are used when planning a radiotherapy treatment. Figure 1.1 contains the boundaries for each of the structures to be discussed. The gross tumour

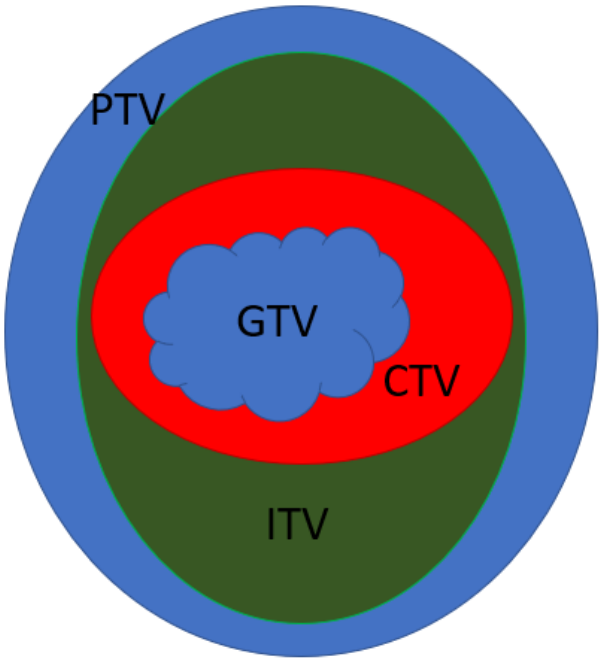

Figure 1.1: Margins used for radiotherapy treatment planning. The ITV encompasses motion (mainly physiological such as breathing), and the PTV includes all other uncertainties to ensure the CTV receives the dose prescribed by the oncologist.

volume (GTV) contains the absolute mass or tumour volume, this is most often contoured by the oncologist. There may, however, be microscopic sub-clinical cancerous tissue surrounding the visible tumour mass; the region that contains the GTV plus

the additional sub-clinical tissue is known as the clinical target volume (CTV). In order to account for tumours in regions of motion (physiological or otherwise), an additional internal target volume (ITV) may be incorporated into the planning. The ITV is often asymmetric depending on the location of the tumour. A tumour located in the lung, for example, will generally experience far greater motion in the superior-inferior direction and will have a larger ITV in that direction to account for it. Finally, the planning target volume (PTV) uses yet a larger region in order to incorporate the patient set-up error, to ensure the dose prescribed by the oncologist is delivered to the CTV. Thus, if the ITV can be minimized, the PTV will have a margin that is tightly bounded to the CTV, reducing the dose to the healthy surrounding tissue and sparing other organs at risk (OARs). While ensuring dose is delivered to the cancerous region is important, we must also be cognizant of the OARs surrounding the cancerous region. The OARs generally have a strict dose limit, which should be minimized whenever possible.

1.3.4 Current State of External Beam Radiotherapy

Modern MV EBRT is generally conducted using linacs, although cobalt 60 teletherapy systems still see use in developing countries. Depending on the linac manufacturer the radiation may initially be collimated using large tungsten jaws followed by the MLC or vice versa. As mentioned earlier, the MLC is comprised of several tungsten alloy leaves that are used to shape the radiation field to avoid healthy structures and/or conform the radiation close to the tumour. Prior to treatment, patients are scanned using a CT simulator. The CT simulator is required to determine the electron densities needed to generate a dosimetric treatment plan. The CT images also provide the baseline images for all linac treatments to be aligned to. Current linac units contain an on-board imaging (OBI) device, which generally consists of a cone

beam CT (CBCT) system. As a patient will often be treated several times within the course of a month or two, the OBI is required to ensure the alignment of the patient (at the time of their treatment) to their initial planning CT. This interfraction alignment is important, as the radiotherapy planning software uses the anatomy of the planning CT to generate the treatment plan. The treatment plan is developed to ensure the cancerous tissue receives the prescribed dose requested by the oncologist, while minimizing the radiation delivered to nearby organs at risk. Alignment ensures the planning region matches the anatomy of the patient during each round of treatment. However, as the OBI acquires static images to align gross anatomy, it is unable to cope with any intrafraction motion that may occur from the patient (either voluntary or involuntary) during the radiation delivery. Thus, the current state of linac technology does not have the ability to monitor intrafraction motion using imaging; however, there are technologies that can monitor tumour motion surrogates, as will be described in the next section.

1.3.5 Image Guidance for External Radiotherapy

As discussed above, imaging is a crucial requirement for radiotherapy for planning and patient alignment. Since the 1990's, linacs have made use of electronic portal imaging devices (EPIDs) as a replacement for MV film [14]. An EPID consists of an amorphous-silicon detector that can generate images from MV photons delivered by the linac. While these images generally have poor contrast, they reveal bony anatomy that can be used to aid in the alignment of the patient. In addition to localization, EPIDs are also used to dosimetrically verify volumetric modulated arc therapy (VMAT) treatment plans [15]. Following the development of the EPID, linac manufacturers began including kV imaging systems coupled to the linac, which has better contrast than the EPID. During the mid 2000's the CBCT was introduced,

enabling 3D imaging of the patient prior to treatment [16]. Given the additional dose delivered to the patient via the use of CBCT, it cannot be used for real-time monitoring purposes.

A few products exist that aim to increase the accuracy of linac treatment, through the monitoring of intrafraction motion. An excellent review on the topic was recently published by Bertholet *et al.* [17]. There are devices, such as respiratory bellows, that can be used to monitor the breathing phase during treatment to allow for gated treatment. A more advanced gating technique known as AlignRT (VisionRT, London, Great Britain) uses 3D stereo cameras to track the surface of the patient, to ensure accurate positioning before and during treatment [18]. Isocenter accuracy of 3mm has been recently reported using AlignRT for breast cancer treatments [19]. Another product known as the Calypso system (Varian Medical Systems, Palo Alto, USA) uses three electromagnetic transponders (1.85mm diameter, 8.7mm long) that are surgically implanted and tracked using a portable detector to monitor its centroid location (stated as 'GPS for the body') [20]. Calypso is an accurate technology (sub-millimeter) for real-time tumour monitoring, but can only be implanted into certain sites, only monitors the centroid of the transponders, and has the further complication of requiring a surgical procedure [21]. The BrainLab ExacTrac (BrainLab, Feldkirchen, Germany) uses a combination of stereoscopic kV imaging along with external (IR reflective fiducials) breathing monitoring to accurately track regions; however, the use of kV imaging increases the overall patient dose [22]. While these advanced systems increase the accuracy of EBRT treatments, they do not have the capability to continuously visualize the treatment location during treatment without further increasing the dose to the patient.

Recently, there has been a strong interest in image guided linac treatment, partic-

ularly with the use of magnetic resonance imaging (MRI). MRI is an ideal imaging modality for external beam radiation treatment (EBRT). The benefits of using MRI include excellent soft tissue contrast (compared to other imaging technologies), various contrast via pulse sequence, and that it does not contribute to an increased patient dose. MRI also has the ability to modify the contrast by changing the imaging parameters or pulse sequence. Additionally, MRI is capable of both anatomical and functional imaging, such as fMRI and diffusion tensor imaging.

By monitoring the tumour during treatment, more conformal and accurate treatments can be achieved. Several groups around the world are currently developing systems that combine a linac with an MRI system [23, 24, 25, 26]. The Cross Cancer Institute (CCI) located in Edmonton, Alberta, has recently developed a system that combines a 6 MV linac with a 0.5T bi-planar MRI system. This unique system, known as the Linac-MR (LMR), contains a standard linac treatment unit that is mounted onto a rotating 0.5T bi-planar magnet. The entire system rotates around the patient, similar to a conventional linac; however, the MR system allows for simultaneous imaging during radiation delivery. The CCI group was the first in the world to demonstrate the feasibility of simultaneous MR imaging during irradiation [27]. The ultimate goal for the Linac-MR is to use real-time MR imaging to adapt the treatment on the fly, to ensure the tumour receives the required dose, while sparring the surrounding healthy tissue.

1.4 Linac-MR

Recently, the Cross Cancer Institute at the University of Alberta has developed a Linac-MR system that combines MR imaging with a linac [23]. This system–known as the Aurora RT–is the flagship product created by the recently formed company,

MagnetTx. Figure 1.2 contains a rendering of the system to scale, demonstrating how a patient would be treated. The bi-planar magnet is unique and was created

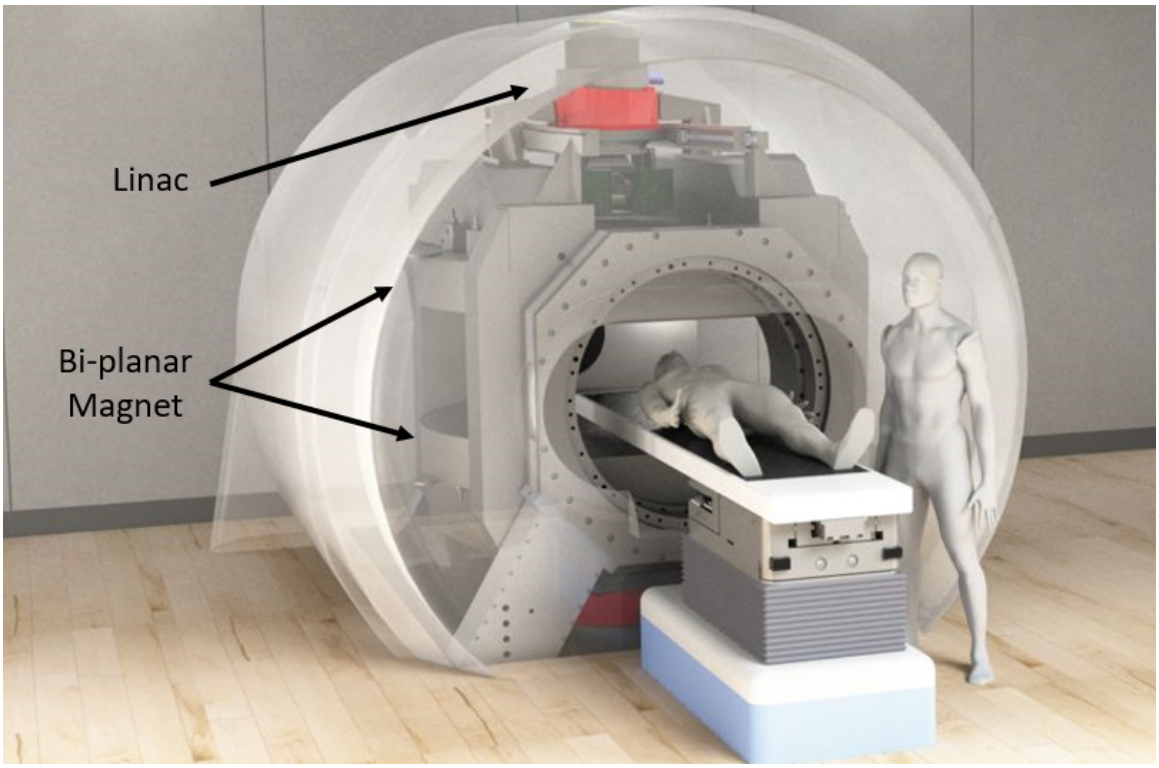

Figure 1.2: Complete rendering of the Linac-MR Aurora RT system by MagnetTx. The bi-planar magnet and the linac head are highlighted. The linac is mounted on the top pole plate of the bi-planar MR system. The bi-planar MR system is unique, in that it allows for the radiation and the magnetic field to be parallel. *Image courtesy of Dr. Gino Fallone, Chairman of MagnetTx.*

specifically for the development of an LMR system [28], whereas a majority of MR systems use cylindrical magnets. The entire Linac-MR system is free to rotate around the patient and will allow for the simultaneous delivery of radiation while imaging in real-time. The Edmonton group was the first to demonstrate the feasibility of simultaneous MR imaging during irradiation [27]. The design is unique as the radiation is delivered in parallel to the magnetic field, which has been shown to be advantageous compared to the perpendicular orientation such as Elekta's Unity and ViewRay's MRIdian cylindrical magnet based systems [29]. The Australian LMR

has the ability to operate using either the perpendicular or parallel orientations [25]; however, the parallel orientation involves the rotation of the patient, which introduces deformations that must be accounted for. Furthermore, the Aurora RT system is the only system that has the capability to shift the couch laterally, enabling the treatment of peripheral tumours. Two LMR systems, the Aurora RT system and the MRIdian, operate at low field strengths (0.5T and 0.35T, respectively) [30]. There are many benefits for having an MRI system operate at low field [31, 32]. While the signal-to-noise ratio (SNR) generally decreases at lower field strengths, this is not true for the contrast-to-noise ratio (CNR) [33]. Lower field strengths are also less effected by magnetic susceptibility artifacts and specific absorption rate (SAR), which can result in tissue heating. In addition, conductive devices and implants are less prone to heating at lower fields.

Having the radiation field parallel to the magnetic field produces a focusing of contaminant electrons which may increase the skin dose [34]. In fact, the focusing of the electrons may be used as an advantage, as it increases the dose along the direction of the beam [29]. However, when the radiation field is perpendicular, electrons will curl via the Lorentz force

$$\vec{F} = q(\vec{E} + \vec{v} \times \vec{B}). \tag{1.1}$$

The Lorentz force ($\vec{F}$) is equal to the charge of the electron (q) multiplied to the electric field $\vec{E}$ in addition to the curl of the speed of the electron and the magnetic field $\vec{v} \times \vec{B}$. The curling of electrons may complicate planning in regions with air, such as the lung or bowel. This effect is known as the electron return effect, which is prominent in tissue interfaces such as the lung [29].

A major benefit of combining MRI with a linac, is the ability to track tumours via real-time imaging and deliver radiation simultaneously. Aside from the excellent soft

tissue contrast offered by MRI, it acquires images without the use of ionizing radiation and is therefore safe for prolonged use. A lower field strength is ideal for prolonged continuous real-time MR imaging; at higher field strengths, tissue heating becomes a limiting factor (due to the RF power deposited into the body) [32]. This is due to the SAR being proportional to the square of the main magnetic field ($SAR \propto B_o^2$).

While having adaptive radiotherapy is ideal, real-time MR imaging can also be used as an advanced gating technique, which may be easier to implement, clinically. A recent gating study investigated having the patient view the reconstructed images (during their treatment) using specially constructed glasses allowing them to view a TV outside the magnet, in order to help control their breathing and guide the tumour into the treatment region [35]. Real-time imaging can also allow for adaptive radiotherapy by having the MLC leaves update based on the tumour location calculated from the real-time images [36]. Image guidance for adaptive radiotherapy will allow for continuous radiation to be delivered the tumourous region, resulting in a shorter treatment time compared to current clinical respiratory gating techniques. Another implementation for real-time MR imaging is the use of a radiation beam stop if an OAR was to enter the location of radiation delivery.

1.5 Real-time MR

As discussed above, perhaps the most valuable aspect of the LMR is the ability to image for the purpose of tumour tracking. To achieve this, the MR system must be capable of real-time imaging. I refer the reader to Appendix A for a discussion on the nomenclature surrounding the term "real-time" for MRI. In short, it was found that there is a discrepancy with the terminology surrounding real-time MRI. Ideally, real-time imaging will generate images that are available for immediate use with low

latency. Real-time MR presents a challenge, given the relatively slow acquisition of raw MR data. The crux being, as one decreases the imaging time (via undersampling the raw MR data), there is an increased presence of artefacts within the reconstructed images. Using a 3T Phillips Achieva system, a fast balanced steady-state free precession (bSSFP) sequence can acquire fully sampled 2D images (128×128 pixels) at 275 ms per dynamic frame using an echo/repetition time of 2.2/1.1 ms, respectively. This is faster than the recommended imaging latency of 500 ms presented in the American Association of Medical Physicists (AAPM) Task Group 76a (TG-76a), which focused on respiratory motion in radiotherapy [13]. Further accelerating the MR data acquisition, however, may be beneficial in several ways: increasing the frame rate for better temporal resolution, increase in spatial resolution, and the ability to acquire multiple 2D-orthogonal images or full 3D imaging volume.

There are many reconstruction techniques that perform well for dynamic (temporal) data reconstruction; however, these techniques are generally focused on reconstructing the entire k-t domain (using the full temporal dataset) [37, 38, 39]. The k-t reconstruction techniques are not applicable to real-time (online) imaging, which is an important feature for MR-Linac systems. Depending on the MR hardware constraints of the LMR system, it may be limited to a small number of coil channels, reducing the applicability of parallel imaging reconstruction techniques [40, 41, 42], including the more recent development of the non-linear inverse (NLINV) technique [43]. Therefore, reconstruction techniques that do not require the use of multi-coil sensitivity maps are ideal for systems with hardware limitations. There have been a few studies that have the capability to reconstruct MR images in real-time without the requirement for large arrays of receive coils. The use of compressed sensing (CS) does not require the use of multiple coils, and has been addressed for real-time MR reconstruction at reconstruction times of 160 ms for 128×128 sized images [44]. The real-time

CS work comprised of accelerated radial acquisition with an undersampling factor of 5x [44]. Another CS method that utilizes spatial and previous temporal information can achieve reconstruction times of 710 ms per frame.

In particular, it has been suggested that ideally 3D real-time imaging should be utilized to track lung tumours (for example) [13, 45]. In order to achieve real-time 3D acquisitions, undersampling of raw MR data will be necessary. While 3D tumour tracking is ideal, it is also far more difficult to achieve. Not only will the acquisition time be increased, but the reconstruction time as well, resulting in a large latency. For adaptive radiotherapy (via tumour tracking) to be feasible, the latency from the start of imaging to the delivery of radiation needs to be minimal. This includes image acquisition, image reconstruction, image segmentation, updating of the MLC, and the delivery of radiation. Thus, having a higher imaging frame rate with a low latency is necessary. Solving this unique problem is the bulk in my PhD work, where I focused on the image reconstruction portion specifically. I developed and investigated two real-time image reconstruction techniques for the use in adaptive radiotherapy (tumour tracking).

The first reconstruction technique I investigated utilized principal component analysis (PCA). The use of PCA is fast and was discovered to be capable of reconstructing highly undersampled MR images for real-time imaging purposes. This technique is suited to dynamic data sets, such as lung motion, and uses a previously acquired fully sampled data set to find the principal components (PCs) of a dynamic data set. The largest PCs correspond to the direction of greatest variance, which captures the region of largest dynamic motion. These PCs can be used to fill in the missing k-space from the undersampled data in under 10 ms per image. However, it was discovered that this technique has increasing noise as the database becomes outdated. This led

me to investigate a more robust technique.

The second reconstruction technique I investigated used convolutional neural networks (CNNs). The CNNs utilizes graphical processing units (GPUs) to train a neural network to learn how to remove the aliasing by convolving filters to the aliased image. For this study a CNN was trained for each patient, which is a paradigm shift from the CNN training that uses data acquired from several patients to create a generalized model. Having a CNN trained to each patient (separately) allows the model to become tuned to each patient's anatomy and pathology. We furthermore investigated two undersampling phase-encoding schemes for CNN reconstruction: (1) incoherent, where the central portion of k-space is fully-sampled and the higher spatial resolution is pseudo-randomly undersampled, and (2) coherent low-resolution, where only the central region of k-space is acquired.

Chapter 2

Theory

2.1 Magnetic Resonance Imaging (MRI)

Nuclear magnetic resonance (NMR) can be traced back to the resonance experiments performed by Isidor Rabi in 1938, who was awarded the 1944 Nobel prize in Physics for his discovery [46]. Further progress was made in 1946 when Felix Bloch and Edward Purcell independently published studies investigating NMR in condensed matter. Their work focused on using NMR for the detection of liquids and solids, which won them the Nobel prize in Physics in 1952. It was not until 1973 when Lauterbur and Mansfield independently published studies investigating the localization of the NMR signal using gradient coils, that it was realized that NMR could be used for imaging [47, 48]. Lauterbur and Mansfield are ultimately credited for the discovery of magnetic resonance imaging (MRI) as it is used today, and in 2003 they shared the Nobel prize in Physiology or Medicine for their work.

Magnetic resonance imaging (MRI) is an imaging modality capable of anatomical and functional imaging. The clinical benefits for the use of MRI include superior soft

tissue contrast, ability to image at various contrasts (T_1, T_2, PD, etc.), functional capability using fMRI and diffusion tensor imaging, as well as image acquisition with the use of non-ionizing radiation. These benefits have resulted in MRI becoming a useful clinical tool for monitoring and diagnosing cancer, since the superior soft tissue contrast makes MRI an ideal imaging technique for imaging masses such as tumours. Furthermore, functional techniques such as spectroscopy and diffusion imaging have been used to aid in the treatment for certain cancers [49, 50]. The non-ionizing aspect of MRI allows for real-time (fluoroscopic) imaging, without worry of delivering excess radiation as with standard x-ray devices such as fluoroscopy and computed tomography (CT), make MRI the ideal choice for real-time dynamic imaging studies.

2.1.1 NMR Signal

It is well known that subatomic particles contain an intrinsic property known as spin. Fermions are particles that have a spin value of $n/2$, which are of interest for NMR experiments. Baryons are subatomic particles that contain an odd number of fermionic quarks, such as protons and neutrons. Protons are abundant in the human body due to the hydrogen composition in water and will have an angular momentum $\vec{J}$ and a magnetic moment $\vec{\mu}$ according to the following equation [51]

$$\vec{\mu} = \gamma \vec{J}. \tag{2.1}$$

The magnetic moment $\vec{\mu}$ is equal to the gyromagnetic ratio γ multiplied to the spin angular momentum $\vec{J}$. The magnitude of the magnetic moment has a value

$$|\vec{\mu}| = \gamma h \sqrt{s(s+1)}. \tag{2.2}$$

Here we have used $\eth$ which is the gyromagnetic ratio divided by 2π. The factor h is Planck's constant (6.6×10^{-34} J $\cdot$ s) and s is the spin quantum number.

When protons are placed into a magnetic field (such as one produced in a MRI system), the magnetic moments will align to the magnetic field. The aligned protons can be oriented parallel or anti-parallel to the field. The parallel orientation is more favoured, creating a net magnetization, as it has a lower energy than the anti-parallel orientation. This well-known quantum phenomenon is known as the Zeeman effect, which splits the quantum spin degeneracy into two states: spin up or spin down (i.e. parallel or anti-parallel). The energy of the magnetic moment $\vec{\mu}$ within the main magnetic field $\vec{B}$ is

$$\begin{aligned} E &= -\vec{\mu} \cdot \vec{B} \\ &= -\mu_z B_z \\ &= -\gamma \, \hbar \, m_s \, B_z. \end{aligned} \tag{2.3}$$

The factor $\hbar$ is the reduced Planck's constant ($h/2\pi$), and m_s is often regarded as the magnetic quantum number, which yields $2s+1$ values such that $m_s \in \{-s, ..., s\}$. For particles with spin $1/2$, there are two possible energy levels

$$m_s = \pm\frac{1}{2}. \tag{2.4}$$

The Zeeman effect is fundamental to NMR, and is the primary reason we are able to detect a bulk magnetization. Figure 2.1 demonstrates how the single degenerate energy state is split into two separate energy states.

The gyromagnetic ratio is dependent on the nuclei and is quantum mechanical in

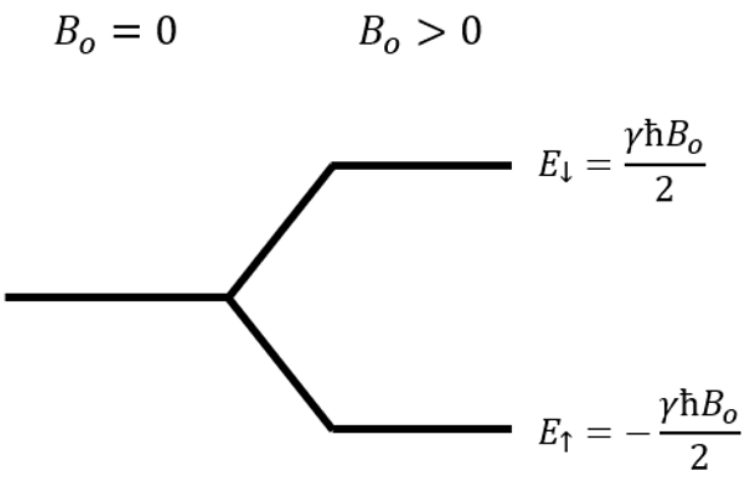

Figure 2.1: In the presence of an external magnetic field, the single degenerate state is split into two separate energy levels; one that is slightly more favourable (lower energy) where the spins align parallel (spin up) and one where the spins align anti-parallel (spin down).

nature. For protons, the gyromagnetic ratio can be calculated using [52],

$$\gamma = \frac{\omega_o}{B_o} = \frac{\mu_N \, g_p}{\hbar}. \tag{2.5}$$

Where g_p is a dimensionless quantity that characterizes the magnetic moment and angular momentum of the proton, μ_N is the nuclear magnetron, and $\hbar$ is the reduced Planck's constant. Furthermore, here we can define the Larmor equation,

$$\omega_o = \gamma B_o \tag{2.6}$$

which is the precession angular frequency for the proton magnetic moment. As the Larmor frequency (and equation) are critical to NMR, we will take a moment to show a simple geometric derivation [51]. Figure 2.2 shows a magnetic moment precessing in a magnetic field. The change in the magnetic moment can be expressed as,

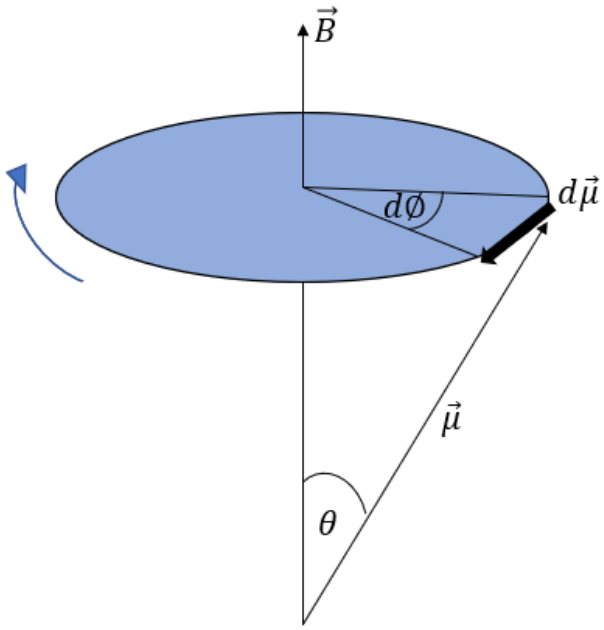

Figure 2.2: A simple geometric schematic of a magnetic moment $\vec{\mu}$ interacting with a magnetic field $\vec{B}$. The magnetic moment is tipped from the direction of $\vec{B}$ by an angle θ.

$$|d\vec{\mu}| = |d\phi|\mu \sin\theta. \tag{2.7}$$

However, we can also write the torque such that

$$\begin{aligned} |d\vec{\mu}| &= \gamma|\vec{\mu} \times \vec{B}|dt, \\ &= \gamma\mu B \sin\theta dt. \end{aligned} \tag{2.8}$$

Comparing the RHS of the two equations above, we see that

$$\frac{|d\phi|}{dt} = \gamma B. \tag{2.9}$$

Which for the constant magnetic field B_o, and denoting the angular frequency $|d\phi|/dt$ as ω_o, we arrive at the Larmor equation stated in Equation 2.6.

It is the manipulation and detection of the net magnetization within a sample that comprises an NMR experiment. There are three components in any NMR experiment which are:

1. Preparation of the bulk spin system,
2. Excitation of the bulk spin system,
3. Acquisition of the signal as the spins precess while the bulk magnetization relaxes back to equilibrium.

The preparation component occurs when the spins are manipulated to be aligned in the same direction, producing a net magnetization, which occurs when a sample is placed into a strong magnetic field. The excitation component involves applying energy to the system in order to excite the spins with the use of an additional magnetic field, known as the RF or B_1 field. The acquisition component occurs immediately after excitation, where (via Faraday's law of magnetic induction) we measure the exponentially decaying precession (resulting in an alternating voltage) as the sample (bulk magnetization) relaxes back to equilibrium.

The signal produced from a MR imaging experiment originates from hydrogen protons. The human body is comprised mostly of water (H_2O), of which each molecule contains two hydrogen nuclei that each contain one proton. When put into a strong magnetic field, nearly 99.9% of the sample will be randomly orientated in the main magnetic field. While this may seem inefficient, to our benefit there are approximately 1.6×10^{19} proton spins contained within 1 μL of water! The abundance of these hydrogen protons, results in the signal (net magnetization) we detect during an MR experiment.

The number of protons that favour the lower energy state can be calculated

$$\begin{aligned} \Delta N &= N\uparrow - N\downarrow \\ &= N(P_+ - P_-). \end{aligned} \tag{2.10}$$

The term $P_\pm$ is the normalized Boltzmann factor,

$$\begin{aligned} P_\pm &= \frac{e^{\pm\frac{E}{kT}}}{Z} \\ &= \frac{e^{\pm\frac{-m_s\hbar\omega_o}{kT}}}{\sum_{m=-s}^{s} e^{\frac{-m_s\hbar\omega_o}{kT}}} \end{aligned} \tag{2.11}$$

where the energy $E = -m_s\hbar\omega_o$. The term m_s is the spin quantum number, $\hbar$ is the reduced Planck's constant, and ω_o is the angular (Larmor) frequency of the proton. There are only two spin states (m_s) for a proton with an intrinsic spin of 1/2: +1/2, -1/2. Therefore, we can write

$$P_\pm = \frac{e^{\pm\frac{\hbar\omega_o}{2kT}}}{e^{\frac{\hbar\omega_o}{2kT}} + e^{\frac{-\hbar\omega_o}{2kT}}}. \tag{2.12}$$

The terms on the exponential denominator are the Boltzmann constant (k) and the temperature in Kelvin (T). Since $kT >> \hbar\omega_o$ we can approximate the exponential terms such that $e^{\pm ax} \approx 1 \pm ax$. After a bit of algebra, we arrive at

$$\begin{aligned} \Delta N &= \frac{N\hbar\omega_o}{2kT}, \\ \frac{\Delta N}{N} &= \frac{\hbar\gamma B_o}{2kT}. \end{aligned} \tag{2.13}$$

For example, solving Eq. 2.13 at a field strength of 3 Tesla ($\hbar$ = 1.05 x 10^{-34} m^2 kg s^{-1}, γ = 42.58 MHz/T, T = 300 K, and k = 1.38 x 10^{-23} m^2 kg s^{-2} K^{-1}), we get a value of 10 spins per million. This may not sound like much, but it is important to

note that we are dealing with a magnitude of Avogadro's constant of protons (6.02 x 10^{23})!

2.1.2 Bloch Equations

In 1946, Felix Bloch published a paper describing the classical mathematical framework for the NMR signal [53]. The equation for the manipulation of the magnetization can be written [51],

$$\frac{d\vec{M}}{dt} = \gamma\vec{M} \times \vec{B}_{ext} + \frac{1}{T_1}(M_o - M_z)\hat{k} - \frac{1}{T_2}\vec{M}_{xy}. \tag{2.14}$$

Separating the above equation into individual components yields the laboratory reference frame equations,

$$\begin{aligned}
\frac{dM_x}{dt} &= \gamma(\vec{M} \times \vec{B}_{ext})_x - \frac{M_x}{T_2}, \\
\frac{dM_y}{dt} &= \gamma(\vec{M} \times \vec{B}_{ext})_y - \frac{M_y}{T_2}, \\
\frac{dM_z}{dt} &= \gamma(\vec{M} \times \vec{B}_{ext})_z + \frac{M_o - M_z}{T_1}.
\end{aligned} \tag{2.15}$$

The T_1 and T_2 values were defined by Bloch as relaxation times. T_1 is known as the longitudinal, thermal, or spin-lattice relaxation time. The term longitudinal comes from the fact that T_1 is the time constant (unique for each tissue) indicative of the time it takes for the magnetization to recover back to the 63% of the maximal M_o value. Thermal, meaning that the energy is irreversibly lost to the surrounding nuclei. Spin-lattice, is an older term that originated back when NMR experiments were primarily conducted in crystalline structures [54]; however it is also used to explain the quantum mechanical nature of the spin phenomena [55]. The time constant T_2, is known as transverse or spin-spin relaxation. After time T_2 the transverse component of the

magnetization (M_{xy}) has lost 63% of its original value, due to dephasing. These time constants are tissue-dependent, where the T_2 values are always less than the T_1 values.

Assuming the case of a constant field along the z-direction ($\vec{B}_{ext} = B_o\hat{z}$), we can simplify Eq. 2.15 to,

$$\begin{aligned}
\frac{dM_x}{dt} &= \gamma\big(M_y B_z - M_z B_y\big) - \frac{M_x}{T_2}, \\
\frac{dM_y}{dt} &= \gamma\big(M_z B_x - M_x B_z\big) - \frac{M_y}{T_2}, \\
\frac{dM_z}{dt} &= \frac{M_o - M_z}{T_1}.
\end{aligned} \tag{2.16}$$

Equation 2.16 can be simplified by noting that the magnetic field is directed along the z-axis ($B_x = 0$, $B_y = 0$, $B_z = B_o$). The solution to Eq. 2.16 is,

$$\begin{aligned}
M_x(t) &= \big(M_x(0)\,\cos(\gamma B_o t) + M_y(0)\,\sin(\gamma B_o t)\big)e^{-t/T_2}, \\
M_y(t) &= \big(M_y(0)\,\cos(\gamma B_o t) - M_x(0)\,\sin(\gamma B_o t)\big)e^{-t/T_2}, \\
M_z(t) &= M_z(0)e^{-t/T_1} + M_o(1 - e^{-t/T_1}).
\end{aligned} \tag{2.17}$$

This can be simplified with a bit of algebra. Using the definition that $M_{xy} \equiv M_x(t) + iM_y(t)$ we can rewrite the above to:

$$\begin{aligned}
M_{xy}(t) &= M_{xy}(0)e^{-t/T_2}, \\
M_z(t) &= M_z(0)e^{-t/T_1} + M_o(1 - e^{-t/T_1}).
\end{aligned} \tag{2.18}$$

2.1.3 Signal Excitation and Detection

In order to excite the sample, we need to tip the magnetization down into the x-y plane. To do so we use a B_1 RF pulse that is delivered perpendicular to the main

magnetic field B_0. The B_1 field can be written as,

$$B_1(t) = B_1\, a(t)\, e^{-i\omega_o t} e^{-i\phi} \tag{2.19}$$

where B_1 is the RF field strength (order of μT), $a(t)$ is a unitless pulse envelope and the exponential terms are the carrier frequency (which is at the Larmor frequency) and the RF phase, respectively.

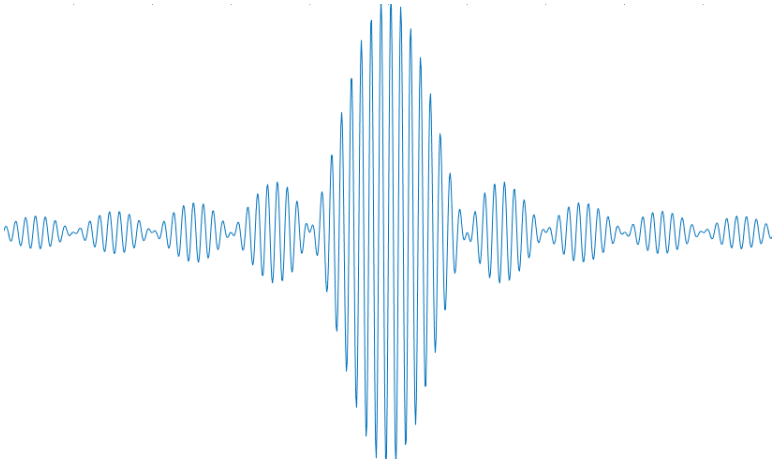

Figure 2.3: A common B_1 pulse envelope is the Sinc pulse. Here you can see that the carrier frequency is applied at the Larmor frequency, within the Sinc envelope.

The role of the RF (B_1) pulse is to manipulate or rotate the magnetization M_o by some angle. The angle at which the magnetization M_o will rotate about B_1 (in the rotating frame of reference) is known as the flip angle, which is given by (assuming B_1 is on resonance),

$$\alpha = \gamma \int_0^t B_1(t)\, dt. \tag{2.20}$$

If we apply a 90° flip angle, we can rewrite the Bloch solutions (Eq. 2.18) as

$$\begin{aligned} M_{xy}(t^+) &= M_o e^{-t/T_2}, \\ M_z(t^+) &= M_o(1 - e^{t/T_1}). \end{aligned} \tag{2.21}$$

The equation $M_{xy}(t^+) = M_o e^{-t/T_2}$ is known as the free induction decay (FID) signal, which is of great importance for NMR spectroscopy experiments. However, due to static inhomogeneities, the signal decays with a time constant of $1/T_2^*$

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_2'}. \tag{2.22}$$

Where T_2 is dependent on the dipole-dipole (thermodynamic) interactions and T_2' is due to the field inhomogeneities. The factor T_2' arises from imperfections in the magnetic field, or at air/tissue interfaces with large susceptibility differences. The additional signal lost to T_2' can be refocused through the use of a spin-echo sequence. The spin-echo sequence removes the effect of T_2' by refocusing the spins as shown in Figure 2.4. After exciting the excess spins to the xy-plane with a 90° RF pulse (along the x-axis), a second 180° RF pulse is applied (along the y-axis) at time τ, which reverses the phase causing them to refocus at time 2τ.

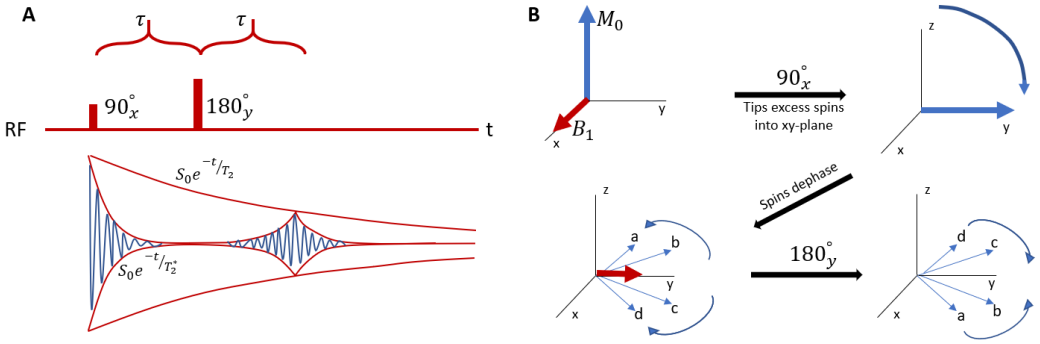

Figure 2.4: Demonstration of how a spin-echo sequence removes the T' signal loss by refocusing the dephasing spins with the use of an additional 180° RF pulse.

As M_{xy} precesses about the transverse plane (after the B_1 field is turned off), the signal can be detected using a receive coil via Faraday's law. Faraday's law states

that a changing magnetic flux will induce an *emf* in a coil such that,

$$emf = -\frac{d\Phi}{dt}. \tag{2.23}$$

The flux Φ through the coil can be written as,

$$\Phi = \int_A \vec{B} \cdot d\vec{S}. \tag{2.24}$$

Assuming the coil is oriented perpendicular to the x-axis, $\Phi = AB_x$, where A is the area of the receive coil and B_x is the magnetic field along the x-direction. This can be expressed as,

$$\Phi = \mu_{\mathrm{o}} A M_x \tag{2.25}$$

where μ_{o} is the permeability of free space and M_x is the magnetization along the x-direction. After a 90$^{\mathrm{o}}$ pulse, we can then write,

$$\Phi = \mu_{\mathrm{o}} A \cdot M_{xy} \sin(\omega_{\mathrm{o}} t). \tag{2.26}$$

Calculating the induced voltage, we get the following

$$emf = -\mu_{\mathrm{o}} A \, M_{xy} \, \omega_{\mathrm{o}} \, \cos(\omega_{\mathrm{o}} t). \tag{2.27}$$

This is an important result, as it states that the induced voltage is proportional to the square of the main magnetic field B_{o}^2. This can be seen by the fact that $\omega_{\mathrm{o}} = \gamma B_{\mathrm{o}}$ and $M_{xy} \propto M_{\mathrm{o}} \propto B_{\mathrm{o}}$.

2.1.4 Signal Localization

In order to generate signal, the RF B_1 field is used to manipulate the magnetization. When the B_1 field is applied, however, it is applied to the entire sample and there is no way to localize where the signal is coming from spatially. In order to introduce spatial localization, additional magnetic fields via gradient coils are used. The gradient coils induce a small linearly changing magnetic field in all three Euclidean axes, a schematic of basic gradient coil design is shown in Figure 2.5.

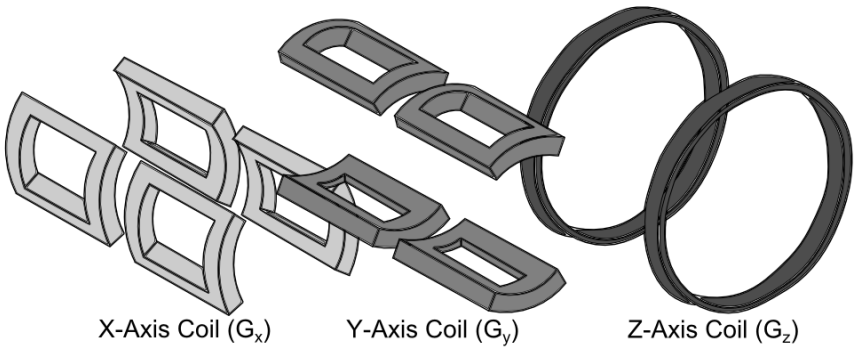

Figure 2.5: Standard basic gradient coil design to create a linearly changing magnetic field along each Euclidean axis. The x-axis and y-axis gradients are a pair of saddle coils, while the z-axis gradient is an example of a Maxwell pair.

The additional gradient magnetic fields, denoted by G_i, create a spatially dependent magnetic field whose z-components can be written as,

$$\begin{aligned} \mathbf{B}_z(\mathbf{r}) &= (B_o + G_x x + G_y y + G_z z)\hat{k} \\ \mathbf{B}_z(\mathbf{r}) &= (B_o + \mathbf{G}\cdot\mathbf{r})\hat{k}. \end{aligned} \tag{2.28}$$

Solving this via the Bloch equations, we get a signal equation where the frequency will vary depending on the spatial position,

$$M(\mathbf{r},t) = M_o e^{-t/T_2} e^{-i\gamma B_o t} e^{-i\gamma(\mathbf{G}\cdot\mathbf{r})t}. \tag{2.29}$$

The first exponential is the T_2 or spin-spin relaxation, the second exponential is from the B_o precession, and the third exponential is the gradients.

By convention, the third exponential is often written in terms of k-space,

$$\mathbf{k} = \gamma \int_0^t \mathbf{G}(\mathbf{r}, t) dt. \tag{2.30}$$

Slice Selection Gradient

As mentioned above, the B_1 field is non-specific, meaning it will excite the entire sample when applied. In order to excite only the slab or slice we would like to image we use a slice selection gradient. For the following discussion, the slice selection (SS) gradient is assumed to be applied along the z-axis.

In the presence of the SS gradient the frequency ω varies as a function of position,

$$\omega(z) = \omega_o + \gamma G_z z. \tag{2.31}$$

Thus, the slice thickness can be determined via,

$$\Delta f = \gamma G_z \Delta Z. \tag{2.32}$$

The slice selection process is demonstrated in Fig. 2.6.

In order to excite the band of frequencies corresponding the slice location we apply an RF pulse at the desired frequency at the same time the gradient is turned on. The bandwidth BW of the RF pulse is simply the Δf in Fig. 2.6. Therefore, to excite the slice shown in Fig. 2.6, the RF frequency would be equal to $f_0 + \gamma G_z Z_0$.

In order to excite only the protons within a given slice, we seek to find a B_1 that

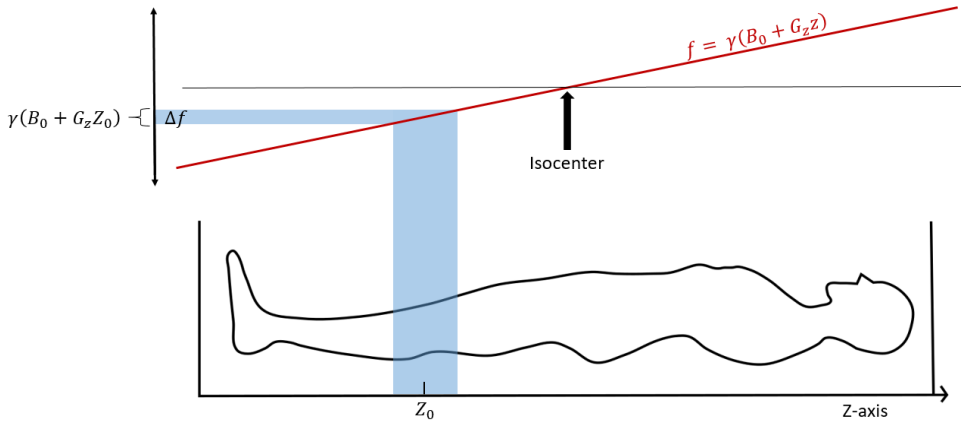

Figure 2.6: Demonstration of how the gradient, slice thickness, and bandwidth are related.

will excite a rectangular frequency profile. It is well known from Fourier analysis, that this can be achieved with a Sinc function as shown in Fig. 2.3. Mathematically,

$$\int_{-\infty}^{\infty} \text{Sinc}(t) e^{-i2\pi ft} dt = \sqcap(f). \tag{2.33}$$

Unfortunately, it would take infinite time to generate a Sinc that would create a perfectly rectangular slice excitation. Thus, there remains a trade-off between the number of Sinc lobes that are included (i.e. time) and the profile of the excited slice.

There are several RF pulses that may be used, dependent on the imaging experiment. Sinc pulses (often windowed or truncated) are a common pulse; however, other notable RF pulses include adiabatic, composite, and Shinnar-Le-Roux [56, 57].

Frequency Encoding Gradient

Slice selection allows the user to image a slice region of a subject through the use of a gradient along with the RF pulse (along the z-axis, for example). If the signal were acquired after slice selection, we would not be able to infer any spatial dependence of the spins within the imaged slice. In order to create spatial dependence along another

axis (generally taken to be the x-axis), we need to apply another gradient.

Applying a gradient along the axis within the excited slice (i.e. the x-axis) creates a spatial frequency variation, such that $\omega = \gamma G_x x$. Figure 2.7 contains a schematic of how the x-gradient creates spatial dependence.

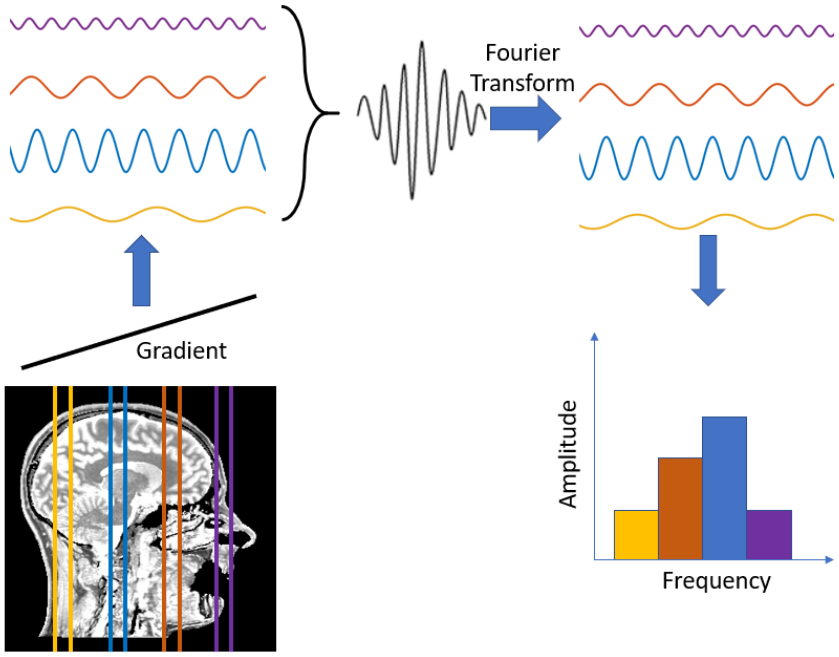

Figure 2.7: Frequency encoding creates a spatial frequency dependence along the x-axis. We are still not able to resolve the image; however, a one dimensional inverse Fourier transform does generate projections of the slice.

As shown in Fig 2.7, the signals are summed into one signal that is decomposed back into individual components. Figure 2.8 demonstrates where the acquisition occurs and how the gradient is applied. An initial dephasing gradient (negative lobe) is applied prior to the positive lobe, when the data is acquired. The dephasing gradient is used to pick the starting k-space location for the acquisition. Generally, the k-space will be read from k_{min} to k_{max} with the echo-time located within the center of the positive lobe. The frequency of samples are acquired such that $\Delta k_x = \gamma G_x \Delta t$, where Δt is the inter-sample spacing.

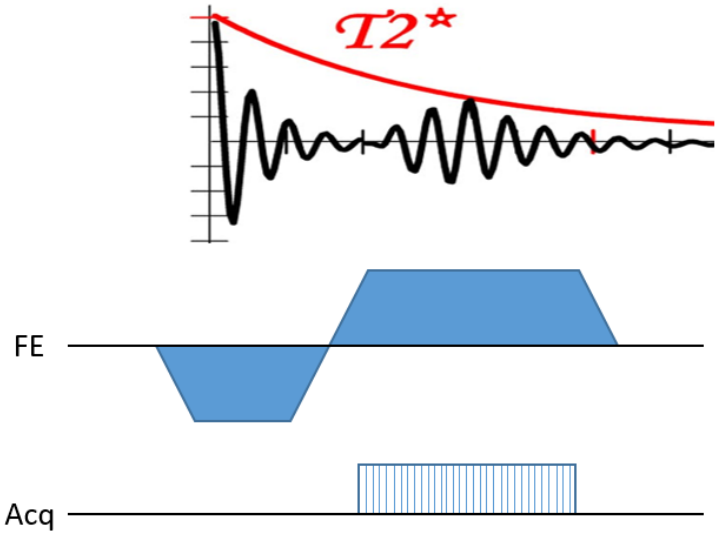

Figure 2.8: Demonstration on how the frequency is acquired and how the signal is rephased during the acquisition. The Acquisition occurs during the rephasing, where the signal is sampled every Δt.

Phase Encoding Gradient

The phase encoding (PE) gradient is required in order to create spatial dependence along the y-axis. As described above, the frequency encoding (FE) requires the acquisition of a sample every Δt; however, phase encoding is more time consuming as we acquire a separate FE while stepping with an incremental constant gradient. This can be reflected in an equation similar to that of frequency encoding, but instead of an increment in time, the gradient is incremented $\Delta k_y = \gamma \Delta G_y \tau$. An example of the PE gradient is shown in Figure 2.9.

Compared to FE, this is a time consuming process, as the phase encode gradient needs to be applied prior to each frequency encode and incremented every TR (to acquire the signal with a different phase increment). Thus, for an image of size 256 x 256, 256 phase encode steps are required (and therefore $256 \times TR$. When it comes to reducing the acquisition time, the PE is an important parameter to consider. As

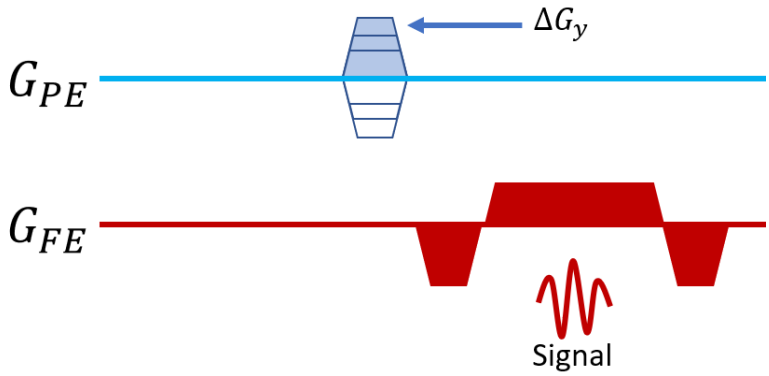

Figure 2.9: Demonstration of the PE gradient. For each ΔG_y, the FE gradient is turned on and the signal is acquired (at a rate of Δt).

will be discussed later, the PE can be undersampled, as is done with techniques such as parallel imaging and compressed sensing.

2.1.5 K-space Acquisition

Signal Equation

The general formalism for acquiring the time varying signal from the sample can be written as,

$$S(k) = \iint M(x,y)\, e^{-i2\pi(k_x \cdot x + k_y \cdot y)}\, dx\, dy. \tag{2.34}$$

The term $M(x,y)$ is the transverse magnetization and k_x and k_y are the corresponding k-space coordinates. The signal is acquired from the FID following the application of the slice selection gradient and during the frequency encoding gradient. This equation is an example of a Fourier pair, where we can extract $M(x,y)$ with the use of the Fourier transform,

$$M(x,y) = \iint S(k) e^{i2\pi(k_x \cdot x + k_y \cdot y)}\, dk_x\, dk_y. \tag{2.35}$$

k-space Sampling

The raw data acquired from an MRI system is acquired in frequency space, or k-space. As discussed previously, in order to localize the NMR signal, gradient magnetic fields are used to apply spatially dependent linear magnetic fields in each cardinal direction. The spatially dependent magnetic fields cause the protons to precess at differing frequencies and phases depending on their location within the MR bore. To ensure that we image the full extent of the required region, the Nyquist criterion must be met. The Nyquist criterion simply states that to ensure the signal is not aliased; the sampling step size must be less than the inverse of the FOV.

$$\Delta k < \frac{1}{\mathrm{FOV}}. \tag{2.36}$$

This can also be written as,

$$\begin{aligned} \mathrm{FOV}_x &= \frac{1}{\Delta k_x} = \frac{1}{\gamma G_x \Delta t}, \\ \mathrm{FOV}_y &= \frac{1}{\Delta k_y} = \frac{1}{\gamma \Delta G_y \tau}. \end{aligned} \tag{2.37}$$

Violation of the Nyquist criteria will result in aliasing, which ultimately results in artifacts occurring throughout the image.

Mathematically, this is due to the Fourier Transform of the finite sampling distribution of signal in k-space. The sampled k-space signal can be written as [58],

$$\bar{M}(k_x, k_y) = \frac{M(k_x, k_y)}{\Delta k_x \Delta k_y} \mathrm{III}_{\mathrm{2D}}\Big(\frac{k_x}{\Delta k_x}, \frac{k_y}{\Delta k_y}\Big) \sqcap_{\mathrm{2D}} \Big(\frac{k_x}{W_{k_x}}, \frac{k_y}{W_{k_y}}\Big). \tag{2.38}$$

Here $\bar{M}(k_x, k_y)$ is the sampled version of the continuous distribution of magnetization in frequency space, $M(k_x, k_y)$. The Dirac comb III_{2D} is the density of the k-space

sampling. The box car (Rect) function $\sqcap_{2\mathrm{D}}$ represents the maximum sampling range. The k-space widths W_{k_i} are defined as,

$$W_{k_i} = 2\big(k_{i_{max}} + \frac{\Delta k_i}{2}\big). \tag{2.39}$$

Taking the FT of Eq. 2.38 results in the desired image,

$$\bar{m}(x,y) = m(x,y) * \mathrm{III}_{2\mathrm{D}}(\Delta k_x x, \Delta k_y y) *_{2\mathrm{D}} \mathrm{Sinc}(W_{k_x} x)\,\mathrm{Sinc}(W_{k_y} y). \tag{2.40}$$

This equation demonstrates the importance of Nyquist. The image m is convolved with the Dirac comb, meaning there will be infinite replicas of m; this does not prove to be an issue so long as the distance is large enough that they do not overlap. In particular, if $\frac{1}{\Delta k_x} > \mathrm{FOV}_x$, then the spacing of $\mathrm{III}(\Delta k_x x)$ will be too small, causing the infinite replicas to overlap.

The Sinc function introduces ringing into the reconstructed image. As the Rect function ($\sqcap$) becomes smaller (i.e. more truncation of k-space), the more the ringing will be apparent. This is a well-known property of the truncation of an infinite Fourier series, also known as Gibbs ringing.

2.1.6 Rapid MR Imaging

There are many sequences that can be used for rapid MRI. Before delving into the rapid sequences, let us first discuss the two basic pulse sequences: gradient echo (GRE) and spin echo (SE).

There are several imaging parameters that affect the signal and contrast generated in an imaging sequence. The repetition time (TR) is the time between subsequent

RF excitations. The echo time (TE) is the time from the delivered RF pulse to the acquisition of the signal (or echo). Depending on the TR and TE times, the contrast may be T_1 weighted, T_2 weighted, or proton density weighted.

Spin Echo

As discussed previously, the transverse relaxation decays at a rate of T_2^*. This effect, however, can be altered with the use of a spin echo (SE) pulse sequence. Figure 2.10 contains a pulse sequence diagram of a SE sequence.

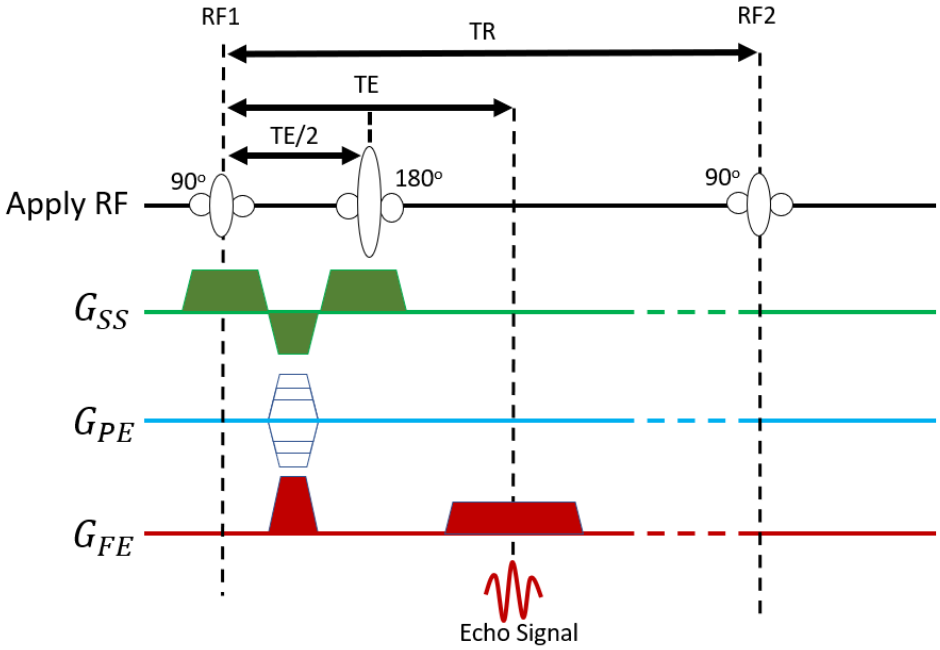

Figure 2.10: The SE sequence uses a 90° RF pulse to tip M_o into the transverse plane. After a time TE/2, a 180° RF pulse is delivered, which refocused the dephasing at time TE.

The benefit to using a SE sequence is that the images contain T_2 contrast due to the rephasing of the T_2', which reduces the susceptibility artifacts occurring at tissue interfaces.

The signal intensity for a SE sequence is

$$S(x,y) \propto \rho(x,y)\big(1 - e^{-TR/T_1(x,y)}\big)e^{-TE/T_2(x,y)}. \tag{2.41}$$

Although SE sequences typically have a high signal, the addition of the 180° refocusing pulse and long TR make the SE sequence not ideal for fast (real-time) imaging.

Gradient Echo

A gradient echo–also known as gradient-recalled echo (GRE)–sequence only uses a single excitation pulse and does not utilize a 180° refocusing pulse that was required for the SE sequence. The result of this is a much shorter echo time (TE), resulting in a contrast that will (generally) be T_2^* weighted; however, a short enough TE will minimize the T_2^* contribution allowing other contrast mechanisms to dominate, such as T_1. Figure 2.11 contains the pulse sequence diagram for a typical GRE acquisition. The RF pulse is applied along with the slice gradient, to excite the 2D plane (slice) of interest. Following the slice selection, the phase gradient is applied, and is applied for every phase encode line. The readout (or frequency) encoding is then applied, where the signal is acquired during the positive lobe.

The major benefit to the GRE sequence is that the acquisition time is much shorter since the TR values are much shorter. Typical TR values for a SE may be hundreds to thousands of ms, whereas for a GRE can be as short as 2 ms!

There are dozens of rapid pulse sequences that utilize GRE; which can be effectively split into two distinct streams: steady-state incoherent (SSI) and steady-state coherent (SSC) [59]. The SSI sequences generally spoil the residual M_{xy} every TR, whereas for SSC sequences the residual M_{xy} is refocused. Examples of common SSI sequences include fast low-angle shot (FLASH) and spoiled gradient-recalled echo

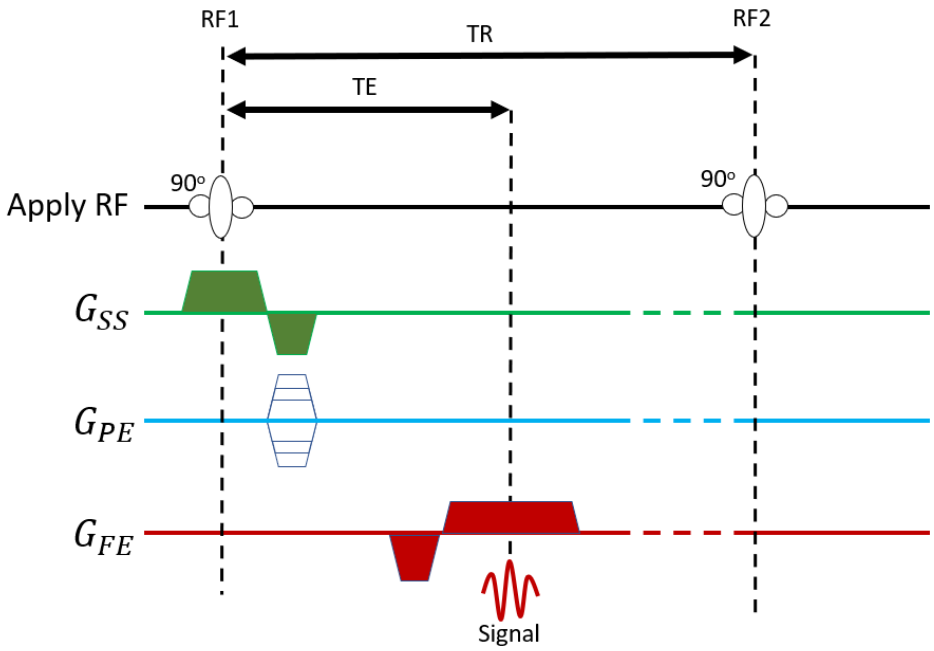

Figure 2.11: The GRE sequence uses a 90° RF pulse to tip M_o into the transverse plane. Unlike the SE, there is no refocusing with a 180° RF pulse and the acquisition occurs almost immediately after the RF pulse is delivered.

(SPGR). The sequence used to acquire the data presented within this thesis is the SSC sequence known as balanced steady state free precession (bSSFP).

Balanced Steady State Free Precession (bSSFP)

Steady state free precession is an SSC sequence that occurs when a repetitive chain of RF pulses is applied at a time TR that is much shorter than the T_2 for the tissues. This continuously refocuses the residual M_{xy} causing the "FID" and the "Echo" to merge into a continuous fluctuating signal as demonstrated in Figure 2.12. The bSSFP pulse sequence is shown in Figure 2.13

The use of the bSSFP sequence relies on the quantity known as the resonance

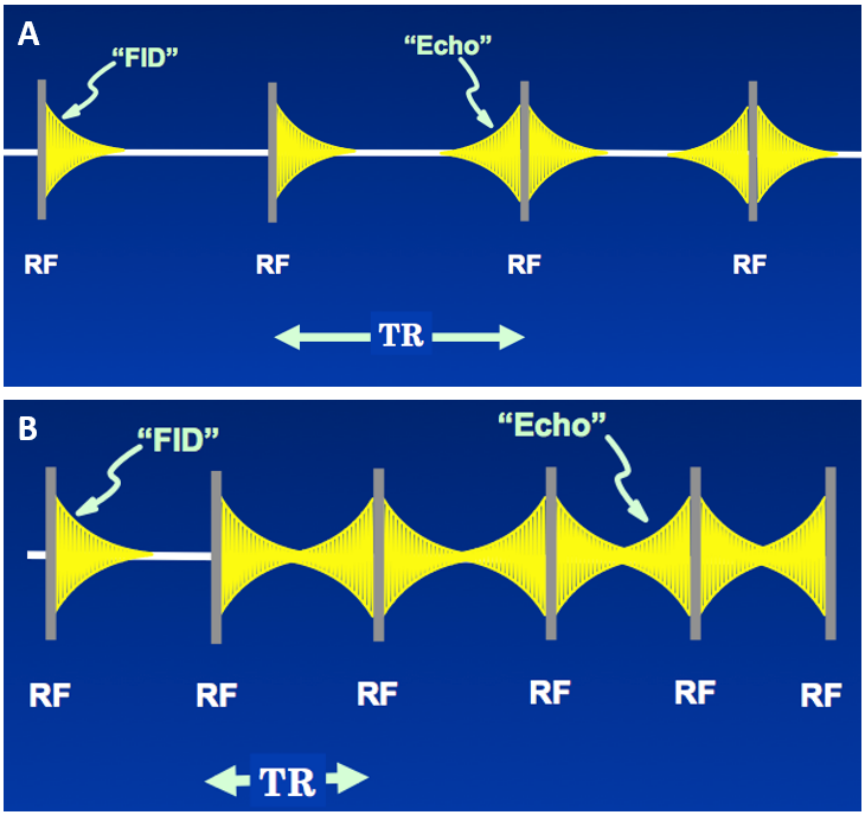

Figure 2.12: Here we demonstrate how the signal of the FID and echos (A) are merged into a continuous signal (B). Figure courtesy of Allen D. Elster, MRIquestions.com.

offset angle [51],

$$\begin{aligned}\beta &= \gamma\,\Delta B\,T_R + \gamma\vec{r}\cdot\int_0^{T_R}\vec{G}(t)dt \\ &= \gamma\frac{\Delta\omega}{\gamma}\,T_R + \gamma\vec{r}\cdot\cancelto{0}{\int_0^{T_R}\vec{G}(t)dt} \\ &= 2\pi\,\nu\,T_R.\end{aligned} \tag{2.42}$$

The integral is null for the bSSFP sequence, as the zeroth gradient moments are zero (this is where the "balanced" term comes from). In general, the β term is set to $\pm\pi$ [51] through the RF phase. This results in a larger signal (than the non-alternating case) and shifts the passband from off-resonance 180° to on-resonance

0^o [56]. When $\beta = \pi$ the signal for the SSC limit can be written as [60],

$$M_y^+(\infty) = \frac{M_0(1 - E_1)\sin(\theta)}{(1 - E_1\cos(\theta)) - E_2(E_1 - \cos(\theta))} \tag{2.43}$$

where E_1 and E_2 are (1-T_R/T_1) and (1-T_R/T_2), respectively. This is simplified in the limit where $T_R << T_2$,

$$M_y^+(\infty) = M_0\frac{\sin(\theta)}{(1 + \cos(\theta)) + (1 - \cos(\theta))\frac{T_1}{T_2}}. \tag{2.44}$$

Furthermore, the optimal flip angle is given by [51],

$$\cos(\theta_{opt}) = \frac{E_1(1 - E_2\cos(\beta) + E_2(\cos(\beta) - E_2)}{(1 - E_2\cos(\beta)) + E_1E_2(\cos(\beta) - E_2)}. \tag{2.45}$$

The β term is the phase of the successive RF pulses. When $\beta = \pi$ and in the limit where $T_R \ll T_2$, we can approximate the optimal flip angle to the following simplified expression

$$\cos(\theta_{opt}) \approx \frac{T_1/T_2 - 1}{T1/T_2 + 1}. \tag{2.46}$$

Inserting the optimal flip angle into the signal equation, we find

$$M_y^+(\infty)|_{\theta=\theta_{opt}} \approx \frac{M_0}{2}\sqrt{\frac{T_2}{T_1}}. \tag{2.47}$$

This result is the reason why bSSFP is an ideal fast imaging sequence. The ability to achieve 50% of M_0 (for liquids that have $T_2 \approx T_1$) results in bSSFP providing a high SNR compared to other fast imaging techniques [60]. Even for the case when $T_2/T_1 \ll 1$, 10%–30% of M_0 can still be achieved, which is greater than other FLASH sequences [61]. One possible downside to using bSSFP is that because the signal is

dependent on T_2/T_1, liquids and fat (which have differing T_2 and T_1 values) will have similar contrast, given the ratio of T2/T1 is similar [60].

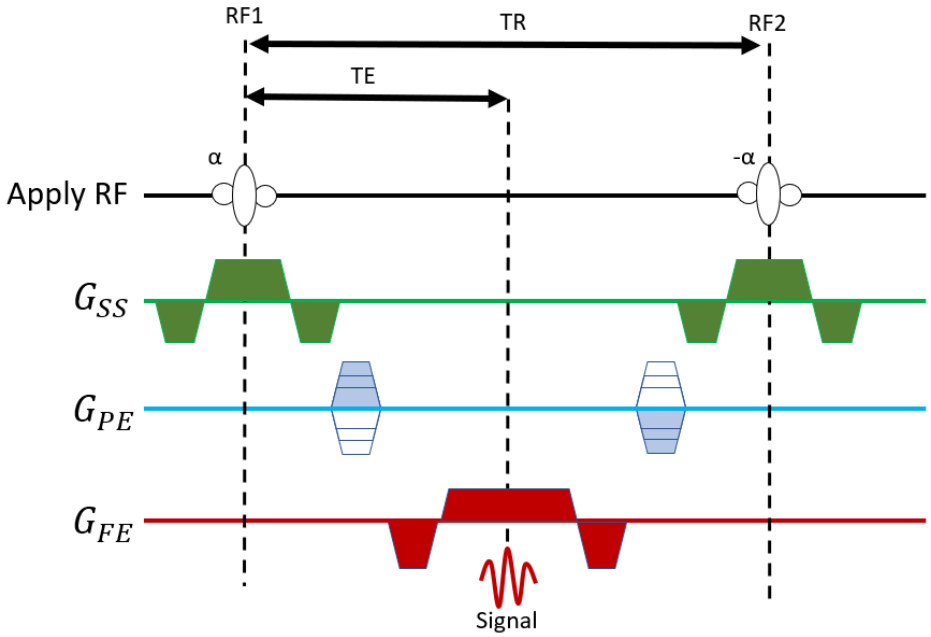

Figure 2.13: The pulse sequence diagram for a balanced steady state free precession acquisition. The slice, phase, and frequency encodings are all balanced.

2.1.7 Noise & Contrast

Signal to Noise Ratio

Every imaging technique is subjected to noise, and MRI is no exception. The signal to noise ratio (SNR) is a general measure of how well an imaging system can detect signal over noise.

In terms of imaging parameters, we can write the SNR/voxel as,

$$\frac{SNR}{\text{Voxel}} \propto \Delta x \Delta y \Delta z \sqrt{\frac{N_{acq} N_x N_y N_z}{\text{BW}}}. \tag{2.48}$$

This essentially boils down to SNR being proportional to the voxel size, the square

root of the number of phase encodes (along the y and z directions) and the square root of the read encode duration $T_s = N_x \Delta t$ (since $1/\mathrm{BW} = \Delta t$).

Although it looks as though you can increase the SNR by simply decreasing the BW, doing so requires a long acquisition time (T_s) given the inverse relationship. Long acquisition times result in increased T_2^* effects, susceptibility artifacts, as well as other detrimental artefacts such as B_o geometric distortion [62]. This is due to having more time for intravoxel phase dispersion to occur (given the small readout gradient).

In terms of the main magnetic field strength B_o, the SNR will vary depending whether the sample noise or coil noise dominates. A general equation can be written as,

$$SNR = \frac{B_o^2}{\sqrt{\alpha B_o^2 + \beta B_o^{1/2}}}, \tag{2.49}$$

where the first and second terms in the denominator are due to noise in the sample and coil, respectively. The factors α and β are constants. At clinical field strengths, the sample noise is the most dominant, resulting in [58]

$$SNR \propto B_\mathrm{o}. \tag{2.50}$$

At lower field strengths (generally $\leq$ 0.5T), coil noise becomes the more dominant factor and the SNR proportionality becomes $SNR \propto B_o^{7/4}$.

As the field strength is increased, so too does the SNR. It is for this reason that 3.0 Tesla MR systems and becoming more popular in a clinical setting. In recent years, clinical systems with field strengths up to 7.0 Tesla have been developed and installed in sites worldwide.

There are also many issues when it comes to imaging with higher fields. One which is known as the dielectric effect. As the field strength is increased, the RF wavelength is shortened creating constructive (bright spots) and destructive (dark spots) standing waves. It is important to keep in mind, that SNR is only one piece of the image quality puzzle. While being able to resolve the signal over the noise floor is important, equally important is the ability to distinguish between various tissues / structures in the acquired images.

Contrast to Noise Ratio

The contrast to noise ratio (CNR) is defined as the ability to distinguish signal from two objects (A and B). A simple CNR equation is often written as,

$$\mathrm{CNR}_{AB} = \mathrm{SNR}_A - \mathrm{SNR}_B. \tag{2.51}$$

Similar to SNR, the CNR is dependent on factors such as the pulse sequence (bSSFP, FLASH, GRE, etc.), flip angle, TE, and TR. The CNR is also dependent on the relative relaxation parameters (T_1, T_2, T_2^*) of the tissues being imaged [33]. While SNR may increase with increasing field strength, CNR is more complex. It has been shown that for some tissue / tumour combinations that a lower field will increase the CNR compared to a higher field [33].

2.1.8 K-space Undersampling

As discussed in the previous chapter, k-space is generally sampled to ensure that it satisfies the Nyquist criteria to avoid aliasing artifacts. A common method to save time in MR imaging is to reduce the number of phase encodes (PE's) required to meet the Nyquist criteria. Skipping lines of k-space coherently; however, results in a

fold-over artifact. The technique of coherently skipping PE's is commonly employed in parallel imaging. Parallel imaging is a popular technique that is often used in clinical imaging. [63] Another way to reduce the imaging time is to undersample the PE's in an incoherent manner. This more recently developed technique is known as compressed sensing [64].

Nyquist Sampling

Before discussing the effects of undersampling k-space, we should first understand how to fully sample k-space to ensure aliasing does not occur. The measurement of the continuous MR signal results in discretization via finite sampling using a Dirac comb function,

$$\mathrm{III}(k) = \Delta k \sum_{p=-\infty}^{\infty} \delta(k - p\Delta k), \tag{2.52}$$

where Δk is the k-space sampling distance. Of interest, is the Fourier transform of Equation 2.52,

$$\mathrm{III}(x) = \sum_{q=-\infty}^{\infty} \delta(x - q/\Delta k). \tag{2.53}$$

The comb function $\mathrm{III}(x)$ is convolved with the spin density $\rho(x)$ (in k-space $\mathrm{III}(k)$ is multiplied to the signal $s(k)$), which yields infinite replicas of the signal separated by $1/\Delta k$. Thus, we can write

$$\mathrm{FOV} = \frac{1}{\Delta k}. \tag{2.54}$$

In order to ensure the replicas do not overlap, the following criterion must be met,

$$\Delta k < \frac{1}{\mathrm{FOV}} \tag{2.55}$$

where Δk is the sampling spacing. This is known as the Nyquist criterion.

Coherent Undersampling: Parallel Imaging

By coherently skipping phase encode lines of k-space, aliasing occurs such that the image is folded over onto itself. This is due to mathematical properties of finite sampling as discussed in the above section. When a continuous signal is sampled into discrete points, an infinite number of replicas of the signal are generated. When the sampling step size is too large and does not meet the Nyquist criteria, the replicas are shifted onto one another resulting in the aliased signal. Figure 2.14 demonstrates how the aliasing occurs.

A requirement for all parallel imaging reconstruction techniques, is that the signal must be acquired using a multi-coil array. There are two main techniques that are used to reconstruct images that have been coherently undersampled, SENSitivity Encoding (SENSE), and GeneRalized Autocalibrating Partial Parallel Acquisition (GRAPPA).

SENSE is an image reconstruction technique that is done post reconstruction and utilizes coil sensitivity maps to aid in the unfolding of the aliased signal [40]. A sensitivity map is simply a map containing the signal strength for each of the coils used to acquire the signal. The general equation for SENSE can be written as,

$$\vec{I} = \mathbf{C} \cdot \vec{\rho} \tag{2.56}$$

where $\vec{I}$ are the aliased pixel values, $\mathbf{C}$ is the coil sensitivity matrix, and $\vec{\rho}$ are the true pixel values. This will result in linear equations equal to the number of coils N_c. It should be noted that the acceleration factor (R) cannot be greater than the number of coils. In order to calculate the true (unaliased) pixel values ($\vec{\rho}$), we need to invert Equation 2.56. The solution to this problem is through the pseudo-inverse

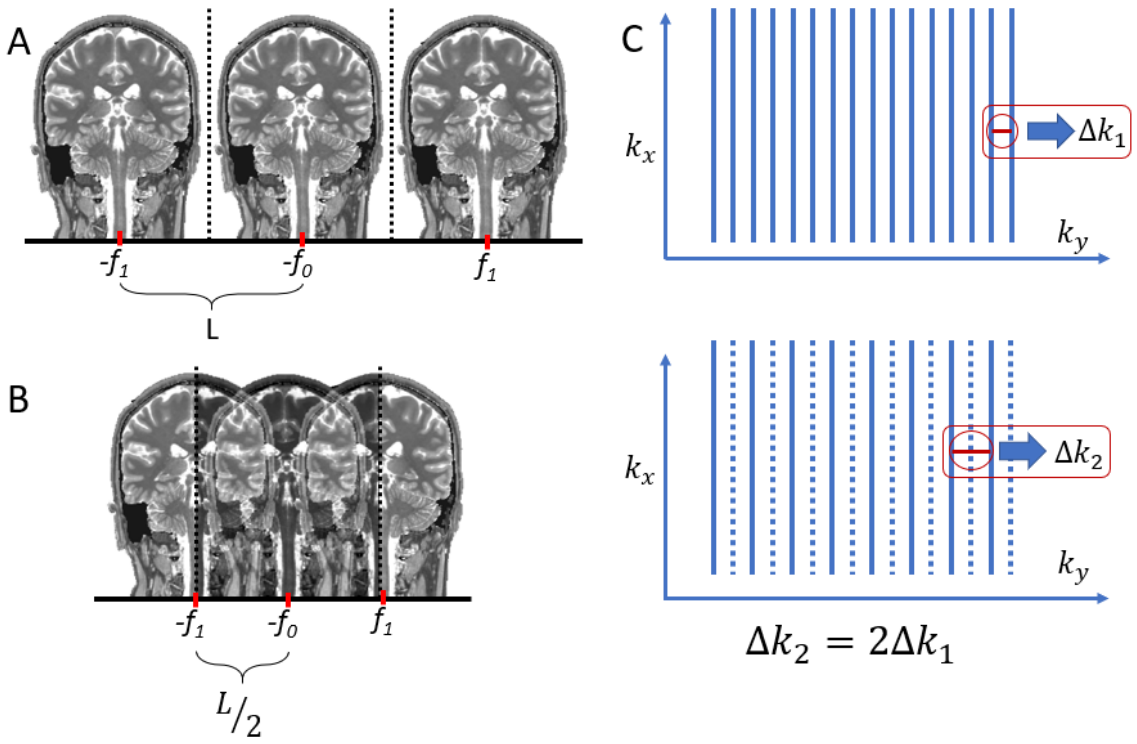

Figure 2.14: Demonstration of signal aliasing. (A) demonstrates when the Nyquist criterion is met, where no overlap between the infinite replications occurs. Meanwhile, when the Nyquist criterion is not satisfied, the replicas move closer together and aliasing (or fold-over) occurs as shown in (B). The k-space sampling schemes for (A) and (B) are shown in (C). The fully sampled k-space is sampled adequately in both k_x and k_y (k_y sampling is Δk_1), however by doubling the k-space sampling (dashed lines are zero-filled in this case), the distance between the replicas (A) reduces by half resulting in (B).

(ignoring noise),

$$F = |I - \mathbf{C} \cdot \rho|^2, \tag{2.57}$$

$$= I^\dagger I - \rho^\dagger \mathbf{C}^\dagger I - I^\dagger \mathbf{C} \rho + \rho^\dagger \mathbf{C}^\dagger \mathbf{C} \rho. \tag{2.58}$$

Note that we have dropped the vector notation for simplification of the notation. We

wish to calculate the minimum given by $\partial F/\partial \rho^{\dagger}$, which gives us

$$\rho = \{(\mathbf{C}^{\dagger}\mathbf{C})^{-1}\mathbf{C}\}I. \tag{2.59}$$

This equation is applied pixel-wise to the aliased signal to reconstruct the full FOV image.

In contrast with SENSE, which is applied in the image domain, GRAPPA is a technique that performs the correction in k-space, prior to image reconstruction. The GRAPPA technique is a generalized form of the SMASH (SiMultaneous Acquisition of Spatial Harmonics) parallel imaging technique [65]. GRAPPA does not use coil sensitivity maps, but instead uses an auto-calibration signal (ACS) to correct for the missing lines of k-space, which is generally the central region of k-space that is fully sampled. The ACS and acquired lines of k-space from each coil are used to calculate the spatial harmonic weights in order to fit the missing lines of k-space for each coil [41]. This is demonstrated in Figure 2.15.

Incoherent Undersampling: Compressed Sensing

As discussed previously, parallel imaging is a technique used to coherently undersample k-space. This creates an aliasing effect that can be corrected using algorithms such as SENSE or GRAPPA. More recently, a reconstruction technique that utilizes incoherent undersampling has been developed. Known as compressed sensing (CS), this technique has proven to be an effective method of k-space undersampling [64, 66]. The CS algorithm generally has three requirements:

(a) The data must be representable in a sparse domain (compressible),

(b) The k-space must be acquired incoherently,

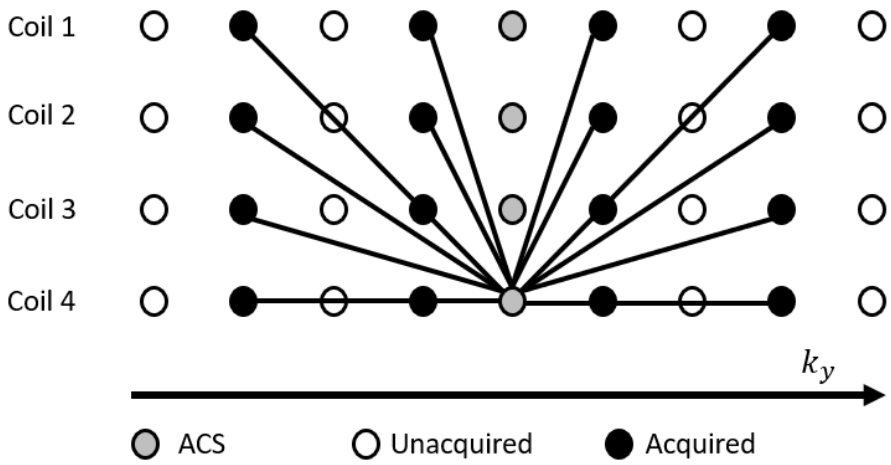

Figure 2.15: Demonstration of how GRAPPA is used to calculate the coil weights. Data from all coils are used to estimate the missing k-space data for each coil. The global weights calculated from the ACS line form a kernel, which is then used to calculate the missing k-space data.

(c) The image reconstruction is non–linear, enforcing both sparsity and data consistency.

Of the 1000's of images we view on the internet in our day-to-day activities, nearly all of them have been compressed. Without compression, the amount of data required to store these images would be very large. The goal of CS is to utilize a transform space where the data is sparse. One of the most common compression algorithms was developed by the joint photographic experts group (JPEG). The JPEG compression algorithm is based on the discrete cosine transform (DCT). Simply put, an image can be transformed into DCT space, where much of the data can be set to zero (depending on the level of compression). To take the data from DCT space back into the image, we apply the inverse DCT. Generally, nearly 90% of the data can be set to zero without a large loss in image quality [67]. A more advanced method (JPEG-2000) uses the wavelet transform. Another sparse representation can simply be the image pixel space for images of blood vessels, for example. Finite difference can also be used

(especially true in the temporal domain). The goal of these transforms is to find a space where information is sparse. Mathematically, this can be written as

$$\begin{aligned} \text{minimize} \quad & ||\Psi m||_1 \\ \text{s.t.} \quad & ||F_s m - y||_2 < \epsilon. \end{aligned} \tag{2.60}$$

This optimization problem states that we want to minimize the l_1 norm of $||\Psi m||$ which promotes sparsity. The term Ψ is an operator that transforms the image (m) to the chosen sparsifying transform (DCT, wavelet, etc.). The second term utilizes the l_2 norm to enforce data fidelity by ensuring the undersampled image ($F_s m$, where F_s is the undersampled Fourier transform), remains close to the data acquired from the scanner (y); where ϵ controls the fidelity of the data.

An important aspect for CS is the incoherent undersampling. Generally, only the phase encode is incoherently undersampled, given that the frequency encoding portion is fast enough to leave fully sampled. This is demonstrated in Figure 2.16, where a fully sampled brain image has been undersampled only along the PE direction. For 3D data sets, incoherent acquisition can be exploited further, since there are essentially two phase encoding directions. Similarly for dynamic data, the temporal domain can be undersampled such that each dynamic frame has a different sampling pattern; which can then incorporate previous data to build a more complete k-space (i.e. lower apparent acceleration).

It has been previously demonstrated that CS alone can achieve good results for real-time lung tumour segmentation at 3 T and simulated low SNR of 0.5 T [68]. Yip *et al.* developed a CS technique that utilized a secondary fidelity term that constrained the current acquired dynamic frame to the mean prior data. Yip demonstrated reconstruction speeds of under 300 ms, with acceleration factors up to 7x

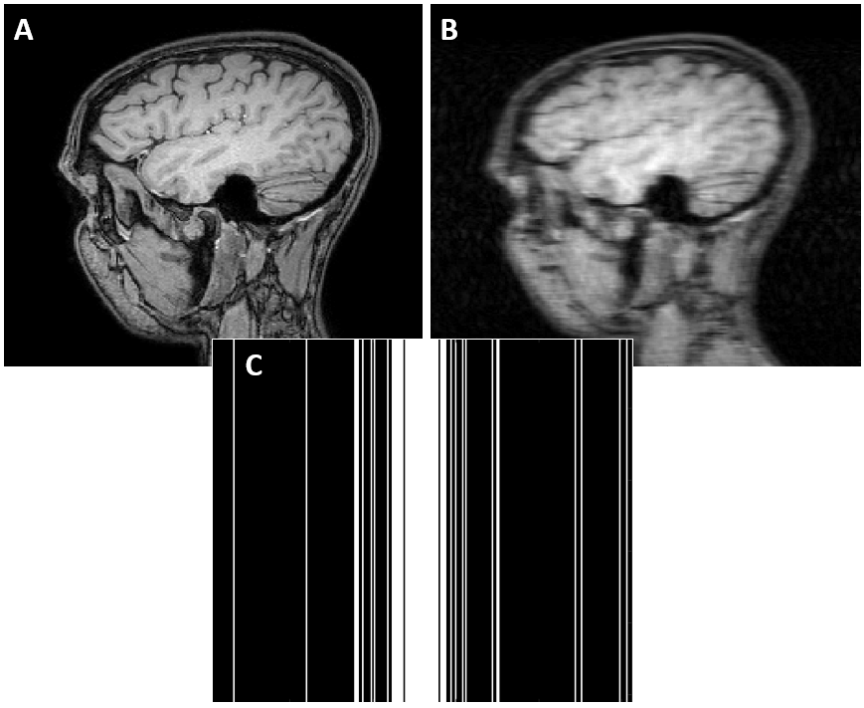

Figure 2.16: Demonstration of incoherent artifacts. (A) demonstrates when the Nyquist criteria is met (fully sampled). (B) demonstrates the noise-like artifact that occurs when the Nyquist criteria is violated incoherently. (C) demonstrates how the k-space is sampled for a general incoherent acquisition at 5x acceleration (20% of k-space is acquired).

acceleration via Cartesian acquisitions. Majumdar *et al.* developed a fast nonconvex solution to the CS problem for dynamic data with a sparse temporal difference [44]. Majumdar investigated real-time CS and demonstrated reconstruction times of 130 ms for 128×128 sized images acquired radially and mapped to Cartesian space via the non-uniform fast Fourier transform (NUFFT).

2.2 Principal Component Analysis (PCA)

Principal component analysis (PCA) is a statistical technique that uses an orthogonal transformation to convert a set of possibly correlated variables into a set of linearly uncorrelated variables called principal components. In other words, PCA reduces the

dimensionality of large data sets to filter out noise and reveal hidden dynamics. For example, PCA is often used for denoising fMRI data prior to performing statistical analysis [69]. When applying PCA to a set of dynamic MR images, the largest principal components will contain the greatest variance, which will generally be the patient motion in a particular direction (superior-inferior breathing motion, for example).

When applying PCA to a data set such as dynamic imaging data, the data should first be transformed into a Casorati matrix to make the application to PCA simple and straightforward. The Casorati matrix takes (in our case) three-dimensional data, for example $256 \times 256 \times 30$, and reshapes it into a two-dimensional space 65536×30. A schematic of the Casorati matrix transformation is shown in Figure 2.17 for p dynamic MR images of size $n \times m$.

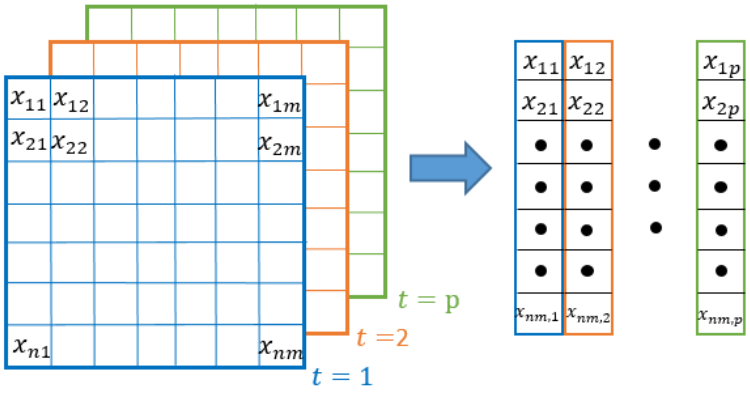

Figure 2.17: An example of how a Casorati matrix is constructed using several dynamic images. Each dynamic frame is vectorized into an $nm \times 1$ vector and then concatenated together to form an $nm \times p$ matrix.

The process of calculating the principal components is fast and rather simple to implement. Lets assume we have a matrix (B) defined as above via the Casorati matrix ($nm \times p$), then we can calculate the principal components through the use of the covariance matrix (C) or through the Gram matrix (G). The use of C or G are

ideal, as they are square symmetric matrices.

$$\begin{aligned} C &= B \cdot B^{\mathrm{T}} \\ G &= B^{\mathrm{T}} \cdot B. \end{aligned} \tag{2.61}$$

It should be noted that if B is comprised of complex numbers the transpose becomes the conjugate transpose. A simple proof demonstrates that either C or G can be used for PCA. The principal components are the eigenvectors of the matrix C,

$$Cv_i = \lambda_i v_i. \tag{2.62}$$

This can be problematic, given that the matrix C will be massive ($nm \times nm$). Alternatively, we can show that the G matrix can also be used with a small caveat,

$$\begin{aligned} Gv_i &= \lambda_i v_i \\ B^T B v_i &= \lambda_i v_i \\ BB^T(Bv_i) &= \lambda_i (Bv_i) \\ C(Bv_i) &= \lambda_i (Bv_i). \end{aligned} \tag{2.63}$$

The second last step in Equation 2.63 we multiplied both sides by the matrix B, which demonstrates that the same eigenvectors and eigenvalues exist for both G and C; however, when using G we must project the eigenvectors onto the matrix B. Thus, both G and C share the same non-zero eigenvalues. The benefit being that the matrix B will be of size $p \times p$ instead of $nm \times nm$! For large matrices (such as the Casorati matrix above), calculation of the covariance matrix may not be possible as there may not be enough memory to store such a large matrix.

2.3 Artifical Neural Networks

When writing code to solve a computational problem, we generally have input values x and some function $f(x)$ that we program in order to solve for the output value y. Artificial neural networks (ANNs) switch the paradigm, such that, given the input x and output y, we attempt to solve for the function $f(x)$. ANNs are used as a computational tool to solve problems by "mimicking" the process of biological neurons. Essentially, ANNs can be thought of as universal function approximators that tweak a set of parameters to take an input to generate a desired output. It is important to note that the parameters are the weights and biases that are updated during the training optimization. These are not to be confused with hyperparameters, which are not updated during the training optimization, and generally need to be set prior to training (or optimized separately). In general, an ANN algorithm requires: a dataset (generally input and output values), a cost function as a way to calculate the error, an optimization procedure (most commonly gradient descent), and a model [70]. The idea is to train the network using a training dataset, such that the model becomes generalized; meaning, it can perform well on previously unobserved data. Interestingly, it has been proven that there is no best machine learning algorithm and no best form of regularization [71, 72]. However, if we make assumptions about the data distribution encountered in real-world data, we can then design ANN algorithms that perform better than others.

Initially, ANNs contained only one layer and could only handle linear classification, which does not apply to many imaging problems. The following section will give a brief introduction to neural networks and their utility in image processing. A historical overview of ANNs is presented followed by the mathematics behind the perceptron, which is the first supervised learning NN. Modern ANNs will then be

presented, which use the mathematical operation of backpropagation and multiple layers to create deep non-linear networks.

2.3.1 Brief Historical Introduction

Artificial neural networks were initially developed in the 40's to mimic the process of biologic neurons based on a branch of mathematics known as threshold logic [73]. These initial neurons were named after the authors and is known as the McCulloch-Pitts (MCP) neuron, which is synonymous with the term Threshold Logic Unit (TLU). This early neuron could only perform binary operations and required that the inputs be Boolean. Furthermore, it treated all inputs equally, meaning there were no weights. If the summation of the input data did not exceed some threshold, the result would be null; if the summation did exceed the threshold, the result would be unity. Thus, MCP neurons were generally used to perform Boolean operations.

The first "real" artificial neural network was the perceptron developed in 1957 by Frank Rosenblatt [74]. This single layer neural network essentially relaxed the constraints of the MCP neuron, which allowed the neurons to "learn" from the data. The input was no longer required to be Boolean, and variable weights were introduced, the output of which would generally be a Boolean classifier (i.e. dog or cat). Rosenblatt's discovery included the development of a supervised learning algorithm where the neurons "learn" the correct weights by training the data. Given a data set of inputs and corresponding correct outputs (known as examples) the error for the neuron was calculated via the difference of the neuron output and the correct output. Training the data is a technique used to update the weights by adding the error to the previous weights.

Another important aspect is the addition of a bias ($w_0 = -\theta$), which is set equal

to the threshold value; this additional parameter allows for the network to be further trainable, by shifting where the threshold lies. Mathematically, the perceptron can be written as,

$$f(x) = \begin{cases} 1 \text{ if } \sum w_i \cdot x_i + w_0 \geq 0 \\ 0 \quad \text{otherwise.} \end{cases} \tag{2.64}$$

The main issue with perceptrons is that they are linear binary classifiers, which is not very helpful for most imaging tasks. A schematic of this simple network (for one neuron and one layer) is shown in Figure 2.18.

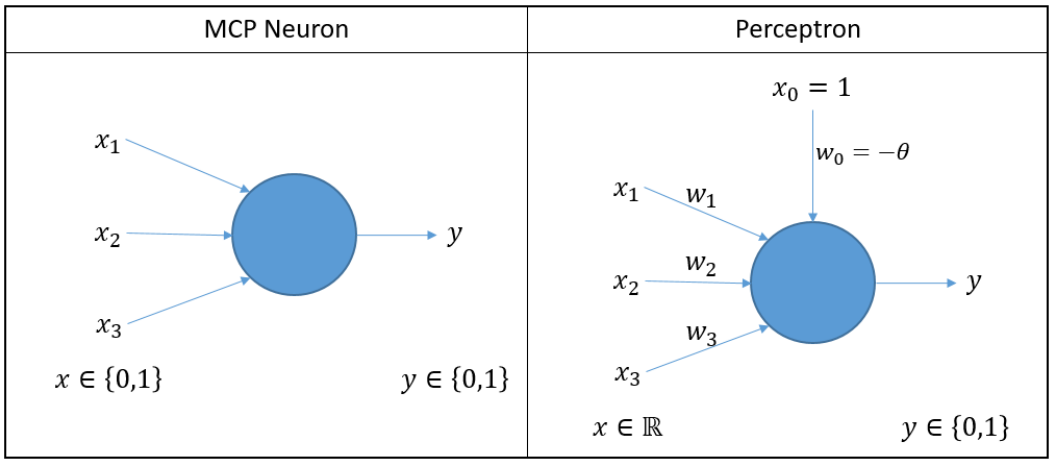

Figure 2.18: Simple neural network consisting of three neurons to demonstrate the differences between an MCP neuron and the perceptron. The MCP neuron does not contain weights or a bias parameter, whereas the perceptron does.

2.3.2 Modern Neural Networks

It is important to state that an ANN is simply a universal function approximator, that takes some input and produces the desired (known) output. Finding the optimal parameters and layers often requires experience and expertise, but ultimately the goal is to find the global minimum of some cost function by fitting the input data appropriately. It has been mathematically proven that a single layer ANN that is

arbitrarily wide (the number of parameters within a layer is defined as the width) can compute any arbitrary function [75]; a similar result has been proven for a NN that is width bounded, but arbitrarily deep [76]. Thus, a NN can be trained to compute any function (this is mathematically proven); but finding the optimal solution is entirely dependent on how well the ANN has been designed.

In general, building an ANN requires:

- A data set (the larger the better)
- Activation function (usually non-linear)
- A cost function (such as least squares)
- Optimization procedure (such as gradient descent)
- A model

These will be discussed further in the proceeding sections; however, it is important to understand the goal of training an ANN. In short, the data set (for supervised learning) will consist of input data with the associated solution (synonymous with labelled data set). The activation function is used to implement non-linearity into the model and is used to determine whether a neuron should be activated or not. The cost function determines the error of the NN calculated output compared to the actual output, which is then fed into the optimization algorithm. The optimizer then tunes the weights and biases to hopefully reduce the error during the next iteration.

Artificial neural networks often consist of several layers. When many layers are used, the ANN is then termed a deep neural network (DNN). For dense layers, each neuron from the previous layer is connected to each neuron in the current layer. After the input, each neuron is a weighted sum of the previous layers output along

with an added bias. The output from each neuron is passed through an activation function (generally a non-linear function) prior to being fed into the following layer. Mathematically,

$$a_j^l = \sigma\Big(\sum_k w_{jk}^l a_k^{l-1} + b_j^l\Big) \tag{2.65}$$

where the activation function is denoted by σ for the neuron a_j^l, where l is the layer and j is the neuron within the l^{th} layer. The weights are denoted as w_{jk}^l; l is the layer the weight is influencing, k is the neuron from the $(l\text{-}1)^{st}$ layer, which connects to the j^{th} neuron in the l^{th} layer. The term b_j^l is the added bias, which is an additional training parameter. Figure 2.19 contains an example of a dense ANN. This process continues for however many layers are contained within the network. The more layers, the deeper the network (and more trainable parameters and computational complexity).

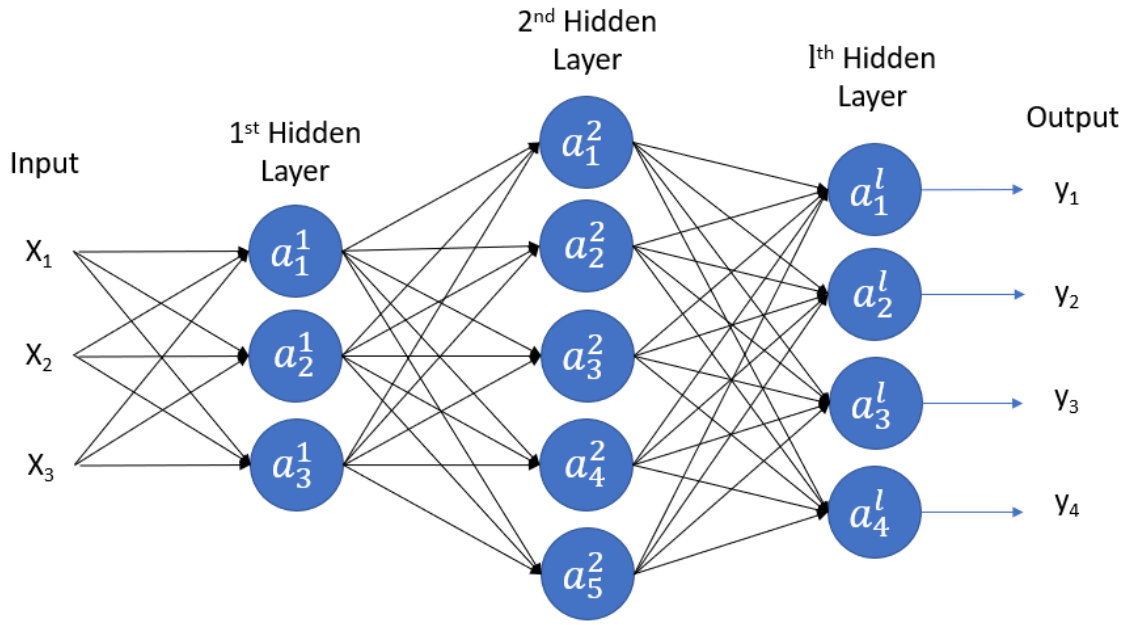

Figure 2.19: Diagram of modern neural networks. The input x_i is fed into multiple hidden layers. The cost of the output y_i and the input is calculated and used to adjust the weights via backpropagation.

Once the output is calculated, the cost C of the output and the input is calculated. This will be used to tune the weights and biases through the operation known as backpropagation. Backpropagation is a technique used to calculate the cost for each weight (and bias), and ultimately is what adjusts the parameters to a more acceptable solution. This will be discussed in detail in the following section.

2.3.3 Backpropagation

This is the mathematical framework required to transfer the error calculated at the output layer back to the previous layers, in order to update the weights. Backpropagation is ultimately just the application of the calculus chain rule. The heart of the backpropagation algorithm is the expression for $\partial C/\partial w_i$; the partial derivative of the cost function C with respect to the network weights w_i. Using the notation from Section 2.3.2, backpropagation can be developed using the following methodology [77]. Note, we will not be deriving the fundamental equations, rather simply stating them and their significance to NNs.

The goal of backpropagation is to find the associated error (δ) for each weight (and bias) with respect to the cost C calculated from the ANN output and the known (gold standard) data. This can be calculated such that

$$\delta_j^l = \frac{\partial C}{\partial a_j^l}\sigma'(z_j^l), \tag{2.66}$$

where δ_j^l is the error in the j^{th} neuron at the l^{th} layer and C is the cost function. The $\partial C/\partial a_j^l$ term on the RHS is a measure of how fast the cost is changing at the j^{th} output activation. The σ' term is the derivative of the activation function and is a measure of how the activation function σ is changing at the weighted input z_j^l (for simplicity, $z_j^l = \sum_k w_{jk}^l a_k^{l-1} + b_j^l$). We can re-write Eq. 2.66 in a more convenient

format using the Hadamard product ($\odot$)

$$\delta^l = \nabla_a C \odot \sigma'(z^l), \tag{2.67}$$

where the Hadamard product is an element-wise multiplication of matrices. We can also write the error at layer l in terms of the error in the next layer $l+1$,

$$\delta^l = \left((w^{l+1})^T \delta^{l+1}\right) \odot \sigma'(z^l). \tag{2.68}$$

Eq. 2.68 propagates the error backward through the network via the transposed weighting matrix $(w^{l+1})^T$. Taking the Hadamard product propagates the error back through the activation function. The combination of Eq. 2.66 and Eq. 2.68 allows us to compute the error δ^l for all layers in the network.

The rate of change in the cost function with respect to the network bias can be written as

$$\frac{\partial C}{\partial b^l_j} = \delta^l_j. \tag{2.69}$$

The final equation describes how the cost function changes with respect to any weight in the neural network,

$$\frac{\partial C}{\partial w^l_{jk}} = a^{l-1}_k \delta^l_j. \tag{2.70}$$

These equations succinctly describe the technique known as backpropagation. In words,

1. Input data X and randomize the weights

2. Feedfoward. Compute $w^l a^{l-1} + b^l$ for each layer $(l = 2, 3, ..., L)$

3. Output the error δ^l via Eq. 2.67

4. Backpropagate the error

5. Calculate the output via Eq. 2.69 and Eq. 2.70

The output is the gradient of the cost function which steers the solution toward the local minimum.

To show a simple example, let's look at the simple NN presented in Figure 2.20. This example breaks down a NN into each step. The input is multiplied to a weight ($x_A \times w_1 = P_1$, for example) and summed with all other inputs (multiplied to the corresponding weights), which is fed into a non-linear activation function; in this case it is modelled as a sigmoid. The output is multiplied to another weight ($y_1 \times w = q_1$, for example), which is fed into another sigmoid. Let us see how we would determine the weights w_1 and w_3, as an exercise in backpropagation.

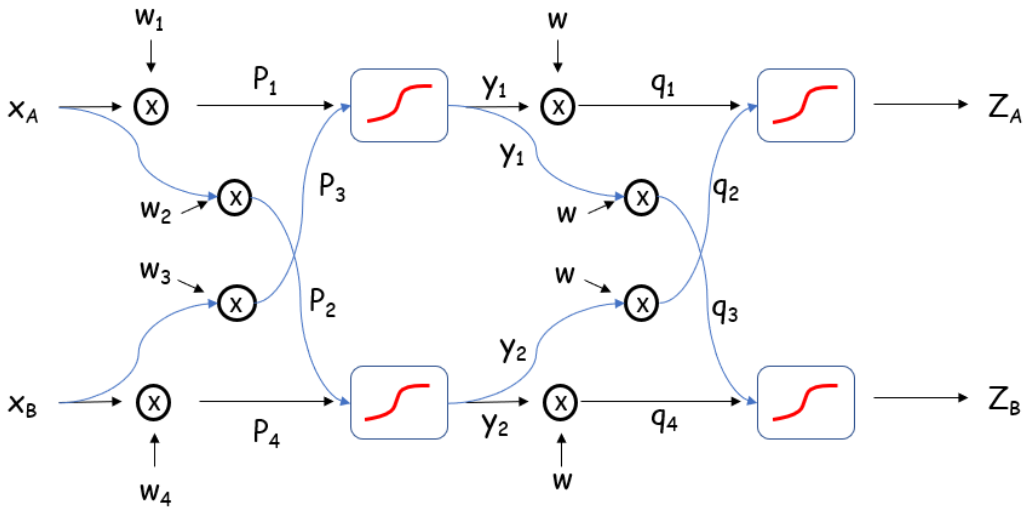

Figure 2.20: Simple NN with two inputs, two layers, and two outputs.

To start, lets calculate the derivative of w_1 with respect to the cost function C. Working our way back from the outputs, we have two avenues that will influence this

weight.

$$\frac{\partial C}{\partial w_1} = \frac{\partial C}{\partial Z_A}\frac{\partial Z_A}{\partial q_1}\frac{\partial q_1}{\partial y_1}\frac{\partial y_1}{\partial P_1}\frac{\partial P_1}{\partial w_1} + \frac{\partial C}{\partial Z_B}\frac{\partial Z_B}{\partial q_3}\frac{\partial q_3}{\partial y_1}\frac{\partial y_1}{\partial P_1}\frac{\partial P_1}{\partial w_1}. \tag{2.71}$$

Similarly for w_3,

$$\frac{\partial C}{\partial w_3} = \frac{\partial C}{\partial Z_A}\frac{\partial Z_A}{\partial q_1}\frac{\partial q_1}{\partial y_1}\frac{\partial y_1}{\partial P_3}\frac{\partial P_3}{\partial w_3} + \frac{\partial C}{\partial Z_B}\frac{\partial Z_B}{\partial q_3}\frac{\partial q_3}{\partial y_1}\frac{\partial y_1}{\partial P_3}\frac{\partial P_3}{\partial w_3}. \tag{2.72}$$

Here it can be seen that to calculate the appropriate change in the weight from a previous layer, we need to work our way back via the chain rule. Comparing Eqs. 2.71 and 2.72, we see that many of the derivatives are repeated. This greatly simplifies the computation, since we can take advantage of terms that have been previously computed.

The derivatives are with respect to the output of the activation function, as well as the cost function. These are generally simple to compute, but are completely dependent on the chosen cost function and activation function. Due to the computational simplicity, the rectified linear unit (ReLU) activation function is often used and is generally the default activation function for most problems.

ReLU Activation Function

When ANNs initially gained popularity (with the introduction of backpropagation), the most common activation function for ANNs was the sigmoid function as it was continuous, differentiable, and resembled a binary step function. This has since been replaced with the Rectified Linear Unit (ReLU). In fact, it is generally the default

activation function for most ANN code packages. This function is given as,

$$f(x) = \max(0, x), \tag{2.73}$$

and the derivative is simple to calculate,

$$\begin{aligned} f'(x) &= 0 \ \forall \ x \leq 0 \\ &= 1 \ \forall \ x > 0. \end{aligned} \tag{2.74}$$

Therefore, the ReLU activation function has an all or none response during back-propagation. The reason ReLU has become the most popular activation is that it reduces the vanishing gradient issue that is experienced when using the once popular sigmoid function. The vanishing gradient associated with the sigmoid occurs when the activation is either large or small, the gradient of the sigmoid trends to zero as shown in Figure 2.21. Furthermore, the derivative of the ReLU is constant, which

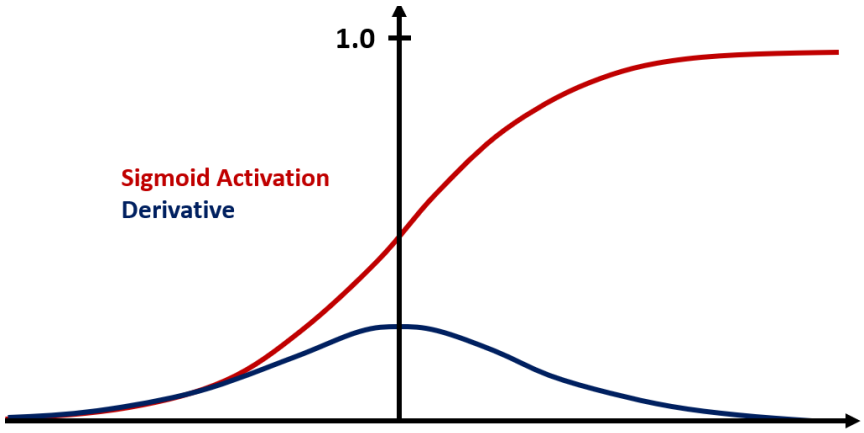

Figure 2.21: Demonstration of the sigmoid activation function, and the derivative, which results in the vanishing gradient problem. As the activation gets high or low, the gradient vanishes, resulting in poor training of the ANN.

is far more efficient to compute and hence, results in faster training (compared to the sigmoid function). Figure 2.22 shows a plot of the ReLU function along with the

derivative.

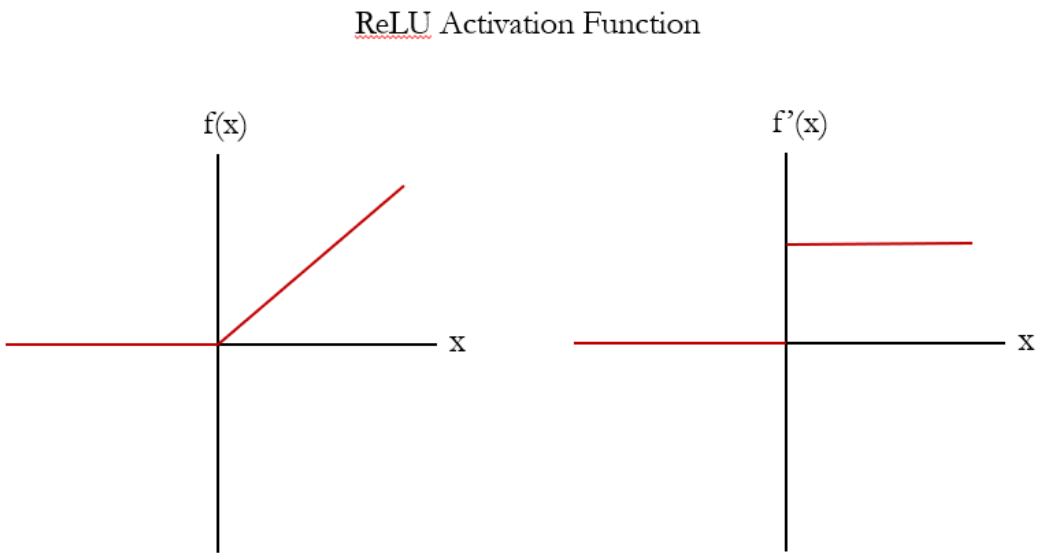

Figure 2.22: Plot of the ReLU activation function along with the derivative.

If the neuron sum is less than zero, the neuron "dies" and will no longer produce weights going forward. This has been addressed with some alternative activation functions, such as the leaky-ReLU. The leaky-ReLU introduces a small slope to the zero-portion of the function, which keeps the "dead" neurons alive. However, these other activation functions are out of the scope of this thesis and will not be addressed.

Training an Artificial Neural Network

The goal of a supervised ANN is to train it such that it can reliably learn the function required to transform the input into the gold standard data, using large data sets that consist of the input and the associated gold standard (also known as labelled data). In general, the ANN is trained to minimize the error between the input and output data by adjusting the weights and biases of the neurons. However, there needs to be some way to test how well the ANN is learning with "real world" data (data that is not used within the training data). This is done using test data, which is a

portion of the data that is not included into training the ANN. The goal is to simulate unseen "real world" data, to help aid the training. Therefore, the output from the trained ANN will be compared to the test data, giving a more realistic measure of how well the ANN is learning. Depending on the size of the data set, the input data is separated into an 80/20 split, where 80% is used for training and the remaining 20% is used to test the NN; however, this is highly dependent on the task. The problem with this technique, is that the test data eventually gets "memorized" by the NN, resulting in a bias [70]. A common way to remove this bias is with the use of k-fold cross-validation [78]. This is a technique where the training data is split into k groups where (k-1) groups are used for training. Thus, a 5-fold CV will train on 4 of the 5 groups, then test on the one group left out. This will be done 5 times, as we can simply shift the test group across the training data. After each k-fold iteration, the errors can be averaged over the k groups. This technique is ideal as it allows you to incorporate more data for training, and greatly reduces the chances of "memorizing" the test data. A downside to this technique, is that you end up increasing your training time by the k-fold factor.

Another technique that can be used to aid the ANN training process is known as data augmentation. Data augmentation is technique where additional training data is generated by applying various transformations to the training data. These transformations may include cropping, rotation, translation, and deformation. The type of transformation applied to the training is dependent on what the CNN is modelled to do. Data augmentation essentially creates "new" training data, which is ideal when dealing with smaller training data sets.

Optimization of the Neural Network

There are two sets of parameters that need to be considered with respect to an ANN. The weights and biases that are iteratively updated through the training of an ANN are known as parameters. There are several other parameters that are chosen prior to training the ANN, which are known as hyperparameters. Hyperparameters may consist of: the optimization learning rate, number of layers, batch size, number of epochs, and activation function. The following discussion details the optimization algorithms that are used to train the weights and biases (parameters).

There are several optimization techniques that are used for ANNs such as (stochastic) gradient descent (SGD) [79], SGD with momentum [80], RMSProp [81], Adam [82], and many more. For brevity, we will focus on SGD and Adam, which improves upon SGD.

Gradient descent applied to the entire data set can be computationally expensive and slow, so often SGD is used. Stochastic gradient descent optimization separates the training data into mini-batches to compute the gradient estimate. Instead of calculating the gradient over the entire data set, it applies it to the mini-batches such that the gradient estimate G

$$G \leftarrow \frac{1}{m}\nabla_\theta \sum_{i=1}^{m} L(f(\vec{x};\vec{\theta}),\, \vec{y}) \tag{2.75}$$

where m is the number of mini-batches and L is the loss or cost function of the CNN output $f(\vec{x};\vec{\theta})$ and the corresponding true output $\vec{y}$. The CNN parameters are denoted via $\vec{\theta}$. When SGD mini-batch has a size of 1, it is known as online-SGD, which can be quite noisy when approaching the minimum. Generally, it is advised not to use a large mini-batch, but also one that is not too small [83]. The parameters

$(\vec{\theta})$ are updated such that,

$$\vec{\theta} = \vec{\theta} - \epsilon G \tag{2.76}$$

where ϵ is the learning rate.

The optimization algorithm Adam, which stands for adaptive moments, combines SGD with momentum and the RMSProp optimization algorithms. It has become one of the most popular optimization algorithms; however, it has been shown to behave poorly in some situations [84]. Nevertheless, it remains the default optimization technique for most CNNs. In short, Adam updates the parameters based on both the first and second moments;

$$\begin{aligned} P_1 &\leftarrow \rho_1 P_1 + (1 - \rho_1) G \\ P_2 &\leftarrow \rho_2 P_2 + (1 - \rho_2) G^2, \end{aligned} \tag{2.77}$$

where ρ_1 and ρ_2 are generally taken to be 0.9 and 0.999; respectively. The parameters are then updated such that

$$\vec{\theta} = \vec{\theta} - \epsilon \frac{P_1}{\sqrt{P_2} + \delta} \tag{2.78}$$

where δ is a small constant (usually 10^{-8}) and ϵ is the step size (learning rate).

2.3.4 Artificial Neural Network Models

When developing a supervised artificial neural network (ANN) you will generally have some labelled input and output data, and want to find a function capable of calculating the output from the given input. There are several ANN models that can be used and are dependant on the task. The most basic neural network models use dense layers. A dense layer means the each neuron from the previous layer is connected to the neurons of the next layer. There dense networks were used almost

exclusively early on. Figure 2.23 contains a schematic of a dense NN layer. Dense

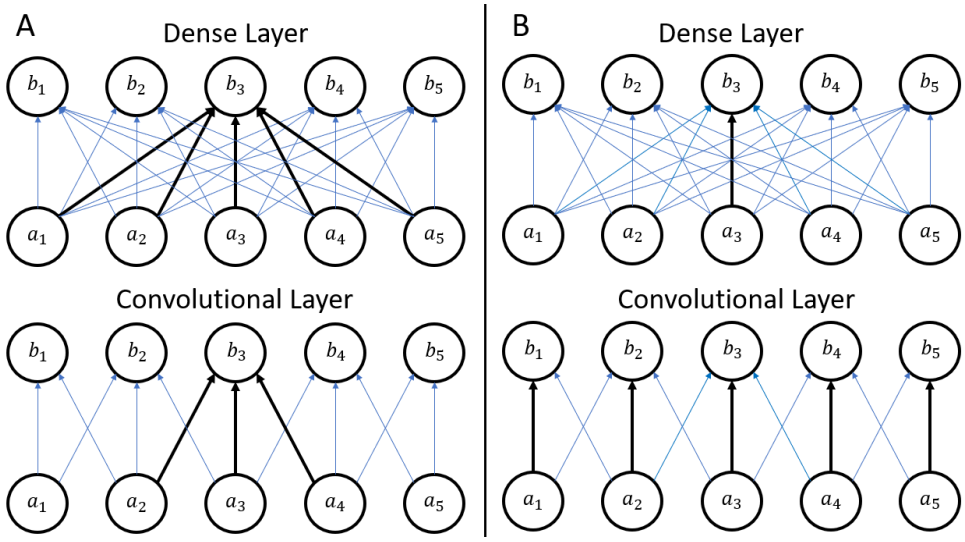

Figure 2.23: Schematics displaying the dense layer configuration (top) and the convolutional layer configuration (bottom). (A) contains an example of how the neurons are connected, (B) demonstrates how convolutional layers utilize tied (or shared) weights.

layers quickly become computationally expensive when evaluating images, especially when considering standard image sizes today. A dense layer of a 256×256 image would result in $(256^2)^2$ weights! Initially developed in 1998 by LeCun [85], it was not until 2012 when convolutional neural networks (CNNs) were found to be superior for accurate classification of images, that CNNs gained in popularity [86]. The use of CNNs are advantageous as they have built-in invariance to translation or local distortions (to some degree), which dense layers lack [85]. CNNs have the added benefit in taking advantage of local pixel correlations. A recently published review on deep learning in medical imaging and radiation therapy demonstrates that CNNs are utilized in nearly all aspects of medical imaging (segmentation, organ detection, tissue characterization, registration, etc.) [87].

The work presented in Chapters 4, 5, and 6, focus on the convolutional neural network (CNN) model, as it is ideal for the analysis of imaging data [70]. As shown

in Figure 2.23, CNNs require far fewer weights and therefore are much faster to train than traditional dense ANN models. A CNN uses several convolutional filters (kernels) to pull out the features of interest. In the case of classification, the kernels (which are a group of weights) find abstract features of the image training dataset. The weights of the kernels are updated through the same process of backpropagation as discussed previously. The CNN model requires knowledge of how many convolutional kernels are to be used and the size of kernels. A CNN is generally made up of several layers, each of which is convolved with a number of kernels (of a given kernel size). These hyperparameters can be determined in various ways, such as a manual search, grid search, or random search. Generally, initial hyperparameter values are determined from previous studies that have demonstrated success in their implementation.

A few other common ANN models are perceptrons [74], recurrent neural networks [88], general adversarial networks [89]. Perceptrons are generally considered the first generation of neural networks, and consist of a single layer feeding into a single output; however, multi-layer perceptrons (MLPs) have also been developed. MLPs may be used for tasks such as classification or regression. Recurrent NNs (RNNs) are most prevalent for speech recognition [90], such as those used for Google and Apple. RNNs have to ability to "remember" previous inputs making them ideal for time-series data such as speech analysis. Finally, general adversarial networks (GANs) use multiple NNs to work together to generate new data. One NN generates data, while the other NN is tasked with judging the data to ensure it appears natural. These have become famous by the scientists at NVIDIA for creating life-like portraits of people who do not exist [91]! GANs have even been used to create portraits of people based on the sound of their voice [92].

2.3.5 Convolutional Neural Networks for MRI

As mentioned above, ANNs and in particular, CNNs are becoming common for use in MRI analysis. It has been used for image segmentation [93], super-resolution [94, 95], classification [96], and image reconstruction [97, 98, 99]. CNNs have also been used to reduce the dosage of contrast enhancing gadolinium by an order of magnitude for contrast-enhanced brain MRI, without sacrificing image quality [100]. It has been demonstrated that CNN reconstructed cardiac studies provide similar results to other "state-of-the-art" techniques, such as kt-SLR and low-rank sparse (L+S) [97]. CNNs are optimal for imaging studies, as they learn features from the input data [101]. For example, the first layer may typically represent the presence or absence of edges throughout the images. The second layer may detect motifs by arrangements of edges. The third layer may combine the motifs into larger combinations corresponding to parts of familiar objects throughout the images.

As briefly discussed above, there are several hyperparameters specific to CNNs. These depend on the task being investigated. The kernel size, or size of the convolution operation, is often chosen to be sufficiently small to detect small features in the image data set. The convolution stride, or how many pixels the convolution jumps, is another adjustable hyperparameter (this is almost always taken to be unity). Kernel dilation is another hyperparameter that can be modified. Typically, a convolution is taken to be a region of pixels adjacent to one another; however, the kernel can be dilated to allow separation between the convolved pixels. Zero padding can also be used to ensure the convolved image is the same size (or larger) as the image the convolution is applied to. Figure 2.24 contains examples of various convolutional operations.

In order to benefit from the use of ANNs for MRI reconstruction (or any other

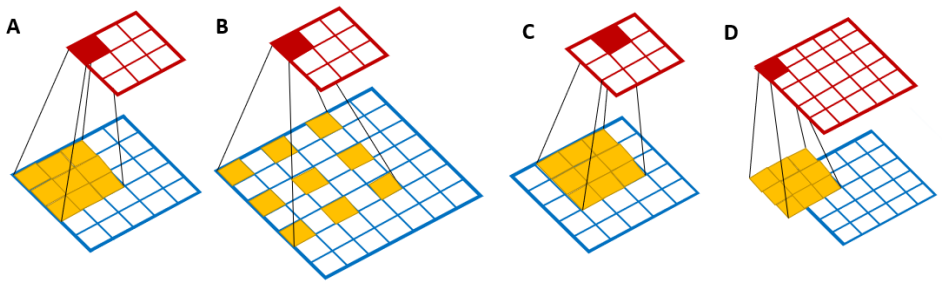

Figure 2.24: Examples of various convolutional operations. (A) example of a 3x3 kernel, (B) dilated convolution (with dilation equal to 1), (C) example of kernel (A) with a stride equal to 1, (D) "same" padding to ensure the output image is the same dimension as the input.

analysis), the use of graphical processing units (GPUs) is generally required. The GPU has the ability to perform many of the mathematical operations in parallel, resulting in a vast speed up in training time. A limitation for the use of ANN is that it is highly dependant on the training data. For studies that aim to have a generalized model (that can be applied to multiple patients, for example), having hundreds, if not thousands of imaging data sets are required. Having large data sets, however, also greatly increases the computational complexity and time to train the ANN. As will be discussed in the upcoming chapters, I developed a CNN model that uses data only from one patient, to reconstruct aliased data from the same patient at another imaging session. This is known as a patient-specific CNN, which is ideal for dynamic data, where several temporal frames can be used for training the CNN.

Chapter 3

Compressed Sensing and Principal Component Analysis (CS-PCA)

Parts of this chapter have been adapted from a published article: Dietz B, Yip E, Yun J, Fallone BG, Wachowicz K. Real-time dynamic MR image reconstruction using compressed sensing and principal component analysis (CS-PCA): Demonstration in lung tumor tracking. Medical Physics. 2017;44(8):3978–3989 [102].

3.1 Introduction

Real-time adaptive radiotherapy requires continuous imaging throughout treatment to track and monitor the tumour over time. Tracking a tumour in real-time may allow for decreased dose to organs at risk surrounding the tumor by minimizing the treatment margin within the planning target volume (PTV). Adaptive radiotherapy is ideal for tumours situated near organs of motion, such as the lung, due to the potential for reduced treatment margins that are conventionally used to compensate

for motion. The development of hybrid linac radiotherapy systems combined with magnetic resonance imaging (LMR) [27, 24, 25, 26], opens the path towards real-time MR guided adaptive radiotherapy [36]. The use of MRI is ideal for real-time imaging: aside from the excellent soft tissue contrast, it does not contribute any additional dose to the patient.

It has been recommended by the American Association of Physicists in Medicine Task Group (AAPM's TG-76), that real-time adaptive radiotherapy should have a total system latency (image acquisition, reconstruction, contouring, and MLC collimation) under 500 milliseconds [13]. The total system latency goal that our group aims to achieve is 250 milliseconds, or four frames per second (FPS). The focus of this study is to minimize the time required for the image acquisition and reconstruction processes. Previous work by our group has already considered the contouring speed, and to some degree the image acquisition and reconstruction using a prior-data assisted compressed sensing (PDACS) [103, 68].

Unfortunately, MRI is an inherently slow imaging modality in comparison to real-time x-ray techniques. Image reconstruction times can be lengthy and for this reason, many real-time reconstruction techniques, such as k-t sparse techniques, are performed offline [104, 105, 106]. Offline methods refer to reconstruction techniques that are performed after the data has been acquired; whereas online methods refer to techniques that can reconstruct the data during imaging. Real-time is a special case of online method, where the images are reconstructed as they are acquired with minimal latency, allowing the image data to reflect the current status of the subject. Real-time methods ideally have reconstruction times on the order of tens of milliseconds and have clinically useful images (from a tumour tracking perspective). Parallel imaging (PI) techniques, such as sensitivity encoding (SENSE) or generalized autocalibrat-

ing partial parallel acquisition (GRAPPA), require several channels in order to take advantage of coil geometry; however, LMR systems may not be capable of having enough coils to allow for highly accelerated PI. Current fast imaging techniques, such as balanced steady-state free precession (bSSFP), can acquire images within about 275 ms. By imaging faster (acquisition and reconstruction) we can benefit from: (1) an increased frame rate, (2) an increased image resolution, (3) an acquisition of multi-slice orthogonal planes or full 3D volume while keeping the frame rate minimal. To achieve real-time online imaging with low latency and fast frame-rate, the MR data acquisition needs to be undersampled. Undersampling the data; however, violates the Nyquist criteria resulting in unwanted artifacts throughout the image.

Compressed sensing (CS) is a technique used to reconstruct a signal from undersampled data [64]. For compressed sensing to be successful, two conditions must be met: the data must be compressible in a known sparse domain, and the undersampled data must be acquired in an incoherent manner. The method of CS essentially states that if data can be represented in a sparse domain, or can be transformed into a known sparse domain, then the incoherently undersampled signal (that violates the Nyquist criteria) can be recovered. Incoherent sampling results in an artifact resembling noise to appear throughout the image, which can be removed by transforming the data into a sparse domain. For the present study, principal component analysis (PCA) is used as the sparse domain. Previous studies have demonstrated that CS can reconstruct images at 300 ms per dynamic frame [68, 44]. The latency caused by this reconstruction (along with the k-space acquisition) is still too time consuming for real-time radiotherapy. Further, the aforementioned CS techniques are limited in the k-space acceleration they can achieve.

Principal component analysis (PCA) is a common data analysis technique that is

often used to reduce the dimensionality of a data set. The dimensionality is reduced by finding the orthogonal basis set that orders the dimensions by redundancy. The least redundant dimensions, or principle components, can be attributed to noise in a data set [107]. Examples of PCA being applied to images include the generation of eigenfaces for facial recognition, and preprocessing resting-state functional MRI data [108, 69].

A recent study by Zong *et al.* investigated the combination of compressed sensing (CS) and principal component analysis (PCA) for the use in reconstructing undersampled images [109]. Zong demonstrated that static undersampled images could be reconstructed from a database of similar images, using CS-PCA with reconstruction times of around one second. Using a similar CS-PCA methodology, a real-time reconstruction algorithm has been developed capable of reconstructing undersampled dynamic images in tens of milliseconds. This speedy reconstruction is an essential for real-time adaptive radiotherapy treatment.

The study presented proposes a novel CS-PCA reconstruction method, capable of real-time reconstruction for adaptive radiotherapy for use with our hybrid Linac-MR system. The algorithm was evaluated using retrospective data acquired from six patients with non-small cell lung cancer. Undersampled reconstruction was investigated at several acceleration factors. The acceleration factor indicates by how much the data has been undersampled; an acceleration factor of 2 specifies that 50% of the fully sampled data is used for reconstruction (4x = 25%, etc.). The undersampled reconstructions were quantitatively evaluated using several metrics. Using a neural-network contouring software developed by our group, the contoured tumours calculated from the undersampled data were compared to the original contours derived from the fully sampled data [103]. The contours were compared using the dice

coefficient (DC) [110], normalized mean square error (NMSE) [111], root mean square error (RMSE), mean absolute percentage error (MAPE), and structural similarity index (SSIM) [112]. Furthermore, the centroid displacement of the tumour contours was calculated to measure any segmentation drift in the accelerated reconstructions.

3.2 Methods

3.2.1 Incoherent Sampling

In order to utilize compressed sensing, it is important that the MR k-space data be acquired in an incoherent manner. We used a pseudo-random sampling pattern that uses a Monte Carlo sampling pattern weighted by a probability function. [64] A distribution of random numbers equal to the number of (image space) pixels along the fully-sampled PE direction are generated, if the random value at a given location is below the probability density function, that PE line is sampled. If the number of sampled PEs does not equal the required value, the random numbers are iterated and the process repeats. The probability function enforces more sampling at the center of k-space, which contains the most energy (the low spatial-frequency components). Figure 3.1 contains an example of an incoherently undersampled phase-encode scheme at 4x acceleration.

3.2.2 Offline PCA Development

Our CS-PCA reconstruction algorithm consists of an offline portion that uses database images to generate principal components (PC's), which take on the order of 10 ms to compute, followed by online real-time reconstruction. Both offline and online portions of the algorithm are implemented in (complex) k-space. Figure 3.2 contains a flow

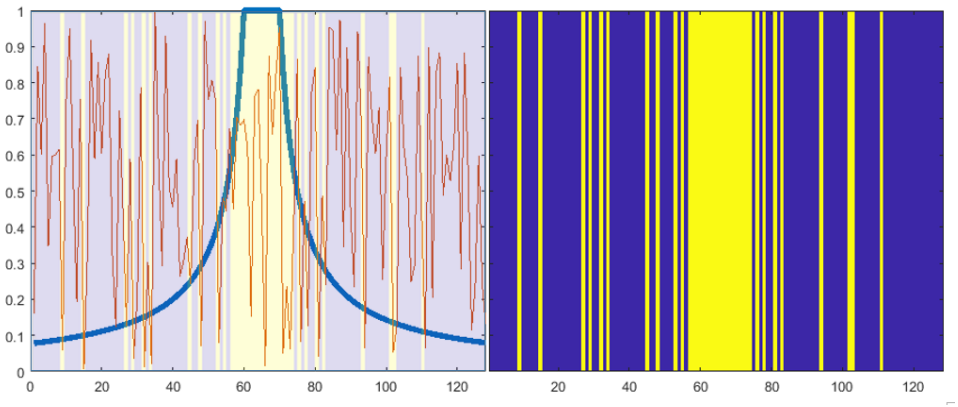

Figure 3.1: The left side of the figure contains the probability density function, which is developed to ensure the central portion (low-resolution) of k-space is always acquired. Then, we repetitively use randomly calculated values (via MATLAB rand) to find the acquired lines. If the random numbers (red) are below the PDF, we sample. This is indicated in the right side of the figure by the yellow lines. The process of generating random numbers continues until we find the number of lines required for a given acceleration.

chart of our proposed algorithm. All data was processed offline using commercially available software (MATLAB 2015a, The MathWorks Inc., Natick, MA).

Prior to real-time imaging (but within the same imaging session), a database of fully-sampled dynamic frames are acquired of size $M \times N$. The database will be used to guide the reconstruction via PCA decomposition. Let the database $\mathbf{DB}$ denote a Casorati matrix of fully-sampled k-space frames, i.e. $\mathbf{DB} = [\mathbf{DB}_1, \mathbf{DB}_2, ..., \mathbf{DB}_J]$, where J is the number of k-space frames in the database and $\mathbf{DB}_i$ is an $MN \times 1$ vector. In general, the Casorati $\mathbf{DB}$ matrix will then be of size $MN \times J$. The mean k-space is calculated and subtracted from each $\mathbf{DB}_i$ (demeaned), which is denoted as $\mathbf{DM}_i$. A J x J Gram matrix $\mathbf{G}$ is calculated,

$$\mathbf{G} = \frac{1}{(J-1)} \mathbf{DM}^{\mathrm{T}} \times \mathbf{DM}. \tag{3.1}$$

The eigenvectors and eigenvalues are then calculated from the Gram matrix. See

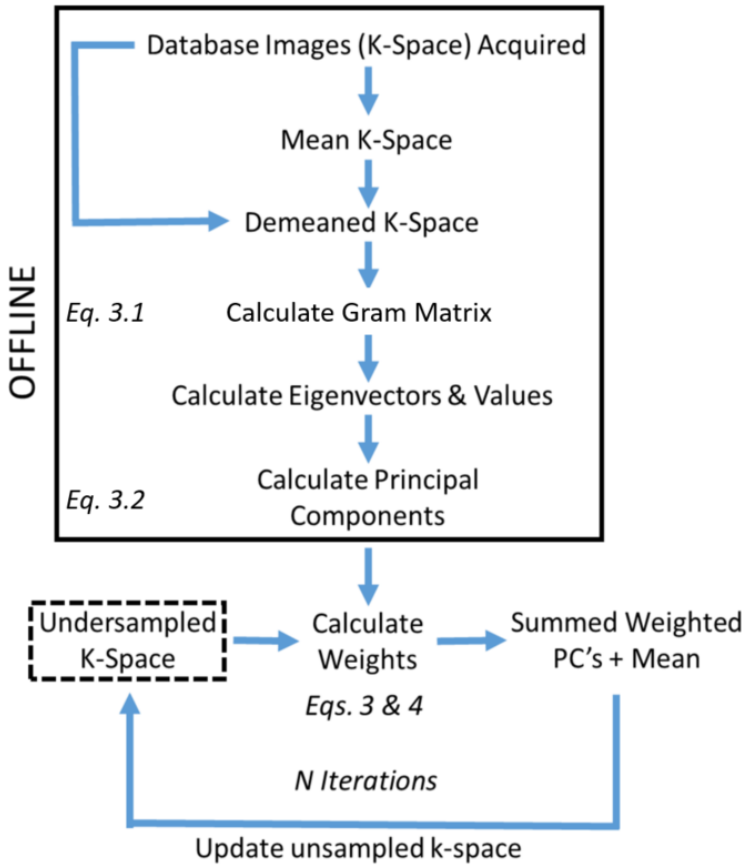

Figure 3.2: Flow chart of the CS-PCA algorithm, which is completely implemented in k-space. The offline portion of the algorithm is to be calculated immediately prior to the acquisition of the undersampled accelerated data. The CS-PCA algorithm iterates N times to update the weights, which are used to calculate and update the unsampled portion of k-space.

Section 2.2, on how the Gram matrix is related to the covariance matrix. To calculate the PC's, **DM** is multiplied by the eigenvectors (**E**) and divided by the sum of the eigenvalues ($\mathbf{e}_i$)

$$\mathbf{PC} = \frac{\mathbf{DM} \cdot \mathbf{E}}{\sum_i \mathbf{e}_i}. \tag{3.2}$$

Which results in PCs that are of the size $MN \times J$. Figure 3.3 contains six computed PC's for one patient (that have been Fourier transformed into image space for visualization purposes), where it is evident that the higher PC's contain more noise and

therefore less variation.

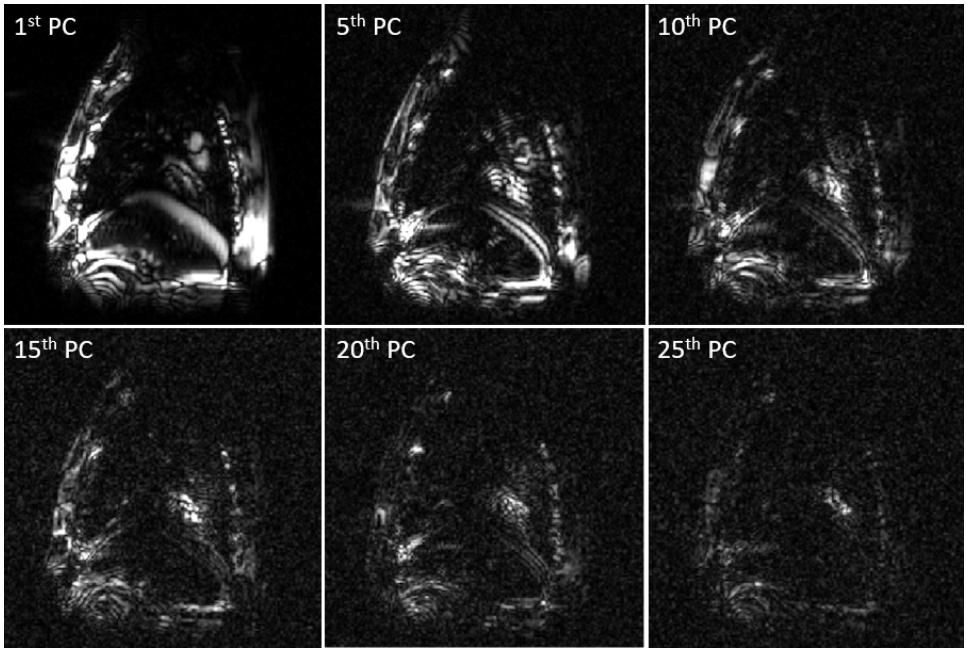

Figure 3.3: Example unweighted PC's for one patient Fourier transformed into image space. Notice how the noise becomes more prevalent in the higher PC components. For this reason, thresholding the PC's help decrease the noise in the reconstructed image by throwing away the noisy components.

3.2.3 Online Real-Time Application of PCA for Undersampled Reconstruction

Everything up to this point is calculated offline, prior to real-time imaging. Now is when we would acquire accelerated (undersampled) k-space data, denoted as **USD**. The undersampled data (**USD**) is zero-filled to be the same size as the fully-sampled data (and hence, **PC**). First, we must demean the **USD** using the mean database that was calculated offline, which we will denote as $\mathbf{USD_{DM}}$. It is important to note, that the accelerated data should not deviate far from the mean database. Each principal component is initially multiplied by the demeaned undersampled $\mathbf{USD_{DM}}$

and a vector of weights ($\mathbf{W}$) is calculated

$$\mathbf{W}_i = \mathbf{PC}_i^{\mathrm{T}} \cdot \mathbf{USD_{DM}}. \tag{3.3}$$

The size of $\mathbf{W}$ is dependent on the number of principal components (which is determined by the J images contained in $\mathbf{DB}$). At this point, we truncate $\mathbf{W}$ by keeping only values above a threshold which we have defined as,

$$\mathbf{W}_i = 0 \ \forall \ \frac{\mathbf{W}_i}{\sum_i \mathbf{W}_i} < T, \tag{3.4}$$

where T is a user defined threshold. The PC's are multiplied by their corresponding weights, summed together along with the mean k-space, and the missing lines of k-space from $\mathbf{USD}$ are filled. This process continues iteratively, such that the $\mathbf{USD_{DM}}$ from above is replaced with the new (fully-filled and demeaned) k-space. The PCA iteration updates the value of the weights using the updated k-space (where unacquired k-space lines are estimated with the previous PC fit). The number of iterations can be determined using a stop condition (i.e. when change between iterations becomes small); however, we chose to use an empirically defined value to ensure consistency of reconstruction time.

3.2.4 Retrospective *in vivo* data

Six retrospective data sets of non-small cell lung cancer patients were used to test our CS-PCA reconstruction technique. These data sets contain fully sampled dynamic images of free breathing cancer patients containing lung tumours. Lung tumours were chosen as they are often in regions where large motion occurs, due to respiratory processes. For this reason, the PTV margins required to treat these tumours can

be large [113]. Tracking the tumour in real-time will allow for decreased margins, sparing healthy tissue that would otherwise have a large radiation dose.

From the fully sampled k-space, we applied a PDF weighted Monte Carlo derived binary mask to mimic an incoherent sampling pattern commonly utilized in compressed sensing [64]. The incoherent sampling mask was applied for acceleration factors ranging from 2x to 10x. An acceleration factor of 2x implies that 50% of the k-space has been masked (set to a value of zero); an acceleration factor of 10x implies that only 10% of the fully sampled k-space is acquired. Having the fully sampled data allows for a quantitative comparison of the undersampled reconstructed data using our CS-PCA technique.

3.2.5 MRI data acquisition

A total of six patients with non-small cell lung cancer were recruited with approval from the institutional board of ethics. The patients were imaged using a 3T Philips Achieva Scanner (Philips Medical Systems, The Netherlands) for three minutes under free breathing. A 2D bSSFP sequence was used to acquire 650 dynamic frames with a field of view (FOV) of 40 x 40 cm^2, voxel size of 3.1 x 3.1 x 20 mm^3, TR/TE = 2.2/1.1 ms, resulting in a total acquisition time for each frame = 275 ms. The data was acquired using a 6-channel torso coil, which was combined via the scanner software using coil sensitivity maps and coil covariance measurements to output single channel combined (complex) data.

Several linac-MR systems have been reported, each with differing magnetic fields, including: 0.35, 0.5, 1.0, and 1.5 T [27, 24, 25, 26]. It is well known that lower field strengths will result in a decrease in detected signal, resulting in decreased SNR. To investigate the robustness of CS-PCA to the presence of increased noise, we increased

the noise of our 3T data by a factor of N = 6 to simulate the SNR at 0.5T ($3T/0.5T$), by adding random Gaussian distributed numbers [114]. This will be representative of the SNR one would expect at lower, given the known-to a first order approximation-linear dependency of the field on SNR. The standard deviation of the noise added was calculated such that,

$$\sigma_{added} = \sqrt{N^2 - 1} \cdot \sigma_{meas}. \tag{3.5}$$

To avoid the effect of the Rician noise distribution, σ_{meas} was calculated using the average σ in a region of zero signal in the real and imaginary images separately. Further details can be found in a previous study from our group [114].

3.2.6 Image Reconstruction

The image reconstruction code was written and implemented using MATLAB. The reconstruction time does depend on the parameters of the CS-PCA algorithm, such as the database size and number of PCA iterations. For a database size of 30 images the average reconstruction time was 9.3 ms using an Intel i7-4710HQ CPU @ 2.5 GHz. As a comparison (using the same CPU), the time for regular zero-filled FFT and CS reconstructions were on average 0.9 ms and 253 ms, respectively.

3.2.7 Split Bregman Compressed Sensing

The Split Bregman CS algorithm was chosen as a comparison method, as it is a fast CS technique with freely available MATLAB code [115]. The Split Bregman formulation solves the compressed sensing formulation

$$\arg\min_{m} \; ||F_s m - y||_2^2 + \lambda ||\Psi m||_1 \tag{3.6}$$

where m is the solution (in image space), y is the measured data, F_s is the masked Fourier transform, and Ψ is the sparsifying transform. Goldstein proposed using the Split Bregman solver to break the total variation ansiotropic form of Equation 3.6

$$\arg\min_{m} \ \mu||F_s m - y||_2^2 + ||\nabla_x m||_1 + ||\nabla_y m||_1 \tag{3.7}$$

up into steps that are easy to solve iteratively. First, Goldstein cast Equation 3.7 into the form

$$\arg\min_{m,dx,dy} \ \mu||F_s m - y||_2^2 + ||dx||_1 + ||dy||_1 + \lambda||dx - \nabla_x m||_2^2 + \lambda||dy - \nabla_y m||_2^2, \tag{3.8}$$

where we constrain $dx \rightarrow \nabla_x m$ and $dy \rightarrow \nabla_y m$. Then, the computationally efficient Split Bregman steps are the following:

$$m^{k+1} = \arg\min_{m} \ \mu||F_s m - y||_2^2 + \lambda||dx - \nabla_x m - b_x^k||_2^2 + \lambda||dy - \nabla_y m - b_y^k||_2^2, \tag{3.9}$$

$$dx^{k+1} = \arg\min_{dx} \ ||dx||_1 - \lambda||dx - \nabla_x m - b_x^k||_2^2, \tag{3.10}$$

$$dy^{k+1} = \arg\min_{dy} \ ||dy||_1 - \lambda||dy - \nabla_y m - b_y^k||_2^2, \tag{3.11}$$

$$b_x^{k+1} = b_x^k + \nabla_x m^{k+1} - dx^{k+1}, \tag{3.12}$$

$$b_y^{k+1} = b_x^k + \nabla_y m^{k+1} - dy^{k+1}. \tag{3.13}$$

The terms dx^0, dy^0, b_x^0, b_y^0, are all initialized to zero when performing the Split Bregman calculation. It was shown that this solution converges faster than alternative methods, such as the gradient descent solution [115].

The algorithm requires the user to set the number of inner and outer loops. We chose conservative values of 30 inner loops and 5 outer loops, as we were not focused

on speed for the CS method. The implemented algorithm also has three parameters that must be chosen for reconstruction. Using normalized mean square error (NMSE) as our metric for optimization, parameter space was searched for each patient at each acceleration factor, using a brute force method to find the three parameters that resulted in the smallest NMSE.

3.3 Quantitative Reconstruction Metrics

3.3.1 CS-PCA Parameters

There are three main parameters that were investigated for optimal CS-PCA reconstruction. The number of PCA iterations (for optimization of the PC weights), the PC weight threshold value (determines how many PCs are kept), and the size of the prior data (database size). To investigate these parameters, the normalized mean square error and Dice coefficients were calculated for all patients at an acceleration factor of 5x (80% of k-space is set to zero). The number of PCA iterations was varied from 1-10 using a constant PC threshold of 0.001 and database size of 30 images. The PC threshold was varied from 0.1-0.00001 keeping the number of PCA iterations at 10, and the size of the database at 30 images. The database parameter was varied from 2-100 using 10 PCA iterations, and a PC threshold value of 0.001.

3.3.2 Dice Coefficient

Using a neural-network based automatic contouring software developed by our group, the tumour segmentations (or contours) from the accelerated undersampled reconstructions can be tested for accuracy against the fully sampled ground truth recon-

structions [104]. The Dice coefficient (DC) [110] is defined as

$$DC = 2 \cdot \frac{ROI_{Full} \cap ROI_{US}}{ROI_{Full} + ROI_{US}}. \tag{3.14}$$

ROI is the region of interest (segmented tumour area in this situation) for the fully sampled data ($Full$) and the undersampled data (US). The DC is a quantitative metric of how much overlap exists; in our case, the overlap between two tumour segmentation masks. The DC is maximum at 1, meaning the segmentations are perfectly matched; whereas a DC of 0 indicates that the segmentations contained no overlapping area. The DC is indicative of image quality at a given ROI; as the number of artifacts increase due to undersampling and acceleration, the segmentation will deviate more from the fully-sampled case, yielding a lower DC. The average DC over the 620 frames was calculated for each patient, at acceleration factors ranging from 2x to 10x (from 50% to 10% of k-space). The total patient aggregate DC for each acceleration factor was also calculated.

3.3.3 Normalized Mean Square Error (Artifact Power)

The amount of artifact appearing in the undersampled reconstructions can be quantitatively determined by calculating the normalized mean square error (NMSE) [111]. An NMSE of zero indicates the images are in perfect agreement, and any value greater than zero is indicative of an increased presence of artefacts from the CS-PCA reconstruction. The NMSE is calculated by taking the square of the difference between the undersampled reconstruction and the fully sampled image, and dividing by the square of the fully sampled image,

$$NMSE = \frac{\sum (I_{Full} - I_{US})^2}{\sum (I_{Full})^2}. \tag{3.15}$$

The average NMSE over the 620 frames was calculated for each patient, at acceleration factors ranging from 2x to 10x. The NMSE was calculated over the entire image, to include all artefacts that may be present. The total patient aggregate NMSE for each acceleration factor was also calculated.

3.3.4 Root Mean Square Error

Another metric commonly used to quantify the amount of artifact appearing throughout the entire image via undersampled reconstruction is the root mean square error (RMSE). The RMSE was calculated over the entire image, to include all artefacts that may be present from the reconstruction. This metric is calculated by taking the square root of the square of the difference between the fully sampled images and the undersampled images, divided by the number of pixels,

$$RMSE = \sqrt{\frac{\sum (I_{Full} - I_{US})^2}{N}}. \tag{3.16}$$

3.3.5 Structural Similarity Index

The structural similarity index (SSIM) is a metric used for measuring image quality. The SSIM determines the quality through the computation of three terms, the luminance, contrast, and structural terms [112]. The combination of these three terms results in the SSIM,

$$SSIM = \frac{(2\mu_F\mu_{US} + C_1)\ (2\sigma_{F,US} + C_2)}{(\mu_F^2 + \mu_{US}^2 + C_1)\ (\sigma_F^2 + \sigma_{US}^2 + C_2)}. \tag{3.17}$$

Where μ_F, μ_{US}, σ_F, σ_{US} are the means and standard deviations of the fully sampled and undersampled data, and $\sigma_{F,US}$ is the cross-correlation of the images. The

constants C_1 and C_2 are typically chosen based on the dynamic range (DR) of the images: $C_1 = (0.01 \cdot \mathrm{DR})^2$, and $C_2 = (0.03 \cdot \mathrm{DR})^2$. The SSIM was calculated over the entire image, to include all artefacts that may be introduced from the reconstruction.

3.3.6 Mean Absolute Percentage Error

The mean absolute percentage error (MAPE) is another common statistic used for measuring image quality. This simple metric evaluates the absolute difference between the fully sampled data and the undersampled data, normalized to the fully sampled data. To calculate it as a percentage, the sum is multiplied by 100 and divided by the number of pixels (N),

$$MAPE = \frac{100}{N} \sum \left| \frac{I_{Full} - I_{US}}{I_{Full}} \right| \tag{3.18}$$

The MAPE was calculated over the entire image, to include all artefacts that may be introduced from the reconstruction.

3.3.7 Centroid Displacement

Another metric that is of importance is the tumour segmentation centroid displacement, which is a measure of how much the contour identified on the undersampled reconstruction has displaced from that of the fully sampled reconstruction. The centroid displacement is simply calculated as the center of mass difference between the tumour segmentations of the fully sampled reconstruction to the undersampled reconstruction. The centroid displacement combined with the dice coefficient and other quantitative metrics, give a complete description of how well the CS-PCA algorithm is performing from a real-time tracking perspective.

3.4 Results

3.4.1 CS-PCA Parameters

To empirically determine the optimal parameters, the Dice coefficient (DC) and normalized mean square error (NMSE) were calculated for all patients, at an acceleration factor of 5x, using varying parameters. Figure 3.4 displays how the various parameters effected the DC and NMSE. Figure 3.4A demonstrates that the NMSE and DC

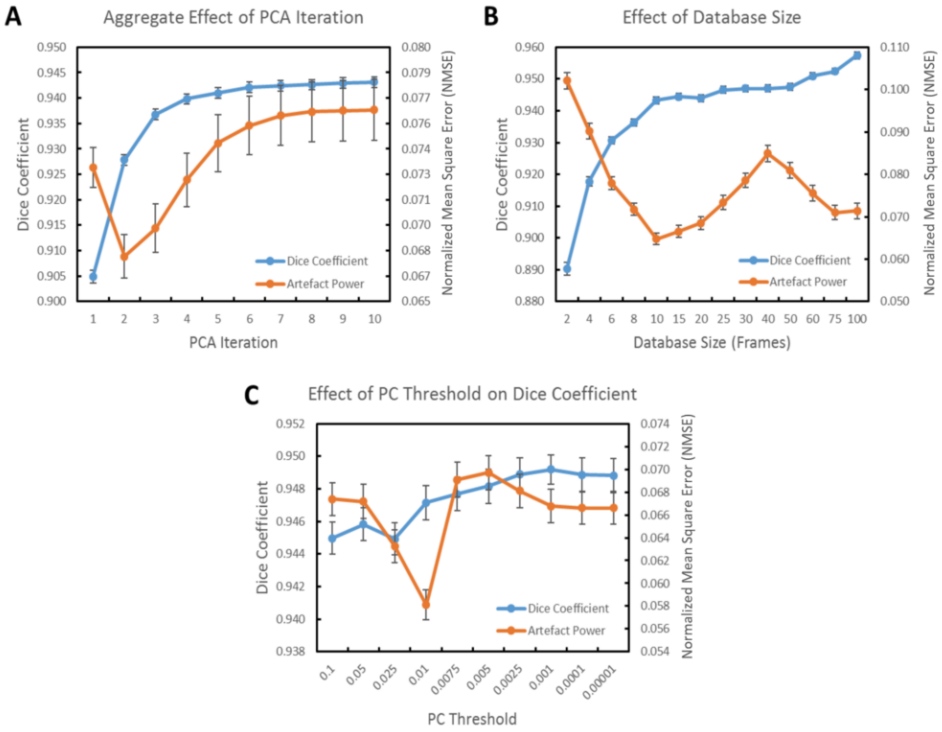

Figure 3.4: Plots demonstrating the effect of CS-PCA parameters on DC and NMSE. All data plotted is the aggregated average over all patients at an acceleration factor of 5x at 3T. Data displayed are mean values with 95% confidence intervals. (A) contains a plot showing how the DC and NMSE change with increasing PCA iteration value. Plot (B) demonstrates how the DC and NMSE are effected by database size. (C) displays how the PC threshold value effects the DC and NMSE. The DC value increases as the threshold is decreased; whereas the NMSE value decreases with increasing threshold.

trend toward a constant value after 5 PCA iterations; however, as the number of

PCA iterations does not greatly increase the computational time, we used 10 PCA iterations as the default. The PC threshold parameter is shown in Figure 3.4C. The NMSE reached a minimum at 0.01; however, the DC steadily increased as the threshold decreased to 0.001. Below 0.001, both the NMSE and DC trend toward a constant value. Since DC is a more important metric of how well the reconstruction is performing, we used a PC threshold of 0.001 as our default. Figure 3.4B displays the effect of the database size. Increasing the database is beneficial as it decreases the NMSE and increases DC; however, it also increases reconstruction time. As the database size increased past 30 the increase in DC became mostly constant. For example, an increase from a database size of 30 to 50 resulted in a 0.5-1% increase in DC.

A database size of 30 was sufficient for all patients; however, one patient benefited in an increased DC from increasing the database size (data not shown). Increasing the database from 30 to 100 increased the DC by 15% (averaged over all acceleration factors); however, the reconstruction time also increased threefold. For the patient grouped data, a database size of 30 was used for all patients to be consistent. The choice of using 30 database images also stems from the fact that our neural-network segmentation algorithm used 30 training images, aside from it being a conscious compromise between reconstruction time, low NMSE, and high DC.

3.4.2 Temporal Evolution

As the proposed technique uses a small database acquired prior to real-time imaging, it is of importance to evaluate the algorithm temporally. Figures 3.5(a) and 3.5(b) contain box and whisker plots of the NMSE and DC of one patient at acceleration factors ranging from 2x to 10x. It is evident that the DC remains largely unaffected by the increase in NMSE at higher acceleration values. Figures 3.5(c) and 3.5(d)

contain the NMSE and DC values calculated for all 620 frames, for one patient. It is evident from Figure 3.5(c) that NMSE increases over time; however, the DC data remained largely unaffected (Figure 3.5(d)). Linear fits for the DC plots shown in Figure 3.5(d) for 2x and 10x acceleration were calculated, which resulted in slope values of -4×10^{-6} frame^{-1} and -8×10^{-6} frame^{-1}, respectively.

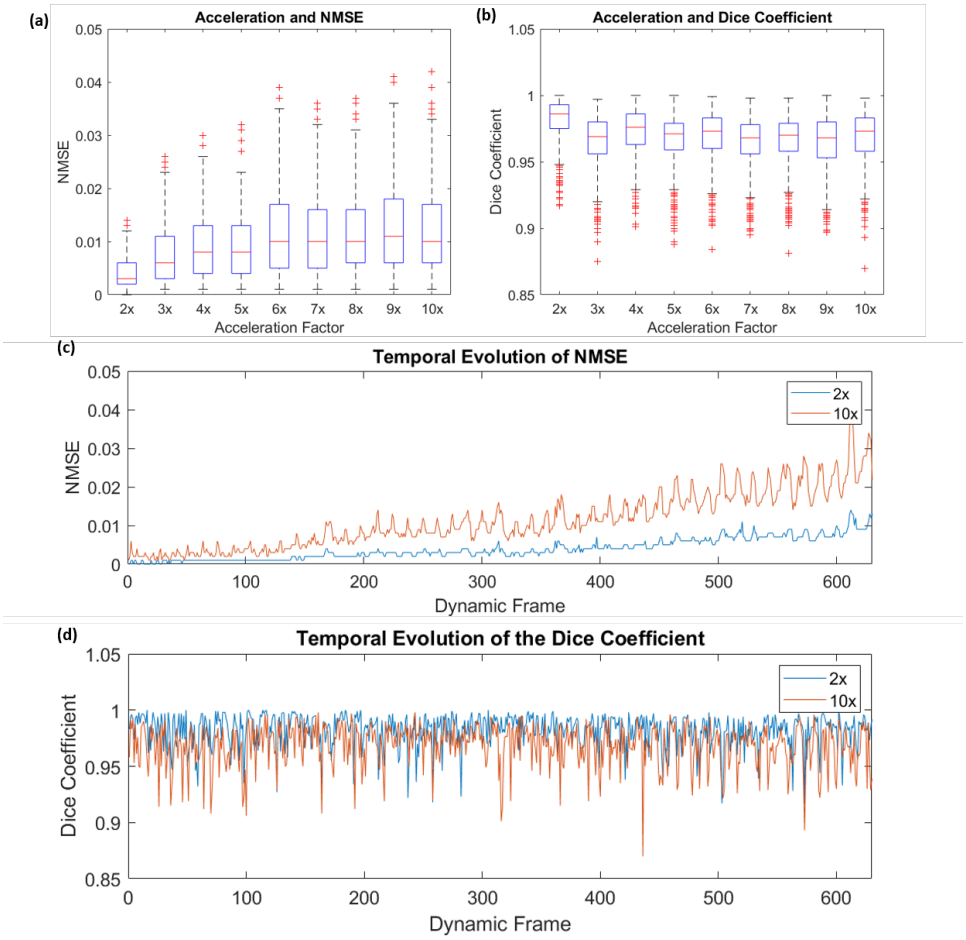

Figure 3.5: Temporal evolution of the CS-PCA NMSE (a), (c) and the Dice coefficient (b), (d) for one patient at 3T. The box plots (a), (b) contain data averaged over the 620 frames for each acceleration factor (red crosses denote outliers). The plots (c), (d) contain the temporal evolution over the duration of scanning for acceleration factors of 2x and 10x. It is evident from (c) that the NMSE increases with increasing time (or dynamic frames); however, the dice coefficient (d) remains relatively unchanged over time.

3.4.3 Dice Coefficients and Centroid Displacement

The Dice coefficient (DC) and centroid displacement were investigated as an aggregated average for each acceleration factor. Figure 3.6A contains the patient aggregated centroid displacement calculated for CS and CS-PCA at 3T and 6x noise degraded data. Both algorithms had a low average centroid displacement for 2x acceleration (50% acquired k-space); however, it is evident that CS-PCA produces a smaller centroid displacement than CS for accelerations up to 10x (10% acquired k-space). Furthermore, the centroid displacement for CS is sporadic, and has a significant increase in displacement at 7x acceleration (∼14% acquired k-space). The centroid displacement for the CS-PCA technique remained under 1.15 mm for all acceleration factors, whereas the displacement for CS remained above 1.49 mm (with the exception of 2x acceleration).

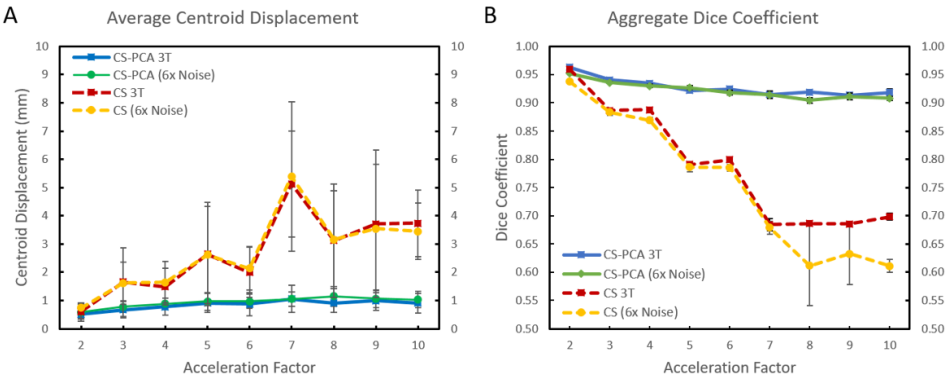

Figure 3.6: Data points represent the mean and 95% confidence interval for each plot. (A) Patient averaged centroid displacement for each acceleration factor. Notice that CS-PCA remained under 1.15 mm displacement for all acceleration factors; using CS alone, the displacement is both larger and more erratic; (B) displays the group averaged DC (all 620 frames for the 6 patients) for CS-PCA and CS. The plot also contains the 6x noise degraded data for both CS-PCA and CS.

Figure 3.7 contains the temporal evolution of the DC and CD. It can be seen that the temporal evolution of the DC and CD follow a similar trend, such that low DC

values do, in general, correspond to a high CD. The aggregate DC averaged over 620

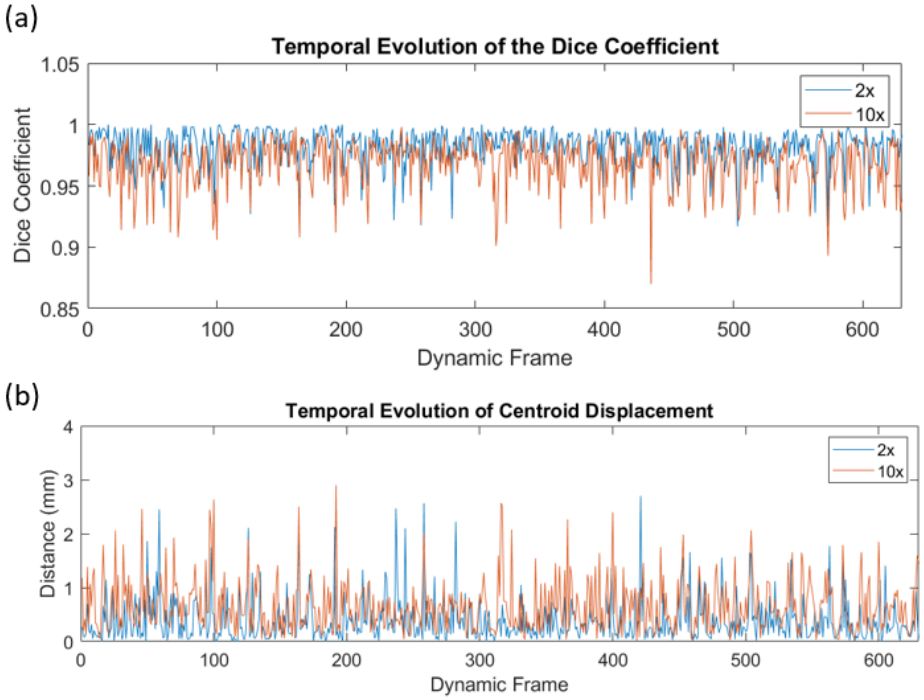

Figure 3.7: Temporal evolution of the DC and CD. It can be seen that the low dips in the DC correspond to high peaks in the CD data.

frames for all six patients at each acceleration factor is shown in Figure 3.6B. The DC for the CS-PCA algorithm consistently performed superior to the CS method. The DC for the CS-PCA algorithm remained above 0.9 for accelerations up to 10x, for both the 3T and 6x noise degraded data; whereas only 2x acceleration for the CS method was above 0.9.

3.4.4 Image Evaluation via Quantitative Metrics

Figure 3.8 contains plots of the patient aggregated MAPE, RMSE, NMSE, and SSIM metrics, averaged over all 620 frames. The MAPE, RMSE, and NMSE data at 3T and 6x noise degraded data, for CS-PCA and CS, are shown in Figure 3.8A, 3.8B, and 3.8C, respectively. It is evident that these three metrics share similar results,

with the exception of MAPE for the CS data, where the 3T data resulted in a higher MAPE value than the 6x noise degraded data. The CS-PCA data had remained relatively unchanged with increasing acceleration factor for all four metrics shown in Figure 3.8. The NMSE for CS-PCA remained below 0.05 for the 3T data, and below 0.06 for the 6x noise degraded data whereas CS greatly increased in NMSE as the acceleration factor increased. It is also interesting to note that the SSIM data in Figure 3.8D displayed a similar trend to the DC shown in Figure 3.6B.

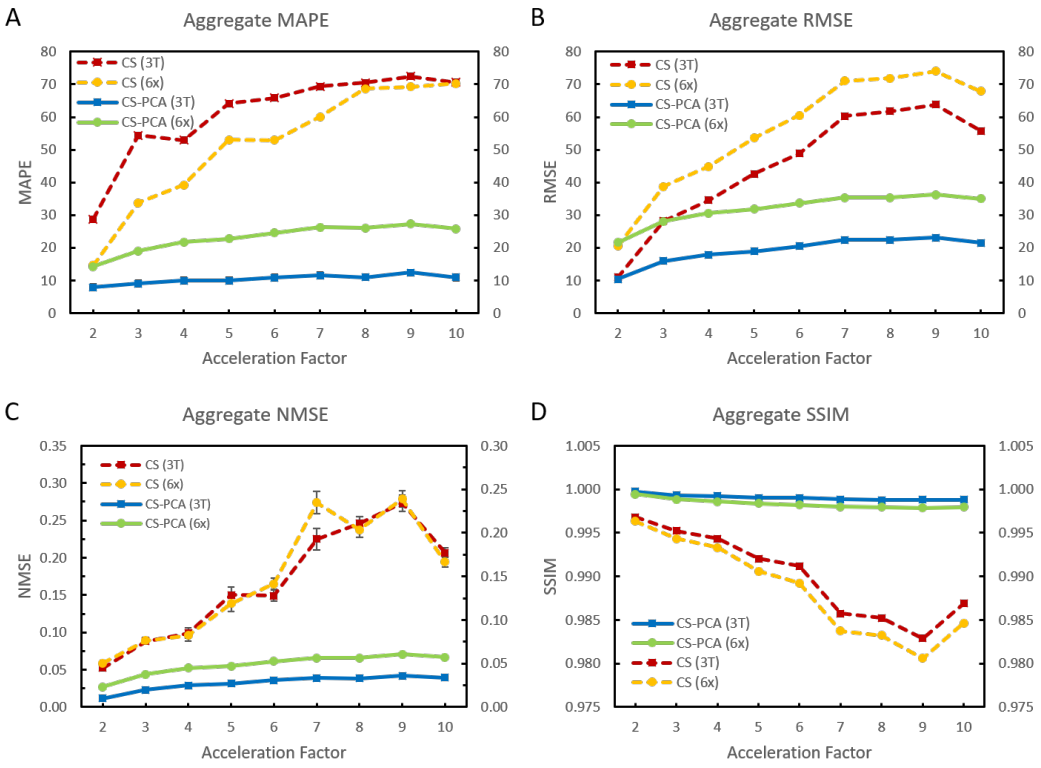

Figure 3.8: Quantitative evaluation of our proposed CS-PCA technique. The quantitative metrics include: (A) MAPE, (B) RMSE, (C) NMSE, and (D) SSIM. Each plotted data point contains the mean and 95% confidence interval for all 620 frames for all six patients, at each acceleration factor. It is evident from the plots that CS-PCA preformed superior than CS alone for all metrics. Further, CS-PCA performs similarly for acceleration factors ranging from 2x to 10x.

3.4.5 Visual Comparison

While quantitative metrics are useful in determining how well a technique is performing, a qualitative visual comparison is also important to see how each technique is handling the data. Figure 3.9 includes two patients each displaying their fully sampled reconstruction, along with the CS-PCA and CS reconstruction at 5x acceleration (20% of k-space used for reconstruction) at both 3T and the 6x noise added data.

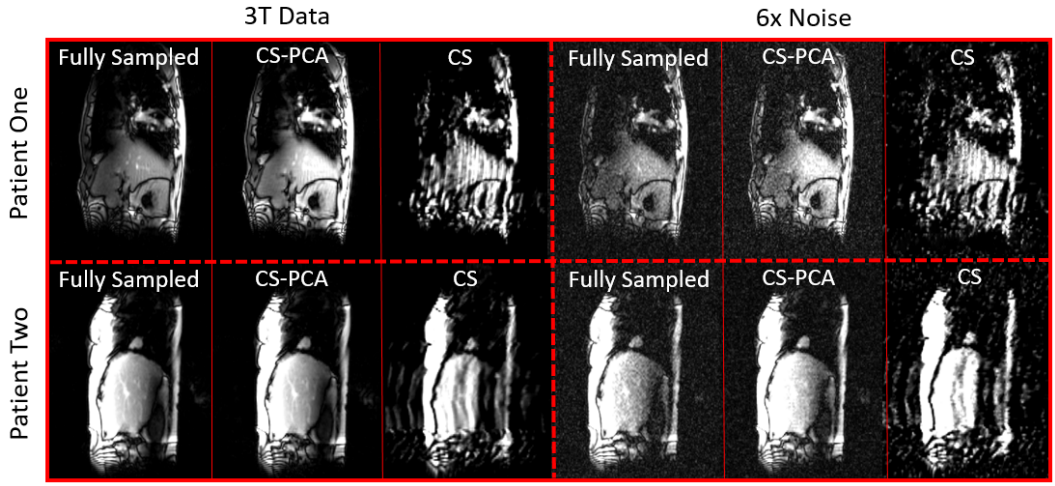

Figure 3.9: Images of two non-small cell lung cancer patients. The CS-PCA and CS images are reconstructed using only 20% of the k-space (5x acceleration). It is evident that CS-PCA retains visual quality, whereas the CS images have artefacts occurring throughout.

3.5 Discussion

3.5.1 Dynamic PCA Image Reconstruction

It is evident that CS-PCA reconstruction is an effective reconstruction technique for real-time dynamic imaging. The CS-PCA algorithm finds the components (from the database) with the greatest variance (i.e. greatest organ motion) and ranks them higher than non-redundant components, which can be attributed to noise. In dynamic

imaging of the lung, for example, acquiring several repetitions of the breathing cycle will result in the greatest variance to be the motion of the diaphragm and surrounding lung structures. Projecting the accelerated (current) image to the PC's produces the PC weights, which are indicative of how relevant each component of the database is to the current image. Therefore, CS-PCA is a technique that is best suited for any dynamic imaging of rhythmic motion such as breathing or cardiac imaging.

An important aspect of the CS-PCA algorithm, is that the central lines of k-space are always acquired and incorporated into the reconstruction. This was evaluated and can be seen in Figure 3.10, where an unexpected square artefact was introduced into the data set, but excluded from the initial offline database. This artefact-imposed image was transformed into k-space and undersampled by an acceleration factor of 5x. Using CS-PCA, the artefact was correctly reconstructed with only minor edge defects, even though the artefact was not present in the database. Thus, CS-PCA is not constrained to data contained in the offline training database images acquired prior to the real-time online undersampled data.

3.5.2 Individual Patient Data

The data shown in the results section contain the patient aggregated data. Figure 3.11 contains the DC and NMSE data for all six patients individually. One patient (patient 5) had NMSE values that were much larger than the other five patients. It can be seen in Figures 3.11B and 3.11D that the cause of the increased NMSE can be attributed to a streaking artefact emanating from the liver region. Although this resulted in larger NMSE values, the DC values for the tumour segmentations were not affected.

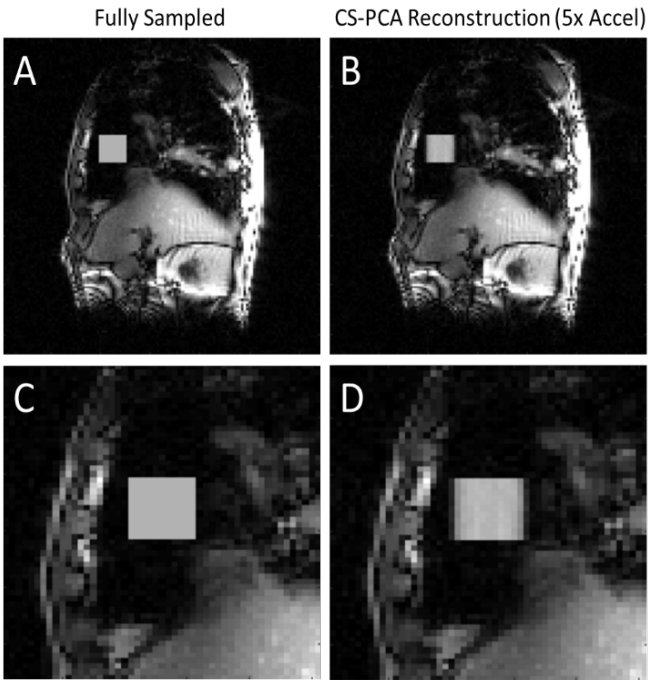

Figure 3.10: Demonstration showing how data not contained within the image database is handled. Images (A, C) contain a simulated artefact introduced to the data set, but not into the offline training database. (B, D) contain the CS-PCA reconstruction of images (A, C) after being transformed into k-space and undersampled, which contain the square artefact that was not present in the database images. Images (C) and (D) contain zoomed in regions of (A) and (B). It is evident in (B, D) that the square artefact is present in the reconstruction, although some of the high frequency information is lost.

3.5.3 Clinical Implementation

A possible clinical implementation of CS-PCA for our Linac-MR system is described as follows. The patient would receive a planning CT/MRI, as is the current clinical standard, along with a dynamic MRI scan prior to treatment. The dynamic MRI would be used to train the neural-network contouring algorithm our group has developed for target (tumour) tracking purposes [103]. For the treatment day(s), the patient would be placed into the Linac-MR and imaged for positioning and alignment. Their images would be registered (as is currently done with CBCT for standard linac treatment) to their planning data. Following patient positioning the database images

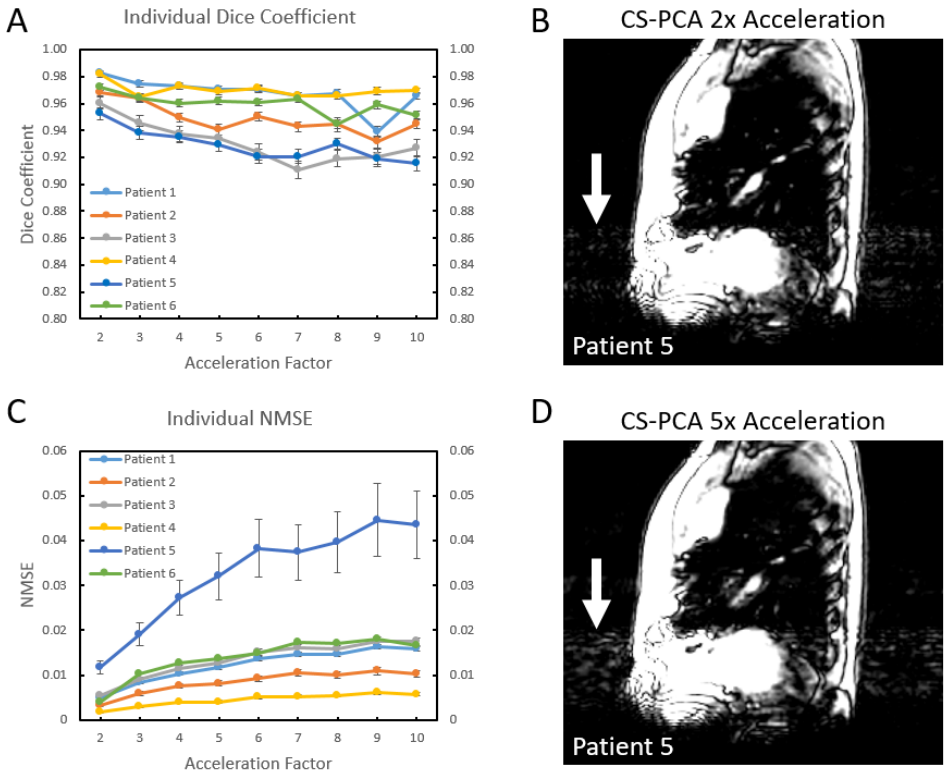

Figure 3.11: Individual patient data for (A) the dice coefficient and (C) the normalized mean standard error (NMSE). It is evident from (C) that patient 5 had larger NMSE values, with a larger variance. This was due to an artefact occurring from the liver, which resulted in a streaking occurring throughout the 620 dynamic images used to calculate the metrics; however, this did not result in dice coefficients that were significantly poorer that the other five patients. (B) and (D) contains a slice from 2x (50% k-space) and 5x (20% k-space), respectively.

would be generated, which takes less than 10 s for a 30-image database (at a TR of 2.2 ms and 128 x 128 matrix size). Immediately following the acquisition of the database, the PCA components are calculated and the accelerated real-time CS-PCA imaging would be initiated for treatment.

It is evident from Figures 3.6 and 3.8 that increasing the noise by a factor of 6, increased the NMSE, RMSE, and MAPE; however, the increase in noise did not affect the more clinically relevant DC and centroid displacement for CS-PCA. The fact that CS-PCA performs well in a variety of noise conditions suggests that this algorithm

will perform well across a wide array of MR platforms and field strengths.

3.5.4 Effect of Increasing Acceleration

As shown in Figure 3.6B, the DC decreases slightly with increasing acceleration factor. The reason for this can be explained by the increasing reliance on the mean database, which creates more motion blurring at higher acceleration factors. Recall that the unacquired data is filled using data from the mean database image plus the sum of the weighted PC's. As the acceleration factor increases there are more lines of k-space that need to be filled, which means that sum of the weighted components has less of an impact as the acceleration increases. The weighted components are reduced due to the decreasing amount of high frequencies acquired and the increasing artefact present from acceleration. Thus, in the case when the acceleration tends to infinity the resultant reconstruction will simply be the mean database image.

3.5.5 Limitations

There are several limitations to this preliminary study. Firstly, the database used in this retrospective study was acquired at the same time as the reconstructed data (i.e. both the database and data for reconstruction had the same acceleration value). While having fully sampled data is ideal for preliminary (proof of concept) studies, it is not truly indicative of how it may perform with prospectively-acquired accelerated data. However, the data used in this study were derived from a steady-state bSSFP sequence, meaning that there will not be inherent k-space signal variation in the phase-encode axis. This makes a retrospective approach more feasible in this proof-of-concept investigation. Regardless, in the clinical case data would be directly undersampled, and any impact this may have will need to be investigated in future

studies.

It was shown in Figure 3.5(c), the artefacts increase both over time and with increasing acceleration factor. Figure 3.12 shows an example for one patient at 5x acceleration. It is evident that the increased artefacts can occur throughout the data; potentially reducing the segmentation accuracy. Techniques to mitigate this issue will be investigated in the future.

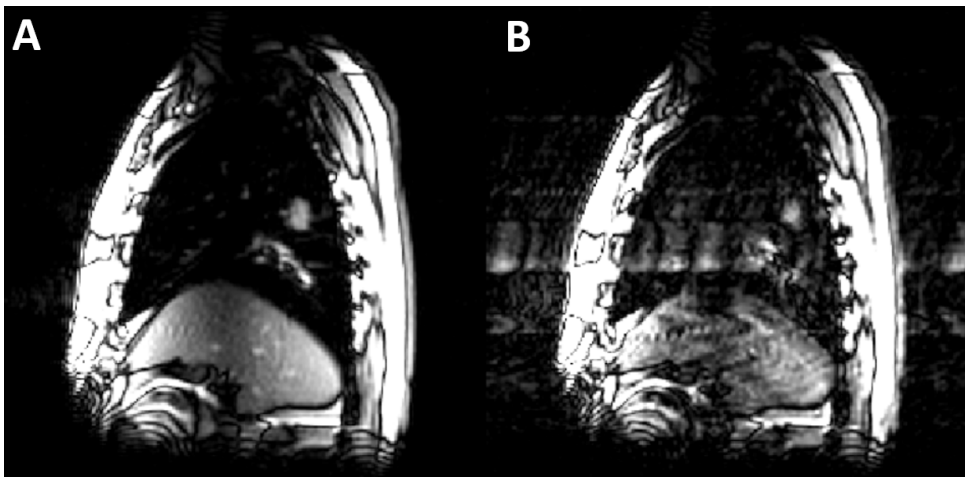

Figure 3.12: Increased artefacts may increase over time. This example contains, (A) the fully-sampled reconstruction, and (B) and CS-PCA reconstruction of the 500^{th} dynamic frame at 5x acceleration.

Further, our optimal parameters were determined empirically. Generally, this should be done by using some optimization scheme that searches throughout parameter space for the parameters that results in the highest DC and lowest NMSE (or another metric such as RMSE and MAPE).

3.6 Conclusion

A real-time online MR image reconstruction algorithm using compressed sensing and principal component analysis (CS-PCA) was developed and tested using six retro-

spective non-small lung cancer data sets. We have shown that the CS-PCA method produces images with low normalized mean square error (NMSE) at acceleration factors from 2x to 10x. The tumour segmentation Dice coefficient (DC) remained above 0.9 for the 3T and the 6x noise degraded data sets. Compared to the Split-Bregman CS technique, CS-PCA had less artefacts, reduced tumour centroid displacement, and increased DC, for all acceleration factors. It was found that the reconstruction noise increases as the database becomes outdated. Developing a technique to correct this will require future studies. Furthermore, the reconstruction time allows for a minimum of 6 frames per second at an acceleration factor of 2x (intel i7-4710HQ CPU @ 2.5 GHz).

Chapter 4

Real-Time MR Image Reconstruction using Convolutional Neural Networks

Parts of this chapter have been adapted from a published article: Dietz B, Yip E, Yun J, Gabos Z, Fallone G, Wachowicz K. Single Patient Convolutional Neural Networks for Real-time MR Reconstruction: A Proof of Concept Application in Lung Tumor Segmentation for Adaptive Radiotherapy. Physics in Medicine and Biology. 2019;64(19):15pp [116].

4.1 Introduction

With the recent advances in hybrid radiation delivery-MRI treatment systems (Linac-MR), real-time MRI is becoming an increasingly important requirement for adaptive radiotherapy, specifically for online tumour tracking [27, 23, 25, 117, 118]. In order to

accurately track a tumour in real-time, the image acquisition needs to be fast and of sufficient quality to allow for segmentation of the tumour. The use of on-the-fly segmentation allows the MLC to update along with the motion of the tumour [114, 119]. Tracking the tumour with the MLC ensures radiation delivery to a smaller planned treatment volume in real-time, reducing the radiation delivered to the surrounding healthy tissue and negating the need for respiratory gating. Accelerated MR acquisition can benefit LMR tumour tracking in many ways: (1) increasing the acquisition frame rate, (2) increasing the image resolution, (3) the ability to image multi 2D orthogonal planes or 3D image volume within the necessary frame rate. Unfortunately, acquiring clinically acceptable MR images with a low latency and high frame rate (including both acquisition and reconstruction) remains a challenge.

Raw MRI data is acquired in k-space, which must be sufficiently sampled to ensure that the Nyquist criterion is satisfied in order to avoid aliasing during image reconstruction. Unfortunately, acquisition of fully sampled k-space is time consuming, which makes real-time imaging at high frame rates challenging. One solution to this problem is to under-sample k-space, necessitating the violation of the Nyquist criterion. Coherent undersampling in k-space is often done when implementing parallel imaging techniques [40, 41]. Alternatively, k-space can be incoherently undersampled, resulting in a noise-like artefact often synonymous with compressed sensing reconstruction techniques [64, 68, 120]. We recently proposed a technique that combines principal component analysis (PCA) with compressed sensing (CS), to reconstruct undersampled (accelerated) MR acquisitions [102]. While this proved to be a fast technique, artefacts increased as the CS-PCA database became "outdated" as demonstrated in Figure 3.12. We believe the use of convolutional neural networks (CNNs) may be a technique that will be temporally robust.

Convolutional neural networks (CNNs) have been gaining popularity recently with the increase in simplified code packages such as: Theano [121], Lasagne [122], and Google TensorFlow [123]. With the increasing accessibility of CNN code and the decreasing cost of graphics processing units (GPUs), there has been more interest in using neural networks for image processing tasks. CNNs are well suited for image processing, due to the ability to find and recognize abstract features of the data [124]. Furthermore, neural networks in general perform calculations that are highly parallel and require large memory bandwidth, allowing for massive speedups using GPUs over traditional CPUs. CNNs have been used for image processing techniques including: segmentation [93], super-resolution [94], classification [96], and image reconstruction [95, 97]. The effectiveness of a CNN depends both on the quality and quantity of data provided to train the network, as well as the optimized hyperparameters.

Patients that undergo cancer treatment via external beam radiation, will generally receive imaging prior to their treatment via CT simulation (CT Sim). The CT Sim data is used by the dosimetrist to generate contours and plan how the radiation will be delivered by the linac. Institutions with MR simulation (MR Sim) or linac-MR systems will have the ability to image the patient with MRI prior to their treatment. This previously acquired MR data can be utilized to train a CNN specifically tailored to an individuals data prior to the scheduled treatment, which can then be used to aid in reconstructing rapidly acquired undersampled data during treatment. We have termed these patient-specific CNNs. Because the MR data would be acquired prior to the scheduled treatment, the training time is generally of not much concern; however, the reconstruction time should be fast to facilitate real-time imaging.

We investigated the use of 3D CNNs to reconstruct retrospectively undersampled data from six non-small cell lung cancer patients on a patient-by-patient basis, to

quantitatively explore the efficacy of 3D CNNs in a real-time imaging scenario. In this approach we use each patients' previously acquired training data to train a neural network that is specific to each patient. This is contrary to the more common neural network technique of training using a large patient cohort, to reconstruct a single patient. We emphasize that our goal is to use patient-specific CNN reconstruction in a real-time on-the-fly scenario for adaptive radiotherapy purposes using a Linac-MR system. We furthermore explore the realistic imaging scenario, where the CNN network is trained using retrospectively undersampled data (from a fully sampled data set) to reconstruct data that has been prospectively undersampled at the time of imaging. We have termed the realistic imaging scenario as the prospective reconstruction case.

While training a CNN may take on the order of hours, the application of the trained CNN parameters is on the order of tens to hundreds of milliseconds, depending on the size of the CNN and the computer hardware used. As respiratory motion remains a significant challenge for radiotherapy targeting lung tumours [13], this proof of concept study investigates patient specific CNNs to reconstruct undersampled patient data for the purpose of real-time tumour segmentation. Our goal for this proof of concept study is not to show that this technique is a significant improvement upon other techniques; rather, we are demonstrating that the use of 3D CNNs for real-time reconstruction is feasible, and worthy of future consideration and development.

4.2 Methods

4.2.1 Neural Network Image Reconstruction

The CNN code used was written by Schlemper *et al.* in Python using Theano and Lasagne packages [97], that has been adapted for the use in real-time dynamic image reconstruction. The original code proposed by Schlemper was written to reconstruct the entire temporal domain in spatial patches, such that, the utilization for on-the-fly reconstruction was not a concern. Our adaptation allows for on-the-fly reconstruction of the entire dynamic frame (not in patches), using the previous images to aid in the reconstruction of the present image.

The complex MRI data is separated into real and imaginary channels where the 3D CNN is applied to each channel separately, which are then recombined prior to reconstruction. The CNN reconstruction can be written as

$$x_{cnn} = f_{cnn}\big(\mathbf{x}_u|\vec{\theta}, \lambda, \mathbf{\Omega}\big), \tag{4.1}$$

where the term $\mathbf{x}_u$ is the zero-filled undersampled image and $f_{cnn}\big(\mathbf{x}_u|\theta, \lambda, \mathbf{\Omega}\big)$ is the CNN reconstructed image. The function f_{cnn} attempts to reconstruct the undersampled image $\mathbf{x}_u$, which depends on the CNN parameters (trained weights and biases) $\vec{\theta}$, a data fidelity weighting term λ, and the known k-space locations $\mathbf{\Omega}$. The data fidelity term λ is calculated such that $\lambda = q/\sigma$, where q is a hyper-parameter and σ^2 is the noise power. The λ term is initialized to 0.025 as it was empirically shown that this value performs well [98]. Given the fully-sampled input training data D, the CNN is trained to reconstruct the data by minimizing the following cost function,

$$Z(\theta) = \sum_{x_u, x_{gnd} \in D} ||x_{gnd} - x_{cnn}||_2^2. \tag{4.2}$$

The terms x_{gnd} and $x_{cnn} = f_{cnn}\left(\mathbf{x}_u|\vec{\theta}, \lambda, \mathbf{\Omega}\right)$ are the ground truth image and the CNN reconstructed images (from the undersampled ground truth data), respectively. Once trained, the parameters can then be rapidly applied directly to the undersampled data (either prospectively or retrospectively). While the CNN utilizes both the k-space and image domain within the network, once trained the CNN takes the aliased image (zero-filled) as an input and outputs a reconstructed (un-aliased) image.

The Adam optimizer was implemented for the CNN training [82]. The details of which can be found in Chapter 3. In short, the Adam optimizer incorporates the use of adaptive moments into the gradient descent optimization. It has been shown to be robust, and is often the default optimizer for most ANN problems. The parameters $\vec{\theta}$ are updated such that,

$$\begin{aligned}
\vec{\theta} &= \vec{\theta} - \epsilon \frac{P_1}{\sqrt{P_2} + \delta}, \\
P_1 &\leftarrow \rho_1 P_1 + (1 - \rho_1) G, \\
P_2 &\leftarrow \rho_2 P_2 + (1 - \rho_2) G^2.
\end{aligned}$$

Where the terms ρ_1 and ρ_2 are set to 0.9 and 0.999, respectively. ϵ is the learning rate and δ is a small constant (10^{-8}), G is the gradient,

$$G \leftarrow \frac{1}{m} \nabla_\theta \sum_{i=1}^{m} L(f_{cnn}(\mathbf{x}_u; \vec{\theta}, \lambda, \mathbf{\Omega}),\, x_{gnd}), \tag{4.3}$$

where m is the number of mini-batches, L is the loss (cost) function of the CNN output f_{cnn} and the corresponding true output x_{gnd}.

The CNN network employed is a deep network that attempts to remove the under-sampled aliasing using alternating convolutional layers and data consistency layers.

The convolution layer convolves the previous layer (or input data) with filters (or kernels) in order to de-alias the undersampled data. The convolution kernels learn as the CNN is trained, to better remove the aliasing artefacts (by reducing the cost error). Compared to classification CNNs, the kernels learn abstract features of the data for the purpose of determining what the picture contains. These abstract features are initially composed of structures such as edges, but as the network becomes deeper with more convolutions, the more abstract these features become [125, 124]. The convolution performed used a single strided convolution with "same" padding. Single strided convolution refers to a convolution kernel that is shifted every pixel, and "same" padding results in an output image that is the same size as the input by adding zeros to the input data to account for the size of the convolution kernel. The data consistency layers are used to ensure that the acquired k-space is incorporated into the reconstruction (enforces data fidelity), which was originally developed for a dictionary learning reconstruction technique [98]. Prior to the data consistency layer, the output from the convolutional layer is added to the input via residual connection [126]. During the data consistency layer, if a given k-space location has no acquired data (zero filled from the undersampling), it is simply equal to the value produced by the Fourier transform of the CNN reconstruction ($\mathbf{K}_{cnn} = F\mathbf{x}_{cnn}$). If the k-space location has been initially acquired, the result will be a weighted sum of the initial value $\mathbf{K}_u$, and the $\mathbf{K}_{cnn}$ weighted by the data fidelity term λ. This can be written as

$$\mathbf{K}_{recon}(j) = \begin{cases} \mathbf{K}_{cnn}(j) & \text{if } j \notin \mathbf{\Omega} \\ \dfrac{\mathbf{K}_{cnn}(j) + \lambda \cdot \mathbf{K}_u(j)}{1+\lambda} & \text{if } j \in \mathbf{\Omega}. \end{cases} \tag{4.4}$$

The CNN reconstructed image is then computed by taking the inverse Fourier transform of K_{recon}.

The neural network performs alternating layers of convolutions and data consistency for N iterations. Each convolution layer performs data sharing followed by M convolutions. Each convolution is followed by the rectifier linear units (ReLU) non-linearity activation function. It has been demonstrated that if the number of convolutional layers is too large, the training data will become overfit [97]. Overfitting reduces the ability of the CNN to generalize to "unseen" data, as such this must be avoided or accounted for. As the number of data consistency layers increase, the reconstruction error decreases; however, this is at the cost of increasing the training time. We empirically investigated the optimal CNN hyperparameters for our data using a systematic grid search of the hyperparameter space. The hyperparameter search consisted of training the CNN for a reduced number of epochs while altering the number of convolutional layers (2-6), data consistency layers (1-5), and number of filters (16, 32, or 64); resulting in 75 combinations (5 convolutional layers $\times$ 5 data consistency layers $\times$ 3 filter values) of hyperparameters.

The data sharing portion combines the k-space of previous acquisitions to generate a more "complete" k-space. By sharing data between previous sequential frames during the training step, we can create a k-space that has a lower apparent acceleration (or a lesser amount of undersampling). The data sharing was implemented by averaging the nearby dynamic k-space frames. If multiple lines of k-space have been acquired, data sharing uses the weighted average of those lines. Figure 4.1 contains a schematic of the CNN along with the data sharing component. The data generated via data sharing are incorporated as additional channels to aid the CNN training. Additionally, data sharing incorporates the dynamic motion directly into the CNN training, allowing the CNN to solve for the joint de-aliasing and dynamic motion estimation.

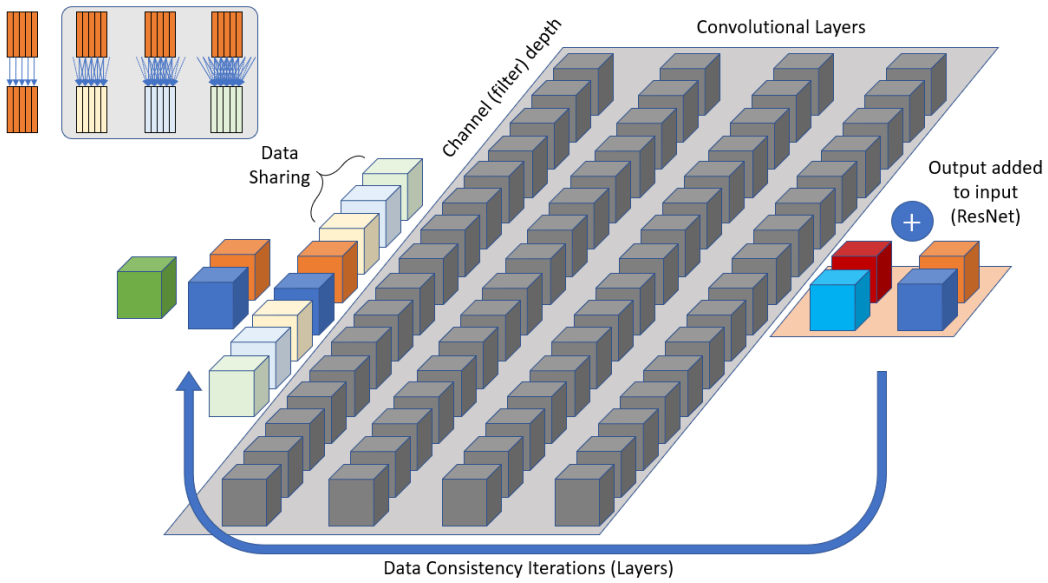

Figure 4.1: Schematic of the CNN. The training (input) data is batched into groups containing five dynamic frames. The data is initially parsed into two channels, the real and imaginary components. Data sharing is employed to encompass the dynamic motion and reduce the incoherent aliasing. This is highlighted in the box shown at the top left of the figure. The eight channels then undergo the convolutional layer (made up of several convolutions) followed by the residual connection (ResNet) and finally the data consistency layer. The output is the same size as the input; however, only the last reconstructed temporal frame would be used in a real-time online scenario.

4.2.2 Neural Network Parameters

There are several hyperparameters that need to be considered when applying a convolutional neural network (CNN), which are dependent on the image processing task. As mentioned above, we performed a systematic hyperparameter search in order to determine optimal parameters. To do so, we trained a network for a reduced number of epochs (25) for each set of parameters. We then investigated the reconstruction quality using the parameters via normalized mean square error (NMSE), peak signal to noise ratio (PSNR), structural similarity index (SSIM), as well as the tumour segmentation via the Dice coefficient (DC) and centroid displacement, as well as the image reconstruction time. The investigated hyperparameters included the number

of convolutional layers, data consistency layers, and filters.

A 3D convolution kernel ($3 \times 3 \times 3$, two spatial dimensions and one temporal dimension) was implemented to incorporate the temporal features of the dynamic data during the CNN training. Having a smaller kernel would increase the training and reconstruction time, while having a larger kernel may not encapsulate the tumour details. The 3×3 convolutional kernel equates to a planar area of 7×7 mm^2, given the resolution of the MR data. The incoherent phase encoding was acquired such that the central 8 phase encodes were always present, while the remaining (undersampled) higher frequency phase encode lines were chosen pseudo-randomly. Figure 4.2 displays 20 dynamic phase encodes temporally along the phase encode axis.

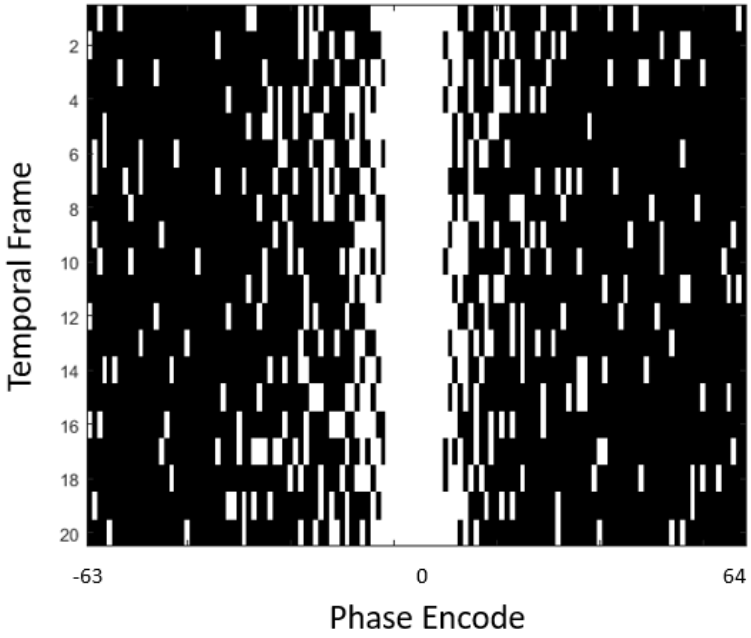

Figure 4.2: Example of the phase encoding scheme for 20 dynamic frames. By having a temporally varying phase encode scheme, data sharing techniques can be employed to create images encompassing both temporal data and less acceleration (by combining neighbouring acquired lines to fill in the zero-filled lines). The black regions indicate the zero-filled k-space, while the white indicates regions that were acquired.

4.2.3 Retrospective Data Analysis

Six non-small cell lung cancer patients previously imaged with our 3T Philips Achieva system (Philips Medical Systems, Cleveland, OH, USA) were used to retrospectively investigate the accuracy of the CNN reconstruction technique, for the purpose of tumour segmentation. The acquired data was subject to approval from our institutional board of ethics. Each patient data set contained 650 dynamic frames of free breathing acquired over a period of 3 minutes with visually distinguishable lung tumours, as shown in Figure 4.3. Assuming a respiratory cycle of 3 seconds, this results in approximately 60 respiratory cycles per imaging session to ensure the maximum range of the tumour is captured.

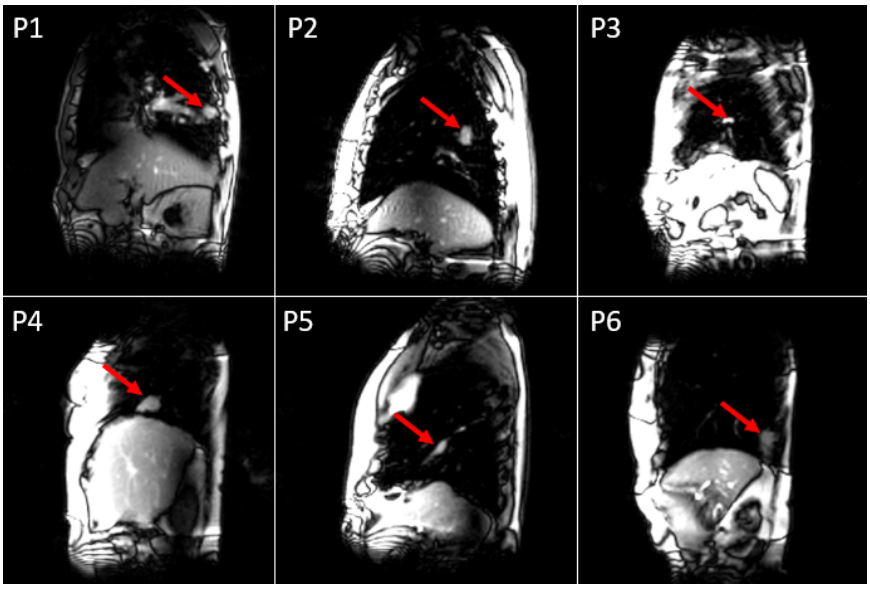

Figure 4.3: Patient data acquired at 3T with our Philips Achieva system. Each patient had a visually distinguishable lung tumour of varying shape and size. All six patients had a dynamic scan of 650 frames of free breathing acquired using a bSSFP sequence with a TR/TE of 2.2/1.1 ms.

The fully sampled data was acquired using a Cartesian bSSFP sequence using a multi-channel chest array. The multi-channel data was combined at the console prior to exporting the complex data. The imaging parameters consisted of 128×128 voxels,

a FOV of 40×40 cm^2, slice thickness of 2 cm, TR/TE = 2.2/1.1 ms, and a flip angle of 20°. The fully sampled frame rate was therefore the acquisition of a single slice at four frames per second. The fully sampled data were retrospectively undersampled in k-space using a pseudo-random Monte Carlo binary mask at acceleration factors of 5x and 10x. At 5x acceleration, 80% of the acquired k-space is masked (set to zero); 10x acceleration implies that 90% of the acquired k-space is masked. Zero-filling is a preferred method of handling unacquired k-space data for the initial image estimate since it is devoid of any assumptions that may add bias to the final result. Zero-filling is a common reconstruction process for generating initial image estimates in undersampled MRI reconstruction. The retrospective mask generation was dynamically calculated, such that each dynamic mask contained different high frequency phase encode lines (see Figure 4.2). Having a dynamically changing mask, allows for optimal use of the data sharing technique discussed above.

The robustness of the CNN reconstruction to noise was investigated by adding random Gaussian noise to the 3T data [102]. The standard deviation of the noise added was calculated as

$$\sigma_{add} = \sqrt{N^2 - 1}\sigma_{meas}. \tag{4.5}$$

The terms σ_{add} and σ_{meas} are the additive noise and measured noise, respectively. A factor of N = 6 was used to emulate the noise of a 0.5T low field. In order to avoid Rician noise distribution, the noise was added to the real and imaginary channels independently. The CNN network was trained on both the 3T data and 0.5T noise simulated data, in order to investigate if training the network on a stronger field strength would negatively affect the reconstruction of lower field data.

CNN Training

The training data consists of the fully-sampled data (initially in image space), which is undersampled using a binary mask in k-space before the training begins. The training cost is then computed using the CNN output and the initial fully-sampled data. Each non-small cell lung cancer patient had two training phases; a coarse training phase followed by a fine-tuning phase, each of which ran for 100 epochs. One epoch is defined as a forward pass and backward pass through the entire training dataset. The difference between training phases is the speed a which the stochastic gradient descent steps towards to minimum. The coarse training used a faster (larger) learning rate of 0.001, whereas the fine-tuning phase used a slower (smaller) learning rate of 0.0005. The CNN parameters generated from the coarse training phase were used to initialize the reconstruction of the fine-tuning phase. The total training time took approximately 6 hours to complete both the coarse and fine-tuning phases.

A CNN was trained for each patient independently for reconstruction of their own undersampled data. The data were separated into groups to ensure the reconstructed images used for quantitative analysis were not incorporated into the CNN training. The first 300 dynamic images for each patient were used to train the CNN and the next 150 dynamic images were separated equally to use 75 for validation, and the other 75 for testing purposes. The remaining 200 frames were undersampled and reconstructed for quantitative analysis. Batching was used along the temporal axis to break the data into small subsets containing 5 dynamic frames per batch.

4.2.4 Prospective Data Analysis

Prospective CNN image reconstruction presents a greater challenge than the retrospectively undersampled case. This is due to differences between the two separately

acquired acquisitions: the fully sampled training data, and the prospectively undersampled data. The prospective case will often have confounding issues, such as increased visibility of eddy current artifacts due to the rapidly changing phase encode steps, which may be different than the eddy current contributions present in the fully sampled training data. Furthermore, the prospectively undersampled data will be acquired at a higher frame rate (due to undersampling), which a retrospective analysis cannot account for.

There are several techniques that can be used to reduce eddy current related artifacts when acquiring k-space data non-sequentially [127]. We initially investigated pairing the phase encodes; however, this reduces the incoherence of the undersampling due to the decrease in randomness of the sampling. We therefore used a "through-slice technique", where the amplitude of the slice preparation lobe is increased. This created a slight unbalancing of the slice-selection gradient which has been shown to suppress eddy-current contributions by inducing a symmetric low-angle dephasing within each voxel [127]. Figure 4.4 contains a pulse sequence diagram demonstrating the through-slice correction. The altered slice-selection gradient waveform is modified in both the fully sampled (training) data and the prospectively undersampled data to ensure the contrast is identical for both data sets.

To test the CNN reconstruction technique with prospectively undersampled data, a software patch was written to enable dynamic incoherent sampling on our 3T Philips Achieva system (shown in Figure 4.2). Data were acquired with a volunteer and Quasar motion phantom (Modus Medical Devices, London, ON, CA) using the same imaging parameters as the non-small cell lung cancer patients. The Quasar motion phantom consists of a static torso with a moving cylindrical insert. The torso was filled with distilled water and doped with 10 mg/L of manganese chloride. The cylindrical

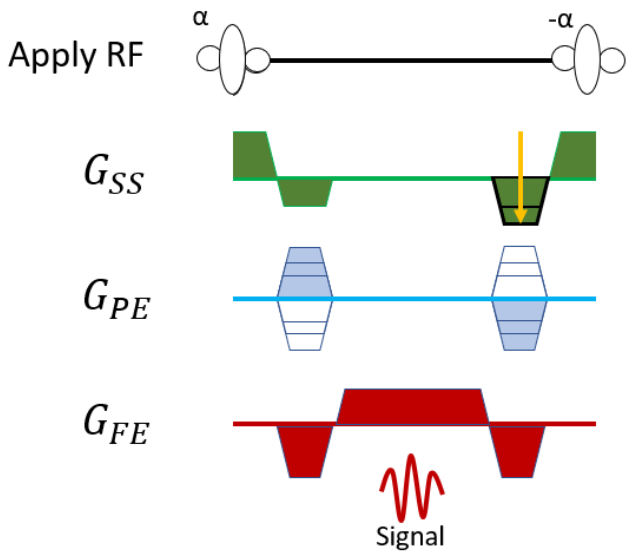

Figure 4.4: Demonstration of the through-slice eddy-current correction. By increasing the slice selection gradient, additional off-resonance is introduced that averages out the signal fluctuations produced by the eddy-currents.

insert was filled with plastic beads as well as doped water (10 mg/L), to mimic the density and signal of lung. Inside the cylinder contained a spherical tumour surrogate, doped to a concentration of 70 mg/L manganese chloride.

Our volunteer data consisted of a fully sampled dynamic acquisition, followed by a prospectively undersampled acquisition at 4x acceleration. Similarly, for the Quasar motion phantom we acquired a fully sampled dynamic acquisition, but we investigated two acceleration factors: 4x and 8x. The Quasar phantom motion consisted of a sinusoidal waveform along the z-axis (to mimic a simple superior-inferior breathing motion) with a total amplitude of 6 cm and a period of 5 seconds. Generally, a breath respiratory cycle may be modelled as a Cosine[4]; however, the use of a Sinusoidal function was chosen for qualitative reconstruction evaluation with a simplified superior-inferior motion.

CNN Training

The data were acquired using a multi-channel chest array consisting of six coils. Our goal is to reconstruct each coil independently, which can then be combined in image space. Therefore, our approach requires a CNN to be trained for each channel independently, reconstructed, then coil combined after reconstruction of the six channels.

The prospective data was trained differently than the retrospective data. For the prospective case, the volunteer and Quasar motion phantom data ran for 200 epochs using only a coarse training phase. Using one coarse training phase with an increased number or epochs was found to be more successful than introducing a fine-tuning phase. The prospective data was split using 300 images for training, and 75/75 for validation and testing, respectively. Only the fully sampled data was included for the CNN training; the prospectively undersampled data was not incorporated into the CNN training. The prospective data was reconstructed using the hyperparameters calculated from the (retrospectively undersampled) fully sampled data.

4.2.5 Quantitative reconstruction metrics

Our study aims to investigate the use of CNNs for MRI reconstruction for the segmentation of lung tumours using an automated tumour segmentation algorithm developed by our group [114]. As such, our metrics are heavily focused on the ability to distinguish the tumour sufficiently for segmentation.

Dice Coefficient

While overall image quality is an important metric, the most important metric for this study is how well the CNN reconstructed data can be used to track a small moving tumour. We chose to evaluate the segmentability as it pertains to the adaptive

radiotherapy scenario using a MR-Linac system, where the region of interest (tumour is this case) is to be consistently segmentable. Therefore, we chose to investigate the Dice coefficient (DC) to compare the Boolean tumour segmentations from the fully sampled data to the CNN reconstructed data [110]. The Dice coefficient can be measured as,

$$DC = 2 \cdot \frac{ROI_{FS} \cap ROI_{US}}{ROI_{FS} + ROI_{US}}. \tag{4.6}$$

ROI_{FS} is the contour region of interest (ROI) for the fully-sampled data and ROI_{US} is the contour for the undersampled CNN reconstructed data. A perfect DC will be equal to unity, while an increasing mismatch will yield a DC closer to zero.

Centroid Displacement

Another metric that was used to assess the tumour segmentation is the tumour displacement, which quantitatively measures the total segmentation displacement of the CNN reconstructed data from the fully sampled data. The distance between the centroid of the binary tumour segmentation from the fully sampled data and the CNN reconstructed data was calculated for each patient, to investigate any shifts that may have occurred during the CNN reconstruction. The centroid displacement alone, provides nearly no information regarding the segmentations; however, used along with the Dice coefficient, we believe it provides a more complete analysis on how the segmentation is performing.

Normalized Mean Square Error

We also investigated the normalized mean square error (NMSE), or artifact power, as an overall image quality reconstruction metric [28]. The NMSE is an image quality metric that measures the amount of artifact appearing in the undersampled recon-

struction compared to the fully sampled reconstruction,

$$NMSE = \frac{\sum (I_{FS} - I_{US})^2}{\sum I_{FS}^2}. \tag{4.7}$$

The terms I_{FS} and I_{US} denote the pixel-wise image data for the fully sampled and undersampled data, respectively. A value of 0 indicates no artifacts are present (i.e. the images are exactly the same), whereas any value above zero indicates present artifacts within the undersampled reconstructed image. The NMSE is used as a comparative metric, to evaluate (compare) the artefact presences across the six patients, and across the two acceleration factors (5× and 10×).

Peak Signal to Noise Ratio

The peak signal to noise ratio (PSNR) is a metric that can be used to evaluate how well the CNN reconstructed compares to the fully sampled data. The PSNR can be written as

$$PSNR = 10 \log_{10} \left(\frac{Max^2}{MSE} \right), \tag{4.8}$$

where Max is the maximum pixel value and MSE is the mean square error, which is the numerator of the NMSE equation above.

Autocontouring

An automatic contouring program that utilizes neural networks has been previously developed by our group [114, 119, 36]. The program takes contours drawn on several phases of the breathing cycle as input data, which is used to train a neural network to find the region of interest (tumour) and segment it. All segmentations were performed using this contouring program, which was trained using contours drawn by a radiation oncologist.

Patient #	Area (cm^2): Mean (SD)	SI Extent (cm)	AP Extent (cm)
1	2.57 (0.09)	1.97	0.40
2	3.76 (0.25)	2.63	0.75
3	1.36 (0.16)	1.57	0.66
4	4.96 (0.31)	0.94	0.39
5	2.09 (0.19)	1.07	0.72
6	4.99 (0.30)	0.87	0.17

Table 4.1: Lung tumour statistics for the six patients used in the study.

4.3 Results

4.3.1 Hyperparameter Selection

Three of the CNN hyperparameters were investigated to determine the optimal network for reconstructing repetitive tumour motion. The hyperparameters included the number of convolutional filters, data consistency layers, and convolutional layers. Figure 4.5 contains the results of the hyperparameter space search, which were quantitatively analyzed using the Dice coefficient (DC), centroid displacement, and NMSE. There were two combinations of hyperparameters that performed better than the other combinations. The two optimal choices for the number of filters, convolutional layers, and data consistency layers, were found to be 32,5,4 and 64,6,1; respectively. We chose to use the 32,5,4 set of parameters going forward, as having only one data consistency layer may result in higher reconstruction artefacts in general.

4.3.2 Tumour Statistics

Table 4.1 contains the tumour details for the six patients. The descriptors include the tumour area, and the maximum extent of the tumour in the superior-inferior and anterior-posterior directions. The data was calculated from the gold standard (fully sampled) contour data.

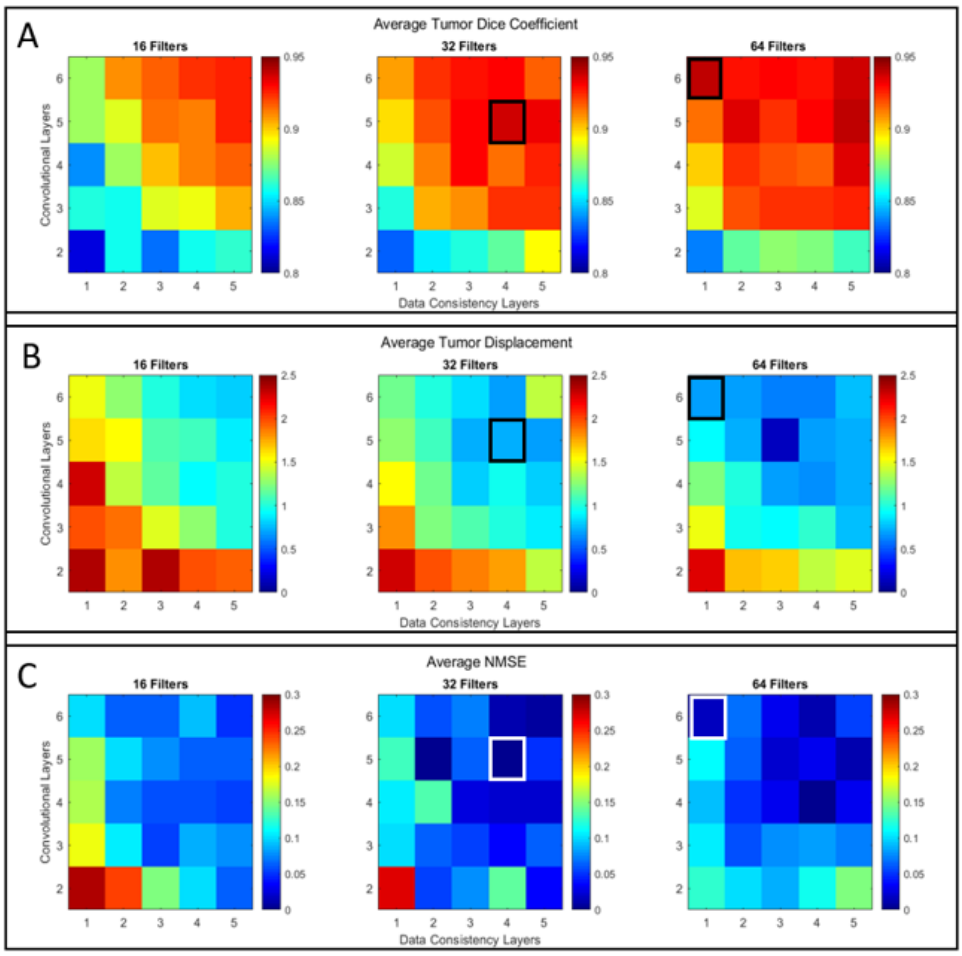

Figure 4.5: Results of a single patient parameter space search. The most important metrics for the purpose of tumour segmentation was the (A) Dice coefficient (DC) and (B) tumour displacement in units of mm. The noise (C) was evaluated via the normalized mean square error (NMSE). There were two combinations of parameters that performed well (large DC, low tumour displacement and NMSE): 32 filters, data consistency layers = 4, and convolutional layers = 5; 64 filters, data consistency layers = 1, convolutional layers = 6.

4.3.3 Retrospective Analysis

Figure 4.6 contains the CNN reconstructions for one patient as an example of the reconstruction quality. For comparison, Figure 4.6 visually compares the zero filled reconstruction, the CNN reconstruction, and the ground truth fully sampled reconstruction. It is evident that the CNN reconstruction qualitatively provides images

comparable to the ground truth data.

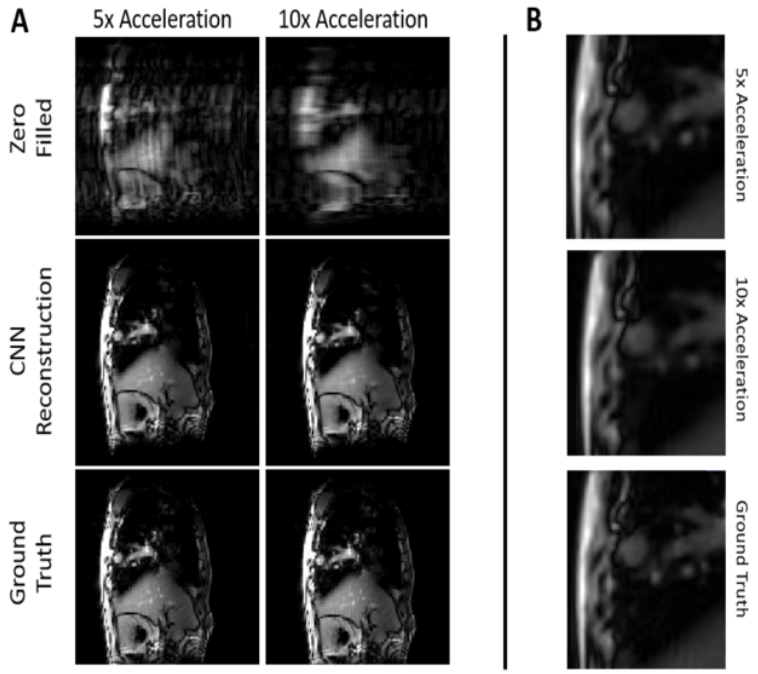

Figure 4.6: (A) Visual results of the CNN reconstruction for one retrospective non-small cell lung cancer patient. The CNN reconstruction removes the aliasing, producing images of similar quality to the ground truth data. (B) Zoomed tumorous region at 5x and 10x acceleration, compared to the ground truth tumour; the tumour structure is comparable between the three zoomed regions.

Dice Coefficient (DC)

Each of the six non-small cell lung cancer patients had CNN training at 5x and 10x acceleration, for both the 3T data and simulated 0.5T data. For each data set (3T and simulated 0.5T), the DC was calculated for each dynamic frame, from which the mean and standard deviation (SD) were calculated. Figure 4.7A contains a plot of the mean and SD values calculated over 200 frames from the CNN reconstructed data for each patient. Figure 4.7B contains a Tukey box and whisker plot of the 3T data at 10x acceleration to provide a more complete representation of the data.

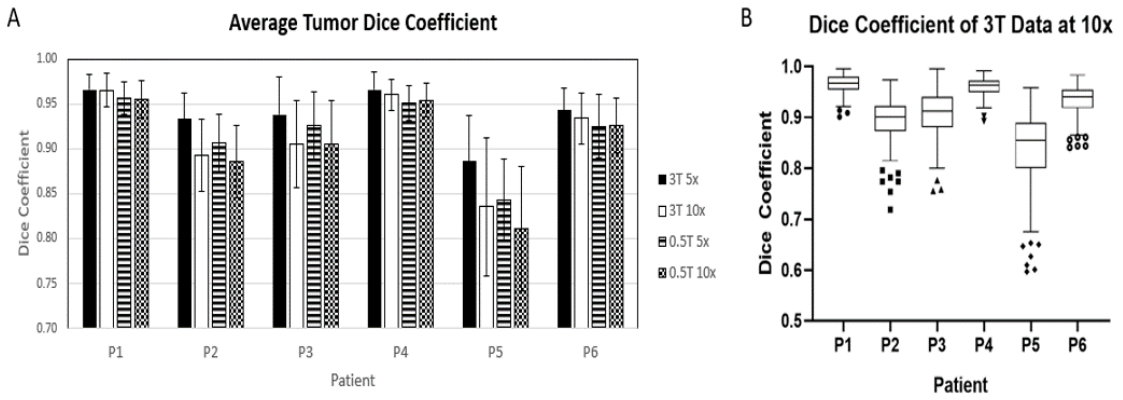

Figure 4.7: Patient averaged segmented tumour DC at 5x and 10x acceleration, for both 3T and 0.5T noise degraded data. (A) contains the mean and standard deviation of the DC computed over the 200 dynamic frames that were not incorporated into the CNN training. (B) contains the DC for 3T at 10x acceleration using a Tukey box plot, where the dots denote outliers.

With the exception of patient 5 (due to erratic tumour shape and motion), every patient had a mean DC above 0.85, most of which being above 0.9. The mean and standard deviation of the DC at 3T across all patients was calculated to be 0.94 $\pm$ 0.06 and 0.92 $\pm$ 0.03 for 5x and 10x acceleration, respectively. The DC at 0.5T across all patients was calculated to be 0.92 $\pm$ 0.03 and 0.91 $\pm$ 0.04 for 5x and 10x acceleration, respectively. The noise degraded CNN data performs similarly to the 3T data, with only slightly lower DC scores. The fine-tuning step resulted in a larger increase in the DC values for 10x acceleration, but only slightly increased the DC for 5x acceleration.

Centroid Displacement

The centroid displacement between the gold standard segmented data and the CNN reconstructed data was calculated for the six non-small cell lung cancer patients. The average displacement was under 2 mm for all patients, except for patient 5 who had an irregular shaped tumour with sporadic motion, which resulted in a larger shifted

centroid. Figure 4.8A contains the results mean and standard deviation from the segmented tumour centroid displacements. Figure 4.8B provides a breakdown of the results for the 3T data at 10x acceleration using a Tukey box and whisker plot.

Normalized Mean Square Error (NMSE)

The NMSE for the non-small cell lung cancer patients were calculated at 3T and the noise degraded (to simulate 0.5T) data at 5x and 10x acceleration. Figure 4.9C contains the plot displaying the mean and standard deviation of the NMSE over the 200 dynamic frames for each patient. The noise degraded (0.5T) data all had higher NMSE values when compared to the 3T data. This was expected due to the additive noise throughout the images. It is evident that the 10x acceleration had larger mean NMSE values, as well as a larger variation in values, due to the increased aliasing throughout the data.

Peak Signal to Noise Ratio (PSNR)

The mean and standard deviation of the PSNR shows similar trends as the NMSE. It is evident from Figure 4.9A that the 3T data had higher PSNR than the corresponding 0.5T noisier data for both 5x and 10x acceleration. Looking at the 0.5T data, patients 1, 4, and 5 had a higher PSNR at 10x acceleration.

Structural Similarity Index (SSIM)

The mean and standard deviation of the SSIM is shown in Figure 4.9B and was higher for the 3T data compared to the 0.5T data. This was to be expected given the increased noise within the 0.5T data. Patient 1 had the lowest SSIM, which can be explained by a pulsatile artifact present; as the artifact was not present near the location of the tumour, there was no decrease in the DC.

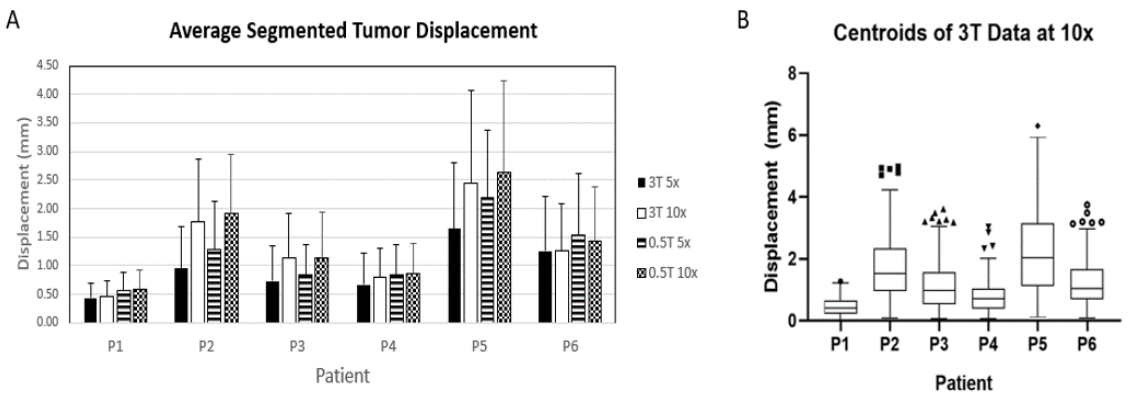

Figure 4.8: (A) Mean and standard deviation of the segmented tumour centroid displacement for each patient, at 3T and the 0.5T noise simulated data for both 5x and 10x acceleration. (B) Tukey box and whisker plot for the 3T data at 10x, to provide more detail into the breakdown of the data.

4.3.4 Qualitative Prospective Artifacts

The CNN reconstruction reduced the artifacts from the prospectively acquired 4x accelerated volunteer data as shown in Figure 4.10, demonstrating the feasibility of using CNN reconstruction in a real-time accelerated acquisition scenario. A single coil reconstruction is shown to highlight the artefacts and demonstrate the feasibility in a single coil situation. A fully sampled data set is shown for comparison (Fig. 4.10A, 4.10D); it is evident that the anatomical structures are clearly distinguishable in the CNN reconstruction. However, as shown in the CNN reconstruction (Fig. 4.10F), there remains some residual aliasing throughout the image.

The Quasar motion phantom was acquired at both 4x and 8x acceleration. The CNN reconstructed images are shown in Figure 4.11. It is evident that the artifacts are reduced in both accelerated CNN reconstructions; however, there is more of a blurring effect evident with the 8x accelerated reconstruction.

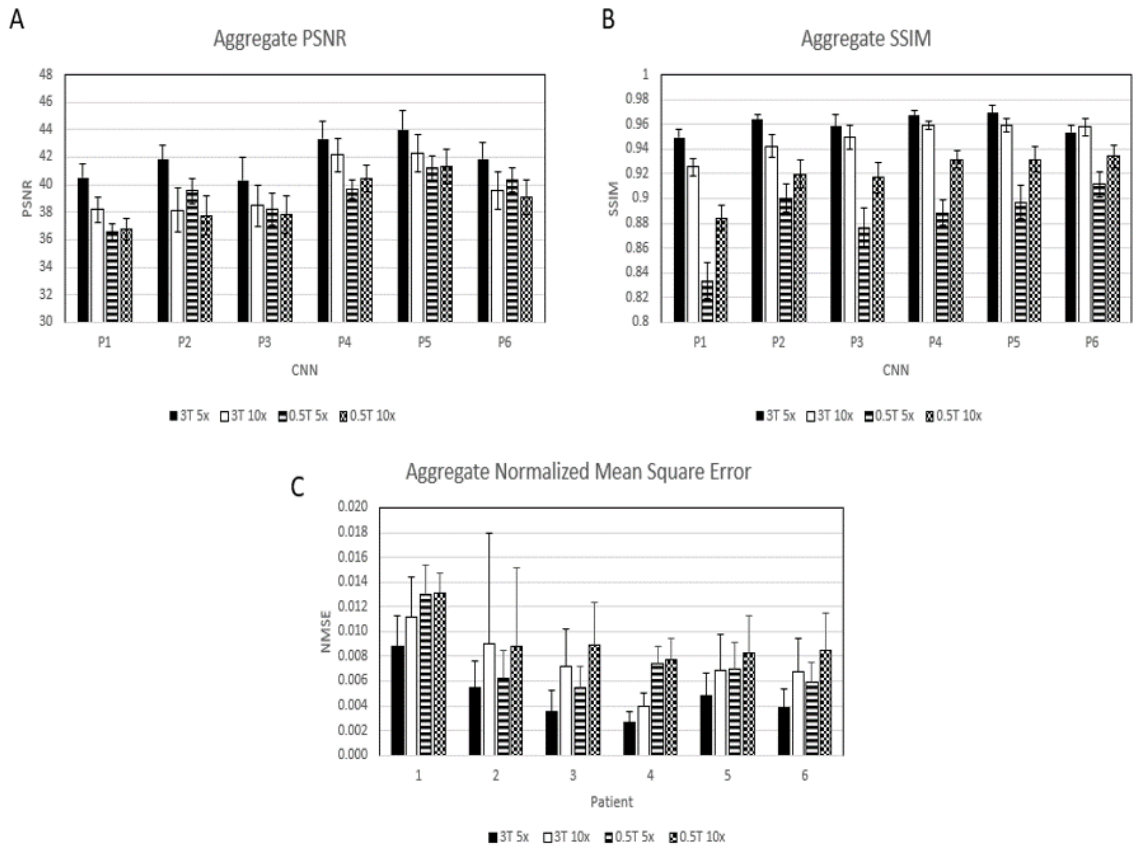

Figure 4.9: Plots displaying the (A) PSNR, (B) SSIM, and (C) NMSE for the six non–small lung cell patients. The data was aggregated over the 200 dynamic images that were not incorporated into the CNN training.

4.3.5 Reconstruction and Training Time

It is important to keep in mind that the reconstruction and training time are highly dependent on the chosen hyper-parameters and GPU hardware. Using an NVIDIA GeForce 1080Ti GPU and hyper-parameters stated earlier, the CNN training time for the retrospective and prospective reconstructions each took only 6 hours. We believe the training time can be shortened further with careful selection of hyper-parameters. Ideally, an automated technique would be used to search parameter space to find the optimal reconstruction. Determining the minimum number of parameters to yield a satisfactory image reconstruction will need to be investigated in a future study.

For the parameters chosen for this preliminary study, the reconstruction time took approximately 65 ms per image. Figure 4.14 shows a parameter plot of the reconstruction times dependency on the training parameters. It should be noted, that the Schlemper *et al.* study did not investigate real-time dynamic reconstruction; rather they investigated dynamic reconstruction (using the entire k-t domain to reconstruct the entire k-t domain) and 2D reconstruction, which does not incorporate temporal data [97].

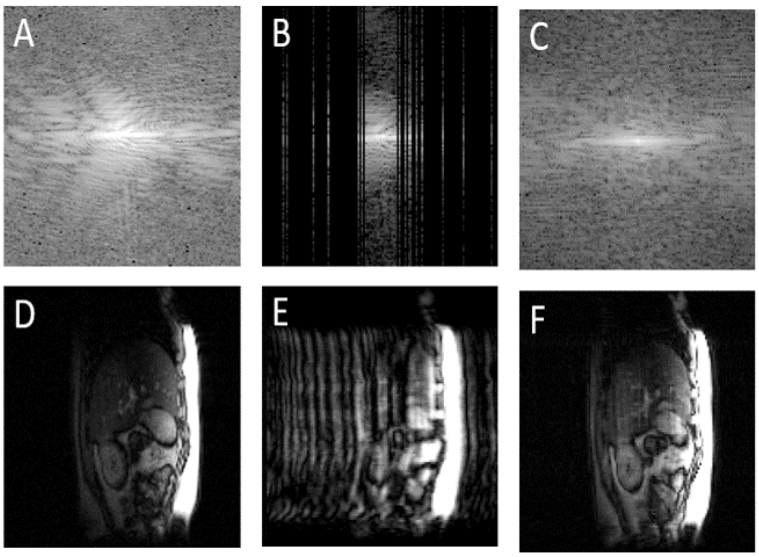

Figure 4.10: Single channel reconstruction of volunteer data. (A, B, and C) contain the k–space of the fully sampled, undersampled, and CNN reconstructed k–space data, respectively. (D, E, and F) contain the images via the inverse Fourier transform of the k–space shown in A–C. It is evident the artifacts prevalent in (E) have been reduced in (F).

4.4 Discussion

Clinical implementation of the proposed CNN reconstruction requires a planning MRI to be acquired prior to treatment. The additional free breathing MRI acquisition serves two purposes: firstly, the data would be used to train the CNN for

real-time image reconstruction during treatment; secondly, the free breathing motion data would be used to train our in-house automated segmentation algorithm. The patient would likely receive their planning MRI during the same visit they receive their planning CT. For the case of lung tumours (or any region with large motion), this would require a few minute scan to acquire a suitable training database of free breathing data. Following the acquisition, the data would be sent for CNN training and the parameters could be made available for testing the next day. For clinical simplicity, the training time could be set to an empirically determined value or made to be dynamic depending on the tumour location and size.

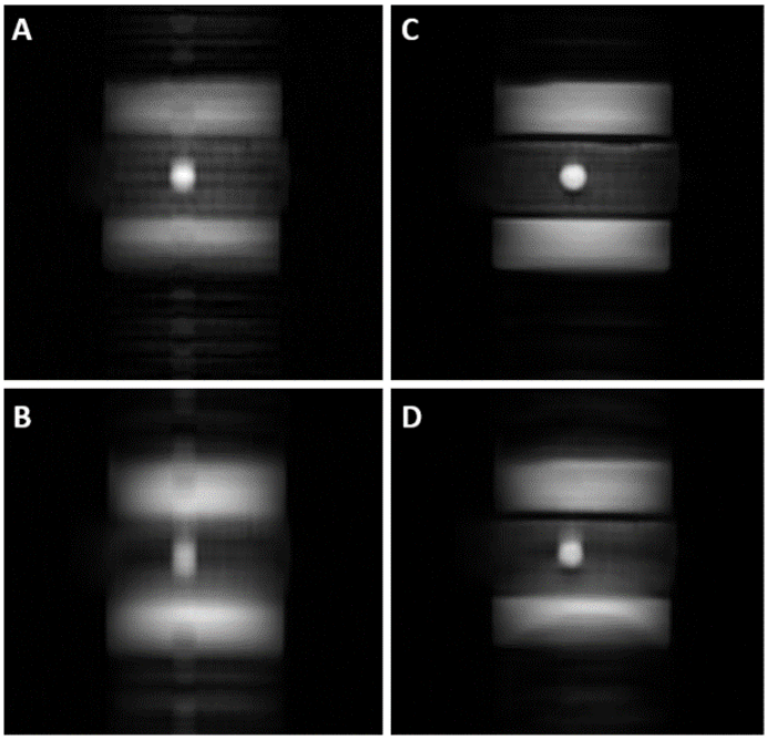

Figure 4.11: Coil combined prospective phantom reconstruction. Each channel was trained using a separate CNN. The 6 channels were combined post CNN reconstruction. (A, C) contain the 4x acceleration of the zero filled and CNN reconstructions, respectively. Similarly, (B, D) contain the 8x acceleration reconstructions.

The mean Dice Coefficient (DC) across the patient group was greater than 0.9 for

nearly all cases. This compares favorably to the variation between contours drawn by the same expert radiation oncologist (ROs) in different sessions (intra-observer) on the same fully sampled data sets (mean DC of 0.88) as reported in the literature [128]. Further, inter-observer contour variation between multiple ROs was reported to have a mean DC of 0.87 [128]. Thus, the mean contour DC produced using our CNN reconstruction of accelerated acquisitions (and noise addition in the case of the 0.5T simulated data) was greater than that seen in manual segmentation by an RO. Furthermore, the CNN technique was found to be an improvement to the CS-PCA technique developed in Chapter 3. Figure 4.12 compares the two techniques via the analysis of the DC for all six patients and the temporal evolution of the NMSE of two patients (patient 1 and patient 5). The analysis consisted of the last 200 dynamic frames of the data set.

Another metric that is of importance to investigate temporally is the centroid displacement. In particular, how the centroid displacement temporally evolves between two real-time techniques; CS-PCA and CNN. Figure 4.13 contains temporal data between the two techniques for patients 1 and 5, to evaluate a simple and worst case. The distributions for the two techniques are similar; however, the CNN technique did have smaller mean CD values.

We also investigated the use of the 3T trained parameters to reconstruct the accelerated noisy (0.5T SNR simulated) data. We found that it made no significant difference using parameters trained using 3T data for the 0.5T SNR simulated data. The importance of this being that the field strength of the MRI system used to acquire the training data may differ from the field strength of the treatment Linac-MR. However, images acquired with different field strengths will additionally have differing contrast, which may decrease the quality of the reconstruction. Training with

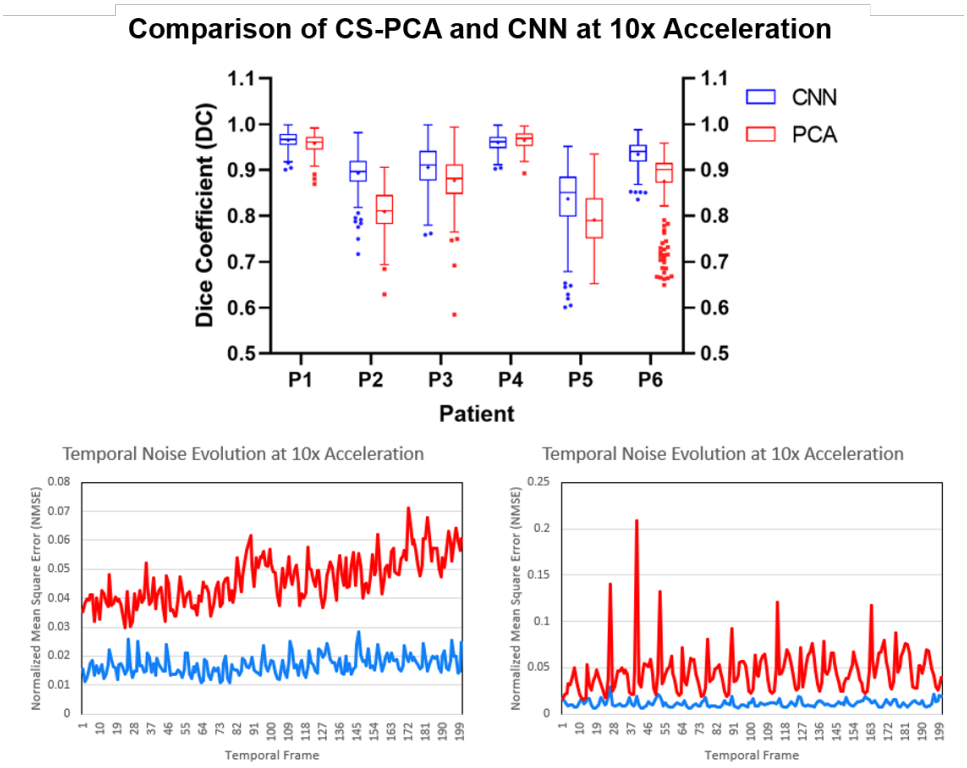

Figure 4.12: The DC was compared for all six patients at 10x acceleration, using a Tukey box-and-whisker plot. The NMSE plots contain the data from two patients, the left plot is of patient 1 and the right plot is of patient 5, which had a more difficult tumour to segment. It is evident that CNN reduces the noise and removes the pulsatile effect seem in the CS-PCA data.

images at a different field strength than the acquired prospective data will need to be investigated.

Our current implementation of the CNN reconstruction required a total CNN training time of 6 hours. Our computer hardware consisted of a Dell Precision tower equipped with an Intel Xeon E5-2650 CPU, 64 GB of RAM, and a single NVIDIA GTX 1080Ti GPU. The training time can be reduced by determining the optimal CNN hyper-parameters. More sophisticated techniques such as competing objectives [129] or adaptive particle swarm optimization [130] can be used to search parameter space for the optimal hyper-parameters. These hyper-parameters may vary by tumour site

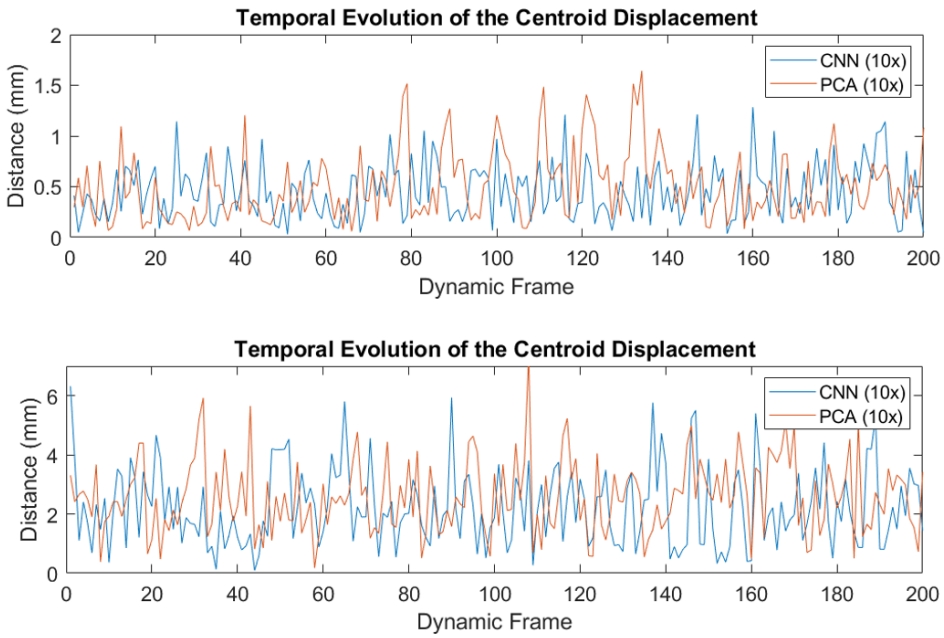

Figure 4.13: The centroid displacement is plotted over 200 dynamic frames for both the CNN and PCA technique. Both techniques had a similar distribution of centroid displacement values for each patient; although the CNN technique had smaller (global) mean values.

and per patient; however, future studies would be required to investigate this in depth.

One patient (patient five) had a tumour that resulted in sub-optimal results, in terms of the DC. As shown in Figure 4.3, this tumour has a larger region of higher signal along with a long skinny region of lower signal. The CNN reconstruction tends to blur out the long skinny region of the tumour, resulting in the segmentation algorithm only contouring the larger high signal portion. We have recently developed a more rigorous segmentation tool, which we believe may solve this issue and will be investigated in future studies.

There are several limitations to this study. We did not include data augmentation within our CNN training regimen. Often used to avoid over fitting in neural network models, the input training data can be augmented via translations, rotations,

and mild warps. While we did not implement this with our current study, we believe it would result in a more robust training dataset, further minimizing the effects of unwanted patient motion during imaging. The use of data augmentation will be investigated in the future, the only downside being a longer training time. Furthermore, the quantitative data was retrospectively undersampled, and may not be a true representation compared to the prospectively undersampled case. Prospectively undersampled patient data will be acquired for future studies.

We have shown that the use of CNNs for real-time on-the-fly image reconstruction is a viable technique via our prospective analysis. In the prospective case, each MRI coil should be trained and reconstructed independently, meaning multiple GPU's (one for each coil) would be ideal for real-time reconstruction.

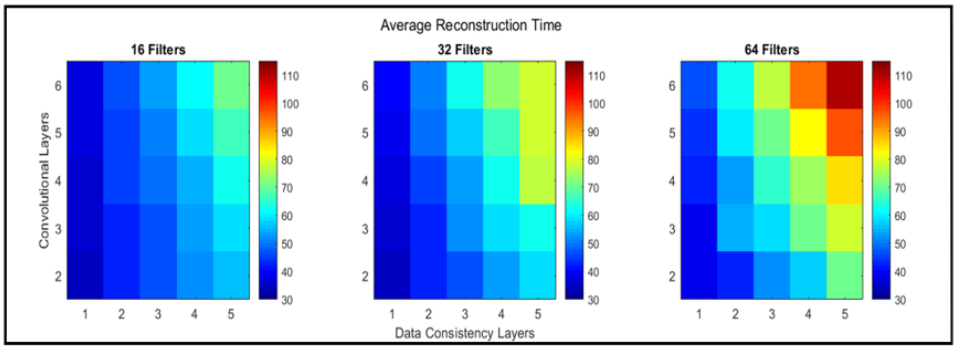

Figure 4.14: Reconstruction time dependency (units of ms) on the CNN training parameters. As the number of layers increase, so too does the reconstruction time.

4.5 Conclusion

We have demonstrated that 3D convolutional neural networks (CNNs) can be used for real-time on-the-fly dynamic image reconstruction utilizing both spatial and temporal data. We have specifically demonstrated the ability to use 3D CNNs trained on a single patient to reconstruct undersampled MR data from the same patient. A

retrospective analysis was conducted using data from six non-small cell lung cancer patients, where we evaluated tumour segmentation. The mean Dice coefficient for our tumour segmentation remained above 0.82 for all CNN reconstructed data, with five of six patients having a Dice coefficient above 0.9. Further validation was conducted using prospectively undersampled data from a volunteer at 4x acceleration and a motion phantom at acceleration factors of 4x and 8x. We achieved reconstruction times of 65 ms using a single NVIDIA GTX 1080Ti GPU. The total training time of 6 hours can be reduced with optimal parameter selection; however, a more in-depth study into determining the optimal parameters is required. We intend to investigate higher acceleration factors in the future. As computer technology keeps advancing, it is likely that CNNs reconstruction using GPUs will become faster, reducing both the training and reconstruction times further.

Chapter 5

Low Resolution Phase Encoding for Real-Time CNN Reconstruction

Parts of this chapter have been adapted from a published article: Dietz B, Yip E, Yun J, Gabos Z, Fallone G, Wachowicz K. Single Patient Convolutional Neural Networks for Real-time MR Reconstruction: Coherent Low-Resolution versus Incoherent Undersampling. Physics in Medicine and Biology. 2020;65(8) [131].

5.1 Introduction

As discussed in the previous chapter, the use of convolutional neural networks (CNNs) is becoming an increasingly popular trend in image processing and reconstruction [99, 93, 95, 94, 97, 116, 132].

The previous chapter investigated the use of CNN using incoherently undersampled MRI acquisition, to achieve a higher frame rate. Reconstruction of an incoherently undersampled k-space results in a noise-like artefact occurring throughout

the image domain, however, the implementation of incoherent undersampling is not without challenges as its use can present confounding eddy current artefacts. The increased eddy current contributions are caused by the rapid gradient switching during the inconsistent phase-encoding jumps required by the pseudo-random sampling. These eddy current artefacts may result in suboptimal reconstruction when using techniques that rely on previously acquired (fully sampled) data, such as CNN reconstruction techniques, which may have differing eddy current contributions. The suboptimal reconstruction is due to the fully sampled acquisition (used for training) having a differing eddy current contribution from the incoherently acquired (prospectively undersampled) data. Techniques can be used to mitigate eddy currents during imaging [127]; however, they may not reduce all the eddy current artefacts and can be restrictive in terms of sequence playout.

One way to avoid issues associated with incoherent encoding, is to acquire only the central portion of k-space and exclude the outer regions collected during the incoherent acquisition. This alternative k-space acquisition technique results in a low-resolution reconstruction in image space. While this does not give sufficient data to generate a fully-resolved reconstruction on its own, when reconstructed using a CNN trained on a previously acquired fully-sampled data set the unsampled data can be inferred during reconstruction, whether in the high spatial resolution region alone or scattered throughout k-space as in the incoherent-sampled case. Examples of the two phase encoding (PE) schemes are shown in Figure 5.1. This coherent low-resolution (hereafter known as coherent-LR) undersampling is ideal, as it is simple to implement on any MR system and does not lead to eddy current artefacts, due to both the fully sampled and coherent-LR undersampled data having identical sampling intervals. The previously acquired fully sampled data, on which the CNN is trained, will more closely resemble the data acquired via the coherent-LR undersampled acquisition.

It is the purpose of the trained CNN to generate this high resolution data based on fully sampled data, from which relationships between low resolution structure and high resolution detail can be established. It is the authors' hypothesis, that this will be a more straightforward task for the CNN to fulfill and may result in a more robust algorithm. Our study investigates and compares the use of incoherent and coherent-LR undersampled k-space acquisitions for real-time on-the-fly 3D CNN reconstruction, both retrospectively and prospectively.

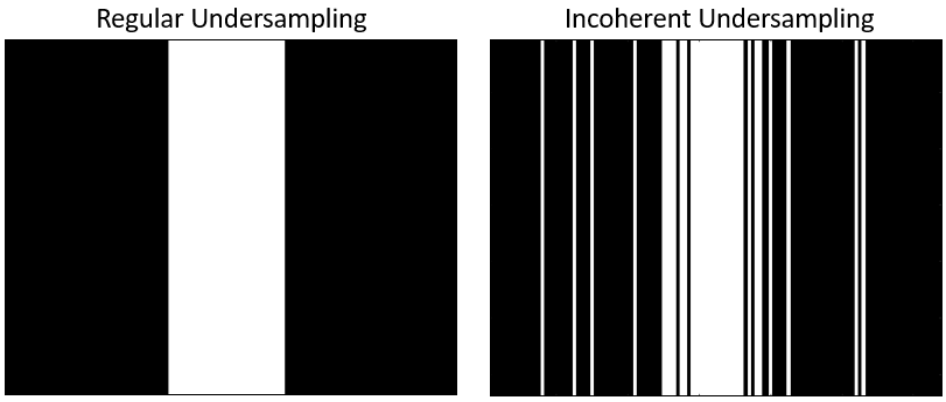

Figure 5.1: Coherent-LR verses incoherent phase encoding schemes. At 4x acceleration, each undersampling scheme uses 32 phase encodes (assuming 128 x 128 image). The coherent-LR undersampling only samples the central band of k-space, whereas the incoherent undersampling is focused on the central band, but also utilizes a pseudo-random sampling of the outer k-space regions.

A common paradigm for the use of (convolutional) neural networks involves training a network using a large cohort of patient data in order to recover information from a single patient. Our paradigm differs, in that we focus on a personalized healthcare approach, such that a CNN is specifically trained for each patient. Training the CNN requires a previously acquired fully sampled data set for training the CNN, which is used to reconstruct the accelerated data acquisition.

In this study we present the use of 3D CNNs (two spatial dimensions, and one temporal dimension) to retrospectively reconstruct undersampled dynamic data from

six non-small cell lung cancer patients on a patient-by-patient basis. A quantitative comparison of coherent-LR and incoherent acquisitions for the use of CNN MR image reconstruction was conducted. We focus on the ability to track and segment a tumorous region located in the lung. Lung was specifically chosen, as breathing motion is a challenging component to account for using conventional linac systems. We further validate and compare the two encoding schemes in the realistic prospective imaging scenario, where the fully sampled training data is acquired separately from the accelerated undersampled data. By reducing the images acquisition and reconstruction time, a higher frame rate with low latency can be achieved, which is of importance for adaptive radiotherapy. The aim of this study is to develop a patient-specific 3D CNN reconstruction technique for use in a real-time, on-the-fly, scenario for LMR adaptive radiotherapy purposes.

5.2 Methods

5.2.1 Convolutional Neural Network

The original CNN code used for this study was written by Schlemper *et al.* in Python using Theano and Lasagne packages [97], that has been adapted for the use in real-time dynamic image reconstruction using 3D CNNs [116]. A dynamic CNN network was proposed by Schlemper that aimed to reconstruct the entire temporal domain in spatial patches; however, the application to rapid real-time reconstruction was not addressed. Our adaptation allows for real-time on-the-fly reconstruction that incorporates both spatial and temporal data. We further adapted the CNN code to allow the use of coherently undersampled data.

The CNN image reconstruction can be expressed as

$$\mathbf{x}_{cnn} = f_{cnn}(\mathbf{x}_u|\vec{\theta}, \lambda, \mathbf{\Omega}) \tag{5.1}$$

where the term $\mathbf{x_u}$ is the zero-filled undersampled image and $f_{cnn}(\mathbf{x_u}|\vec{\theta}, \lambda, \mathbf{\Omega})$ is the CNN reconstructed image. The function f_{cnn} attempts to reconstruct the undersampled image $\mathbf{x_u}$ and depends on the trained CNN parameters $\vec{\theta}$, a data fidelity weighting term λ, and the known k-space locations $\mathbf{\Omega}$. The data fidelity term λ is calculated such that $\lambda = q/\sigma$, where q is a hyper-parameter and σ^2 is the noise power. The λ term is initialized to 0.025 as it was empirically shown that this value performs well [98]. Given the fully sampled input training data D, the CNN is trained to reconstruct the data by minimizing the following cost function,

$$Z(\theta) = \sum_{(\mathbf{x_u}, \mathbf{x_{gnd}}) \in D} ||\mathbf{x}_{gnd} - \mathbf{x}_{cnn}||_2^2. \tag{5.2}$$

The terms $\mathbf{x}_{gnd}$ and $\mathbf{x}_{cnn} = f_{cnn}(\mathbf{x}_u|\vec{\theta}, \lambda, \mathbf{\Omega})$ are the ground truth (fully-sampled) image and the CNN reconstructed image (reconstructed from the retrospectively undersampled data), respectively. Once trained, the parameters can then be rapidly applied directly to the undersampled data (either prospectively or retrospectively). The Adam optimizer was implemented for the CNN training [82].

The CNN is comprised of a deep network of alternating convolutions layers and data consistency layers. The convolutional layer convolves filters with the undersampled data in order to inhibit the aliasing artefacts or low-resolution blur corresponding to the raw undersampled data. The convolutional layer applies N convolutions sequentially, which are each followed by the rectifier linear units (ReLU) non-linearity activation function. The final N^{th} convolution output is summed with the initial

input data, via the process known as residual connection [126]. Following the convolutional layer the data is passed to the consistency layer, which enforces data fidelity by ensuring that the acquired portion of k-space is always incorporated into the 3D CNN. The data consistency step takes the image $\mathbf{x}_{cnn}$ (output from the convolutional layer), which is Fourier transformed (FT) into k-space. If no k-space was initially acquired, it is equal to the value from the FT of $\mathbf{x}_{cnn}$; however, if k-space was initially acquired then the value is equal to a weighted sum of the original k-space and the FT of $\mathbf{x}_{cnn}$. The process of convolutions followed by data consistency continues for M data consistency layers.

5.2.2 CNN Hyperparameter Optimization

Finding the optimal hyperparameters for our CNN was done systematically by setting the number of convolutional layers, data consistency layers, and number of convolution filters. The CNN was trained for a reduced number of iterations (25), and the reconstructed data was quantitatively evaluated using metrics described below. The systematic hyperparameter search was conducted for each retrospective patient. We evaluated the convolutional layers from two to six, data consistency layers from one to five, and the number of filters was set to either 16, 32, or 64. This resulted in a total of 75 sets of hyperparameters for each patient.

5.2.3 Tumour Segmentation via Autocontouring

An automatic contouring program previously developed by our group was used for the automatic segmentation of the lung tumours [114, 36, 119]. Using contours drawn on several phases of the breathing cycle as input data (a total of 30 images were used), the program trains a neural network to find the region of interest (tumour)

Patient	Area (cm^2)	Sup-Inf Extent (cm)	Ant-Post Extent (cm)
1	2.57 (0.09)	1.97	0.40
2	3.76 (0.25)	2.63	0.75
3	1.36 (0.16)	1.57	0.66
4	4.96 (0.31)	0.94	0.39
5	2.09 (0.19)	1.07	0.72
6	4.99 (0.30)	0.87	0.17

Table 5.1: Tumour statistics for each of the six patients. The area contains the mean and standard deviation over the 650 fully sampled dynamic images. It is evident from the motion extents (superior-inferior and anterior-posterior) that the tumours ranged in size and motion.

and segment it. This contouring program was used for all segmentations performed in this study, which was trained using contours drawn by a radiation oncologist.

5.2.4 Retrospective Data Analysis

The acquisition of data from six non-small cell lung cancer patients was subject to approval by our institutional board of ethics. Lung cancer was chosen as it often involves large tumour motion from breathing, making it an ideal challenging case for real-time imaging and segmentation for LMR tumour tracking applications. Figure 4.3 contains images displaying the field of view (FOV) with the tumour present. Table 5.1 contains tumour specific information for each patient.

The retrospective data sets comprised of 650 dynamic frames acquired with a Cartesian bSSFP sequence using a multi-channel chest array. The data were acquired using our 3T Philips Achieva system (Philips Medical Systems, Cleveland, OH, USA). The MR imaging parameters consisted of a 40 x 40 cm^2 FOV, a slice thickness of 2 cm, 128 x 128 voxels, a flip angle of 20^o, and TR/TE = 2.2/1.1 ms. The dynamic data were acquired at a rate of 4 frames per second. The data was split up such that the first 450 dynamic frames were used to train the CNN and the remaining 200 dynamic frames were used to quantitatively evaluate the CNN. In order to robustly train the

CNN, a 5-fold cross validation was used on the training data portion (450 dynamic frames). Thus, each CNN k-fold iteration used 90 dynamic frames for test data and trained on the remaining 360 dynamic frames. Figure 5.2 contains a schematic of the k-fold training technique. The training data was mini-batched into groups of 5 images and shuffled to further increase the generalization of the model.

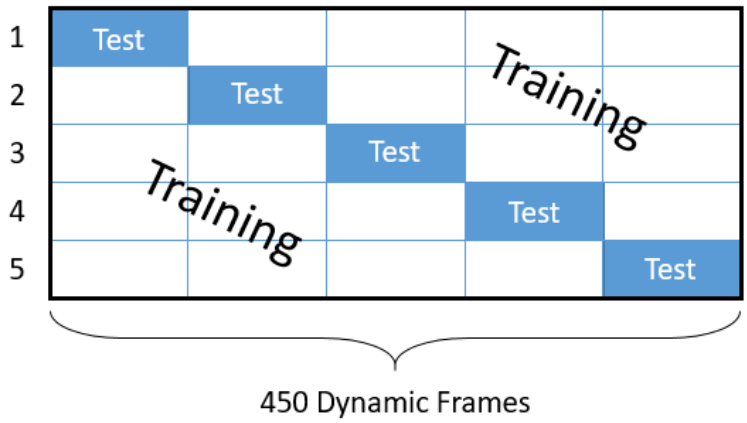

Figure 5.2: Schematic of the 5-fold cross-validation training paradigm.

5.2.5 Prospective Data Analysis

Prospectively undersampled data (at 4x acceleration, 32 phase encodes instead of 128), along with a fully sampled data set (for CNN training), was acquired from a volunteer. The acquisition imaging parameters were identical to the retrospective case. However, here we have two separate acquisitions; a fully sampled acquisition (exact same parameters as the retrospective case) and an undersampled acquisition at 4x acceleration. The training of the CNN was conducted similarly to the retrospective case; except the training data comprised of 650 frames instead of the 450 frames as in the retrospective training (since the undersampled data is acquired separately). The undersampled acquisition contained the same parameters, except that only the central 32 lines of k-space were acquired, resulting in data of the size 32 x 128 pixels

per dynamic frame. The remaining 96 unacquired lines of k-space were zero-filled (48 on either side of the acquired 32 lines), ultimately resulting in a matrix size of 128 x 128 voxels.

5.2.6 Quantitative Analysis

The metrics we have chosen are specific for the analysis of tumour segmentations, being that our focus lies on the ability to segment a tumour for the purpose of real-time adaptive radiotherapy. The image quality, while an important aspect for image reconstruction, is secondary to the tumour segmentation.

Dice Coefficient

The Dice coefficient is an ideal way to compare how well the tumour can be resolved in the CNN reconstructed data, by comparing the Boolean segmentation with the fully sampled segmentation [110]. This metric compares how well the two segmentations overlap. A DC is equal to unity if the two segmentations overlap entirely, meaning there is no depreciable difference in the image reconstruction using the CNN technique. The Dice coefficient can be measured as,

$$DC = 2 \cdot \frac{ROI_{FS} \cap ROI_{US}}{ROI_{FS} + ROI_{US}} \tag{5.3}$$

where ROI_{FS} is the contour region of interest (ROI) for the fully sampled data and ROI_{US} is the contour for the undersampled CNN reconstructed data.

Centroid Displacement

The centroid displacement offers another metric to evaluate how well the CNN reconstructed segmentation matches the fully sampled segmentation. The centroid

displacement is simply the difference in the center of masses between the two binary segmentations. The smaller the value, the closer the two overall segmentations overlap. This metric evaluated in concert with the DC gives a more complete picture of how well the CNN reconstruction is performing, for the purpose of segmentability.

Structural Similarity Index

The structural similarity index (SSIM) attempts to evaluate the perceived quality through the comparison of two data sets [112]. The SSIM requires the computation of three distinct terms: the luminance, contrast, and structural terms. Combining these terms, we get an expression

$$SSIM = \frac{(2\mu_F\mu_{US} + C_1)\ (2\sigma_{F,US} + C_2)}{(\mu_F^2 + \mu_{US}^2 + C_1)\ (\sigma_F^2 + \sigma_{US}^2 + C_2)}. \tag{5.4}$$

The terms μ_{FS}, μ_{US}, σ_{FS}, and σ_{US} are the means and standard deviation of the fully sampled and undersampled data respectively. The term $\sigma_{(FS,US)}$ is the cross-correlation of the fully and undersampled images. The constants C_1 and C_2 are based on the dynamic range (DR) of the two image sets; these values are typically set to $C_1 = (0.01 \cdot DR)^2$ and $C_2 = (0.03 \cdot DR)^2$.

5.3 Results

5.3.1 CNN Hyperparameter Search

An exploration of the optimal CNN hyperparameters was conducted using the DC metric. Figure 5.3 contains a visual representation of the searched hyperparameter space. In general, increasing the number of layers and/or filters did increase the DC; however, the reconstruction time also increased. Based on the results of the hyper-

parameter search we chose to use four convolutional layers, four data consistency layers, and 16 filters. This proved to be robust on all retrospective patients and had a short reconstruction time of 54 ms per dynamic frame. For the prospectively acquired volunteer data, we chose to use four convolutional layers and four data consistency layers; however, we chose 64 filters. While this resulted in a slightly increased reconstruction time, it provided a qualitatively better result (data not shown) than using 16 or 32 filters.

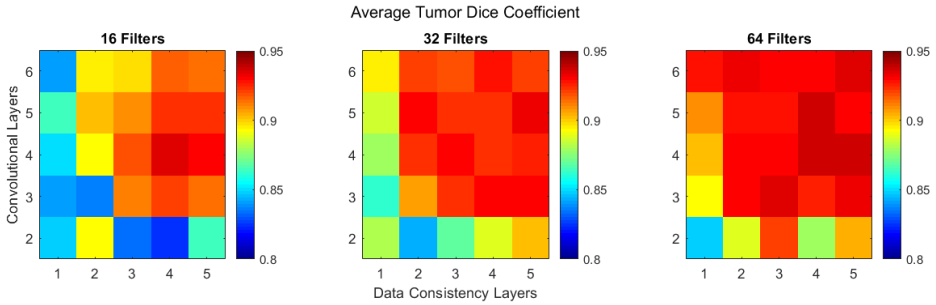

Figure 5.3: Dice coefficient results of the hyperparameter search. Based on the results we chose: 4 convolutional layers, 4 data consistency layers, and 16 filters. The reconstruction time increases as the number of filters, convolutional layers, or data consistency layers increase.

5.3.2 Retrospective CNN Analysis

Dice Coefficient

The DC was calculated for both the incoherent and coherent-LR undersampled phase encoding schemes at 5x and 10x acceleration for the six retrospective lung tumour patients. Figure 5.4 contains Tukey box and whisker plots of the DC, performed on the 200 dynamic frames that were not included in the training of the network.

The percent changes for the mean and standard deviations of the DC and CD (from incoherent to coherent-LR) are shown in Figure 5.5. It is evident from Figures 5.4 and 5.5 that the DC from the coherent-LR undersampled PE scheme had similar or

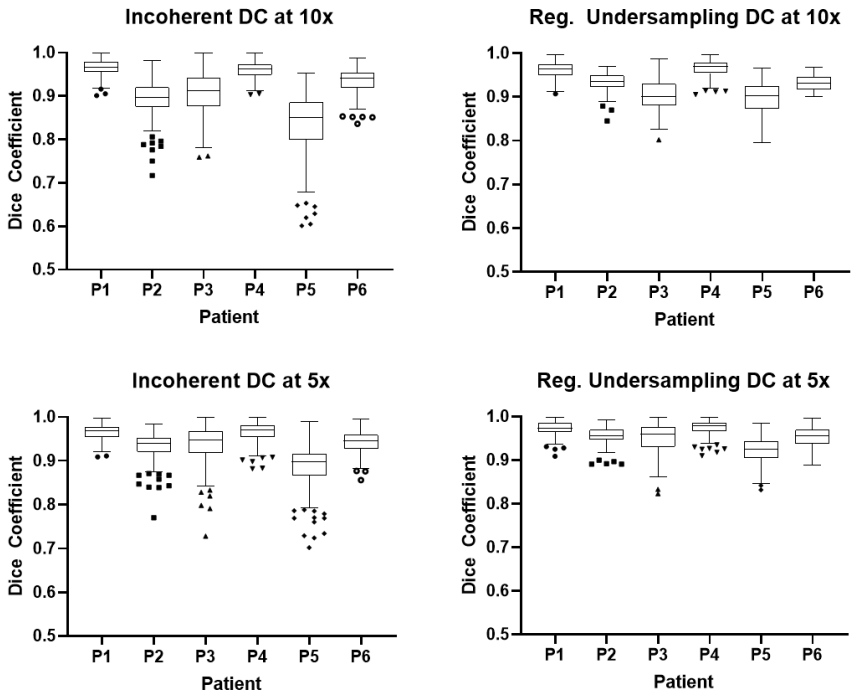

Figure 5.4: Tukey box plots containing the DC results from the 200 CNN reconstructed images using the incoherent and coherent-LR undersampled PE schemes at 5x and 10x acceleration. It is evident that that the coherent-LR undersampling scheme for both 5x and 10x acceleration provide a larger mean DC, and less variation than the incoherent phase encoding.

larger mean DCs compared to the incoherent PE scheme. The coherent-LR PE also had less variability (smaller standard deviations) than the incoherent PE. Patients 2 and 5 had the greatest increase in mean DC (2.3% and 4.2% at 5x acceleration; 4.4% and 7.0% at 10x acceleration) and reduction in variance of the DC values (-48% and -55% at 5x acceleration; -71% and -67% at 10x acceleration); patients 1, 3, and 6 had a small decrease in mean DC (-0.4%, -0.2%, and -0.3% at 10x acceleration; respectively). However, the standard deviation for patient 1, 3, and 6 decreased by 8.6%, 36.5%, and 49.1%; respectively.

Furthermore, we have included an example of a poor segmentation for the most difficult patient (P5), as shown in Figure 5.6 for both phase-encoding schemes. It is

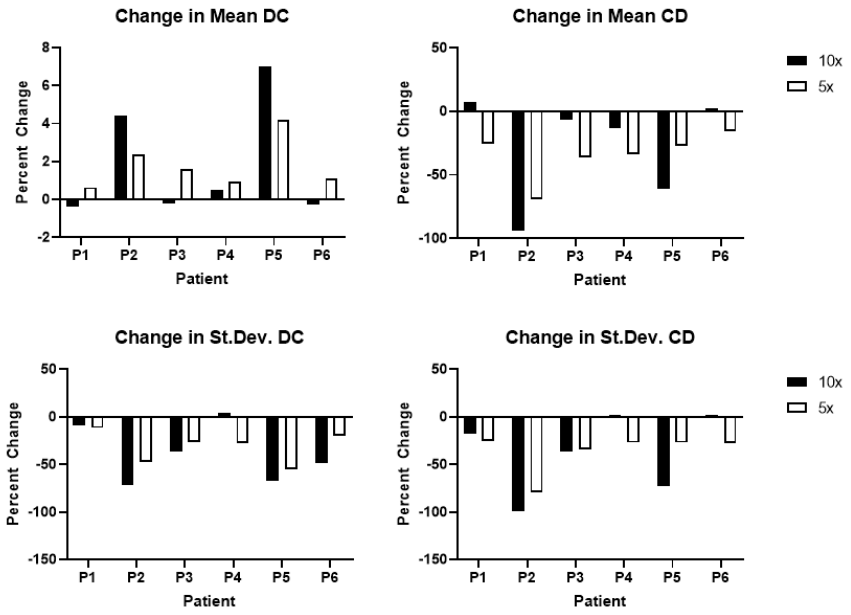

Figure 5.5: The percent change in the mean and standard deviation for the Dice coefficient (DC) and centroid displacement (from incoherent to coherent-LR). Positive percent change indicates that the coherent-LR PE had higher DC values. Nearly all patients had an increased DC when using of the coherent-LR phase-encoding scheme, with the exception of P1, P3, and P6 at 10x acceleration. The DC standard deviation decreased or remained unchanged for all patients except P4 at 10x acceleration. The mean CD decreased for all patients except for P1 and P6 at 10x acceleration. The CD standard deviation decreased for all patients, except for P4 and P6 which only very slightly increased.

evident that a blurring of the tumorous region occurred, resulting in a reduced DC of 0.610 and 0.795 for the incoherent PE and coherent-LR PE, respectively. The reduced DC may be a potential limitation of the acceleration and segmentation itself; however, this example demonstrates how the coherent-LR improves upon the incoherent PE. This patient was particularly difficult to segment, given the long tumour shape and sporadic motion.

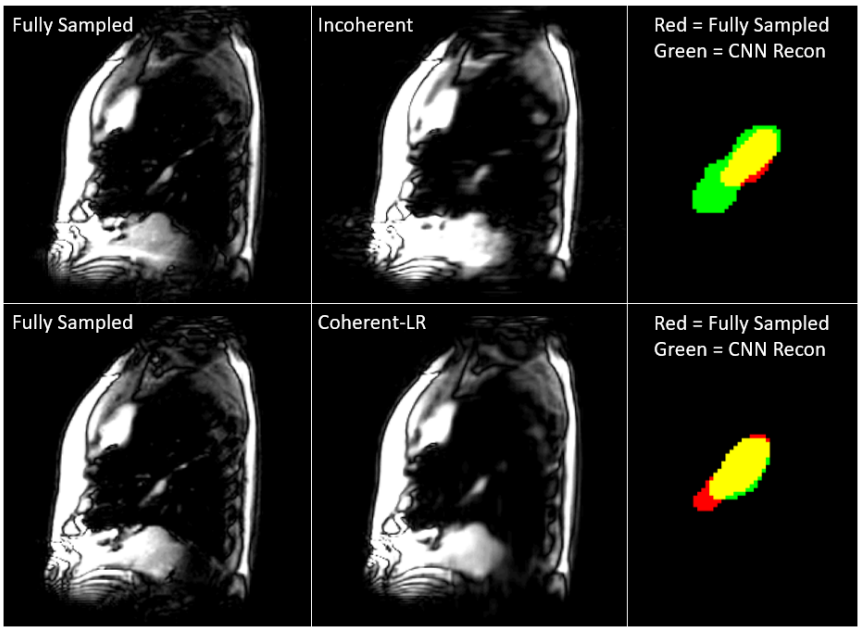

Figure 5.6: Example of the worst-case DC for patient 5 for the incoherent and coherent-LR phase-encoding with values of 0.610 and 0.795, respectively. It is evident for both PE schemes that some fine detail was lost; however, the incoherent had more significant blurring.

Centroid Displacement

The centroid displacement (CD) was calculated for both the incoherent and coherent-LR undersampled phase encoding schemes at 5x and 10x acceleration for the six retrospective lung tumour patients for the 200 dynamic frames. Figure 5.5 contains the percent changes for the mean and standard deviation values for both phase-encoding schemes for each patient and acceleration factor. As shown in Figure 5.5, the centroid displacement is reduced (negative values indicate that coherent-LR had lower CD values) for the coherent-LR undersampled case for both acceleration factors, with the exception of patients 1 and 6 at 10x acceleration. The variation of the centroid displacement values was reduced, similarly as the DC values, suggesting the coherent-LR undersampled phase encoding scheme is superior to the incoherent phase encoding scheme for our purpose. Furthermore, the CD had a greater improvement using the

coherent-LR phase-encoding scheme than the DC for both acceleration factors.

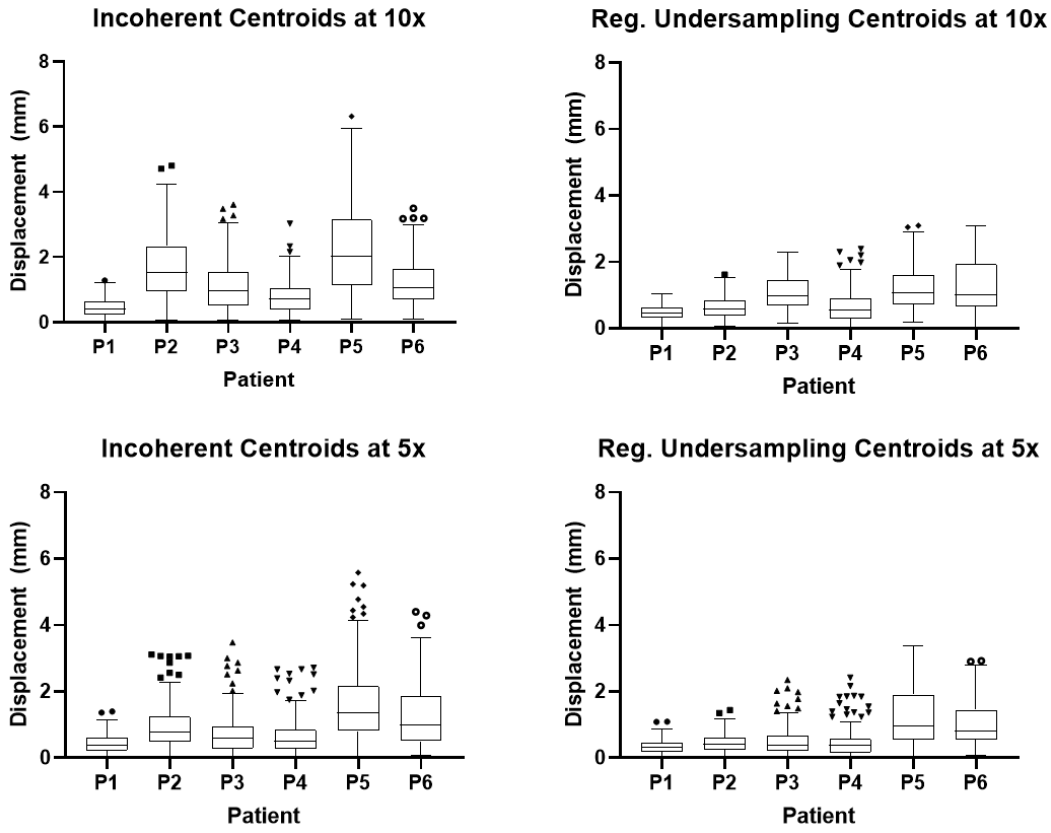

Figure 5.7: Tukey box plots containing the results of the centroid displacement between the CNN reconstructed data and the fully sampled data. The mean and variation of the segmented tumour was reduced when using the coherent-LR undersampling phase encoding scheme, compared to the incoherently undersampled data.

Structural Similarity Index

The SSIM was averaged over the 200 reconstructed frames at 5x and 10x acceleration for both phase encoding schemes. The mean and standard deviation are plotted in Figure 5.8, which demonstrates that the coherent-LR image quality improved for all patients at both acceleration factors.

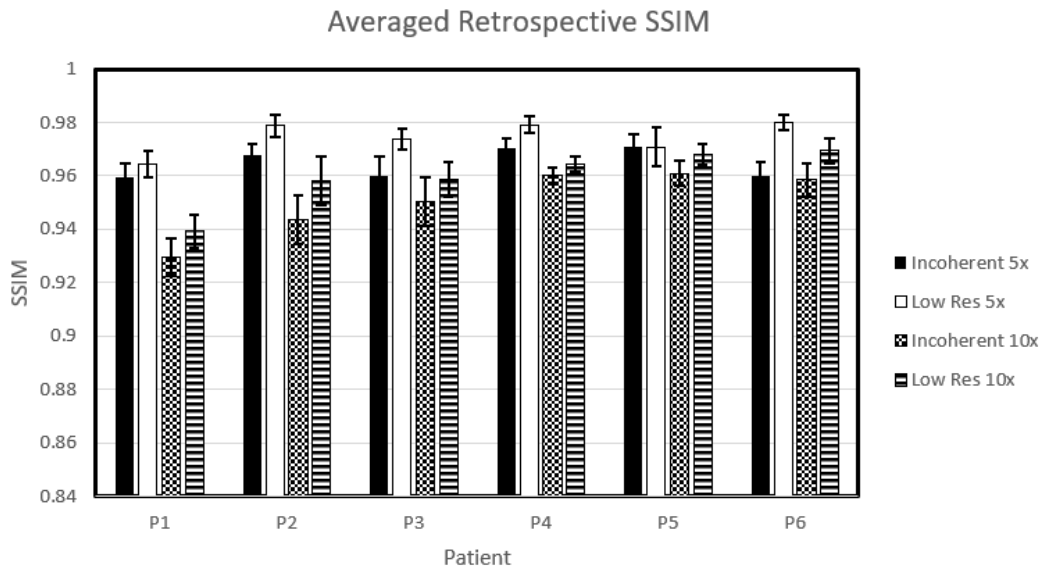

Figure 5.8: The SSIM showed improvement for all patients when using the coherent-LR undersampled phase encoding scheme, with the exception of patient 5 at 5x acceleration.

5.3.3 Prospective CNN Analysis

Figure 5.9 contains spatial and temporal views (taken through the center of the spatial domain images) of the CNN reconstructed volunteer data using the coherent-LR and incoherent undersampled phase encoding schemes. The fully sampled training data, and the prospectively acquired (zero-filled) data are also shown for comparison. While the CNN reconstruction removed much of the aliasing from the incoherently acquired data, there were still artefacts present, which manifests as a faint rippling throughout the image. Visual inspection of the temporal domain demonstrates that the coherent-LR provides a qualitatively better reconstruction. The artefacts did not appear to be present in the coherent-LR reconstruction. This demonstrates that the CNN technique does in fact work with real world accelerated data that has been trained on a separate fully sampled data set (which is a different scenario than the retrospective case). This figure thus has only qualitative value, since there is no fully

sampled "ground-truth" data set to compare to.

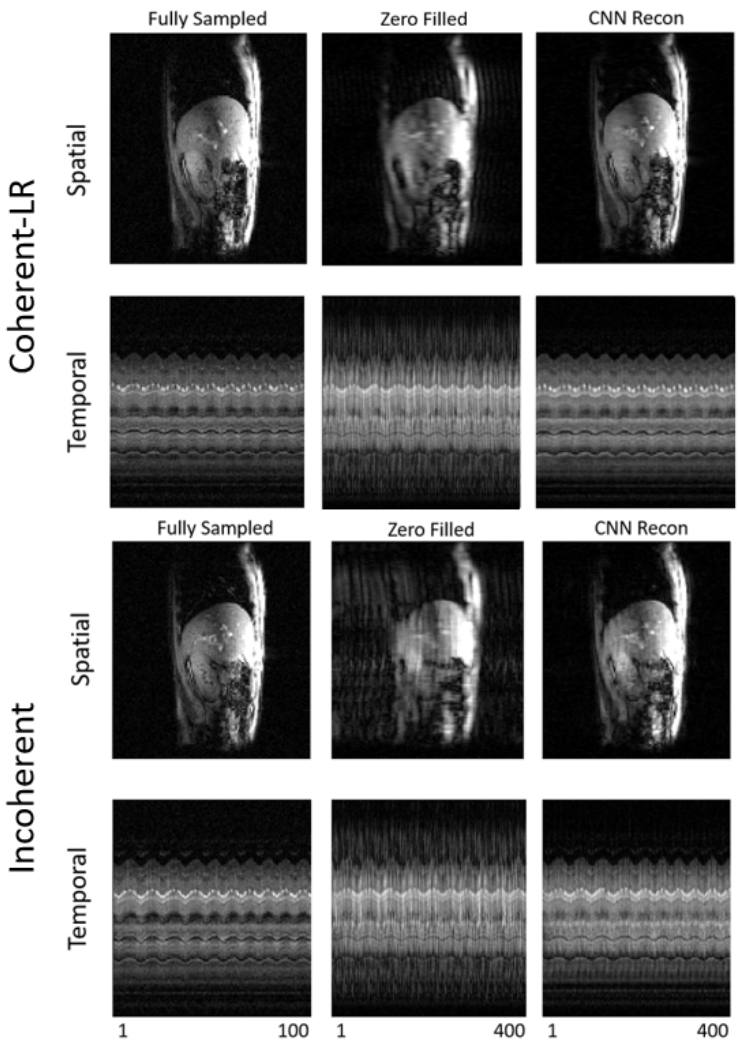

Figure 5.9: Prospective reconstruction for the coherent-LR and incoherent undersampled CNN reconstructions of a volunteer data set. The first column contains fully sampled data, the second column is the zero-filled prospectively acquired data, and the third column contains the CNN reconstructed data. For each sampling scheme, the top row contains a spatial view, while the bottom row contains a temporal view; note that there are 4 times more temporal frames for the prospectively acquired data due to the reduced acquisition time. It is evident that the artefacts have been removed from both the spatial and temporal views; however, the coherent-LR reconstruction produced a qualitatively better reconstruction.

5.4 Discussion

By reducing the imaging time through undersampling, a higher frame rate and reduced latency can be achieved, allowing for real-time adaptive radiotherapy treatment. The use of patient-specific trained CNNs can be interpreted as personalized healthcare for the patient. This differs from the common NN paradigm, where a generalized solution is sought to be applied to any new patient data.

For the purpose of the study, clinical interpretation of the real-time tracking is not the primary objective. Any required clinical interpretation would be achieved with other sequences in the planning or diagnostic stage of the patient's treatment. Further, delineation of the contours for training purposes would be performed on the fully-sampled images, not the CNN reconstructed data in our proposed workflow. As such, a successful match of the CNN reconstructed contour to the fully sampled case (as demonstrated via the DC and CD) is the primary concern.

As mentioned previously, the CNN training time is on the order of several hours; however, this does not pose any issue for clinical implementation, given that the pre-treatment training data would be acquired prior to the patients scheduled treatment. Patients undergoing external beam radiotherapy will generally receive a pre-treatment CT for the dosimetrist to generate patient contours and plan how the radiation will be delivered by the linac (or LMR). Institutes treating with LMR systems will also acquire pre-treatment MRI data. This previously acquired MRI data could be utilized to train a patient-specific CNN prior to the scheduled treatment, which can then be used to aid in reconstructing rapidly acquired (prospectively) undersampled data during treatment. Because the MRI data would be acquired prior to the scheduled treatment (often days prior), the training time is generally not much of a concern; however, the reconstruction time must be fast enough to facilitate real-time on-the-fly

imaging.

The benefits to using a coherent-LR undersampled acquisition is that it is simple to implement on an MRI scanner and has reduced artefacts as compared to the incoherent sampling scheme. We have shown that the realistic prospective imaging scenario qualitatively appears superior when using the coherent-LR undersampled acquisition (compared to the incoherent undersampling). The coherent-LR undersampling does result in a slight blurring of the reconstructed image; however, we believe that the segmentation program handles this case better than the incoherent noise-like artefact. Furthermore, not all MR systems have the software capability to implement an incoherent acquisition.

The use of low-resolution undersampling may evoke memories of keyhole imaging [133], since only the central region of k-space is acquired during the real-time imaging portion of the exam. Indeed, both the incoherent and coherent-LR undersampling CNN strategies described in this work can be considered sophisticated versions of view-share, since patient-specific fully-sampled data is used to train a CNN to reconstruct a sub-sampling of that data during the real-time acquisition. However, unlike the keyhole strategy, where blocks of k-space from fully-sampled acquisitions are simply imported into newly-acquired frames; the trained CNN computes the unacquired k-space for the data acquired in the undersampled frame. This computed k-space will be different for every frame, based only on patterns of correspondence seen in the training data.

Generally, the incoherent sampling scheme requires some correction for the increased artefacts caused by eddy currents, either pre- or post-reconstruction. For our incoherent acquisition we used a "through-slice technique", where the slice preparation amplitude is increased. This creates a slight unbalancing of the slice-selection

gradient in order to suppress the eddy-current contributions [127]. When using the coherent-LR undersampling scheme, this is not a concern, given the coherent-LR sampling pattern does not require rapidly changing gradients with large PE steps (high frequency components).

The tumour specifics are stated in Table 5.1; however, Figure 5.10 displays the centroid motion for the six patients along the 650 fully sampled dynamic images. It can be seen that patients 2, 4, 5, and 6 contain rhythmic motion for the most part; while patients 1 and 3 had tumour motion that changed during the imaging. As shown from Figures 5.4, 5.5, 5.7, these patients did not correlate to a reduced DC or increased centroid displacement.

There are several limitations to this study. Our patient cohort was small with only six patients; however, the tumours varied in location, size, and motion. Further, our study only evaluated data acquired at one field strength (3T); however, the scenario of having training data acquired by a different system than the treatment Linac-MR should be investigated as the contrast may differ between field strengths. While future studies will be required to assess varying contrast etc.; as a proof of concept this study provides encouraging results, and a trend toward better segmentation when using coherent-LR undersampling with the CNN. Regarding the SNR, we have previously demonstrated that the presence of increased noise does not decrease the accuracy of the CNN reconstruction [116]. We only investigated one specific region for tumour tracking of lung tumours, however, as outlined in Table 5.1, the tumours varied in motion and size. In the future, other regions should be investigated, such as the liver or pancreas. Furthermore, the smallest tumour present in our data is 1.36 cm^2. Future studies should investigate the extent of how small the tumorous region can be, while still providing an accurate CNN reconstruction with either phase encoding

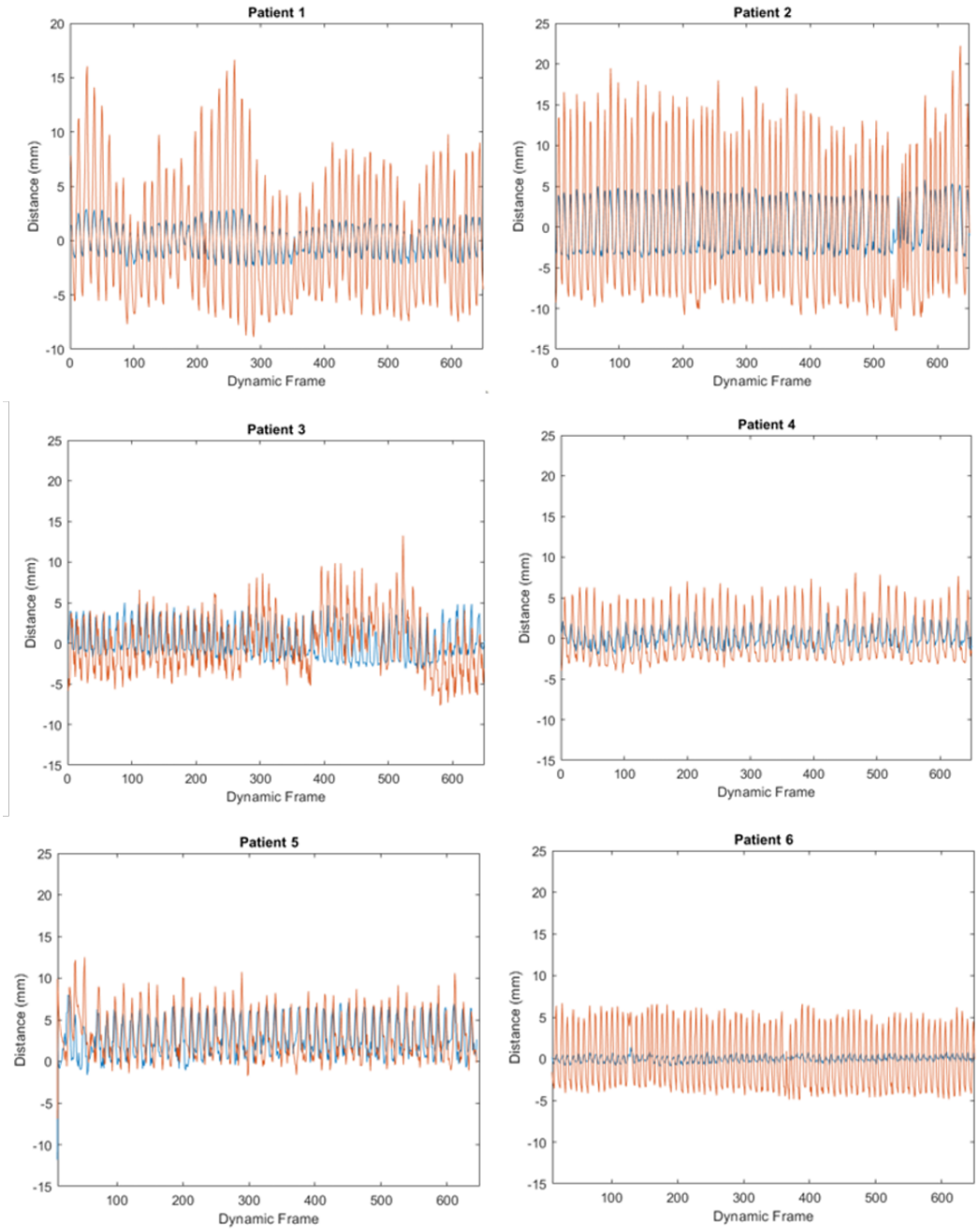

Figure 5.10: Centroid plots for the six patients lung tumours. The orange displays the superior-inferior motion, while the blue displays the anterior-posterior motion.

scheme.

5.5 Conclusion

Incoherent undersampling results in a noise-like aliasing artefact, while coherent-LR undersampling results in a low spatial resolution blurring. We investigated the suitability of using these phase-encoding schemes for MRI reconstruction using CNNs. In particular, we investigated the use for lung tumour segmentation for adaptive radiotherapy purposes using an MR-Linac system. We believe our results support the hypothesis that the CNN with coherent undersampling performs at least as well as, and often better than the CNN with the more typical incoherent undersampling, at least with regards to this particular real-time tracking scenario. We retrospectively evaluated the technique by comparing tumour segmentations of six non-small cell lung cancer patients. The coherent-LR undersampled acquisition demonstrated larger DC values, but more importantly, there was less variation in the DC across the temporal domain (200 dynamic frames). We applied the coherent-LR undersampled acquisition in a prospectively undersampled volunteer data set. The qualitative results from the prospective case demonstrate that the artefacts were removed, with only a slight blurring occurring throughout the reconstructed images. The reconstruction speed of the 3D CNN technique was 54 ms per image.

Chapter 6

Investigation into robustness of CNN reconstruction with respect to patient shifts and rotations

6.1 Introduction

Cancer patients who undergo external beam radiotherapy (EBRT) often require several treatments over an extended period of time. In order to deliver the treatment accurately, the interfractional motion must be accounted for. This is generally done with the use of CBCT, which is used to confirm patient positioning through registration with their planning CT. However, the intrafractional motion should also be taken into account. This becomes especially important for tumours located in regions of large motion (such as the lung or liver). This can be done with equipment such as breathing bellows or stereoscopic cameras to track the patients' surface. Ideally, imaging would be used to accurately track the tumour during treatment. With

the increasing popularity of linac-MRI (LMR) systems [27, 118, 24], real-time MR imaging is of great importance for tumour tracking [36, 103] and advanced gating techniques. A recent example of an advanced gating technique uses the real-time image reconstruction as a guide to help aid the patient control their breathing; the patient uses specially constructed glasses to view their images to help focus the tumour into the treatment region [35]. The use of real-time imaging for intrafraction monitoring can be used to reduce dose to healthy surrounding tissue. Furthermore, tracking intrafraction motion may allow for more accurate dose reconstruction either during or post fractional treatment.

The application of a convolutional kernel to an image is a shift-invariant process. This is beneficial for intrafraction and interfraction imaging for tumour tracking; if the patient is translated or shifts during scanning, the convolution operation (using kernels trained on previous data) will be (mostly) unaffected. This is an important feature for robust image reconstruction, especially for patients who are imaged during multiple sessions. The convolution process, however, will not be invariant to rotations as they will inevitably change the relative positions of the pixels, resulting in a non-trivial change in the output of the convolution operation. Training a CNN using data acquired at one imaging session may not have the exact same patient placement when undergoing another imaging session, which will be a common scenario for LMR treatment. This study aims to explore the effect of patient shifts using data acquired from six non-small cell lung cancer patients.

It has been previously demonstrated in Chapters 4 and 5, that the use of patient-specific CNNs can be used for real-time reconstruction of dynamic imaging data [116, 131]. However, the patient may be somewhat shifted or rotated during successive imaging sessions. This work aims to investigate how shifts and rotations may effect

the reconstruction.

6.2 Methods

6.2.1 MRI Acquisition

Data was acquired from six non-small cell lung cancer patients, with approval from our institutional board of ethics. The data was acquired with our 3T Philips Achieva Scanner (Philips Medical Systems, The Netherlands) using a 2D bSSFP sequence. Each patient was imaged for three minutes, resulting in 650 dynamic imaging frames. The parameters included: FOV = 40 × 40 cm^2, voxel size = 3.1 × 3.1 × 20 mm^3, TR/TE = 2.2/1.1 ms, and acquisition time per frame = 275 ms. The six channel data was channel combined and exported as complex data via at the console using coil sensitivity maps and the coil covariance measurements (to preserve the phase data). The retrospective data was incoherently undersampled by a factor of 10×.

6.2.2 Convolutional Neural Network

A single convolutional neural network (CNN) was trained separately for each patient, following the methodology of previous studies presented in Chapters 4 and 5 [116, 131]. The CNN model is based off the CNN cardiac study presented by Schlemper *et al.* [97]. The CNN model uses a cascading network of alternating convolution and data consistency layers.

In short, each patients' complex MR data in image space is passed to the CNN. The data is split into two channels containing the real and imaginary components of the data. The temporal data undergoes a data sharing step, where multiple temporal frames are combined and added to the training data along the channel dimension.

Following data sharing, the alternating convolution and data consistency layers are implemented. Following our previous study, we used 4 convolution layers, with 16 kernels in each layer, along with 4 data consistency layers. The layers are cascading such that after every 4 convolution layers, the data is concatenated along the channel dimension and FT back to k-space for the data consistency layer to enforce data fidelity. After the data consistency is computed, the data is IFT back to image space and undergoes the convolution layer. This repeats for the number of data consistency layers (4 in this case). A schematic of the CNN model is shown in Figure 4.1. The chosen activation function was the rectified linear unit (ReLU). The optimization algorithm was Adam, with the default parameters ($\beta_1 = 0.9$ and $\beta_2 = 0.999$).

Each dynamic data set (consisting of 650 free-breathing dynamic images) was split such that the first 450 temporal frames were used for training. The remaining 200 frames were used to provide an unbiased quantitative evaluation of the CNN reconstruction. The 450 temporal frames were trained using a k-fold cross-validation technique with a k-fold factor of 5 (as shown in Figure 5.2). Thus, the test data consisted of a different set of 90 frames for each cross-validation.

6.2.3 Shifts and Rotations

To investigate the robustness, we first trained the CNN on the acquired patient data as discussed above. Then, using the parameters computed from the (non-shifted, non-rotated) training data, we applied the CNN reconstruction technique to shifted and rotated data (in-plane). All shifts were implemented in MATLAB. The translational shifts were implemented in k-space via complex modulation, while the rotations were done in image space using bi-cubic interpolation in MATLAB.

Given the confining space of an MR system, the translation and rotation are inher-

ently limited; however, in order to investigate a wide range of effects, larger motion values were used. For translation, the data from the six patients were shifted from 10 mm to 30 mm in increments of 10 mm. For rotations, we investigated 5° up to 20° in increments of 5°. The patient data was shifted prior to CNN reconstruction, either via translation or rotation. After CNN reconstruction, the patient data was shifted back for comparison to the fully sampled data via the Dice coefficient (DC) and structural similarity metric (SSIM) metrics.

In order to quantify the reconstruction, we utilized the tumour segmentations using a segmentation software previously developed by our group [114, 36, 103, 128]. We quantified the reconstruction by comparing the CNN reconstructed segmented tumour to the tumour segmentations from the fully-sampled gold standard data.

6.2.4 Quantitative Metrics

Dice Coefficient

The Dice coefficient is a metric that defines the overlap between two binary structures. For this study, we used it to evaluate how well the CNN reconstructed images can be used to segment the tumours from the six non-small cell lung cancer patients. By comparing the binary segmentation from the fully sampled contours to the CNN reconstructed contours, we can determine how faithful the CNN reconstruction is in the tumourous region.

$$DC = 2 \cdot \frac{ROI_{FS} \cap ROI_{US}}{ROI_{FS} + ROI_{US}} \tag{6.1}$$

where ROI_{FS} is the binary segmentation (ROI) of the fully sampled data and ROI_{US} is the binary segmentation of the undersampled CNN reconstructed data.

Structural Similarity Index

The structural similarity index (SSIM) is a metric that evaluates the perceived quality difference by comparing two data sets [112]. The SSIM computes three distinct terms: luminance, contrast, and structure. By combining these terms, we get an expression for SSIM,

$$SSIM = \frac{(2\mu_F\mu_{US} + C_1)\ (2\sigma_{F,US} + C_2)}{(\mu_F^2 + \mu_{US}^2 + C_1)\ (\sigma_F^2 + \sigma_{US}^2 + C_2)}. \tag{6.2}$$

The terms μ_{FS}, μ_{US}, σ_{FS}, and σ_{US} are the means and standard deviation of the fully sampled and undersampled data respectively. The term $\sigma_{(FS,US)}$ is the cross-correlation of the fully and undersampled images. The constants C_1 and C_2 are based on the dynamic range (DR) of the two image sets; these values are typically set to $C_1 = (0.01 \cdot DR)^2$ and $C_2 = (0.03 \cdot DR)^2$.

6.3 Results

The results contain the CNN reconstruction for the six non-small cell lung cancer patients.

6.3.1 Dice Coefficient

As expected, the translational shifts did not reduce or degrade the image quality. Figure 6.1 displays violin plots of the DC from the six patients. The violin plot display the median (bold dashed line in the center of the plots), the 25^{th} and 75^{th} quartiles (dotted lines), as well as the entire distribution of the data; the width of the plot at a given y-axis displays the relative number of values at that point. It is evident that the shifting did not reduce the DC. This demonstrates that CNNs are indeed robust to translations.

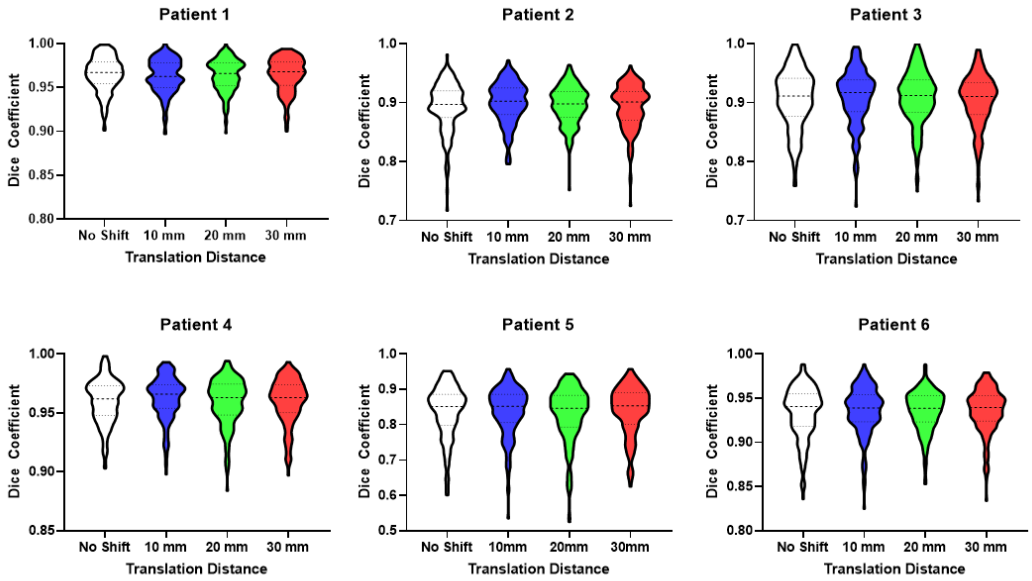

Figure 6.1: Violin plots of the six patients with a range of translations. The plots contain the DC from the 200 CNN reconstructed dynamic frames. A t-test analysis showed no significance between any rotation to the no shift case. Thus, translational shifts do not have a significant effect on the reconstruction.

The effects of rotation on the DC are shown in Figure 6.2. Here it can be seen that the DC decreases with increasing rotation for all patients. The decrease in DC is more drastic for patients 2, 4, and 5 where the variance was also increased. From the data presented in Figure 6.2 it appears that some patients are more sensitive to rotational shifts, while other patients are not. This is likely dependent on several factors such as tumour morphology, signal relative to surrounding structures, and location of the tumour.

6.3.2 Structural Similarity Index (SSIM)

The mean and standard deviation of the SSIM is shown in Figure 6.3 for the translations and rotations. The SSIM remained constant for the translational shifts, indicating that there was no decrease in image quality.

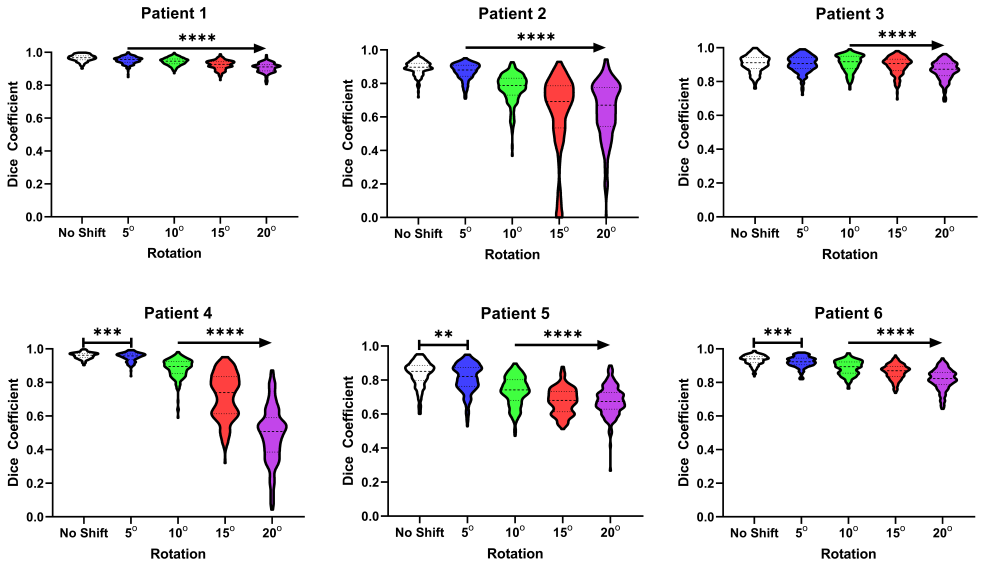

Figure 6.2: Violin plots of the six patients with various rotational shifts. The plots contain the DC from the 200 CNN reconstructed dynamic frames. The values along the arrows denote the t-test significance (compared to the no shift case); while nearly all rotations were statistically different, it is evident that the rotations decrease the DC more significantly for patients 2, 4, and 5.

The rotational shifts did, however, decrease the overall image quality of the images. Furthermore, the decrease in SSIM trended similarly for all six patients.

6.4 Discussion

It is evident from Figure 6.1 that translational shifts have little to no impact on the reconstruction quality of the CNN data. This is due to the convolutional invariance to shifts. However, the rotational shifts did decrease the DC for all patients as shown in Figure 6.2. The DC decreased only slightly when rotated by 5°; however, at 10° or larger there were large decreases in DC, specifically for patients 2, 4, and 5. Patients 1, 3 and 6 were less effected by the rotations. Figure 6.4 contains the worst rotated CNN reconstructions for patients 1 and 4. As shown in Figure 6.4 the image quality

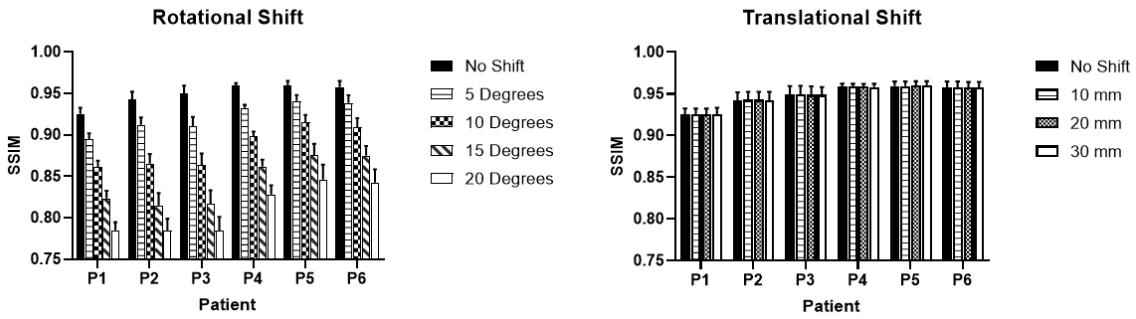

Figure 6.3: The mean and standard deviation of the SSIM metric for the six patients with various translations and rotations.

for patient 1 decreases with increasing angle; however, the tumour remains visible at all angles. The tumour in patient 4 begins to disappear at 15° and is almost completely removed at 20°.

While it is evident the rotational shifts decrease the image quality, and ability to segment tumours; care must be taken to ensure the patient set-up is accurate ($\pm$ 5° to ensure accurate CNN reconstruction. Rotations can be included in the training step via data augmentation, which will increase the training time, but may result in a more robust reconstruction model. This generalization may reduce the average in DC and SSIM; however, the model would be robust to rotational shifts and may not suffer from large artifacts. The balance of robustness versus accuracy will need to be explored in future studies.

This study only investigated the linear shifts and rotations; however, warps (deformations) and other non-linear shifts would provide a more realistic scenario and will need to be investigated in the future. Furthermore, contrast changes (such as imaging at multiple field strengths), different pulse sequences, etc. also need to be investigated.

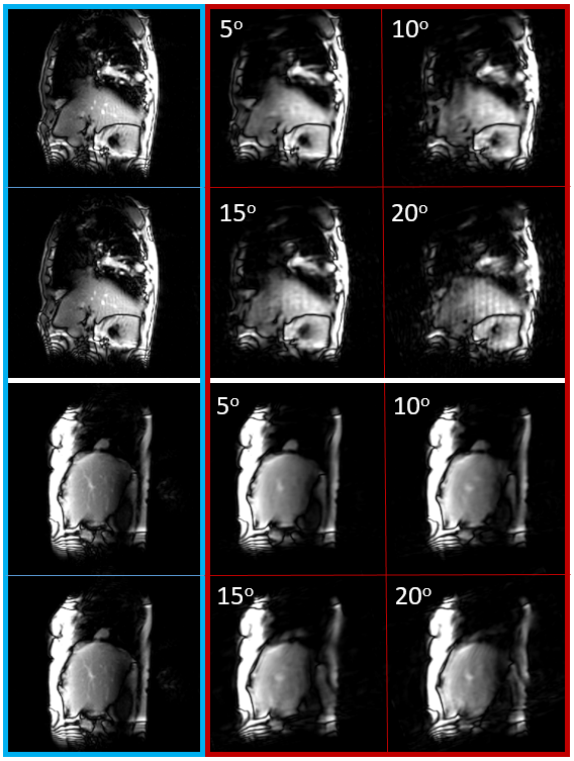

Figure 6.4: Images for the worst DC for patients 1 and 4, as an example. Patient 1 remained quite robust to the rotational changes, while patient 4 had the tumourous region nearly completely disappear. The left hand bounding box contains the fully sampled images.

6.5 Conclusion

We quantified the effects of translational shifts and rotations on the CNN reconstruction of six non-small cell lung cancer patients. As slight shifts and rotations may occur when imaging a patient undergoing successive EBRT treatments, we investigated how these shifts may reduce the segmentability of the CNN reconstructed data. We verified that translational shifts do not degrade the quality of the images via the DC and SSIM metrics. Rotational shifts can drastically degrade the quality of the images and segmentability it the patient is rotated 10° or more; however, this was found to be dependent on the tumour location and signal. If rotational shifts of 10° or greater are unavoidable, augmented rotational shifts should then be incorporated into

the CNN training to account for and reduce the artefacts associated with rotation.

Chapter 7

Conclusion and Future Work

The use of real-time MRI for adaptive radiotherapy is becoming of increasing interest as more LMR systems are being installed world wide. The utility of real-time MRI allows for the tracking of intra-fractional tumour motion during treatment. Achieving a high frame-rate with low latency is of great importance in order to conform the radiation field to the segmented structure via the MLC. The goal of which is to increase the accuracy of the radiation and reduce the ITV margins, reducing damage to nearby healthy tissue.

The work presented within this thesis aims to solve the issue surrounding real-time on-the-fly MRI, specifically for LMR adaptive radiotherapy. Real-time MR imaging is a rather challenging problem to solve, due to the trade-off between image quality and data acquisition speed. Increasing the data acquisition speed via k-space undersampling, will necessarily decrease the image quality with an increased presence of aliasing artifacts. Within this body of work, two distinct techniques have been investigated to solve this problem: CS-PCA and 3D CNNs.

Chapter 1 presented the pertinent background and motivation such as external

beam radiotherapy and current linac technologies. Image guidance and the LMR development at the Cross Cancer Institute was also discussed. Chapter 2 provided the theory relevant to the research within this work including magnetic resonance imaging (MRI), principal component analysis (PCA), and convolutional neural networks (CNNs).

Chapter 3 presented the compressed sensing principal component analysis (CS-PCA) technique developed for real-time MRI. This technique uses a small fully-sampled dyanamic dataset acquired prior to prospective undersampling, to calculate the principal components, which are then used to aid in the image reconstruction. Using free-breathing MRI data acquired from six non-small cell lung cancer patients, this study compared the image quality and tumour segmentation of the CS-PCA reconstructed data (at several acceleration factors) to the fully sampled (ground truth) data. This proved to be a fast (under 10 ms per image) technique that achieved Dice coefficients of 0.9 or greater for acceleration factors up to 10x. It was found that the noise increases with increasing time (due to the database becoming outdated)

Chapter 4 presented a 3D CNN real-time MRI reconstruction technique. We demonstrated that the CNN could reconstruct MR images at 65 ms using a single standard GPU (NVIDIA GTX 1080Ti). The technique was validated using free-breathing data acquired from six retrospective non-small cell lung cancer patients, using an incoherent phase-encoding scheme. By training an individual patient-specific CNN, this novel technique can be seen as a patient centered care reconstruction technique. We furthermore validated the technique using prospectively undersampled data from a volunteer.

Chapter 5 presented an investigation into the use of coherent low resolution (coherent-LR) phase-encoding (PE) for real-time CNN reconstruction. The coherent-LR PE

scheme was quantified using the six non-small cell lung cancer patients, retrospectively. Furthermore, the coherent-LR was qualitatively investigated using a prospectively undersampled volunteer. The coherent-LR phase-encoding proved to be superior over the incoherent phase-encoding scheme, for both the retrospective and prospective imaging scenario. Coherent-LR phase-encoding is ideal, as it is simple to implement on most clinical MRI systems and does not suffer from eddy currents present with incoherent sampling.

Chapter 6 investigated the robustness of CNN reconstruction to translation and rotation. It is was found that the CNN reconstruction does exhibit translational invariance. Rotation, however, did degrade the image quality and segmentability of the tumour. It was found that the degradation following a rotation was dependant on the signal and location of the tumour.

We have demonstrated that these techniques can be used to reconstruct MR images with a latency that may allow for adaptive radiotherapy. With the increasing popularity of Linac-MR systems, having the ability to image in real-time will be a desired feature. In particular, we investigated lung tumours, as they present a large period motion that results in a more difficult treatment with large treatment volumes (or time consuming treatment via respiratory gating). Other regions need to be investigated in the future, for example the liver, breast, and pancreas. While we presented a study utilizing CNNs for real-time MRI, we believe that more work in finding the optimal network and parameters will result in a reduction in reconstruction time and increase in image quality.

The techniques developed within this work rely on previously acquired fully sampled data. However, accelerated data could also be acquired and used to increase the size of the training data. For incoherently undersampled data, reconstruction

techniques that utilize the kt-domain could be used to reconstruct the accelerated data [37, 38, 39]. Thus, accelerated data could be acquired (along with the fully sampled data) and reconstructed using a time consuming kt-domain technique to be incorporated with the fully sampled data to train the CNN. This additional training data may aid in the CNN reconstruction, as it more closely resembles the truly accelerated data.

Future work for the CS-PCA technique demonstrated in Chapter 3 would include developing a way to update the previously acquired (fully sampled) database, used to generate the PCs. Updating the database during the imaging session may reduce the increasing noise that was found to occur over time. Utilizing an alternating PE scheme, such that the entire k-space is acquired every four frames. This way pseudo fully sampled data can continually be used to update the database. This is currently being investigated by a colleague.

An investigation into CNN localized reconstruction may improve upon our CNN technique for real-time prospective imaging. By confining the CNN training only on the region of interest (region directly surrounding the tumour motion), as opposed to the entire image domain, it may be possible to reconstruct the smaller region of interest with high fidelity, while sacrificing image quality elsewhere. This would pertain strictly to techniques that aim to segment a particular region, such as a tumour that is known in advance.

The CNN technique investigated within this work attempts to reconstruct any image, regardless of the respiratory cycle in real-time. Another future research project could investigate the use of respiratory monitoring along with a patient-specific CNN reconstruction that is focused on specific phases of the breathing cycle. This would require several CNNs trained on one patient, each focused on a portion of the breath-

ing cycle. For real-time imaging, the breathing monitoring would be used to trigger which breathing phase CNN would be used for the image reconstruction. While the training time would increase, it would be interesting to investigate if this technique may be more accurate.

Another ANN technique that may prove to be successful for real-time tumour tracking is known as inception. Developed by Google, the inception model aims to reduce the depth of a CNN model, by concatenating convolutions with differing kernel sizes [134, 135]. This makes the model wider (i.e. more neurons); however, having a wider model is computationally cheaper than a deep network. The inception model also factorizes $n \times n$ convolutions into $1 \times n$ and $n \times 1$ convolutions, to save time. Similarly, the 5×5 convolutions are factored into two 3×3 convolutions, which was found to be 2.5 times faster. The fact that this model can capture multiple kernels of varying sizes may be an interesting avenue to pursue. The inception technique was demonstrated to be successful at segmenting and classifying lung nodules [136].

Lastly, another interesting avenue of research is the use of neuromorphic computing and spiking neural nets (SNNs) [137, 138]. Spiking neural networks are similar to modern ANNs, except that they use binary impulses as opposed to real-valued numbers; which is far more efficient that standard ANNs. SNNs also exploit time as an additional dimension, making it a spatio-temporal NN structure that is similar to the brain's response. Given the impulse design of SNNs, backpropagation remains an issue (impeding training algorithms), as we can no longer utilize gradient descent (the heart of the issue lies with the calculation of derivatives). One problem with modern NNs (which has led to advances in SNNs) is that they require a huge amount of energy consumption via GPUs. They are also limited by the "memory wall bottleneck", meaning the compute capability has risen to a greater rate than the memory storage.

Therefore, there is a lag as we need to wait for memory transfer which hinders the advancements that have been made in CPU and GPU technology. One way around the immense power requirement and "memory wall" is with the use of neuromorphic chips [139].

Bibliography

[1] C. C. S. A. Committee, Canadian Cancer Statistics, Canadian Cancer Society (2018).

[2] C. C. S. A. Committee, Canadian Cancer Statistics, Canadian Cancer Society (2017).

[3] F. M. Khan and J. P. Gibbons, *The Physics of Radiation Therapy*, Lippincott Willians & Wilkins, Philidelphia, PA, 2014.

[4] S. Christensen, L. Wilson, and D. Leffell, *DeVita, Hellman, and Rosenberg's Cancer: Principles & Practice of Oncology*, Lippincott Williams & Wilkins, Philadelphia, PA, 2019.

[5] Y. Xiao, S. Kry, R. Popple, E. Yorke, N. Papanikolaou, S. Stathakis, P. Xia, S. Hug, J. Bayouth, J. Galvin, and F. Yin, Flattening filter–free accelerators: a report from the AAPM therapy emerging technology assessment work group, Journal of Applied Clinical Medical Physics **16**, 12–29 (2015).

[6] C. F. Njeh, K. C. Snyder, and J. Cai, The use of six degrees of freedom couch is only clinically beneficial in stereotactic radio surgery, Medical Physics **46**, 415–418 (2019).

[7] W. C. Röntgen, ON a new kind of rays, Science **3**, 227–231 (1896).

[8] E. Grubbe, Priority in the use of x-rays, Radiology **21**, 156–162 (1933).

[9] J. Thariat, J. Hannoun-Levi, A. Myint, T. Vuong, and J. Gerard, Past, present, and future of radiotherapy for the benefit of patients, Nat Rev Clin Oncol **10**, 52–60 (2012).

[10] B. Cho, Intensity-modulated radiation therapy: a review with a physics perspective, Radiation Oncology Journal **36**, 1–10 (2018).

[11] K. Otto, Volumetric modulated arc therapy, Med Phys **35**, 310–317 (2008).

[12] D. A. Jaffray and M. K. Gospodarowicz, *Cancer: Disease Control Priorities*, The International Bank for Reconstruction and Development, Washington, DC, 2015.

[13] P. Keall, G. Mageras, J. Balter, R. Emery, K. Forster, S. Jiang, J. Kapatoes, D. Low, M. Murphy, B. Murray, C. Ramsey, M. V. Herk, S. Vedam, J. Wong, and E. Yorke, The Management of respiratory motion in radiation oncology: Report of AAPM task group 76, Med Phys **12**, 3874–3900 (2006).

[14] G. Ibbott, The need for, and implementation of, image guidance in radiation therapy, Annals of the ICRP **47**, 160—-176 (2018).

[15] M. Bakhtiari, L. Kumaraswamy, D. Bailey, S. de Boer, H. Malhotra, and M. Podgorsak, Using an EPID for patient-specific VMAT quality assurance, Med Phys **38**, 1366–1373 (2011).

[16] U. Oelfke, T. Tücking, S. Nill, A. Seeber, B. Hesse, P. Huber, and C. Thilmann, Linac-integrated kV-cone beam CT: Technical features and first applications, Medical Dosimetry **31**, 62–70 (2006).

[17] J. Bertholet, A. Knopf, B. Eiben, J. McClelland, A. Grimwood, E. Harris, M. Menten, P. Poulsen, D. T. Nguyen, P. Keall, and U. Oelfke, Real-time intrafraction motion monitoring in external beam radiotherapy, Phys Med Biol **64**, 15TR01 (2019).

[18] C. Bert, K. G. Metheany, K. Doppke, and G. T. Chen, A phantom evaluation of a stereo-vision surface imaging system for radiotherapy patient setup, Med Phys **32**, 2753–2762 (2005).

[19] M. Laaksomaa, S. Sarudis, M. Rossi, T. Lehtonen, J. Pehkonen, J. Remes, H. Luukkanen, T. Skytta, and M. Kapanen, AlignRT and Catalyst in whole-breast radiotherapy with DIBH: Is IGRT still needed?, J Appl Clin Med Phys **20**, 97–104 (2019).

[20] T. R. Willoughby, P. A. Kupelian, J. Pouliot, K. Shinohara, M. Aubin, M. Roach, L. L. Skrumeda, J. M. Balter, D. W. Litzenberg, S. W. Hadley, J. T. Wei, and H. M. Sandler, Target localization and real-time tracking using the Calypso 4D localization system in patients with localized prostate cancer, Int. J. Radiat. Oncol. Biol. Phys. **65**, 528–534 (2006).

[21] M. Belanger, Z. Saleh, T. Volpe, R. Margiasso, X. Li, M. Chan, X. Zhu, and X. Tang, Validation of the Calypso Surface Beacon Transponder, J Appl Clin Med Phys **17**, 223–234 (2016).

[22] T. Gevaert, D. Verellen, K. Tournel, N. Linthout, S. Bral, B. Engels, C. Collen, T. Depuydt, M. Duchateau, T. Reynders, G. Storme, and M. De Ridder, Setup accuracy of the Novalis ExacTrac 6DOF system for frameless radiosurgery, Int. J. Radiat. Oncol. Biol. Phys. **82**, 1627–1635 (2012).

[23] B. Fallone, The rotating biplanar linac-magnetic resonance imaging system,

Semin Radiat Oncol **24**, 200–202 (2014).

[24] J. Lagendijk, B. Raaymakers, A. Raaijmakers, J. Overweg, K. Brown, E. Kerkhof, R. van der Put, B. Hardemark, M. van Vulpen, and U. van der Heide, MRI/linac integration, Radiother Oncol **86**, 25–29 (2008).

[25] P. Keall, M. Barton, and S. Crozier, The Australian magnetic resonance imaging-linac program, Radiother Oncol **24**, 203–206 (2014).

[26] F. Dempsey, D. Benoit, J. Fitzsimmons, A. Haghighat, J. Li, D. Low, S. Mutic, J. Palta, H. Romeijn, and G. Sjoden, A device for realtime 3D image–guided IMRT, Int J Radiat Oncol **63**, S202 (2005).

[27] B. Fallone, B. Murray, S. Rathee, T. Stanescu, S. Steciw, S. Vidakovic, E. Blosser, and D. Tymofichuk, First MR images obtained during mega-voltage photon irradiation from a integrated Linac-MR system, Med Phys **36**, 2084–2088 (2009).

[28] S. Yaghoobpour Tari, K. Wachowicz, and B. Gino Fallone, A non-axial superconducting magnet design for optimized patient access and minimal SAD for use in a Linac-MR hybrid: proof of concept, Phys Med Biol **62**, N147–N160 (2017).

[29] C. Kirkby, B. Murray, S. Rathee, and B.G.Fallone, Lung dosimetry in a linac-MRI radiotherapy init with a longitudinal magnetic field, Med Phys **37**, 4722–4732 (2010).

[30] G. P. Liney, B. Whelan, B. Oborn, M. Barton, and P. Keall, MRI-linear accelerator radiotherapy systems, Clin Oncol **30**, 686–691 (2018).

[31] J. P. Marques, F. F. J. Simonis, and A. G. Webb, Low-field MRI: An MR physics perspective, J Magn Reson Imaging **49**, 1528–1542 (2019).

[32] A. E. Campbell-Washburn et al., Opportunities in interventional and diagnostic imaging by using high-performance low-field-strength MRI, Radiology **29**, 384–393 (2019).

[33] K. Wachowicz, N. D. Zanche, E. Yip, V. Volotovskyy, and B. G. Fallone, CNR considerations for rapid real-time MRI tumor tracking in radiotherapy hybrid devices: Effects of Bo field strength, Med Phys **43**, 4903–4919 (2016).

[34] A. Keyvanloo, B. Burke, B. Warkentin, T. Tadic, S. Rathee, C. Kirkby, D. Santos, and B. Fallone, Skin dose in longitudinal and transverse linac-MRI's using Monte Carlo and realistic 3D MRI field models, Med Phys **10**, 6509–6521 (2012).

[35] B. Sahin, T. Mustafayev, G. Gungor, G. Aydin, B. Yapici, B. Atalar, and E. Ozyar, First 500 Fractions Delivered with a Magnetic Resonance-guided Radiotherapy System: Initial Experience, Cureus **11** (2019).

[36] J. Yun, K. Wachowicz, M. Mackenzie, S. Rathee, D. Robinson, and B. Fallone, First demonstration of intrafractional tumor–tracked irradiation using 2D phantom MR images on a prototype Linac–MR, Med Phys **40**, 1–12 (2013).

[37] S. Lingala, Y. Hu, E. DiBella, and M. Jacob, Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR, IEEE Trans Med Imaging **30**, 1042–54 (2011).

[38] L. Feng, L. Axel, H. Chandarana, K. Block, D. Sodickson, and R. Otazo, XD-GRASP: Golden-angle radial MRI with reconstruction of extra motion-state dimensions using compressed sensing, Magn Reson Med **75**, 775–88 (2016).

[39] R. Otazo, D. Kim, L. Axel, and D. Sodickson, Combination of compressed sensing and parallel imaging for highly accelerated first–pass cardiac perfusion MRI, Magn Reson Med **64**, 767–76 (2010).

[40] K. Pruessmann, M. Weiger, M. Scheidegger, and P. Boesiger, SENSE: Sensitivity encoding for fast MRI, Magn Reson Med **42**, 952–962 (1999).

[41] M. Griswold, P. Jakob, R. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, Generalized autocalibrating partially parallel acquisitions (GRAPPA), MRM **47**, 1202–1210 (2002).

[42] X. Li, Y. Lee, S. Mikaiel, J. Simonelli, T. Tsao, and H. H. Wu, Respiratory motion prediction using fusion-based multi-rate Kalman filtering and real-Time golden-angle radial MRI, IEEE Transactions on Biomedical Engineering , 1–11 (2019).

[43] M. Uecker, S. Zhang, D. Voit, A. Karaus, K. Merboldt, and J. Frahm, Real–time MRI at a resolution of 20 ms, NMR in Biomedicine **23**, 986–994 (2010).

[44] A. Majumdar, R. Ward, and T. Aboulnasr, Compressed sensing based real-time dynamic MRI reconstruction, IEEE Trans Med Imaging **31**, 2253–66 (2012).

[45] Y. Seppenwoolde, H. Shirato, K. Kitamura, S. Shimizu, M. van Herk, J. V. Lebesque, and K. Miyasaka, Precise and real-time measurement of 3D tumor motion in lung due to breathing and heartbeat, measured during radiotherapy, Int. J. Radiat. Oncol. Biol. Phys. **53**, 822–834 (2002).

[46] I. Rabi, J. Zacharias, S. Millman, and P. Kusch, A new method for measuring nuclear magnetic moment, Phys Rev **53**, 318 (1938).

[47] P. Lauterbur, Image formation by induced local interactions: examples of employing nuclear magnetic resonance, Nature **242**, 190–191 (1973).

[48] P. Mansfield and P. Grannell, NMR "diffraction" in solids, J Phys C: Solid state Phys **6**, L422–L426 (1973).

[49] J. Luo, D. Hippe, H. Rahbar, S. Parsian, M. Rendi, and S. Partridge, Diffusion tensor imaging for characterizing tumor microstructure and improving diagnostic performance on breast MRI: a prospective observational study, Breast Cancer Res **102**, 1–16 (2019).

[50] M. Guimaraes, A. Schuch, B. Hochhegger, J. Gross, R. Chojniak, and E. Marchiori, Functional magnetic resonance imaging in oncology: state of the art, Radiol Bras **47**, 1–10 (2014).

[51] R. W. Brown, Y.-C. N. Cheng, E. M. Haacke, M. R. Thompson, and R. Venkatesan, *Magnetic Resonance Imaging: Physical Principals and Sequence Design*, John Wiley & Sons, Inc., Hoboken, New Jersey, 2014.

[52] R. Scherrer, *Quantum Mechanics: An Accessible Introduction*, Pearson Education Inc., Glenview, IL, 2006.

[53] F. Bloch, Nuclear Induction, Phys Rev **70**, 460–474 (1946).

[54] A. D. Elster and J. H. Burdette, *Questions and Answers in Magnetic Resonance Imaging*, Mosby, Philadelphia, PA, 2001.

[55] M. Goldman, Formal theory of spin-lattice relaxation, J. Magn. Reson. **149**, 160–187 (2001).

[56] M. A. Bernstein, K. F. King, and X. J. Zhou, *Handbook of MRI Pulse Sequences*, Elsevier Academic Press, Burlington, MA, 2004.

[57] J. Pauly, P. L. Roux, D. Nishimura, and A. Macovski, Parameter relations for the Shinnar-Le Roux selective excitation pulse design algorithm [NMR imaging], IEEE Trans Med Imaging **10**, 53–65 (1991).

[58] D. G. Nishimura, *Principles of magnetic resonance imaging*, Stanford University, Stanford, CA, 2010.

[59] G. Chavhan, P. Babyn, B. Jankharia, H. Cheng, and M. Shroff, Steady-state MR imaging sequences: Physics, classification, and clinical applications, Radiographics **28**, 1146–1160 (2008).

[60] J. Scheffler and S. Lehnhardt, Principles and applications of balanced SSFP techniques, Eur Radiol **13**, 2409–2418 (2003).

[61] O. Bieri and K. Scheffler, Fundamentals of balanced steady state free precession MRI, J Magn Reson Imaging **38**, 2–11 (2013).

[62] K. Krupa and M. Bekiesinska-Figatowska, Artifacts in magnetic resonance imaging, Polish J of Radiol **80**, 93–106 (2015).

[63] A. Deshmane, V. Gulani, M. Griswold, and N. Seiberlich, Parallel MR imaging, J Magn Reson Imaging **36**, 55–72 (2012).

[64] M. Lustig, D. Donoho, and J. Pauly, Sparse MRI: The application of compressed sensing for rapid MR imaging, Megn Reson Med **58**, 1182–1195 (2007).

[65] D. Sodickson and W. Manning, Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays, MRM **38**, 591–603 (1997).

[66] O. Jaspan, R. Fleysher, and M. Lipton, Compressed sensing MRI: a review of the clinical literature, Br J Radiol **88**, 1–12 (2015).

[67] A. M. Raid, W. M. Khedr, M. A. El-dosuky, and W. Ahmed, Jpeg image compression using discrete cosine transform - a survey, 2014.

[68] E. Yip, J. Yun, K. Wachowicz, A. Heikal, Z. Gabos, S. Rathee, and B. Fallone, Prior data assisted compressed sensing: A novel MR imaging strategy for real time tracking of lung tumors, Med Phys **41**, 1–12 (2014).

[69] F. G. Ashby, *Statistical Analysis of fMRI Data*, The MIT Press, Boston, MA, 2011.

[70] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016, `http://www.deeplearningbook.org`.

[71] D. Wolpert, The lack of a prior distinction between learning algorithms, Neural Computation **8**, 1341–1390 (1996).

[72] D. Wolpert and W. MacReady, No free lunch theorems for optimization, IEEE Transactions on Evolutionary Computation **1**, 67–82 (1997).

[73] W. McCulloch and P. Walter, A logical calculus of ideas immanent in nervous activity, Bulletin of Mathematical Biophysics **5**, 115–133 (1943).

[74] F. Rosenblatt, The perceptron - A perceiving and recognizing automation, Cornell Aeronautical Laboratory , 85–460–1 (1957).

[75] G. Cybenko, Approximations by superpositions of sigmoidal functions, Math Control Signals Systems **2**, 303–314 (1989).

[76] Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang, The expressive power of neural networks: A view from the width, NIPS , 1–9 (2017).

[77] M. A. Nielsen, *Neural Networks and Deep Learning*, Determination Press, 2015.

[78] T. Fushiki, Estimation of prediction error by using K-fold cross-validation, Statistics and Computing **21**, 137–146 (2011).

[79] L. Bottou, Online learning and stochastic approximations, AT&T Labs Research (1998).

[80] B. Polyak, Some methods of speeding up the convergence of iteration methods, USSR Computational Mathematics and Mathematical Physics **4**, 1–17 (1964).

[81] T. Tieleman and G. Hinton, Lecture 6.5-RMSProp: Divide the Gradient by a Running Average of its Recent Magnitude, COURSERA: Neural Networks for Machine Learning **4**, 26–31 (2012).

[82] D. Kingma and J. Ba, Adam: A method for stochastic optimization, CoRR **abs/1412.6980** (2015).

[83] D. Masters and C. Luschi, Revisiting small batch training for deep neural networks, arXiv e-prints **abs/1804.07612v1** (2018).

[84] S. Reddi, S. Kale, and S. Kumar, On the convergence of adam and beyond, arXiv e-prints **abs/1904.09237** (2019).

[85] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE **86**, 2278–2324 (1998).

[86] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, in *Advances in Neural Information Processing Systems 25*, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, pages 1097–1105, Curran Associates, Inc., 2012.

[87] B. Sahiner, A. Pezeshk, L. M. Hadjiiski, X. Wang, K. Drukker, K. H. Cha, R. M. Summers, and M. L. Giger, Deep learning in medical imaging and radiation

therapy, Med Phys **46**, e1–e36 (2019).

[88] J. Elman, Finding structure in time, Cognitie Science **14**, 179–211 (1990).

[89] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative Adversarial Networks, 2014.

[90] A. Graves, A. Mohamed, and G. Hinton, Speech recognition with deep recurrent neural networks, in *IEEE International Conference on Acoustics*, pages 6645–6649, 2013.

[91] T. Karras, S. Laine, and T. Aila, A style-based generator architecture for generative adversarial networks, 2018.

[92] T. Oh, T. Dekel, C. Kim, I. Mosseri, W. T. Freeman, M. Rubinstein, and W. Matusik, Speech2Face: Learning the Face Behind a Voice, 2019.

[93] S. Valverde, M. Cabezas, E. Roura, S. Gonzalez-Villa, D. Pareto, J. C. Vilanova, L. Ramio-Torrenta, A. Rovira, and A. Oliver, Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach, Neuroimage **15**, 159–168 (2017).

[94] K. Zeng, H. Zheng, C. Cai, Y. Yang, K. Zhang, and Z. Chen, Simultaneous single- and multi-contrast super-resolution for brain MRI images based on a convolutional neural network, Comput Biol Med **14**, 133–141 (2018).

[95] U. Kensuke, O. Junko, and I. Takayuki, Application of super-resolution convolutional neural network for enhancing image resolution in chest CT, J Digit Imaging **4**, 441–450 (2018).

[96] G. L. da Silva, T. L. Valente, A. C. Silva, A. C. de Paiva, and M. Gattass, Convolutional neural network-based PSO for lung nodule false positive reduction

on CT images, Comput Methods Programs Biomed **162**, 109–118 (2018).

[97] J. Schlemper, J. Caballero, J. Hajnal, A. Price, and D. Rueckert, A deep cascade of convolutional neural networks for dynamic MR image reconstruction, IEEE Trans on Med Imaging **37**, 491–503 (2018).

[98] J. Caballero, D. Rueckert, and J. Hajnal, Dictionary learning and time sparsity for dynamic MR data reconstruction, IEEE Trans Med Imag **33**, 3201–3226 (2014).

[99] C. Qin, J. Schlemper, J. Caballero, A. Price, J. Hajnal, and D. Rueckert, Convolutional recurrent neural networks for dynamic MR image reconstruction, IEEE Trans of Med Imaging **38**, 280–290 (2019).

[100] E. Gong, J. M. Pauly, M. Wintermark, and G. Zaharchuk, Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI, J Magn Reson Imaging **48**, 330–340 (2018).

[101] S. Soffer, A. Ben-Cohen, O. Shimon, M. M. Amitai, H. Greenspan, and E. Klang, Convolutional Neural Networks for Radiologic Images: A Radiologist's Guide, Radiology **290**, 590–606 (2019).

[102] B. Dietz, E. Yip, J. Yun, B. Fallone, and K. Wachowicz, Real-time dynamic MR image reconstruction using compressed sensing and principal component analysis (CS-PCA): Demonstration in lung tumor tracking, Med Phys **63** (2017).

[103] J. Yun, E. Yip, Z. Gabos, K. Wachowicz, S. Rathee, and B. Fallone, Neural-network based autocontouring algorithm for intrafractional lung-tumor tracking using Linac-MR, Med Phys **42**, 2096–2310 (2015).

[104] J. Tsao and S. Kozerke, MRI temporal acceleration techniques, J Magn Reson Imaging **36**, 543–560 (2012).

[105] R. Otazo, E. Candes, and D. Sodickson, Low–rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components, Magn Reson Med **73**, 1125–36 (2015).

[106] L. Feng, M. B. Srichai, R. P. Lim, A. Harrison, W. King, G. Adluru, E. V. R. Dibella, D. K. Sodickson, R. Otazo, and D. Kim, Highly accelerated real-time cardiac cine MRI using k-t SPARSE-SENSE, Magn Reson Med **70**, 64–74 (2013).

[107] G. H. Dunteman, *Principal Component Analysis*, Sage Publications, Newbury Park, CA, 1989.

[108] M. Turk and A. Pentland, Eigenfaces for recognition, J Cognitive Neurosci **3**, 71–86 (1991).

[109] F. Zong, M. Nogueira, and P. Galcosas, Fast reconstruction of highly undersampled MR images using one and two dimensional principal component analysis, Magn Reson Imag **34**, 227–37 (2016).

[110] L. Dice, Measures of the amount of ecologic association between species, Ecology **26**, 297–302 (1945).

[111] Z. Xiao, W. Hoge, R. Mulkern, L. Zhao, G. Hu, and W. Kyriakos, Comparison of parallel MRI reconstruction methods for accelerated 3D fast spin-echo imaging, Magn Reson Med **60**, 650–660 (2008).

[112] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans Image Process **13**,

600–12 (2004).

[113] F. J. L. S. Senan, O. Chapet and R. K. T. Haken, Defining target volumes for non-small cell lung carcinoma, Semin Radiat Oncol **14**, 308–314 (2004).

[114] J. Yun, K. Wachowicz, S. Rathee, M. Mackenzie, D. Robinson, and B. Fallone, Evaluation of a lung tumor autocontouring algorithm for intrafractional tumor tracking using low–field MRI: a phantom study, Med Phys **39**, 1481–1494 (2012).

[115] T. Goldstein and S. Osher, The split Bregman method for L1-regularized problems, SIAM J Imaging Sci **2**, 323–243 (2009).

[116] B. Dietz, J. Yun, E. Yip, Z. Gabos, B. Fallone, and K. Wachowicz, Single patient convolutional neural networks for real-time MR reconstruction: a proof of concept application in lung tumor segmentation for adaptive radiotherapy, Phys Med Biol **64** (2019).

[117] B. Raaymakers, J. Lagendijk, J. Overweg, J. Kok, A. Raaijmakers, E. Kerkhof, R. van der Put, S. Crijns, F. Benedosso, M. van Vulpen, C. Graaff, J. Allen, and K. Brown, Integrating a 1.5 T MRI scanner with a 6 MV accelerator: proof of concept, Phys Med Biol **54**, N229–N237 (2009).

[118] S. Mutic and J. Dempsey, The Viewray system: magnetic resonance guided and controlled radiotherapy, Semin Radiat Oncol **24**, 196–199 (2014).

[119] J. Yun, E. Yip, Z. Gabos, K. Wachowicz, S. Rathee, and B. Fallone, Improved lung tumor autocontouring algorithm for intrafractional tumor tracking using 0.5 T linac-MR, Biomed Phys Eng Express **2**, 1–5 (2016).

[120] E. Yip, J. Yun, K. Wachowicz, Z. Gabos, S. Rathee, and B. Fallone, Sliding

window prior data assisted compressed sensing for MRI tracking of lung tumors, Med Phys **44**, 84–98 (2017).

[121] Theano Development Team, Theano: A Python framework for fast computation of mathematical expressions, arXiv e-prints **abs/1605.02688** (2016).

[122] S. Dieleman et al., Lasagne: First release., 2015.

[123] M. Abadi et al., TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, 2015, Software available from tensorflow.org.

[124] W. Rawat and W. Zenghui, Deep convolutional neural networks for image classification: A comprehensive review, Neural Computation **29**, 2352–2449 (2017).

[125] S. M, Simple Introduction to Convolutional Neural Networks, 2019.

[126] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, IEEE CVPR , 770–778 (2016).

[127] O. Bieri, M. Markl, and K. Scheffler, Analysis and compensation of eddy currents in balanced SSFP, Magn Reson Med **54**, 129–137 (2005).

[128] E. Yip, J. Yun, K. Wachowicz, Z. Gabos, S. Rathee, and B. Fallone, Sliding window prior data assisted compressed sensing for MRI tracking of lung tumors, Med Phys **44**, 84–98 (2017).

[129] S. Tartakovsky, S. Clark, and M. McCourt, Deep Learning Hyperparameter Optimization, 2017.

[130] Z. H. Zhan, J. Zhangr, Y. Li, and H. S. Chung, Adaptive particle swarm optimization, IEEE Transactions on Systems, Man, and Cybernetics, Part B **39**, 1362–1381 (2009).

[131] B. Dietz, J. Yun, E. Yip, Z. Gabos, B. Fallone, and K. Wachowicz, Single patient convolutional neural networks for real-time MR reconstruction: Low resolution phase encoding scheme, Phys Med Biol **65** (2020).

[132] K. Eppenhof, M. Maspero, M. Savenije, J. de Boer, J. van der Voort van Zyp, B. Raaymakers, A. Raaijmakers, M. Veta, C. van den Berg, and J. Pluim, Fast contour propagation for MR-guided prostate radiotherapy using convolutional neural networks, Medical Physics (2019).

[133] J. van Vaals, M. Brummer, W. Dixon, H. Tuithof, H. Engels, R. Nelson, B. Gerety, J. Chezmar, and J. den Boer, "Keyhole" Method for Accelerating Imaging of Contrast Agent Uptake, JMRI **3**, 671–675 (1993).

[134] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, Rethinking the Inception Architecture for Computer Vision, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) , 2818–2826 (2016).

[135] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning, in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, page 4278–4284, AAAI Press, 2017.

[136] G. Kang, K. Liu, B. Hou, and N. Zhang, 3D multi-view convolutional neural networks for lung nodule classification, PLOS ONE **12**, 1–21 (2017).

[137] K. Roy, A. Jaiswal, and P. Panda, Towards spike-based machine intelligence with neuromorphic computing, Nature **575**, 607–617 (2019).

[138] W. Maass, Networks of spiking neurons: the third generation of neural network models, Neural Netw **10**, 1659–1671 (1997).

[139] M. Davies et al., Loihi: A neuromorphic many core processor with on-chip learning, IEEE Micro **38**, 82–99 (2018).

[140] S. Lingala, B. Sutton, M. Miquel, and K. Nayak, Recommendations for real-time speech MRI, J Magn Reson Imaging **43**, 28–44 (2016).

[141] S. Riederer, S. Fain, D. Kruger, and R. Busse, Real-time imaging and triggering of 3D contrast-enhanced MR angiograms using MR fluoroscopy, MAGMA **8**, 129–206 (1999).

[142] S. Zhang, A. Joseph, D. Voit, S. Schaetz, K. Merboldt, C. Unterberg-Buchwald, A. Hennemuth, J. Lotz, and J. Frahm, Real-time magnetic resonance imaging of cardiac function and flow – recent progress, Quant Imaging Med Surg **4**, 313–329 (2014).

[143] R. Wright, S. Riederer, F. Farzaneh, P. Rossman, and Y. Liu, Real-time MR fluoroscopic data acquisition and image reconstruction, MRM **12**, 407–415 (1989).

[144] A. Wilman, S. Riederer, B. King, J. Debbins, P. Rossman, and R. Ehman, Fluoroscopically triggered contras-enhanced three-dimensional MR angiography with elliptical centric view order: application to the renal arteries, Radiology **205**, 137–146 (1997).

[145] X. Feng, S. Blemker, J. Inouye, C. Pelland, L. Zhao, and C. Meyer, Assessment of velopharyngeal function with dual-planar high-resolution real-time spiral dynamic MRI, Magn Reson Med **75**, 1467–1474 (2018).

[146] Y. Kim, M. Proctor, S. Narayanan, and K. Nayak, Improved imaging of lingual articulation using real-time multislice MRI, JMRI **35**, 943–948 (2012).

[147] L. Traser, A. Ozen, F. Burk, M. Burdumy, M. Bock, B. Richter, and M. Echternach, Respiratory dynamics in phonation and breathing-A real-time MRI study, Respir Physiol Nerurobiol **236**, 69–77 (2017).

[148] N. Seiberlich, P. Ehses, J. Duerk, R. Gilkeson, and M. Griswold, Improved radial GRAPPA calibration fro real-time free-breathing cardiac imaging, Magn Reson Med **65**, 492–505 (2011).

[149] Y. Kim, R. Lebel, Z. Wu, S. Ward, M. Khoo, and K. Nayak, Real-time 3D magnetic resonance imaging of the pharyngeal airway in sleep apnea, Magn Reson Med **71**, 1501–1510 (2014).

[150] Z. Wu, W. Chen, M. Khoo, S. Ward, and K. Nayak, Evaluation of upper airway collapsibility using real–time MRI, J Magn Reson Imaging **44**, 158–167 (2016).

[151] A. Niebergall, S. Zhang, E. Kunay, G. Keydana, M. Job, M. Uecker, and J. Frahm, Real-time MRI of speaking at a resolution of 33 ms: undersampled radial FLASH with nonlinear inverse reconstruction, Magn Reson Med **69**, 477–485 (2013).

[152] S. Krohn, N. Gersdorff, T. Wassmann, K. Merboldt, A. Joseph, R. Buergers, and J. Frahm, Real-time MRI of the temporomandibular joint at 15 frames per second-A feasibility study, Eur J Radiol **85**, 2225–2230 (2016).

[153] S. Zhang, M. Uecker, D. Voit, K. Merboldt, and J. Frahm, Real–time cardiovascular magnetic resonance at high temporal resolution: radial FLASH with nonlinear inverse reconstruction, Journal of Cardiovascular Magnetic Resonance **12**, 1–7 (2010).

[154] S. Zhang, K. Block, and J. Frahm, Magnetic resonance imaging in real time: advances using radial FLASH, J Magn Reson Imaging **31**, 101–109 (2010).

[155] A. Campbell-Washburn, M. Tavallaei, M. Pop, E. Grant, H. Chubb, K. Rhode, and G. Wright, Real-time MRI guidance of cardiac interventions, J Magn Reson Imag **46**, 935–950 (2017).

Appendices

Appendix A

Real-time Nomenclature

This section has been adapted from a published article: Dietz B, Fallone BG and Wachowicz K. Nomenclature for real-time magnetic resonance imaging. Magnetic Resonance in Medicine. 2018;81(3):1483-1484 [116].

Real-time MRI (RT-MRI), also known as online or on-the-fly imaging, is a heavily researched area in the field of MRI. Over 700 manuscripts were published in 2017 via PubMed search of "real-time"' and"MRI". Although RT-MRI is an active area of research, there remains ambiguity regarding the term real-time imaging. Real-time imaging is most often synonymous with dynamic imaging studies, such as speech and swallowing [140], dynamic contrast enhanced imaging [44, 141], and cardiac studies [142]. There are two key criteria that can be used to differentiate the use of the term "RT-MRI" in literature from all other MRI as a whole:

1. The generated images represent a specific time-point in an on-going process, as in physical motion or an evolving biochemical event.

2. The generated images are available for immediate use with low latency.

With virtually no exception, the situations described in the literature meet the first criterion. However, there is a large disparity among the cases in meeting the second, and the authors believe this is an important distinction affecting both acquisition and reconstruction methods that needs to be captured in terminology. While a range of works using the RT-MRI label in the past decades have satisfied both the above criteria [140, 44, 141, 142, 143, 144], many do not, resulting in potential for confusion. For example, applications where imaging is used for on-the-fly treatment guidance will call for RT-MRI techniques. However, many studies often label their technique as "RT-MRI", but lack any indication of reconstruction time or latency for the entire reconstruction pipeline [145, 146, 43, 147, 148]. We propose to the community that techniques satisfying the first criterion alone adhere to the established description of "dynamic MRI". Techniques that satisfy both use the terminology "real-time MRI", since these are capable of real-time use, not just high frame rate dynamic visualization.

Another distinction often used in the literature as an automatic rationale for the use of the RT-MRI label is that of online reconstruction. Image reconstruction can be conducted either offline or online, the latter being reconstructed at the console. An important requirement in our proposed definition of RT-MRI is the ability to visualize the acquired data with minimal latency. Specifically, the term latency is used here to describe the delay between a time-frame of interest represented by a particular data acquisition, and its final reconstruction and visualization. Several offline reconstruction techniques utilize k-space acquired from the entire dynamic data set (in the k-t domain). The k-t reconstruction algorithms are a perfect demonstration of these (often time consuming) offline reconstruction techniques [37, 38, 39]. However, strictly speaking RT-MRI can still be achieved frame-by-frame using offline reconstruction methods, provided the combined acquisition, data transfer, and reconstruction time is fast enough to constitute RT (i.e. low latency). As such these authors believe that

the offline/online distinction is an independent concern to that of real-time imaging, and that one should not be implicitly linked to the other.

It has become apparent that researchers may claim their reconstruction technique is capable of RT-MRI, although the reconstruction time may take tens of seconds [149, 150], or several minutes [106, 151], which violates the second proposed criterion (low latency). This misuse (in our opinion) is due to researchers considering only the acquisition time, also known as the true temporal resolution [56], as opposed to the total imaging time (including reconstruction) [145, 146, 43, 147, 148, 149, 150, 106, 151, 152, 153, 154]. There have been a few recent papers that aim to define RT-MRI for specific experiments. Lingala *et al.* recommends a definition for RT-MRI in application to speech imaging [140], which proposes a maximum latency criterion of 500 msec. Campbell *et al.* outlines RT-MRI for cardiac interventions, which recommends a latency of <200 msec for tolerable device navigation [155]. While the value for minimum latency will clearly be heavily dependent on the anatomic site and intended image use (and hence require careful and rigorous determination), these publications underscore the importance of the latency criterion in real-time applications.

RT-MRI is a heavily researched area in MR, and as such, we propose that care must be taken as to what experiments are considered RT-MRI. It should be noted that RT-MRI is inherently coupled to computational hardware (i.e. CPU's, GPU's) used for image reconstruction. Techniques that currently have reconstruction times that result in a long delay between data acquisition and visualization may one day be reasonably considered real-time based on improved hardware and algorithm development alone. However, predictions of increased acquisition/reconstruction times alone do not warrant a claim of RT-MRI, although predictions of that nature could

well be included in discussion. We look forward to hearing responses from the MR community regarding the definition of RT-MRI.

Appendix B

CNN Diagram

Below is a pictorial description of the convolutional neural network utilized throughout the study from Chapter 4. Starting with the first convolution (using 32 filters) of the first data consistency step, to the final reconstructed image.

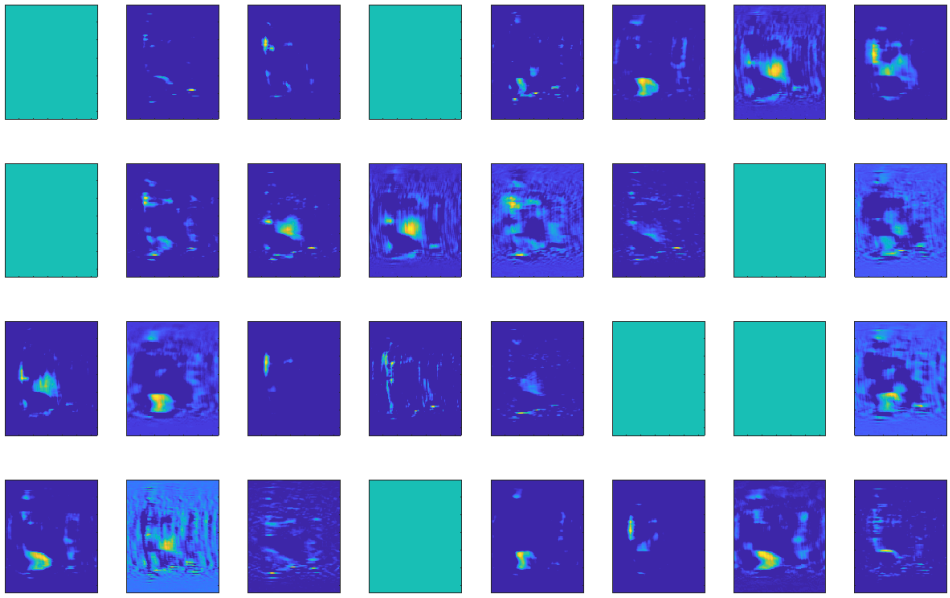

Figure B.1: First convolution and first data consistency step.

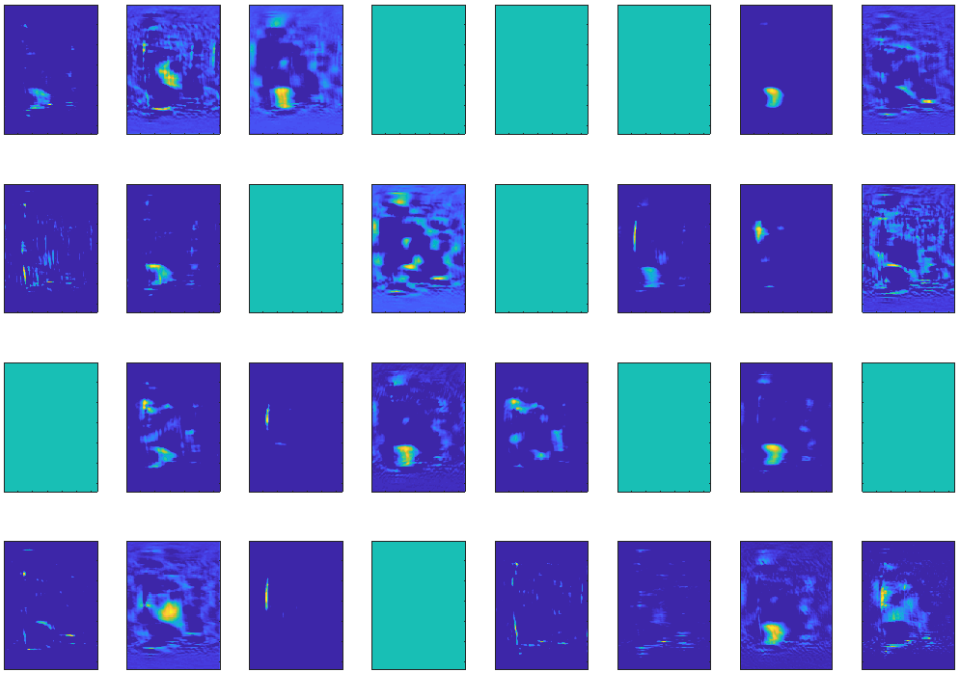

Figure B.2: Second convolution first data consistency step.

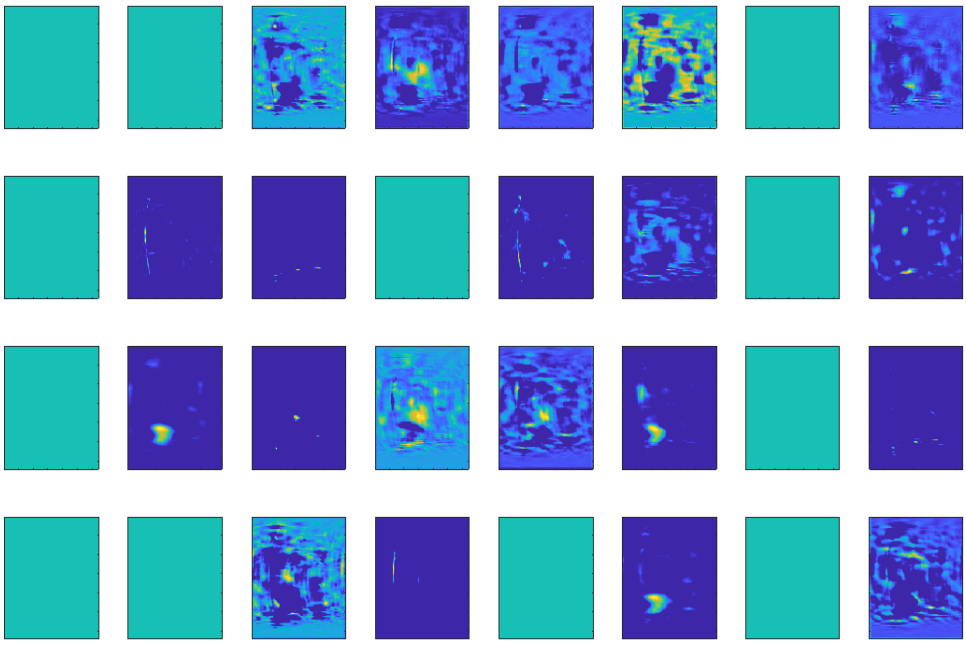

Figure B.3: Third convolution first data consistency step.

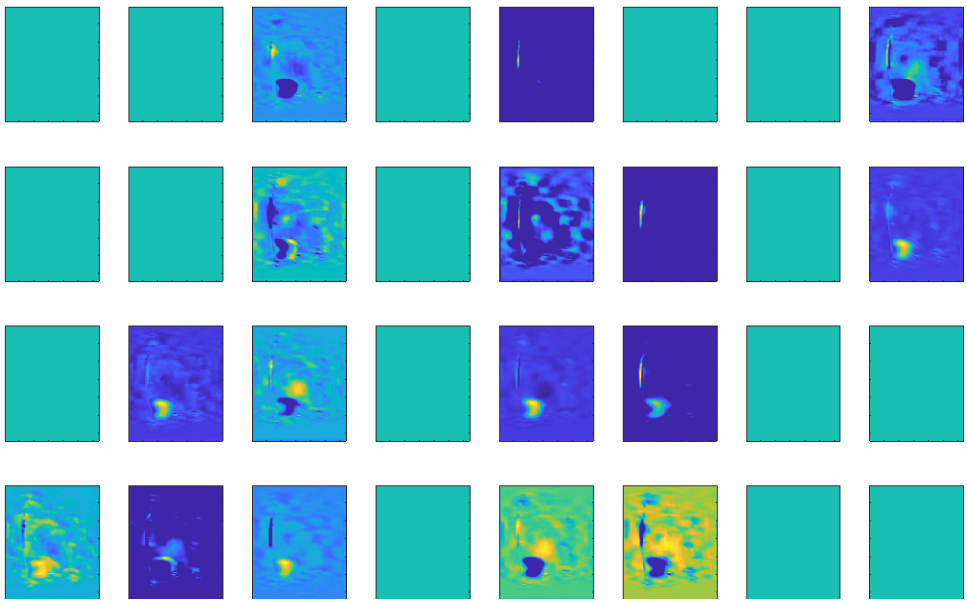

Figure B.4: Fourth convolution first data consistency step.

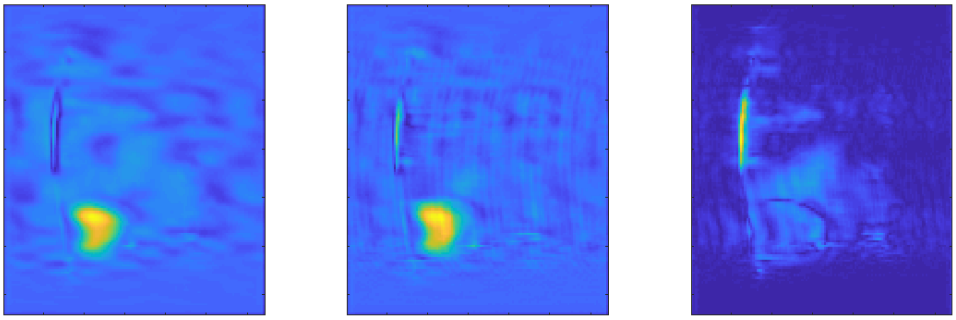

Figure B.5: Images of the aggregated convolution, residual connection (initial data added to the aggregated convolution), and after the first data consistency step.

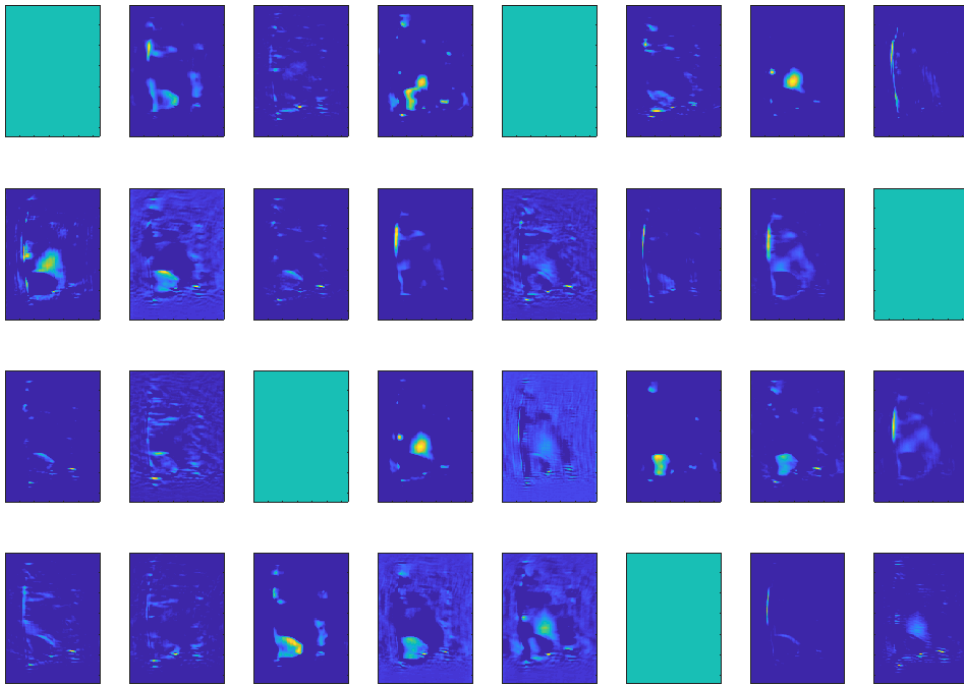

Figure B.6: First convolution and second data consistency step.

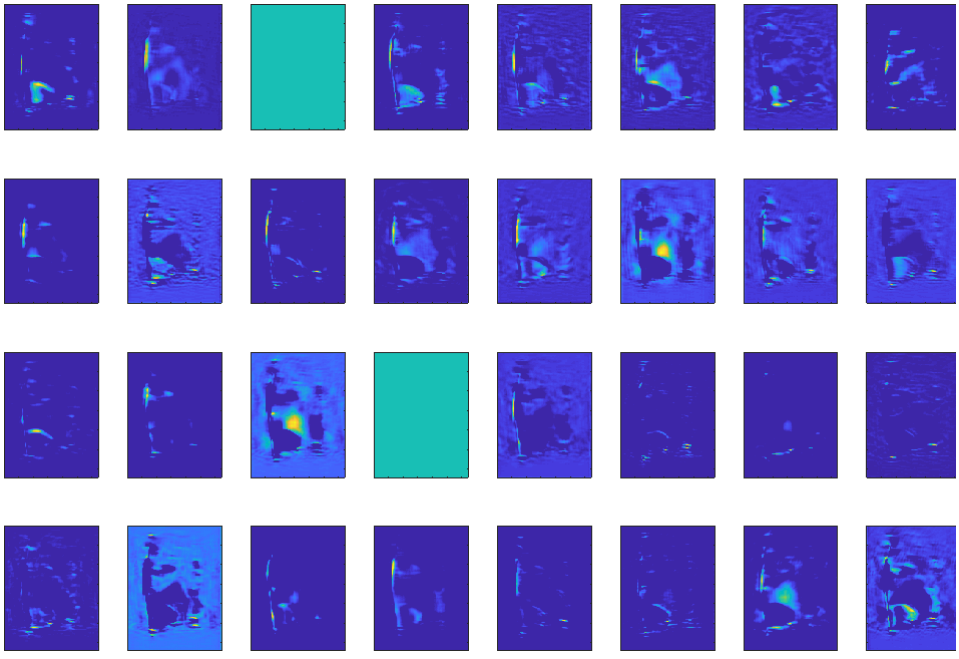

Figure B.7: Second convolution second data consistency step.

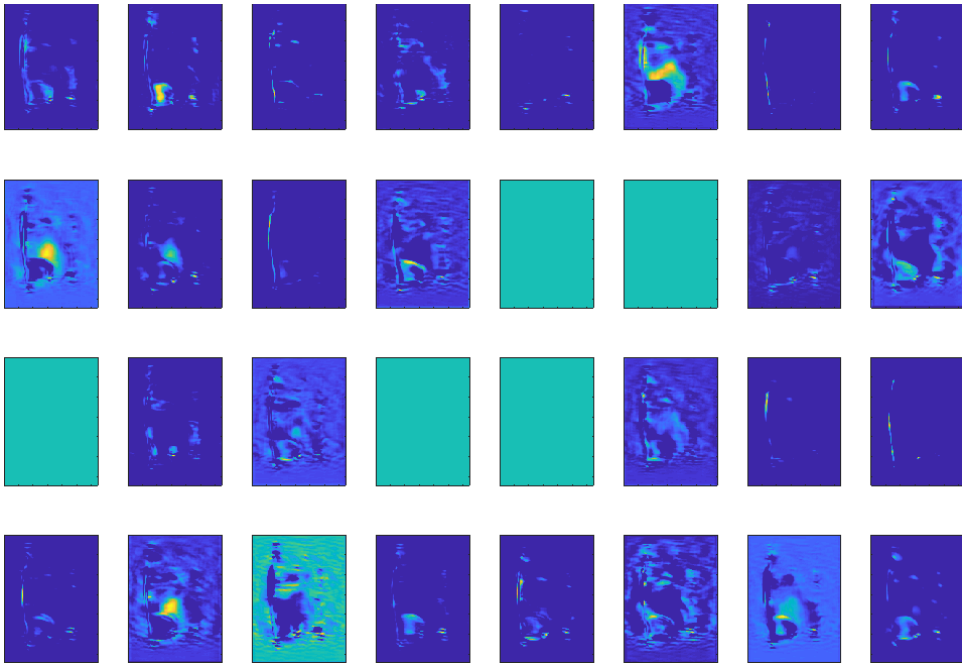

Figure B.8: Third convolution second data consistency step.

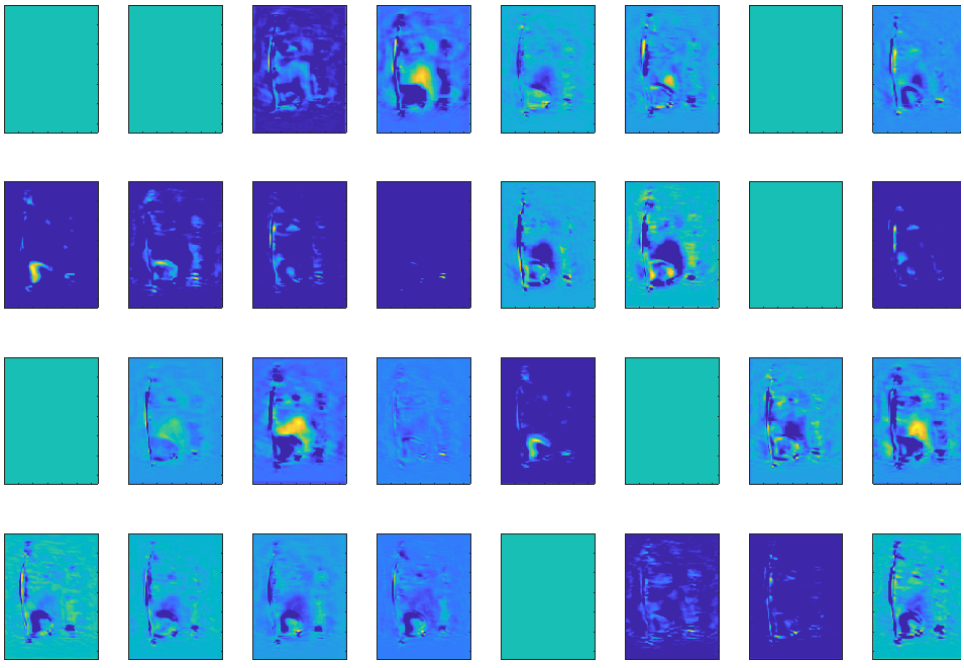

Figure B.9: Fourth convolution second data consistency step.

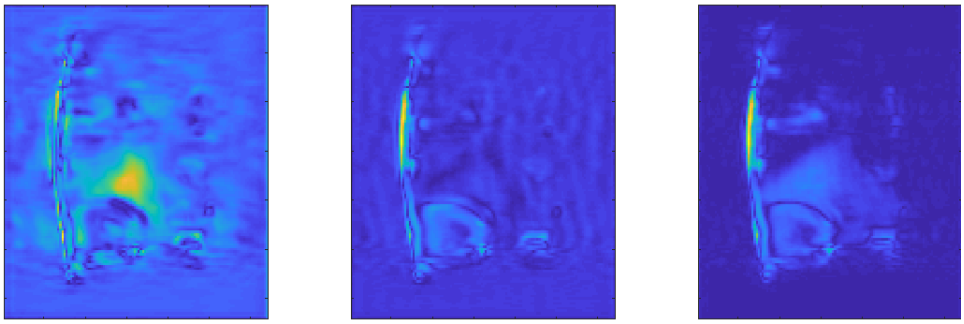

Figure B.10: Images of the aggregated convolution, residual connection (initial data added to the aggregated convolution), and after the second data consistency step.

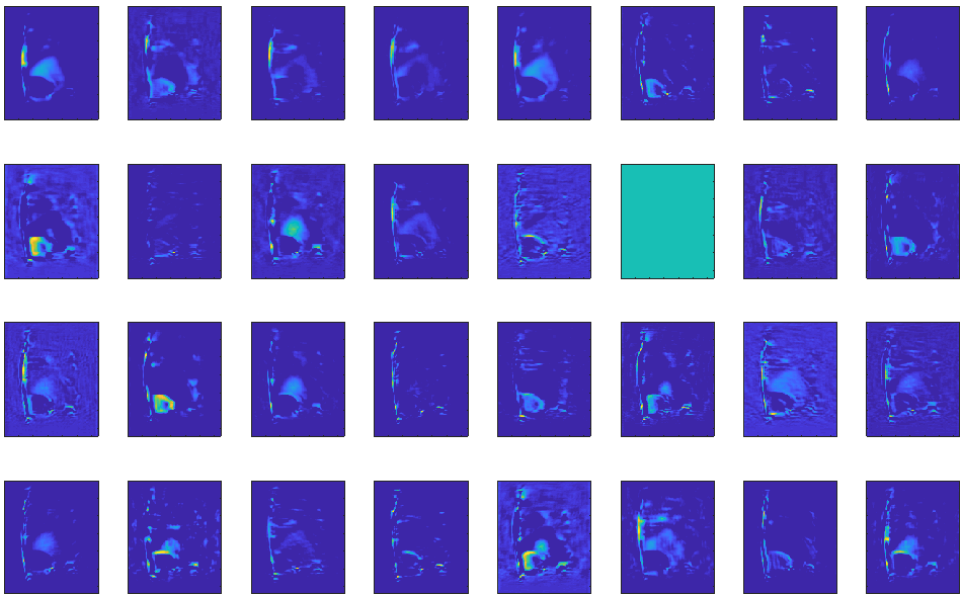

Figure B.11: First convolution and third data consistency step.

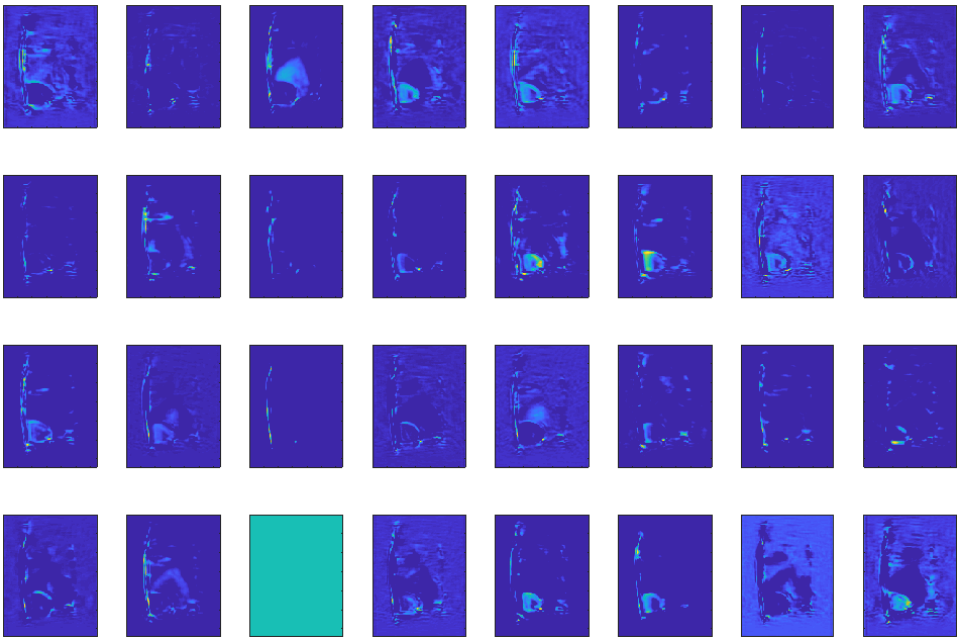

Figure B.12: Second convolution third data consistency step.

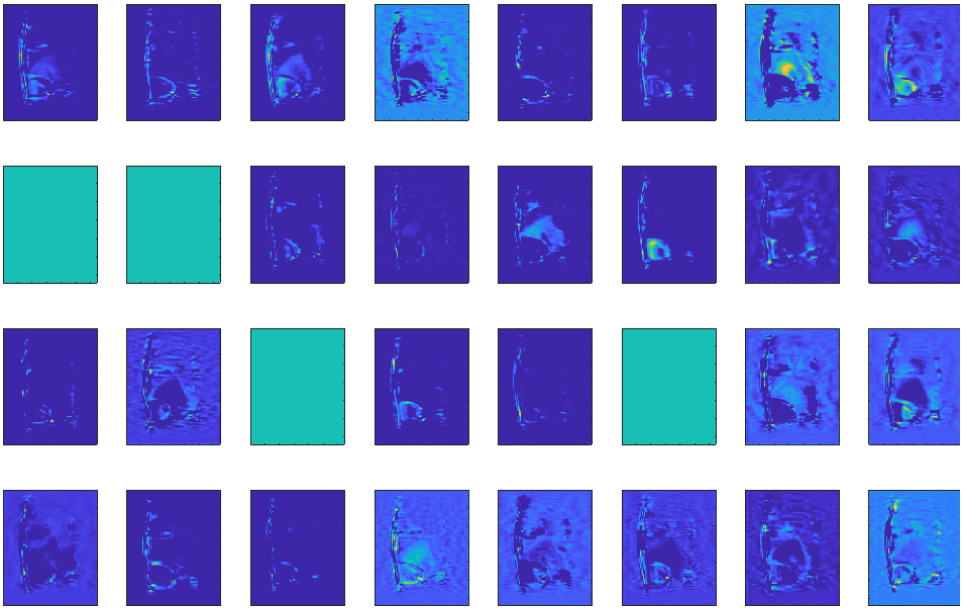

Figure B.13: Third convolution third data consistency step.

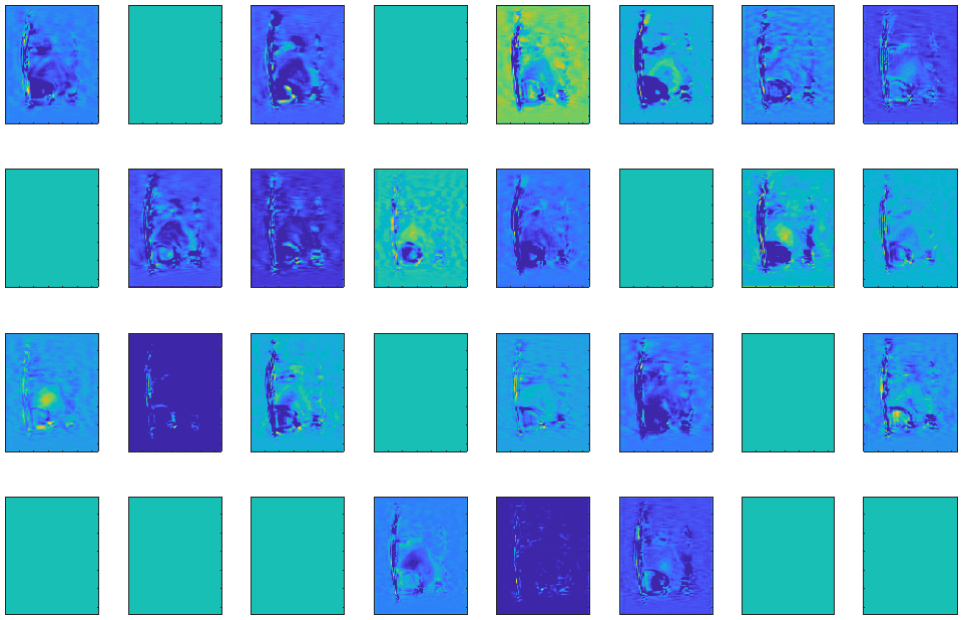

Figure B.14: Fourth convolution third data consistency step.

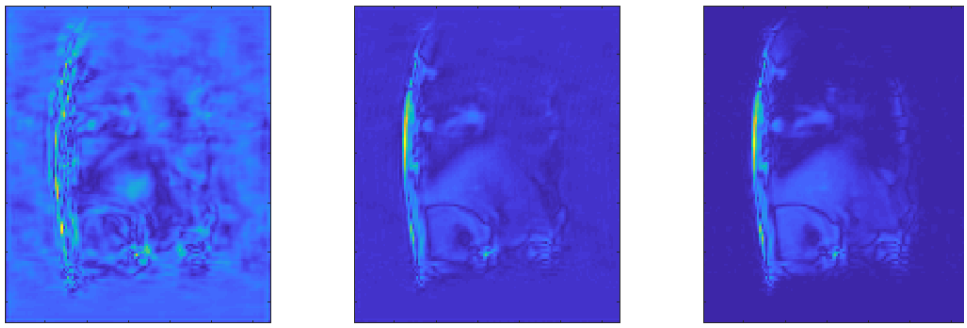

Figure B.15: Images of the aggregated convolution, residual connection (initial data added to the aggregated convolution), and after the third data consistency step.

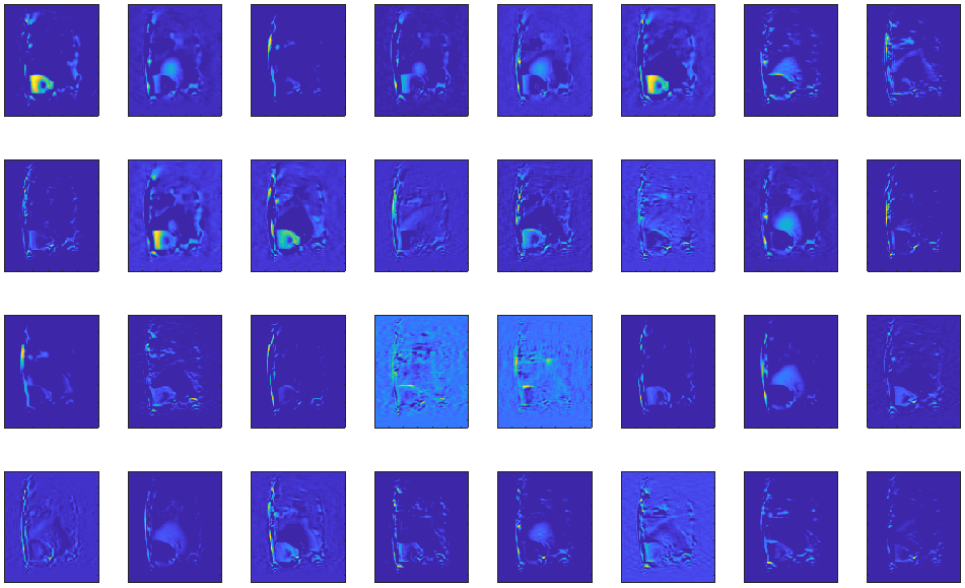

Figure B.16: First convolution and fourth data consistency step.

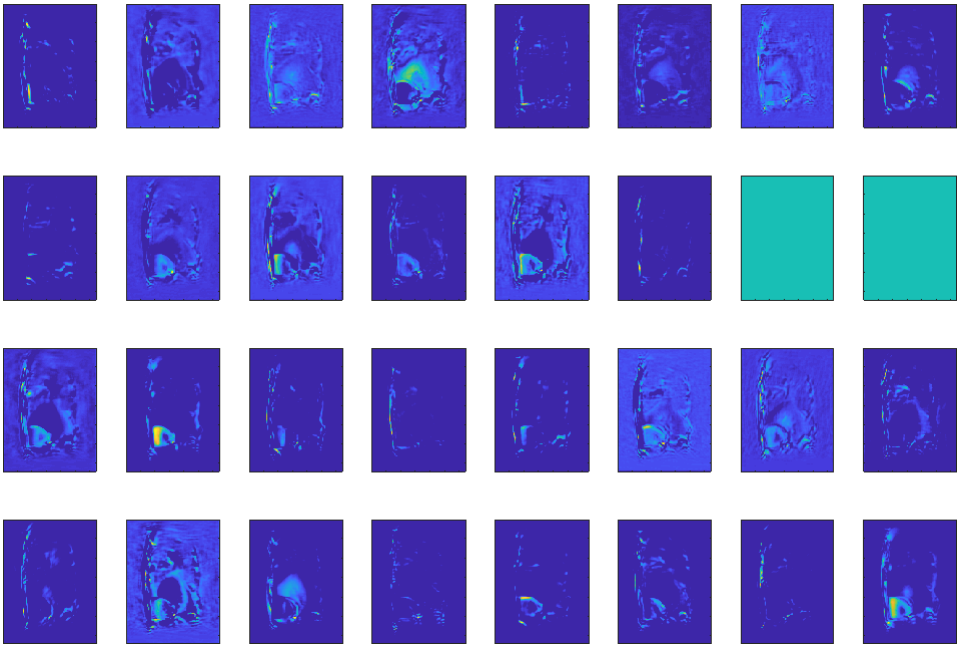

Figure B.17: Second convolution fourth data consistency step.

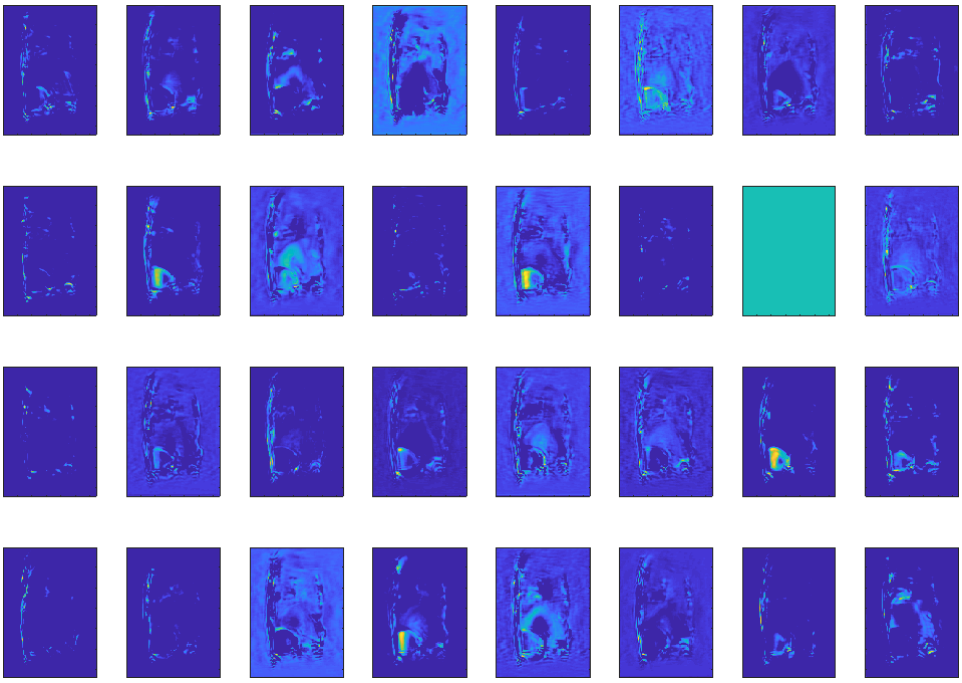

Figure B.18: Third convolution fourth data consistency step.

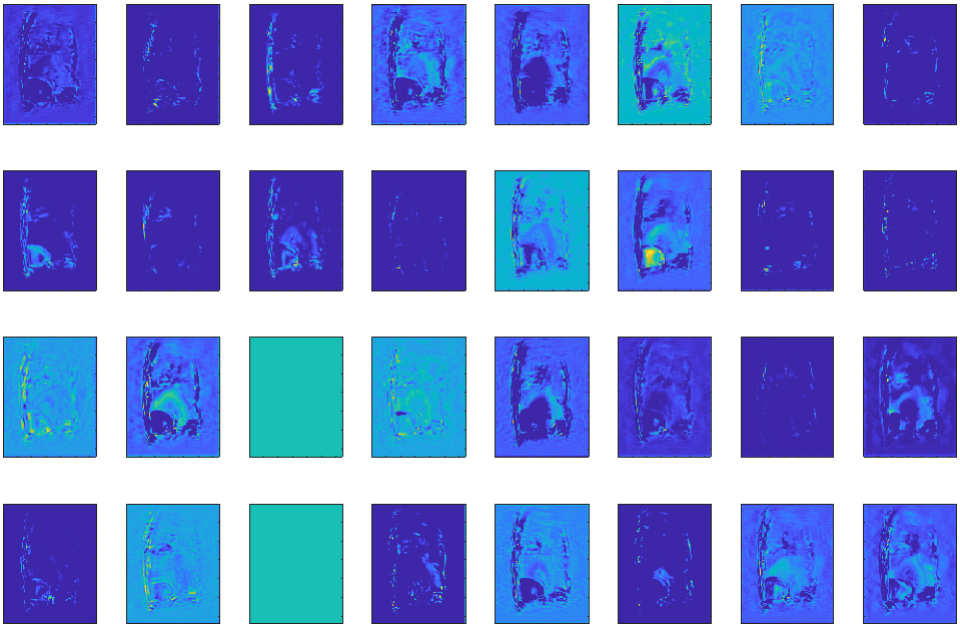

Figure B.19: Fourth convolution fourth data consistency step.

Figure B.20: Images of the aggregated convolution, residual connection (initial data added to the aggregated convolution), and after the fourth data consistency step.