Real-time MRI-guided radiotherapy with deep learning

Maarten Lennart Terpstra

Cover
The stylistic interpretation by deep neural networks of a proton in a magnetic field. The style of "De sterrennacht" (*The starry night*) by Vincent van Gogh was applied using a VGG-19-based neural style transfer algorithm to an image generated by the "Dall-e Mini" model using the text prompt: "A proton undergoing relaxation in a magnetic field".

Real-time MRI-guided radiotherapy with deep learning
PhD Thesis, Utrecht University, The Netherlands
Maarten Lennart Terpstra (2023)

Real-time MRI-guided radiotherapy with deep learning

Tijdsgetrouwe kernspintomografie-gestuurde radiotherapie met neurale netwerken

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. H.R.B.M. Kummeling, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op dinsdag 14 november 2023 des middags te 2:15 uur

door

Maarten Lennart Terpstra

geboren op 19 oktober 1992
te Sneek

Contents

List of Acronyms

3D Three-dimensional

4D Four-dimensional

bSSFP Balanced steady-state free precession

$\mathbb{C}$ The set of complex numbers $z = a + bi$ where $a, b \in \mathbb{R}$ and $i^2 = -1$

CBCT Cone beam computed tomography

CNN Convolutional neural network

CT Computed tomography

CTV Clinical target volume

CS Compressed sensing

CSM Coil sensitivity map

DL Deep learning

DVF Deformation vector field

EBRT External beam radiotherapy

EPE End-point error

FFT Fast Fourier transform

FOV Field of view

GA Golden angle $\left(= \frac{2\pi}{1+\sqrt{5}} \approx 111.25° \right)$

GPU Graphics processing unit

GTV Gross tumor volume

IGRT Image-guided radiotherapy

IMRT Intensity-modulated radiotherapy

ITV Internal target volume

Linac Linear accelerator

MRI Magnetic resonance imaging

MRL Magnetic resonance imaging – linear accelerator

MRIgRT MRI-guided radiotherapy

MLC Multileaf collimator

mm millimeter

ms milliseconds

MSE Mean-squared error $\left(= \frac{1}{n} \sum_{i=0}^{n} (x_i - y_i)^2 \right)$

MV Megavolt

NRMSE Normalized root-mean-squared error

OAR Organ at risk

OF Optical flow

PET Positron emission tomography

PI Parallel imaging

PSNR Peak signal-to-noise ratio

PTV Planning target volume

PSF Point spread function

NUFFT Non-uniform fast Fourier transform

QSM Quantitative susceptibility mapping

ReLU Rectified linear unit $f(x) = \max(0, x)$

RF Radiofrequency

$\mathbb{R}^n$ The set of real numbers. If $n > 1$, $\mathbb{R}^n$ denotes an n-dimensional space.

RT Radiotherapy

RMSE Root-mean-squared error

RNN Recurrent neural network

ROI Region of interest

SABR Stereotactic ablative radiotherapy

SBRT Stereotactic body radiotherapy

SNR Signal-to-noise ratio

SPAIR Spectral Attenuated Inversion Recovery

SSIM Structural similarity

TRE Target registration error

TV Total variation

VIF Visual information fidelity

VMAT Volumetric arc therapy

CHAPTER 1

General introduction

1.1   Radiotherapy

In 2020, more than nineteen million people were diagnosed with cancer worldwide[1,2], and approximately half of these patients were treated using radiation therapy (RT)[3,4]. With external-beam radiotherapy (EBRT), a linear accelerator (linac) is used to generate photons at megavolt (MV) energy targeting the malignant tissue to damage the cell DNA and eventually killing all tumor cells when a sufficient dose is delivered[5]. Ideally, as much dose as possible is deposited in the tumor, as increased dose delivery results in better tumor control[6]. However, this will also damage nearby healthy tissue and organs at risk (OARs), requiring a delicate balancing act of dose delivery and healthy tissue sparing for optimal treatments.

Several strategies are used to spare OARs from the detrimental side effects of radiation as much as possible. First, the radiation is delivered with varying intensity from multiple angles around the patient, using the tumor as the focal point to prevent dose accumulation in healthy tissue. These delivery modes are now called intensity-modulated radiotherapy (IMRT) and volumetric arc therapy (VMAT). Second, radiation treatments are often fractionated, allowing the healthy tissue to recover between treatments as malignant tissue suffers from diminished cell repair ability[7]. Modern linear accelerators allow precise shaping of the radiation beam to match the shape and position of the tumor using a multileaf collimator (MLC). These three strategies facilitate delivering high radiation dose to malignant tissue with extreme precision while sparing nearby healthy tissue and OARs. Nowadays, EBRT with high dose rates per fraction is referred to as stereotactic ablative radiotherapy (SABR) or stereotactic body radiotherapy (SBRT)[8,9]. However, one essential prerequisite to delivering high radiation doses to malignant tissue is accurate knowledge of the tumor and OARs positions during treatment.

1.2   Image-guided radiotherapy

Nowadays, the positions of tumors and OARs are derived from structures visible from medical imaging, enabling image-guided radiotherapy (IGRT). The acquired images are the principal source of information on tumor and OAR localization and determine the dose delivery plan throughout the treatment. A radiation oncologist determines the gross tumor volume (GTV) from imaging acquired during radiation treatment simulation[10]. To account for microscopic tumor invasions near the GTV invisible to the imaging, the GTV is expanded to the clinical target volume (CTV). Then, to compensate for the motion *between* radiation fractions (*inter-fraction* motion), e.g., geometric uncertainties introduced by the imaging, changes to patient anatomy, or errors in patient positioning, the CTV is expanded to the planning target volume (PTV). Finally, motion *during* radiation delivery

(*intra-fraction* motion), e.g., motion due to physiological processes, such as baseline drift due to relaxation[11], respiration, peristalsis, or cardiac activity, can introduce positional uncertainties[12]. Physiological processes can greatly impact the delivered dose due to changes in tumor position. For example, respiratory-induced motion can cause deformations as large as 40 mm in the craniocaudal direction in abdominal organs during treatment[13,14]. This motion causes the tumor to move outside the radiation window and could result in underdosing of the malignant tissue, diminishing treatment efficacy. The simplest option to mitigate large deformations is to increase the treatment margins and enlarge the PTV encompassing the tumor and the full extent of the motion, defining the internal target volume (ITV). For example, respiratory-resolved four-dimensional imaging[15,16], where a 3D volumetric image is obtained for several respiratory phases, can be acquired prior to treatment to estimate the size of the ITV. Resolving respiratory-induced motion for abdominal radiotherapy with minimal additional treatment margins will be the prime focus of this thesis.

The geometric accuracy, positional certainty, and image quality directly influence the treatment outcome: insufficient image contrast between malignant and healthy tissue can cause failures in tumor delineation, reducing treatment quality[17]. Moreover, geometric inaccuracies in the imaging, errors in patient positioning, and physiological motion will create random and systematic errors, causing a cumulative error in the delivered dose[18]. Hence, it is paramount that the images that dictate where radiation is delivered have sufficient contrast, do not suffer from geometric distortion, and are of high quality, i.e., free of image artifacts and high resolution. Several imaging modalities have been used for radiotherapy. For example, computed tomography (CT), positron emission tomography (PET), and magnetic resonance imaging (MRI) have been used for treatment planning[19–21], while CT, cone-beam CT (CBCT), ultrasound (US), and MRI have been used patient positioning[22–25]. These imaging modalities have complementary properties and limitations whose information can be used for radiation treatment planning and dose delivery. For example, the electron density can be derived from CT, which is essential information for calculating the absorption of radiation beams[26]. On the other hand, PET is more sensitive to the metabolic differences between malignant and benign tissue. MRI is a versatile imaging modality that exhibits excellent soft-tissue contrast as it is an endogenous imaging modality. It derives its signal from the number of hydrogen protons with tissue-specific magnetic parameters T_1 and T_2[27]. These parameters are vital for cancer imaging, as they differ between malignant and healthy tissue[28]. Examples of some of these contrasts are shown in Figure 1.1.

1.3   Real-time adaptive MRI-guided radiotherapy

Recently, hybrid treatment systems employing a linear accelerator with an integrated MRI scanner (MR-Linac) have been clinically introduced (Figure 1.2),
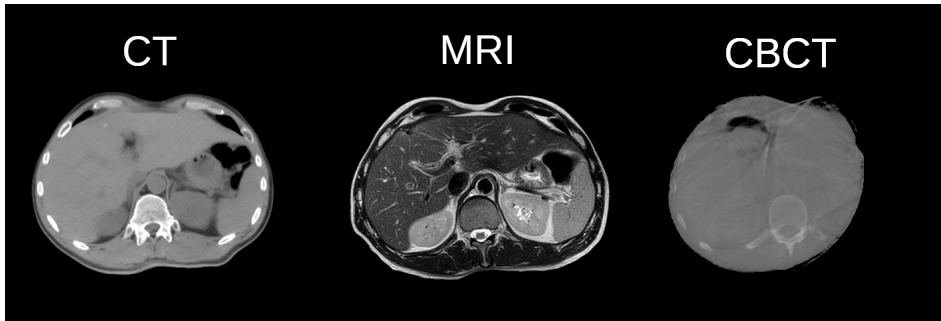
Figure 1.1 Examples of images used in radiotherapy. Anatomical imaging of the liver of the same patient using different imaging modalities in the transversal orientation. On the left is a CBCT, in the middle is a CT, and on the right is a T_2-weighted MRI. The CBCT images are used for position verification before treatment, while the CT and MR images are also used to identify tumors and plan treatment. The CBCT has a limited field of view and limited image contrast. The CT has a larger field of view but limited image contrast. The MRI shows superior soft-tissue contrast and clear visualization of the organs and small structures.

enabling the treatment of patients with MRI-guided radiotherapy (MRIgRT)[29–32]. Treating patients on MRI-Linacs has several significant advantages. Imaging the patient immediately before treatment eliminates the need to move them from an imaging device to an irradiation device, enabling treatment plan adjustment based on the patient's current anatomy without repositioning and eliminating the largest source of geometric uncertainty.[33]

Moreover, acquiring MRI can coincide with irradiation of the patient, which enables the unique opportunity to let the MRI dictate the treatment, succinctly summarized as the "See it and zap it" paradigm[34]. This synergy of simultaneous MRI and radiation delivery can mitigate intra-fraction motion using several strategies. The first strategy is to correct the treatment for baseline drift[11], which is a change in average tumor position occurring during the treatment due to, for example, patient relaxation or changes in the respiratory pattern[35], depending on the tumor location. Accounting for baseline drift by shifting the entire treatment plan can reduce systematic errors and thus greatly reduce the required PTV. The second strategy is *gating*: the tumor position is estimated from the imaging, and the radiation beam is only active when the initial treatment margin encompasses the tumor[36]. Gating is a relatively simple technique to increase treatment conformality while minimizing the irradiation of healthy tissue. However, this will significantly lengthen radiation therapy as the radiation beam is only active when the tumor is in the right place, decreasing treatment efficiency by 50-80%[36,37]. The more advanced option is *tracking*: the radiation beam follows the tumor position using motion estimated from the imaging, achieving maximum efficiency as there is no downtime. Moreover, tracking removes the need for an ITV

Figure 1.2 Example hybrid MRI/Linac scanner. Hybrid MRI-Linac scanners combine an MRI scanner for image guidance with a linear accelerator for dose delivery. The Elekta Unity is pictured, a commercially-available MR-Linac that combines a 1.5 T MRI with a 7 MV linac.

by minimizing the size of the PTV and results in minimal irradiation of healthy tissue[38]. Here, we focus on facilitating real-time MRI-guided tumor tracking as this is the most time-efficient dose delivery mode and could lead to the largest gain in OAR

Tumor tracking requires estimating motion using an image registration algorithm[39–42], which attempts to find a transformation that achieves spatial alignment between an acquired image (also named the *dynamic* or *moving* image) to an image acquired in the past (also named the *fixed* or *static* image)[43]. Some types of motion, e.g., head motion, can be defined using a rigid motion model, describing the translation and rotation of the anatomy. A deformable motion estimation algorithm is required to resolve the locally-varying anatomical changes caused by, for example, respiratory motion. Successful image registration results in a deformation vector field (DVF), a map that denotes the magnitude and direction of the motion for every voxel in the image independently, which is used to alter the position and shape of the MLC[44], focusing the radiation on the malignant tissue. An example of image registration is shown in Figure 1.3.

We are entering the real-time adaptive MRI-guided radiotherapy age using fast, high-contrast, high-quality imaging combined with fast and accurate image registration[45]. With real-time adaptive MRIgRT, the MRI and MLC must operate in a real-time feedback loop to track the tumor. While the MRI signal is acquired, the tumor motion relative to the treatment plan is quickly estimated, and the MLC position is altered to match the current shape and position of the tumor. Real-time tumor tracking results in highly accurate dose delivery, minimizing the
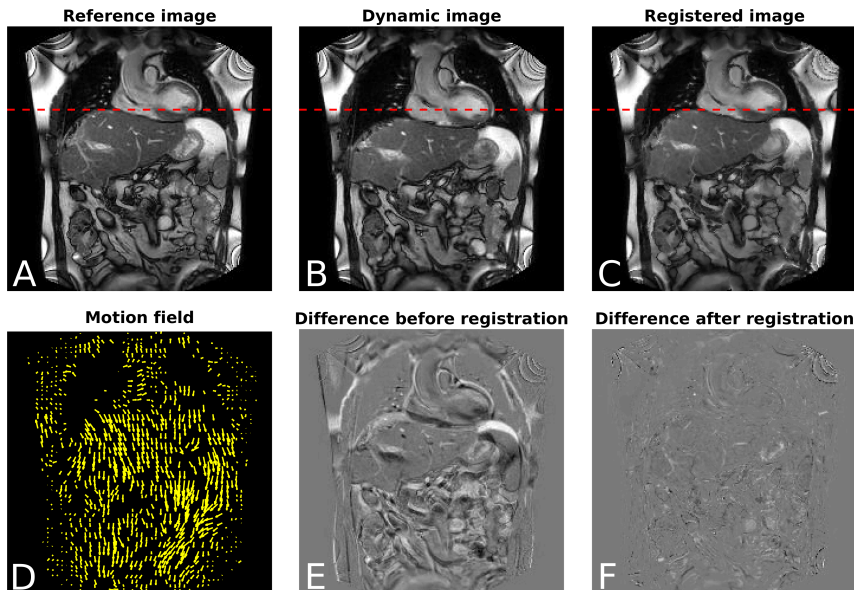
Figure 1.3 Example of image registration for coronal 2D MRI acquired at different time points. Respiratory motion can cause significant deformations between images acquired at different time points, such as images (A) and (B). Image registration finds a DVF to register a dynamic image (B) towards a reference image (A) to resolve the motion (C), re-aligning, for example, the liver dome, as illustrated by the red lines. The found DVF is shown in D, which can be used to drive the MLC. Image registration reduces the intensity difference between images (E,F).

irradiation of healthy tissue and enabling delivery of high dose rates with minimal PTV and ITV margins. As minimal healthy tissue is irradiated due to the removal or reduction of treatment margins, highly fractionated treatments are no longer necessary to achieve a high accumulated dose rate in malignant tissue while sparing healthy tissue. Instead, hypofractionated treatment regimes become possible[46,47], enabling improved local control as a higher tumor dose can be achieved. Also, hypofractionation may reduce the treatment cost by reducing the patient's hospital visits.

1.4 Advances in accelerated MRI acquisition

Enabling real-time adaptive MRIgRT requires that MRI acquisition and motion estimation are performed in real-time, displaying high spatiotemporal resolution and low latency[48]. In this work, we consider real-time methods for resolving the abdominothoracic motion to have a maximum latency of 400 ms[49,50] and a minimum temporal resolution of 200 ms ($\geq$ 5 Hz)[51]. Resolving motion from MRI within these time constraints becomes challenging due to the limited sampling

speed of MRI. The MRI acquisition involves measuring the harmonic oscillations of precessing hydrogen atoms at specific locations $\vec{k}$ in the frequency domain (hence, the frequency domain is also called k-space)[52]. Generally, the k-space is sampled line-by-line in multiple shots and must be densely filled to transform k-space into high-quality MRI. Specifically, the Shannon-Nyquist theorem[53] dictates the minimum number of samples required to obtain an MRI with a specific resolution. Failing to meet this requirement can result in image defects due to aliasing, i.e., image artifacts. A sufficiently dense k-space is reconstructed to an MR image by applying the *inverse fast Fourier transform* (FFT)[54,55], transforming the frequency domain into the spatial domain.

Sampling k-space is time-consuming, and densely filling k-space typically exceeds the maximum allowable latency. There are essentially four ways to accelerate the MRI acquisition:

1. Reducing the resolution or field of view, requiring fewer k-space samples.
2. Traverse k-space more efficiently to increase the sampling rate per read-out.
3. Alter the imaging sequence to sample shots faster.
4. Sample fewer k-space points, violating the Shannon-Nyquist theorem.

Typically, the first option is unavailable due to the resolution and field of view size requirements to accurately image and treat the patient. It is well-known that no free lunch can be offered for the remaining options: acquisitions cannot be infinitely shortened without sacrificing SNR or resolution. The third option involves changing the used scan parameters to, for example, acquire multiple k-space lines in a single shot or shorten the time between shots. However, this will also alter the image contrast and SNR, which might diminish the image quality.

1.4.1 Beyond Cartesian MRI acquisitions

Significant research has been dedicated to exploring option 2, efficient sampling of k-space. So far, we have assumed that k-space is traversed in a line-by-line fashion with equal distance to other k-space points in every direction, also called Cartesian imaging, as the sampled points will lie on a Cartesian grid. However, MRI is not limited to this Cartesian constraint, and non-Cartesian trajectories, such as spiral or radial trajectories, to traverse k-space have been proposed[56]. Examples of these trajectories are shown in Figure 1.4.

Non-Cartesian MRI offers several advantages over Cartesian imaging. First, non-Cartesian sequences can potentially sample more k-space points in a single shot, increasing the sampling efficiency and reducing the acquisition time per location $\vec{k}$ in the frequency domain. Second, a Cartesian trajectory samples the low-frequency region only once, while non-Cartesian acquisitions may sample the low-frequency region multiple times. While this sounds like a disadvantage, as repeated sampling should result in longer acquisition times, oversampling the
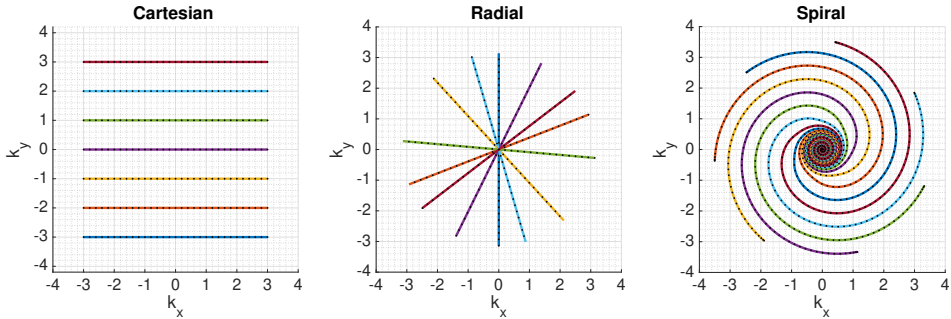
Figure 1.4 Examples of undersampled 2D trajectories. In every trajectory, seven shots are shown, with the sampled points as black dots. On the left, a Cartesian trajectory is shown. Notice that all points lie on intersections of the grid. In the middle is a golden-angle radial trajectory. On the right is a variable-density spiral. The latter two are non-Cartesian trajectories, as the distance between the sampled points on different shots is not constant, and the points do not necessarily lie on grid intersections. A NUFFT is required for reconstruction rather than an FFT for these trajectories. However, the center of the k-space is densely oversampled, and some trajectories achieve better coverage of the k-space. Moreover, non-Cartesian acquisitions promote incoherence, as shown in Figure 6.1.

low-frequency region of k-space makes these acquisition strategies significantly more motion-robust than their Cartesian counterparts[57]. In this sense, the ability for fast spatio-temporal imaging and the ability to resolve motion-induced spatial uncertainty make them particularly suited for MRIgRT and are further explored in this thesis.

However, there are also disadvantages to using non-Cartesian readouts. Hardware imperfections in the gradient system, such as gradient delays or non-linearities, can cause deviations from the nominal trajectory, leading to signal loss or geometric distortions[58,59]. Moreover, the rapid switching of the gradient field can cause eddy currents due to dB/dt effects, resulting in shifts in the k-space. While benign for Cartesian acquisitions, they can cause significant image artifacts during non-Cartesian acquisitions[60]. Finally, non-Cartesian acquisitions must be reconstructed using a more costly non-uniform Fourier transform (NUFFT) algorithm as they violate the equidistant sampling assumption of the FFT[61].

For MRIgRT, hybrid Cartesian/non-Cartesian acquisitions are often used for volumetric imaging[62]. These trajectories use a discretized non-Cartesian sampling pattern to obtain Cartesian spacing[63], or use a Cartesian spacing in the slice dimension and a 2D non-Cartesian radial or spiral trajectory is used in-plane, i.e., a stack-of-stars or stack-of-spirals trajectory[64]. However, k-space can also be sampled using three-dimensional non-Cartesian trajectories, such as a golden-mean kooshball[65].

1.4.2 Undersampling

Violating the Shannon-Nyquist criterion and acquiring fewer k-space samples is perhaps the simplest way to accelerate MRI acquisitions. However, it will result in severe aliasing artifacts as the reconstruction from k-space to image-space becomes ill-conditioned. However, the Shannon-Nyquist sampling theorem is not a necessary condition to reconstruct artifact-free images[66]. High-quality MRI can be obtained from undersampled acquisitions by exploiting complementary information during sampling or by exploiting prior knowledge of the signal properties, even if the Shannon-Nyquist criterion is violated, making this an active area of research.

One of the earliest techniques to accelerate the MRI acquisition exploits the complementary signal encoding introduced by using multiple receiver coils to remove the aliasing artifacts or interpolate missing k-space points, a technique known as parallel imaging (PI)[67–69]. However, the achieved acceleration factor (R) is typically limited to a factor of 3-4 in a single direction[70]. Another way to reconstruct high-quality MRI from accelerated k-space is by using prior information about the imaged subject. For example, compressed sensing (CS)[71] exploits transform-domain sparsity by assuming natural images are compressible while carefully constructed k-space sampling trajectories introduce non-compressible aliasing artifacts. A combination of PI/CS can achieve a reduction factor of up to 12.5 for reconstructing volumetric MRI while maintaining diagnostic quality[72]. An undersampling factor of 20 can be achieved for reconstructing dynamic imaging with diagnostic image quality by exploiting the spatio-temporal sparsity of the signal[73].

At first sight, CS seems suitable for online adaptive MRIgRT: the high acceleration factor enables MRI reconstruction with high spatiotemporal resolution. However, CS suffers from long reconstruction time (i.e., several minutes for volumetric MRI) and introduces a large latency between the event and availability between images and is not feasible for online adaptive MRIgRT[48,49].

While MRI can achieve the required spatiotemporal resolution, resolving abdominothoracic motion with minimal latency is a challenging problem with traditional image reconstruction and motion estimation methods. Alternative model-based techniques, such as machine learning, might provide a solution to overcome this problem. Recent advancements in deep learning have shown promising results with low inference time, as the model training phase is performed only once before usage. This thesis will focus on developing model-based image reconstruction and motion estimation methods for real-time adaptive MRIgRT using deep learning.

1.5   Deep learning

Deep learning (DL) involves developing highly-parameterized models trained to "learn" to detect implicit features from data[74,75]. This is a radically different approach from traditional algorithms for regression and classification, which rely on the features and inductive biases as defined by their human implementer.

Supervised machine learning involves training on a non-linear model f that is parameterized by Θ using a large dataset consisting of pairs of desired input and output (x_i, y_i). The training phase aims to find the weights Θ such that a loss function $\mathcal{L}(f_\Theta(x_i), y_i)$ is minimized for the entire dataset. These weights are found by performing *gradient descent*. In the simplest form, the parameters are found by $\Theta_{t+1} = \Theta_t - \eta \nabla \mathcal{L}$, where η is the step size or learning rate, and $\nabla \mathcal{L}$ is the gradient of the cost function over the dataset. However, more advanced optimization schemes can also be used[76]. The most widely-used algorithm for gradient computation in deep neural networks is backpropagation[77]. On GPUs, this is efficiently implemented using reverse-mode automatic differentiation[78]. The choice of cost function $\mathcal{L}$ depends on the problem to be solved but is commonly chosen as the mean-squared error $\mathcal{L}_{\text{MSE}} = \frac{1}{N} \sum_{i=0\ldots N-1} ||f_\Theta(x_i) - y_i||^2$ (MSE) for regression tasks. The power of neural networks is that they are universal function approximators[79]. A model can *learn* the transform function to map some x_i to y_i after sufficient training, depending on the quality and size of the data, the model capacity[80], the specific model architecture, and the hyperparameters such as batch size or learning rate. However, depending on the size of the dataset, the optimization procedure, and the specific model parameters, finding the minimum of $\mathcal{L}$ is a costly operation, and training such models on commodity hardware can take days to weeks. When the relatively costly training step of deep learning models has been completed, inference of deep convolutional neural networks using GPUs is fast (e.g., milliseconds or seconds).

With the advent of high-performance graphics processing units (GPUs) and the availability of large datasets, deep learning has become a feasible approach for regression and classification. Modern deep neural networks for image tasks are implemented using the convolutional inductive bias[81,82], consisting of multiple hierarchical layers with non-linear activation functions to learn robust, high-level, and expressive features[83,84]. By training efficiently on large datasets, DL models with millions of trainable parameters have been demonstrated to outperform handcrafted models for regression and classification problems[85]. Specifically for IGRT, convolutional neural networks have been applied to image segmentation[86], image registration[87], contour propagation[88], MRI reconstruction[89–94], CBCT-to-CT synthesis[95], and MRI-to-CT synthesis[96], among other tasks, achieving state-of-the-art performance.

The capability to perform fast and accurate regression makes DL a promising approach for real-time MRI reconstruction and motion estimation from
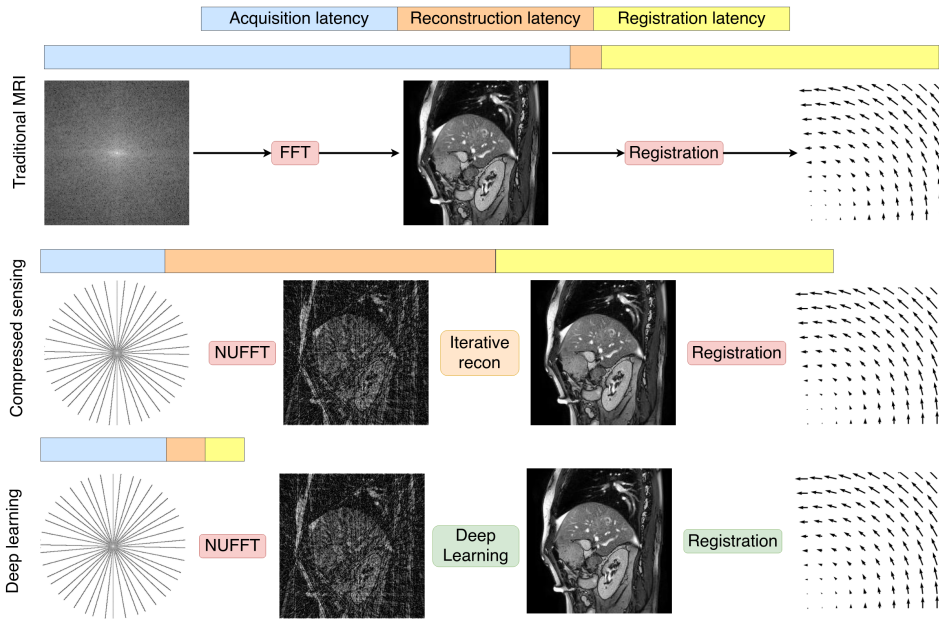
Figure 1.5 Trade-offs in image reconstruction and registration. Three pathways are shown to obtain DVFs from MRI along with a bar visualizing the latency to acquire the MRI (blue), reconstruct the k-space (orange), and estimate the motion (yellow). The original approach (top) acquires a fully-sampled k-space. While image reconstruction is fast, MRI acquisition image registration induces a long latency. The middle pathway uses compressed sensing to reduce the acquisition latency. However, this induces a long reconstruction latency due to its iterative nature. The bottom pathway uses deep learning (green) to replace the traditional counterparts (red) and acquires, reconstructs, and registers highly undersampled MRI with minimal latency.

highly-undersampled k-space, leveraging the parallel processing capabilities of modern GPUs during inference while the time-consuming training phase is performed offline. This potential benefit is illustrated in Figure 1.5, where DL is a low-latency alternative to computationally expensive image reconstruction and motion estimation algorithms.

However, deep learning has not yet been applied to real-time image reconstruction and motion estimation for real-time adaptive MRI-guided radiotherapy. This thesis explores this exciting intersection of technology and opportunity.

1.6   Thesis outline

This thesis explores the use of deep learning models to enable real-time adaptive MRI-guided radiotherapy using MRI-Linacs. With deep learning models, we aim to obtain high-quality MRI and accurate motion estimation from highly undersampled acquisitions with low latency and high temporal resolution, traditionally very challenging with classical approaches. In this work, we develop and evaluate DL-based methods for resolving the abdominothoracic motion to enable online adaptive MRIgRT. Moreover, we investigate the role of loss functions when training these models and apply deep learning to high-dimensional image reconstruction to accelerate the radiotherapy workflow.

DVFs must be accurately estimated from undersampled acquisitions to enable real-time adaptive radiotherapy. Deep learning can be individually applied to the image reconstruction and motion estimation steps. Alternatively, deep learning can replace both steps simultaneously in an end-to-end approach. **Chapter 2** examines where to best use deep learning by comparing all four combinations. We examine the latency and performance of the best-performing approach for retrospectively-undersampled radial 2D cine imaging and establish the maximum achievable undersampling factor.

Based on these results, we extended this approach to volumetric imaging in **Chapter 3** using MRI acquired with a golden-angle stack-of-stack radial trajectory. To overcome the low temporal resolution of the acquisition, the model was trained with respiratory-resolved four-dimensional abdominal MRI. A volumetric, multi-resolution motion network was proposed, and a network hyperparameter and model architecture optimization was conducted. The stability and performance at high undersampling factors were evaluated using time-resolved 3D MRI, a digital phantom, a physical motion phantom, and four-dimensional respiratory-resolved CT.

Chapter 4 examines the properties of conventional loss functions for deep neural networks when applied to noisy regression in the complex domain. Using previously-trained motion networks, we empirically observed a systematic underestimation in the reconstructed magnitude of the DVFs. We propose and examine a new loss function for complex regression tasks to resolve this magnitude underestimation and evaluate its properties for complex-valued regression tasks, such as image reconstruction and image registration.

Respiratory-resolved 4D-MRI is a versatile tool in the MRIgRT workflow: it can quantify the motion extent to determine the ITV and is used to develop personalized motion models for online adaptive MRIgRT. However, 4D-MRI suffers from long acquisition and reconstruction times. In **Chapter 5**, we propose and evaluate a new deep learning architecture that employs low-dimensional convolution kernels to reconstruct 4D-MRI efficiently.

Finally, we summarize and discuss the main findings of this work in **Chapter 6**, reflecting on the potential impact, limitations, and challenges of applying deep learning to real-time adaptive MRI-guided radiotherapy.

CHAPTER 2

Deep learning-based image reconstruction and motion estimation from undersampled radial k-space for real-time MRI-guided radiotherapy

ML Terpstra
M Maspero
F d'Agata
B Stemkens
MPW Intven
JJW Lagendijk
CAT van den Berg
RHN Tijssen

Abstract

Purpose: To enable magnetic resonance imaging (MRI)-guided radiotherapy with real-time adaptation, motion must be quickly estimated with low latency. The motion estimate is used to adapt the radiation beam to the current anatomy, yielding a more conformal dose distribution. As the MR acquisition is the largest component of latency, deep learning (DL) may reduce the total latency by enabling much higher undersampling factors compared to conventional reconstruction and motion estimation methods. The benefit of DL on image reconstruction and motion estimation was investigated for obtaining accurate deformation vector fields (DVFs) with high temporal resolution and minimal latency.

Methods: 2D cine MRI acquired at 1.5T from 135 abdominal cancer patients were retrospectively included in this study. Undersampled radial golden angle acquisitions were retrospectively simulated. DVFs were computed using different combinations of conventional- and DL-based methods for image reconstruction and motion estimation, allowing a comparison of four approaches to achieve real-time motion estimation. The four approaches were evaluated based on the end-point-error and root-mean-square error compared to a ground-truth optical flow estimate on fully-sampled images, the structural similarity (SSIM) after registration and time necessary to acquire k-space, reconstruct an image and estimate motion.

Results: The lowest DVF error and highest SSIM were obtained using conventional methods up to $\mathcal{R} \leq 10$. For undersampling factors $\mathcal{R} > 10$, the lowest DVF error and highest SSIM were obtained using conventional image reconstruction and DL-based motion estimation. We have found that, with this combination, accurate DVFs can be obtained up to $\mathcal{R} = 25$ with an average root-mean-square error up to 1 millimeter and an SSIM greater than 0.8 after registration, taking 60 milliseconds.

Conclusion: High-quality 2D DVFs from highly undersampled k-space can be obtained with a high temporal resolution with conventional image reconstruction and a deep learning-based motion estimation approach for real-time adaptive MRI-guided radiotherapy.

2.1 Introduction

Magnetic resonance imaging-guided radiotherapy (MRIgRT) is increasingly adopted in clinical practice. Hybrid MRI scanners with an integrated linear accelerator (MR-linac) have shown to be very efficient in dealing with inter-fraction anatomical changes by employing online re-planning prior to each treatment session[30,33].

The future promise of hybrid MR-linac systems is to not only account for inter-fraction motion but also adapt the radiation delivery in real-time during treatment to accommodate for respiration or cardiac-induced intra-fraction motion[45,97–100], peristaltic motion and tissue deformation, e.g. due to bladder filling or passing air bubbles.

Real-time adaptive radiotherapy requires imaging with extremely high temporal resolution as well as a very low total latency (i.e., the time between an event and response) of the MR-linac feedback chain[49]. The most significant source of latency in the MR-linac feedback chain is MR image acquisition[101]. If acquisitions could be significantly undersampled, motion could be estimated with minimal latency. Although dense array radio-lucent receiver coils improve the acquisition speed of MR-linac systems by use of parallel imaging[102], most motion quantification techniques are image-based and rely on high-quality images, which limits the maximum acceleration factors achievable with parallel imaging[70]. Regularized reconstruction methods like compressed sensing[71] may achieve even higher acceleration factors, but the iterative nature of compressed sensing reconstruction algorithms make it unsuitable for real-time applications.

Recently, deep learning (DL) has become a popular technique in many scientific fields due to its high-quality results and speed. The use of neural networks to generate a hierarchical representation of the input data to achieve high task-specific performance without the need of hand-engineered features has proven extremely powerful for imaging applications[103–105]. In computer vision, various DL methods have been developed that outperform traditional motion estimation algorithms[106–108], while for MRI several DL methods have been proposed to replace the computationally expensive compressed sensing reconstructions[89,90,93].

In this paper, we investigate the performance of DL for image reconstruction and motion quantification on highly undersampled golden-angle (GA) radial acquisitions for real-time MRIgRT with the goal of providing accurate motion quantification with minimal latency. We hypothesize that the benign undersampling artifacts in GA radial MRI in combination with DL image reconstruction provides high acceleration factors with image quality on par with CS reconstruction but at a fraction of the computation time. The addition of a DL-based motion quantification approach is believed to relax the requirements for high-quality images, potentially allowing even greater image acceleration factors.
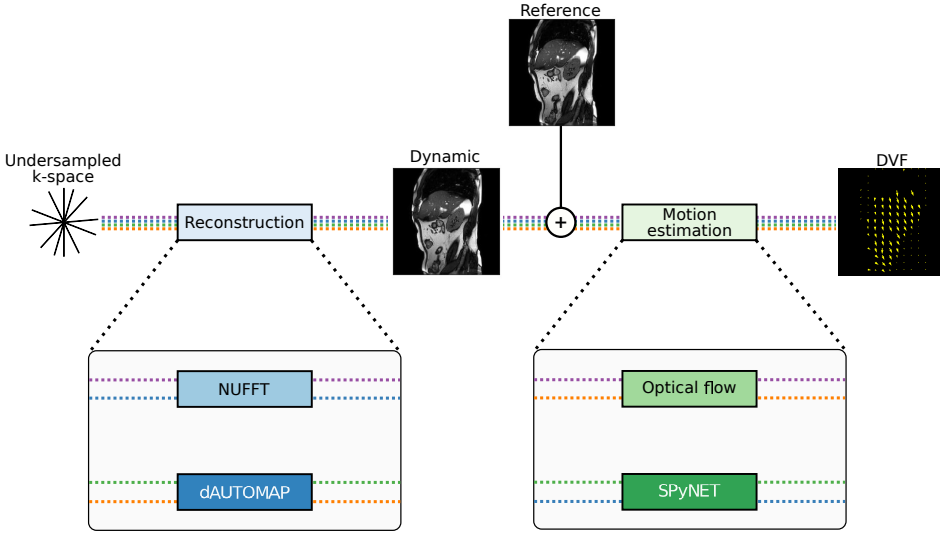
Figure 2.1 Schematic overview of the study design. Generation of undersampled k-space (top) and the DVFs from fully-sampled k-space (ground-truth) and undersampled k-space (bottom). Reconstruction can happen with a NUFFT or a DL-based image reconstruction model. Motion estimation with the reference image happens with optical flow or a DL-based motion estimation approach.

In this work, we investigate a two-step process in which retrospectively undersampled dynamic GA radial data are reconstructed by classical methods or using DL models. With this approach, we assess the individual and the combined performance of DL-based image reconstruction and processing on computation time and motion estimation accuracy for acceleration factors of up to 50.

2.2   Materials & Methods

The study design is illustrated in Figure 2.1. Image reconstruction from undersampled dynamic GA radial k-space was performed with either a classical non-uniform fast Fourier transform (NUFFT)[61] or with dAUTOMAP[109], a convolutional neural network designed for image reconstruction. Subsequently, motion is estimated on the reconstructed images via a classical optical flow (OF) based motion estimation algorithm, or a modified version of SPyNET, a multi-resolution layered deep network that computes deformation vector fields (DVFs) at multiple resolutions, similar to OF[106]. This allowed us to compare four approaches using varying degrees of DL to estimate motion from undersampled dynamic GA radial k-space.

Parameter	Sagittal	Coronal
TE (ms)	1.4	1.4
TR (ms)	2.8	2.7
Flip angle	50°	50°
Resolution (mm^2)	1.4×1.4	2.0×2.0
FOV (mm^2)	320×320	450×450
Reconstruction resolution (px^2)	224×224	224×224
Slice thickness (mm)	7	7
Readout direction	FH	FH
Bandwidth (Hz/px)	724 - 2034	1431 - 2034
Temporal resolution (ms)	500-570	500-570
Number of dynamics	50-300	100-300

Table 2.1 Scan parameters for sagittal and coronal 2D Cartesian balanced steady-state free precession MRI used in this work.

2.2.1 Patient data collection

Patients diagnosed with cancer in the abdomen undergoing radiotherapy simulation at our department between June 2015 and December 2019 were included in this study when sagittal cine MRI were acquired. In total, 135 patients were included, of whom 83 were male and 52 were female and were diagnosed with tumors to the abdomen (7), liver (40), kidneys (62) and pancreas (26). The patients were between 37 and 89 years old with a mean age of 67 ± 11 years old. Two-dimensional (2D) Cartesian balanced steady-state free precession (bSSFP) cine MRIs were acquired on a 1.5T MRI scanner (Ingenia MR-RT, Philips, Best, the Netherlands). Table 2.1 lists the acquisition parameters. The total acquisition time was between 25 s and 2.5 min, according to the number of dynamics acquired per scan, which varied between 50 and 300. Patients were scanned on a flat tabletop in the supine position using a 16-channel anterior and a 12-channel posterior phased-array coil. Two in-house built coil bridges supported the anterior coil to avoid skin contour deformation and not to affect natural motion. In total, 31750 magnitude-only dynamics were collected from 200 cine MRIs, as for some patients the cine data were acquired multiple times. Of these 200 cine MRIs, 126 were scanned after contrast agent injection.

For 30 of the 135 patients, 42 coronal cine MRIs were also acquired. Coronal cine MRIs were used for model validation. The scan parameters of these cine MRIs are also detailed in Table 2.1.

2.2.2 Data preparation

The signal intensity over all dynamics was linearly rescaled to an output range of [0, 1], clipping to the 99th percentile of intensity values of the dynamics in a cine MRI. Complex k-space was obtained by adding simulated phase to

the magnitude-only images and computing the non-uniform Fourier transform (NUFFT)[61] using PyNUFFT version 2019.1.1[110] with an undersampled GA radial readout trajectory. The simulated phase was generated per dynamic, as suggested by [91], i.e. by generating two two-dimensional sinusoids with a randomly-chosen spatial frequency between 0.05 Hz and 0.25 Hz and rotating these sinusoids separately with a random angle around the origin. These sinusoids were added together and the amplitude normalized to $[-\pi, \pi]$ such that the intensity represents phase values. K-space was density-compensated with a Ram-Lak filter and gridded to a Cartesian grid.

To ensure that representative noise was present in the retrospectively undersampled k-space, additional Gaussian noise $X \sim \mathcal{N}\left(0, \epsilon \cdot |k_0|\right)$ was added separately to the real and imaginary channels, where ϵ was randomly chosen between $\left[3 \cdot 10^{-3}, 5 \cdot 10^{-3}\right)$. The range for ϵ was determined from separate noise scans as the magnitude of the noise divided by the magnitude of the DC component.

The undersampling factor $\mathcal{R}$ was determined by dividing the number of spokes required for a Nyquist-sampled radial acquisition at the reconstruction resolution by the undersampling factor, i.e. $\lceil 224 \cdot \pi/2 \rceil \cdot \mathcal{R}^{-1} = 352 \cdot \mathcal{R}^{-1}$. The induced latency of this acquisition scheme is half of the acquisition time, i.e. $352 \cdot \mathcal{R}^{-1} \cdot \text{TR}/2$[101]. Data was prepared for the undersampling factors $\mathcal{R} = 1, 5, 10, 16, 20, 25, 30, 40,$ and 50.

2.2.3 Image reconstruction

The generated k-space of each dynamic was reconstructed with a conventional method and a DL-based approach.

2.2.3.1 Conventional

Non-Cartesian k-space was reconstructed with a NUFFT adjoint reconstruction, obtaining a fast reconstruction at the cost of undersampling artifacts compared to an iterative reconstruction algorithm.

2.2.3.2 Deep learning

For image reconstruction from undersampled k-space dAUTOMAP[1][109] was trained on a GPU (Tesla P100, NVIDIA, Santa Clara, CA, USA). dAUTOMAP is a model that performs non-iterative reconstruction with low parameter count, which makes it suitable for real-time image reconstruction. As dAUTOMAP assumes that the k-space points lie on a Cartesian grid, the k-space was re-gridded and density-compensated, as illustrated in Figure 2.2 (top). The

---

1. Reference implementation as found on `https://github.com/js3611/dAUTOMAP`
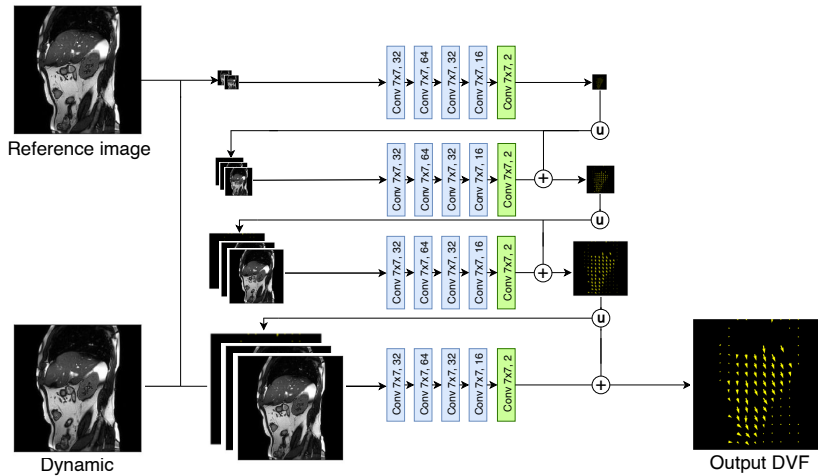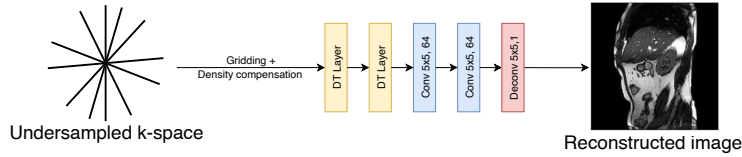
Chapter 2

Figure 2.2 Schematic of the image reconstruction and motion estimation models. The dAUTOMAP model (top) reconstructs the re-gridded and density-compensated undersampled k-space to an image. SPyNET (bottom) is a multi-resolution approach that estimates a DVF between a reference image and dynamic using multiple CNNs. Blue and green layers are two-dimensional convolution layers with and without non-linear activation, respectively.

model was implemented in PyTorch 1.0.1 and had 913473 trainable parameters. dAUTOMAP was initialized using Xavier initialization[111] and trained using the Adam optimizer[76] using $\beta_1 = 0.9, \beta_2 = 0.999$ and a learning rate of 10^{-3} with a batch size of 64 on an undersampled k-space with $\mathcal{R} = 10$ to minimize the mean-square-error (MSE) between reconstruction and target. After 50 epochs, the high-frequency error norm (HFEN)[112] was added to the loss function as it was found to improve performance. dAUTOMAP was trained until validation loss converged. $\mathcal{R} = 10$ was chosen as the undersampling factor for training as a balance between a fast acquisition and image quality, as training with higher undersampling factors became unstable. The learning rate was halved if the validation error plateaued, i.e. if the validation error has not improved with at least 10^{-8} in the last ten epochs. dAUTOMAP was trained on 119 cine MRIs from 81 patients comprising 60% of all sagittal dynamics. The hyper-parameters were validated on 38 cine MRIs from 26 patients comprising 20% of all sagittal

dynamics. The final model was tested on 43 cine MRIs from 28 patients comprising 20% of all sagittal dynamics.

2.2.4 Motion estimation

For every sagittal cine MRI, a reference image was chosen by randomly selecting a dynamic after ensuring that the dynamic was acquired in the steady-state. This was ensured by excluding the first 30 dynamics of the cine MRI from the selection of reference images. Then, DVFs were computed between every dynamic and the reference image.

Five reference images were randomly selected per cine MRI as data augmentation strategy and to ensure that the reference images were not only on an "extreme" point of the respiratory phase, e.g. inspiration or expiration.

This yielded a total of 130475 DVFs for training and validation and 28275 DVFs for testing.

2.2.4.1 Conventional

DVFs were computed using optical flow[39,40,113].

Optical flow is a registration algorithm that assumes the DVF to be smooth and the brightness of the images is preserved over time. Optical flow estimates DVFs by minimizing the energy function given in Equation 2.1:

$$E = \iint_\Omega |I_x u + I_y v + I_t| + \beta^2 \left(||\nabla u||_2^2 + ||\nabla v||_2^2 \right) \mathrm{d}x \mathrm{d}y \qquad (2.1)$$

where $\Omega \subseteq \mathbb{R}^2$ is the image domain, u and v are components of the DVF, I_x, I_y, I_t are the spatial and temporal partial derivatives of the images, respectively, and β is the regularization parameter enforcing smoothness.

Optical flow refines the motion estimate through iteration and estimating motion at multiple resolution levels in a pyramid approach in order to resolve large displacements.

In a preliminary study that is presented in 2.6, we compared an implementation of optical flow and Elastix[42] to assess the registration performance on our dataset.

As a result of this preliminary study, we opted to use optical flow as implemented with RealTITracker[40,113] in this work. In particular, ground-truth DVFs were computed on the fully-sampled dynamics by computing optical flow between every dynamic/reference image pair with RealTITracker with $\beta = 0.6$.

2.2.4.2 Deep learning

For motion estimation, the convolutional neural network called SPyNET[106] was trained on a GPU (Tesla P100, NVIDIA, Santa Clara, CA, USA). SPyNET is a multi-resolution pyramid approach. At every resolution level in the pyramid, a small CNN of 233778 parameters is employed to estimate motion from the input images together with an upsampled motion estimate from the previous pyramid level. The motion estimation approach is illustrated in Figure 2.2 (bottom). The model was implemented in PyTorch 1.0.1 and was serially trained with four pyramid levels, for a total of 935112 trainable parameters. The image pyramid had an image size of $224{\times}224$ pixels at the highest resolution level down to $28{\times}28$ pixels at the lowest resolution level. SPyNET was trained separately on pairs of images reconstructed with either a NUFFT or dAUTOMAP reconstruction with $\mathcal{R} = 10$ to learn the ground-truth optical flow DVFs by minimizing the end-point-error $\left(\text{EPE} = \sqrt{(u_{\text{est}} - u_{\text{gt}})^2 + (v_{\text{est}} - v_{\text{gt}})^2}\right)$. The model weights of all networks were initialized using Kaiming uniform initialization[114].

The effect of the warping operator as defined in the original implementation of SPyNET, which registers the images at lower resolution levels to resolve larger displacements, was evaluated and was found to be detrimental to the motion estimation quality and therefore omitted.

Data augmentation was performed on the images consistent with the ground-truth DVF by random horizontal and vertical flips and contrast jitter to prevent overfitting[115]. The EPE was minimized using the Adam optimizer $\beta_1 = 0.9, \beta_2 = 0.999$ with a learning rate of $5 \cdot 10^{-4}$ until convergence of the validation loss. The batch size was limited by the available GPU memory and was 1024 for the lowest resolution level, and 32 for the highest resolution level.

Every SPyNET level was trained, tested, and validated on the same data partition as dAUTOMAP. That is, 119 cine MRIs from 81 patients comprising 60% of all sagittal dynamics were used for training. The hyper-parameters were validated on 38 cine MRIs from 26 patients comprising 20% of all sagittal dynamics. The final model was tested on 43 cine MRIs from 28 patients comprising 20% of all sagittal dynamics.

2.2.5 Experiment setup

As image reconstruction and motion estimation can be computed with conventional or DL-based methods, we investigated four different combinations to obtain DVFs from k-space:

- Using NUFFT reconstruction and optical flow motion estimation (NUFFT/OF);

- Using NUFFT reconstruction and SPyNET motion estimation (NUFFT/SPyNET);
- Using dAUTOMAP reconstruction and optical flow motion estimation (dAUTOMAP/OF);
- Using dAUTOMAP reconstruction and SPyNET motion estimation (dAUTOMAP/SPyNET).

As the goal of these methods is to estimate motion from undersampled k-space, quality is defined solely by the correctness of the DVF. The four approaches were evaluated using the following criteria:

Registration performance The image similarity after registration of fully-sampled dynamics using a DVF estimated on undersampled images was evaluated over the whole image. This was quantified by the structural similarity (SSIM)[116] over the whole image. In particular, the mean ($\pm$ std) of the SSIM after registration was computed for 100 dynamic/reference image pairs of each cine MRI for every approach. In total, a sample of 2975 dynamic/reference pairs were considered.

DVF quality The quality of the DVF was measured by the mean absolute displacement error, as well as the root-mean-square error (RMSE) compared to the ground truth in a region of interest (ROI) that was manually generated to include relevant structures, e.g. liver veins, kidney structures or tumors. The ROIs of all patients in the test set are presented in 2.6. The root-mean-square error of displacement within the ROIs was considered as well. Bland-Altman plots[117] of the mean absolute displacement error were calculated to compare the average DVF magnitude within an ROI to the ground-truth optical flow. These plots reveal the bias of a model for undersampled motion estimation in the generated DVFs, computing statistical error bounds. The statistical significance was estimated using the Wilcoxon signed-rank test.

Time The time necessary to estimate motion, including MR acquisition, was reported. For a fair comparison of the different approaches, only GPU timings were considered. Given that RealTITracker, the optical flow implementation that we adopted, is available only for CPUs, we obtained the timing of conventional motion estimation using a CUDA implementation of optical flow that is part of the OpenCV library [2]. Note that such implementation uses a different algorithm[118] than the optical flow implementation used to generate ground-truth data.

All the metrics were computed on the test set, consisting of 28275 sagittal image pairs as well as 27900 coronal image pairs, for undersampling factors $\mathcal{R} = 1, 5,$

---

2. https://github.com/NeerajGulia/python-opencv-cuda

10, 16, 20, 25, 30, 40, and 50 without retraining of the DL models, which were trained on $\mathcal{R} = 10$.

2.3 Results

dAUTOMAP was trained on $\mathcal{R} = 10$ for 300 epochs in approximately six hours. After training, inference of the model to reconstruct a dynamic from gridded k-space was performed in 5 ms, making it as fast as NUFFT adjoint reconstruction. Examples of NUFFT and dAUTOMAP reconstructions are shown in Figure 2.3d and Figure 2.3g, respectively. It can be observed that NUFFT reconstructions at $\mathcal{R} = 20$ suffer from considerable streaking artifacts and dAUTOMAP reconstructions are overly smoothed with intensity patches, as highlighted by the red arrows. Every SPyNET level was trained $\mathcal{R} = 10$ for 12 hours until the validation error converged which took between 200 and 1000 epochs, depending on the resolution level. After training, inference of the four-level pyramid including resizing the input images and upsampling the intermediate DVFs was performed in 15 ms, which is slower than a GPU optical flow implementation that estimates motion in 5 ms. Example DVFs estimated by SPyNET are shown in Figure 2.3f and Figure 2.3i, on NUFFT and dAUTOMAP reconstructions, respectively. Example DVFs estimated by optical flow are shown in Figure 2.3e and Figure 2.3h, on NUFFT and dAUTOMAP reconstructions, respectively. In the supplementary material, an animation of Figure 2.3 is reported. It can be observed that optical flow DVFs in the liver are comparable to the ground-truth, but in this case SPyNET is able to improve the motion estimate in the spine, which seems more physiologically plausible than for optical flow.

2.3.1 Registration performance

Using the DVFs as generated by the four proposed methods to register the fully-sampled dynamics, the SSIM quantifies the registration performance across the entire image. Figure 2.4 shows the SSIM as a function of the undersampling factor. DVFs generated by SPyNET lead to a significantly higher SSIM after registration compared to optical flow for $\mathcal{R} > 10$ (Wilcoxon, $p < 0.001$), even though the models were trained at $\mathcal{R} = 10$. At $\mathcal{R} = 30$ an average SSIM of 0.8 is achieved using NUFFT/SPyNET, whereas using NUFFT/optical flow results in an average SSIM of 0.72. Interestingly, Using SPyNET with NUFFT reconstruction shows a similar performance when evaluated on coronal acquisitions even though SPyNET was trained on sagittal dynamics, as presented in Figure 2.4. Using dAUTOMAP for image reconstruction results in a 5-25% drop in performance when registering coronal images compared to sagittal images depending on the undersampling factor.
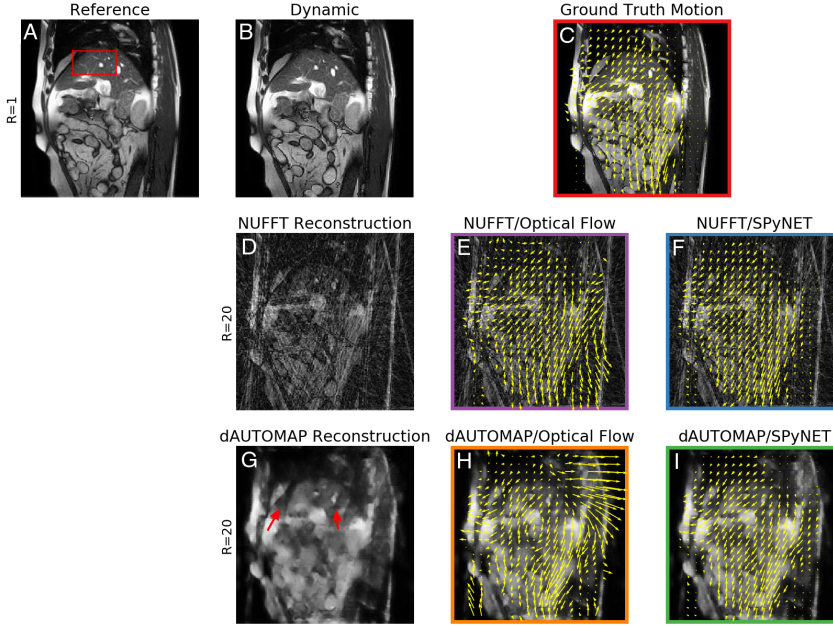
Figure 2.3 Example of a dynamic with image reconstruction and motion estimation. Figure 2.3a and Figure 2.3b show the fully-sampled sagittal reference image with region-of-interest in the red box and dynamic, respectively. The corresponding ground-truth DVF is shown in Figure 2.3c. Figure 2.3d shows the NUFFT adjoint reconstruction of the 20-fold retrospectively undersampled dynamic. The DVFs computed with the optical flow or SPyNET with adjoint reconstructions are shown in Figure 2.3e and f, respectively. Figure 2.3g, h, and i show the same as Figure 2.3d, e, and f, respectively but using dAUTOMAP for image reconstruction instead of a NUFFT adjoint. The arrows in Figure 2.3g indicate pseudo-random intensity patches introduced by dAUTOMAP.

2.3.2 DVF quality

The root-mean-square displacement error of the DVF generated with conventional methods compared to the ground-truth within an ROI on sagittal images significantly increases for acceleration factors $\mathcal{R} \geq 20$, as presented in Figure 2.5. For the NUFFT/SPyNET approach, the RMSE shows a slower rise as the undersampling factor increases, indicating robustness to undersampling artifacts. For NUFFT/SPyNET the root-mean-square displacement is lowest among all approaches at high undersampling factors ($\mathcal{R} \geq 20$) and remains within 1 mm with a narrower standard deviation, even for $\mathcal{R} = 30$.

Figure 2.6 reports Bland-Altman plots of the mean absolute displacement error within an ROI compared to the ground-truth on sagittal images. At $\mathcal{R} = 10$, there is no clear improvement of using DL rather than conventional methods. The mean
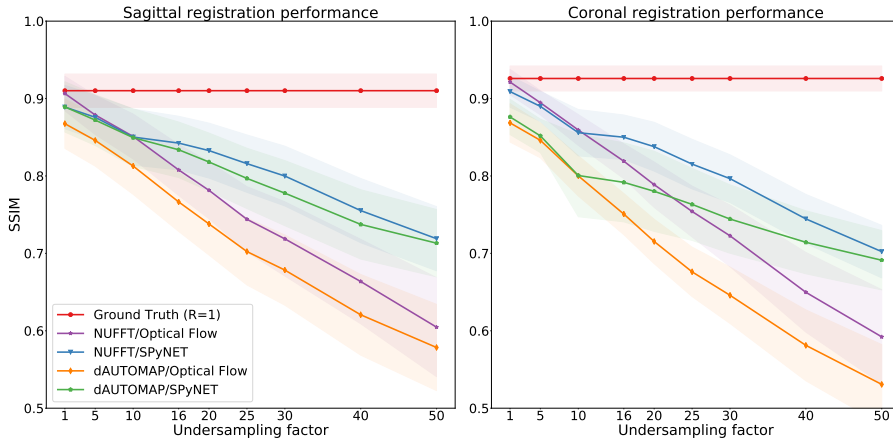
Figure 2.4 Comparison of the SSIM after registration over the whole image for sagittal images (left) and coronal images (right). Shaded regions indicate standard deviation.

difference is zero for the fully conventional method and has standard deviations within 0.95 mm, compared to a bias of -0.28 mm and a standard deviation up to 1.6 mm for dAUTOMAP/SPyNET. However, at $\mathcal{R} = 25$ the smallest error is obtained when using NUFFT reconstruction with SPyNET motion estimation as the bias is reduced to -0.1 mm and the standard deviation of the absolute error remains within 2 mm, compared to a standard deviation up to 3.5 mm for NUFFT in combination with optical flow.

2.3.3 Time

At $\mathcal{R} = 25$, approximately 40 ms would be spent acquiring k-space of a single dynamic with TR=2.8 ms. Combined with a NUFFT adjoint reconstruction which takes 5 ms and a SPyNET forward evaluation of 15 ms, DVFs can be computed with high quality in 60 ms, which is more than adequate for real-time MRIgRT of respiratory induced moving targets.

Table 2.2 summarizes all quantitative results in the sagittal plane. It can be observed that almost 94% of all vectors have a root-mean-square error of less than 2 mm when computed with a NUFFT adjoint reconstruction and SPyNET for motion estimation.

27

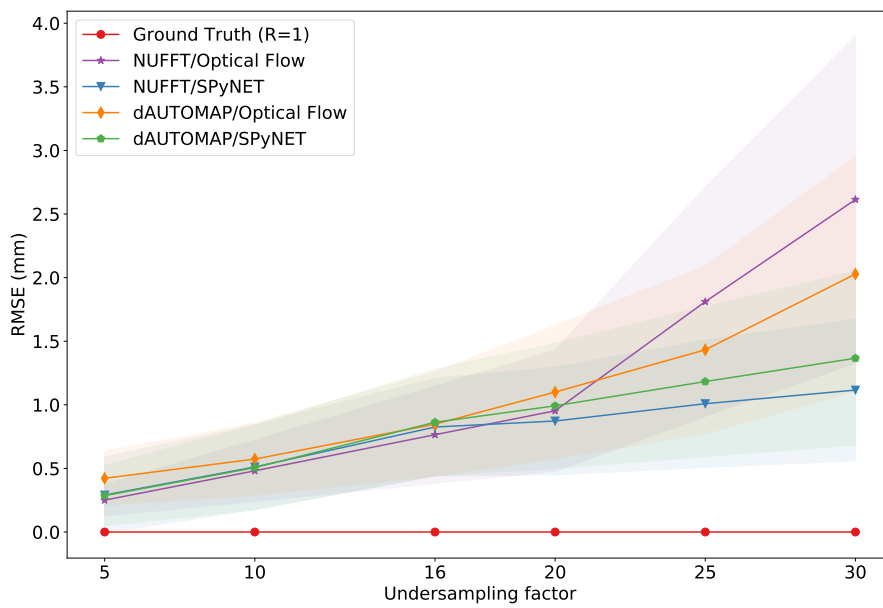**Figure 2.5** Root-mean-square displacement error within an ROI on sagittal images. Shaded regions indicate standard deviation.
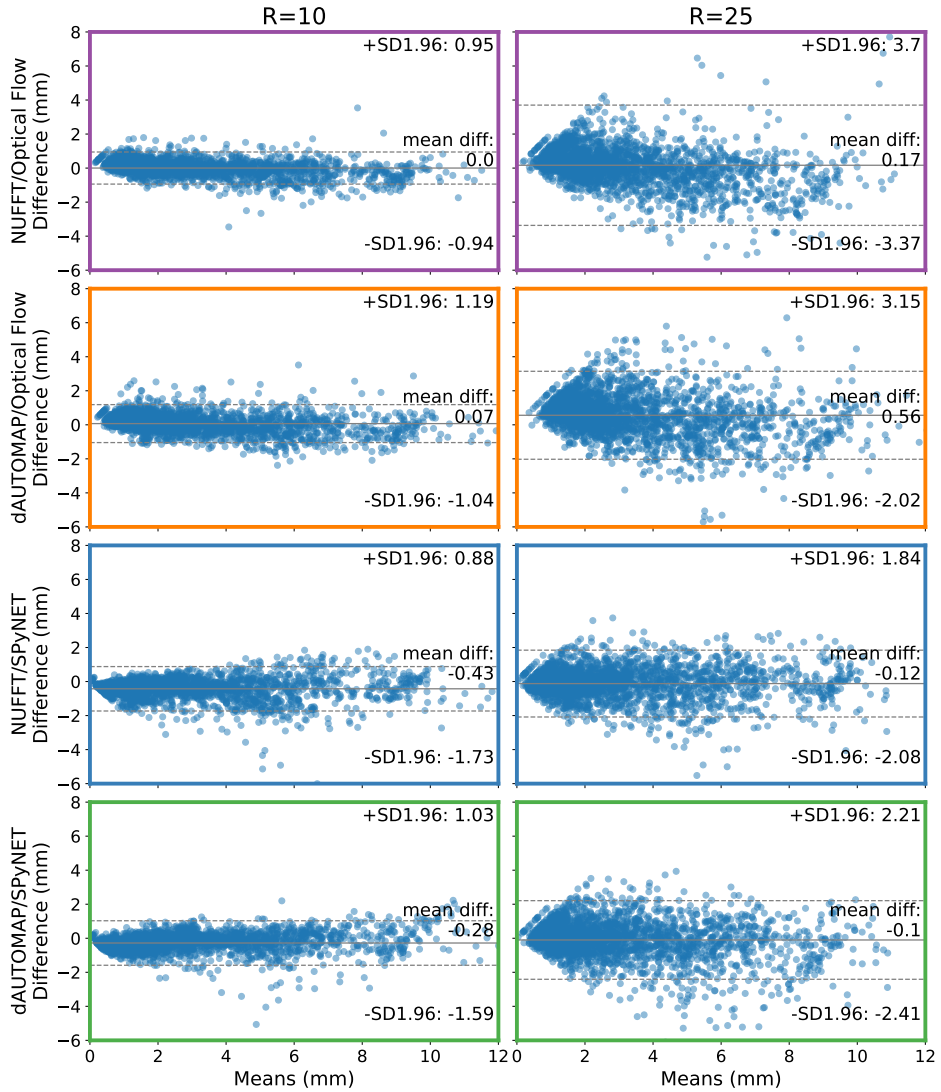
Figure 2.6 Bland-Altman plots of the average vector magnitude within an ROI on sagittal images as generated by the various model configurations at $\mathcal{R} = 10$ and $\mathcal{R} = 25$ compared to the ground-truth. A positive value indicates an overestimation compared to the ground-truth.

	Ground Truth (R=1)	NUFFT Optical Flow	NUFFT SPyNET	dAUTOMAP Optical Flow	dAUTOMAP SPyNET
		SSIM after registration			
$\mathcal{R} = 1$	0.91 ± 0.04	$\mathbf{0.91 \pm 0.04}$	0.89 ± 0.06	0.87 ± 0.06	0.89 ± 0.06
$\mathcal{R} = 5$	0.91 ± 0.04	$\mathbf{0.88 \pm 0.05}$	0.88 ± 0.06	0.85 ± 0.07	0.87 ± 0.07
$\mathcal{R} = 10$	0.91 ± 0.04	$\mathbf{0.85 \pm 0.06}$	0.85 ± 0.07	0.81 ± 0.07	0.85 ± 0.07
$\mathcal{R} = 16$	0.91 ± 0.04	0.81 ± 0.07	$\mathbf{0.84 \pm 0.07}$	0.77 ± 0.08	0.83 ± 0.07
$\mathcal{R} = 20$	0.91 ± 0.04	0.78 ± 0.07	$\mathbf{0.83 \pm 0.07}$	0.74 ± 0.08	0.82 ± 0.07
$\mathcal{R} = 25$	0.91 ± 0.04	0.74 ± 0.08	$\mathbf{0.82 \pm 0.07}$	0.70 ± 0.09	0.80 ± 0.08
$\mathcal{R} = 30$	0.91 ± 0.04	0.72 ± 0.10	$\mathbf{0.80 \pm 0.08}$	0.68 ± 0.09	0.78 ± 0.08
$\mathcal{R} = 40$	0.91 ± 0.04	0.66 ± 0.11	$\mathbf{0.76 \pm 0.08}$	0.62 ± 0.10	0.74 ± 0.09
$\mathcal{R} = 50$	0.91 ± 0.04	0.60 ± 0.13	$\mathbf{0.72 \pm 0.08}$	0.58 ± 0.11	0.71 ± 0.09
		RMSE ≤ 1 mm (within ROI)			
$\mathcal{R} = 5$	100%	$\mathbf{99.5\%}$	95.5%	98.0%	95.5%
$\mathcal{R} = 10$	100%	$\mathbf{95.3\%}$	92.3%	92.8%	92.9%
$\mathcal{R} = 16$	100%	86.1%	86.1%	81.4%	$\mathbf{88.5\%}$
$\mathcal{R} = 20$	100%	78.1%	$\mathbf{83.4\%}$	70.1%	80.6%
$\mathcal{R} = 25$	100%	69.9%	$\mathbf{76.6\%}$	56.0%	71.1%
$\mathcal{R} = 30$	100%	61.8%	$\mathbf{70.3\%}$	48.1%	64.6%
		RMSE ≤ 2 mm (within ROI)			
$\mathcal{R} = 5$	100%	$\mathbf{100.0\%}$	99.2%	100.0%	99.1%
$\mathcal{R} = 10$	100%	$\mathbf{99.7\%}$	98.8%	99.2%	98.8%
$\mathcal{R} = 16$	100%	97.8%	97.7%	96.8%	$\mathbf{98.1\%}$
$\mathcal{R} = 20$	100%	95.6%	$\mathbf{96.8\%}$	93.9%	96.0%
$\mathcal{R} = 25$	100%	91.7%	$\mathbf{94.8\%}$	86.7%	91.8%
$\mathcal{R} = 30$	100%	85.8%	$\mathbf{93.9\%}$	80.1%	90.1%
		Time (acquisition/reconstruction/motion (ms))			
$\mathcal{R} = 10$	500/1/5	$\mathbf{100/5/5}$	100/5/15	$\mathbf{100/5/5}$	100/5/15
$\mathcal{R} = 20$	500/1/5	$\mathbf{50/5/5}$	50/5/15	$\mathbf{50/5/5}$	50/5/15
$\mathcal{R} = 25$	500/1/5	$\mathbf{40/5/5}$	40/5/15	$\mathbf{40/5/5}$	40/5/15

Table 2.2 Quantitative results for the four approaches in the sagittal plane, displaying the structural similarity index (SSIM) after registration for various undersampling factors, the root-mean-square error (RMSE) of the motion magnitude within an ROI (ROIs displayed in 2.6), and the time it takes for MRI acquisition, image reconstruction, and motion estimation. Best results per metric per undersampling factor are marked in boldface, excluding ground-truth.

2.4 Discussion

In this work, we have investigated the impact of conventional and DL-based approaches to estimate 2D DVFs from highly undersampled k-space for real-time MRIgRT applications. In particular, we have quantified how much specific deep learning models can accelerate MRI acquisition and processing over conventional techniques and in which step deep learning is beneficial to obtaining high-quality motion estimates. We have shown that motion can be estimated from heavily undersampled k-space with high temporal resolution and low error compared to the ground-truth when images are reconstructed with a conventional NUFFT and motion is estimated with deep learning. For example, the mean absolute displacement error remained within 2 mm and the RMSE remained within 1 mm

at $\mathcal{R} = 25$ while the SSIM after registration remained above 0.8 when motion is estimated with NUFFT adjoint image reconstruction and SPyNET is used. Our method can compute DVFs with these errors within 60 ms and induces a total latency of 40 ms of which 20 ms comes from MRI acquisition[101] and 20 ms comes from processing, but extra overhead may present itself in a prospective setting. This demonstrated that reconstruction of DVFs is feasible at very high undersampling factors despite severe artifacts in the reconstructed images, indicating that accurate motion estimation is more resilient to undersampling than high-quality image reconstruction.

Results show that using SPyNET for motion estimation rather than optical flow significantly improves DVF quality at undersampling factors $\mathcal{R} \geq 10$. Also, we observe that the best DL-based approach can achieve the same SSIM after registration as the fully conventional approach with approximately two times more undersampling.

Interestingly, applying SPyNET to NUFFT-reconstructed images also outperforms applying SPyNET to dAUTOMAP-reconstructed images. This indicates that general-purpose trained DL-based image reconstruction obtained with dAUTOMAP does not have added value for motion estimation. We observed that dAUTOMAP favored overly smoothed reconstructions at high undersampling factors. We hypothesize that this may be detrimental to recover motion information.

We believe we have designed a robust approach to motion estimation. Augmenting the input images with flips and rotations makes dAUTOMAP and SPyNET robust against slight angulations. Moreover, the NUFFT/SPyNET approach shows near-equivalent registration performance on coronal images compared to registration of sagittal images without retraining. When dAUTOMAP is used for image reconstruction, the performance is significantly lower on coronal images than on sagittal images as it fails to reconstruct high-quality coronal images when trained on sagittal images. Even though the networks were trained at $\mathcal{R} = 10$, evaluation at higher undersampling factors seems to have a low impact on the registration quality.

NUFFT/SPyNET is thus able to resolve incoherent streaking artifacts introduced by radial sampling. An interesting exploration would be to investigate whether other sampling strategies (e.g., variable-density spirals) achieve similar results, but this was considered out of the scope of this paper. This robustness of NUFFT/SPyNET could suggest that the model is well generalizable and might transfer to other body sites and contrasts without retraining, which is currently under investigation.

This method of a radial readout with NUFFT image reconstruction and SPyNET motion estimation could find its application in real-time MRI-guided radiotherapy

applications. Keall *et al.*[49] suggest that acquisition, motion estimation and dose delivery needs to happen within 200 milliseconds to maintain accuracy. By using NUFFT/SPyNET, accurate DVFs can be obtained at $\mathcal{R} = 25$ in 60 ms with a latency of 40 ms, including MR acquisition. This leaves ample time for adaptation of the radiation beam to counteract the motion. This could enable real-time tumor tracking to account for intra-fraction motion.

One of the limitations of our approach is that it requires a ground-truth motion estimate to learn. While computing a ground-truth is feasible for retrospectively undersampled data, obtaining a high-quality ground-truth motion estimate for prospectively undersampled in-vivo MR data is challenging. Prospective data will also be acquired with multiple receiver coils while this work is focused on single-coil images. Considering multi-coil images might be beneficial for motion estimation quality but also introduces new challenges. It requires more data needs to be evaluated, which might result in more parameters to train and higher inference times. Future work may investigate unsupervised approaches to learning motion or find another way to obtain motion estimates from k-space acquired with multiple receiver coils. Another limitation is that our networks were only trained at $\mathcal{R} = 10$. Performance might be improved at high undersampling factors if they are retrained at $\mathcal{R} > 10$.

When compared to other works, our method is significantly faster while achieving similar accuracy at $\mathcal{R} = 25$, even when compared to other deep learning-based methods[119–121]. Seegoolam *et al.*[119] investigated motion estimation on 2D cardiac cine MRI for $\mathcal{R} = 9$ and $\mathcal{R} = 50$ achieving an average SSIM after registration of 0.93 at $\mathcal{R} = 9$ versus 0.86 in this work and an SSIM of 0.776 at $\mathcal{R} = 51.2$ versus 0.72 in this work. Also, they indicate that the motion estimation network shows better generalization than the reconstruction network for various undersampling factors, which is in accordance with what we observed. However, their reconstruction method takes approximately 1.8 seconds per frame, excluding MR acquisition which is a significant performance penalty.

Stemkens *et al.*[120] obtained a 3D motion estimation with an RMSE of 1 mm using a 360 ms 2D acquisition and a few seconds of motion calculation. This error is comparable with we observed, even though their work estimates motion in three dimensions. This is, however, not a "full" 3D method but uses multi-2D cine scans in conjunction with a 4D MRI to obtain 3D motion estimates, limiting the accuracy of the method.

The approach by Haskell *et al.*[121] significantly reduces motion artifacts in image space by combining a CNN with a physics-based model. This approach of combining DL to remove artifacts with conventional SENSE reconstruction[67] produces the best results, which is in line with what we found. However, their approach requires fully-sampled data, and the full motion correction

model requires several minutes to evaluate, making it unsuitable for real-time applications.

In this work, we showed that acquisition, reconstruction and motion estimation can be performed in approximately 60 ms for $\mathcal{R} = 25$ achieving a root-mean-square displacement error of less than 1 millimeter compared to a ground-truth motion estimate. This is of particular interest for applications with crucial time constraints, such as MRIgRT[122]. We believe that deep learning models play an important role in facilitating real-time motion management on MR-Linacs, but should be carefully assessed, taking into account the entire feedback chain. Replacing an individual "classic" step in the processing pipeline by a DL alternative does not necessarily result in improved performance. We did show that using a DL-based motion estimation network in conjunction with a NUFFT yields a robust and generic method for motion estimation. The combination of highly undersampled k-space with DL-based methods yields high-quality motion estimation for a real-time MRIgRT with low latency, which makes it a worthwhile area of ongoing research.

In **Chapter 3**, we will investigate a "full" three-dimensional extension of this method for real-time motion estimation. We hypothesize this will have a higher accuracy and performance than a multi-2D approach. Motion has been successfully estimated from fully-sampled 3D MR cardiac images[123], but the method has not been demonstrated for real-time applications. We will investigate whether the use of multi-channel MRI may further improve the current performances.

2.5   Conclusions

The performance of DL-based image reconstruction and motion estimation was assessed on retrospectively undersampled GA radial MRI to allow real-time motion estimation with minimal latency. It was found that DL-based motion estimation (SPyNET) allowed far greater acceleration factors than traditional optical flow based motion estimation. DL-based image reconstruction of undersampled radial data, however, did not result in better performance compared to standard NUFFT reconstructions in combination with SPyNET motion estimation. The NUFFT/SPyNET approach produced an acceptable performance for 25-fold accelerated data, thereby achieving an imaging frame rate of 25 Hz while the root-mean-square error remained within 1 millimeter.

2.6   Acknowledgement

Appendix A: Choice of ground-truth motion estimation algorithm

High-quality ground-truth motion estimates are required for training SPyNET. To determine which motion estimation is best suited for the dataset used in this work, a preliminary study was conducted comparing the motion estimation quality of optical flow[40] and Elastix[42]. Both methods were compared and evaluated based on registration performance. This was measured by the structural similarity (SSIM) metric[116], the mean-squared-error (MSE) between the reference image and the registered image, and evaluation of the error spectrum plot (ESP)[124] of the registered images compared to the reference image were calculated over the entire image.

Optical flow and Elastix DVFs were computed for all fully-sampled cine MRIs in the training dataset used in this work. Optical flow was computed as described in section 2.2.4.1 with $\beta = 0.6$. Elastix DVFs[3] were computed on four resolution levels using rigid, affine, and deformable motion estimation using B-splines. For rigid and affine motion was estimated using the mutual information metric. For deformable motion estimation, mutual information was used with weight 1, and a transform bending energy penalty was added with weight 2. For every cine MRI, 100 dynamic/reference image pairs were randomly sampled to ensure representative measurements. The average SSIM, MSE, and ESP were computed over 8100 dynamic/reference image pairs.

It was found that optical flow yielded an average SSIM of 0.920 ± 0.045, which was significantly higher than the average SSIM of Elastix registrations 0.899 ± 0.053 (Wilcoxon, $p < 0.001$). The average MSE was 3.63 ± 1.86 for optical flow, which was significantly lower than the MSE of Elastix (Wilcoxon, $p < 0.001$), which was 5.08 ± 2.45. The averaged ESP is shown in Figure 2.7. It can be observed that for nearly all frequencies, the error of optical flow is lower than for Elastix, except for the very highest frequencies.

Based on these results have selected optical flow as ground-truth for evaluation and learning target for deep learning models.

Appendix B: ROIs of the validation set

---

3. The exact parameter files can be found here: `http://elastix.bigr.nl/wiki/index.php/Par0060`
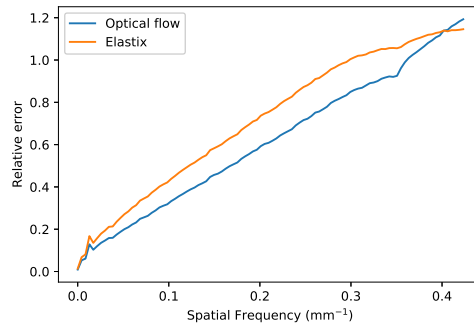
Figure 2.7 The average error spectrum plot of optical flow and Elastix registrations.
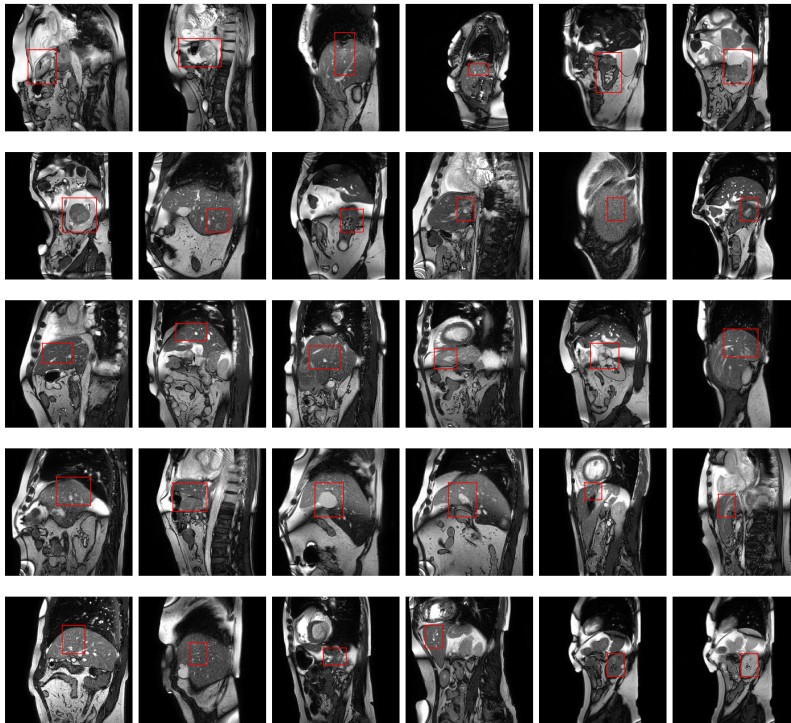
Figure 2.8 Manually generated regions-of-interest (ROIs) of the 30 patients used in the test set. These ROIs were used for the RMSE computation in Table 2.2 and the Bland-Altman plots in Figure 2.6. The ROIs were generated to include relevant structures and have an average size of 1010 ± 442 mm^2 or $4.1 \pm 1.8\%$ of the image.

CHAPTER 3

Real-time 3D motion estimation from undersampled MRI using multi-resolution neural networks

ML Terpstra
M Maspero
T Bruijnen
JJC Verhoeff
JJW Lagendijk
CAT van den Berg

Abstract

Purpose: To enable real-time adaptive MRI-guided radiotherapy (MRIgRT) by obtaining time-resolved 3D deformation vector fields (DVFs) with high spatio-temporal resolution and low latency (< 500 ms).

Theory & Methods: Respiratory-resolved T_1-weighted 4D-MRI of 27 patients with lung cancer were acquired using a golden-angle radial stack-of-stars readout. A multi-resolution convolutional neural network called TEMPEST was trained on up to 32x retrospectively undersampled MRI of 17 patients, reconstructed with a non-uniform fast Fourier transform, to learn optical flow DVFs. TEMPEST was validated using 4D respiratory-resolved MRI, a digital phantom, and a physical motion phantom. The time-resolved motion estimation was evaluated *in-vivo* using two volunteer scans, acquired on a hybrid MR-scanner with integrated linear accelerator. Finally, we evaluated the model robustness on a publicly-available 4D-CT dataset.

Results: TEMPEST produced accurate DVFs on respiratory-resolved MRI at twenty-fold acceleration, with the average end-point-error < 2 mm, both on respiratory-sorted MRI and on a digital phantom. TEMPEST estimated accurate time-resolved DVFs on MRI of a motion phantom, with an error < 2 mm at 28x undersampling. On two volunteer scans, TEMPEST accurately estimated motion compared to the self-navigation signal using 50 spokes per dynamic (366x undersampling). At this undersampling factor, DVFs were estimated within 200 ms, including MRI acquisition. On fully-sampled CT data, we achieved a target registration error of 1.87 ± 1.65 mm without retraining the model.

Conclusion: A CNN trained on undersampled MRI produced accurate 3D DVFs with high spatio-temporal resolution for MRIgRT.

3.1   Introduction

Real-time adaptive radiotherapy aims to increase the accuracy with which radiation is delivered, leading to increased sparing of healthy tissue and organs-at-risk (OARs)[45]. By rapidly acquiring images and estimating tumor motion during dose delivery, the radiation beam can be adapted to follow the current anatomy. To facilitate treatment adaptation, magnetic resonance imaging-guided radiotherapy (MRIgRT) is increasingly adopted in clinical practice, e.g. with the introduction of hybrid MR-Linac devices[25,30,33,99,125]. With its superior soft-tissue contrast, MRI facilitates direct visualization of tumors and OARs[17,126].

For *real-time* treatment adaptation, image acquisition and motion estimation must occur with low latency and a high spatio-temporal resolution[49], i.e. the maximum time between a (respiratory) motion event and dose delivery should be $\leq$ 400 ms[50,127]. However, real-time acquisition of 3D MRI and computation of a non-rigid deformation vector field (DVF) is challenging due to the long acquisition times of fully-sampled MRI (seconds to minutes) and the ill-posed and under-determined nature of motion estimation, hindering real-time motion estimation[39,128].

Several methods have been presented to accelerate MR acquisition and motion estimation, such as parallel imaging[67,68,102], simultaneous multi-slice acquisitions[129], advanced image reconstruction algorithms allowing for greater undersampling factors, such as compressed sensing[71], or novel motion estimation methods model from 2D MRI[120]. Recent works proposed using low-rank models to reconstruct highly undersampled MRI with sub-second temporal resolution[130], but these methods currently have long reconstruction times. Currently, none of these methods can achieve the required acceleration factor combined with low-latency reconstruction to estimate motion within 500 ms[127].

Recently, deep learning (DL) has been proposed to speed up MRI reconstruction and motion estimation, achieving performances on par, if not higher, than its non-DL counterparts[89,90,123,131–133]. Specifically, DL models allow for fast inference, leaving the time-consuming step to the training phase, which can take hours or days.

In a previous work, we introduced a supervised DL-based framework for real-time 2D motion estimation[134]. By reconstructing highly undersampled golden-angle radial acquisitions with a non-uniform fast Fourier transform (NUFFT), motion was estimated by a multi-resolution convolutional neural network (CNN), allowing for fast and accurate motion estimation.

Ideally, we could extend this approach to real-time three-dimensional motion estimation by training a 3D network on 3D cine-MRI acquired with high

spatio-temporal resolution. However, it is challenging to obtain high-quality ground-truth DVFs from in-vivo MRI acquired at a high spatio-temporal resolution as the images will suffer from severe artifacts due to undersampling and respiratory motion. One way to circumvent this problem is by performing *respiratory-sorted* image reconstruction instead of *time-resolved* image reconstruction. Respiratory-sorted MRI displays physiological motion similar to time-resolved MRI, maintaining higher image quality as fewer motion artifacts due to less severe undersampling.

In this work, we extend the previously-introduced 2D approach to 3D by training a DL model named TEMPEST (real-**t**ime 3D motion **e**stimation fro**m** undersa**m**ple**d** MRI using multi-re**s**olution neural ne**t**works) to estimate DVFs from highly accelerated 3D-MRI. We train TEMPEST on respiratory-sorted 4D-MRI to learn ground-truth DVFs computed using conventional registration methods. The trained network is subsequently used to estimate motion from highly accelerated time-resolved MRI. We investigate the optimal model hyperparameters, and evaluate the model performance on digital and physical phantoms and 4D respiratory-resolved CT data. Moreover, we estimate the performance of TEMPEST on time-resolved MRI of two healthy volunteers acquired on an MR-Linac.

3.2 Methods

We trained a supervised multi-resolution deep learning model (TEMPEST) to estimate a DVF ($DVF_{TEMPEST}$) between two undersampled MRI volumes acquired with a golden-angle radial stack-of-stars readout. The model requires MRI for training, together with ground-truth DVFs (DVF_{GT}) describing the motion between a dynamic and static volume.[1]

3.2.1 Patient data collection and preparation

Twenty-seven patients undergoing radiotherapy for lung cancer between February 2019 and February 2020 at the RT department were retrospectively included under the approval of the local medical ethical committee with protocol number 20-519/C.

Free-breathing 3D golden-angle radial stack-of-stars (GA-SOS) T_1-weighted spoiled gradient echo MRI of the thorax were acquired for 7 min on a 1.5T MRI (MR-RT Philips Healthcare, Best, the Netherlands) during gadolinium injection (Gadovist, 0.1 ml/kg). The acquisition was fat-suppressed using Spectral Attenuated Inversion Recovery (SPAIR). The relevant scan parameters are listed in Table 3.1 (4D MRI).

---

1. Code for training and inference will be made available at `https://gitlab.com/computational-imaging-lab/tempest` after acceptance.

| Parameter | 4D MRI | Phantom experiments | | Time-resolved MR-Linac experiments | |
		Stationary phantom	Moving phantom	Volunteer 1	Volunteer 2
Readout	GA-SOS	GA-SOS	GA-SOS	GA-SOS	Kooshball
Number of coils	28	8	8	8	8
TR/TE (ms)	3.2/1.3	4.3/1.8	3.4/1.5	3.4/1.5	3.5/1.4
Flip angle (°)	8	10	10	10	10
Bandwidth (Hz/px)	866	866	868	865	868
FOV (mm^3)	$440 \times 440 \times 270$	$440 \times 440 \times 270$	$440 \times 440 \times 270$	$525 \times 525 \times 270$	$525 \times 525 \times 525$
Resolution (mm^3)	$2.13 \times 2.13 \times 3.50$	$2.0 \times 2.0 \times 3.5$	$4.9 \times 4.9 \times 3.5$	$5.0 \times 5.0 \times 3.5$	$4.9 \times 4.9 \times 4.9$
Matrix size	$206 \times 206 \times 77$	$220 \times 220 \times 77$	$90 \times 90 \times 77$	$106 \times 106 \times 77$	$108 \times 108 \times 108$
Slice direction	FH	FH	FH	FH	-
Scan time (s)	438	60	28	163	40

Table 3.1 Relevant scan parameters for three experiments. The "4D MRI" column describes the MR parameters for the respiratory-resolved 4D-MRI used for training, validation and testing. The "Phantom experiments" columns refer to the two experiments acquiring a stationary and moving motion phantom on an MR-linac. The "Time-resolved MRI experiments" columns refer to MRI data acquired on an MR-linac of two healthy volunteers for evaluation of real-time motion estimation performance.

Patients were scanned in supine position using a 16-channel anterior and 12-channel posterior phased-array coil. In total, 1312 radial spokes per slice were acquired, corresponding to approximately four times oversampling compared to a fully-sampled volume, which requires $206 \cdot \pi/2 \approx 324$ spokes. To train and evaluate the motion estimation model, patients were divided in a train set (17 patients), validation set (5 patients) to find optimal hyperparameters and prevent overfitting, and a test set (5 patients) to evaluate the final model performance.

3.2.2 Image reconstruction

To train TEMPEST with physiological motion, we reconstructed respiratory-resolved MRI based on the self-navigation signal present in the 4D-MRI[135], as illustrated in Figure 3.1A. An example of respiratory-resolved reconstruction versus free-breathing image reconstruction is shown in Figure 3.1B,C,E,F. The self-navigation signal was obtained by sampling radial spokes and performing a 1D Fourier transform of center of k-space, i.e. k_0, along the slice direction. The respiratory motion surrogate was obtained by performing principal component analysis on the concatenated navigators[136,137].

As contrast was injected, the relative magnitude of the self-navigation signal changed over time. To account for the contrast wash-in phase, we discarded the first 200 spokes of every scan to prevent contrast mixing. The remaining spokes were sorted based on the respiratory phase and relative amplitude using a hybrid binning algorithm[138].

After sorting, k-space was density-compensated using a Ram-Lak filter, interpolated onto a twice-oversampled Cartesian grid using a 3×3
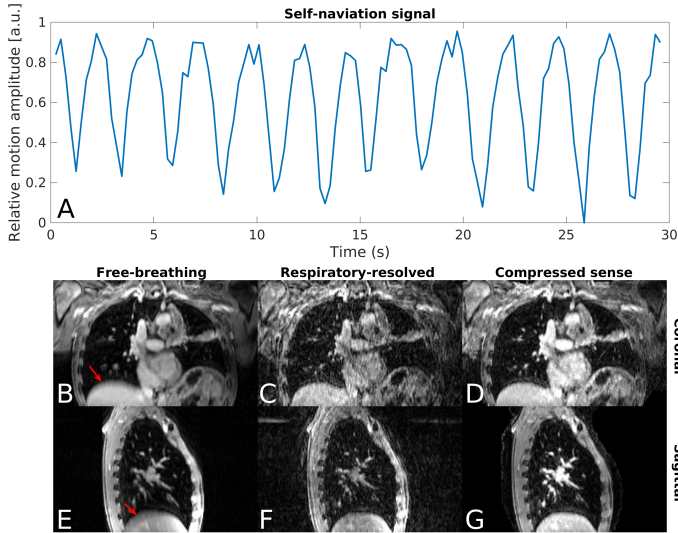
Figure 3.1 Examples of the self-navigation signal and the data. A typical example of the self-navigation signal during the first 30 seconds of the acquisition (A). In a free-breathing reconstruction (B,E), blurring due to motion can be observed near the red arrows. With a respiratory-resolved NUFFT reconstruction (C,F), the blurring is resolved at the cost of undersampling artifacts. Compressed-sense reconstructions (D,G), show improved image quality at the cost of longer reconstruction times.

Kaiser-Bessel kernel, and transformed to image-space with a NUFFT-adjoint reconstruction[61,139] with a weighted coil combination. Four-dimensional respiratory-resolved magnitude reconstructions were made for 1, 3, 5, 10, 20, 30, 40, 50, 75, and 100 respiratory phases. As 1312 spokes were sampled in total, and 324 sampled spokes are required to fulfill the Nyquist criterion, the undersampling factor of the respiratory-resolved MRI is computed as $\mathcal{R} = (324 \cdot n_{\text{phases}})/1312$, corresponding to approximately 0.25, 0.75, 1, 3, 5, 10, 13, 15, 18, and 27-fold undersampling, respectively. As we aimed to train a multi-resolution motion estimation model, we also reconstructed images at a lower spatial resolution, i.e. 2x and 4x spatial downsampling, by radially cropping the k-space around k_0, reducing the spatial resolution in the left-right and anterior-posterior direction. Along the feet-head (Cartesian) direction, resolution was maintained. The reconstructed images were normalized by scaling the image intensity to an output range of [0, 1] by clipping to the 99th percentile of the image intensity. The percentiles were computed on a patient basis over all respiratory phases.

To validate TEMPEST at high undersampling factors, i.e. $\mathcal{R} = 10, 13, 15, 18, 27$, we required ground-truth DVFs for comparison. However, traditional methods were unable to provide accurate DVFs based on the adjoint-reconstructed

images due to the undersampling artifacts. Therefore, MRI was also reconstructed using compressed sense with temporal total variation (TV) regularization, $\lambda = 0.03$[71,137]. An example of these reconstructions is shown in Figure 3.1D,G.

3.2.3 Ground-truth motion

Ground-truth DVFs were computed using optical flow[40,113], as it provided a good balance between computation time, registration performance and number of hyperparameters. Optical flow computes motion by assuming spatial smoothness of the DVF, regularized by the β hyperparameter. A preliminary study, which is presented in Supporting Information Figure 4, was performed to select the optimal value for $\beta = 0.4$ for our training data.

We calculated optical flow DVFs ($\mathrm{DVF_{OF}}$) for each respiratory-resolved dynamic to three static volumes: full inhale, full exhale, and halfway inhale and exhale. This increased the training data and to ensure that the network learned to compute motion in multiple principal directions.

Optical flow was computed up to 20 respiratory phases (i.e. $R \approx 7$) at full resolution. For $\mathcal{R} > 10$ optical flow DVFs ($\mathrm{DVF_{CS,OF}}$) were computed on the compressed sense reconstructions as the motion estimate became unreliable due to the artifacts present in the undersampled NUFFT-adjoint reconstructed MRI.

3.2.4 Network architecture

TEMPEST was designed as a multi-resolution 3D convolutional neural network (CNN) operating on the entire volume to learn the DVF between a static and dynamic image. The complete motion model consisted of $\mathcal{K}$ multi-resolution motion networks, operating on different spatial resolution levels. Each motion network had a fixed architecture and consisted of five 3D convolution layers with 32, 64, 32, 16, and 3 filters of size $c_k \times c_k \times c_k$, respectively. The motion network that operated at the lowest resolution directly attempted to learn a DVF from a static and dynamic volume. The motion network that operated at higher resolution levels received a static volume, dynamic volume, and upsampled DVF from the previous motion network as input, and attempted to learn a residual DVF to refine the upsampled DVF from the previous motion network. Several works perform intermediate warping of the dynamic images according to the estimated DVF[106,140,141]. In a previous work[134], we identified that warping undersampled images using the estimated DVF was detrimental to the motion estimation performance. As performing three-dimensional image interpolation is a costly operation, we opted to omit this operation. Each convolution layer in a motion network, except for the final layer, was followed by a ReLU non-linear activation function[142]. Figure 3.2 depicts the model architecture.
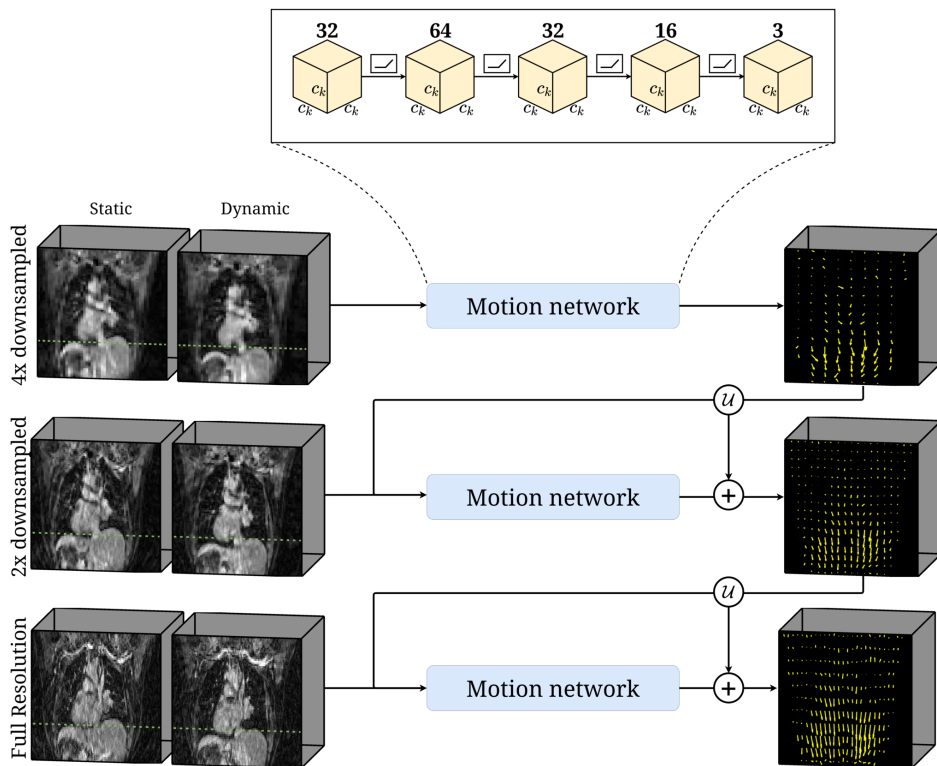
Figure 3.2 Model overview. TEMPEST computes the DVF between a static and dynamic volume, where the green line indicates the motion magnitude. TEMPEST starts at 4x spatially downsampled resolution. A motion network, consisting of a 3D CNN of five layers (32, 64, 32, 16, and 3 learned filters respectively) which operates on the whole volume, estimates the DVF between a static and a dynamic volume. This first motion estimate is upsampled through $\mathcal{U}$ and serves as additional input for the motion network operating at the next resolution level. The subsequent layers learn the residual DVF which improves the previous estimate. The size of all convolution kernels is $c_k \times c_k \times c_k$, depending on the resolution level. All layers but the last are followed by a ReLU non-linear activation.

TEMPEST was trained to minimize the end-point-error ($\text{EPE} = ||\text{DVF}_{\text{GT}} - \text{DVF}_{\text{TEMPEST}}||_2$) by considering the magnitude error and angle error as separated terms and penalizing non-smooth DVFs. The full loss function that was minimized during training was

$$\mathcal{L} = \alpha \cdot (\mathcal{L}_{\text{mag}} + \mathcal{L}_\phi) + (1 - \alpha) \cdot \mathcal{L}_{\text{EPE}} + \lambda \cdot \nabla\text{DVF}$$

where $\mathcal{L}_{\text{mag}}$ is the ℓ_2-norm of the magnitude difference between the target and output DVF, $\mathcal{L}_\phi$ is the ℓ_2-norm of the difference in angle between the target and output DVF, $\mathcal{L}_{\text{EPE}}$ is the EPE i.e., the ℓ_2-norm of the difference of the output DVF and target DVF, and ∇DVF enforces smoothness of the DVF by penalizing the mean Laplacian of the DVF. For our experiments, we used $\lambda = 10^{-5}$.

The motion networks were trained sequentially, starting at the lowest resolution level. When the validation loss of this network converged, the motion network operating at the next-higher resolution level was trained. At that point, two training strategies were considered for training the next motion networks: conventional "serial" training and "end-to-end" training. During serial training, no backpropagation was performed over the low-resolution motion networks when training the higher-resolution motion networks. With end-to-end-training, however, backpropagation *was* performed over the lower resolution levels. We investigated this scheme based on the idea that it allows the low-resolution network to learn features which are more expressive for high-resolution motion estimation than the DVF at that level.

The final performance depended partially on the model hyperparameters. Good hyperparameters were found through a representative grid-search, searching among the following values:

- α: The weight factor between the EPE and variable-split terms, $\alpha \in [0.0, 0.1, 0.2, \ldots, 1.0]$
- $\mathcal{K}$: The number of resolution levels to use, $\mathcal{K} \in [3, 4]$
- c_k: The sizes of the convolution kernels in the convolution kernels for every resolution level $k \in \mathcal{K}$, $c_k \in [3, 5, 7]$

This resulted in a total of 1188 different model configurations, which were trained on 5 patients and evaluated on 3 patients. For each of the 1188 combination of hyperparameters, a model was trained for 50 epochs on 5 patients with a fixed random seed. We selected the hyperparameters corresponding to the model that achieved the lowest average EPE on three unseen patients. With these hyperparameters we trained TEMPEST with serial and end-to-end training strategies on the training set of 17 patients. Both models were identically initialized and trained deterministically to prevent unintended advantages on the train set of 17 patients. Respiratory-resolved MRI was made for every patient in the train set with multiple undersampling factors. In total, the train set consisted of 2108 static/dynamic/DVFs samples with undersampling factors $\mathcal{R} \in [1, 3, 5, 7]$.

Chapter 3

To prevent overfitting, the model performance on the validation set was evaluated after every epoch. The models were trained using the Adam optimizer with a base learning rate of $lr = 10^{-4}$ and with 10^{-3} ℓ_2 weight decay on a GPU (Tesla V100, NVIDIA, Santa Clara, CA, USA) with 32GB VRAM. We also used a learning rate schedule that halved the learning rate if the average validation loss did not decrease with at least $\Delta\mathcal{L} = 10^{-8}$ during 10 epochs.

Both models were trained until convergence of the validation loss was observed, i.e. the average validation loss did not decrease more than $\Delta\mathcal{L} = 10^{-8}$ during 10 epochs and the learning rate was smaller than 10^{-8}. During training we performed augmentation on the static and dynamic MRI and the DVFs using TorchIO[143] by performing random flips along an axis ($p = 0.5$), applying a random bias field ($p = 0.25$, order $\in [0, 1, 3, 5]$), and adding random Gaussian noise to the volumes ($p = 0.25$, $\mu = 0$, $\sigma \sim \mathcal{U}(0, 0.05)$). After initial training, the full model was fine-tuned for 100 epochs on a dataset consisting for 25% of image pairs from the training set up to $\mathcal{R} = 7$ *with* motion (i.e. non-zero ground-truth DVF), and for 75% of image pairs of the training set between 7 and 32-fold undersampling *without* motion (i.e. the ground-truth DVF is zero everywhere) to decrease sensitivity to undersampling artifacts.

To increase inference speed, the fully-trained models were quantized from full-precision (fp32) to half-precision (fp16) after fine-tuning by rounding the weights and biases to the nearest 16-bit floating-point number without retraining.

3.2.5 Evaluation

After training, fine-tuning and quantization of the model, we evaluated the model performance on several motion estimation tasks. The accuracy of the DVF$_{\text{TEMPEST}}$ was assessed using two metrics: the voxel-wise EPE compared to a ground-truth DVF, and the mean and standard deviation of the target registration error (TRE). The mean and standard deviation of the EPE was computed over the entire field-of-view, within the body contour, and within the lungs. The body mask was obtained by thresholding the normalized MR image > 0.1, selecting the largest connected component, and performing a morphological closing. The lung mask was obtained by thresholding the normalized MR image within the body < 0.03, selecting the largest connected component, and performing a morphological closing.

The impact of end-to-end training versus serial training was measured by comparing the mean EPE of the two models on the test set over the entire FOV, within the body contour, and within the lungs. Statistical significance ($p < 0.01$)of the difference in mean EPE was established by the Wilcoxon signed-rank test.

The registration performance was evaluated by applying the DVF to the moving, CS-reconstructed volume and estimating the similarity between this warped

volume I_{warped} and the static, CS-reconstructed volume I_{static}. This similarity was computed using the SSIM metric[116] and the normalized root-mean-squared error $\text{NRMSE} = \frac{\sqrt{\text{MSE}(I_{\text{static}}, I_{\text{warped}})}}{\rho_{\text{warped}}}$ where ρ_{warped} is the mean image intensity of I_{warped}.

Respiratory-resolved volumes TEMPEST was evaluated on the four-dimensional respiratory-resolved test set consisting of 5 patients. Model output was compared using the EPE metric ($\mu \pm \sigma$) against the DVF computed with optical flow computed on CS reconstructions ($\text{DVF}_{\text{OF,CS}}$). Moreover, we measure registration performance by registering the CS-reconstructed dynamic volume to the static volume using the $\text{DVF}_{\text{TEMPEST}}$. The registration performance was quantified using the SSIM metric and the NRMSE ($\mu \pm \sigma$) over the entire FOV, within the body contour and within the lungs.

Digital phantom TEMPEST was evaluated without retraining on a digital phantom, as this allows for comparison to a ground-truth DVF. The XCAT digital phantom[144,145] was simulated with MR contrast with equal voxel size and field-of-view size as our training data, as described in Table 3.1, column "4D MRI". The phantom was simulated for 5 frames with respiratory motion up to 50mm in the anterior-posterior direction and 100 mm in the feet-head direction. Motion with a magnitude this large is unlikely to occur in patients, but allows us to evaluate TEMPEST in situations with large deformations. We compared $\text{DVF}_{\text{TEMPEST}}$ to ground-truth DVFs (DVF_{GT}) provided by the digital phantom, which were post-processed using the framework by Eiben et al.[146] for improved accuracy. Retrospective undersampling was performed using a GA-SOS readout for undersampling factors 1, 4, 8, 10, 20, 30, 40, 50. For every undersampling factor, the quality of $\text{DVF}_{\text{TEMPEST}}$ was evaluated using the EPE ($\mu \pm \sigma$) compared to DVF_{GT} over the entire FOV, within the body contour and within the lungs.

Physical phantom Time-resolved 3D cine-MRI of a physical phantom (QUASAR MRI 4D Motion Phantom, Modus QA, Ontario, Canada) was acquired on a 1.5 T hybrid MRI-Linac (Unity, Elekta AB, Sweden). The phantom consisted of an insert in a water tank and was acquired with and without motion applied to the insert. During the "moving phantom" acquisition, the insert moved according to a sinusoidal trajectory with 1/7Hz frequency and 20 mm amplitude. The relevant scan parameters are listed in Table 3.1, column "Moving phantom". During the "stationary phantom" acquisition, for which the relevant scan parameters are listed in Table 3.1, column "Stationary phantom", we tested the sensitivity of the streaking artifacts on the motion estimation performance. The performance of TEMPEST and optical flow were assessed by computing the mean absolute error and Pearson correlation between the ground-truth phantom motion

and the z-magnitude of $\text{DVF}_{\text{TEMPEST}}$ and DVF_{OF}, without retraining TEMPEST.

Fully-sampled CT data To test the generalizability, we evaluated TEMPEST on a publicly accessible 4D respiratory-resolved CT dataset[147] without retraining the model. The quality of $\text{DVF}_{\text{TEMPEST}}$ was assessed using the EPE metric ($\mu \pm \sigma$) within the body contour compared to DVF_{GT}, which was provided by the dataset. Moreover, the registration performance was evaluated using the TRE ($\mu \pm \sigma$) using 41 landmarks within the lungs, which were provided by the dataset for every frame[147].

Real-time motion estimation To evaluate the time-resolved motion estimation performance, we acquired undersampled MRI from two healthy volunteers on an MR-Linac using a GA-SOS readout and a golden-mean radial "kooshball" readout. Both scans were acquired without contrast agent injection and were reconstructed using the NUFFT-adjoint operator after performing radial view-sharing between a dynamic and the two adjacent dynamics[148]. During the kooshball acquisition, a feed-head spoke was acquired every 25 spokes, which provided a self-navigation signal in the feed-head direction.

The relevant scan parameters are listed in Table 3.1, columns "Volunteer 1" and "Volunteer 2", respectively.

For volunteer 1, we evaluated TEMPEST performance by comparing magnitude of $\text{DVF}_{\text{TEMPEST}}$ in the feet-head direction (i.e., motion trace) to the self-navigation signal present in GA-SOS acquisitions.

For volunteer 2, we evaluated TEMPEST performance by computing the Pearson correlation between the magnitude of $\text{DVF}_{\text{TEMPEST}}$ in the feet-head direction to the self-navigation signal obtained from navigation spokes, as no reliable quantification of the true motion is available at this high undersampling factor. The undersampling factor for kooshball MRI was given by $\mathcal{R} = (M_x \cdot M_y \cdot \pi/2) / N_{sp}$, where N_{sp} is the number of spokes per dynamic and $M_x = M_y = 108$ is the matrix size in the x and y direction.

Time We measured whether TEMPEST is fast enough for real-time applications by reporting the time for MR acquisition, image reconstruction, and motion estimation. We measured the model inference time ($\mu \pm \sigma$) at fp32 and fp16 resolution over 50 evaluations for static/dynamic volume pairs from the test set with a matrix size of $206 \times 206 \times 77$ at full resolution. We considered our approach fast enough for real-time MRIgRT if the total time ≤ 400 ms, as suggested by Keall *et al.*[50,127].

3.3 Results

Based on the hyperparameter evaluation we found that $\alpha = 0.8$ and $\mathcal{K} = 3$ to be optimal among those that we evaluated. The full results are presented in Supporting Information Figures 1-3. For the sizes of the convolution kernels we found that $c_0 = 3, c_1 = 5$, and $c_2 = 3$ to be best performing, where $k = 0$ is the lowest spatial resolution, and $k = 2$ is full spatial resolution, resulting in a model with 859660 trainable parameters.

We trained two variants of TEMPEST with these hyperparameters: with serial training and with end-to-end training. The network operating at the lowest resolution level was trained for 250 epochs in approximately 4 hours with a batch size of 8. The network operating at the second resolution level was trained for 150 epochs in approximately 8 hours with a batch size of 4, while also performing backpropagation over the lowest resolution level. The network operating the highest resolution level was trained for 125 epochs in approximately 12 hours with a batch size of 4, while also performing backpropagation over both models operating at lower resolution levels.

An example of a DVF produced by TEMPEST from undersampled MRI ($\mathcal{R} = 23$) is shown in Figure 3.3A-C. $\text{DVF}_{\text{TEMPEST}}$ shows good agreement with $\text{DVF}_{\text{CS,OF}}$ (D-F). In this particular case, the mean EPE was 2.78 mm. Animated figures of TEMPEST DVFs computed on 4D MRI are provided in Supporting Information Videos S1-3.

The quality of TEMPEST DVFs significantly increased when using end-to-end training compared to serial training on our test set, as shown in Figure 3.4. For example, the average EPE at $\mathcal{R} = 15$ reduced from 3.47 ± 0.76 mm using serial training to 2.25 ± 0.70 mm using end-to-end training within the body contour (Wilcoxon, $p \ll 0.001$). At the same time, the average SSIM increased with $\geq 6\%$ at $\mathcal{R} \geq 15$ (Wilcoxon, $p \ll 0.001$) when using end-to-end training, indicating better registration performance.

We quantized the weights and biases of TEMPEST from full-precision (fp32) to half-precision (fp16). Our analysis revealed that weight quantization step has negligible impact on the model performance, increasing the mean EPE with only $3.7 \cdot 10^{-4}$ mm. However, weight quantization reduced the inference time of a static/dynamic volume-pair of matrix size $206 \times 206 \times 77$ from 81 ± 7.4 ms to 31 ± 2.9 ms on a NVIDIA V100 GPU, reducing the total latency.

Based on these results, the quantized, end-to-end-trained TEMPEST model has been adopted for further performance evaluation.
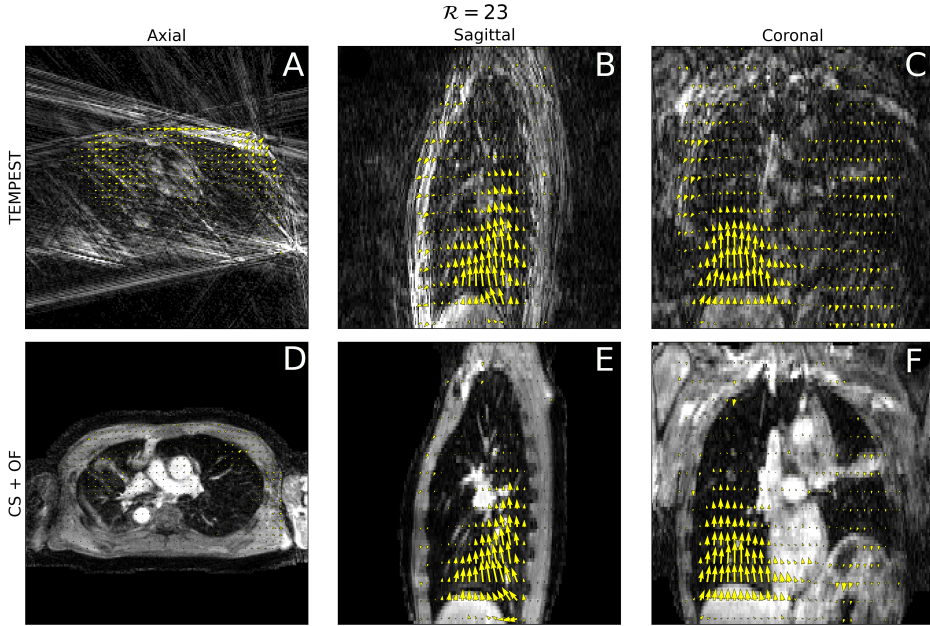
Figure 3.3 Example of motion estimation. An example of motion estimation in 3D by TEMPEST (A-C) on 23-fold undersampled MRI. Good correspondence can be observed between the motion estimated by TEMPEST and motion estimated by optical flow computed on compressed sense reconstructions (D-F). In the quasi-static region, TEMPEST estimates slightly larger residual motion. In the Supporting Information Videos S1-3, animated figures are provided.

3.3.1 4D respiratory-resolved motion estimation

The performance of TEMPEST on respiratory-resolved MRI is shown in Figure 3.5. We found that the EPE remained within 1.9 ± 0.6 mm within the lungs at $\mathcal{R} = 18$ compared to $\text{DVF}_{\text{CS,OF}}$. The mean NRMSE and mean SSIM plateau at $\mathcal{R} > 15$ at approximately 0.51 and 0.63 within the lungs, respectively. Surprisingly, even though TEMPEST has been trained on MRI with an undersampling factor up to $\mathcal{R} \approx 7$, the mean EPE only moderately increases from 1.5 mm to 1.9 mm at $\mathcal{R} = 18$ within the lungs.

3.3.2 Digital phantom

Evaluation of TEMPEST on a digital phantom showed results similar to the respiratory-resolved test set, as shown in Figure 3.6. At low undersampling factors, e.g. $\mathcal{R} = 4$, the mean EPE of $\text{DVF}_{\text{TEMPEST}}$ compared to DVF_{GT} was 0.8 ± 0.12 mm within the body contour. At higher undersampling factors beyond those seen

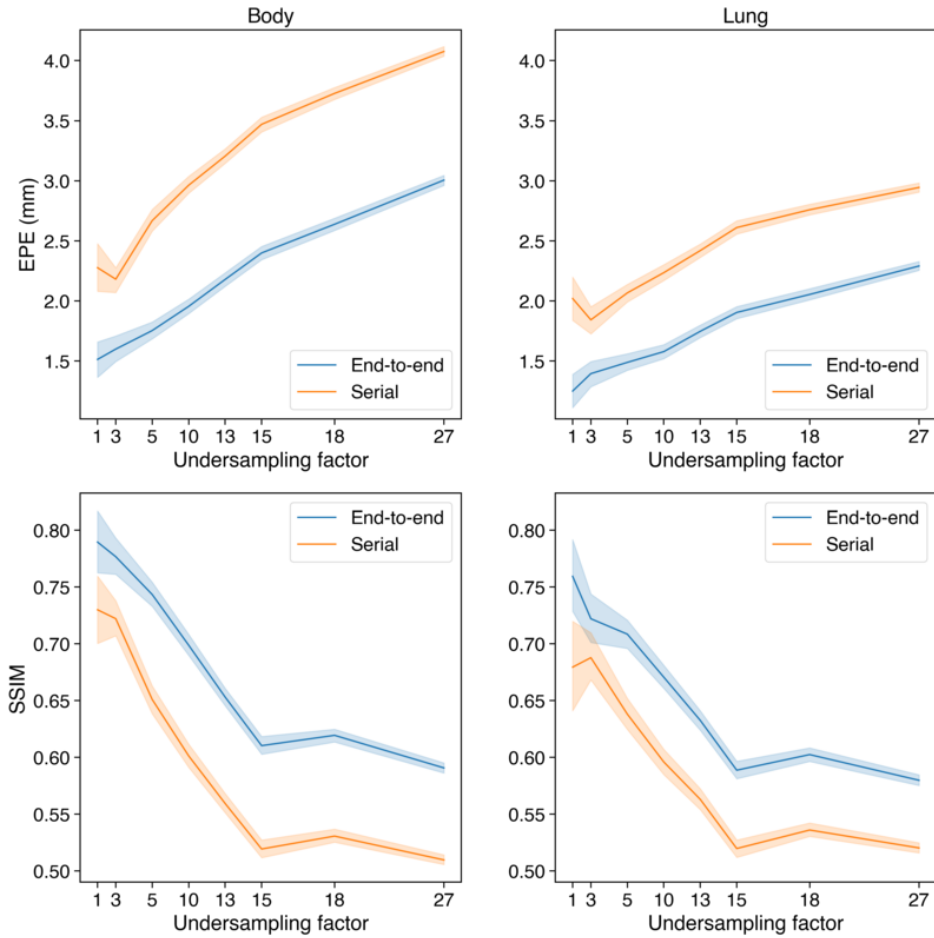**Figure 3.4** Comparison of serial vs. end-to-end training on 4D MRI. Two variants of TEMPEST were trained: with end-to-end backpropagation (blue) and serial training (orange). EPE (top row) and the SSIM of registered compressed sense reconstructions (bottom row) as a function of the undersampling factor. Evaluation was done within the body contour (left column) and within the lungs (right column).
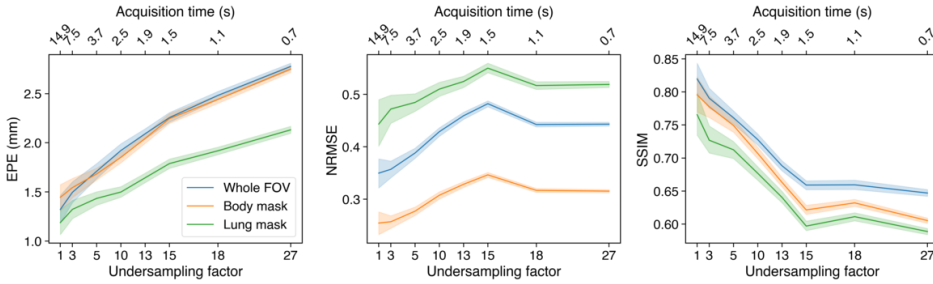
Figure 3.5 4D Evaluation. The EPE (left), NRMSE of compressed sense reconstructions after registration (middle), and SSIM of compressed sense reconstructions after registration (right) as a function of the undersampling factor. These metrics were evaluated over the whole FOV (blue), within the body contour (orange), and within the lungs (green). Secondary x-axis show the approximate acquisition time in seconds using GA-SOS for that undersampling factor.

during training, the mean and standard deviation of the EPE increases, yielding a mean EPE of 2.0 ± 0.76 within the body contour at $\mathcal{R} = 30$.

3.3.3 Physical phantom

In Figure 3.7 we show the results of the phantom experiments. In the moving phantom experiment (top row), TEMPEST (red) produces motion traces similar to the ground-truth self-navigation motion signal (yellow), indicated by the Pearson correlation factor of 0.911 at $\mathcal{R} = 20.2$. At $\mathcal{R} = 28.3$, TEMPEST estimates motion with an absolute error of 1.75 ± 1.3 mm versus 2.66 ± 1.7 mm for optical flow. For the stationary phantom (middle), optical flow shows significantly more residual motion than TEMPEST at high undersampling factors while there should be no motion. At $\mathcal{R} = 20.3$, TEMPEST produces motion traces with an error of 1.03 ± 0.6 mm, whereas optical flow produces an error of 3.65 ± 2.4 mm.

3.3.4 Generalization to CT data

Surprisingly, when applied to 4D CT, TEMPEST estimates motion with low EPE compared to the ground-truth DVF without retraining the model for this modality. For example, Figure 3.8 shows that TEMPEST produces DVFs with a mean EPE of 1.23 mm over all respiratory phases, and is able to register CT with little residual motion. When registering images with no motion (e.g. estimating motion from exhale to exhale) the mean EPE was 0.29 mm. The largest mean EPE was observed when registering the inhale CT to exhale, resulting in a mean EPE of 2.01 mm. Registration of the landmarks yielded an average TRE of 1.87 ± 1.65 mm.
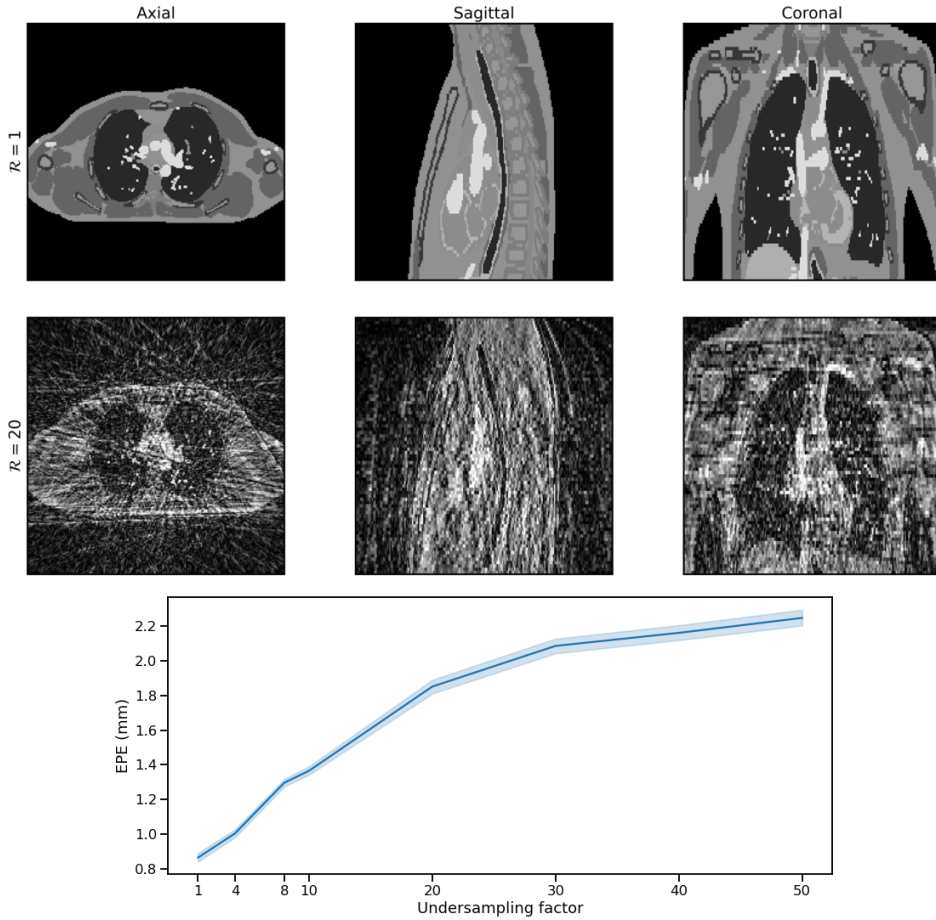
Figure 3.6 Digital phantom results. An XCAT digital phantom was simulated with up to 50mm AP motion and 100mm FH motion. An example of a fully-sampled dynamic in exhale is shown in the top row. Volumes were retrospectively undersampled using a GA-SOS trajectory, e.g. $\mathcal{R} = 20$, as shown in the middle row. The bottom row shows the magnitude error $(\mu \pm \sigma)$ between the model output and the post-processed ground-truth XCAT DVF over 100 reconstructions per undersampling factor, using different azimuthal angles and noise.

Figure 3.7 Phantom results. In the background is the progression of a single location along the slice direction over time. In yellow is the ground-truth motion trace produced by the phantom. In red, the motion trace produced by TEMPEST. In cyan, the motion trace produced by optical flow. This was computed for the phantom in motion (top row) and the stationary phantom (bottom row) 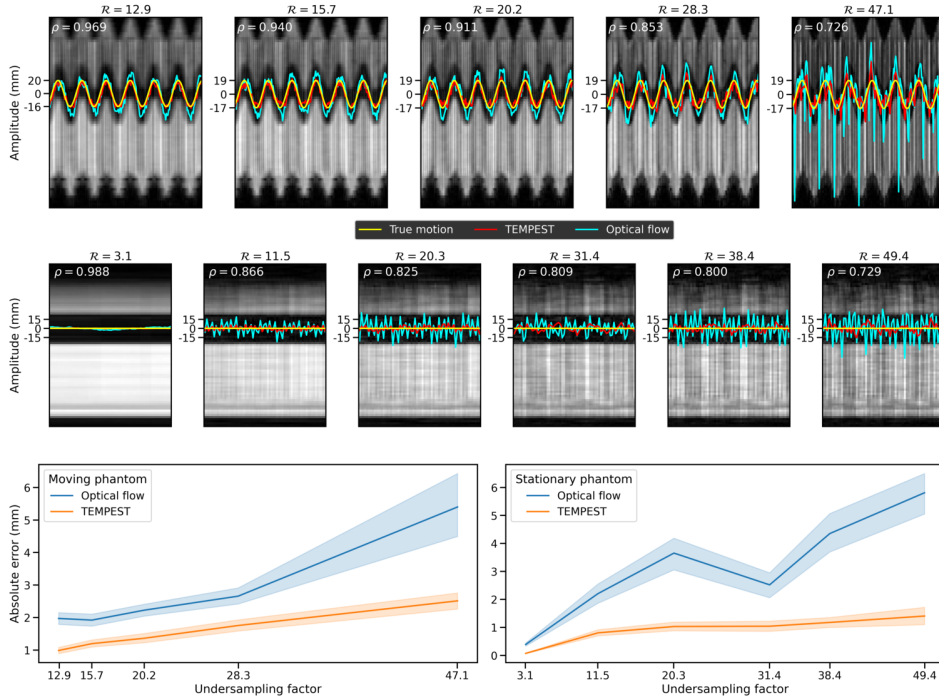for several undersampling factors. The Pearson correlation between the ground-truth motion and the TEMPEST motion is displayed above the plot as ρ. In the bottom row, the absolute error ($\mu \pm \sigma$) is shown for both experiments as a function of the undersampling factor.

3.3.5 Real-time motion estimation

Evaluation of TEMPEST on time-resolved MRI is shown in Figure 3.9. On GA-SOS k-space acquired on an MR-Linac, TEMPEST produces motion similar to the self-navigation signal, indicated by the Pearson correlation of 0.93 at $\mathcal{R} = 18.5$. Animated figures of TEMPEST DVFs computed on time-resolved GA-SOS MRI are provided in Supporting Information Video S4-6.

With the golden-means kooshball readout we achieve good correlation between the 40 and 75 spokes per dynamic, achieving a Pearson correlation of $\sim$0.80. Animated figures of TEMPEST DVFs computed on time-resolved golden-means kooshball MRI are provided in Supporting Information Video S7-10.
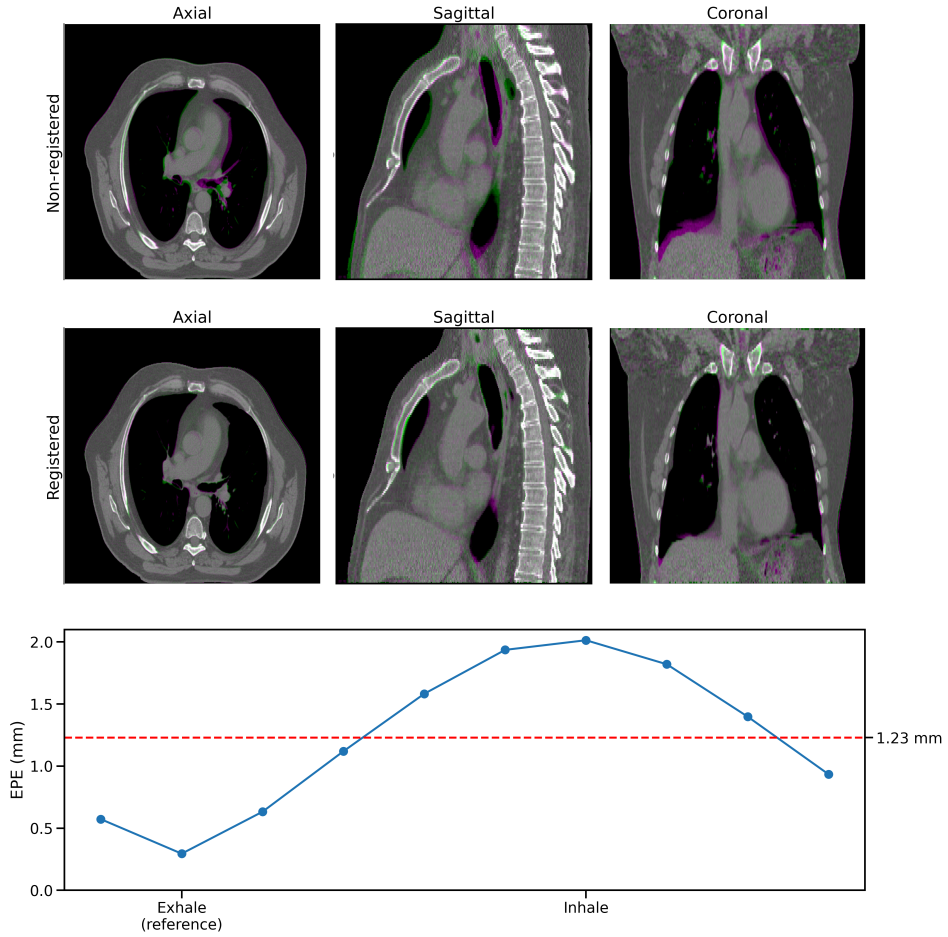
Figure 3.8 CT results. TEMPEST was evaluated on a fully-sampled respiratory-resolved 4D-CT dataset. Non-registered CT shows large differences in image space, especially in the liver dome (top row). In the middle row, it can be seen that TEMPEST is able to register the images with little residual error. On the bottom row, the EPE of TEMPEST compared to the ground-truth DVF is shown as a function of the respiratory phase. End exhale was the reference phase. The mean EPE was 1.23 mm, shown in the red horizontal line.
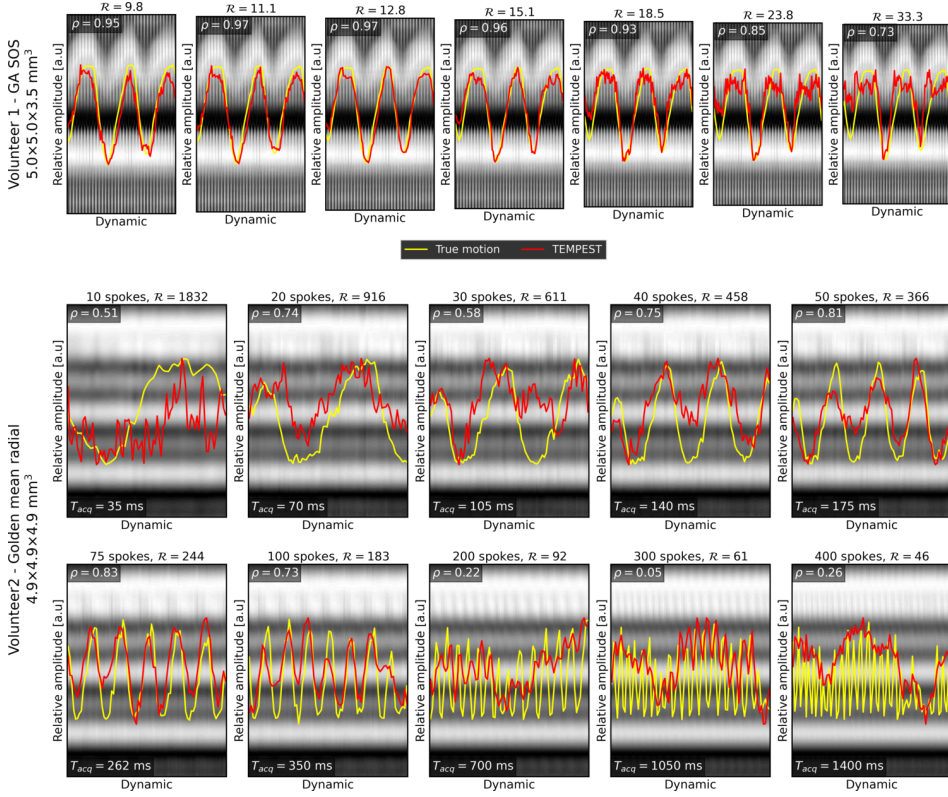
Figure 3.9 Real-time results. TEMPEST was evaluated on time-resolved GA-SOS volunteer data (top, red line) and compared to the self-navigation signal (yellow) for multiple undersampling factors. The Pearson correlation between these lines is shown in the same figure. This was also done for another volunteer using a golden-mean radial kooshball acquisition (middle, bottom). The self-navigation in a surrogate for the motion in the feet-head direction in both scans.

Acquiring one spoke per slice using a GA-SOS readout takes ∼270 ms for 77 slices. Reconstruction of GA-SOS k-space at three resolution levels, where full-resolution is $206 \times 206 \times 77$ takes about 50 ms per slice using a simple GPU-based NUFFT algorithm. As TEMPEST takes 30 ms to estimate motion, DVFs can be computed using our methods well within 400 ms, which is the maximum affordable latency for radiotherapy, as suggested by Keall *et al.*[50,127].

3.4   Discussion

In this work, we have proposed a deep learning model called TEMPEST to estimate 3D DVFs from highly undersampled acquisitions to facilitate real-time

MRIgRT applications. In particular, we have presented a multi-resolution model that has been trained on respiratory-resolved MRI that can be used to estimate motion with low latency and high spatio-temporal resolution in time-resolved MRI. This model is an extension from 2D to 3D of our previously-presented approach that estimates motion from undersampled 2D golden-angle radial acquisitions using NUFFT reconstruction and deep learning-based motion estimation[134]. To the best of our knowledge, this is the first deep learning model that enables real-time 3D motion estimation from highly undersampled MRI, with a total latency of less than 400 ms.

We have shown that at $\mathcal{R} = 18$ motion was estimated in respiratory-resolved imaging with less than 2 mm error. The model was validated with various experiments, such as a digital phantom, a physical motion phantom and 4D respiratory-resolved CT data. In all these experiments motion could be accurately estimated with undersampling factors up to $\mathcal{R} = 20$. Inference of the model took only ~30 ms, which is acceptable for MRIgRT[49,127]. We found that "end-to-end" training improved DVF quality compared to "serial" training, decreasing the mean EPE with over 1 mm compared to serial training.

Our experiments with the physical phantom show that TEMPEST is able to accurately estimate motion from data acquired on an MR-linac compared to the ground-truth, even though the model was trained on patient data. Compared to optical flow, TEMPEST DVFs seem to display a greater robustness against the incoherent streaking artifacts present in radially undersampled images. Especially in the stationary phantom experiment optical flow produces a response to aliasing, most notably at higher undersampling factors. Presumably, this is due to the introduced image artifacts present in highly undersampled images.

Even though TEMPEST has been trained on T_1-weighted spoiled gradient echo lung MRI we have demonstrated that our model also performs surprisingly well on different imaging modalities, such as CT, without retraining and yields a TRE of 1.87 ± 1.65 mm. While these results are promising, state-of-the-art image registration methods or specialized neural networks trained solely on CT images report lower TREs. For example, Marstal et al.[149] showed that Elastix is able to obtain a TRE of 1.58 ± 0.59 mm and Eppenhof et al.[150] obtained a TRE of 1.38 ± 1.24 mm using CNNs. However, these results indicate that the model may generalize well and demonstrates that model has not over-fit to a specific imaging contrast. Further experiments are needed to investigate whether TEMPEST also generalizes beyond T_1-weighted MRI contrasts, radial MRI, or to different body sites. For example, TEMPEST could be applied to MRI acquired with other non-Cartesian acquisitions like a stack-of-spirals[151], golden-mean cones[152], or even Cartesian readouts such as variable-density Cartesian spirals[153]. However, this may require retraining as the aliasing changes depending on the sampled trajectory.

For time-resolved imaging, TEMPEST is able to produce motion traces with high correlation to the self-navigation signal, as demonstrated in the phantom experiments and the time-resolved MRI experiments. For GA-SOS MRI acquired using volunteer 1 at $\mathcal{R} = 18.5$, TEMPEST produces DVFs with a motion trace correlating 93% to the self-navigation signal. For the golden-mean radial kooshball data, good motion traces can be produces between 40 and 75 spokes per dynamic, corresponding to extreme undersampling factors between $\mathcal{R} = 458$ and $\mathcal{R} = 244$. We hypothesize for this dataset, this number of spokes provides a good trade-off between image quality and acquisition latency. With fewer spokes, the undersampling artifacts presumably dominates the motion. With more spokes, longer acquisitions introduce temporal aliasing, as shown in the Supporting Information Videos S7-10. However, radial view-sharing reduces the effective undersampling factor such that this approach becomes feasible. Moreover, the spatial resolution of 5 mm was significantly larger than the training data. Using larger voxels significantly accelerates the MR acquisition, but reduces image quality as fine details are lost. However, it has been demonstrated that larger voxels have little impact on the estimated optical flow[97]. With 50 spokes per dynamic, acquisition took ∼175 ms, while motion estimation took 30 ms. Our approach thus took 205 ms plus time for image reconstruction, which was within the time budget of 400 ms for real-time MRIgRT applications, possibly enabling real-time adaptive MRI-guided radiotherapy by resolving motion during radiotherapy. For these experiments, we used GPU-NUFFT implementations that were not fully optimized and assume serial reconstruction of slices. We reckon that highly-optimized, parallel NUFFTs can significantly reduce image reconstruction time.

Compared to other works, our method is significantly faster while achieving similar accuracy. For example, Stemkens et al.[120] obtained a 3D motion estimation with a RMSE of 1 mm using a 360 ms 2D acquisition and a few seconds of motion calculation, which is comparable with what we observed. However, this method is not a 'full' 3D method but uses multi-2D cine scans in conjunction with a 4D MRI to obtain 3D motion estimates, possibly limiting the accuracy of the method. Moreover, the computation time of multiple seconds is not fast enough for MRIgRT. Morales et al.[123] proposed an unsupervised deep learning method to learn 3D DVFs in cardiac imaging with a mean EPE of 2.25 mm. However, their method operated on fully-sampled images and needs 9 seconds of computation for a single DVF, which is not fast enough for MRIgRT. At ten-fold radial undersampling, we achieve a lower error with an approximately 300 times shorter computation time. Navest et al.[154] used another method to detect motion in MRI. They detected motion from the variance in the noise present in MRI acquisition, achieving fast computation and accurate detection of bulk movement, respiratory motion, cardiac motion and swallowing. However, while this method may be useful for gated dose delivery, it did not provide absolute motion information per voxel and can therefore currently not be used for real-time adaptive radiotherapy.

The method we propose is a supervised method and requires ground-truth DVFs for learning, which could be considered as a limitation given that obtaining high-quality ground-truth DVFs for time-resolved 3D MRI is challenging. We have opted to use optical flow to generate ground-truth DVFs. While this is a simple and well-known motion estimation method, the assumptions optical flow uses to compute DVFs has limitations in, for example, regions with piecewise constant image intensities[41]. Using other motion estimation methods, such as Elastix[42] or demons[155], might improve results. Another way to overcome this challenge is by training on synthetic DVFs[132]. However, the model may learn non-physiologically plausible DVFs. Also, the training data is then limited to retrospectively undersampled k-space, which does not suffer from imperfect MRI acquisitions observed in practice. A different way to overcome this challenge is by using an *unsupervised* method[123]. However, these approaches often use the registration performance as loss metrics[156], which may be hindered by for undersampled acquisitions due image artifacts.

Due to the highly undersampled nature of the time-resolved MRI experiments and the lack of ground-truth DVFs, high-quality validation of TEMPEST is challenging. As severe image artifacts preclude the computation of accurate ground-truth DVFs, the self-navigation signal is the most reliable surrogate for ground-truth motion. However, this is a one-dimensional motion which only provides relative motion information along one direction, rather than an absolute displacement per voxel along the three axes. Moreover, comparison of global motion information does not allow for motion quality evaluation of specific sites, such as tumors or OARs. In the future, realistic deformable motion phantoms might provide more insight in the motion estimation quality and evaluation on a large patient population could give a better characterization of tumor or OAR motion by using metrics based on anatomical information, such as the Dice score or Hausdorff distance between estimated and ground-truth segmentations.

Even though TEMPEST was fine-tuned on highly undersampled images, there is still a response to undersampling artifacts at very high undersampling factors. This could be mitigated by using more sophisticated image reconstruction algorithms, e.g. compressed sense or DL-based image reconstruction. However, as no additional latency is permitted for MRIgRT, these methods are currently not suitable. While the presented multi-resolution approach has proven to produce good results, different deep learning model architectures incorporating concepts from 2D optical flow, such as cascaded flow inference[141] and optical flow cost volumes[140], have the potential to improve DVF quality at the cost of increased inference times. Another possible cause of the residual undersampling response could be the relatively small training set of 17 patients. Moreover, the estimated hyperparameters might not be optimal as they were optimized for 3 patients, which might allow for selection of hyperparameters for those three patients instead of all patients. Increasing the number of training samples for hyperparameter estimation and model training might yield improved results at high undersampling factors.

Chapter 3

59

An alternative approach may foresee omitting image reconstruction and aim at obtaining DVFs directly from k-space, as proposed with model-based methods by Huttinga et al.[157]. However, reconstructing DVFs from k-space with deep learning might prove challenging as convolutional operators have strong *local* priors, whereas k-space contains *global* information.

We believe that deep learning models are a promising way to facilitate real-time adaptive MRIgRT, where latency in spatio-temporal resolution have paramount importance. Also, we foresee that TEMPEST could be used for applications that require fast motion estimation or registration of images with artifacts, e.g. dose accumulation[158], image registration[42] or motion-compensated image reconstruction[159,160]. In the future, we aim to investigate possibilities to further increase the DVF accuracy at extreme undersampling factors and the spatio-temporal resolution of TEMPEST. TEMPEST could be extended to include temporal information, based on the fact that motion can be represented with spatially and temporally low-rank models[130].

3.5   Conclusion

We have presented TEMPEST, a deep learning model that estimates time-resolved 3D DVFs from undersampled 3D MRI with high spatio-temporal resolution for real-time adaptive MRI-guided radiotherapy. To the best of our knowledge, this is the first method to perform real-time 3D motion estimation from highly undersampled MRI. We have shown that this model can estimate DVFs with high accuracy (< 2 mm), low latency, and high spatio-temporal resolution from undersampled radial MRI. TEMPEST estimated DVFs within 200 ms, including MRI acquisition, complying with the requirements for online adaptive MRIgRT. We have evaluated the model performance *in-silico* using digital and physical motion phantoms and applied the model to 4D CT without retraining. Also, we have shown that TEMPEST can estimate accurate DVFs and achieves good performance in two healthy volunteers.
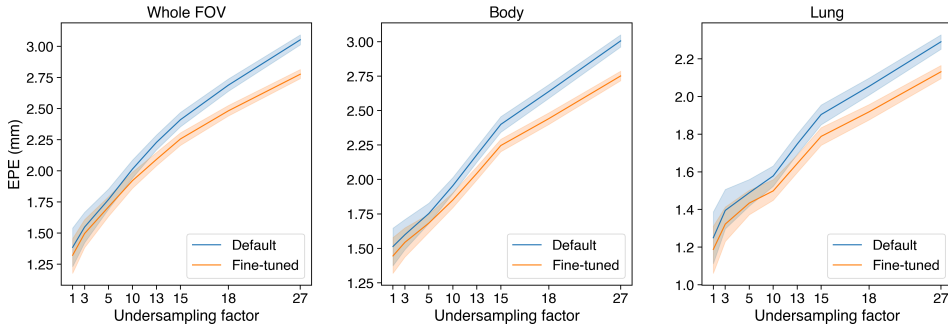
3.6   Acknowledgment

Figure 3.10 Fine-tuning results. Two TEMPEST models were fine-tuned: one using the regular data schedule (blue) and one using our proposed data schedule, featuring highly undersampled images with no motion (orange).

3.7 Appendix: Impact of fine-tuning

We performed an ablation study to evaluate the impact of fine-tuning a fully-trained TEMPEST on a dataset consisting for 25% of image pairs from the training set up to $\mathcal{R} = 7$ *with* motion (i.e. non-zero ground-truth DVF), and for 75% of image pairs of the training set between 7 and 32-fold undersampling *without* motion (i.e. the ground-truth DVF is zero everywhere), as described in Section 3.2.4.

We hypothesized that this training schedule improves robustness against severe undersampling artifacts. We tested this hypothesis by fine-tuning TEMPEST two times:

- Continue training TEMPEST for 100 epochs on the training set, making no changes to the dataset and only presenting MRI undersampled up to $\mathcal{R} = 7$.
- Continue training TEMPEST for 100 epochs on the training set using the proposed data schedule.

We evaluated both models after fine-tuning on the test set using the EPE metric ($\mu \pm \sigma$). The EPE was evaluated over the entire FOV, within the body contour, and within the lungs. The results are presented in Figure 3.10.

Here, we see that the model fine-tuned using our data schedule outperforms the default data schedule for every undersampling factor. At low undersampling factors, the impact is small (0.06 mm EPE reduction within the body contour at $\mathcal{R} = 1$), but at high undersampling factors the impact increases (0.25 mm EPE reduction within the body contour at $\mathcal{R} = 27$).

We conclude that fine-tuning TEMPEST by exposing the model to extremely undersampled MRI with no motion increases image artifact robustness compared to fine-tuning using the unmodified data schedule.

CHAPTER 4

⊥-loss: A symmetric loss function for magnetic resonance imaging reconstruction and image registration with deep learning

ML Terpstra
M Maspero
A Sbrizzi
CAT van den Berg

Abstract

Convolutional neural networks (CNNs) are increasingly adopted in medical imaging, e.g., to reconstruct high-quality images from undersampled magnetic resonance imaging (MRI) acquisitions or estimate subject motion during an examination. MRI is naturally acquired in the complex domain $\mathbb{C}$, obtaining magnitude and phase information in k-space. However, CNNs in complex regression tasks are almost exclusively trained to minimize the L2 loss or maximizing the magnitude structural similarity (SSIM), which are possibly not optimal as they do not take full advantage of the magnitude *and* phase information present in the complex domain. This work identifies that minimizing the L2 loss in the complex field has an asymmetry in the magnitude/phase loss landscape and is biased, underestimating the reconstructed magnitude. To resolve this, we propose a new loss function for regression in the complex domain called $\perp$-loss, which adds a novel phase term to established magnitude loss functions, e.g., L2 or SSIM. We show $\perp$-loss is symmetric in the magnitude/phase domain and has favourable properties when applied to regression in the complex domain. Specifically, we evaluate the $\perp+\ell^2$-loss and $\perp+$SSIM-loss for complex undersampled MR image reconstruction tasks and MR image registration tasks. We show that training a model to minimize the $\perp+\ell^2$-loss outperforms models trained to minimize the L2 loss and results in similar performance compared to models trained to maximize the magnitude SSIM while offering high-quality phase reconstruction. Moreover, $\perp$-loss is defined in $\mathbb{R}^n$, and we apply the loss function to the $\mathbb{R}^2$ domain by learning 2D deformation vector fields for image registration. We show that a model trained to minimize the $\perp+\ell^2$-loss outperforms models trained to minimize the end-point error loss.

4.1 Introduction

Magnetic resonance imaging (MRI) is a noninvasive imaging technique to obtain an anatomical image with high resolution and excellent soft-tissue contrast. These properties have made MRI an indispensable diagnostic tool and is increasingly used for interventional guidance, such as high-intensity focused ultrasound[161], catheter guidance during surgery[162], and radiotherapy[25,30].

MRI must be acquired, reconstructed, and processed with high accuracy in real-time for interventional guidance applications. However, MRI acquisition can be time-consuming due to hardware limitations that ensure patient safety, e.g., low gradient slew rate to avoid peripheral nerve stimulation[163] or limited radio-frequency power to prevent patient heating[164]. However, the limited shot-encoding power of MR acquisitions is the most time-consuming part of image formation[165], resulting in repeated sampling of the frequency domain (k-space) to fulfil the Nyquist sampling criterion[53]. These constraints limit patient throughput and prohibit MRI applications that require high spatio-temporal resolution, e.g., cardiac imaging[166], speech imaging[167], or tracking abdominal motion[168]. One way to accelerate MRI is to acquire fewer data, i.e., undersampling the frequency domain as the number of k-space samples determines acquisition time. However, undersampling violates the Nyquist criterion, introducing image artifacts. Several methods have been proposed to remove these undersampling artifacts. For example, parallel imaging[67,68] exploits information redundancy using multiple receiver elements. Also, iterative reconstruction algorithms as compressed sensing have been proposed[71], which resolve image artifacts by casting MR reconstruction as a sparse denoising problem. However, the resulting acceleration factor of parallel imaging remains limited[70], e.g., up to four-fold undersampling, and compressed sensing reconstructions cause a significant reconstruction latency, precluding real-time applications.

Recently, machine learning has been proposed as an alternative to traditional methods to solve inverse problems[169]. Specifically, several methods have been proposed to use convolutional neural networks (CNNs) to accelerate undersampled MRI reconstruction[89,90], perform organ segmentation[170], or estimate motion from undersampled MRI[171]. These methods have attractive properties compared to traditional methods, such as the ability to obtain high-quality MRI reconstructions with high undersampling factors (eight-fold or higher), reducing acquisition time. Moreover, CNNs exhibit low inference times by taking advantage of parallel GPU architectures, even though training of CNNs can take hours or days. Using CNNs to accelerate the acquisition, reconstruction, and processing of MRI could enable new applications such as real-time interventional guidance using MRI[172] or real-time adaptive MRI-guided radiotherapy[45].

In this work, we consider complex regression with CNNs for MRI applications. These CNNs are trained by finding parameters that minimize a loss function

over a training set. The quality of the estimated solution largely depends on the loss function, as these loss functions determine the impact of residual artifacts and the importance of specific (image) features, such as texture, contrast, or the effects of noise propagation. It has been shown that loss functions with desirable properties, such as monotonicity, smoothness[173], or symmetry[174,175], can lead to better generalization, noise robustness, and faster convergence[176], depending on the task at hand. As MRI is an inherently complex signal, loss functions for reconstructing or processing MRI are naturally defined in the complex domain.

For example, for image reconstruction, popular choices of loss functions between a complex target image $\mathbf{Y} \in \mathbb{C}^{m \times n}$ and a complex estimated image $\hat{\mathbf{Y}} \in \mathbb{C}^{m \times n}$ include minimizing the complex difference using the $\ell_{\mathbb{C}}^p$-norm, e.g., $\ell_{\mathbb{C}}^2(\mathbf{Y}, \hat{\mathbf{Y}}) = ||\Re(\mathbf{Y} - \hat{\mathbf{Y}})||^2 + ||\Im(\mathbf{Y} - \hat{\mathbf{Y}})||^2$ or $\ell_{\mathbb{C}}^1(\mathbf{Y}, \hat{\mathbf{Y}}) = ||\Re(\mathbf{Y} - \hat{\mathbf{Y}})||_1 + ||\Im(\mathbf{Y} - \hat{\mathbf{Y}})||_1$, where $|| \cdot ||^2$ is the squared Frobenius norm, $|| \cdot ||_1$ is the ℓ_1 norm, $\Re(x)$ is the real part of x, and $\Im(x)$ is the imaginary part of x, or $\ell_{\mathrm{SSIM}} = 1 - \mathrm{SSIM}(|\mathbf{Y}|, |\hat{\mathbf{Y}}|)$, where SSIM is the structural similarity index measure[116]. For image registration, the most common loss function between two deformation vector fields (DVFs) is the end-point error (EPE), which is equivalent to $\ell_{\mathbb{C}}^2$ in the $\mathbb{R}^2$ domain[107,177]. The geometric interpretation of minimizing the $\ell_{\mathbb{C}}^2$-norm is illustrated in fig. 4.1.

These loss functions do not take full advantage of the phase structure of the data; for example, ℓ_{SSIM} discards phase information, while it has been shown that using this information could improve image reconstruction performance[178]. On the other hand, the $\ell_{\mathbb{C}}^p$ is separately defined on the real and imaginary components of the complex-valued signal, possibly precluding the $\ell_{\mathbb{C}}^p$ from taking full advantage of the magnitude and phase properties. Moreover, these loss functions do not necessarily produce a symmetrically-distributed error, while it has been demonstrated that using symmetric loss functions can improve task performance[174]. Specifically, a symmetrically-distributed error when estimating motion for image registration is a desirable property as the registration error is equally distributed in all directions.

We hypothesize that a new loss function based on a complex signal's magnitude and phase components could lead to improved regression in the complex domain using deep learning models, thus improving image quality when reconstructing undersampled MRI and leading to better image registration. Therefore, we propose ⊥-loss (pronounced ˌpɜːpˈlɒs), a symmetric loss function defined in the magnitude and phase domain. In this work, we will:

1. Analyze the behavior of $\ell_{\mathbb{C}}^p$ in the complex plane and investigate the symmetry of the loss landscape.
2. Introduce the ⊥-loss function, which operates on the polar representation of complex numbers and adds a novel phase term to magnitude loss functions. We examine the loss landscape produced by this loss function, comparing it to $\ell_{\mathbb{C}}^p$.
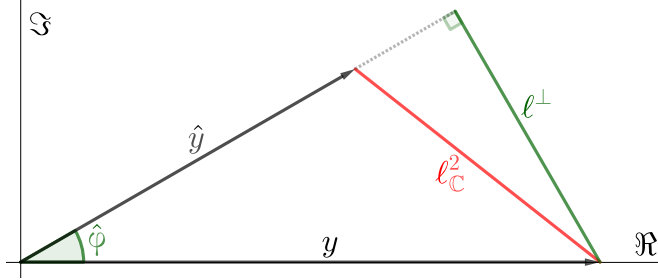
Figure 4.1 Geometric illustration of regression in the $\mathbb{R}^2$ domain. An estimated vector $\hat{Y}$ compared to a target vector Y, with magnitude ratio $\lambda = |\hat{Y}|/|Y|$ and phase difference $\hat{\varphi}$. The $\ell_{\mathbb{C}}^2$ loss is the magnitude of the red vector. Our proposed loss function ⊥-loss determines the phase error $\hat{\varphi}$ as the scalar rejection $\ell^{\perp}$, i.e., the length of the perpendicular line from $\hat{Y}$ to Y.

3. Study the performance of the $\ell_{\mathbb{C}}^2$ loss, ℓ_{SSIM}, ⊥+ℓ^2-loss, and ⊥+SSIM-loss functions for undersampled MRI reconstruction tasks using deep neural networks.
4. Explore the application of the ⊥-loss to other domains, considering regression in the $\mathbb{R}^2$ domain for 2D image registration.

4.2 Theory

This section examines the topology of the $\ell_{\mathbb{C}}^p$ loss landscape and proposes a new loss function. We consider a complex image $\mathbf{Y} \in \mathbb{C}^{m \times n}$, representing the signal of each voxel as $Y_{jk} = a + bi$ where $a, b \in \mathbb{R}$ are the real and imaginary components (Cartesian representation), or $Y_{jk} = |Y_{jk}|e^{i\varphi_{jk}}$ where $|Y_{jk}|$ is the *magnitude* and φ_{jk} as the *phase* (polar representation).

4.2.1 Proof of asymmetry of $\ell_{\mathbb{C}}^2$

It has been shown that least-squares regression in the $\mathbb{R}^2$-space in the presence of noise results is biased. In particular, performing least-squares regression on noisy data results in parameters that underestimate the recovered magnitude[179–183]. We make the case that this bias is also present for regression tasks in the complex domain.

Suppose we wish to estimate a complex value, e.g., estimating a voxel $\hat{Y}$ against the ground-truth voxel Y in the target image $\mathbf{Y}$. The reconstructed voxel $\hat{Y}$ is typically estimated by regression, minimizing a loss function such as the complex extension of the ℓ^2-norm, i.e., $\ell_{\mathbb{C}}^2(Y, \hat{Y}) = ||\Re(Y - \hat{Y})||^2 + ||\Im(Y - \hat{Y})||^2$. The reconstructed magnitude bias is expressed as $|\hat{Y}| = \lambda|Y|$ with $\lambda > 0$, and the remaining phase error is denoted as $\hat{\varphi}$.
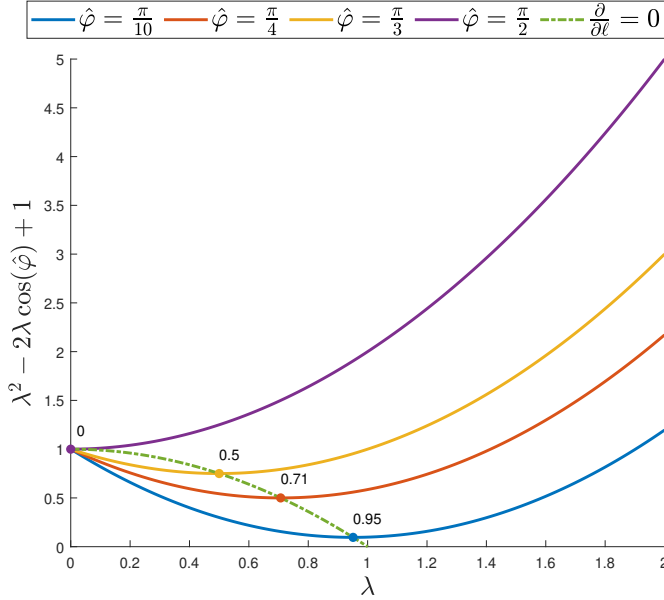
Figure 4.2 The loss of $\ell_{\mathbb{C}}^2$. The parabola of the loss for a given value of $\hat{\varphi}$ is shown as a function of λ. The minimal value is indicated in every line plot, and the analytical curve of minimal loss is shown in green. It is clear that as $\hat{\varphi}$ increases, the minimum of the loss function goes to $\lambda \to 0$.

In fig. 4.1, this estimation is illustrated using vectors in the complex plane, where the $\ell_{\mathbb{C}}^2$ error, $\hat{Y}$, and Y form a triangle. Therefore $\ell_{\mathbb{C}}^2$ can be expressed using the law of cosines as

$$\ell_{\mathbb{C}}^2 = |\hat{Y}|^2 + |Y|^2 - 2|\hat{Y}||Y|\cos\hat{\varphi} \to \ell_{\mathbb{C}}^2 \propto \lambda^2 - 2\lambda\cos\hat{\varphi} + 1.$$

This expression is zero if and only if Y and $\hat{Y}$ are identical, i.e., $\lambda = 1$ and $\hat{\varphi} = 0$. We can observe that the $\ell_{\mathbb{C}}^2$ follows a parabola, depending on the value of λ and $\cos\hat{\varphi}$. For $\hat{\varphi} = 0$, this function is minimal when $\lambda = 1$. However, the minimum of the curve when $\hat{\varphi} \neq 0$ is at $\lambda < 1$ and tends toward $\lambda \to 0$ as $|\hat{\varphi}|$ increases, as shown in fig. 4.2. This shift depending on $\hat{\varphi}$ implies that $\ell_{\mathbb{C}}^2$ underestimates the reconstructed magnitude, assigning a lower loss to estimates with $\lambda < 1$ than to estimates with $\lambda > 1$.

We can show that $\ell_{\mathbb{C}}^2$ is biased towards reconstructions with $\lambda \leq 1$ by examining the gradient of $\ell_{\mathbb{C}}^2$. The loss of $\ell_{\mathbb{C}}^2$ is minimal when $\frac{\ell_{\mathbb{C}}^2}{\partial\lambda} = 0$. Taking the derivative of $\ell_{\mathbb{C}}^2$ for λ yields

$$\frac{\partial\ell_{\mathbb{C}}^2}{\partial\lambda} = 0 \to 2\lambda - 2\cos\hat{\varphi} = 0$$
$$2\lambda = 2\cos\hat{\varphi}$$
$$\lambda = \cos\hat{\varphi}.$$

As $\cos\hat{\varphi} \leq 1$ for every value of $\hat{\varphi}$, the loss function is minimal when $\lambda \leq 1$, and $\lambda = 1$ if and only if $\hat{\varphi} = 0 \pmod{2\pi}$.

This analysis proves that using the $\ell_{\mathbb{C}}^2$ loss function to reconstruct $\hat{\mathbf{Y}}$ typically favors reconstructions where $|\hat{Y}| \leq |Y|$.

4.2.2 Proposed solution

We hypothesize that this bias occurs because the angle and magnitude errors are minimized simultaneously by manipulating the real and imaginary components. We introduce a possible solution by proposing a new loss function that separates the loss in angle and magnitude losses.

We propose representing the phase error between $\hat{Y}$ and Y as $\ell^\perp$, the length of the line segment which is perpendicular to $\hat{Y}$, ending at Y, as visualized in fig. 4.1 as the green line $\ell^\perp(\hat{Y}, Y)$. More specifically, $\ell^\perp$ is the scalar rejection $\ell^\perp = ||Y - \text{proj}_{\hat{Y}} Y||$. This segment is independent of the magnitude error λ and computed as

$$\ell^\perp(Y, \hat{Y}) = \frac{Y \cdot \hat{Y}^T}{|\hat{Y}|}. \tag{4.1}$$

In the complex plane, this function is efficiently computed[184] as

$$\ell^\perp(Y, \hat{Y}) = \frac{|\Re(\hat{Y}) \cdot \Im(Y) - \Im(\hat{Y}) \cdot \Re(Y)|}{|\hat{Y}|}. \tag{4.2}$$

The magnitude error can then be independently represented from $\ell^\perp$ in the full loss function.

The definition of $\ell^\perp$ assumes $\hat{\varphi} \leq \pi/2$ and $\ell^\perp$ decreases when $\hat{\varphi}$ increases beyond $\pi/2$. To ensure $\ell^\perp$ smoothly increases when $\hat{\varphi} > \pi/2$, a smooth continuation of $\ell^\perp$ when $\hat{\varphi} \geq \pi/2$ has been defined in eq. (4.4). Moreover, it was assumed that $\hat{Y}$ and Y have nonzero magnitude, which is not always the case. Therefore, we add $\epsilon = 10^{-8}$ to the denominator of eq. (4.2) during implementation as defined in eq. (4.3).

The complete loss function for complex images $\mathbf{Y}, \hat{\mathbf{Y}} \in \mathbb{C}^{m \times n}$ is then defined as

$$\ell^\perp(\hat{\mathbf{Y}}, \mathbf{Y}) = \frac{1}{mn} \sum_j^m \sum_k^n \frac{|\Re(\hat{Y}_{jk})\Im(Y_{jk}) - \Im(\hat{Y}_{jk})\Re(Y_{jk})|}{|\hat{Y}_{jk}| + \epsilon} \tag{4.3}$$

$$\mathcal{L}^\perp(\hat{\mathbf{Y}}, \mathbf{Y}) = \begin{cases} \ell^\perp(\hat{\mathbf{Y}}, \mathbf{Y}) & |\hat{\varphi}| < \frac{\pi}{2} \\ 2|\mathbf{Y}| - \ell^\perp(\hat{\mathbf{Y}}, \mathbf{Y}) & |\hat{\varphi}| \geq \frac{\pi}{2} \end{cases} \tag{4.4}$$

$$\perp(\hat{\mathbf{Y}}, \mathbf{Y}) = \mathcal{L}^\perp(\hat{\mathbf{Y}}, \mathbf{Y}) + f(|\mathbf{Y}|, |\hat{\mathbf{Y}}|) \tag{4.5}$$

In eq. (4.5), f is the loss function operating on the magnitude part of $\hat{\mathbf{Y}}$ and $\mathbf{Y}$. For example, we can define $\perp + \ell^2$-loss where f is the Frobenius norm of $|\hat{\mathbf{Y}}| - |\mathbf{Y}|$,

or $\perp$+SSIM-loss where $f = 1 - \text{SSIM}(|\hat{\mathbf{Y}}|, |\mathbf{Y}|)$. As $\ell^{\perp}$ is independent of λ and the magnitude term is independent of $\hat{\varphi}$, $\perp$-loss is symmetric and assigns equal loss to vectors with magnitude bias λ and λ^{-1} for the same $\hat{\varphi}$.

$\perp$-loss as defined in eq. (4.5) is currently only defined in the complex domain but can also be applied to different domains. In particular, as $\mathbb{C}$ is isomorphic to $\mathbb{R}^2$, $\perp$-loss may find application in many regression tasks, e.g., image registration[134,183] or fluid flow analysis[185]. Moreover, the concept of the scalar rejection that forms the basis of $\ell^{\perp}$ in eq. (4.1) allows for direct extension to $\mathbb{R}^n$, enabling the application of $\perp$-loss to higher-dimensional problems.

4.3 Methods

First, we examine the symmetry of the loss landscapes of the $\ell_{\mathbb{C}}^1$, $\ell_{\mathbb{C}}^2$ and $\perp$+ℓ^2-loss functions. Subsequently, to study the image reconstruction quality and dataset or model dependence, we apply $\perp$+ℓ^2-loss and $\perp$+SSIM-loss to complex image reconstruction using two different networks and two different datasets. Finally, we study how $\perp$+ℓ^2-loss generalizes to problems defined in the $\mathbb{R}^2$ vector space, focusing on image registration[1].

4.3.1 Loss landscape examination

To verify the symmetry of the loss functions, we have performed a simulation to visualize the loss landscape of $\ell_{\mathbb{C}}^1$, $\ell_{\mathbb{C}}^2$, and $\perp$+ℓ^2-loss loss functions. We generated a vector of $n = 5000$ complex numbers $x \in \mathbb{C}^n$, $x_j = a + bi$ with a and b independently drawn from a uniform distribution between -5 and 5. Next, we perturbed every x_j with every $(\lambda, \hat{\varphi})$-pair generated by the Cartesian product between $\Lambda = \{0, 0.02, 0.04, \ldots, 2\}$ and $\Phi = \{0, \frac{\pi}{100}, \frac{2\pi}{100}, \frac{3\pi}{100}, \ldots, \pi\}$. That is, $\hat{x}_k = \lambda_k x_j e^{i\hat{\varphi}_k}$, $(\lambda_k, \hat{\varphi}_k) \in \Lambda \times \Phi$. This resulted in a vector $\hat{x}$ with 10,000 perturbed complex numbers for every x_j. Finally, we computed the $\ell_{\mathbb{C}}^1$, $\ell_{\mathbb{C}}^2$ and $\perp$+ℓ^2-loss loss between x_j and $\hat{x}$ to obtain a loss value per $(\lambda, \hat{\varphi})$-pair. Calculating this loss for every x_j and taking the mean resulted in the loss landscape of λ and $\hat{\varphi}$. These landscapes were compared for the $\ell_{\mathbb{C}}^1$, $\ell_{\mathbb{C}}^2$ and the $\perp$+ℓ^2-loss loss.

4.3.2 MRI reconstruction

We have trained two deep CNNs on two different datasets to examine the performance of $\perp$-loss when optimizing deep neural networks for undersampled MRI reconstruction. An "end-to-end variational network" (E2E-VarNet)[186] was trained to reconstruct complex MRI from the fastMRI challenge[187] (Experiment A). Also, a recurrent inference machine[188,189] (RIM) was trained to reconstruct

---

1. Code will be made publicly available on `https://gitlab.com/computational-imaging-lab/perp_loss`

complex MRI from the Calgary-Campinas MRI reconstruction challenge[190] (Experiment B).

4.3.2.1 Datasets

The fastMRI dataset[187] is a large, open dataset of knee MRI, providing 34742 MRI slices of 973 volumes for training and 7135 slices of 199 volumes for validation. The MRI was acquired using 15 receiver channels at 1.5T and 3T systems, providing various contrasts at 0.5 mm^2 resolution. The MRI was acquired using an Cartesian 2D turbo spin echo sequence (TE=27-34 ms, TR=2200-3000 ms). Approximately half of the scans were acquired with fat suppression, while the other half did not employ fat suppression. The dataset provided unprocessed, fully-sampled, complex multi-coil k-space for every slice, along with fully-sampled ground-truth images. However, these ground-truth images only contained the magnitude component as they were computed using a root-sum-of-squares coil combination, which precluded using a complex-valued loss function. We computed coil-sensitivity maps (CSMs) and generated fully-sampled complex coil-weighted reconstructions as target images[191]. Finally, the multi-coil k-space was retrospectively undersampled by multiplying it with an equispaced Cartesian undersampling mask (R=4), preserving 8% of the center lines.

The Calgary-Campinas multi-channel MR dataset[190] is an open, 2D brain MRI dataset providing 12-channel k-space of 167 volunteers, acquired using a T_1-weighted gradient-recalled echo sequence (TE=2.6-3.1 ms, TR=6.3-7.4 ms, TI=400-650 ms). The MRI was acquired using a 3T system at 1 mm^3 isotropic resolution. The dataset provided unprocessed, fully-sampled, complex multi-coil k-space for every slice, along with fully-sampled ground-truth images. As with the FastMRI dataset, the ground-truth images only contained the magnitude component as they were computed using a root-sum-of-squares coil combination, which precluded using a complex-valued loss function. Compuation of CSMs allowed generation of fully-sampled complex coil-weighted reconstructions as target images[191]. Data were undersampled using provided random Poisson disc undersampling masks, yielding an acceleration factor R=5. In total, the dataset provided 12,032 slices for training, 5,120 slices for validation, and 7,800 slices as the test set.

4.3.2.2 Architectures

The E2E-VarNet[186] is an unrolled network for undersampled MRI reconstruction consisting of multiple cascades. Each cascade computes

$$k^{t+1} = k^t - \eta^t M(k^t - \hat{k}) + G(k^t)$$

where k^t is the k-space per coil from the previous cascade, M is the sampling mask, $\hat{k}$ is the k-space sampled by the MRI, η is a learnable parameter, and G is a

convolutional neural network operating on the coil-combined, Fourier-transformed k^t. Its output was then again coil-weighted and Fourier-transformed. The CSMs were estimated from the central lines of k-space using a U-Net of depth 4 and 8-channel input. In our case, G is a U-Net of depth 4 with 18 input channels[192]. The entire model consisted of 8 cascades. The reconstructed image was obtained by transforming the coil-combined k-space of the final cascade to image-space. Sriram et al. showed that end-to-end estimation of the CSMs allows higher reconstruction quality than using than using traditional methods to estimate CSMs from the undersampled MRI, as the ground-truth CSMs are not available as these were derived from fully-sampled k-space.

The RIM is a recurrent neural network proposed for inverse problems, such as undersampled MRI reconstruction[188,189]. The network starts with an initial estimate of the of the image x_0, which is the Fourier transform of the undersampled, coil-combined k-space. Then, this estimate is updated such that

$$x_{t+1}, s_{t+1} = x_t + h_\theta \left(\nabla_{y|x_t}, x_t, s_t \right).$$

Here, $\nabla_{y|x_t}$ is the data fidelity term, h_θ is the recurrent neural network parameterized by θ, and s is the internal state of the recurrent neural network. The final reconstruction is given by x_T for some predefined number of steps T.

A RIM was used for complex image reconstruction with $T = 8$ and h_θ is a 2-layer convolutional gated recurrent unit (GRU) with 64 hidden features. Additionally, a U-Net of depth 4 was used to estimate CSMs from the undersampled k-space, as the ground-truth CSMs were not available during training and CSM estimation using deep CNNs showed improved performance compared to traditional methods to estimate CSMs.

4.3.2.3 Experiments

Two experiments were performed: The E2E-VarNet model was used to reconstruct complex undersampled MRI of the fastMRI dataset to examine the performance of $\perp$-loss compared to $\ell_{\mathbb{C}}^2$ and ℓ_{SSIM} (Experiment A), and the RIM was used to reconstruct complex undersampled MRI with added noise of the Calgary-Campinas dataset to examine the noise robustness of $\perp$-loss compared to $\ell_{\mathbb{C}}^2$ and ℓ_{SSIM} (Experiment B).

For every experiment, four separate models were trained:

1. A model trained to minimize the $\ell_{\mathbb{C}}^2$ loss function between the ground truth and the estimated image.
2. A model trained to minimize the $\perp + \ell^2$-loss between the ground truth and the estimated image.

3. A model trained to minimize the function ℓ_{SSIM} loss function between the magnitude ground-truth and the magnitude estimated image, discarding phase information.

4. A model trained to minimize the function $\perp + \mathrm{SSIM} = \ell^{\perp} \cdot \psi + (1 - \psi) \cdot \ell_{\mathrm{SSIM}}$ between the ground-truth and the estimated image. Here, ψ is a learnable parameter between 0 and 1.

Each model was trained on an NVIDIA V100 GPU with 32GB VRAM. The models were trained deterministically with a fixed random seed and identical hyperparameters for a fair comparison. The models from experiment A were trained for 50 epochs using a batch size of 1. At each epoch, 25% of the volumes were randomly sampled from all training data to manage training time, using the Adam optimizer[76] with a learning rate of 10^{-3}. After 40 epochs, the learning rate was reduced to 10^{-4}. The models in experiment B were trained for 200,000 steps with a batch size of 4 using the Adam optimizer. The base learning rate was 10^{-4} and was halved after every 50,000 steps. During training of the models for Experiment B, Gaussian noise $\epsilon \sim \mathcal{N}(0, \xi)$ was added to the sampled k-space points, where $\xi \geq 0$ was drawn from a uniform distribution $\xi \sim \mathcal{U}(0, 0.2 \cdot |k_0|)$, with $|k_0|$ as the magnitude of the central point in k-space.

4.3.2.4   Evaluation

The reconstruction quality of the models was evaluated over the entire field of view using the mean and standard deviation of the peak signal-to-noise ratio (PSNR), SSIM, and visual information fidelity (VIF)[193] metrics of the magnitude estimate to the magnitude target. VIF is a multi-resolution image quality metric based on the mutual information between two images and has been shown to strongly correlate with MRI quality as assessed by radiologists[194]. The VIF between a target image I_{target} and estimated image I_{est} is computed as

$$\mathrm{VIF}(I_{\mathrm{target}}, I_{\mathrm{est}}) = \sum_{j} \frac{\log_{10}\left(C_j\left(I_{\mathrm{target}}\right) \cdot g_j\left(I_{\mathrm{target}}, I_{\mathrm{est}}\right)^2\right)}{\log_{10}\left(C_j\left(I_{\mathrm{target}}\right)\right)}$$

Here, $j = 0, 1 \ldots, J$ is the resolution level, $C_j(I) = 1 + \frac{(I_j)^2 - (I^2)_j}{\sigma_N^2}$ is the information in an image I at resolution level j, and $g_j(I_{\mathrm{target}}, I_{\mathrm{est}}) = \frac{C_j(I_{\mathrm{est}})}{C_j(I_{\mathrm{target}})}$ is the mutual information between I_{target} and I_{est}. The jth sub-band of an image I is approximated by blurring I using a zero-meaned Gaussian kernel with $\sigma^2 \propto 2^{J-j}$ and downsampling the image by factor 2^j. σ_N^2 is a parameter of the vision model and was chosen as 0.4 for MRI images[190]. The VIF is bounded by 0, but can reach values greater than one if the reconstructed image shows less noise or improved contrast compared to the target image.

Besides magnitude quality, we evaluated the mean and standard deviation of the mean-squared error between the estimated phase map and the target phase map.

Wilcoxon signed-rank tests with $\alpha < 0.01$ were performed when comparing the results. During the evaluation of experiment B, Gaussian noise was added to the input data using $\xi \sim \mathcal{U}(0, 0.5 \cdot |k_0|)$, allowing comparison of the loss functions based on image quality depending on the noise level.

4.3.3 Image registration

To investigate whether the proposed $\perp$-loss generalizes beyond MR image reconstruction, we trained a model to learn deformation vector fields (DVFs) from a pair of 2D MR images. A DVF is a vector field $V \in \mathbb{R}^{2 \times m \times n}$, where the first dimension is the displacement in the x-direction, while the second dimension is the displacement in the y-direction. A transformation $f : \mathbb{R}^{2 \times m \times n} \rightarrow \mathbb{C}^{m \times n}$ mapped x-displacements of the DVF to the real part of the complex field and y-displacements of the DVF to the imaginary part of the complex field.

4.3.3.1 Data acquisition and processing

We have used magnitude-only sagittal cine-MRI of 135 patients with abdominal cancer undergoing radiotherapy simulation at our department[134] to train the image registration model on simulated ground-truth deformations. The data were acquired using a two-dimensional Cartesian balanced steady-state free precession (bSSFP) sequence on a 1.5 T MRI scanner (Ingenia MR-RT, Philips, Best, the Netherlands) using 28 receiver channels, TE/TR=1.3/2.8 ms, a flip angle of 50 degrees, a resolution of $1.4mm^2$ and a field of view of $320mm^2$, yielding an acquisition matrix size of 224×224 pixels. In total, 31750 magnitude-only dynamics were collected. The signal intensity over all dynamics was linearly rescaled to [0, 1], clipping the top 99th intensity percentile of the dynamics in a cine-MRI. Images were augmented using random affine transformations (rotation between $[-20, 20]$ degrees, translations between $[-10, 10]\%$ of the image size, scaling between $[75, 125]\%$, and shearing between $[-10, 10]\%$), random horizontal and vertical flips, and cropping to a random region of 224×224 pixels.

The ground-truth deformation was generated using Gryds[87], which generates random, smooth DVFs $\mathcal{D} \in \mathbb{R}^{2 \times m \times n}$ using a B-spline basis from $\mathbb{R}^{2 \times 3 \times 3} \sim \mathcal{N}(0, \mathcal{U}(0.001, 0.025))$. The motion parameters were selected such that the determinant of the Jacobian was higher than 0 everywhere to avoid folding by the DVF, which would have resulted in implausible motion[195]. Warping randomly-chosen cine frames using the ground-truth DVF $\mathcal{D}$ yielded a magnitude reference image, magnitude warped image. The ground-truth DVF was mapped to the complex domain by applying the transformation $f : \mathbb{R}^{2 \times m \times n} \rightarrow \mathbb{C}^{m \times n}$ to $\mathcal{D}$.

4.3.3.2 Model architecture, training, and evaluation

We trained a residual U-Net[196] of depth 5 with two input channels and two output channels, with two residual units per level. Every residual unit consisted of a 3×3 convolution, a two-dimensional instance norm[197], and PReLU non-linear activation[114]. The first convolution used stride 2, while the latter convolution used stride 1. Two variants of this image registration model were trained:

1. One model was trained to minimize the end-point error (EPE), which is equivalent to the $\ell_{\mathbb{C}}^2$ loss function in the $\mathbb{R}^2$-domain. This is the de-facto loss function for training image registration models[107,177].
2. One model was trained to minimize the $\perp + \ell^2$-loss.

The models used the reference and warped image as input to reconstruct the DVF. The models were trained using the AdamW optimizer[198] for 150 epochs with a base learning rate of $1 \cdot 10^{-4}$ and a weight decay of $5 \cdot 10^{-4}$ using a batch size of 16 on an NVIDIA V100 GPU with 32GB VRAM. After this training, the models were fine-tuned for 50 epochs using a learning rate of 10^{-6}.

The models were evaluated on the residual EPE and SSIM after image registration with the estimated DVF. Moreover, the models were evaluated on λ, i.e., the ratio between magnitude overestimation and magnitude underestimation. Wilcoxon signed-rank tests with $\alpha < 0.01$ were performed when comparing the results.

4.4 Results

4.4.1 Loss landscape

The loss landscapes of the $\ell_{\mathbb{C}}^1$, $\ell_{\mathbb{C}}^2$ and $\perp + \ell^2$-loss functions are shown in fig. 4.3. The $\ell_{\mathbb{C}}^1$ and $\ell_{\mathbb{C}}^2$ loss functions both display an asymmetry, assigning a higher loss to vectors with $\lambda > 1$ and a lower loss to vectors with $\lambda < 1$ with the same phase error. For $\perp + \ell^2$-loss, this loss landscape is symmetric with steep gradients far from $\lambda = 1$ and a large region where the loss is low, as can be observed from the size of the area within the first isocontour around $\lambda = 1$ and $\hat{\varphi} = 0$.

4.4.2 Image reconstruction

4.4.2.1 Experiment A: E2E-VarNet

Each E2E-VarNet model was trained in approximately 48 hours. As shown in table 4.1, the proposed $\perp + \ell^2$-loss yields reconstructions with significantly higher SSIM and VIF than $\ell_{\mathbb{C}}^2$ (Wilcoxon, $p \ll 0.01$) – with an SSIM of 0.90 ± 0.07 versus 0.86 ± 0.10 – while the phase error is equally low. Example reconstructions are shown in Figure 4.4, where it can be seen that the $\ell_{\mathbb{C}}^2$ reconstruction shows
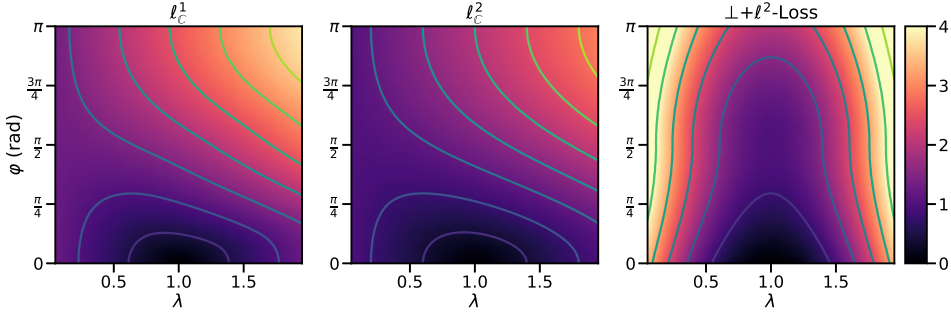
Figure 4.3 The loss landscapes for the $\ell_{\mathbb{C}}^1$ loss function (left), $\ell_{\mathbb{C}}^2$ loss function (center) and $\perp + \ell^2$-loss (right). Lines through the loss landscape are isolines of the losses. The $\ell_{\mathbb{C}}^1$ and $\ell_{\mathbb{C}}^2$ loss functions show an asymmetry, resulting in a higher loss value for $\lambda > 1$, while $\perp + \ell^2$-loss is symmetric.

an apparent magnitude underestimation in the reconstructed knee tissue and a more significant error in the background region. The reconstruction from the $\perp + \ell^2$-loss model shows no clear preference for magnitude overestimation or underestimation. However, less signal is present in the background in this specific instance, indicating better denoising properties.

The model trained to maximize the magnitude SSIM results in even higher magnitude quality but does not reconstruct usable phase information, as the network received no loss for the phase term. Combining the SSIM loss with $\perp$-loss using the $\perp + $SSIM-loss results in reconstructions of equal quality as the SSIM model but has a similar phase reconstruction performance as the $\ell_{\mathbb{C}}^2$ loss and $\perp + \ell^2$-loss. However, there seems to be an increased residual magnitude error compared to the SSIM reconstruction, indicating a trade-off between denoised and dealiased magnitude reconstructions and high-quality phase maps.

4.4.2.2   Experiment B: RIM

Each RIM was trained in approximately 8 hours. Similar to the results of the E2E-VarNet model, using $\perp + \ell^2$-loss yields significantly higher reconstruction quality in the noise-free case than $\ell_{\mathbb{C}}^2$ (Wilcoxon, $p \ll 0.01$) – with SSIM values of 0.89 ± 0.04 versus 0.93 ± 0.03. Additional quantitative results are presented in table 4.1. The model trained to maximize the magnitude SSIM outperforms both the $\perp + \ell^2$-loss and $\ell_{\mathbb{C}}^2$ models, but these reconstructions have high phase error. Based on the VIF metric, the $\perp + $SSIM-loss model outperforms the SSIM model (with a mean VIF of 1.0 ± 0.011 versus 0.99 ± 0.11) while having low phase error.

An example reconstruction without added noise is shown in fig. 4.5. It can be observed that both SSIM and $\perp + $SSIM-loss models yield high-quality magnitude reconstructions. However, the SSIM model reconstructs poor phase images, while

Loss \ Model	E2E-VarNet/FastMRI				RIM/Calgary-Campinas			
	SSIM (↑)	PSNR (↑)	Phase MSE (↓)	VIF (↑)	SSIM (↑)	PSNR (↑)	Phase MSE (↓)	VIF (↑)
$\ell_{\mathbb{C}}^2$	0.86 ± 0.10	31.2 ± 3.1	$\mathbf{0.04 \pm 0.03}$	0.80 ± 0.17	0.89 ± 0.04	31.2 ± 2.5	$\mathbf{0.03 \pm 0.01}$	0.98 ± 0.09
⊥+ℓ^2-loss	0.90 ± 0.07	33.0 ± 3.0	$\mathbf{0.04 \pm 0.03}$	0.83 ± 0.16	0.93 ± 0.03	$\mathbf{31.4 \pm 2.6}$	$\mathbf{0.03 \pm 0.01}$	0.99 ± 0.09
SSIM	$\mathbf{0.91 \pm 0.07}$	$\mathbf{33.9 \pm 3.5}$	0.12 ± 0.03	$\mathbf{0.86 \pm 0.15}$	0.93 ± 0.03	31.0 ± 2.6	0.15 ± 0.01	0.99 ± 0.11
⊥+SSIM-loss	$\mathbf{0.91 \pm 0.07}$	33.1 ± 3.0	$\mathbf{0.04 \pm 0.03}$	$\mathbf{0.86 \pm 0.16}$	0.93 ± 0.03	30.9 ± 2.5	$\mathbf{0.03 \pm 0.01}$	$\mathbf{1.0 \pm 0.11}$

Table 4.1 Evaluation of E2E-VarNet and RIM models trained to minimize $\ell_{\mathbb{C}}^2$, ⊥+ℓ^2-loss, SSIM or ⊥+SSIM-loss functions using the structural similarity, peak-signal-to-noise ratio, phase mean-squared error, and VIF metrics. Best results per model and metric are marked in boldface.

the phase images of the ⊥+SSIM-loss model are similar to the ground-truth phase maps.

When evaluated with added noise, i.e., $\xi > 0$, the quality of the magnitude reconstructions degrade for all models. In fig. 4.6, the performance of the four RIM models with added noise is evaluated based on the magnitude PSNR metric. Here, in the noise-free case, ⊥+ℓ^2-loss achieves the highest median PSNR of 32.8. However, the models were trained up to a noise level of 0.2. Up to this noise level, ⊥+ℓ^2-loss achieves the highest median PSNR. Beyond this noise level, the image quality of ⊥+ℓ^2-loss degrades faster than the competing models, yielding a median PSNR of 22.8 at noise level 0.5 versus 23.1 for the $\ell_{\mathbb{C}}^2$ model. The ⊥+SSIM-loss model performs similarly to the SSIM model, which both yield an median PSNR of 32.1 in the noise-free case and 23.3 at noise-level 0.5.

The model trained to maximize the magnitude SSIM shows high phase error for every noise level and produced unusable phase maps. For the $\ell_{\mathbb{C}}^2$ models, the phase error slowly increased as the added noise increases. For the ⊥+ℓ^2-loss and ⊥+ℓ^2-loss, the phase error remained near-constant as the noise level was increased, showing superior phase reconstruction for all noise levels.

4.4.3 Image registration

Each image registration model was trained in approximately 6 hours. An example comparing the DVFs of the ⊥+ℓ^2-loss and EPE models to the ground-truth is shown in fig. 4.7, where the model trained to minimize ⊥+ℓ^2-loss produces a lower residual error and a lower registration error, with a registration SSIM of 0.931 for the ⊥+ℓ^2-loss model versus 0.768 for the EPE model. Moreover, it can be seen from the difference DVF that the EPE model yields a relative magnitude underestimation of up to 25% compared to the target DVF in regions with significant deformation. Quantitative evaluation of both models demonstrates that minimizing the ⊥+ℓ^2-loss produces significantly better DVFs, reducing the mean EPE from 1.39 mm to 0.89 mm (Wilcoxon, $p \ll 0.01$), as presented in table 4.2 Moreover, the mean registered SSIM increased by approximately 0.03 (Wilcoxon, $p \ll 0.01$), and the mean value λ is significantly closer to 1. This indicates
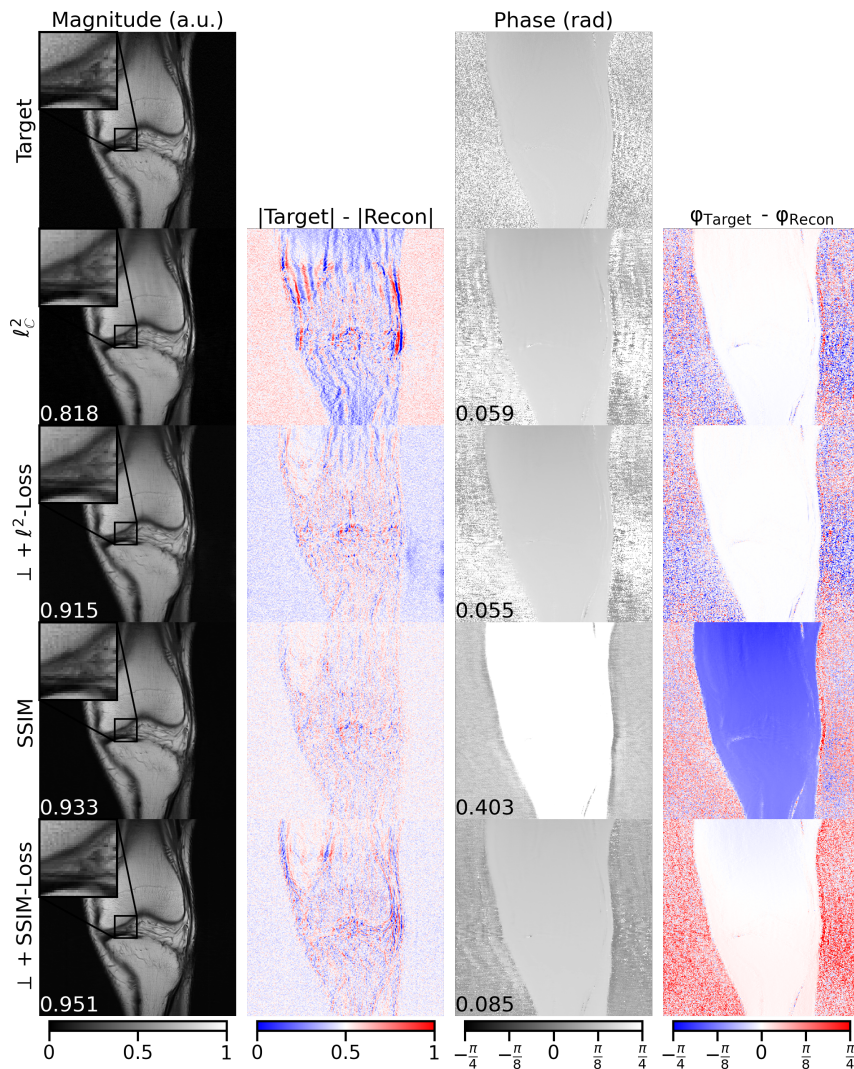
Chapter 4

Figure 4.4 Example reconstructions using an E2E-VarNet (R=4). Examples of reconstructions of the same slice with normalized magnitude using four networks, trained to minimize the $\ell_{\mathbb{C}}^2$, $\perp+\ell^2$-loss, SSIM-Loss, and $\perp+$SSIM-loss loss, respectively. The bottom-left number in magnitude reconstructions shows the foreground VIF (magnitude) while showing the foreground mean squared error for phase images. Using the $\perp+\ell^2$-loss instead of the $\ell_{\mathbb{C}}^2$ loss significantly improves image quality. In the zoomed region of the $\ell_{\mathbb{C}}^2$ model, a hallucinated lesion is visible and is much less severe using the $\perp+\ell^2$-loss model. Using the SSIM as a loss function yields even higher magnitude image quality and higher phase error than the $\ell_{\mathbb{C}}^2$ and $\perp+\ell^2$-loss models. Using the $\perp+$SSIM-loss function obtains the highest magnitude image quality with low phase error. In the zoomed region, higher image quality and more contrast for the $\perp+$SSIM-loss model can be observed compared to the SSIM model.
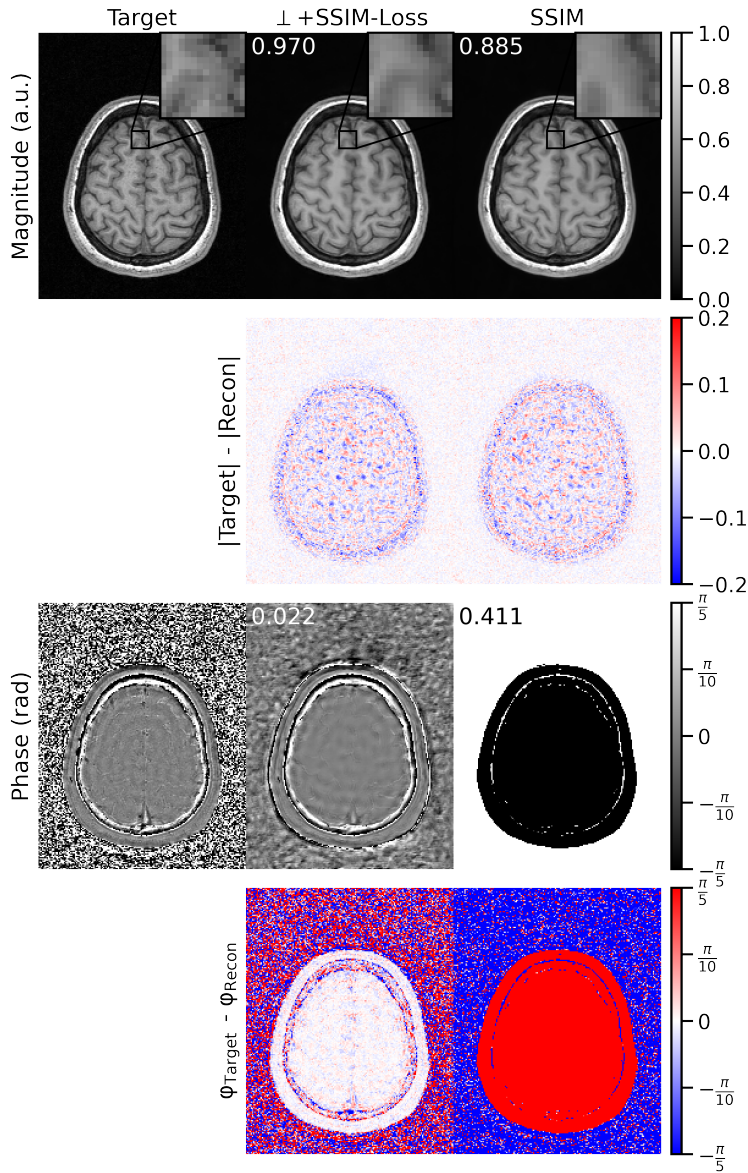
Figure 4.5 Example reconstructions of the RIM (R=5). Examples of reconstructions with the normalized magnitude of the same slice, using a RIM trained to minimize the ⊥+SSIM-loss and SSIM-Loss, respectively. Using ⊥+SSIM-loss as a loss function produces significantly higher-quality phase information (Indicated by the phase MSE, shown in the top-left corner of the phase images) and similar magnitude reconstructions (Indicated by the VIF on the magnitude images). The zoomed region indicates an artifact in the SSIM reconstruction.
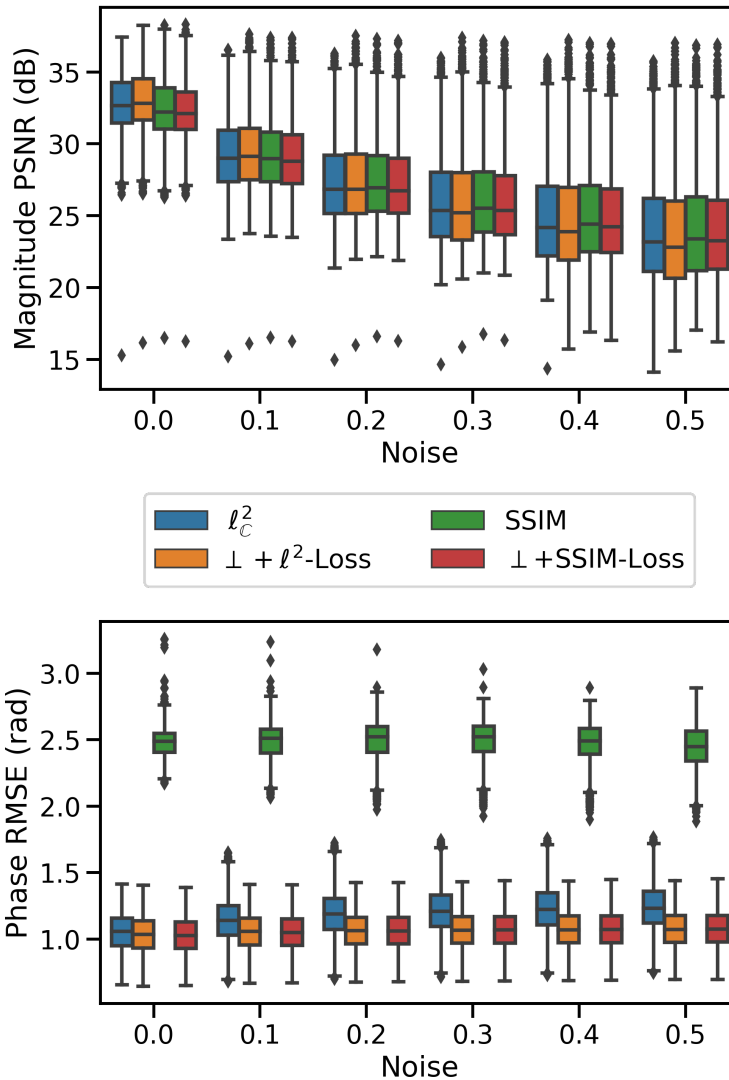
Figure 4.6 Quantitative RIM results. The PSNR of the magnitude reconstructions (top) and the root-mean square error of the reconstructed phase (bottom) were evaluated for the four RIM models trained with different loss functions. On the horizontal axis is the magnitude of the added noise. In the low-noise regime, $\perp+\ell^2$-loss achieves a median PSNR of 32.8 versus 32.6 for $\ell^2_{\mathbb{C}}$. The PSNR of $\perp+\ell^2$-loss decreases faster at higher noise levels, while the $\perp+$SSIM-loss model achieves a similar PSNR as the SSIM RIM. The SSIM model show very high phase error, while the $\perp$-loss models show superior phase reconstruction compared to the $\ell^2_{\mathbb{C}}$ and SSIM models.

	SSIM ($\uparrow$)	EPE (mm, $\downarrow$)	λ (≈ 1)
$\perp + \ell^2$-loss	**0.927 ± 0.053**	**0.898 ± 0.598**	**0.960 ± 0.304**
EPE	0.895 ± 0.071	1.391 ± 0.704	0.720 ± 0.632

Table 4.2 Evaluation of image registration models trained to minimize $\perp + \ell^2$-loss or the EPE loss functions using the registered SSIM, remaining EPE, and λ, the ratio between magnitude overestimation and magnitude underestimation. Best results per metric are marked in boldface.

that the magnitude error is more symmetrically-distributed, i.e., magnitude underestimation is about as likely as magnitude overestimation for the $\perp + \ell^2$-loss model. The error is significantly more biased towards magnitude underestimations for the EPE model than the $\perp + \ell^2$-loss model (Wilcoxon, $p < 0.01$).

4.5 Discussion

In this work, we have identified that the $\ell^2_{\mathbb{C}}$ loss function exhibits a magnitude bias when applied to regression in the $\mathbb{C}$ and $\mathbb{R}^2$ domains, which may impact the performance of deep neural networks. Specifically, for the reconstruction and processing of complex MRI, we have shown this bias is detrimental to the performance of CNNs for complex MRI reconstruction and MRI registration. To address this issue, we have proposed a new loss function called $\perp$-loss, which produces a symmetric loss space when applied to regression in the $\mathbb{C}$ and $\mathbb{R}^2$ domains.

We have shown that $\perp + \ell^2$-loss reconstructs undersampled MRI with higher quality than models trained to minimize the $\ell^2_{\mathbb{C}}$ loss. A hybrid $\perp$+SSIM-loss function allowed image reconstruction with similar image quality as networks trained to maximize the magnitude SSIM while generating high-quality phase maps. The E2E-VarNet model trained to minimize $\perp + \ell^2$-loss as loss function increased the magnitude SSIM by approximately 0.04 and the PSNR by approximately 1.8 dB compared to the $\ell^2_{\mathbb{C}}$ loss. Using $\perp + \ell^2$-loss as the loss function for the RIM model increased the magnitude SSIM by approximately 0.04 and the VIF by approximately 0.01 compared to the $\ell^2_{\mathbb{C}}$ loss.

These models trained to optimize for $\perp + \ell^2$-loss display competing performance to state-of-the-art image reconstruction models while achieving high-quality phase reconstruction. However, some literature currently reports higher scores than the values we reported here. For example, an SSIM of 0.930 and PSNR of 40 when reconstructing the fastMRI "challenge" dataset using an E2E-VarNet model has been reported[186]. On the validation dataset, they reported an SSIM of 0.923 whereas we found an SSIM of 0.91 ± 0.07. This difference in performance could be caused by multiple reasons, such as a better training regime with larger batch
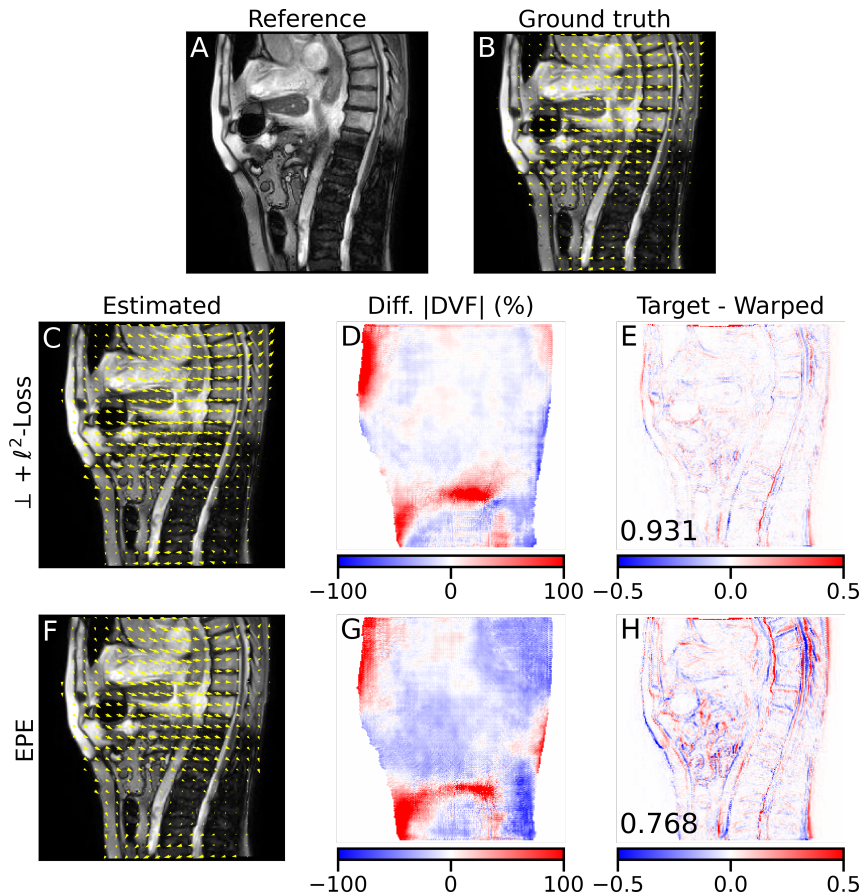
Figure 4.7 Image registration results. The reference image (A), warped image (B) were input to the models. The ground-truth DVF is also shown in (B), but was not presented to the model. Figures C and F show the estimates produced by the $\perp+\ell^2$-loss and EPE models, respectively. Figures D and G show the relative magnitude error between the produced DVF and target DVF for the $\perp+\ell^2$-loss and EPE models, respectively. Finally, figures E and H show the normalized intensity difference (a.u.) between the images registered by the $\perp+\ell^2$-loss and EPE DVF and the ground-truth image, respectively. In the bottom-left of C and H, the SSIM between the estimated registration and the ground-truth registration is shown. The EPE model shows a larger registration error (H), as visible by larger intensities in the difference image and the lower registration SSIM. Also, the estimated DVF by the EPE model shows a larger negative error (G), which indicates a magnitude underestimation, while the $\perp+\ell^2$-loss shows errors closer to zero.

sizes, longer training times, larger models, or our training with complex target images instead of standard magnitude-only training.

These results indicate that deep learning models can use phase information from the complex signal to improve magnitude reconstructions. This observation aligns with other works, which found that including phase information or designing a network that uses complex-valued model weights – thus taking full advantage of the complex information present in the data – improves both magnitude and phase reconstruction performance[199,200]. Moreover, improved phase reconstruction might enable deep learning applications, such as quantitative susceptibility mapping (QSM)[201] or four-dimensional flow MRI[202]. We speculate that reconstructing phase information makes the inversion problem less ill-posed as $\ell^\perp$ acts as a phase regularizer.

Finally, we demonstrated that ⊥-loss is not limited to complex image reconstruction but can be applied to problems in other domains. In particular, we have shown that using $\perp + \ell^2$-loss as the loss function outperforms the EPE when applied to an image registration task. The DVFs are estimated with lower residual error, reducing from approximately 1.4 mm to 0.9 mm, better image registration, and a more symmetrically distributed error.

In past literature[179,183], it has been established that applying least squares regression to the $\mathbb{R}^2$ domain in the presence of noise yields a magnitude underestimation. Our findings align with these previous results, and we have shown that these issues also apply to the complex domain in general and occur during non-linear regression.

Currently, we have only used real-valued networks for comparison. Recently, networks with complex-valued weights have been proposed, and it has been shown that these models achieve superior performance[200]. Future work might investigate whether these complex-valued networks could enable even better performance with ⊥-loss, as every part from input to output is defined in the complex plane.

We have identified that it is beneficial to have a smooth and symmetric loss function for MR image reconstruction and image registration. However, it is crucial to consider the domain of the problem, and these qualities do not necessarily transfer to other tasks. For example, it has been identified that non-convex loss functions achieve better performance for image classification tasks[203] while convex loss functions are generally preferred, and for anomaly detection, where it is beneficial to use an asymmetric loss function that penalizes outliers harder than inliers[204].

A symmetrically-distributed error in image registration is an attractive property for many applications. For example, in radiotherapy, accurate tracking of tumor motion enables real-time adaptive treatments, where the position and shape of the radiation beam are adapted to the current anatomy[45]. A biased image registration algorithm might underdose a tumor while risking increased toxicity to increased

dose delivery to healthy tissue[205]. With a symmetric loss function like $\perp$-loss, a more conformal dose delivery could be obtained.

For future work, it would be helpful to further study the properties of $\perp$-loss. For example, it would be interesting to analyze further the noise robustness of $\perp$-loss to conventional loss functions. Currently, $\perp+\ell^2$-loss does not outperform $\ell^2_{\mathbb{C}}$ in the magnitude domain when subjected to more noise than seen during training. In the phase domain, however, the performance of $\perp+\ell^2$-loss and $\perp+$SSIM-loss is minimally perturbed by the added noise. The cause of this phenomenon is unknown, but we speculate this might be caused by a low gradient in the phase direction in high-noise environments when $\lambda \approx 1$ (fig. 4.3). Moreover, the loss function consists of weighing between the magnitude term and phase term. This parameter is currently optimized as part of the network, while more manual tuning could change the emphasis of the loss between the domains. Further experiments could investigate the performance of $\perp$-loss in a low SNR regime or at higher undersampling factors.

For image registration, it would be interesting to extend the definition of $\perp$-loss to three dimensions. It is simple to express the scalar rejection as defined in eq. (4.1) in $\mathbb{R}^n$, thus facilitating extension to $\mathbb{R}^3$ and beyond. Further analysis could establish whether the attractive properties of the $\perp$-loss function in the complex plane, e.g., symmetric output or higher image registration performance, translates to higher dimensions.

Finally, it would be interesting to apply $\perp$-loss to non-deep learning image reconstruction. For example, $\perp$-loss could be used as a cost function in an iterative compressed sense algorithm. Future research could investigate whether using $\perp+\ell^2$-loss or $\perp+$SSIM-loss could improve image reconstruction or result in faster iteration convergence.

As using $\perp+$SSIM-loss allows for better image reconstruction model training, further undersampling of k-space with equal image quality could be achieved by such models. This would increase MRI efficiency, increasing patient throughput. Moreover, reconstructing images with higher quality could lead to better diagnosis by radiologists.

4.6   Conclusion

We have identified that the conventional $\ell^2_{\mathbb{C}}$ loss function gives rise to an asymmetric loss landscape in the complex field, resulting in an underestimation bias of the reconstructed magnitude.

To resolve this problem, we have presented $\perp$-loss, a loss function defined in a complex vector space for MRI reconstruction and image registration with

a symmetric magnitude/phase loss landscape. We have applied $\perp+\ell^2$-loss to undersampled complex MR image reconstruction, obtaining higher-quality reconstructions than when minimizing the $\ell^2_{\mathbb{C}}$ loss. Compared to state-of-the-art models that maximize the SSIM, we achieved competitive performance in the magnitude domain and superior performance for phase reconstruction using $\perp+$SSIM-loss.

Finally, we showed that $\perp+\ell^2$-loss generalized beyond complex image reconstruction and could be applied for image registration. We achieved higher performance than models minimizing the end-point error, with lower residual error, better image registration performance, and a symmetric output error.

4.7   Acknowledgments

Chapter 4

CHAPTER 5

Accelerated respiratory-resolved 4D-MRI with separable spatio-temporal neural networks

ML Terpstra
M Maspero
JJC Verhoeff
CAT van den Berg

Abstract

Background: Respiratory-resolved four-dimensional magnetic resonance imaging (4D-MRI) provides essential motion information for accurate radiation treatments of mobile tumors. However, obtaining high-quality 4D-MRI suffers from long acquisition and reconstruction times.

Purpose: To develop a deep learning architecture to quickly acquire and reconstruct high-quality 4D-MRI, enabling accurate motion quantification for MRI-guided radiotherapy.

Methods: A small convolutional neural network called MODEST is proposed to reconstruct 4D-MRI by performing a spatial and temporal decomposition, omitting the need for 4D convolutions to use all the spatio-temporal information present in 4D-MRI. This network is trained on undersampled 4D-MRI after respiratory binning to reconstruct high-quality 4D-MRI obtained by compressed sensing reconstruction. The network is trained, validated, and tested on 4D-MRI of 28 lung cancer patients acquired with a T1-weighted golden-angle radial stack-of-stars sequence. The 4D-MRI of 18, 5, and 5 patients were used for training, validation, and testing. Network performances are evaluated on image quality measured by the structural similarity index (SSIM) and motion consistency by comparing the position of the lung-liver interface on undersampled 4D-MRI before and after respiratory binning. The network is compared to conventional architectures such as a U-Net, which has 30 times more trainable parameters.

Results: MODEST can reconstruct high-quality 4D-MRI with higher image quality than a U-Net, despite a thirty-fold reduction in trainable parameters. High-quality 4D-MRI can be obtained using MODEST in approximately 2.5 minutes, including acquisition, processing, and reconstruction.

Conclusion: High-quality accelerated 4D-MRI can be obtained using MODEST, which is particularly interesting for MRI-guided radiotherapy.

5.1   Introduction

Respiratory motion poses a significant challenge in abdominal and thoracic imaging, causing large displacements in the liver[206], lung[207], kidney[208], and pancreas[209], introducing disruptive image artifacts that may preclude an accurate diagnosis[210,211].   In radiation therapy, respiratory-induced motion can lead to sub-optimal treatment because it may influence the shape and position of tumors[49,212]. Consequently, the target may receive a different dose than planned while delivering hazardous radiation to nearby healthy tissue and organs-at-risk[213]. In the past, respiratory-resolved imaging has been proposed to improve treatments, using imaging with high spatial resolution and accurate motion information to enable the definition of treatment margins that encompass the tumor displacement[214,215].   In particular, four-dimensional respiratory-resolved computed tomography (4D-CT) is the standard imaging modality in current clinical practice and is part of radiation treatment planning[216]. However, 4D-CT can be affected by artifacts that negatively influence the treatment outcome and local control[217,218].

Recently, magnetic resonance imaging (MRI) has been proposed as an alternative to CT for radiotherapy guidance, leveraging the superior soft-tissue contrast that facilitates accurate target identification and dose deposition. With the clinical introduction of MRI-guided radiotherapy (MRIgRT)[25,30], MRI acquired prior to treatment can be used to adapt the treatment plan to the daily anatomy, while fast MRI during treatment can be used to track the tumor position[45,97,125,157,171,219].

In MRIgRT, respiratory-resolved four-dimensional MRI (4D-MRI) is used in the treatment planning phase to adapt the radiation treatment based on the quantified tumor motion[16]. The 4D-MRI must be high-quality and quickly available to ensure treatment efficiency and patient comfort, i.e., acquired and reconstructed within five minutes[220]. However, obtaining high-quality 4D-MRI remains challenging due to the limited acquisition speed of MRI.

A straightforward way to accelerate MRI is by undersampling the acquisition, violating the Shannon-Nyquist data sufficiency criterion[53], and introducing image artifacts that may preclude accurate motion quantification[220]. Several techniques have been proposed to reconstruct high-quality MRI from undersampled acquisitions, such as parallel imaging[67,68], simultaneous multi-slice acquisitions[221–223], or compressed sensing[71].   Some algorithms have been specifically developed to reconstruct high-quality respiratory-resolved 4D-MRI by taking advantage of all spatio-temporal information in the images, such as XD-GRASP[137] or HDTV-MoCo[160].   However, these reconstruction algorithms have a large computational cost and can take from 15 minutes up to 8 hours[16,160], which is insufficient in clinical practice as long treatment times are detrimental to patient comfort and treatment efficiency.

Recently, convolutional neural networks (CNNs) have been proposed as a data-driven alternative to classic iterative algorithms to reconstruct undersampled MRI quickly[89,90,94,186,224]. With CNNs, the time-consuming model training can be performed offline before treatment. Then, the trained model can be used for fast, online inference, achieving reconstruction quality on par or better than compressed sensing within tens of milliseconds for 2D imaging[225].

Training such models requires large amounts of GPU memory to optimize the model parameters. As GPU memory is limited, training CNN-based reconstruction models is feasible for 2D and 3D MRI but challenging for 4D-MRI as these models require prohibitively costly four-dimensional convolutions to take advantage of the spatio-temporal information and obtain high image and motion quality. Several approaches have been proposed to avoid using 4D convolutions, e.g., by performing slice-by-slice reconstruction or carefully using multiple views of the spatio-temporal data[133,226–228]. However, training such models to obtain high-quality 4D-MRI remains challenging due to the computational cost or requirement for large datasets.

We propose an unrolled model to reconstruct 4D-**M**RI using l**o**w-**di**m**e**nsional **s**ubne**t**works (MODEST), which exploits the spatio-temporal nature of 4D-MRI by separating the reconstruction problem into spatial and temporal components. Two independent subnetworks with few trainable parameters have been designed to learn these components without using 4D convolutional kernels. This allows the model to access the complete spatio-temporal information in 4D-MRI while maintaining low computational cost.

This work investigates the application of the proposed spatio-temporal decomposed network to accelerate the acquisition and reconstruction of undersampled 4D respiratory-resolved lung MRI, which is of particular interest for MRI-guided radiation treatments. The model is evaluated on reconstructed image quality and consistency of the respiratory motion compared to compressed sensing reconstructions. Moreover, MODEST is compared to standard deep learning architectures such as a U-Net. Finally, we estimate the minimum acquisition length for high-quality 4D-MRI with MODEST.

5.2 Methods

We considered two networks to reconstruct 4D-MRI: a baseline residual U-Net, and our newly proposed architecture. After patient data was collected and pre-processed, the model hyperparameters were optimized. Then, the U-Net and MODEST were trained. To investigate the impact of the model architecture rather than the number of trainable parameters, the optimized parameters of the U-Net were pruned to match MODEST. The three models (MODEST, the

baseline U-Net, and pruned U-Net) were evaluated using undersampled 4D-MRI before and after respiratory binning.

5.2.1 Patient data collection and preparation

Twenty-eight patients undergoing radiotherapy for lung cancer between February 2019 and February 2020 at the radiotherapy department were retrospectively included under the approval of the local medical ethical committee with protocol number 20-519/C. The male/female ratio was 16/12, and the mean age was 66 ± 13 years (range = 20-81). Patients affected by squamous cell carcinoma (11), adenoma & adenocarcinoma (7), small cell/large cell carcinoma (4), neoplasm (1), thymoma (1), and a mix of other rare tumors (4) were included.

Free-breathing 3D golden-angle radial stack-of-stars (GA-SOS) T_1-weighted spoiled gradient echo MRI (TR/TE=3.2/1.3 ms, FA=8°, bandwidth=866Hz/px, resolution=2.13 × 2.13 × 3.5 mm³, FOV=440 × 440 × 270 mm³, feet-head slices) of the thorax were acquired for 7 min on a 1.5T MRI (MR-RT Philips Healthcare, Best, the Netherlands) during gadolinium injection (Gadovist, 0.1 ml/kg). The acquisition was fat-suppressed using spectral attenuated inversion recovery (SPAIR).

Patients were scanned in the supine position using a 16-channel anterior and 12-channel posterior phased-array coil. In total, 1312 radial spokes per slice were acquired, corresponding to approximately four times oversampling compared to a fully-sampled volume, which requires $206 \cdot \pi/2 \approx 324$ spokes. However, as the contrast agent was injected, the relative magnitude of the self-navigation signal changed over time. To account for the contrast pickup phase, we discarded the first 200 spokes of every scan to prevent contrast mixing.

For every patient, 4D-MRI was created based on a self-navigation signal by sorting k-space into ten respiratory-correlated bins for a final matrix size of $V_x, V_y, n_{\text{slice}}, n_{\text{phase}} = 206 \times 206 \times 77 \times 10$. The self-navigation signal was obtained by performing a 1D Fourier transform of the center of k-space (i.e., k_0) along the slice direction and principal component analysis on the concatenated navigators[136,137]. Then, radial spokes were sorted into respiratory bins using a hybrid binning algorithm[138] based on the phase and relative amplitude of the motion surrogate. For training purposes, undersampled 4D-MRI was obtained by undersampling the respiratory bins, i.e., "phase undersampling", ensuring motion consistency between the target reconstruction and undersampled MRI. The fully-sampled 4D-MRI contained n spokes per bin for every patient. Phase-undersampled 4D-MRI was created by retaining the first n/k spokes per bin, where $k \in \mathbb{N}$ is the acceleration factor, for undersampling factors R_{4D} = 1, 2, and 4. This corresponded to a true undersampling factor R_{Nyquist} of approximately 3.7, 7.4, and 14.8 per respiratory phase, respectively. After sorting, k-space was density-compensated using a Ram-Lak filter, interpolated

onto a twice-oversampled Cartesian grid using a 3×3 Kaiser-Bessel kernel, and transformed to image-space using a non-uniform fast Fourier transform (NUFFT)[61,139] with a weighted coil combination. Coil sensitivity maps were estimated using ESPiRIT[69]. The patients were randomly split into a train (18), validation (5), and test (5). The training target was generated by performing an XD-GRASP reconstruction of the fully-sampled 4D-MRI using temporal total variation, using a regularization weight $\lambda = 0.03$[71,137].

To match the effect of a shorter acquisition time, we have also created undersampled 4D-MRI by removing spokes prior to respiratory binning and discarding the final j sampled spokes, with $j \in \{100, 200, \ldots, 1000\}$, i.e., "free-breathing undersampling". These reconstructions were used to estimate the maximum achievable undersampling factor in a clinical setting, comparing the motion consistency of the free-breathing undersampled 4D-MRI to the fully-sampled reconstruction. We selected the maximum value of j where the zero-filled reconstruction has a mean EPE < 1 mm and the mean SSIM of MODEST was > 0.85.

5.2.2 Model architectures

We propose MODEST, which uses two subnetworks to learn the spatial and temporal features[1]. We trained a network to reconstruct 4D-MRI on a per slice basis rather than per volume to reduce memory usage, which allowed using 2D convolutions. The model input consisted of the zero-filled undersampled 4D-MRI and deformation vector fields (DVFs) computed on zero-filled, undersampled 4D-MRI, registering the exhale phase to every other respiratory phase. The DVFs were obtained using a deep learning model[134]. They were added as additional input as we hypothesize that adding DVFs improves the reconstruction performance as they provide additional spatial information when considering the respiratory phase dimension. To reconstruct a $V_x \times V_y \times n_{\mathrm{phase}}$ volume, the subnetwork learning the spatial component $\hat{\Xi}$ was implemented using $k \times k \times 1$ convolution kernels, while the network learning the temporal component $\hat{\Psi}$ was implemented using $1 \times 1 \times n_{\mathrm{phase}}$ convolutions. Both subnetworks used five convolutional layers and a cardioid non-linear activation function[229]. The model hyperparameters and architecture were optimized using Bayesian optimization. Details for this optimization are provided in Supplementary Document 1. An estimate of the 4D-MRI is then obtained as $f(\hat{\Xi}, \hat{\Psi})$, using some combination function f, which was chosen as the point-wise multiplication operator. We implemented the model to perform an unrolled optimization using three iterations. Data consistency was enforced between the reconstructed image and the sampled k-space after every iteration except the final iteration by computing

$$x^{t+1} = x^t - \eta \mathcal{F}^{-1} \left(\mathcal{F}\left(x^t \right) - \mathbf{y} \right) + \mathrm{M}_t \left(x^t \right), \tag{5.1}$$

---

1. Code available at `https://gitlab.com/computational-imaging-lab/modest`

where t is the iteration, x^t is the image at iteration t, $\mathbf{y}$ is the measured, undersampled radial k-space, $\mathcal{F}$ is the multi-coil non-uniform Fourier transform operator, η is a learned parameter, and M_t is the deep learning model for iteration t of the unrolled model. The model architecture is illustrated in Figure 5.1 and had 312,782 trainable parameters. To investigate the impact of data consistency and adding DVFs as model input, we have trained four variants of MODEST: a variant that only uses the zero-filled 4D-MRI, a variant that uses 4D-MRI and DVFs, a variant that uses 4D-MRI and data consistency, and a variant that uses 4D-MRI, DVFs, and data consistency.

MODEST was compared to a baseline residual U-Net[196,230] that reconstructs 4D-MRI from the undersampled images, where every residual unit consisted of a 3D convolution layer, followed by a PReLU non-linear activation, instance normalization, and a residual connection. The residual U-Net consisted of four resolution levels and five residual units per resolution level. Depending on the resolution level, the residual unit's convolution layers learned 32, 64, 128, and 256 filters. The residual U-Net had 11,793,289 trainable parameters. The model architecture and hyperparameters were found after a Bayesian hyperparameter search. Details for this optimization are provided in Supplementary Document 1.

5.2.2.1 *Training and evaluation*

Both the residual U-Net and MODEST were implemented using PyTorch 1.10. The data consistency operator was implemented using TorchKbNUFFT 1.3.0[231]. The U-Net and MODEST with optimized hyperparameters and architectures were trained on phase-undersampled MRI to reconstruct XD-GRASP 4D-MRI from zero-filled undersampled 4D-MRI. Both models were trained using 20,000 randomly-sampled batches of zero-filled 4D-MRI with undersampling factors R_{4D} = 1, 2, and 4 to minimize the $\perp$ +SSIM-loss[232]. In total, 18 patients · 77 slices · 3 undersampling factors = 4185 samples were used for training, and 1155 samples were used for testing and validation, respectively. MODEST was trained using a batch size of 7 with the AdamW optimizer using a learning rate of 10^{-3} and 10^{-4} weight decay. The baseline residual U-Net was trained using a batch size of 3 using the AdamW optimizer using a learning rate of 10^{-3} and 10^{-4} weight decay. To investigate the impact of the model architecture rather than the number of trainable parameters, we performed iterative pruning of the trained U-Net model (Pruned U-Net), matching the number of parameters of MODEST[233].

The model reconstructions were evaluated on image quality, sharpness, motion quality, and processing time. The image quality was measured by the average SSIM and the normalized root-mean-square error (NRMSE) over the respiratory phases between the model reconstruction and the XD-GRASP reconstruction. The NRMSE was computed as $\mathrm{NRMSE}\,(I_{\mathrm{est}}, I_{\mathrm{target}})$ = $\sqrt{1/M \sum (I_{\mathrm{est}} - I_{\mathrm{target}})^2} / \overline{|I_{\mathrm{target}}|}$, where M is the number of voxels and $\overline{|I_{\mathrm{target}}|}$ is
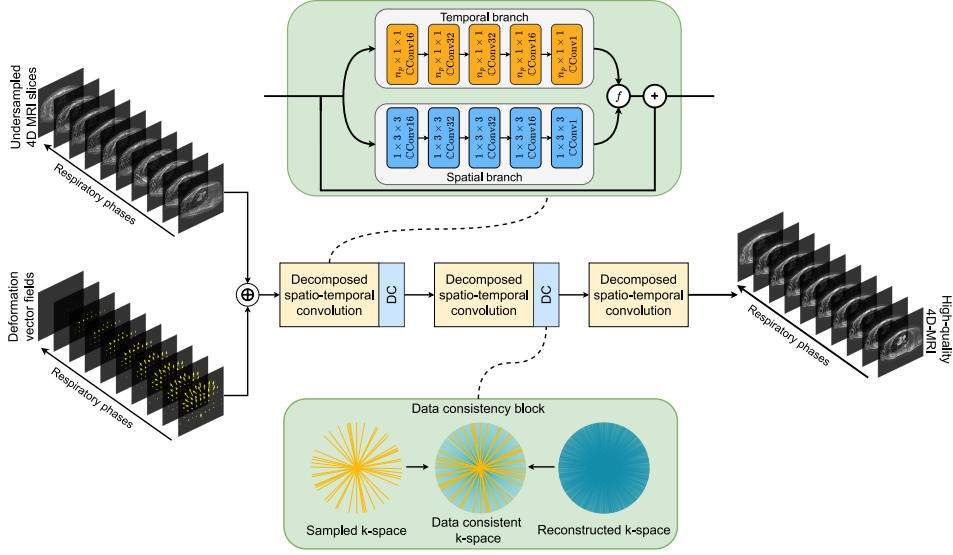
Figure 5.1 Illustration of the proposed MODEST model. The unrolled model reconstructs undersampled 4D-MRI into high-quality 4D-MRI. The undersampled, zero-filled 4D-MRI and deformation vector fields derived from the undersampled 4D-MRI are concatenated and enter a decomposed spatio-temporal convolution block with 104,260 parameters. The spatio-temporal convolution block performs low-dimensional convolution over the spatial domain (blue) and the temporal domain (orange), recombining into a 4D-MRI using a combination function f. After every iteration of the unrolled model, data consistency is enforced on the reconstructed radial k-space using the sampled radial k-space using Equation 5.1.

the mean absolute value of I_{target} within the anatomy[234]. The motion estimation quality was quantified in two ways:

1. DVFs based on XD-GRASP reconstructions and the deep learning reconstructions were estimated using a neural network trained on undersampled MRI[134], registering the first respiratory phase (exhale) to every other respiratory phase. The motion error was then quantified as the mean end-point error (EPE).

2. The position of the hepatic dome in the reconstruction was compared to the hepatic dome position in the ground-truth XD-GRASP reconstruction. The hepatic dome position was manually extracted by computing the median intensity along the AP direction and thresholding the gradient image[235]. Then, the liver position was estimated for every dynamic as the mean of the binary thresholded image along the LR direction within a manually delineated region, ensuring a similar delineation volume among the patients in the test set. The hepatic dome position was normalized by subtracting
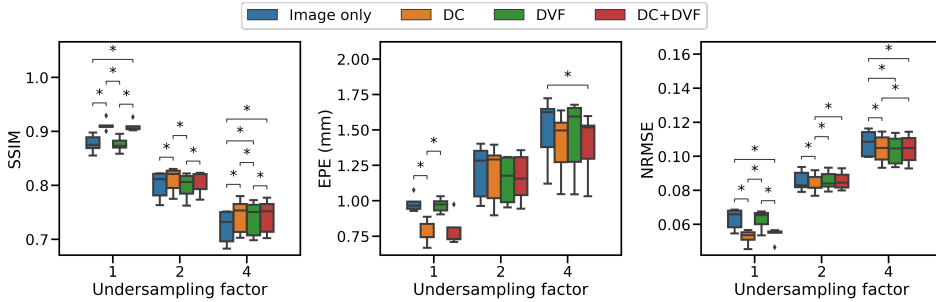
Figure 5.2 Impact of data consistency and DVFs. Four models were compared on the SSIM, registration error, and NRMSE in the foreground for the reconstructed 4D-MRI. The models used only the 4D-MRI, 4D-MRI and data consistency, 4D-MRI and DVFs, or all information to reconstruct the data. A star indicates a statistically significant result $p < 0.05$.

the position of the hepatic dome in the free-breathing zero-filled acquisition. Finally, the error was determined as the absolute error between the hepatic dome of XD-GRASP reconstructions and MODEST.

The image sharpness was evaluated over all the 4D-MRI phases by computing the variance of one 3D respiratory phase after convolution with a 3D Laplacian kernel[236]. The final sharpness was estimated as the mean variance over all respiratory phases. Sharper images have a higher variance.

The metrics' statistical significance ($p < 0.05$) was established using a paired t-test, comparing MODEST to the U-Net and parameter-pruned U-Net.

5.3 Results

Based on the model architecture and hyperparameter search, we found that adding non-Cartesian data consistency and motion information increased the reconstruction quality, as shown in Figure 5.2. Using data consistency increased the validation SSIM from 0.88 ± 0.04 to 0.90 ± 0.04 ($p = 10^{-6}$), while adding DVFs did not significantly improve the SSIM compared to image-only reconstruction or in addition to using data consistency. However, using DVFs decreased the mean EPE from 1.23 ± 0.28 mm to 1.18 ± 0.27 mm ($p = 0.0008$) and the NRMSE from 0.086 ± 0.02 to 0.084 ± 0.18 ($p = 0.0009$), indicating increased motion consistency. Therefore, we opted to use data consistency and DVFs for MODEST.

5.3.1 4D-MRI reconstruction

Phase-undersampled zero-filled reconstructions were created using a NUFFT in approximately 5 seconds, while the XD-GRASP reconstruction took about one

hour. MODEST took 15 seconds to process the zero-filled reconstructions on an NVIDIA V100 GPU, while the U-Net took approximately 30 seconds to reconstruct the 4D-MRI. The parameter-pruned U-Net took about 25 seconds to perform a reconstruction.

In the example of phase-undersampled 4D-MRI at $R_{4D} = 1$ in the test set (Figure 5.3), MODEST produced reconstructions with an SSIM of 0.92 over the entire 4D volume, considering XD-GRASP as reference. This has significantly higher quality than the zero-filled reconstruction, which already shows undersampling artifacts and an SSIM of 0.82 ($p = 0.0017$). Despite having over thirty times fewer trainable parameters, MODEST also produces higher image quality for the considered subject than the U-Net. Compensating for the increase in parameters of the U-Net, the pruned U-Net reconstructs 4D-MRI with low image and low motion consistency, as identified by the hepatic dome position. At $R_{4D} = 4$, MODEST and U-Net showed comparable performance. However, the reconstructions by the U-Net seemed to suffer more from temporal blurring, as observable in the error maps of Figure 5.3. Videos of phase-undersampled reconstructions are provided for $R_{4D} = 1$, 2, and 4 in Supplementary Videos V1, V2, and V3, respectively. In these videos, it can be observed that MODEST and U-Net display similar image quality. However, in Supplementary Video V3, it can be seen that the U-Net reconstruction suffers from significantly reduced respiratory amplitude at the anterior chest wall, while MODEST shows better motion consistency.

The U-Net and MODEST outperformed the zero-filled reconstruction based on the SSIM and EPE metrics ($p = 10^{-9}$), as visible in the quantitative evaluation in Figure 5.4. However, no statistically significant difference was found between the U-Net and proposed architecture, except for the SSIM at $R_{4D} = 1$. Both models outperformed the parameter pruned U-Net for the SSIM metric ($p = 10^{-8}$). For the NRMSE metric, MODEST outperformed the U-Net, parameter pruned U-Net, and zero-filled reconstruction ($p = 10^{-7}$). MODEST showed sharper reconstructions for all under-sampling factors than the U-Net ($p = 10^{-7}$).

Using MODEST led to reconstructions with increased motion consistency, as found by the increased correspondence of the hepatic dome position, as presented in Figure 5.5. At $R_{4D} = 4$, the proposed architecture accurately tracked the hepatic dome position within 1.56 ± 1.98 mm compared to the XD-GRASP reconstruction versus 4.73 ± 2.48 for the U-Net. We observed that MODEST performed worse at exhale than inhale. However, the mean hepatic dome error was approximately 1.2 mm, significantly smaller than the voxel size of 3.5 mm in the feet-head direction.

Retaining fewer spokes for the free-breathing undersampled 4D-MRI decreased model performance due to an increased undersampling factor and increased intra-bin variability of the motion, as presented in Figure 5.6. The sharpness of
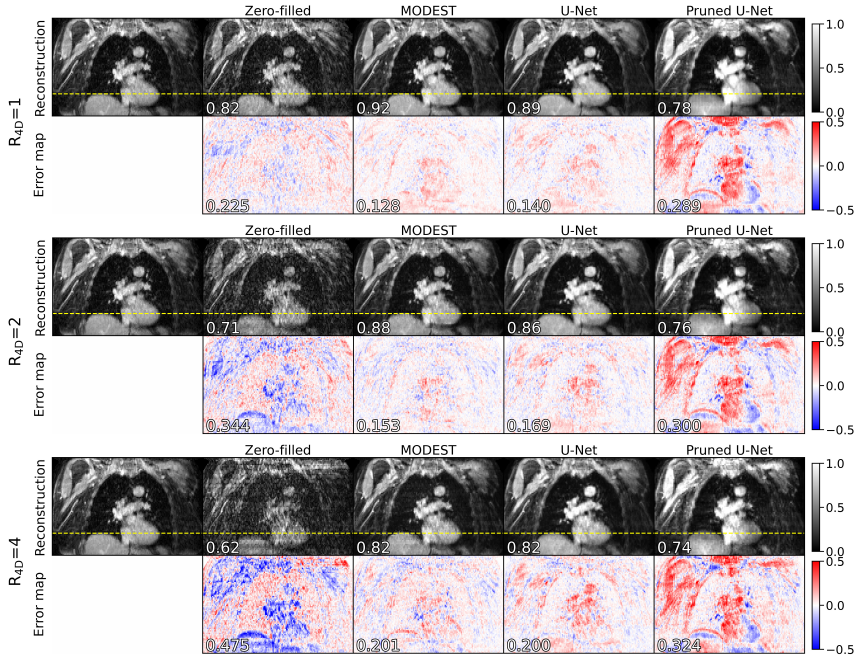
Figure 5.3 Example reconstructions. 4D-MRI was acquired of a female, 81 years old, affected by adenoma (T2N3M1). Reconstructions of phase-undersampled 4D-MRI inhaling by zero-filling, MODEST, the U-Net, and the parameter-pruned U-Net are shown for several undersampling factors and are compared to the XD-GRASP reconstruction. The top row shows the magnitude reconstructions and the SSIM, while the bottom row shows the NRMSE map and the mean NRMSE value for the 4D reconstruction. In Figure 5.4, a quantitative evaluation for the entire test set is shown.

the U-Net reconstruction decreased due to temporal blurring as the undersampling factor increased. In contrast, the sharpness of MODEST reconstruction is more stable. Based on the criterion that the shortest acquisition needed to have an EPE < 1 mm for the zero-filled reconstruction and an SSIM > 0.85 for the MODEST reconstruction, using the first 500 spokes is the shortest free-breathing acquisition that allowed reconstructing high-quality 4D-MRI using MODEST, corresponding to an acquisition time of approximately two minutes.

An example reconstruction for this acquisition is shown in Figure 5.7. Here, it can be seen that MODEST can reconstruct 4D-MRI with high quality with a mean SSIM of 0.92 and a mean NRMSE of 0.137 for this patient, which is of higher quality than the U-Net and pruned U-Net reconstruction. This model also shows good motion correspondence, as indicated by the alignment of the hepatic dome position. The quantitative results for the test set are presented in Table 5.1, showing that MODEST can achieve superior reconstructions compared to the

Chapter 5
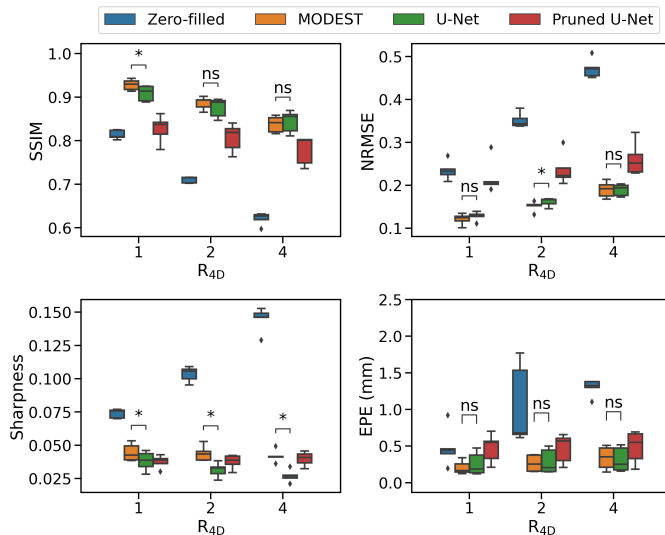
Figure 5.4 Quantitative comparison. All reconstruction methods are evaluated on the test set compared to the XD-GRASP reconstruction based on image similarity, measured by the SSIM and NRMSE, and motion similarity, measured by the EPE. All deep learning models perform significantly better than the zero-filled reconstruction, but MODEST outperforms the U-Net models based on image sharpness and NRMSE. A star indicates the t-test resulted in statistically significant differences with $p < 0.05$.
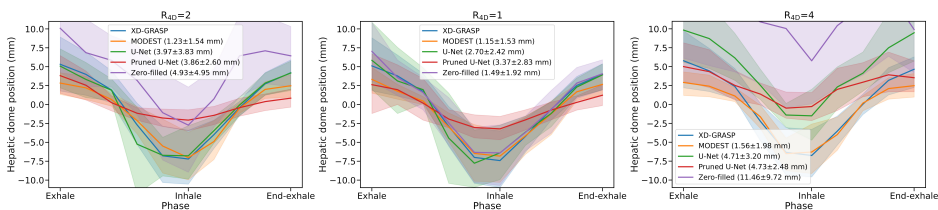
Figure 5.5 Hepatic dome analysis. MODEST closely follows the XD-GRASP reconstruction, especially at inhale. At high undersampling factors, MODEST is able to reconstruct motion-consistent 4D-MRI as measured by the hepatic dome, while the other reconstruction methods show significant errors.
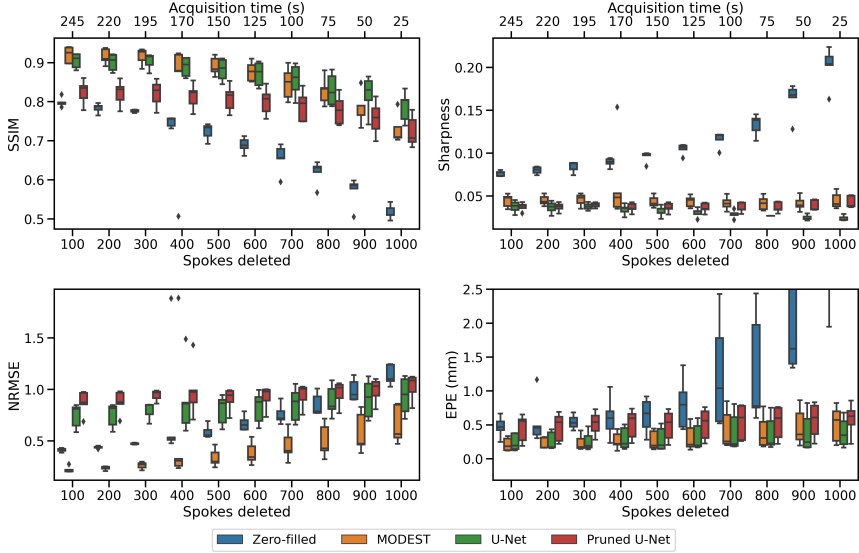
Figure 5.6 Impact of free-breathing undersampling. The impact of free-breathing undersampling was evaluated by continually removing n spokes from the acquisition and compared to the fully-sampled XD-GRASP reconstruction using the SSIM, EPE, and NRMSE metrics. As the increased significantly beyond removing 600 spokes, the minimum acquisition length was determined as 500 spokes. The approximate acquisition time is shown on top.

|	SSIM ($\uparrow$)	NRMSE ($\downarrow$)	Sharpness ($\uparrow$)	EPE ($\downarrow$)
Zero-filled	0.689 ± 0.019	0.673 ± 0.07	$\mathbf{0.104 \pm 0.006}$	0.813 ± 0.39
U-Net	0.871 ± 0.032	0.824 ± 0.16	0.030 ± 0.005	0.326 ± 0.20
MODEST	$\mathbf{0.877 \pm 0.025}$	$\mathbf{0.383 \pm 0.11}^*$	$0.043 \pm 0.007^*$	$\mathbf{0.313 \pm 0.20}$
Pruned U-Net	0.801 ± 0.036	0.920 ± 0.11	0.036 ± 0.006	0.512 ± 0.24

Table 5.1 Quantitative evaluation of the test set for free-breathing undersampled 4D-MRI using 500 spokes. Best result per metric marked in boldface, results with a star for MODEST indicate a statistically significant improvement compared U-Net ($p < 0.05$).

U-Net and pruned U-Net, with an NRMSE of 0.383 ± 0.11 versus 0.824 ± 0.16 and 0.920 ± 0.11, respectively. A video of free-breathing undersampled reconstructions is provided in Supplementary Videos V4.

5.4 Discussion

In this work, we have proposed an architecture called MODEST for efficient 4D-MRI reconstruction by splitting the model into spatial and temporal components. We designed a model that exploits all spatio-temporal information

Figure 5.7 Example free-breathing undersampled reconstructions. 4D-MRI was acquired of a female, 71 years old, affected by squamous cell carcinoma (T2N1M0). Reconstructed 4D-MRI by MODEST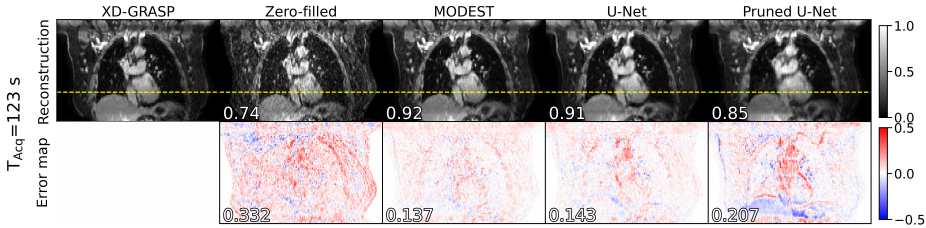, a U-Net, and the parameter-pruned U-Net are shown using an acquisition of 500 spokes (approx. 2 minutes) and are compared to the target XD-GRASP reconstruction. The top row shows the magnitude reconstructions and the SSIM, while the bottom row shows the NRMSE map and the mean NRMSE value for the 4D reconstruction.

of 4D-MRI using only low-dimensional convolution layers. High-quality 4D-MRI was obtained using this model from highly undersampled acquisitions in only 25 seconds and outperforms an optimized residual U-Net, despite having 3% of its trainable parameters. We have shown that the model can accurately reconstruct 4D-MRI from shortened acquisitions for up to two minutes while maintaining high image quality (SSIM of 0.877 ± 0.025) and motion-consistency with the fully-sampled 4D-MRI. These properties have some advantages over other models: models with few trainable parameters are less likely to overfit than larger models and have the potential to generalize better on unseen data due to less parameter variance[237]. Moreover, small models typically require fewer training samples converge[238], which is particularly interesting for MRI, as large datasets are difficult to acquire.

Our hyper-parameter optimization and model architecture search found that performing data consistency improved image quality, and adding motion information increased the reconstructed image quality. These findings are in line with previously published literature[239]. However, only adding the DVFs without adding data consistency can be detrimental to the image reconstruction quality. At $R_{4D} < 4$, adding DVFs to the images resulted in a lower SSIM, as indicated in Figure 5.2. However, at $R_{4D} = 4$ and in combination with data consistency, increased SSIM, lower EPE, and lower NRMSE was observed by adding DVFs. This could indicate that adding motion information at higher undersampling helps image reconstruction but provides less benefit at lower undersampling factors. This latter aspect could be due to the better conditioning of the inverse problem at higher sampling factors and due to imperfections in the motion estimation model. Currently, we only present the DVFs to the model as generated by a pre-trained network[134], which could limit the model performance. Based on previous literature, we foresee that performance may be improved by jointly

learning the image reconstruction and DVFs during training[160,240], improving image registration and image reconstruction performance.

Also, it would be interesting to investigate whether combining the spatial and temporal features by a learnable operator, e.g., convolution or self-attention[241], would impact, possibly improving the model performance and leading to even shorter MRI acquisitions. Alternatively, one could optimize the imaging protocol whenever possible by refining the image contrast and reducing scan time by decreasing the number of slices while maintaining the large field of view by slice interpolation.

This work used XD-GRASP reconstructed 4D-MRI as a ground truth since it demonstrated sufficient accuracy for radiotherapy applications[16,220,242]. However, this algorithm's regularization over the respiratory phases can introduce errors by overly smoothing the respiratory motion. This could introduce differences in motion amplitude compared to the measured data, and this uncertainty might limit the reconstructed motion quality by deep learning models. Using iterative joint image and motion reconstruction as ground truth could be a viable way to improve image quality[160] and remove residual artifacts in the ground truth. When comparing to XD-GRASP we considered a GPU implementation using commodity hardware, which might not be optimal. Technological developments have accelerate the XD-GRASP algorithm with specialized "Processing-in-memory" hardware[243], curtailing the computational bottleneck for XD-GRASP which enables a speed-up factor of 11, or 90 seconds of processing time. However, while this is a promising approach, these speed-ups have only been achieved in simulation and such hardware has not been clinically demonstrated.

The models presented in this manuscript have been trained on data obtained from eighteen patients, which is a limited training set size and could limit the performance of the presented models. Large training sets can offer several advantages, such as better performance and improved generalization capabilities. Several steps can be taken to increase the size of our training set. First, more patient data could be acquired, but this process is slow and costly, resulting in limited extra data. Second, digital phantoms could be used to generate 4D-MRI from numerical anatomy[144]. However, these samples might not be accurate compared to 4D-MRI acquired in-vivo. Future work will investigate the impact of different data augmentation approaches and dataset size.

MODEST is not the only architecture able to reconstruct 3D+t MRI. Freedman et al. proposed the so-called Dracula framework[226], consisting of a U-Net reconstructing zero-filled radial 4D-MRI to a high-quality 4D-MRI dataset and a mid-position image. Dracula produced 4D-MRI similar to HDTV-MoCo-based 4D-MRI in 28 seconds. However, this model was only investigated with a five-minute acquisition. Moreover, the network consisted of approximately 90,000,000 trainable parameters and took 11 days of training. Given the number of

Chapter 5

trainable parameters and their related GPU memory consumption, extending the model from a slice-by-slice reconstruction to a four-dimensional reconstruction is challenging. Küstner et al. proposed CINENet: a complex-valued unrolled U-Net that performs 4D spatio-temporal convolutions to reconstruct cardiac phase-resolved 4D-MRI[133]. They achieve the 4D convolutions by interspersing 3D convolutions with 1D convolutions. CINENet used an approach somewhat similar to ours by decomposing the 4D convolution into lower-dimensional convolution kernels, but we separated the spatial and temporal domains, whereas in CINENet they are interspersed. It is currently unclear whether interspersing or separating the spatial and temporal features would result in better performance, and it may be the object of future investigations.

MODEST has been specifically constructed to take advantage of the spatio-temporal information in 4D-MRI to obtain high-quality reconstructions. Interestingly, spatial and temporal information from MRI is relevant in other applications, such as cardiac imaging[133,225] or dynamic contrast-enhanced MRI[244,245]. Future work could investigate the application of MODEST, retraining the currently used model for these applications.

The availability of fast, accurate, and high-quality 4D-MRI is of particular interest for MRI-guided radiotherapy, where 4D-MRI is used for treatment adaptation of mobile tumors. With fast acquisition and reconstruction of 4D-MRI, treatment efficiency and patient comfort can be improved, eliminating the acquisition of a 4D-CT for motion quantification. By treating such patients on a hybrid MRI-Linac, motion can quickly be quantified without repositioning the patient. Moreover, high-quality 4D-MRI can also be used for high-quality time-resolved imaging[242,246] and could be helpful for real-time intra-fraction radiation treatment adaptation[45].

5.5 Conclusion

We proposed a deep learning architecture called MODEST that efficiently reconstructs high-quality 4D-MRI by decomposing the reconstruction into spatial and temporal components. This approach yielded superior performance than conventional models such as U-Nets, despite having only 3% of the trainable parameters. We found that high-quality 4D-MRI can be obtained with an MR acquisition of two minutes and 15 seconds of model inference, shortening the time for MRI-guided radiation treatments while improving treatment quality and incorporating accurate motion quantification.

5.6   Acknowledgement

Chapter 5

CHAPTER 6

General discussion and summary

This thesis has explored methods for real-time abdominothoracic motion estimation to enable online adaptive MRI-guided radiotherapy. Specifically, we have focused on deep learning, given its potential to speed up image reconstruction and registration. This chapter summarizes the main findings of the work presented in this thesis. In **Section 6.1**, we discuss this thesis's findings on real-time motion estimation with deep learning and discuss the implications, clinical impact, and possible limitations of these techniques. **Section 6.2** considers the training of deep learning models with the radiotherapy application in mind, while in **Section 6.3**, we apply these lessons to accelerate the MRIgRT workflow.

Finally, in **Section 6.4**, we reflect on the findings of this thesis and how they could translate into clinical practice. We discuss the limitations of deep learning and give directions for future work to enable deep learning-based real-time adaptive MRIgRT fully.

6.1   Real-time motion estimation

Intra-fraction motion has remained a challenging problem for the accurate irradiation of mobile tumors. To mitigate the effects of intra-fraction motion, the three-dimensional location of the tumor and its surrounding tissue must be (non-invasively) estimated with high accuracy in real-time. **Chapter 2** focused on real-time 2D motion quantification for irradiating abdominal tumors with the primary objective of examining the feasibility of deep learning-based motion estimation. While 2D imaging is generally fast to acquire and reconstruct (e.g., within 250 ms), resolving abdominothoracic motion from these images remains challenging within the allowable timeframe, depending on the registration algorithm[49,50,247]. Moreover, estimating motion in real-time, i.e., a latency ≤ 400 ms and a temporal resolution ≥ 5 Hz, requires significant acceleration of the acquisition, potentially introducing motion field artifacts due to signal aliasing[248]. Deep learning provided an advantage over traditional image registration algorithms in this setting. End-to-end image reconstruction and motion estimation allowed for high-quality deformation vector fields from aliased images. Moreover, we established where in this pipeline it is most effective to apply deep learning and found that replacing the motion estimation algorithm benefits the DVF quality while replacing regular Fourier-based image reconstruction is detrimental to the motion quality. With this approach, we could significantly accelerate the image acquisition and reconstruction and estimate motion with 25 Hz (R=25) while maintaining an average RMSE below 1 mm.

In **Chapter 3**, we have extended the findings of the 2D motion estimation approach to a volumetric real-time motion estimation method called TEMPEST. Here, we demonstrated that TEMPEST could estimate motion within 2 mm (i.e., smaller than the voxel size) despite an acceleration factor of 30 using a golden-angle radial stack-of-stars (GA-SOS) acquisition in digital and physical

phantoms. Moreover, TEMPEST could provide accurate volumetric motion estimates using a 3D golden-mean kooshball acquisition with an acceleration factor of 244. The total latency of TEMPEST was estimated to be approximately 200 ms with a temporal resolution of approximately 175 ms, which is sufficient to resolve respiratory motion and enable online adaptive MRIgRT in the abdominothoracic region. However, the latency is insufficient for resolving faster types of motion in real time, such as cardiac motion or speech imaging.

Deep learning models can estimate motion from highly undersampled images with high spatiotemporal resolution. This seems surprising as one intuitively expects that high-quality MRI is required for motion estimation. However, the experiments in **Chapter 2** reveal that using a de-aliasing network before estimating the DVF is detrimental to directly estimating the DVF from the aliased images. One possible explanation for this phenomenon could be that conventional (NU)FFT-based reconstruction preserves the critical information necessary for motion estimation. In contrast, DL-based image reconstruction algorithms that maximize the image quality might retain detrimental residual image artifacts. For example, minimizing the mean-squared error for image reconstruction leads to blurring[249], while motion estimation algorithms often rely on sharp edges[250]. Therefore, when performing end-to-end training from the aliased images to the DVFs, the network may learn to perform improved de-aliasing while preserving salient information for image registration. This provides unique benefits to these DL-based image algorithms as traditional image registration methods minimize a manually-defined cost function, whose effectiveness depends on the image quality. Moreover, as traditional image registration often requires online iterated gradient descent, which can take seconds or minutes, DL models have a significant speed advantage as the training happens offline.

The presented motion estimation models are supervised deep neural networks requiring ground-truth motion. However, in practice, *in-vivo* ground-truth motion is difficult to obtain. Instead, surrogate ground-truth motion is obtained through conventional image registration. This approach introduces the risk of transferring the imperfections of the surrogate ground truth to the CNNs. It is, therefore, worthwhile to generate high-quality ground truth DVFs, as it is unlikely that a neural network will perform better than its training target. In this thesis, we have used optical flow[40] as the surrogate ground truth, which assumes smoothness in its deformation vector field to register images. This assumption fails when registering sliding interfaces, such as abdominal organs sliding along the spine[251]. However, in Section 2.6, we compared optical flow to a medical image registration toolbox[42] and found that optical flow showed better image registration performance. This could suggest that the impact of failing to register sliding motion is relatively benign or that the image similarity metrics do not sufficiently reflect the impact. Another way to quantify motion estimation failures could be by evaluating the dosimetric impact[252].

Chapter 6

Alternatively, unsupervised image registration has been proposed as a suitable image registration technique, mitigating the need for ground-truth DVFs and the corresponding parameter tuning of the registration algorithm[247,253]. These approaches attempt to learn a DVF that maximizes the image similarity (e.g., the SSIM[116] or mutual information[156]) between a reference image and the dynamic image warped by the DVF. While these approaches are promising, they perform similarly to traditional, non-DL-based image registration techniques and have yet to outperform the traditional methods[247,253,254]. However, they are significantly faster than traditional motion estimation techniques, which can take seconds in 2D and seconds or minutes for volumetric image registration. On the other hand, DL-based image registration takes milliseconds or seconds. The quality of the DVFs estimated by unsupervised models depends on the fixed and moving image quality, the image similarity metric, and the model architecture. Unsupervised real-time image registration is challenging when considering undersampled MRI due to severe aliasing artifacts that hinder image-based similarity metrics. Unsupervised image registration could be feasible for real-time motion estimation if a suitable image similarity metric that is only sensitive to motion and insensitive to aliasing artifacts is developed.

The advantages and disadvantages of traditional, deep learning-based supervised and unsupervised image registration are summarized in Table 6.1. Based on the needs of real-time adaptive MRIgRT, which requires accurate DVFs from highly undersampled MRI, supervised deep learning networks seem most suited for this task.

One of the essential ingredients to the success of the presented motion estimation models is the use of a multi-resolution approach, a well-known technique in motion estimation algorithms to promote convergence when resolving large deformations[255]. However, when registering MRI, there are two additional advantages: first, due to the variable-density nature of a radial k-space acquisition, the effective undersampling factor and subsequent aliasing are reduced as spatial downsampling is equivalent to cropping in k-space, increasing the effective sampling rate of low resolution images. This provides the network tasked with registering high-resolution, highly aliased images with good initial estimates of the motion, making the problem less underdetermined. Second, reducing the image resolution facilitates resolving large displacements as the motion is captured within the receptive field of the convolutional operators of the network. Despite the reduction in resolution, image registration algorithms can still accurately estimate motion[97]. However, the presented motion models do not consider the temporal information while exploiting prior knowledge on the temporal variation of MRI could result in improved motion estimation quality and faster MRI acquisition[157,256]. Recurrent neural network (RNN) architectures[257] and self-attention networks like Transformers[241,258] have been proposed to learn spatiotemporal representations. However, they could be challenging to apply to

Traditional image registration	Supervised deep learning	Unsupervised deep learning
+ Can produce high-quality DVF	+ Fast	+ Fast
+ Well understood	+ Easily validated	+ Does not need ground-truth
+ Extensively validated	+ Works well with aliased/non-perfect images	+ No parameter tuning after training
– Slow, iterative computation (seconds to minutes)	+ No parameter tuning after training	– Needs image-based similarity metric
– Needs high-quality images	– Needs ground-truth DVF	– Needs high-quality images
– Needs manually-defined cost functions		
– Manual parameter tuning		

Table 6.1 Comparison of image registration techniques. Deep learning-based image registration is a fast, high-quality alternative to traditional image registration that is parameter-free after training the model. DL-based image registration can be supervised or unsupervised, which have their respective trade-offs: Supervised image registration can work with imperfect images, but unsupervised image registration does not need a ground-truth DVF, which can be difficult to obtain.
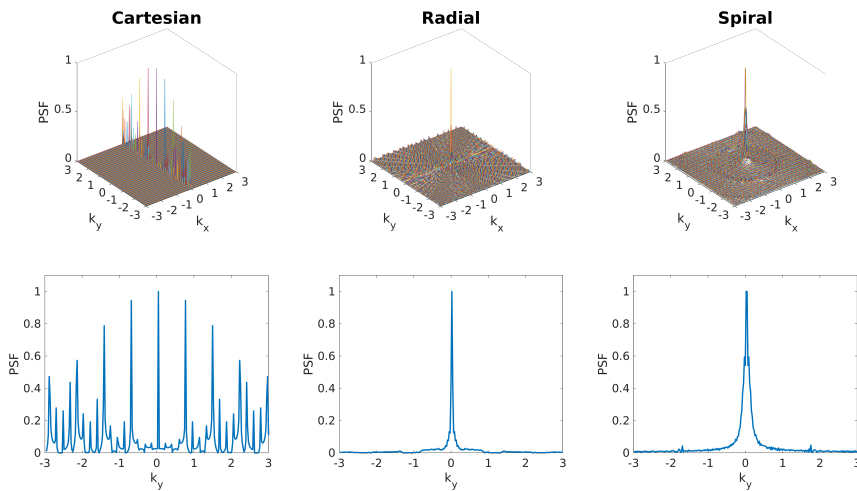
Figure 6.1 The point-spread function (PSF) of the trajectories that are shown in Figure 1.4. The top row shows the full PSFs, while the bottom row shows line profiles along $k_x = 0$. The PSF visualizes the incoherence of a sampling trajectory. The value of PSF at location i, j represents how much the signal at the ith position contributes to the jth location. When fully sampled, the PSF is $\delta_{(0,0)}$, where δ is the Dirac delta function. If the Nyquist criterion is violated, non-zero off-center values appear. The spiral and radial samplings show incoherency as they have little off-center structure and low off-center energy. In contrast, the Cartesian undersampling shows high coherency in the phase-encode direction k_y and no undersampling in the frequency-encode direction k_x.

2D+t or 3D+t problems due to the computational cost of long sequences[259,260]. Further investigation is needed to exploit the temporal information efficiently.

We have solely focused on non-Cartesian methods to perform real-time motion estimation. Using these trajectories is motivated by the motion-robust nature of radial acquisitions; repeated sampling and averaging of k_0 makes motion artifacts

Chapter 6

109

appear as blurring rather than ghosting artifacts using sequential Cartesian acquisitions[261]. The streaking artifacts of radial acquisitions are easily resolved due to the unstructured spreading of the aliasing energy. This spread of the aliasing energy is measured by the trajectory's point spread function (PSF) of the trajectory, as shown in Figure 6.1. The low off-center energy of the PSFs of radial or spiral trajectories reveals that undersampling results in relatively benign aliasing artifacts. However, the non-Cartesian acquisition also comes at a cost. These readouts are more sensitive to gradient system imperfections, e.g., gradient delays, eddy currents, and gradient nonlinearities, which can introduce geometric uncertainties. Previous studies have shown that gradient nonlinearities can significantly impact the radiation treatment plan[262]. While errors in the gradient system could be compensated[60] and are generally minor for commercially-available MRI-Linac devices[263], this downside makes them seem less suited for radiotherapeutic applications where geometric accuracy is paramount.

Moreover, reconstruction of the k-space to image space cannot be performed by a conventional fast Fourier transform (FFT) as frequency locations $\vec{k}$ are not sampled on a rectilinear grid. Instead, image reconstruction is typically performed using a non-uniform fast Fourier transform (NUFFT), which is one order of magnitude slower in 2D and three orders of magnitude slower in 3D compared to a regular FFT. While tractable in 2D, computing a multi-coil 3D-NUFFT for a golden-mean kooshball acquisition is intractable for high-resolution acquisitions. It is worthwhile to revisit whether the established non-Cartesian acquisitions are optimal for real-time adaptive MRIgRT.

One option is using hybrid Cartesian/non-Cartesian acquisitions to get the best of both worlds. For example, the recently-proposed (r)CASPR[63] trajectory proposes to acquire 3D k-space points at Cartesian grid locations following a discretized spiral arm, retaining the motion robustness of non-Cartesian MRI while allowing MRI acquisition with high temporal resolution and FFT-based reconstruction[153,264]. The other option is to embrace the non-Cartesian nature and increase the complexity of the sampling trajectory[265]. With this approach, the k-space sampling trajectory is optimized along the parameters of a DL reconstruction network to maximize sampling efficiency per shot, given the constraints of the gradient system. The synergy between optimal encoding and regularized regression with DL could increase the maximum acceleration factor and motion estimation accuracy. The most radical approach would be to disband the linear gradient system design and embrace the non-linear gradients[266]. It has been shown that non-linear gradients have the potential to achieve greater SNR and noise robustness than conventional parallel imaging reconstructions in predetermined regions of the image and could be used to maximize the signal and resolution in the GTV. Combined with learned regularization by deep learning, the physiological motion could be resolved with optimal spatial encoding, resulting in a minimal motion estimation latency.

6.2 Deep learning model training

The performance of deep neural networks is dependent on a few high-level parameters: the training data distribution, the model architecture, parameter initialization, the model hyperparameters, and the learning objective. The role of the learning objective on the performance of deep learning models to estimate motion for MRIgRT was evaluated in **Chapter 4**. The main finding is that traditional minimization of the ℓ^2-norm (or mean-squared-error minimization) in the $\mathbb{C}$ or $\mathbb{R}^2$ domain in the presence of noise results in a systematic bias of the error term, promoting estimates with lower magnitude, as observed by previous literature[179,183]. We proposed $\perp$-loss (pronounced ˌpɜːpˈlɒs), a symmetric loss function in the complex domain, to resolve this problem.

While the systematic underestimation of magnitude is present for every regression task in the $\mathbb{C}$ or $\mathbb{R}^2$ space, it is particularly concerning for MRI-guided radiotherapy. Ideally, the DVFs used to perform MLC tracking have a magnitude and directional error distribution with a mean of zero and are distributed equally around the true value, resulting in a uniform distribution of spurious dose delivery to nearby OARs. A non-uniform error distribution, such as an underestimation in reconstructed magnitude, could result in increased dose delivery to a nearby OAR while delivering insufficient radiation to the tumor, decreasing treatment efficacy and local control. This effect is illustrated in Figure 6.2. Using a symmetric loss function could lead to more accurate motion estimation models (both DL-based and non-DL-based) that estimate DVFs with a uniformly-distributed magnitude and direction error, leading to improved dose delivery. Moreover, the identified bias when minimizing the ℓ^2 norm highlights a need to validate motion estimation models. These methods should validate the accuracy of such models and the systematic bias in the output of these models[18,267].

Curiously, we have found that using $\perp$-loss reduces the systematic bias of the reconstructions while improving the reconstruction quality. However, we found that the loss function underperforms at extreme noise levels. With realistic noise levels, $\perp$-loss-based regression outperforms ℓ^2-based regression while exhibiting reduced systematic bias. We can speculate about the origin of reduced performance at high noise levels by considering Figure 4.3. Here, it can be observed that $\perp$-loss shows a strong gradient in the λ or magnitude error direction. However, the gradient in the phase error direction is relatively flat, even for high phase errors. The lack of gradient magnitude could stymie the reconstruction of noisy data. However, in the noise-free case, a strong gradient coupled with momentum-based optimizers could explain the increased convergence speed compared to MSE-based optimization.

The $\perp + \ell^2$-loss function was only evaluated for regression in the $\mathbb{C}$ or $\mathbb{R}^2$ domain using deep learning models. Non-DL-based regression techniques such as least squares regression or compressed sensing MRI reconstruction could benefit from
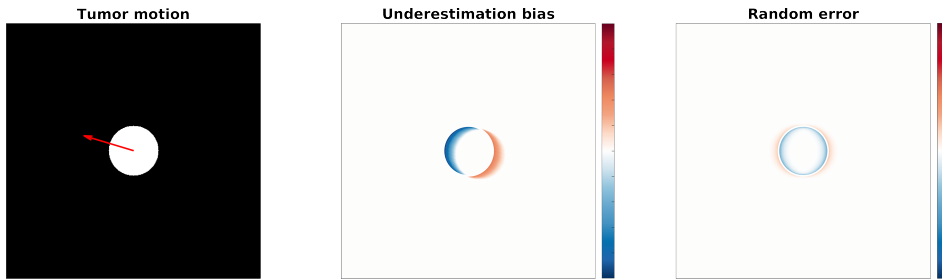
Figure 6.2 Illustration of the effect of biased motion estimation. On the left, a tumor (white) moves along the red arrow. In the middle, the relative dose deposited when an image registration algorithm is biased towards magnitude underestimations. Here, white means the dose is delivered as planned, blue indicates under-dosage, and red indicates over-dosage. A dose increase is visible next to the tumor, where the OARs typically reside, while the tumor is under-dosed. On the right is the effect of an image registration algorithm with a random error of equal magnitude, which results in a more uniform dose distribution.
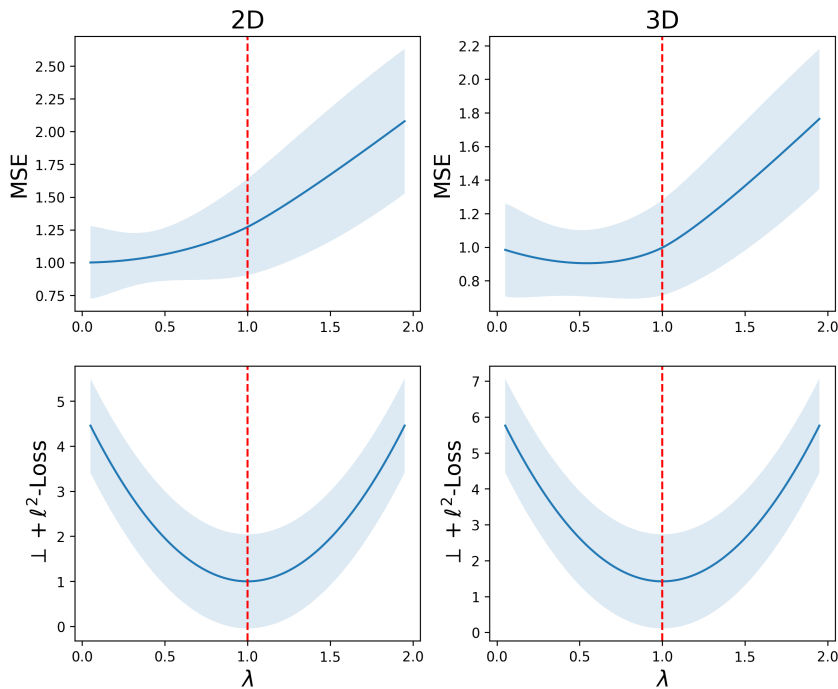
Figure 6.3 Comparison between the MSE and $\perp + \ell^2$-loss in 2D and 3D. Fifty thousand random vectors were generated and perturbed by a random magnitude scaling λ and random angle. On the horizontal axis is the λ value, and on the vertical axis is the loss value ($\mu \pm \sigma$) when averaged over all vector pairs with a specific λ value. In 2D and 3D, the minimum value of the MSE is at $\lambda < 1$, while it is precisely at $\lambda = 1$ for the $\perp + \ell^2$-loss.

$\perp$-loss. This loss function seems especially suited for phase-based applications such as chemical shift MRI techniques such as Dixon MRI[268], quantitative susceptibility mapping[201], and phase-contrast MRI techniques such as 2D velocity encoding[269] or 4D flow MRI[202]. Outside medical imaging, $\perp$-loss could find applications for audio processing[270] or crystallography[271], which both involve the analysis of complex-valued signals. Even though the $\perp + \ell^2$-loss was only evaluated in the complex domain, Equation 4.1 suggests that it can be extended to higher domains such as $\mathbb{R}^3$ and applied to volumetric or spatiotemporal regression. It is currently unclear whether $\perp$-loss outperforms MSE-based regression in higher dimensions. However, repeating the loss landscape investigation described in Section 4.3.1 in 3D shows that the asymmetry of the ℓ^2-norm also manifests itself in $\mathbb{R}^3$, as illustrated in Figure 6.3. The $\perp + \ell^2$-loss remains symmetric, indicating that its beneficial properties extend to volumetric or spatiotemporal regression problems. Incorporating the time domain might be especially interesting as it could lead to a symmetric dose distribution error over time, which has yet to be considered part of a single loss function.

6.3 Accelerating the MRL workflow

The online adaptive MRIgRT workflow for mobile tumors will consist of multiple time-consuming steps: patient set-up, a (4D-MRI) pre-treatment scan, treatment planning, radiation delivery, and post-treatment evaluation. In this chain, obtaining 4D-MRI takes too long, limiting the efficiency of the treatment and reducing patient comfort. DL can be used to obtain high-quality 4D-MRI from undersampled acquisitions by taking advantage of the spatio-temporal correlations present in the data. In **Chapter 5**, we proposed a new deep learning-based approach called MODEST that uses the spatio-temporal information by decomposing the reconstruction into separate spatial and temporal components. The motion estimates are then used to generate a personalized motion model and to define the necessary treatment margins. However, these acquisitions can take up to five minutes, and reconstructing high-quality 4D-MRI can take prohibitively long. With MODEST, high-quality 4D-MRI can be acquired and reconstructed in 150 seconds.

The proposed model can efficiently reconstruct 4D-MRI by decomposing the reconstruction problem into spatial and temporal subnetworks. This has resulted in a small model with few trainable parameters as the decomposition allowed using $(n-1)$-dimensional convolution kernels for reconstructing n-dimensional MRI. Unlike a baseline model, MODEST achieved a 97% reduction in the trainable parameters and a 99.6% reduction in trainable parameters compared to a previously-published technique[226]. A model with few trainable parameters has theoretical benefits. A small model has less variance than more extensive neural networks and is less likely to overfit[272]. Second, smaller models should require less data to converge than large models as they contain few unknowns.

113

However, other factors, such as data quality and diversity have to be considered for the final performance.

The decomposition into spatial and temporal subnetworks is assumed to be allowed as the model can leverage the prior knowledge on the data that every temporal volume utilizes a mutual spatial 3D basis. At the same time, the motion deforming the subject for every respiratory phase is low-rank, allowing separate processing of the spatial and dynamic domains. This separability property and the low-rank assumption have often been used in dynamic MRI reconstruction[137,273–275]. Recently, attempts have been made to formalize these properties mathematically. Kofler et al.[228] compared the complexity of the manifold as learned by auto-encoders when trained on only spatial data and spatio-temporal cine slices. The spatio-temporal manifold was significantly less complex and easier to learn. Moreover, Mardani et al.[276] performed tensor analysis on four-dimensional multi-coil cardiac MRI data and found that the temporal domain can be captured using a rank-one tensor. These results indicate that it is mathematically sound to decompose the spatiotemporal problem into its constituent parts, as low-dimensional approximations can capture the dynamics. However, we assume a lower limit to the data required for successful decomposition, and the 4D-MRI should contain sufficient spatiotemporal correlation. The 4D-MRI considered in this work contains sufficient spatiotemporal correlation. However, we found a lower limit on the acquisition time, and violating this limit precludes the generation of reliable 4D-MRI when decomposed. However, it results in a "motion snapshot" instead of a valid respiratory motion model with sufficient motion quality for radiotherapy planning[277].

Fast, DL-based 4D-MRI could be used for high-quality mid-position imaging. The mid-position image is the time-weighted average image during respiration. It is beneficial for MRIgRT as it can be used as the treatment planning image, resulting in minimal PTV margins[278]. 4D-MRI can also be used as a generic approach to personalize motion models[120,279]. These personalized priors could be combined with generic 3D motion models, improving DVF quality or enabling further MRI acquisition during real-time tracking. Moreover, the 4D-MRI could be used for time-resolved imaging. Recently, MR-SIGMA[246] has been proposed as a real-time time-resolved volumetric imaging method by performing online matching of the acquired k-space to the 4D-MRI. However, the work presented in this thesis has found that high-quality motion estimates can be obtained in real-time from highly-undersampled 3D MRI. This way, high-quality time-resolved 3D MRI could be obtained by deforming the mid-position MRI image with the estimated DVF. While the MR-SIGMA method depends on the motion patterns in the 4D-MRI, generating time-resolved MRI using the presented motion model does not rely on this assumption.

6.4   Clinical applicability and recommendations

We are at the dawn of deep learning for image-guided interventions. The central research question was to develop methods to enable real-time adaptive MRI-guided radiotherapy. The works presented in this thesis have shown that this is a feasible and promising way forward toward real-time time-resolved motion estimation for online adaptive MRIgRT on MRI-Linac devices. However, several necessary steps must be taken before deep learning-guided interventions make their way to clinical practice. Here, we address some critical issues that must be solved to ensure the safe and effective use of high-quality deep learning models for motion tracking.

Validation and clinical acceptance

In this thesis, we have presented deep learning models that can quickly estimate DVFs from highly-undersampled MRI. We have demonstrated that these models can resolve large, non-rigid deformations with high accuracy (e.g., error $\leq$ voxel-size, or $\leq$ 1 mm), depending on the undersampling factor. While these results are promising, the acceptance criteria for clinical implementation still need to be determined.

Speed
First, the method has to be fast enough with a low enough latency. While staying within 200-400 ms of latency for accurate respiratory motion tracking is generally recommended[49,50], it is still being determined what the maximum accepted latency is for MLC tracking. In either case, the tracking latency could be partially mitigated by *predicting* rather than *estimating* the respiratory motion[280], exploiting the temporal smoothness of respiratory motion. Moreover, the requirements for the *sensitivity* of the motion estimation method for MLC tracking have yet to be determined, i.e., the minimum motion magnitude that needs to be resolved by the motion estimation method. The sensitivity is limited by a few factors, such as the voxel size and the image contrast, thus limiting the MRI acquisition parameter set.

Robustness
Another factor to consider is the *robustness* of the motion estimation method. Deep learning methods carry the intrinsic risk of unexpected failure when applied to unseen data or data that diverges sufficiently far from the training data distribution. This fear is partially corroborated by the black-box nature of deep learning models that might learn non-robust features[281] and the potential that the training data does not reflect the heterogeneity of clinically-acquired MRI[282]. There are, however, some potential solutions to preemptively mitigate these problems. One solution is uncertainty estimation[283]. Here, the deep learning method attempts to perform the classification or regression task it is trained for while simultaneously attempting to estimate the aleatoric uncertainty (or data

uncertainty) and epistemic uncertainty (or model uncertainty)[284]. The epistemic uncertainty can be reduced by increasing the dataset used for training the model. In contrast, the aleatoric uncertainty cannot be reduced due to measurement noise or partially observed features. However, the model input during deployment could be compared to the data distribution of the training set and halting irradiation if the likelihood that the sample follows the training distribution decreases below a threshold.

Another solution to verify the robustness of DL models is by employing explainable machine learning[285,286]. This approach attempts to open the metaphorical black box by, for example, offering a post-hoc explanation of what input features *caused* a particular output. The explanation allows operators to (mis)trust the machine learning algorithms by verifying that the algorithm reached a solution based on sound information. While this seems like an attractive method to gain confidence in the model output, I believe this holds deep learning models to an unreasonably high standard where operators expect functionality from these models that we do not expect from non-DL-based methods or humans. Moreover, given the real-time nature of the problem, there is no time for humans to inspect or correct the output, limiting the efficacy of explainable machine learning for this application. Instead, it is my opinion that the successful implementation of deep learning models needs strict requirements: the output of a deep learning model is only clinically used within a specific set of operating parameters (e.g., for specific anatomy and image contrast). Within this parameter space, deep learning methods must maintain a clinically-acceptable performance. For example, a tracking algorithm could be required to maintain a specific performance relevant to radiotherapy, such as the gamma index[287]. These algorithms could be required to, for example, consistently achieve a gamma pass rate of $1\%/1\,\text{mm} > 95\%$ during validation. While restricting DL models to a parameter set is not as generic as traditional models (e.g., the Fourier transform), it removes the need for careful parameter tuning as required for methods relying on traditional mathematical modeling[72,288–290] while quickly producing high-quality output.

While serious concerns were raised on the robustness and uncertain nature of DL, we have observed a surprising ability of the presented motion models to generalize beyond the training data. Applying deep learning to MRI with significantly higher undersampling factors showed a gradual degradation of motion estimation performance, as shown in **Chapters 2 and 3**. For example, applying a model trained on undersampled MRI achieved a target registration error of 1.87 ± 1.65 mm on CT. While dedicated CT registration methods and networks perform better than the presented model[291], the observed cross-modality image registration performance is remarkable despite being solely trained on MRI. Moreover, increasing the undersampling factor from 5 to 18 increased the registration error by merely 0.5 mm despite being trained only up to R=7. These results suggest that the models generalize very well, even beyond the MRI contrast it was trained on or even the imaging modality. However, this thesis only demonstrated

the technical feasibility of real-time motion estimation with deep learning and performed preliminary validation of the models. Further validation is needed to verify that the DVF quality is maintained as the MRI is acquired using different pulse sequences, image contrasts, coil arrays, or MRI acquired using the same subject on MRI constructed by different vendors.

In-vivo validation

Even if the models can robustly obtain specific results with certainty, the validation is one of the main issues with real-time MRI-based motion estimation. The models are tasked to estimate motion from highly-undersampled MRI. As the exact forces and physiological models that deform and displace tissue due to these forces are unknown *in-vivo*, it is near-impossible to come to the correct solution to this highly ill-posed inverse problem. This problem implies that it is near-impossible to *validate* the correctness of the model estimates as there is no correct ground-truth motion field to estimate the quality of the model' DVFs. However, the estimated DVF must represent the true physiological motion during irradiation to enable accurate tumor tracking and dose accumulation. The correctness of the global motion could be independently validated by interleaving navigators with the acquisition[292], but these can only be used to validate the bulk density changes rather than the local deformations. The lack of in-vivo validation methods is a grave concern for verifying volumetric tracking algorithms, including those based on deep learning. Future work should investigate a reliable, standard method for verifying the quality of *in-vivo* motion estimation[293]. The only methods to validate motion estimation algorithms are physical[294–296] and digital motion phantoms[144,297]. However, no physical phantom currently generates ground-truth DVFs of non-rigid motion. While digital phantoms provide a ground-truth DVF, they often lack the intricate hardware imperfections acquired during MRI acquisition, possibly resulting in overly optimistic model performance. Moreover, machine learning models are typically fitted to a specific data distribution, such as contrast, anatomy, or motion pattern. Then, the phantom measurements are considered out-of-distribution samples and could induce failure in the DL models. However, it could be argued that a DL model that cannot be validated using phantoms is not trustworthy enough to use in clinical practice.

Datasets

One of the ingredients that helped the interest and success of deep learning is the availability of high-quality datasets. These large, humanly-curated datasets have been the standard for many domains, such as image classification[298], object detection[299], natural language processing[300], and image generation[301]. Even in the medical domain, open datasets have become more common, despite the challenges of acquiring and releasing the data in compliance with privacy regulations. This has resulted in a surge of deep learning datasets to train models to perform, for example, disease detection from chest X-rays[302], brain tumor

Chapter 6

117

segmentation and classification[303,304], abdominal MRI segmentation[305], or MRI image reconstruction[187,190,306].

However, specifically for MRIgRT, the availability of large, high-quality datasets for real-time motion estimation or image reconstruction has remained largely absent. Datasets that could be used for these problems are either too small (e.g., few patients from a single treatment center), are for different anatomies than the intended application, lack realistic ground-truth data (e.g., a description of the true motion), or lack ground-truth measurements (e.g., prospectively-undersampled k-space). As a result, most publications now use private datasets, which hinders reproducing the results of proposed deep learning methods and comparing different approaches. This lack of publicly-available datasets of this data stymies new research.

Creating large, open datasets is essential to advance the field of MRI-guided radiotherapy and produce deep learning models with clinical impact. Such datasets must include many patients and consist of longitudinal real-time volumetric MRI acquired in multiple treatment facilities using all commercially-available MR-Linac devices. Moreover, the corresponding motion fields and segmentations of organs-at-risk, GTV, and CTV must be accurate and available for all timesteps. Finally, methods for accurate treatment planning with MLC tracking based on the real-time estimated DVFs must be provided to evaluate these models. The MOMENTUM study is one example of the efforts to create an MRIgRT-first dataset[307]. This multi-institutional, international registry of technical and clinical data of cancer patients treated on MRI-linacs aims to include 6000 subjects. While the registry does include the pre-treatment MRI and treatment margins, it does not include intrafraction MRI. However, it can still provide valuable data for training and validating DL models.

Creating a dataset with intrafraction MRI, ground-truth DVFs, segmentations, and treatment margins will be a daunting effort but is a prerequisite to advancing the field of model-based MRI-guided radiotherapy. The lack of publicly available, large, realistic datasets to evaluate methods stymies the field and makes it difficult to gauge the impact of allegedly improved tumor-tracking algorithms.

Radiotherapy-first methods

Medical imaging has particular needs and requirements that differ strongly from computer vision or the general DL community, and radiotherapy is no different. The imaging is necessary to obtain accurate geometric information about the tumor and OARs with a sufficient spatial and temporal resolution to deliver radiation as accurately as possible in the tumor while sparing OARs.

However, most deep learning methods applied to MRI-guided radiotherapy use approaches originating in radiology or computer vision. While these methods

show adequate performance in practice, the trade-offs that these models present are not necessarily beneficial for the needs of radiotherapy. For example, the geometric accuracy of the tumor and OARs positions is paramount, whereas models trained for radiology favor blurred reconstructions in the presence of undersampling artifacts[249]. Moreover, motion estimation algorithms are optimized for *global* motion estimates rather than *local* ones in the tumor and OARs. It is conceivable that a model that estimates the motion of a liver tumor reduces the motion quality in this region to improve the motion estimate in a less critical region.

Deep learning models must use all the domain knowledge available to make optimal trade-offs for the radiotherapy domain. For image reconstruction, this would mean favoring geometric accuracy over blurring when reconstructing undersampled acquisitions. This could be enabled by presenting additional information to the model, such as a segmentation of the patient in a pre-treatment image, a map of the main magnetic field (B_0-map) to characterize local field inhomogeneity due to magnetic susceptibility[308], or a characterization of the local gradient system to correct for geometric inaccuracies due to gradient delays[59]. For motion estimation algorithms, it is essential to consider the patient's segmentation to assign local importance where the DVFs must be correct.

Finally, current DL models are optimized for surrogate metrics for radiotherapy applications. For example, image reconstruction models are optimized for global image quality metrics such as the mean-squared error or structural similarity. Motion estimation models optimize for a supervised DVF loss or the image similarity after registration to a fixed image. However, in the end, the only relevant metric is the dose delivered to the tumor and the OARs. Future work could focus on the holistic training of these models to optimize tracking performance and dose delivery accuracy.

6.5   Real-time adaptive MRIgRT

MRI-guided radiotherapy brings radiation treatments to a new level and significantly improves the quality of care. However, the real potential of MRIgRT is yet to come when real-time adaptive radiotherapy enters routine clinical care worldwide. If widely available, the door is open for exciting new treatment paradigms, such as extreme hypofractionation for mobile tumors[47] and reduced side effects due to minimal treatment margins[309], improving patient comfort, treatment cost, and treatment outcome. However, this will require accurate tracking and significant workflow improvements[310]. This thesis shows that deep learning is a feasible technique to estimate motion accurately, removing one of the technological roadblocks standing in the way of real-time adaptive MRIgRT. However, to truly enable this vision, we must go much further. Given that the previously mentioned challenges are solved, and deep learning-based tracking
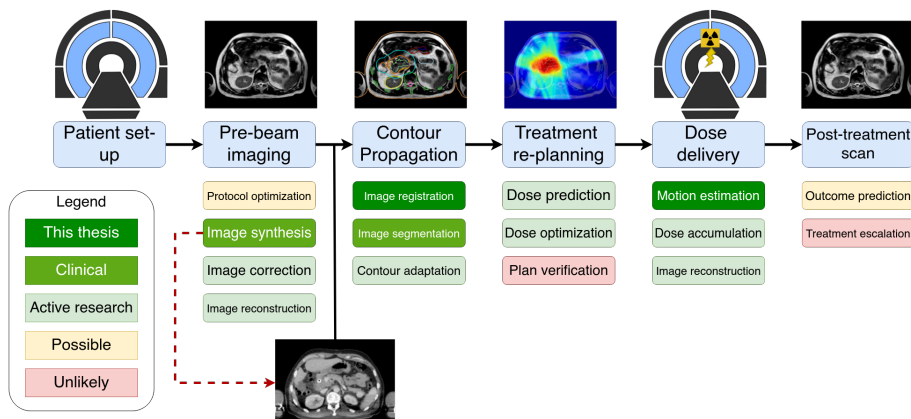
Figure 6.4 Overview of the MRIgRT workflow. After patient set-up, a pre-treatment (4D-)MRI is acquired to propagate the contours. The treatment plan is then optimized and delivered on an MRI-Linac. In dark green are the parts of the workflow that have been addressed in this thesis. We have also highlighted the parts of the workflow that are already clinically addressed in our clinic by deep learning and areas of active research. We have also identified parts of the workflow that are not under active research for deep learning solutions but could benefit from DL or where deep learning is unlikely to be beneficial. Finally, speeding up the workflow (e.g., within 15 minutes) also enables new opportunities that were too time-consuming, such as acquiring quantitative MRI before treatment to aid outcome prediction.

enters the clinical workflow of MRIgRT, the MRI-Linac workflow of the future should be significantly accelerated such that patients are treated in fifteen minutes. The envisioned workflow is graphically depicted in Figure 6.4. However, several technical innovations must occur to achieve this acceleration.

First, the patient setup-up and pre-treatment phase should be significantly accelerated, and acquire motion-resolved MRI to build a motion model of the patient. Based on this information, deep learning models can fully re-segment the patient, propagate the GTV, CTV, and PTV from the initial treatment plan, and fully re-plan the treatment based on the pre-treatment MRI, taking the accumulated dose and treatment response into account. The pre-treatment phase should be completed within 10 minutes, reducing the burden of labor of technical staff and improving treatment efficiency. The prescribed dose and linear accelerator's properties limit the speed of irradiation. Treatment can commence with the consent of radiation oncologists, medical physicists, and RTTs (and possible manual plan adaptations). During irradiation, fast acquisitions use the pre-treatment motion model to perform accurate, real-time MLC tracking to deliver a high radiation dose to the tumor with (near)-zero treatment margin. During treatment, visual monitoring and feedback enables the operators to verify

the progress on the treatment and acting upon potentially impactful or unexpected events, such as bulk motion.

Despite this real-time nature, visual monitoring and feedback is still useful for operators to make decisions during the treatment, such as pausing the treatment during unexpected or potentially impactful events, such as bulk motion.

Deep learning can help facilitate an online adaptive MRIgRT workflow. For example, DL-based automatic segmentation is currently used clinically[86]. However, deep learning is unlikely to replace every step in the treatment workflow. For example, before treatment can commence, quality assurance should remain a manual task. Implementing a DL-enabled online adaptive MRIgRT workflow is a daunting task, but if proven successful will increase the quality of the treatment, increase the efficiency of radiation facilities, and reduce the workload on radiotherapy staff.

6.6 Concluding remarks

In the past decade, MRI-guided radiotherapy has made its clinical introduction. By performing inter-fraction treatment adaptation, the accuracy of radiation treatments is greatly improved when faced with sizable day-to-day anatomy variations. Resolving intra-fraction motion requires, depending on the motion source, accurate motion estimation with high spatio-temporal frequency. Deep learning is a good candidate for this problem due to its high inference speed.

In this thesis, we have presented techniques for managing intra-fraction motion with deep learning, demonstrating that deep learning can accurately and reliably guide radiation treatments. The technological developments of deep learning can enable real-time adaptive MRI-guided radiotherapy on MR-Linac devices. While intra-fraction motion tracking is one technical challenge for real-time adaptive MRIgRT, there are still challenges, such as the quality assurance of deep learning or real-time treatment planning. Real-time adaptive MRIgRT will unlock the true potential of MRI-Linac devices and redefine radiotherapy's impact as a pillar of cancer treatment. Even though deep learning needs significant efforts for validation and quality assurance, the technological advancements presented in this thesis show that real-time adaptive MRI-guided radiotherapy has the potential to have a clinical impact.

Chapter 6

CHAPTER 7

Samenvatting

Radiotherapie is een veelgebruikte methode voor de behandeling van kanker waarbij ioniserende straling wordt gebruikt om het DNA van kankercellen te vernietigen. Hierbij is het belangrijk om zoveel mogelijk dosis af te geven in de tumor, terwijl het gezonde weefsel zo min mogelijk geraakt wordt. Omdat gezond weefsel een groter herstellend vermogen heeft dan tumorweefsel, wordt een behandeling opgedeeld in fracties waardoor het gezonde weefsel kan herstellen tussen de behandelfracties. Als een tumor met zeer grote precisie kan worden bestraald, komt er minder dosis in gezond weefsel en zal dit resulteren in minder bijwerkingen. Bovendien zouden er minder behandelfracties nodig zijn om dezelfde dosis te bereiken.

Een complicerende factor in de behandeling is de beweging van de anatomie. Dit kan gebeuren tussen de fracties (interfractie beweging) door bijvoorbeeld verandering in de anatomie zoals blaasvulling, of verandering in de positionering van de patiënt op het behandelapparaat. Daarnaast vindt ook beweging van de anatomie plaats gedurende de fractie (intrafractie beweging) door bijvoorbeeld ademhaling, hartslag, of abdominale peristalsis.

Voor de optimale bestraling van bewegende tumoren is een belangrijke rol weggelegd voor beeldvormende technieken, zoals *computed tomography* (CT), *positron emission imaging* (PET), of *magnetic resonance imaging* (MRI). Binnen radiotherapie is MRI de laatste jaren een steeds grotere rol gaan spelen vanwege het uitmuntende zachte-weefsel contrast, veelzijdigheid in beeldvorming, en mogelijkheid tot beeldvorming tijdens de behandeling. Deze voordelen hebben geleid tot de ontwikkeling van een hybride MRI/bestraler (MRI-Linac), die de mogelijkheid biedt tot gelijktijdige beeldvorming en bestraling.

Met real-time adaptieve MRI-gestuurde radiotherapie (MRIgRT) is het mogelijk om de positie van de bestralingsstraal te synchroniseren met de positie en vorm van de tumor (tracking) door continu MRI te acquireren en de beweging ten opzichte van het behandelplan te bepalen. In deze behandelmodus, kan er zeer efficiënt bestraald worden met minimale foutmarges, wat leidt tot betere en snellere behandelingen.

Een grote uitdaging is dat deze synchronisatie in *real-time* moet gebeuren om de tumor accuraat te kunnen volgen. Hiertoe moet de MRI met minimale vertraging geacquireerd en gereconstrueerd worden, waarna een bewegingsveld uitgerekend moet worden die de grootte en richting van de beweging van de anatomie ten opzichte van een referentiepunt representeert, waarna de straal wordt bijgesteld. Voor het volgen van ademhalingsbeweging mag dit hele proces niet langer 400 milliseconden duren, wat ons dwingt tot een sterke onderbemonstering van het MRI signaal omdat de opname van hoge-kwaliteit MRI enkele minuten kan duren. Hoewel er veel onderzoek gedaan is naar de versnelling van de MRI acquisitie en reconstructie, is er momenteel geen methode die bewegingsvelden van hoge kwaliteit kan bepalen van MRI data binnen het beschikbare tijdsbudget.

Neurale netwerken lijken een veelbelovende oplossing voor dit probleem. Dit zijn non-lineaire modellen die grote hoeveelheden data gebruiken om te *leren* hoe een probleem opgelost moet worden. Hiertoe wordt de gemiddelde fout tussen de oplossing van het netwerk en de echte oplossing geminimaliseerd tijdens de trainingsfase. Na de training is het model direct bruikbaar, wat resulteert in zeer snelle oplossing. Zo kunnen deze modellen bijvoorbeeld worden ingezet voor tekstgeneratie of beeldherkenning, maar ook voor de verwerking van onderbemonsterde MRI beelden.

In dit proefschrift is onderzocht hoe neurale netwerken gebruikt kunnen worden om real-time MRI-gestuurde radiotherapie mogelijk te maken. In **Hoofdstuk 2** hebben we gekeken tijdens welke fase neurale netwerken het best ingezet kunnen worden: voor de beeldreconstructie, voor het berekenen van de beweging, voor allebei de fases, of helemaal niet. Hiertoe hebben wij gevonden dat neurale netwerken het effectiefst zijn als zij gebruikt worden voor de beeldregistratie in combinatie met traditionele beeldreconstructie. Dit leidt tot een model dat beweging kan bepalen met een fout van minder dan 1 millimeter terwijl er 25x minder data wordt gebruikt dan bij een traditionele MRI scan. Al met al kan er door inzet van neurale netwerken accurate bewegingsinformatie beschikbaar zijn binnen ongeveer 60 milliseconden, wat toereikend is voor real-time adaptieve MRI-gestuurde radiotherapie van organen die bewegen door ademhaling.

Dit gepresenteerde model rekent echter de beweging in twee dimensies, terwijl een tumor beweegt in drie dimensies. Daarom hebben we de resultaten uit **Hoofdstuk 2** gebruikt om het model uit te breiden naar een 3D bewegingsmodel in **Hoofdstuk 3**. Ook voor dit model laten we zien dat neurale netwerken in staat zijn om op een accurate manier beweging uit te rekenen: de fout van het model blijft binnen een pixel ten opzichte van minutenlange MRI scans terwijl dit model de beweging binnen 200 milliseconden kan uitrekenen. De accuraatheid van het model is gevalideerd met fantoomscans: MRI scans van niet-anatomische objecten waarvan we exact de beweging weten. Met deze fantoomscans hebben we laten zien dat zelfs bij een extreem snelle MRI scan de fout beperkt blijft binnen een pixel. Daarnaast krijgen neurale netwerken vaak de kritiek dat de training van zo'n model ervoor zorgt dat deze alleen maar werkt op een specifiek type scan. Hier hebben we het model getraind op MRI beelden toegepast op CT beelden, waarbij we zeer goede resultaten waarnemen. Dit geeft aan dat zulke modellen vaak goed blijven werken, zelfs in onvoorziene omstandigheden, wat klinische implementatie mogelijk maakt.

Neurale netwerken worden getraind om de parameters te vinden die een minimale fout opleveren gemeten over een gehele dataset. Een veelgekozen manier om de fout te bepalen is het kwadratische verschil tussen de goede waarde en modelwaarde $\epsilon = (f_\Theta(x) - y)^2$. In **Hoofdstuk 4** hebben we gekeken hoe deze fout zich gedraagt wanneer deze wordt toegepast op bewegingsvelden. Hiertoe hebben wij opzettelijk fouten in de richting en grootte van bewegingsvelden

geïntroduceerd. De analyse van deze foutbepalingsfunctie laat zien dat er een lagere fout aan bewegingsvelden wordt toegekend die de grootte van de beweging *onderschatten* ten opzichte van bewegingsvelden die de grootte *overschatten*, oftewel deze is asymmetrisch. Dit kan grote gevolgen hebben voor de toepassing van neurale netwerken voor MRI-gestuurde radiotherapie, aangezien een structurele fout in de geschatte beweging een structurele fout in de bestraling zal veroorzaken. Om dit op te lossen hebben we een nieuwe functie ontwikkeld om de fout in bewegingsvelden te bepalen, genaamd $\perp$-Loss. Deze foutbepalingsfunctie is symmetrisch en resulteert in modellen die betere bewegingsvelden produceren dan wanneer het kwadratisch verschil wordt geminimaliseerd.

Ten behoeve van de behandeling van mobiele tumoren wordt vaak een zogenaamde vierdimensionale MRI gemaakt. Dit wil zeggen, een 3D-MRI voor verschillende fases van de ademhaling. De acquisitie en reconstructie van dit soort scans met hoge kwaliteit duurt echter zeer lang, wat 4D-MRI lastig toe te passen maakt in de kliniek. De acquisitie en reconstructie van 4D-MRI zou versneld kunnen worden met neurale netwerken, maar dit soort modellen zijn computationeel erg lastig te trainen vanwege de grootte van de MRI scan. Daardoor kan er geen gebruik worden gemaakt van alle spatiële en temporele informatie beschikbaar in 4D-MRI. In **Hoofdstuk 5** presenteren we een zeer efficiënte architectuur voor de reconstructie van 4D-MRI, genaamd MODEST, die ruim dertig maal kleiner is (d.w.z., 30x minder trainbare parameters) dan vergelijkbare technieken, terwijl nog steeds alle spatio-temporele informatie gebruikt wordt. Hiermee laten we zien dat deze 4D-MRI veel sneller opgenomen kunnen worden en in slechts twee minuten al beschikbaar kunnen zijn met dezelfde kwaliteit als de langzame scans.

De gepresenteerde werken in dit proefschrift laten zien dat neurale netwerken een veelbelovende techniek zijn voor alle stappen van MRI-gestuurde radiotherapie. De toepassing van neurale netwerken belooft een versnelling van de MRI scans die nodig zijn om een behandeling te starten, automatisering van bepaalde tijdrovende taken zoals de detectie van tumoren en segmentatie van tumoren en gezond weefsel, en het mogelijk maken van nieuwe toepassingen zoals tijdsgetrouwe adaptieve MRI-gestuurde radiotherapie. Indien succesvol toegepast in de kliniek zal dit resulteren in efficiëntere, betere, goedkopere en minder behandelingen, wat een verbetering is voor de patiënt (betere behandeling en minder vaak naar de kliniek), het personeel (lagere werkdruk), en de samenleving (goedkopere behandelingen en hogere kwaliteit van leven). Dit proefschrift heeft laten zien dat neurale netwerken toepassen voor het oplossen van deze problemen technisch mogelijk is, maar er zal nog een lange weg te gaan zijn voordat deze in de praktijk gebruikt kunnen worden. Een heikel punt blijft de *in-vivo* validatie van de real-time bewegingsmodellen, een tot op heden onopgelost probleem. Niettegenstaande deze uitdagingen staat het buiten kijf dat neurale netwerken een geschikte techniek zijn om real-time adaptieve MRI-gestuurde radiotherapie mogelijk te maken.

Bibliography

1. Ferlay, J. *et al.* Cancer statistics for the year 2020: An overview. *International Journal of Cancer* **149,** 778–789. ISSN: 0020-7136 (Aug. 2021).

2. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians* **71,** 209–249. ISSN: 0007-9235 (May 2021).

3. Yap, M. L., Zubizarreta, E., Bray, F., Ferlay, J. & Barton, M. Global Access to Radiotherapy Services: Have We Made Progress During the Past Decade? *Journal of Global Oncology* **2,** 207–215. ISSN: 2378-9506 (Aug. 2016).

4. Chandra, R. A., Keane, F. K., Voncken, F. E. & Thomas, C. R. Contemporary radiotherapy: present and future. *The Lancet* **398,** 171–184. ISSN: 1474547X (July 2021).

5. Curie-Skłodowska, M. Sur l'étude des courbes de probabilité relatives à l'action des rayons X sur les bacilles. *Comptes rendus l'Académie des Sci* **188,** 202–204 (1929).

6. Schultheiss, T. E., Orton, C. G. & Peck, R. A. Models in radiotherapy: Volume effects. *Medical Physics* **10,** 410–415. ISSN: 00942405 (July 1983).

7. Mitchell, G. The Rationale for Fractionation in Radiotherapy. *Clinical Journal of Oncology Nursing* **17,** 412–417. ISSN: 1092-1095 (Aug. 2013).

8. Chang, J. Y. Stereotactic ablative radiotherapy: aim for a cure of cancer. *Annals of translational medicine* **3,** 12. ISSN: 2305-5839. arXiv: 25705644 (Jan. 2015).

9. Folkert, M. R. & Timmerman, R. D. Stereotactic ablative body radiosurgery (SABR) or Stereotactic body radiation therapy (SBRT). *Advanced Drug Delivery Reviews* **109,** 3–14 (Jan. 2017).

10. ICRU50. *Prescribing, recording, and reporting photon beam therapy* (International Commission on Radiation Units and Measurements, 1993).

11. Seppenwoolde, Y. *et al.* Precise and real-time measurement of 3D tumor motion in lung due to breathing and heartbeat, measured during radiotherapy. *International Journal of Radiation Oncology*Biology*Physics* **53,** 822–834 (July 2002).

12. Liu, H. H. *et al.* Assessing Respiration-Induced Tumor Motion and Internal Target Volume Using Four-Dimensional Computed Tomography for Radiotherapy of Lung Cancer. *International Journal of Radiation Oncology*Biology*Physics* **68,** 531–540. ISSN: 03603016 (June 2007).

13. Heerkens, H. D. *et al.* MRI-based tumor motion characterization and gating schemes for radiation therapy of pancreatic cancer. *Radiotherapy and Oncology* **111,** 252–257. ISSN: 0167-8140 (2014).

14. Von Siebenthal, M. *et al.* 4D MR imaging of respiratory organ motion and its variability. *Physics in Medicine and Biology* **52,** 1547–1564 (Feb. 2007).

15. Ford, E. C., Mageras, G. S., Yorke, E. & Ling, C. C. Respiration-correlated spiral CT: A method of measuring respiratory-induced anatomic motion for radiation treatment planning. *Medical Physics* **30,** 88–97. ISSN: 00942405 (Dec. 2002).

16. Paulson, E. S. *et al.* 4D-MRI driven MR-guided online adaptive radiotherapy for abdominal stereotactic body radiation therapy on a high field MR-Linac: Implementation and initial clinical experience. *Clinical and Translational Radiation Oncology* **23,** 72–79. ISSN: 24056308 (July 2020).

17. Dirix, P., Haustermans, K. & Vandecaveye, V. The Value of Magnetic Resonance Imaging for Radiotherapy Planning. *Seminars in Radiation Oncology* **24.** Magnetic Resonance Imaging in Radiation Oncology, 151–159. ISSN: 1053-4296 (2014).

18. Van Herk, M. Errors and margins in radiotherapy. *Seminars in Radiation Oncology* **14,** 52–64. ISSN: 10534296 (Jan. 2004).

19. Aird, E. G. A. & Conway, J. CT simulation for radiotherapy treatment planning. *The British Journal of Radiology* **75,** 937–949. ISSN: 0007-1285 (Dec. 2002).

20. Jarritt, P. H., Carson, K. J., Hounsell, A. R. & Visvikis, D. The role of PET/CT scanning in radiotherapy planning. *The British Journal of Radiology* **79,** S27–S35. ISSN: 0007-1285 (Sept. 2006).

21. Lagendijk, J. J., Raaymakers, B. W. & van Vulpen, M. The Magnetic Resonance Imaging–Linac System. *Seminars in Radiation Oncology* **24,** 207–209. ISSN: 10534296 (July 2014).

22. Kreß, J., Minohara, S., Endo, M., Debus, J. & Kanai, T. Patient position verification using CT images. *Medical Physics* **26,** 941–948 (June 1999).

23. Barney, B. M. *et al.* Image-Guided Radiotherapy (IGRT) for Prostate Cancer Comparing kV Imaging of Fiducial Markers With Cone Beam Computed Tomography (CBCT). *International Journal of Radiation Oncology*Biology*Physics* **80,** 301–305. ISSN: 0360-3016 (2011).

24. Morr, J., DiPetrillo, T., Tsai, J.-S., Engler, M. & Wazer, D. E. Implementation and utility of a daily ultrasound-based localization system with intensity-modulated radiotherapy for prostate cancer. *International Journal of Radiation Oncology,Biology,Physics* **53,** 1124–1129 (Aug. 2002).

25. Raaymakers, B. W. *et al.* Integrating a 1.5 T MRI scanner with a 6 MV accelerator: proof of concept. *Physics in Medicine and Biology* **54,** N229–N237 (May 2009).

26. Dobbs, H. J. & Parker, R. P. The respective roles of the simulator and computed tomography in radiotherapy planning: A review. *Clinical Radiology* **35,** 433–439. ISSN: 00099260 (Jan. 1984).

27. Lauterbur, P. C. Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance. *Nature* **242,** 190–191. ISSN: 0028-0836 (Mar. 1973).

28. Damadian, R. Tumor Detection by Nuclear Magnetic Resonance. *Science* **171,** 1151–1153. ISSN: 0036-8075 (Mar. 1971).

29. Raaymakers, B. W. *et al.* First patients treated with a 1.5 T MRI-Linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment. *Physics in Medicine & Biology* **62,** L41–L50. ISSN: 1361-6560 (Nov. 2017).

30. Mutic, S. & Dempsey, J. F. The ViewRay System: Magnetic Resonance–Guided and Controlled Radiotherapy. *Seminars in Radiation Oncology* **24,** 196–199. ISSN: 10534296 (July 2014).

31. Keall, P. J., Barton, M. & Crozier, S. The Australian Magnetic Resonance Imaging–Linac Program. *Seminars in Radiation Oncology* **24,** 203–206. ISSN: 10534296 (July 2014).

32. Fallone, B. G. The Rotating Biplanar Linac–Magnetic Resonance Imaging System. *Seminars in Radiation Oncology* **24,** 200–202. ISSN: 10534296 (July 2014).

33. Winkel, D. *et al.* Adaptive radiotherapy: The Elekta Unity MR-linac concept. *Clinical and translational radiation oncology* **18,** 54–59. ISSN: 2405-6308 (Sept. 2019).

34. Winkel, D. *et al.* Individual lymph nodes: "See it and Zap it". *Clinical and Translational Radiation Oncology* **18,** 46–53. ISSN: 24056308 (Sept. 2019).

35. Takao, S. *et al.* Intrafractional Baseline Shift or Drift of Lung Tumor Motion During Gated Radiation Therapy With a Real-Time Tumor-Tracking System. *International Journal of Radiation Oncology*Biology*Physics* **94,** 172–180 (Jan. 2016).

36. Keall, P. J., Kini, V. R., Vedam, S. S. & Mohan, R. Potential radiotherapy improvements with respiratory gating. *Australasian Physics & Engineering Sciences in Medicine* **25,** 1–6. ISSN: 0158-9938 (Mar. 2002).

37. Ehrbar, S. *et al.* MR-guided beam gating: Residual motion, gating efficiency and dose reconstruction for stereotactic treatments of the liver and lung. *Radiotherapy and Oncology* **174,** 101–108. ISSN: 0167-8140 (2022).

38. Cerviño, L. I., Du, J. & Jiang, S. B. MRI-guided tumor tracking in lung cancer radiotherapy. *Physics in Medicine and Biology* **56,** 3773–3785. ISSN: 0031-9155 (July 2011).

39. Horn, B. K. & Schunck, B. G. Determining optical flow. *Artificial Intelligence* **17** (ed Pearson, J. J.) 185–203. ISSN: 00043702 (Aug. 1981).

40. Zachiu, C., Papadakis, N., Ries, M., Moonen, C. & Denis de Senneville, B. An improved optical flow tracking technique for real-time MR-guided beam therapies in moving organs. *Physics in Medicine and Biology* **60,** 9003–9029. ISSN: 0031-9155 (Dec. 2015).

41. Zachiu, C., de Senneville, B. D., Moonen, C. T. W., Raaymakers, B. W. & Ries, M. Anatomically plausible models and quality assurance criteria for online mono- and multi-modal medical image registration. *Physics in Medicine & Biology* **63,** 155016 (Aug. 2018).

42. Klein, S., Staring, M., Murphy, K., Viergever, M. & Pluim, J. elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Transactions on Medical Imaging* **29,** 196–205. ISSN: 0278-0062 (Jan. 2010).

43. Foskey, M. *et al.* Large deformation three-dimensional image registration in image-guided radiation therapy. *Physics in Medicine and Biology* **50,** 5869–5892. ISSN: 0031-9155 (Dec. 2005).

44. Ruan, D., Kupelian, P. & Low, D. A. Image-Guided Positioning and Tracking. *The Cancer Journal* **17,** 155–158. ISSN: 1528-9117 (May 2011).

45. Keall, P., Poulsen, P. & Booth, J. T. See, Think, and Act: Real-Time Adaptive Radiotherapy. *Seminars in Radiation Oncology* **29,** 228–235. ISSN: 10534296 (July 2019).

46. Brand, D., Kirby, A., Yarnold, J. & Somaiah, N. How Low Can You Go? The Radiobiology of Hypofractionation. *Clinical Oncology* **34,** 280–287 (May 2022).

47. Gough, J., Hall, W., Good, J., Nash, A. & Aitken, K. Technical Radiotherapy Advances – The Role of Magnetic Resonance Imaging-Guided Radiation in the Delivery of Hypofractionation. *Clinical Oncology* **34,** 301–312 (May 2022).

48. Dietz, B., Fallone, B. G. & Wachowicz, K. Nomenclature for real-time magnetic resonance imaging. *Magnetic Resonance in Medicine* **81,** 1483–1484. ISSN: 0740-3194 (Mar. 2019).

49. Keall, P. J. *et al.* The management of respiratory motion in radiation oncology report of AAPM Task Group 76. *Medical Physics* **33,** 3874–3900. ISSN: 00942405 (Sept. 2006).

50. Keall, P. J. *et al.* ICRU REPORT 97: MRI-Guided Radiation Therapy Using MRI-Linear Accelerators. *Journal of the ICRU* **22,** 1–100 (Dec. 2022).

51. Bjerre, T. *et al.* Three-dimensional MRI-linac intra-fraction guidance using multiple orthogonal cine-MRI planes. *Physics in Medicine and Biology* **58,** 4943–4950 (June 2013).

52. *Magnetic Resonance Imaging* (eds Brown, R. W., Cheng, Y.-C. N., Haacke, E. M., Thompson, M. R. & Venkatesan, R.) ISBN: 9781118633953 (John Wiley & Sons Ltd, Chichester, UK, Apr. 2014).

53. Shannon, C. E. A Mathematical Theory of Communication. *The Bell System Technical Journal* **27,** 379–423 (1948).

54. Fourier, J. B. J., Darboux, G., *et al. Théorie analytique de la chaleur* (Didot Paris, 1822).

55. Cooley, J. W. & Tukey, J. W. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation* **19,** 297–301. ISSN: 0025-5718 (1965).

56. Wright, K. L., Hamilton, J. I., Griswold, M. A., Gulani, V. & Seiberlich, N. Non-Cartesian parallel imaging reconstruction. *Journal of Magnetic Resonance Imaging* **40,** 1022–1040. ISSN: 10531807 (Nov. 2014).

57. Nishimura, D. G., Jackson, J. I. & Pauly, J. M. On the nature and reduction of the displacement artifact in flow images. *Magnetic Resonance in Medicine* **22,** 481–492. ISSN: 07403194 (Dec. 1991).

58. Robison, R. K., Devaraj, A. & Pipe, J. G. Fast, simple gradient delay estimation for spiral MRI. *Magnetic Resonance in Medicine* **63,** 1683–1690 (June 2010).

59. Peters, D. C., Derbyshire, J. A. & McVeigh, E. R. Centering the projection reconstruction trajectory: Reducing gradient delay errors. *Magnetic Resonance in Medicine* **50,** 1–6 (June 2003).

60. Moussavi, A., Untenberger, M., Uecker, M. & Frahm, J. Correction of gradient-induced phase errors in radial MRI. *Magnetic Resonance in Medicine* **71,** 308–312 (Feb. 2013).

61. Fessler, J. & Sutton, B. Nonuniform fast fourier transforms using min-max interpolation. *IEEE Transactions on Signal Processing* **51,** 560–574. ISSN: 1053-587X (Feb. 2003).

62. Stemkens, B., Paulson, E. S. & Tijssen, R. H. N. Nuts and bolts of 4D-MRI for radiotherapy. *Physics in Medicine & Biology* **63,** 21TR01 (Oct. 2018).

63. Bruijnen, T. *et al. Free-breathing motion compensated 4D (3D+respiration) T2-weighted turbo spin-echo MRI for body imaging* arXiv. 2022.

64. Peters, D. C. *et al.* Undersampled projection reconstruction applied to MR angiography. *Magnetic Resonance in Medicine* **43,** 91–101. ISSN: 0740-3194 (Jan. 2000).

65. Barger, A. V., Block, W. F., Toropov, Y., Grist, T. M. & Mistretta, C. A. Time-resolved contrast-enhanced imaging with isotropic resolution and broad coverage using an undersampled 3D projection trajectory. *Magnetic Resonance in Medicine* **48,** 297–305. ISSN: 0740-3194 (Aug. 2002).

66. Landau, H. J. Necessary density conditions for sampling and interpolation of certain entire functions. *Acta Mathematica* **117,** 37–52 (1967).

67. Pruessmann, K. P., Weiger, M., Scheidegger, M. B. & Boesiger, P. SENSE: sensitivity encoding for fast MRI. *Magnetic resonance in medicine* **42,** 952–62. ISSN: 0740-3194 (Nov. 1999).

68. Griswold, M. A. *et al.* Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med* **47,** 1202–1210 (June 2002).

69. Uecker, M. *et al.* ESPIRiT-an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA. *Magnetic Resonance in Medicine* **71,** 990–1001. ISSN: 07403194 (Mar. 2014).

70. Wiesinger, F., Boesiger, P. & Pruessmann, K. P. Electrodynamics and ultimate SNR in parallel MR imaging. *Magnetic Resonance in Medicine* **52,** 376–390. ISSN: 0740-3194 (Aug. 2004).

71. Lustig, M., Donoho, D. & Pauly, J. M. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine* **58,** 1182–1195. ISSN: 07403194 (Dec. 2007).

72. Jaspan, O. N., Fleysher, R. & Lipton, M. L. Compressed sensing MRI: a review of the clinical literature. *The British Journal of Radiology* **88,** 20150487. ISSN: 0007-1285 (Dec. 2015).

73. Usman, M. *et al.* Motion corrected compressed sensing for free-breathing dynamic cardiac MRI. *Magnetic Resonance in Medicine* **70,** 504–516 (Aug. 2012).

74. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review* **65,** 386–408. ISSN: 1939-1471 (1958).

75. McCulloch, W. S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics* **5,** 115–133. ISSN: 0007-4985 (Dec. 1943).

76. Kingma, D. P. & Ba, J. *Adam: A Method for Stochastic Optimization* in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (eds Bengio, Y. & LeCun, Y.) (2015).

77. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323,** 533–536. ISSN: 0028-0836 (Oct. 1986).

78. Linnainmaa, S. Taylor expansion of the accumulated rounding error. *BIT* **16,** 146–160. ISSN: 0006-3835 (June 1976).

79. Hornik, K., Stinchcombe, M. & White, H. Multilayer feedforward networks are universal approximators. *Neural Networks* **2,** 359–366. ISSN: 08936080 (Jan. 1989).

80. Vapnik, V. N. & Chervonenkis, A. Y. On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities. *Theory of Probability & Its Applications* **16,** 264–280. ISSN: 0040-585X (Jan. 1971).

81. Fukushima, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics* **36,** 193–202. ISSN: 0340-1200 (Apr. 1980).

82. LeCun, Y. *et al.* Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation* **1,** 541–551. ISSN: 0899-7667 (Dec. 1989).

83. Hinton, G. E., Osindero, S. & Teh, Y.-W. A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation* **18,** 1527–1554. ISSN: 0899-7667 (July 2006).

84. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural networks* **61,** 85–117 (2015).

85. Krizhevsky, A., Sutskever, I. & Hinton, G. E. *ImageNet Classification with Deep Convolutional Neural Networks* in *Advances in Neural Information Processing Systems* (eds Pereira, F., Burges, C., Bottou, L. & Weinberger, K.) **25** (Curran Associates, Inc., 2012).

86. Savenije, M. H. *et al.* Clinical implementation of MRI-based organs-at-risk auto-segmentation with convolutional networks for prostate radiotherapy. *Radiation Oncology* **15,** 104. ISSN: 1748717X (Dec. 2020).

87. Eppenhof, K. A. J. & Pluim, J. P. W. Pulmonary CT Registration through Supervised Learning with Convolutional Neural Networks. *IEEE Transactions on Medical Imaging* **38,** 1097–1105 (2019).

88. Eppenhof, K. *et al.* Fast contour propagation for MR-guided prostate radiotherapy using convolutional neural networks. *Medical Physics* **47,** 1238–1248. ISSN: 0094-2405 (Mar. 2020).

89. Hammernik, K. *et al.* Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine* **79,** 3055–3071. ISSN: 07403194 (June 2018).

90. Schlemper, J., Caballero, J., Hajnal, J. V., Price, A. N. & Rueckert, D. A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction. *IEEE Transactions on Medical Imaging* **37,** 491–503. ISSN: 0278-0062 (Feb. 2018).

91. Zhu, B., Liu, J. Z., Cauley, S. F., Rosen, B. R. & Rosen, M. S. Image reconstruction by domain-transform manifold learning. *Nature* **555,** 487–492. ISSN: 0028-0836 (Mar. 2018).

92. Biswas, S., Aggarwal, H. K. & Jacob, M. Dynamic MRI using model-based deep learning and SToRM priors: MoDL-SToRM. *Magnetic Resonance in Medicine* **82,** 485–494. ISSN: 0740-3194 (July 2019).

93. Lønning, K. *et al.* Recurrent inference machines for reconstructing heterogeneous MRI data. *Medical Image Analysis* **53,** 64–78. ISSN: 13618415 (Apr. 2019).

94. Zhang, C. *et al. Scan-Specific Residual Convolutional Neural Networks for Fast MRI Using Residual RAKI* in *2019 53rd Asilomar Conference on Signals, Systems, and Computers* (IEEE, Nov. 2019), 1476–1480. ISBN: 978-1-7281-4300-2.

95. Harms, J. *et al.* Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography. *Medical Physics* **46,** 3998–4009. ISSN: 0094-2405 (Sept. 2019).

96. Maspero, M. *et al.* Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy. *Physics in Medicine & Biology* **63,** 185001. ISSN: 1361-6560 (Sept. 2018).

97. Glitzner, M., de Senneville, B. D., Lagendijk, J. J. W., Raaymakers, B. W. & Crijns, S. P. M. On-line 3D motion estimation using low resolution MRI. *Physics in Medicine and Biology* **60,** N301–N310 (Aug. 2015).

98. Fast, M. F. *et al.* Tumour auto-contouring on 2D cine MRI for locally advanced lung cancer: A comparative study. *Radiotherapy and Oncology* **125,** 485–491. ISSN: 01678140 (Dec. 2017).

99. Kontaxis, C. *et al.* Towards fast online intrafraction replanning for free-breathing stereotactic body radiation therapy with the MR-linac. *Physics in Medicine & Biology* **62,** 7233–7248. ISSN: 1361-6560 (Aug. 2017).

100. Dietz, B. *et al.* 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 & Biology* **64,** 195002. ISSN: 1361-6560 (Sept. 2019).

101. Borman, P. T. S. *et al.* Characterization of imaging latency for real-time MRI-guided radiotherapy. *Physics in Medicine & Biology* **63,** 155023. ISSN: 1361-6560 (Aug. 2018).

102. Zijlema, S. E. *et al.* Design and feasibility of a flexible, on-body, high impedance coil receive array for a 1.5 T MR-linac. *Physics in Medicine & Biology* **64,** 185004. ISSN: 1361-6560 (Sept. 2019).

103. Litjens, G. *et al.* A survey on deep learning in medical image analysis. *Medical Image Analysis* **42,** 60–88. ISSN: 13618415 (Dec. 2017).

104. Meyer, P., Noblet, V., Mazzara, C. & Lallement, A. Survey on deep learning for radiotherapy. *Computers in Biology and Medicine* **98,** 126–146. ISSN: 00104825 (July 2018).

105. Sahiner, B. *et al.* Deep learning in medical imaging and radiation therapy. *Medical Physics* **46,** e1–e36. ISSN: 00942405 (Jan. 2019).

106. Ranjan, Anurag and Black, Michael J. *Optical Flow Estimation Using a Spatial Pyramid Network* in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (July 2017), 4161–4170.

107. Dosovitskiy, A. *et al. FlowNet: Learning Optical Flow with Convolutional Networks* in *2015 IEEE International Conference on Computer Vision (ICCV)* (IEEE, Dec. 2015), 2758–2766. ISBN: 978-1-4673-8391-2. arXiv: 1504.06852.

108. Ilg, E. *et al.* *FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks* in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, July 2017), 1647–1655. ISBN: 978-1-5386-0457-1.

109. Schlemper, J. *et al.* *dAUTOMAP: Decomposing AUTOMAP to Achieve Scalability and Enhance Performance* in *Proceedings of the International Society of Magnetic Resonance in Medicine* (2019).

110. Lin, J.-M. Python Non-Uniform Fast Fourier Transform (PyNUFFT): An Accelerated Non-Cartesian MRI Package on a Heterogeneous Platform (CPU/GPU). *Journal of Imaging* **4,** 51. ISSN: 2313-433X (Mar. 2018).

111. Glorot, X. & Bengio, Y. *Understanding the difficulty of training deep feedforward neural networks* in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (eds Teh, Y. W. & Titterington, M.) **9** (PMLR, Chia Laguna Resort, Sardinia, Italy, 2010), 249–256.

112. Ravishankar, S. & Bresler, Y. MR Image Reconstruction From Highly Undersampled k-Space Data by Dictionary Learning. *IEEE Transactions on Medical Imaging* **30,** 1028–1041. ISSN: 0278-0062 (May 2011).

113. Zachiu, C., Denis de Senneville, B., Moonen, C. & Ries, M. A framework for the correction of slow physiological drifts during MR-guided HIFU therapies: Proof of concept. *Medical Physics* **42,** 4137–4148. ISSN: 00942405 (June 2015).

114. He, K., Zhang, X., Ren, S. & Sun, J. *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification* in *2015 IEEE International Conference on Computer Vision (ICCV)* (IEEE, Dec. 2015), 1026–1034. ISBN: 978-1-4673-8391-2.

115. Bloice, M. D., Roth, P. M. & Holzinger, A. Biomedical image augmentation using Augmentor. *Bioinformatics* **35** (ed Murphy, R.) 4522–4524. ISSN: 1367-4803 (Nov. 2019).

116. Wang, Z., Bovik, A., Sheikh, H. & Simoncelli, E. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing* **13,** 600–612. ISSN: 1057-7149 (Apr. 2004).

117. Altman, D. G. & Bland, J. M. Measurement in Medicine: The Analysis of Method Comparison Studies. *The Statistician* **32,** 307–317. ISSN: 00390526 (Sept. 1983).

118. Farnebäck, G. Two-Frame Motion Estimation Based on Polynomial Expansion. *Lecture Notes in Computer Science* **2749,** 363–370 (2003).

119. Seegoolam, G. *et al.* Exploiting Motion for Deep Learning Reconstruction of Extremely-Undersampled Dynamic MRI. *Lecture Notes in Computer Science* **11767,** 704–712 (2019).

120. Stemkens, B., Tijssen, R. H. N., de Senneville, B. D., Lagendijk, J. J. W. & van den Berg, C. A. T. Image-driven, model-based 3D abdominal motion estimation for MR-guided radiotherapy. *Physics in Medicine and Biology* **61,** 5335–5355 (June 2016).

121. Haskell, M. W. *et al.* Network Accelerated Motion Estimation and Reduction (NAMER): Convolutional neural network guided retrospective motion correction using a separable motion model. *Magnetic Resonance in Medicine* **82,** 1452–1461. ISSN: 0740-3194 (Oct. 2019).

122. Lagendijk, J. J. W. *et al.* MR guidance in radiotherapy. *Physics in Medicine and Biology* **59,** R349–R369. ISSN: 0031-9155 (Nov. 2014).

123. Morales, Manuel A. and Izquierdo-Garcia, David and Aganj, Iman and Kalpathy-Cramer, Jayashree and Rosen, Bruce R. and Catana, Ciprian. Implementation and Validation of a Three-dimensional Cardiac Motion Estimation Network. eng. *Radiology. Artificial intelligence* **1.** 32076659[pmid], e180080–e180080. ISSN: 2638-6100 (July 2019).

124. Kim, T. H. & Haldar, J. P. *The Fourier radial error spectrum plot: A more nuanced quantitative evaluation of image reconstruction quality* in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (IEEE, Apr. 2018), 61–64. ISBN: 978-1-5386-3636-7.

125. Menten, M. J., Wetscherek, A. & Fast, M. F. MRI-guided lung SBRT: Present and future developments. *Physica Medica* **44,** 139–149. ISSN: 1120-1797 (2017).

126. Schmidt, M. A. & Payne, G. S. Radiotherapy planning using MRI. *Physics in Medicine and Biology* **60,** R323–R361 (Oct. 2015).

127. Keall, P. J. *et al.* AAPM Task Group 264: The safe clinical implementation of MLC tracking in radiotherapy. *Medical Physics* **48.** ISSN: 0094-2405 (May 2021).

128. Gibson, J. & Marques, O. *Sparse Regularization of TV-L1 Optical Flow* in *Image and Signal Processing* (eds Elmoataz, A., Lezoray, O., Nouboud, F. & Mammass, D.) (Springer International Publishing, Cham, 2014), 460–467. ISBN: 978-3-319-07998-1.

129. Barth, M., Breuer, F., Koopmans, P. J., Norris, D. G. & Poser, B. A. Simultaneous multislice (SMS) imaging techniques. *Magnetic Resonance in Medicine* **75,** 63–81. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.25897` (2016).

130. Ong, F. *et al.* Extreme MRI: Large-scale volumetric dynamic imaging from continuous non-gated acquisitions. *Magnetic Resonance in Medicine* **84,** 1763–1780. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.28235` (2020).

131. Dietz, B. *et al.* Single patient convolutional neural networks for real-time MR reconstruction: coherent low-resolution versus incoherent undersampling. *Physics in Medicine & Biology* **65,** 08NT03 (Apr. 2020).

132. Eppenhof, K. A. J., Lafarge, M. W., Veta, M. & Pluim, J. P. W. Progressively Trained Convolutional Neural Networks for Deformable Image Registration. *IEEE Transactions on Medical Imaging* **39,** 1594–1604 (2020).

133. Küstner, T. *et al.* CINENet: deep learning-based 3D cardiac CINE MRI reconstruction with multi-coil complex-valued 4D spatio-temporal convolutions. *Scientific Reports* **10,** 13710. ISSN: 2045-2322 (Aug. 2020).

134. Terpstra, M. L. *et al.* Deep learning-based image reconstruction and motion estimation from undersampled radial k-space for real-time MRI-guided radiotherapy. *Physics in Medicine & Biology* **65,** 155015 (Aug. 2020).

135. Larson, A. C. *et al.* Self-gated cardiac cine MRI. *Magnetic Resonance in Medicine* **51,** 93–102. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.10664 (2004).

136. Zhang, T. *et al.* Robust self-navigated body MRI using dense coil arrays. *Magnetic Resonance in Medicine* **76,** 197–205. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.25858 (2016).

137. Feng, L. *et al.* XD-GRASP: Golden-angle radial MRI with reconstruction of extra motion-state dimensions using compressed sensing. *Magnetic Resonance in Medicine* **75,** 775–788. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.25665 (2016).

138. Stemkens, B. *et al.* Optimizing 4-Dimensional Magnetic Resonance Imaging Data Sampling for Respiratory Motion Analysis of Pancreatic Tumors. *International Journal of Radiation Oncology*Biology*Physics* **91,** 571–578. ISSN: 0360-3016 (2015).

139. Knoll, F., Schwarzl, A., Diwoky, C. & Sodickson, D. K. *gpuNUFFT-an open source GPU library for 3D regridding with direct Matlab interface* in *Proceedings of the 22nd annual meeting of ISMRM, Milan, Italy* (2014), 4297.

140. Sun, D., Yang, X., Liu, M.-Y. & Kautz, J. *PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume* in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2018).

141. Hui, T.-W., Tang, X. & Loy, C. C. A Lightweight Optical Flow CNN —Revisiting Data Fidelity and Regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **43,** 2555–2569 (2021).

142. Glorot, X., Bordes, A. & Bengio, Y. *Deep Sparse Rectifier Neural Networks* in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* (eds Gordon, G., Dunson, D. & Dudík, M.) **15** (JMLR Workshop and Conference Proceedings, Fort Lauderdale, FL, USA, Apr. 2011), 315–323.

143. Pérez-García, F., Sparks, R. & Ourselin, S. TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *arXiv:2003.04696 [cs, eess, stat]*. arXiv: 2003.04696 (Mar. 2020).

144. Segars, W. P., Mahesh, M., Beck, T. J., Frey, E. C. & Tsui, B. M. W. Realistic CT simulation using the 4D XCAT phantom. *Medical Physics* **35,** 3800–3808. eprint: `https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1118/1.2955743` (2008).

145. Paganelli, C. *et al.* A tool for validating MRI-guided strategies: a digital breathing CT/MRI phantom of the abdominal site. *Medical & Biological Engineering & Computing* **55,** 2001–2014. ISSN: 1741-0444 (Nov. 2017).

146. Eiben, B. *et al.* Consistent and invertible deformation vector fields for a breathing anthropomorphic phantom: a post-processing framework for the XCAT phantom. *Physics in Medicine & Biology* **65,** 165005 (Aug. 2020).

147. Vandemeulebroucke, Jef and Sarrut, David and Clarysse, Patrick. *The POPI-model, a point-validated pixel-based breathing thorax model* in *Proceeding of the XVth International Conference on the use of Computers in Radiation therapy (ICCR)* **2** (2007), 195–199.

148. Zhang, S., Block, K. T. & Frahm, J. Magnetic resonance imaging in real time: Advances using radial FLASH. *Journal of Magnetic Resonance Imaging* **31,** 101–109. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmri.21987` (2010).

149. Marstal, K., Berendsen, F., Dekker, N., Staring, M. & Klein, S. *The Continuous Registration Challenge: Evaluation-as-a-Service for Medical Image Registration Algorithms* in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* (2019), 1399–1402.

150. Eppenhof, K. A. J. & Pluim, J. P. W. Error estimation of deformable image registration of pulmonary CT scans using convolutional neural networks. *Journal of Medical Imaging* **5,** 1–11 (2018).

151. Thedens, D. R., Irarrazaval, P., Sachs, T. S., Meyer, C. H. & Nishimura, D. G. Fast magnetic resonance coronary angiography with a three-dimensional stack of spirals trajectory. *Magnetic Resonance in Medicine* **41,** 1170–1179. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291522-2594%28199906%2941%3A6%3C1170%3A%3AAID-MRM13%3E3.0.CO%3B2-J` (1999).

152. Johnson, K. M. Hybrid radial-cones trajectory for accelerated MRI. *Magnetic Resonance in Medicine* **77,** 1068–1081. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.26188` (2017).

153. Prieto, C. *et al.* Highly efficient respiratory motion compensated free-breathing coronary mra using golden-step Cartesian acquisition. *Journal of Magnetic Resonance Imaging* **41,** 738–746. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmri.24602` (2015).

154. Navest, R. J. M. *et al.* The noise navigator for MRI-guided radiotherapy: an independent method to detect physiological motion. *Physics in Medicine & Biology* **65,** 12NT01 (June 2020).

155. Vercauteren, T., Pennec, X., Perchant, A. & Ayache, N. Diffeomorphic demons: efficient non-parametric image registration. *Neuroimage* **45,** 61–72 (Mar. 2009).

156. De Vos, B. D. *et al. Mutual information for unsupervised deep learning image registration* in *Medical Imaging 2020: Image Processing* (eds Landman, B. A. & Išgum, I.) (SPIE, Mar. 2020).

157. Huttinga, N. R. F., van den Berg, C. A. T., Luijten, P. R. & Sbrizzi, A. MR-MOTUS: model-based non-rigid motion estimation for MR-guided radiotherapy using a reference image and minimal k-space data. *Physics in Medicine & Biology* **65,** 015004 (Jan. 2020).

158. Janssens, G. *et al.* Evaluation of nonrigid registration models for interfraction dose accumulation in radiotherapy. *Medical Physics* **36,** 4268–4276. eprint: `https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1118/1.3194750` (2009).

159. Batchelor, P. G. *et al.* Matrix description of general motion correction applied to multishot images. *Magnetic Resonance in Medicine* **54,** 1273–1280. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.20656` (2005).

160. Rank, C. M. *et al.* 4D respiratory motion-compensated image reconstruction of free-breathing radial MR data with very high undersampling. *Magnetic Resonance in Medicine* **77,** 1170–1183. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.26206` (2017).

161. Wijlemans, J. *et al.* Magnetic resonance-guided high-intensity focused ultrasound (MR-HIFU) ablation of liver tumours. *Cancer Imaging* **12,** 387 (2012).

162. Rogers, T., Ratnayaka, K. & Lederman, R. J. MRI catheterization in cardiopulmonary disease. *Chest* **145,** 30–36 (2014).

163. Zhang, B. *et al.* Peripheral nerve stimulation properties of head and body gradient coils of various sizes. *Magnetic Resonance in Medicine* **50,** 50–58 (2003).

164. Collins, C. M. & Wang, Z. Calculation of radiofrequency electromagnetic fields and their effects in MRI of human subjects. *Magnetic Resonance in Medicine* **65,** 1470–1482 (2011).

165. Wright, G. Magnetic resonance imaging. *IEEE Signal Processing Magazine* **14,** 56–66 (1997).

166. Bustin, A., Fuin, N., Botnar, R. M. & Prieto, C. From Compressed-Sensing to Artificial Intelligence-Based Cardiac MRI Reconstruction. *Frontiers in Cardiovascular Medicine* **7,** 17. ISSN: 2297-055X (2020).

167. Lingala, S. G., Sutton, B. P., Miquel, M. E. & Nayak, K. S. Recommendations for real-time speech MRI. *Journal of Magnetic Resonance Imaging* **43,** 28–44 (2016).

168. Keiper, T. D. *et al.* Feasibility of real-time motion tracking using cine MRI during MR-guided radiation therapy for abdominal targets. *Medical Physics* **47,** 3554–3566 (2020).

169. Adler, J. & Öktem, O. Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Problems* **33,** 124007 (Nov. 2017).

170. Fu, Y. *et al.* A review of deep learning based methods for medical image multi-organ segmentation. *Physica Medica* **85,** 107–122. ISSN: 1120-1797 (2021).

171. Terpstra, M. L. *et al.* Real-time 3D motion estimation from undersampled MRI using multi-resolution neural networks. *Medical Physics* **48,** 6597–6613 (2021).

172. Jaubert, O. *et al.* Real-time deep artifact suppression using recurrent U-Nets for low-latency cardiac MRI. *Magnetic Resonance in Medicine* **86,** 1904–1916 (2021).

173. Kanai, S., Yamada, M., Takahashi, H., Yamanaka, Y. & Ida, Y. Smoothness Analysis of Loss Functions of Adversarial Training. *arXiv preprint arXiv:2103.01400* (2021).

174. Patel, D. & Sastry, P. S. *Memorization in Deep Neural Networks: Does the Loss Function Matter?* in *Advances in Knowledge Discovery and Data Mining* (eds Karlapalem, K. *et al.*) (Springer International Publishing, Cham, 2021), 131–142. ISBN: 978-3-030-75765-6.

175. Kunin, D., Sagastuy-Brena, J., Ganguli, S., Yamins, D. L. & Tanaka, H. *Neural Mechanics: Symmetry and Broken Conservation Laws in Deep Learning Dynamics* in *International Conference on Learning Representations* (2021).

176. Janocha, K. & Czarnecki, W. M. On loss functions for deep neural networks in classification. *arXiv preprint arXiv:1702.05659* (2017).

177. Butler, D. J., Wulff, J., Stanley, G. B. & Black, M. J. *A naturalistic open source movie for optical flow evaluation* in *European Conf. on Computer Vision (ECCV)* (ed A. Fitzgibbon et al. (Eds.)) (Springer-Verlag, Oct. 2012), 611–625.

178. Haji-Valizadeh, H. *et al.* Comparison of Complex k-Space Data and Magnitude-Only for Training of Deep Learning–Based Artifact Suppression for Real-Time Cine MRI. *Frontiers in Physics* **9,** 475 (2021).

179. Fermüller, C., Shulman, D. & Aloimonos, Y. The statistics of optical flow. *Computer Vision and Image Understanding* **82,** 1–32 (2001).

180. Fuller, W. A. *Measurement error models* (John Wiley & Sons, 2009).

181. Gleser, L. J. Estimation in a multivariate "errors in variables" regression model: large sample results. *The Annals of Statistics,* 24–44 (1981).

182. Nagel, H.-H. & Haag, M. *Bias-corrected optical flow estimation for road vehicle tracking* in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)* (1998), 1006–1011.

183. Buczko, M. & Willert, V. *Monocular Outlier Detection for Visual Odometry* in *2017 IEEE Intelligent Vehicles Symposium (IV)* (2017), 739–745.

184. Heckbert, P. *Graphics gems IV (IBM version)* (Elsevier, 1994).

185. Kim, B. & Günther, T. Robust Reference Frame Extraction from Unsteady 2D Vector Fields with Convolutional Neural Networks. *Computer Graphics Forum (Proc. EuroVis)* **38,** 285–295 (2019).

186. Sriram, A. *et al. End-to-End Variational Networks for Accelerated MRI Reconstruction* in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020* (eds Martel, A. L. *et al.*) (Springer International Publishing, Cham, 2020), 64–73. ISBN: 978-3-030-59713-9.

187. Zbontar, J. *et al. fastMRI: An Open Dataset and Benchmarks for Accelerated MRI* 2018.

188. Putzky, P. & Welling, M. Recurrent inference machines for solving inverse problems. *arXiv preprint arXiv:1706.04008* (2017).

189. Teuwen, J., Moriakov, N., Karkalousos, D., Caan, M. & Yiasemis, G. *DIRECT* https://github.com/directgroup/direct. 2020.

190. Beauferris, Y. *et al. Multi-channel MR Reconstruction (MC-MRRec) Challenge – Comparing Accelerated MR Reconstruction Models and Assessing Their Genereralizability to Datasets Collected with Different Coils* 2020. arXiv: 2011.07952 [eess.IV].

191. Inati, S., Hansen, M. & Kellman, P. *A fast optimal method for coil sensitivity estimation and adaptive coil combination for complex images* in *Proceedings of the 22nd Annual Meeting of ISMRM, Milan* (2014), 4407.

192. Ronneberger, O., Fischer, P. & Brox, T. *U-Net: Convolutional Networks for Biomedical Image Segmentation* in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (eds Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F.) (Springer International Publishing, Cham, 2015), 234–241. ISBN: 978-3-319-24574-4.

193. Sheikh, H. R. & Bovik, A. C. *A visual information fidelity approach to video quality assessment* in *The First International Workshop on Video Processing and Quality Metrics for Consumer Electronics* **7** (2005).

194. Mason, A. *et al.* Comparison of Objective Image Quality Metrics to Expert Radiologists' Scoring of Diagnostic Quality of MR Images. *IEEE Transactions on Medical Imaging* **39,** 1064–1072 (2020).

195. Reinhardt, J. M. *et al.* Registration-based estimates of local lung tissue expansion compared to xenon CT measures of specific ventilation. *Med Image Anal* **12,** 752–763 (Dec. 2008).

196. Cardoso, M. J. *et al. MONAI: An open-source framework for deep learning in healthcare* 2022. arXiv: `2211.02701 [cs.LG]`.

197. Ulyanov, D., Vedaldi, A. & Lempitsky, V. *Instance Normalization: The Missing Ingredient for Fast Stylization* 2017. arXiv: `1607.08022 [cs.CV]`.

198. Loshchilov, I. & Hutter, F. *Decoupled Weight Decay Regularization* 2019. arXiv: `1711.05101 [cs.LG]`.

199. Pezzotti, N. *et al. An Adaptive Intelligence Algorithm for Undersampled Knee MRI Reconstruction* 2020. arXiv: `2004.07339 [eess.IV]`.

200. Cole, E., Cheng, J., Pauly, J. & Vasanawala, S. Analysis of deep complex-valued convolutional neural networks for MRI reconstruction and phase-focused applications. *Magnetic Resonance in Medicine* **86,** 1093–1109 (2021).

201. Wang, Y. & Liu, T. Quantitative susceptibility mapping (QSM): Decoding MRI data for a tissue magnetic biomarker. *Magnetic Resonance in Medicine* **73,** 82–101 (July 2014).

202. Markl, M., Frydrychowicz, A., Kozerke, S., Hope, M. & Wieben, O. 4D flow MRI. *Journal of Magnetic Resonance Imaging* **36,** 1015–1036 (2012).

203. Zhao, L., Mammadov, M. & Yearwood, J. *From Convex to Nonconvex: A Loss Function Analysis for Binary Classification* in *2010 IEEE International Conference on Data Mining Workshops* (2010), 1281–1288.

204. Fu, L. & Wang, Y.-G. Robust regression with asymmetric loss functions. *Statistical Methods in Medical Research* **30,** 1800–1815 (2021).

205. Cai, J. *et al.* Estimation of Error in Maximal Intensity Projection-Based Internal Target Volume of Lung Tumors: A Simulation and Comparison Study Using Dynamic Magnetic Resonance Imaging. *International Journal of Radiation Oncology*Biology*Physics* **69,** 895–902. ISSN: 0360-3016 (2007).

206. Park, J. C. *et al.* Liver motion during cone beam computed tomography guided stereotactic body radiation therapy. *Medical Physics* **39,** 6431–6442. eprint: `https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1118/1.4754658` (2012).

207. Ekberg, L., Holmberg, O., Wittgren, L., Bjelkengren, G. & Landberg, T. What margins should be added to the clinical target volume in radiotherapy treatment planning for lung cancer? *Radiotherapy and Oncology* **48,** 71–77. ISSN: 0167-8140 (1998).

208. Song, R., Tipirneni, A., Johnson, P., Loeffler, R. B. & Hillenbrand, C. M. Evaluation of respiratory liver and kidney movements for MRI navigator gating. *Journal of Magnetic Resonance Imaging* **33,** 143–148. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmri.22418` (2011).

209. Feng, M. *et al.* Characterization of Pancreatic Tumor Motion Using Cine MRI: Surrogates for Tumor Position Should Be Used With Caution. *International Journal of Radiation Oncology*Biology*Physics* **74,** 884–891. ISSN: 0360-3016 (2009).

210. Wood, M. L. & Henkelman, R. M. MR image artifacts from periodic motion. *Medical physics* **12 2,** 143–51 (1985).

211. Noterdaeme, O., Gleeson, F., Phillips, R. R. & Brady, M. Quantification of missing and overlapping data in multiple breath hold abdominal imaging. *European Journal of Radiology* **64.** Ultrasound Imaging Special Issue, 273–278. ISSN: 0720-048X (2007).

212. Ozhasoglu, C. & Murphy, M. J. Issues in respiratory motion compensation during external-beam radiotherapy. *International Journal of Radiation Oncology*Biology*Physics* **52,** 1389–1399. ISSN: 0360-3016 (2002).

213. Bussels, B. *et al.* Respiration-induced movement of the upper abdominal organs: a pitfall for the three-dimensional conformal radiation treatment of pancreatic cancer. *Radiotherapy and Oncology* **68,** 69–74. ISSN: 0167-8140 (2003).

214. Vedam, S. S. *et al.* Acquiring a four-dimensional computed tomography dataset using an external respiratory signal. *Physics in Medicine and Biology* **48,** 45–62 (Dec. 2002).

215. Deng, Z. *et al.* Four-dimensional MRI using three-dimensional radial sampling with respiratory self-gating to characterize temporal phase-resolved respiratory motion in the abdomen. *Magnetic Resonance in Medicine* **75,** 1574–1585. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.25753` (2016).

216. Rietzel, E. *et al.* Design of 4D treatment planning target volumes. *International Journal of Radiation Oncology*Biology*Physics* **66,** 287–295. ISSN: 0360-3016 (2006).

217. Sentker, T. *et al.* 4D CT image artifacts affect local control in SBRT of lung and liver metastases. *Radiotherapy and Oncology* **148,** 229–234. ISSN: 0167-8140 (2020).

218. Mutaf, Y. D., Antolak, J. A. & Brinkmann, D. H. The impact of temporal inaccuracies on 4DCT image quality. *Medical Physics* **34,** 1615–1622. eprint: `https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1118/1.2717404` (2007).

219. Paganelli, C., Summers, P., Bellomi, M., Baroni, G. & Riboldi, M. Liver 4DMRI: A retrospective image-based sorting method. *Medical Physics* **42,** 4814–4821. eprint: `https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1118/1.4927252` (2015).

220. Mickevicius, N. J. & Paulson, E. S. Investigation of undersampling and reconstruction algorithm dependence on respiratory correlated 4D-MRI for online MR-guided radiation therapy. *Physics in Medicine and Biology* **62,** 2910–2921 (Mar. 2017).

221. Larkman, D. J. *et al.* Use of multicoil arrays for separation of signal from multiple slices simultaneously excited. *Journal of Magnetic Resonance Imaging* **13,** 313–317. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/1522-2586%28200102%2913%3A2%3C313%3A%3AAID-JMRI1045%3E3.0.CO%3B2-W` (2001).

222. Keijnemans, K. *et al.* Simultaneous multi-slice accelerated 4D-MRI for radiotherapy guidance. *Physics in Medicine & Biology* **66,** 095014 (Apr. 2021).

223. Keijnemans, K. *et al.* A hybrid 2D/4D-MRI methodology using simultaneous multislice imaging for radiotherapy guidance. *Medical Physics* **49,** 6068–6081 (June 2022).

224. Biswas, S., Aggarwal, H. K. & Jacob, M. Dynamic MRI using model-based deep learning and SToRM priors: MoDL-SToRM. *Magnetic Resonance in Medicine* **82,** 485–494. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.27706` (2019).

225. Machado, I. P. *et al.* *A Deep Learning-based Integrated Framework for Quality-aware Undersampled Cine Cardiac MRI Reconstruction and Analysis* `https://arxiv.org/abs/2205.01673`. 2022. arXiv: `2205.01673 [eess.IV]`.

226. Freedman, J. N. *et al.* Rapid 4D-MRI reconstruction using a deep radial convolutional neural network: Dracula. *Radiotherapy and Oncology* **159,** 209–217. ISSN: 0167-8140 (2021).

227. Qin, C. *et al.* Convolutional Recurrent Neural Networks for Dynamic MR Image Reconstruction. *IEEE Transactions on Medical Imaging* **38,** 280–290 (2019).

228. Kofler, A., Dewey, M., Schaeffter, T., Wald, C. & Kolbitsch, C. Spatio-Temporal Deep Learning-Based Undersampling Artefact Reduction for 2D Radial Cine MRI With Limited Training Data. *IEEE Transactions on Medical Imaging* **39,** 703–717 (2020).

229. Virtue, P., Yu, S. X. & Lustig, M. *Better than real: Complex-valued neural nets for MRI fingerprinting* in *2017 IEEE International Conference on Image Processing (ICIP)* (2017), 3953–3957.

230. Kerfoot, E. *et al.* *Left-Ventricle Quantification Using Residual U-Net* in *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges* (eds Pop, M. *et al.*) (Springer International Publishing, Cham, 2019), 371–380. ISBN: 978-3-030-12029-0.

231. Muckley, M. J., Stern, R., Murrell, T. & Knoll, F. *TorchKbNufft: A High-Level, Hardware-Agnostic Non-Uniform Fast Fourier Transform* in *ISMRM Workshop on Data Sampling & Image Reconstruction* Source code available at `https://github.com/mmuckley/torchkbnufft`. (2020).

232. Terpstra, M. L., Maspero, M., Sbrizzi, A. & van den Berg, C. A. ⊥-loss: a symmetric loss function for magnetic resonance imaging reconstruction and image registration with deep learning. *Medical Image Analysis* **48,** 6597–6613 (2022).

233. Jin, X., Yuan, X., Feng, J. & Yan, S. Training Skinny Deep Neural Networks with Iterative Hard Thresholding Methods. *CoRR* **abs/1607.05423.** arXiv: `1607.05423` (2016).

234. Gourdeau, D., Duchesne, S. & Archambault, L. On the proper use of structural similarity for the robust evaluation of medical image synthesis models. *Medical Physics* **49,** 2462–2474. eprint: `https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1002/mp.15514` (Feb. 2022).

235. van de Lindt, T. *et al.* A Self-Sorting Coronal 4D-MRI Method for Daily Image Guidance of Liver Lesions on an MR-LINAC. *International Journal of Radiation Oncology*Biology*Physics* **102.** Imaging in Radiation Oncology, 875–884. ISSN: 0360-3016 (2018).

236. Pech-Pacheco, J., Cristobal, G., Chamorro-Martinez, J. & Fernandez-Valdivia, J. *Diatom autofocusing in brightfield microscopy: a comparative study* in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000* **3** (IEEE Comput. Soc, 2000), 314–317. ISBN: 0-7695-0750-6.

237. Nowlan, S. J. & Hinton, G. E. Simplifying Neural Networks by Soft Weight-Sharing. *Neural Computation* **4,** 473–493 (1992).

238. Miotto, R., Wang, F., Wang, S., Jiang, X. & Dudley, J. T. Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics* **19,** 1236–1246. ISSN: 1477-4054. eprint: `https://academic.oup.com/bib/article-pdf/19/6/1236/27119191/bbx044.pdf` (May 2017).

239. Chen, Y. *et al.* AI-Based Reconstruction for Fast MRI—A Systematic Review and Meta-Analysis. *Proceedings of the IEEE* **110,** 224–245 (2022).

240. Qin, C. *et al. Joint Learning of Motion Estimation and Segmentation for Cardiac MR Image Sequences* in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* (eds Frangi, A. F., Schnabel, J. A., Davatzikos, C., Alberola-López, C. & Fichtinger, G.) (Springer International Publishing, Cham, 2018), 472–480. ISBN: 978-3-030-00934-2.

241. Vaswani, A. *et al. Attention is All you Need* in *Advances in Neural Information Processing Systems* (eds Guyon, I. *et al.*) **30** (Curran Associates, Inc., 2017).

242. Feng, L., Tyagi, N. & Otazo, R. MRSIGMA: Magnetic Resonance SIGnature MAtching for real-time volumetric imaging. *Magnetic Resonance in Medicine* **84,** 1280–1292. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.28200` (2020).

243. Barbone, M. *et al. Efficient Online 4D Magnetic Resonance Imaging* in *2021 IEEE 33rd International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)* (2021), 177–187.

244. Lebel, R. M., Jones, J., Ferre, J.-C., Law, M. & Nayak, K. S. Highly accelerated dynamic contrast enhanced imaging. *Magnetic Resonance in Medicine* **71,** 635–644. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.24710` (2014).

245. Chen, J., Liu, S. & Huang, M. Low-Rank and Sparse Decomposition Model for Accelerating Dynamic MRI Reconstruction. *Journal of Healthcare Engineering* **2017,** 9856058. ISSN: 2040-2295 (Aug. 2017).

246. Kim, N., Tringale, K. R., Crane, C., Tyagi, N. & Otazo, R. MR SIGnature MAtching (MRSIGMA) with retrospective self-evaluation for real-time volumetric motion imaging. *Physics in Medicine & Biology* **66,** 215009 (Oct. 2021).

247. Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J. & Dalca, A. V. VoxelMorph: A Learning Framework for Deformable Medical Image Registration. *IEEE Transactions on Medical Imaging* **38,** 1788–1800 (2019).

248. Küstner, T. *et al.* LAPNet: Non-rigid Registration derived in k-space for Magnetic Resonance Imaging. *IEEE Transactions on Medical Imaging* **40,** 3686–3697 (2021).

249. Hammernik, K., Knoll, F., Sodickson, D. K. & Pock, T. *L2 or not L2: impact of loss function design for deep learning MRI reconstruction* in *ISMRM 25th Annual Meeting* (2017), 0687.

250. Pluim, J., Maintz, J. & Viergever, M. Image registration by maximization of combined mutual information and gradient information. *IEEE Transactions on Medical Imaging* **19,** 809–814 (2000).

251. Berendsen, F. F., Kotte, A. N. T. J., Viergever, M. A. & Pluim, J. P. W. *Registration of organs with sliding interfaces and changing topologies* in *SPIE Proceedings* (eds Ourselin, S. & Styner, M. A.) (SPIE, Mar. 2014).

252. Grimbergen, G. *et al.* Dosimetric impact of intrafraction motion under abdominal compression during MR-guided SBRT for (Peri-) pancreatic tumors. *Physics in Medicine & Biology* **67,** 185016 (Sept. 2022).

253. De Vos, B. D. *et al.* A deep learning framework for unsupervised affine and deformable image registration. *Medical Image Analysis* **52,** 128–143 (Feb. 2019).

254. Zheng, Y., Jiang, S. & Yang, Z. Deformable registration of chest CT images using a 3D convolutional neural network based on unsupervised learning. *Journal of Applied Clinical Medical Physics* **22,** 22–35 (Sept. 2021).

255. Anandan, P. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision* **2,** 283–310 (Jan. 1989).

256. Tsao, J., Boesiger, P. & Pruessmann, K. P. k-t BLAST and k-t SENSE: dynamic MRI with high frame rate exploiting spatiotemporal correlations. en. *Magn. Reson. Med.* **50,** 1031–1042 (Nov. 2003).

257. Hochreiter, S. & Schmidhuber, J. Long Short-Term Memory. *Neural Computation* **9,** 1735–1780 (1997).

258. Dosovitskiy, A. *et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale* in *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021* (OpenReview.net, 2021).

259. Greaves-Tunnell, A. & Harchaoui, Z. *A Statistical Investigation of Long Memory in Language and Music* in *Proceedings of the 36th International Conference on Machine Learning* (eds Chaudhuri, K. & Salakhutdinov, R.) **97** (PMLR, June 2019), 2394–2403.

260. Tay, Y., Dehghani, M., Bahri, D. & Metzler, D. Efficient Transformers: A Survey. *ACM Computing Surveys* **55,** 1–28 (Dec. 2022).

261. Feng, L. Golden-Angle Radial MRI: Basics, Advances, and Applications. *Journal of Magnetic Resonance Imaging* **56,** 45–62 (Apr. 2022).

262. Pappas, E. P. *et al.* MRI-Related Geometric Distortions in Stereotactic Radiotherapy Treatment Planning: Evaluation and Dosimetric Impact. *Technology in Cancer Research & Treatment* **16,** 1120–1129 (Oct. 2017).

263. Bruijnen, T., Stemkens, B., Lagendijk, J. J., van den Berg, C. A. & Tijssen, R. H. *Gradient system characterization of a 1.5 T MRI-Linac with application to UTE imaging* in *International Society for Magnetic Resonance in Medicine (ISMRM)* **235** (2018).

264. Küstner, T. *et al.* Fully self-gated free-running 3D Cartesian cardiac CINE with isotropic whole-heart coverage in less than 2 min. *NMR in Biomedicine* **34** (Sept. 2020).

265. Wang, G., Luo, T., Nielsen, J.-F., Noll, D. C. & Fessler, J. A. B-Spline Parameterized Joint Optimization of Reconstruction and K-Space Trajectories (BJORK) for Accelerated 2D MRI. *IEEE Transactions on Medical Imaging* **41,** 2318–2330 (2022).

266. Schultz, G. *et al.* Reconstruction of MRI data encoded with arbitrarily shaped, curvilinear, nonbijective magnetic fields. *Magnetic Resonance in Medicine* **64,** 1390–1403. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.22393` (2010).

267. Tran, E. H. *et al.* Evaluation of MRI-derived surrogate signals to model respiratory motion. *Biomedical Physics & Engineering Express* **6,** 045015 (June 2020).

268. Dixon, W. T. Simple proton spectroscopic imaging. *Radiology* **153,** 189–194 (Oct. 1984).

269. Gatehouse, P. D. *et al.* Applications of phase-contrast flow and velocity imaging in cardiovascular MRI. *European Radiology* **15,** 2172–2184 (July 2005).

270. Osako, K., Singh, R. & Raj, B. *Complex recurrent neural networks for denoising speech signals* in *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)* (IEEE, Oct. 2015).

271. Wu, L., Juhas, P., Yoo, S. & Robinson, I. Complex imaging of phase domains by deep neural networks. *IUCrJ* **8,** 12–21 (Jan. 2021).

272. Geman, S., Bienenstock, E. & Doursat, R. Neural Networks and the Bias/Variance Dilemma. *Neural Computation* **4,** 1–58 (Jan. 1992).

273. Huttinga, N. R. F., Bruijnen, T., Berg, C. A. T. & Sbrizzi, A. Nonrigid 3D motion estimation at high temporal resolution from prospectively undersampled k-space data using low-rank MR-MOTUS. *Magnetic Resonance in Medicine* **85,** 2309–2326 (Nov. 2020).

274. Otazo, R., Candès, E. & Sodickson, D. K. Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components. *Magnetic Resonance in Medicine* **73,** 1125–1136 (Apr. 2014).

275. Lingala, S. G., Hu, Y., DiBella, E. & Jacob, M. Accelerated Dynamic MRI Exploiting Sparsity and Low-Rank Structure: k-t SLR. *IEEE Transactions on Medical Imaging* **30,** 1042–1054 (May 2011).

276. Mardani, M., Cheng, J., Pauly, J. & Xing, L. *Fast 4D MRI Reconstruction Analytics using Low-Rank Tensor Imputation* in *Proc. Intl. Soc. Mag. Reson. Med* **25** (2017), 3862.

277. Thomas, D. H. *et al.* Is there an ideal set of prospective scan acquisition phases for fast-helical based 4D-CT? *Physics in Medicine & Biology* **61,** N632 (Nov. 2016).

278. Van de Lindt, T. *et al.* MRI-guided mid-position liver radiotherapy: Validation of image processing and registration steps. *Radiotherapy and Oncology* **138,** 132–140 (Sept. 2019).

279. Xiao, H. *et al.* A dual-supervised deformation estimation model (DDEM) for constructing ultra-quality 4D-MRI based on a commercial low-quality 4D-MRI for liver cancer radiation therapy. *Medical Physics* **49,** 3159–3170 (Feb. 2022).

280. Uijtewaal, P. *et al.* First experimental demonstration of VMAT combined with MLC tracking for single and multi fraction lung SBRT on an MR-linac. *Radiotherapy and Oncology* **174,** 149–157. ISSN: 0167-8140 (2022).

281. Zheng, S., Song, Y., Leung, T. & Goodfellow, I. *Improving the Robustness of Deep Neural Networks via Stability Training* in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE Computer Society, Los Alamitos, CA, USA, June 2016), 4480–4488.

282. Qu, L., Balachandar, N. & Rubin, D. L. *An Experimental Study of Data Heterogeneity in Federated Learning Methods for Medical Imaging* 2021.

283. Abdar, M. *et al.* A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion* **76,** 243–297. ISSN: 1566-2535 (2021).

284. Hüllermeier, E. & Waegeman, W. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Machine Learning* **110,** 457–506 (Mar. 2021).

285. van der Velden, B. H., Kuijf, H. J., Gilhuijs, K. G. & Viergever, M. A. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Medical Image Analysis* **79,** 102470. ISSN: 1361-8415 (2022).

286. Barragán-Montero, A. *et al.* Towards a safe and efficient clinical implementation of machine learning in radiation oncology by exploring model interpretability, explainability and data-model dependency. *Physics in Medicine & Biology* **67,** 11TR01 (May 2022).

287. Low, D. A., Harms, W. B., Mutic, S. & Purdy, J. A. A technique for the quantitative evaluation of dose distributions. *Medical Physics* **25,** 656–661 (May 1998).

288. Bauer, S., Markl, M., Honal, M. & Jung, B. A. The effect of reconstruction and acquisition parameters for GRAPPA-based parallel imaging on the image quality. *Magnetic Resonance in Medicine* **66,** 402–409. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.22803` (2011).

289. Sharma, S. D., Fong, C. L., Tzung, B. S., Law, M. & Nayak, K. S. Clinical image quality assessment of accelerated magnetic resonance neuroimaging using compressed sensing. en. *Invest. Radiol.* **48,** 638–645 (Sept. 2013).

290. Feng, L. *et al.* Compressed sensing for body MRI. *Journal of Magnetic Resonance Imaging* **45,** 966–987. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmri.25547` (2017).

291. Kabus, S. *et al. Evaluation of 4D-CT Lung Registration* in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009* (eds Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A. & Taylor, C.) (Springer Berlin Heidelberg, Berlin, Heidelberg, 2009), 747–754. ISBN: 978-3-642-04268-3.

292. Ehman, R. L. & Felmlee, J. P. Adaptive technique for high-definition MR imaging of moving structures. *Radiology* **173,** 255–263 (Oct. 1989).

293. Bertholet, J. *et al.* Real-time intrafraction motion monitoring in external beam radiotherapy. *Physics in Medicine & Biology* **64,** 15TR01. ISSN: 0031-9155 (Aug. 2019).

294. Drangova, M., Bowman, B. & Pelc, N. J. Physiologic motion phantom for MRI applications. *Journal of Magnetic Resonance Imaging* **6,** 513–518 (May 1996).

295. Swailes, N. E., MacDonald, M. E. & Frayne, R. Dynamic phantom with heart, lung, and blood motion for initial validation of MRI techniques. *Journal of Magnetic Resonance Imaging* **34,** 941–946 (July 2011).

296. Dunn, L. *et al.* A programmable motion phantom for quality assurance of motion management in radiotherapy. *Australasian Physical & Engineering Sciences in Medicine* **35,** 93–100 (Nov. 2011).

297. Hanson, H. M., Eiben, B., McClelland, J. R., Herk, M. & Rowland, B. C. Technical Note: Four-dimensional deformable digital phantom for MRI sequence development. *Medical Physics* **48,** 5406–5413 (Aug. 2021).

298. Russakovsky, O. *et al.* ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)* **115,** 211–252 (2015).

299. Lin, T.-Y. *et al.* in *Computer Vision - ECCV 2014* 740–755 (Springer International Publishing, 2014).

300. Brown, T. B. *et al. Language Models are Few-Shot Learners* 2020.

301. Schuhmann, C. *et al. LAION-5B: An open large-scale dataset for training next generation image-text models* 2022.

302. Wang, X. *et al. ChestX-ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases* in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (July 2017).

303. Baid, U. *et al.* The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification. *CoRR* **abs/2107.02314.** arXiv: `2107.02314` (2021).

304. Ghaffari, M., Sowmya, A. & Oliver, R. Automated Brain Tumor Segmentation Using Multimodal Brain Scans: A Survey Based on Models Submitted to the BraTS 2012–2018 Challenges. *IEEE Reviews in Biomedical Engineering* **13,** 156–168 (2020).

305. Zhang, R. *et al. MICCAI Abdominal Multi-Organ Segmentation Challenge 2022* Mar. 2022.

306. Desai, A. D. *et al. SKM-TEA: A Dataset for Accelerated MRI Reconstruction with Dense Image Labels for Quantitative Clinical Evaluation* in *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)* (2021).

307. De Mol van Otterloo, S. R. *et al.* The MOMENTUM Study: An International Registry for the Evidence-Based Introduction of MR-Guided Adaptive Therapy. *Frontiers in Oncology* **10** (Sept. 2020).

308. Matakos, A., Balter, J. M. & Cao, Y. A robust method for estimating B0 inhomogeneity field in the liver by mitigating fat signals and phase-wrapping. en. *Tomography* **3,** 79–88 (June 2017).

309. Henke, L. *et al.* Magnetic Resonance Image-Guided Radiotherapy (MRIgRT): A 4.5-Year Clinical Experience. *Clinical Oncology* **30,** 720–727 (Nov. 2018).

310. Goodburn, R. J. *et al.* The future of MRI in radiation therapy: Challenges and opportunities for the MR community. *Magnetic Resonance in Medicine* **88,** 2592–2608. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.29450` (2022).

Publications

Published articles

M.L. Terpstra, M. Maspero, F. d'Agata, B. Stemkens, M.P.W. Intven, J.J.W. Lagendijk, C.A.T. van den Berg & R.H.N Tijssen (2020). Deep learning-based image reconstruction and motion estimation from undersampled radial k-space for real-time MRI-guided radiotherapy. *Physics in Medicine & Biology*, 65(15), 155015.

M. L. Terpstra, M. Maspero, T. Bruijnen, J.J.C. Verhoeff, J.J.W. Lagendijk & C.A.T. van den Berg (2021). Real-time 3D motion estimation from undersampled MRI using multi-resolution neural networks. *Medical Physics*, 48:6597–6613.

M. L. Terpstra, M. Maspero, A. Sbrizzi & C.A.T. van den Berg (2022). ⊥-loss: A symmetric loss function for magnetic resonance imaging reconstruction and image registration with deep learning. *Medical Image Analysis*, 80, 102509.

M. L. Terpstra, M. Maspero, J.J.C. Verhoeff & C.A.T. van den Berg (2023). Accelerated respiratory-resolved 4D-MRI with separable spatio-temporal neural networks. *Medical Physics*, 50:5331–5342.

Conference proceedings

M.L. Terpstra, F. D'Agata, B. Stemkens, J.J.W. Lagendijk, C.A.T. van den Berg, R.H.N. Tijssen *Estimating 2D Deformation Vector Fields from Golden Angle Radial Undersampled k-space Using Stacked Convolutional Neural Networks.* Proc. ImagO conference, Utrecht, 2019 (oral presentation)

M.L. Terpstra, F. D'Agata, B. Stemkens, J.J.W. Lagendijk, C.A.T. van den Berg, R.H.N. Tijssen *Estimating 2D Deformation Vector Fields from Golden Angle Radial Undersampled k-space Using Stacked Convolutional Neural Networks.* Proc. MRinRT, Toronto, 2019 (oral presentation)

M.L. Terpstra, F. D'Agata, B. Stemkens, J.J.W. Lagendijk, C.A.T. van den Berg, R.H.N. Tijssen *Real-time estimation of 2D deformation vector fields from highly undersampled, dynamic k-space for MRI-guided radiotherapy using deep learning.* Proc. ISMRM Benelux Chapter, Arnhem, 2020 (oral presentation)

M.L. Terpstra, F. D'Agata, B. Stemkens, J.J.W. Lagendijk, C.A.T. van den Berg, R.H.N. Tijssen *Real-time estimation of 2D deformation vector fields from highly undersampled, dynamic k-space for MRI-guided radiotherapy using deep learning.* Proc. ISMRM, On-line, 2020 (oral presentation)

M.L. Terpstra, F. D'Agata, B. Stemkens, J.J.W. Lagendijk, C.A.T. van den Berg, R.H.N. Tijssen *Reconstructing non-Cartesian acquisitions using dAUTOMAP.* Proc. ISMRM, On-line, 2020 (digital poster)

M.L. Terpstra, M. Maspero, T. Bruijnen, J.J.C. Verhoeff, J.J.W. Lagendijk, C.A.T. van den Berg *Time-resolved motion estimation for real-time MRI-guided radiotherapy.* NVKF Kringdag, On-line, 2021 (oral presentation)

M.L. Terpstra, M. Maspero, T. Bruijnen, J.J.C. Verhoeff, J.J.W. Lagendijk, C.A.T. van den Berg *A generic framework for real-time 3D motion estimation from highly undersampled k-space using deep learning.* Proc. ISMRM Benelux chapter, On-line, 2021 (oral presentation)

M.L. Terpstra, M. Maspero, T. Bruijnen, J.J.C. Verhoeff, J.J.W. Lagendijk, C.A.T. van den Berg *A generic framework for real-time 3D motion estimation from highly undersampled k-space using deep learning.* Proc. ISMRM, On-line, 2021 (oral presentation)

M.L. Terpstra, M. Maspero, J.J.W. Lagendijk, C.A.T. van den Berg *Rethinking complex image reconstruction: ⊥-loss for improved complex image reconstruction with deep learning.* Proc. ISMRM, On-line, 2021 (digital poster)

M.L. Terpstra, M. Maspero, J.J.C. Verhoeff, C.A.T. van den Berg *Accelerated respiratory-resolved 4D-MRI with separable spatio-temporal neural networks.* Proc. ISMRM Benelux chapter, Maastricht, 2022 (oral presentation)

M.L. Terpstra, M. Maspero, J.J.C. Verhoeff, C.A.T. van den Berg *Accelerated respiratory-resolved 4D-MRI with separable spatio-temporal neural networks.* Proc. ISMRM, London, 2022 (power pitch)

M.L. Terpstra, M. Maspero, J.J.C. Verhoeff, C.A.T. van den Berg *Accelerated respiratory-resolved 4D-MRI with separable spatio-temporal neural networks.* ISMRM Workshop on Motion detection & correction, Oxford, 2022 (oral presentation)

M.L. Terpstra, S. Verschuren, T. Bruijnen, M. Maspero, C.A.T. van den Berg *Feasibility of automatic patient-specific sequence optimization with deep reinforcement learning.* ISMRM Benelux chapter, Brussels, 2023 (poster presentation)

M.L. Terpstra, S. Verschuren, T. Bruijnen, M. Maspero, C.A.T. van den Berg *Feasibility of automatic patient-specific sequence optimization with deep reinforcement learning.* Proc. ISMRM, Toronto, 2023 (oral presentation)

M.L. Terpstra, K. Hammernik, T. Küstner, M. Maspero, D. Rückert, C.A.T. van den Berg *MERLIN: In-depth investigation on complex-valued image reconstruction in PyTorch and Tensorflow.* Proc. ISMRM, Toronto, 2023 (power pitch)

Awards

Best presentation award - 1st place, ISMRM MR in Radiation Therapy Study Group Meeting, 2020

ISMRM Magna Cum Laude Merit Award (Top 15% abstract) - *Real-time estimation of 2D deformation vector fields from highly undersampled, dynamic k-space for MRI-guided radiotherapy using deep learning.*, 2020

Best abstract award - 3rd place, ISMRM Detection & Correction of Motion in MRI Study Group Meeting, 2021

ISMRM Summa Cum Laude Merit Award (Top 5% abstract) - *A generic framework for real-time 3D motion estimation from highly undersampled k-space using deep learning.*, 2021

Klaas Nicolay Junior Poster Award - *Feasibility of automatic patient-specific sequence optimization with deep reinforcement learning.*, ISMRM Benelux chapter, 2023

Dankwoord

Werken en feesten vormt schoone geesten.

Johanna Westerdijk, buitengewoon hoogleraar te Universiteit Utrecht.

Gedurende het produceren van het werk dat voor u ligt heb ik het voorrecht gehad om veel mensen te ontmoeten die aan beide componenten voor een *schoone geest* hebben bijgedragen. Het is zonder meer duidelijk dat zonder hen dit boekje niet zou hebben bestaan. Daarvoor is mijn dank zeer groot, en ik wil de rest van dit hoofdstuk gebruiken om deze mensen in het bijzonder te bedanken. Het is zonder meer mogelijk dat ik iemand vergeet hier te noemen. Voor deze personen wil ik de traditie van **Matteo Maspero** en **Tom Bruijnen** voortzetten en zijn jullie bij deze uitgenodigd voor een etentje bij ~~hen~~ mij thuis.

Hierbij wil ik beginnen met al mijn supervisoren gedurende mijn tijd in het UMC. Deze begon met mijn gesprek met **Rob** en **Nico**. Hoewel ik bij binnenkomst in het UMC extreem weinig van MRI en radiotherapie afwist, leek dat niet echt uit te maken. Het was meteen duidelijk dat ik binnen het UMC en het RESOLVE project op een gezellige plek was met tomeloze ambitie en enthousiasme: wij gaan met deep learning (toen nog net nieuw in de MRI wereld!) real-time MRI-gestuurde radiotherapie doen. Ik had toen nog niet kunnen bevroeden waar ik "ja" tegen zei toen jullie besloten mij aan te nemen, maar ik ben blij dat ik het gedaan heb.

Allereerst gaat mijn dank uit naar **Rob**. Met eindeloos geduld stond je altijd klaar voor al mijn vragen en ideeën, wat onmisbaar was tijdens het begin van mijn PhD. Ik heb ook grote bewondering hoe je je klinisch en acadamisch werk schijnbaar moeiteloos kon combineren ondanks het dagelijkse retourtje Neerkant - Utrecht. Door jouw inspanning heb ik mijn eerste stapjes mogen zetten in de academische wereld met het schrijven van mijn eerste conferentie abstract en eerste paper. Daarnaast vond ik het ook erg fijn om met je samen te werken en heb ik genoten van onze gezamenlijk reis naar de MRinRT conferentie.

Nico, jouw visie, drive, kennis, en vriendelijke karakter zijn een zeldzame en onmisbare combinatie, en ik ben blij met het voorrecht om jou als promotor te hebben. Ik heb grote bewondering voor jouw brede kennis (van elektromagnetisme en MRI fysica tot deep learning en Moiré patronen) en hoe je deze gebruikt om uit te weiden in wilde ideeën, om ze vervolgens weer in het perspectief te zetten van de realiteit en de patiënt. Daartegenover sta je ook open voor de wilde ideeën van anderen, hoewel ik meestal snel te horen kreeg dat dit misschien niet de beste richting was. Maar als mijn idee beantwoord werd met een schaterlach wist ik dat het goed zat (*"We gaan de deep learning wereld veroveren met middelbare-school*

wiskunde!"). Ik ben dan ook dankbaar dat ik de ruimte kreeg om hier verder aan te werken, wat geresulteerd heeft in Hoofdstuk 4. Daarnaast stond je altijd voor me klaar, ook voor het persoonlijke in plaats van het werk. Dit heb ik enorm gewaardeerd.

Jan, tijdens mijn promotie ben jij meer op de achtergrond gebleven, maar daarom niet minder belangrijk. Jouw visie en vastberadenheid voor de MR-Linac zijn een enorme inspiratie, zonder welke dit werk niet had kunnen bestaan. Jouw inspirerende verhalen en vertrouwen in de toekomst van MRI-gestuurde radiotherapie werken door in dit proefschrift.

Tijdens het begin van mijn PhD waren **Bjorn** en **Federico** een enorme hulp en waren ze voor alles beschikbaar: wegwijs raken in het ziekenhuis, het vinden van de MRI scanner (en de weg terug), en bij wie je moest zijn om praktische zaken geregeld te krijgen. Maar ook de praktische zaken van onderzoek doen, zoals kritisch zijn op je eigen onderzoek en duidelijk overbrengen wat ik niet of wel gedaan heb. Hiervoor waren jullie altijd beschikbaar met groot geduld, waarvoor mijn dank.

Met het vertrek van **Rob**, **Bjorn** en **Federico** in korte tijd, was **Matteo** beschikbaar om de rol van supervisor over te nemen en ik had niemand beters kunnen wensen. Met schijnbaar oneindig veel tijd en geduld heb je mij met zeer effectief gebruik van de *pars destruens* en de *pars construens* alles geleerd en verfijnd wat een goede wetenschapper moet kunnen: effectief werken, kritisch zijn op jezelf en anderen, en effectief communiceren. Ik heb enorme bewondering voor jouw toewijding, kennis en geduld, zowel als onderzoeker als persoon. Daarnaast heb ik ook genoten van onze dinsdag en donderdag meetings die altijd exact een uur duurden, ongeacht hoeveel we moesten bespreken. Zeker tijdens de corona periode waren dit enorm motiverende meetings om naar uit te kijken. Dan rest mij niets te zeggen behalve: "*Grazie mille per tutto il tuo aiuto in questi anni. Non avrei potuto fare questo lavoro senza il tuo supporto e ti auguro ogni successo nel tuo nuovo ruolo*".

Naast mijn supervisoren wil ik ook graag de beoordelingscommissie van dit proefschrift, prof. dr. Martijn Intven, prof. dr. Bas Raaymakers, prof. dr. Daniel Obserki, dr. Alessandro Sbrizzi, en prof. dr. ir. Marius Staring, bedanken voor jullie tijdsinvestering om dit lange proefschrift te lezen.

Tevens wil ik alle co-auteurs (in volgorde van verschijning) van onze manuscripts bedanken voor hun hulp, tijd en moeite: **Matteo**, **Federico**, **Bjorn**, **Martijn**, **Jan**, **Nico**, **Rob**, **Tom**, **Alessandro** en **Joost**.

Het werk in dit proefschrift staat niet op zichzelf maar was onderdeel van het RESOLVE project, een intiutief acroniem wat staat voor "Enabling M**R** guided radioth**E**rapy of moving abdominal tumor**S** and **O**rgans by development of a dense e**L**ement recei**V**er array and highly undersampled, parall**E**l MRI". Hierbij wil

dan ook graag mijn mede-RESOLVErs bedanken: **Tom**, **Stefan**, **Luca**, **Bjorn**, **Matteo**, **Gabrio**, **Alessandro**, **Federico**. In het bijzonder wil ik **Tom** en **Stefan** bedanken. **Tom**, bedankt voor al jouw hulp voor alles wat met reconstructie te maken heeft. Ik heb enorme bewondering voor jouw doorzettingsvermogen en diepgaande kennis over MRI acquisitie en reconstructie. **Stefan**, zonder jou was ik niet in het UMC terecht gekomen. Ik zal altijd dankbaar zijn dat je me hier hebt voorgedragen. Just like the people of the RESOLVE project, I'd also like to thank the RESOLVE user committee: Gerald Schubert, Dave Fuller, Dennis Klomp, Martino Borgo, Catalina Arteaga, Shirley Baert and Margo Beukers.

Bovendien kijk ik altijd uit naar de Computational Imaging Meeting. De kennis die hier zit is ongekend en daar heb ik veel van geleerd, bedankt: **Alessandro**, **Alexander**, **Bart**, **Bas**, **Beau**, **Carel**, **David**, **Edwin**, **Fei**, **Flavio**, **Gabrio**, **Hannah**, **Jordi**, **Mark**, **Mathijs**, **Matteo**, **Max**, **Miha**, **Mike**, **Nico**, **Niek**, **Oscar**, **Peter**, **Robin**, **Seb**, **Soraya**, **Stefano**, **Thierry**, **Thomas**, en **Tom**.

Most time of a PhD is spent in an office (unless there is a work-from-home requirement) and I've really enjoyed my time in Q.02.2.300! So I want to thank all my roommates over the years: **Stefan**, **Federico**, **Robin**, **Georgios**, **Szabolcs**, **Jordi**, **David**, **Stefano**, **Niek** and **Hannah**. **Stefan en Robin**, ik kijk uit naar het moment dat we weer iemand een cadeau mogen geven, gezien veiligheid en handzaamheid onderaan het vaandel staan. **Niek**, bedankt voor alle vele inzichten in beweging en bijbehorende wiskunde, je kunt inderdaad bijna alles vatten in een B-spline. Ik heb genoten van onze trip naar Oxford! **Hannah**, you can always ask me a short question regarding motion, reconstruction, or deep learning! Ook het kantoor aan de ene overkant (**Maureen**, **Ellis**, **Filipa**) wil ik bedanken voor de koffiemomenten. Daarnaast wil ik ook het andere kantoor aan de overkant bedanken voor de gezelligheid en gedeelde liefde voor de muziek van Natasha Bedingfield: **Max**, **Mike**, **David** en **Thierry**. **Mike**, mijn vaste *partner in crime* die net zo enthousiast is als ik om ergens een biertje te zoeken, ongeacht de plaats of tijdstip. Ik kan niet wachten tot de volgende trein vertrekt. **Thierry**, ik ben er volledig van overtuigd dat de "𝔎-space Collective" gaat doorbreken.

Hiernaast wil ik alle onderzoekers bedanken die deelnamen aan onze gezamelijke meetings om over vergelijkbare onderwerpen te presenteren en discussieren, zoals beweging en deep learning, oftewel de Trackers meeting (**Katrinus**, **Paris**, **Osman**, **Cornel**, **Prescilla**, **David**, **Max**, **Lando**, **Guus**, **Pim** en **Niek**) en de DLinRT meeting (**Mark**, **Matteo**, **Jacquelien**, **Flavio** en **Georgios**). Dank jullie wel voor de discussies, inspiratie en hulp.

Een zeer belangrijke traditie binnen de radiotherapie is het 16:00 tafelvoetbal moment. Ik wil alle deelnemers bedanken die met de broodnodige afleiding kwamen en fysieke activiteit brachten naar het werk: **Anna**, **Bart**, **Bjorn**, **David**, **Edwin**, **Ellis**, **Federico**, **Federique**, **Filipa**, **Flavio**, **Gabrio**, **Georgios**, **Guus**, **Jorine**, **Katrinus**, **Mark**, **Matteo**, **Maureen**, **Max**, **Mike** (ik weet eindelijk

welke kant de bal op moet), **Niek**, **Oscar**, **Osman**, **Peter**, **Robin**, **Soraya** en **Stefan**.

Daarnaast wil ik iedereen van Q2 bedanken die de conferenties nòg zoveel leuker hebben gemaakt: **Alexander**, **Alex**, **Ayhan**, **Bart**, **Bas**, **David**, **Dennis**, **Edwin**, **Evita**, **Fei**, **Flavio**, **Giulia**, **Hannah**, **Hans**, **Jordi**, **Katrinus**, **Kyungmin**, **Marius**, **Martijn**, **Mathijs**, **Max**, **Michael**, **Miha**, **Narjes**, **Oscar**, **Peter**, **Pien**, **Rick**, **Sarah** (een ISMRM waar wij niet dansen is geen ISMRM), **Seb**, **Stanley**, **Thierry**, **Thomas Olausson** en **Thomas Roos**.

I'd also like to thank **Kerstin Hammernik** and **Thomas Küstner**, who always showed a kind interest in my work, even resulting in a collaboration on an ISMRM abstract. I look forward to many more collaborations and conferences together.

I'd like to express my gratitude to the editors and peer-reviewers of the manuscripts presented in this thesis, who selflessly dedicated their free time to improve our work with their helpful suggestions.

Finally, I'd like to thank the bachelor and master students that I had the pleasure working with. It's been a joy supervising you and learning from all you. Thanks **Adine**, **Mario**, **Sjors**, and **Vish**.

Tom, **Vincent**, **Jaron**, **Fedde**, **Bas**, en **Roeland**: mijn familie van Groningen. Nog steeds beschouw ik lid worden van Gyas en met jullie een roeiploeg vormen als één van de beste keuzes die ik gemaakt heb. Ik ben er enorm blij mee dat we een hechte groep vrienden zijn die nog steeds door dik en dun voor elkaar klaar staan, bij elkaar langskomen, en samen op vakantie gaan. Ik had de afgelopen twaalf jaar samen (en bijbehorende brakheid) voor geen goud willen missen. Dank jullie wel! Ook kunnen **Sterre**, **Iris**, **Manon**, **Anneleen** en **Lisanne** hierbij niet ontbreken. Dank jullie wel voor de gezelligheid.

Inmiddels acht jaar geleden zijn we in Amersfoort komen wonen, en we zijn zeer blij dat **Folkert**, **Nina**, **Vincent**, **Iris**, **Floris**, **Annemarleen**, **Sybren**, **Merel**, **Stefan** en **Ilona** ons daarin gevolgd zijn. Ik kijk uit naar veel meer bordspelletjes, barbecues, en biertjes in 033 de komende jaren. **Folkert** en **Nina**, elk jaar weer kijk ik uit naar de 24 december sushi traditie. Dank jullie wel voor de inmiddels al vele jaren vrienschap. Het was dan ook een gemakkelijke keuze om jouw als paranimf te kiezen, **Folkert**.

Ook mijn vrienden van Cover wil ik graag bedanken voor de vele gezelligheid die jullie brachten tijdens en na de studie: **Guus**, **Robin**, **Eric**, **Kim**, **Ruben**, **Liese** en **Everhard**. Dank jullie wel voor de campinggezelligheid, pasjes in de Negende Cirkel, schitterende karaoke, vuige muziek en teleurstellende (en natuurlijk ook leuke) museabezoeken!

160

Ook **Tom** en **Ruben** wil ik bedanken voor de afgelopen haast 25 jaar die we al met z'n drieën doorbrengen en dat we samen hebben mogen hockeyen en de basisschool, middelbare school en Sneekweek hebben overleefd. **Ruben**, ik ben nog altijd blij dat je me hebt overgehaald om met je mee te gaan van 4HAVO naar 5VWO. Ik denk dat dit werk er anders nooit was gekomen. **Tom**, ik vind het heel bijzonder dat ik je nu niet alleen vriend kan noemen, maar zelfs collega en paranimf!

Teun en **Lia**, dank jullie wel voor al jullie gezelligheid en hulp de afgelopen jaren. Of het nou gaat om even de (zeer geliefde) VW camper uit te lenen voor onze tripjes of een keuken in elkaar te zetten, jullie zijn er altijd. Dank daarvoor. Ook **Cor** en **Mariska** wil ik bedanken omdat de bezoekjes naar Apeldoorn met jullie erbij nòg leuker zijn.

Lieve **papa** en **mama**, dank jullie wel voor al die jaren zorg en liefde die jullie gegeven hebben. Al in de vroege jaren hebben jullie al duidelijk gemaakt: "Een goed onderwijs is de basis voor de rest van je leven" en ik denk dat deze boodschap toen al aan de basis van dit boekje heeft gestaan. Mijn dank is onmetelijk groot voor de basis die jullie mij hebben gegeven en dat jullie mij altijd aanmoedigde om ervoor te gaan (of er nog even twee keer over na te denken als het een minder goed idee was, natuurlijk). Ook **Lotte** en **Carlijn** wil ik bedanken. Dat jullie iets ouder zijn dan ik gaf me altijd iets om na te streven. Dat blijf ik nog altijd doen. Dank jullie wel dat jullie mijn zussen zijn en alles (vakanties, trips, en het avondeten) leuker hebben gemaakt.

Als laatste blijft de belangrijkste persoon in mijn leven over. Mijn allerliefste **Lianne**. Ik ben zo blij dat wij inmiddels al twaalf jaar samen zijn en dat heeft het leven zo veel leuker gemaakt. Zonder jou had dit boekje er sowieso niet gelegen. Jouw positiviteit, humor, nuchterheid en onvoorwaardelijke steun en liefde zijn voor mij onmisbaar geweest en ik bewonder je enorm om deze eigenschappen. Ik kan niet wachten wat de rest van ons leven ons gaat brengen want ik weet zeker dat het prachtig gaat zijn. Ik kan vol overtuiging zeggen: ik hou van jou.

Curriculum Vitae

Maarten Terpstra was born in Sneek, the Netherlands, on October 19, 1992. After attending 4 years of HAVO and 2 years of VWO at the CSG Bogerman in Sneek, he enrolled in the Bachelor program of Computing Science at the University of Groningen. In 2014, he started the Master's program in Computing Science at the University of Groningen. Maarten followed the Computer Graphics and Scientific Visualization track, which resulted in his Master thesis titled *"Dense Skeletons for Image Compression and Manipulation"* under the supervision of prof. dr. Alex Telea.

After obtaining his Master of Science degree in 2017, he started as a full-stack software developer at the Royal Netherlands Meteorological Institute (*Dutch: Koninklijk Nederlands Meteorlogisch Instituut*). Here, he helped develop a new browser-based software suite to visualize the results of predictive weather models and meteorological observations.

In 2018, Maarten returned to the academic world to start a PhD within the radiotherapy department of the UMC Utrecht, where he developed deep learning models to enable real-time MRI reconstruction and motion estimation. The results of this work are presented in this thesis.

After this work, Maarten will stay at the UMC Utrecht as a post-doctoral researcher. He will continue the research presented in this thesis and also focus on real-time cardiac MRI reconstruction and motion estimation.