Methods for Reducing 3D Non-Cartesian Reconstruction Time

By

Zachary Miller

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

(Biomedical Engineering)

at the

UNIVERSITY OF WISCONSIN-MADISON

2022

Date of Final Oral Examination: 4/21/2022

This dissertation is approved by the following members of the Final Oral Committee:

Kevin M. Johnson, Assistant Professor, Medical Physics and Radiology
John Paul Yu, Assistant Professor, Radiology
Scott Reeder, Professor, Radiology
Elizabeth Meyerand, Professor, Biomedical Engineering and Medical Physics
Sean Fain, Professor, Biomedical Engineering and Medical Physics

Bubbie,
This ain't much in the way of poetry,
But it's for you

Acknowledgements

Its 1 AM and I am almost done with my thesis. By no stretch of the imagination did I do this alone. First, I want to thank my advisor, Kevin Johnson. You gave me the freedom to explore my own ideas, make my own (many) mistakes, and grow over the last five years. Reflecting back, I had a lot of crazy ideas, many which wouldn't come close to working, but you patiently listened to all of them. I really appreciate your broad (and deep) view of the field and science in general, and its been great working with you.

I want to thank all of my committee members, especially JP Yu for putting up with all of my questions about how to actually make an MD/PhD into a coherent career. I want to thank all the members of Kevin's lab for many great years together: Leo Rivera, Luis Torres, Chenwei Tang, Mu-lan Jen, and Muntasir Shamim. I want to thank Garrett and Dali for teaching me what math actually is and for putting up with my sometimes crazy schedule.

I want to thank all the friends I made during graduate school:

- Elle Najt for our long random walks around Madison talking about math and life, and pre-covid trips to Biaggis.
- Adrian Lopez for chess games, trading neuroanatomy for real analysis, and talks on the terrace over beer
- Yohan Sumathipala for challenging me to think deeper and more rigorously about medicine, and for just being a good friend. When I think of dog walks in the Madison winter darkness, its nine degrees and somehow you still have a smile on your face.

I want to thank my MSTP cohort: Michael Rigby, Anna Heffron, Kat Braun, Jon Tai, and Alex Pieper. It's been a long journey. I want to thank all the students I worked with on the medical school board exams, you helped give me purpose when at times grad school felt a little purposeless. I want to thank my friends from forever ago: Joseph Elsakr, Halei Benefield, and Athena Kifah. You influenced me way back when, I am sure the impact carries forward. I want to thank my grandparents: Bubbie and Zadie for everything. Bubbie, I know you wanted to see me graduate. You fought so hard, and if anyone is a role model, you are. Zadie, you basically raised me. Mom, Dad and Lauren, I know its been a really tough year. I love you so much. Adrienne, I met you almost four years ago to the day. You are the love of my life.

Zach Miller
4/13/22

Table of Contents

Abstract vii

Chapter 1:Thesis Overview and Outline 1

1.1 Overview 1

1.2: Outline 2

Chapter 2: Background 3

2.1: MRI experiment: Getting to the Signal Equation 4

2.2 Basic Methods for reconstructing Cartesian and Non-Cartesian Acquisitions 8

2.3 The Undersampling Problem 9

2.4. Parallel Imaging 13

2.5. CS-like methods 17

2.6 Deep Learning 19

2.7: Deep Learning and Image Reconstruction 21

Chapter 3: Memory Efficient Model Based Deep Learning Reconstructions for High Spatial Resolution 3D Non-Cartesian Acquisitions 23

3.1. Introduction 23

3.2 Theory 24

3.2.1 Model Based DL 24

3.2.2 Block-wise Learning Algorithm 25

3.3 Methods 27

3.3.1 Non-Cartesian Data 27

3.3.2 MBDL Architecture 28

3.3.3 Evaluation 28

3.4 Results 29

3.4.1 Hyperparameter Choices: Number of Unrolls 29

3.4.2 Full Resolution Results 32

3.5. Discussion 37

3.6.Conclusion: 39

3.7.Supplemental Section 40

3.7.1 Zero Padding Correction 40

4.1.Introduction 42

4.2 Theory 43

4.2.1 Self-supervised MBDL 43

4.2.2 Dynamic MBDL Architecture 45
4.2.3 Motion Correction Method 46
4.3 Methods 47
4.3.1 Overview 47
4.3.2 Dynamic MBDL Implementation 48
4.3.3 Motion Compensation Workflow 48
4.3.4 Evaluation 51
4.4 Results 52
4.4.1 Impact of Motion Correction during training and testing on dynamic MBDL Image quality 52
4.4.2 Motion Resolved Reconstruction Comparison 55
4.4.3 Motion Compensated Dynamic MBDL Image quality Comparison 56
4.5 Discussion 59
4.6 Conclusions 61
4.7 Supporting Information 61
Chapter 5: Motion Compensated High Spatiotemporal Resolution MRI 64
5.1 Introduction 64
5.2 Theory 66
5.2.1 Extreme MRI: Multi-scale Low Rank Reconstruction Review 66
5.2.2 MoCo-MSLR Reconstruction 67
5.3.Methods 70
5.3.1 Reconstruction Implementation 70
5.3.2 Healthy Volunteer 1 71
5.3.3 Cystic Fibrosis Patient 71
5.3.4 IPF Patient 72
5.3.5 Third Trimester Pregnant Patient 72
5.4 Results 73
5.4.1 Healthy Volunteer Dataset 74
5.4.2 Cystic Fibrosis Lung Dataset 77
5.4.3 Idiopathic Pulmonary Fibrosis Dataset 79
5.4.4 Third Trimester Pregnant Patient Dataset 80
5.5 Discussion 80
5.6 Conclusion: 83
Chapter 6: Summary and Future Directions 84

6.1: Summary of Contributions 84
6.2: Future Directions 85
6.2.1 Pulmonary Lesion Study 85
6.2.2. Validating Motion Dynamics in MoCo-MSLR and Extreme MR 86
6.2.3. Motion Compensated High spatiotemporal resolution Dynamic contrast Enhanced Reconstructions 89
References 91

List of Figures

Figure 2.1: Coil signal from Net Magnetization precessing in transverse plane……13

Figure 2.2: Received Magnitude "Image" in Fourier Domain without Gradients. Shown……13

Figure 2.3: Example Non-Cartesian Trajectories……...15

Figure 2.4: Fourier Transform Pairs……...17

Figure 2.5: Impact of sampling in the time domain on the frequency domain……...19

Figure 2.6: Coherent Aliasing in Undersampled Cartesian Acquisition…….20

Figure 2.7: Noise-like Undersampling Artifacts in 3D Radial UTE Imaging……20

Figure 2.8: 2X Undersampled Cartesian Acquisition viewed from two different coils…….21

Figure 2.9: Coil Combined Reconstruction without Aliasing Artifact……...22

Figure 3.1: MBDL with block-wise learning model……...34

Figure 3.2: MBDL performance vs. number of unrolls. ……...38

Figure 3.3: Reconstruction Quality versus number of MBDL unrolls. …….39

Figure 3.4: L1 Wavelet vs. MBDL performance for all test data. …….40

Figure 3.5: L1 Wavelet vs. MBDL performance for flip angle24°……..41

Figure 3.6: L1 Wavelet vs. MBDL performance for flip angle 6°……41

Figure 3.7: Coronal slices from high resolution volumetric images using different reconstruction strategies. ……42

Figure 3.8: Performance of MBDL on different contrasts. ……43

Figure 3.9: Reconstructions of paired anatomy with different contrasts. ……..44

Figure 3.10: MBDL Reconstructions with varying numbers of spokes. ……...45

Supplemental Figure 3.1: General Method for zero-padding correction. ……48

Figure 4.1: Dynamic MBDL Architecture……..53

Figure 4.2: Motion Compensated Workflow……..55

Figure 4.3: Impact of Registration on Reconstruction Results……...60

Figure 4.4: aSNR and CNR comparison across MBDL Reconstructions……..61

Figure 4.5: MBDL Reconstructions Liver Edge Sharpness……62

Figure 4.6: Motion Compensated Dynamic MBDL from Motion fields estimated using Motion Resolved reconstructions from Dynamic MBDL (MBDL MR) vs XD. Grasp (XD-Grasp MR)……...63

Figure 4.7: End-inspiratory reconstruction comparisons: CG Sense vs. Spatial Self-Super MBDL vs XD-grasp vs iMoCo vs. Motion compensated Dynamic MBDL……64

Figure 4.8: Maximum Intensity Projection comparisons: CG Sense vs. Spatial Self-Super MBDL vs XD-grasp vs iMoCo vs. Motion compensated Dynamic MBDL. ……65

Figure 4.9: aSNR and CNR comparison across CG Sense vs. Spatial Self-Super MBDL vs XD-grasp vs iMoCo vs. motion compensated dynamic MBDL from motion fields estimated using XD-Grasp motion resolved reconstructions (XD-Grasp MR) vs. motion compensated dynamic MBDL from motion fields estimated using dynamic MBDL motion resolved reconstructions (MBDL MR)……...66

Figure 4.10: Liver edge sharpness comparison across CG Sense vs. Spatial Self-Super MBDL vs XD-grasp vs iMoCo vs. motion compensated models with motion fields estimated using different motion resolved reconstruction strategies……67

Figure 5.1: Respiratory Signal Tracking……81

Figure 5.2: Reconstruction Results on Healthy Volunteer……82

Figure 5.3: Cardiac and Respiratory Dynamics at High temporal resolution……83

Figure 5.4: Short axis Cardiac Phases……84

Figure 5.5: Reconstructions Results on Patient with Cystic Fibrosis……85

Figure 5.6: Reconstruction Results on Patient with Cystic Fibrosis at high temporal resolution……...86

Figure 5.7: Reconstructions Results on IPF Patient……....87

Figure 5.8: Reconstructions Results on Healthy Pregnant Patient in Third Trimester……...88

Figure 6.1: Representative Respiratory Signals using Different Navigation Strategies……96

Abstract

Magnetic resonance imaging (MRI) is a powerful imaging modality. Its flexibility allows for both diagnostic and functional imaging with unparalleled soft tissue contrast. In the brain, MRI is the go-to imaging technique for many structural and functional applications. The same, however, cannot be said for the body where computed tomography (CT) remains the imaging modality of choice. This difference is in part a result of MR's slow acquisition speed making it sensitive to the complex, non-rigid motions seen in the body during minutes long scans. CT, on the other hand, is relatively insensitive to these motions, acquiring high resolution images within seconds.

Non-Cartesian sampling trajectories combined with retrospective motion correction and efficient reconstruction techniques have the potential to change this. Compared to Cartesian scans, non-Cartesian trajectories efficiently sample k-space in all dimensions, have intrinsic motion robustness, and generate noise-like aliases when under-sampled making them optimal for applications that require reconstructions with high spatiotemporal resolution [1]. For these reasons, non-Cartesian acquisitions are being developed for free breathing pulmonary [2] and dynamic contrast enhanced imaging [1] (among others). Despite the promise of non-Cartesian trajectories for rapid body imaging, they have seen limited use clinically.

In the first part of this thesis, I take steps toward making non-Cartesian acquisitions easier to integrate into clinical workflows. The first part of this work addresses the lengthy iterative reconstruction times (on the order of 30 minutes to an hour on state of the art GPUs**)** seen with 3D non-Cartesian acquisitions by developing methods to allow robust deep learning methods to be applied to these high dimensional acquisitions. To do this, I address two primary challenges to applying DL to these datasets: extreme GPU memory demand (>250 GB) and lack of supervision.

In the second part of this dissertation, I work towards improving the quality and dynamics captured by time resolved reconstructions for high spatial resolution non-Cartesian acquisitions. Building on the work of [1], I incorporate motion compensation into large scale time-resolved multi-scale low rank reconstructions in a technique called MoCo-MSLR. Although these reconstructions are computationally and memory intensive, and remain difficult to integrate into clinical workflows, simply demonstrating the ability to capture such high temporal resolution dynamics with high fidelity is a step forward.

Chapter 1:Thesis Overview and Outline

1.1 Overview

Magnetic resonance imaging (MRI) is a powerful imaging modality. Its flexibility allows for both diagnostic and functional imaging with unparalleled soft tissue contrast. In the brain, MRI is the go-to imaging technique for many structural and functional applications. The same, however, cannot be said for the body where computed tomography (CT) remains the imaging modality of choice. This difference is in part a result of MR's slow acquisition speed making it sensitive to the complex, non-rigid motions seen in the body during minutes long scans. CT, on the other hand, is relatively insensitive to these motions, acquiring high resolution images within seconds.

Non-Cartesian sampling trajectories combined with retrospective motion correction and efficient reconstruction techniques have the potential to change this. Compared to Cartesian scans, non-Cartesian trajectories efficiently sample k-space in all dimensions, have intrinsic motion robustness, and generate noise-like aliases when under-sampled making them optimal for applications that require reconstructions with high spatiotemporal resolution [1]. For these reasons, non-Cartesian acquisitions are being developed for free breathing pulmonary [2] and dynamic contrast enhanced imaging [1] (among others). Despite the promise of non-Cartesian trajectories for rapid body imaging, they have seen limited use clinically.

One goal of my work has been to take steps toward making non-Cartesian acquisitions easier to integrate into clinical workflows. The first part of my thesis (**chapter 3, chapter 4**) addresses the lengthy iterative reconstruction times (on the order of 30 minutes to an hour on state of the art GPUs**)** seen with 3D non-Cartesian acquisitions by developing methods that allow model based deep learning (MBDL) [3], [4] to be applied to these acquisitions. Model based deep learning has been shown to significantly reduce reconstruction time and improve image quality relative to compressed sensing (CS) methods for 2D Cartesian reconstructions [3]–[5]. **I hypothesized that the benefits seen for 2D Cartesian reconstruction using MBDL would transfer to 3D non-Cartesian reconstructions.** There are two major barriers, however, to applying MBDL to these acquisitions: **extreme GPU memory demand** and **difficulty obtaining ground truth for supervised training.**

In **chapter 3**, I address the problem of GPU memory demand by developing a memory efficient training technique that allows MBDL to be applied to high resolution volumetric, non-Cartesian scans on a single GPU. In **chapter 4**, I address the difficulty of acquiring fully sampled ground truth data for these acquisitions by developing a self-supervised learning method that reconstructs highly accelerated dynamic acquisitions by combining efficient motion correction with an MBDL architecture that leverages correlations across frames.

The second goal of my work has been to improve the quality and dynamics captured by time resolved reconstructions for high spatial resolution non-Cartesian acquisitions (**Chapter 5)**. Building on the work of [1]I incorporate motion compensation into large scale time-resolved multi-scale low rank reconstructions in a technique called MoCo-MSLR. Although these reconstructions are computationally and memory intensive, and remain difficult to integrate into clinical workflows, simply demonstrating the ability to capture such high temporal resolution dynamics with high fidelity is a step forward.

1.2: Outline

Chapter 2: MRI Reconstruction Review

In this chapter I first describe how the reconstruction problem arises from the MR experiment itself. I then review basic methods for reconstructing fully sampled Cartesian and non-Cartesian acquisitions. Following this, I describe well known methods that allow for faster than Nyquist sampling like parallel imaging and compressed sensing type reconstructions. I then introduce deep learning and its applications in image reconstruction.

Chapter 3: Memory Efficient MBDL Reconstructions for High Spatial Resolution 3D Non-Cartesian Acquisitions

Here I tackle the high GPU memory demand (>250 GB) associated with MBDL reconstructions of 3D non-Cartesian data. The algorithm I propose combines gradient checkpointing in place of traditional backpropagation with a block-wise training method that decomposes the input volume into smaller patches, iteratively passes these patches through the neural network regularizer, and then recomposes these patches back into the full volume to then be passed to the data-consistency step. By passing $P_x x P_y x P_z$ patches through the network in place of the full $N_x x N_y x N_z$ volume, GPU memory requirements are reduced $N_x x N_y x N_z / P_x x P_y x P_z$ fold. I apply this algorithm to reconstruction of high resolution (~1mm isotropic) 3D pulmonary magnetic resonance angiography (MRA) datasets on a single 40 GB GPU.

Chapter 4: Self Supervised Deep Learning for Highly Spatial Resolution 3D Non-Cartesian Acquisitions

In this chapter, I address the difficulty obtaining fully sampled 3D non-Cartesian ground truth data for supervised training of MBDL architectures. I extend the self-supervised learning framework proposed by [6] to take advantage of correlations across frames (called dynamic MBDL). I combine this technique with an efficient GPU based registration method to develop motion compensated deep learning methos. I

apply this method to reconstruction of the end-inspiratory phase of high resolution (~ 1mm isotropic) 3D pulmonary magnetic resonance angiography (MRA) data-sets.

Chapter 5: Motion Compensated High Spatiotemporal Resolution MRI

In this chapter, I switch focus from deep learning to developing methods to incorporate motion modeling into high spatiotemporal resolution volumetric iterative reconstructions. I build off the Extreme MRI technique proposed in [1]. I demonstrate that this motion compensated technique results in significantly improved image quality over Extreme MRI at ~500 ms temporal resolution. Further, I show that this technique is able to capture realistic cardiac dynamics at ~100 ms temporal resolution. This work was done in collaboration with Luis Torres.

Chapter 6: Summary and Future Directions

In this chapter I summarize the work completed in this thesis and discuss future work that I hope to complete during my fourth year of medical school.

Chapter 2: Background

The goal of this chapter is to help readers 1) understand reconstruction as a fundamental part of the workflow of MR imaging, 2) recognize the undersampling problem and why undersampled

acquisitions are a necessary part of high spatial resolution scans, and 3) learn about different reconstruction techniques including parallel imaging, compressed sensing-like, and deep learning reconstruction.

2.1: MRI experiment: Getting to the Signal Equation

The goal of reconstruction in MRI is to produce an interpretable image x that is consistent with the data collected in the acquisition space. In this section, I describe how the MR experiment leads to a Fourier transform relationship between image-space x and acquisition space. This Fourier relationship forms the basis for all reconstruction techniques used throughout this thesis.

The fundamental goal of the MRI experiment is to convert a large number of protons in an object of interest into a usable signal that reveals something about the structure or function of that object. The behavior of these protons has been extensively studied by physicists revealing the following properties:

1. Protons have a magnetic dipole moment
2. This magnetic dipole moment makes these protons sensitive to applied magnetic fields
3. If these protons are placed in a large magnetic field B_0 , they precess about the axis of that field. Individual protons precess about a magnetic field at a frequency determined by the Larmor equation: $\omega_{pre} = \gamma B_0$
4. The sum of the dipole moments in aggregate leads to a net magnetization in the direction of the external field.
5. Application of a secondary oscillating field (RF pulse), however, can tip the net magnetization, off axis leading to precession. This leads to emission of radiation that oscillates with the same frequency as this precession. This EM wave representing the sum of many much smaller waves emitted by precessing protons can be picked up by coil elements and measured.

This series of observations does not get us to an image, but hints at how an image might be obtained.

Suppose an RF pulse is applied that tips the net magnetization to an axis orthogonal to B_0. The net magnetization now precesses about B_0. The (simplified) signal picked up by the coil elements (**figure 2.1)** is shown below (ignoring decay for simplicity):

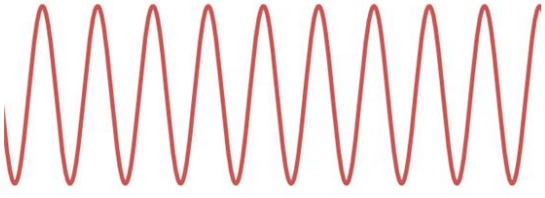


Figure 2.1*: Coil signal from net magnetization precessing in transverse plane.*

All protons in the object are precessing at the same frequency as they all see the same B-field meaning the wave picked up by the coil elements oscillates at single frequency weighted by the proton density. If we look at this from another perspective and take the Fourier transform of this wave, a delta function at the precession frequency is obtained:

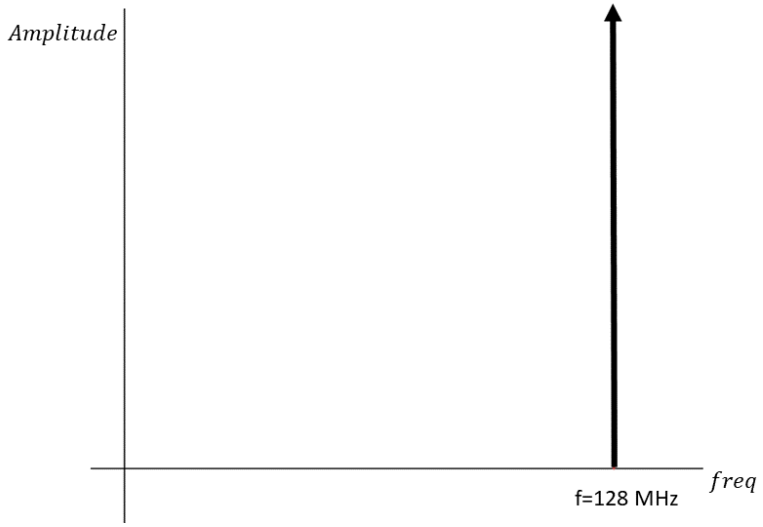


Figure 2.2: *Magnitude "Image" in Fourier Domain without Gradients. Shown here is the Fourier transform of the received time varying signa generated from precession of the net magnetization. This precessional frequency is ~128 Mhz for a 3T Magnet. I only show the positive half plane ($freq > 0$) of the Fourier domain here.*

If we view **figure 2.2** as an "image," all protons have been assigned the same temporal frequency, that is, there is no spatial localization. If the goal is to create an image that gives us useful information, then we must somehow spatially localize protons.

Suppose we replace the constant B-field used above with a B-field that varies linearly over space: $\vec{B} =< G_x x, G_y y, G_z z >$. Where does that get us? From the Larmor equation, it follows that $\omega(r) = \gamma \vec{G} \cdot \vec{r}$ where $\vec{r} =< x, y, z >$ and $\vec{G} =< G_x, G_y, G_z >$ (assuming we filter out the precession frequency

associated with B_0). When we receive the EM wave $X(t)$ emitted by this volume of protons X , we now obtain the sum of waves oscillating at many different frequencies:

$$X(t) = \int_{-\infty}^{+\infty} X(\omega) e^{-i\omega(r)t}\, d\omega \quad \textbf{(2.1)}$$

This sum (over the continuum) of waves of different frequencies is simply the Fourier transform! If we take the inverse Fourier transform of this time varying signa , we obtain $X\big(\omega(r)\big)$, a signal that varies as a function of precession frequency. We have thus intentionally designed the temporal frequency ω to linearly vary with space meaning that $X(\omega)$ is our image of interest.

The expression above relates temporal (precession) frequency to time. For image acquisitions, space vs spatial frequency relationships are more intuitive. The phase term in **2.1**: $\phi = \omega(r)t$ can be expanded out as $\phi = 2\pi\gamma\big(\vec{G}\cdot\vec{r}\big)t$ because $\omega = 2\pi\gamma\big(\vec{G}\cdot\vec{r}\big)$. Note for simplicity I assume here that $\vec{G}$ is constant in time. It turns out that $\overrightarrow{k_r} = \gamma t\big(\vec{G}\big)$ has units of spatial frequency. Thus we can rewrite this term as:

$$X\big(\overrightarrow{k_r}\big) = \int_{-\infty}^{+\infty} X(\vec{r}) e^{-i2\pi\gamma\left(\overrightarrow{k_r}\cdot\vec{r}\right)}\, \mathrm{d}\vec{r} \quad \textbf{(2.2)}$$

Our received signal $X(t)$ has thus been re-parameterized as $X(k_r)$. This suggests an entirely different way of looking at the MR experiment. By applying gradients, we are varying phase across the object literally building a wave in space to probe our object of interest. Near the start of the experiment (assuming no other gradients have been applied), there is little phase variation across the object. A wave with slowly varying phase is a wave with low spatial frequency. A wave with low spatial frequency can only resolve objects far apart. As time flows, the phase separation between protons closer together grows larger allowing us to resolve ever closer protons.

This analysis implies that *X* **our object of interest, is not obtained all at once.** Remember, *X* is parameterized by 1D curve $X(k_r)$ (we are reading out a one-dimensional signal) meaning we obtain X as we traverse curves in spatial frequency-space or k-space. We design these curves $\overrightarrow{k_r} = \gamma < G_x, G_y, G_z > t$ meaning we decide how to traverse k-space for a given acquisition and obtain information about our object X . There is a great deal of freedom then in choosing k-space trajectories. Acquiring along an equispaced grid, otherwise known as a cartesian acquisition is commonly used clinically. All acquisitions that are not equispaced are known as non-Cartesian acquisitions. **Figure 2.3** below demonstrates radial and spiral acquisitions, two commonly used non-cartesian acquisitions.

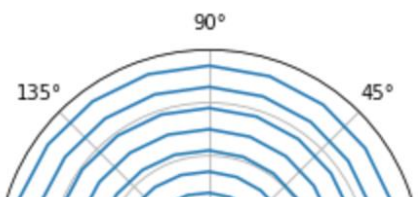


Figure 2.3: *Example Non-Cartesian Trajectories. The 3D radial trajectory on the left resembles the sampling pattern for ultra-short-echo time lung imaging used during acquisition throughout this thesis. The trajectory on the right is a spiral trajectory in 2D.*

In this thesis, I primarily focus on radial acquisitions as these trajectories are used for ultrashort echo time acquisitions for pulmonary imaging.

We have ignored a major component of the MR experiment though. The signal we ultimately analyze is NOT analog, it is sampled in time at some sampling frequency f_s to store and process the signal digitally. Taking this sampling into account, the final signal equation we obtain then that relates sampled k-space data to our image (in just 1D here) is:

$$X_k = \sum_{n=0}^{N-1} X_n \, e^{-\frac{i2\pi\gamma kn}{N}} \quad \textbf{(2.3)}$$

This relationship as designed by the MR experiment is a discrete Fourier transform (DFT). It might appear that we are done: given k-space data (acquisition), compute the inverse DFT (from k-space to image-space), create an image. In **most** cases though, the DFT cannot practically be used for image reconstruction for two reasons. First, the DFT is computationally inefficient for large matrices (applying the DFT to acquisitions with >10000 observations is prohibitively slow even on modern computers). Second, if we wish to accelerate the acquisition by undersampling data, use of the DFT (or any other simple transformation) without more sophisticated approaches leads to under-sampling artifacts.

The limitations of the DFT in fact point to two of the fundamental problems in MR image reconstruction:

1. How do we make reconstructions computationally efficient?

2. How do we do more with less data, that is, can we accelerate the acquisition without sacrificing image quality?

Much of the work in this thesis is focused on providing potential (and very much partial!) answers to these questions.

2.2 Basic Methods for reconstructing Cartesian and Non-Cartesian Acquisitions

In this section, I introduce methods for reconstructing fully sampled cartesian and non-Cartesian acquisitions. Fully sampled datasets are acquisitions that are above the Nyquist limit **(see 2.3 for a more in-depth discussion)** for the range of frequencies that have been excited by the RF pulse. In other words, these data-sets are an accurate representation of the imaging volume . For Cartesian acquisitions, fast Fourier transforms (FFT) are used in place of the DFT as the FFT is computationally efficient and robust implementations are widely available.

The FFT, however, requires equi-spaced input data making reconstruction of non-Cartesian data tricker. It is possible to directly compute the DFT in (**2.3**), but as stated previously this is computationally inefficient for matrix sizes greater than ~10000 entries.

A great deal of effort has gone into developing robust methods to allow for efficient reconstruction of non-Cartesian data. The essential goal of these methods is to figure out ways to interpolate the non-uniform k-space data to an over-sampled equispaced grid so that the inverse FFT can be applied. Methods that use this combination of operations are known as non-uniform fast Fourier transform methods (NUFFTs). NUFFTs are used in place of FFTs throughout this work as I handle primarily non-Cartesian data. For greater detail on the implementation of NUFFTs, I highly recommend [7]. **Practically speaking** for my work, the reader should recognize the terms **forward** and **adjoint NUFFTs**. The **forward NUFFT** has non-uniform k-space as input and uniform image space as output. The **adjoint NUFFT** is simply the reverse operations of the **forward NUFFT**: input is uniform image space and output is non-uniform k-space data.

We have made an important assumption in this discussion, namely, that the data we have acquired is fully sampled. Acquiring fully sampled data is simple if there are no temporally dependent dynamics: simply scan the patient long enough to reach the desired spatial resolution. If there are temporally dependent dynamics, however, there are hard trade-offs between spatial and temporal resolution. To see this, suppose we can collect at most N^2 samples per second. As we lower the spatial resolution, the k-space grid we need to fill to be fully sampled gets smaller and smaller resulting in increasing temporal resolution assuming we acquire continuously. If our grid requires N^2 points to be fully sampled, our temporal resolution is one second, but if our grid only requires $N^2/2$ points to be fully

sampled than our temporal resolution is 500 milliseconds. Now as spatial resolution is increased, temporal dynamics slower than the set temporal resolution are not only lost, they show up as image artifact as multiple dynamic snapshots are being placed on the same k-space grid . In MR, we are interested in capturing dynamics (motion/contrast change) not only for relevant clinical applications, but also because explicitly incorporating dynamics into reconstruction can significantly improve image quality. **Chapter 5** in this thesis is all about incorporating motion dynamics into reconstruction to improve image quality.

Undersampling the acquisition is a way to relax the trade-off between spatial and temporal resolution seen with fully sampled acquisitions. Although undersampling removes constraints on acquisition parameters, it makes reconstruction more involved as we must deal with aliasing artifacts that arise from this undersampling. There is no free lunch!

In the next section, I discuss the undersampling problem in more detail and methods commonly used to reconstruct undersampled acquisitions including compressed sensing (CS) and parallel imaging. Parallel imaging is used throughout this thesis, and is a standard part of modern MR acquisitions. Although I do not explicitly use classical compressed sensing (sparsifying transforms) in this thesis, much of my work is founded on CS-related ideas.

2.3 The Undersampling Problem

Aliasing and the notion of the Nyquist limit come from the need to discretize analog signals for digital storage. Nyquist's theorem states that for a bandlimited signal in frequency space with maximum frequency f_{band} that if we **uniformly** sample this signal at $2f_{band}$ then it is possible to exactly recover the original analog signal through sinc interpolation. Note that nyquist's theorem ONLY applies to uniform sampling. Let's think a little about where Nyquist's theorem comes from though as it helps build intuition for the non-Cartesian case where non-uniform sampling is the rule.

Suppose we have an analog signal $f(t)$, say a $sinc^2(t)$ function with the triangle function as its Fourier transform (**figure 2.4**):

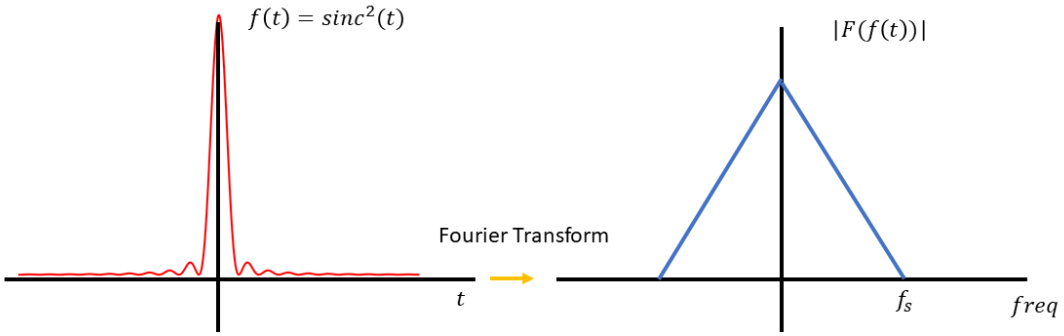


Figure 2.4*: Fourier Transform Pairs. Notice that the triangle function in the frequency domain is bandlimited with maximum frequency* f_s

We can represent **uniformly spaced samples** of this signal as:

$$h(t) = \sum_{l=-\infty}^{+\infty} \delta\left(t - \frac{l}{f_s}\right) \cdot f(t) \quad \textbf{(2.4)}$$

Note that $g(t) = \sum_{l=-\infty}^{+\infty} \delta\left(t - \frac{l}{f_s}\right)$ is a series of spikes at $\frac{1}{f_s}$ intervals known as a comb function. Let us now look at the comb function in the Fourier domain. **Multiplication over time** in image space is equivalent in the Fourier domain to the **convolution** of the Fourier transform of the comb function with the Fourier transform of $f(t)$:

$$F[f(t)g(t)] = F(f) * G(f) \quad \textbf{(2.5)}$$

where the Fourier transform of the comb function is just another comb function. Now as the comb function is made up of an infinite number of translated delta functions, and all these operations are linear, we can get a sense of what the convolution of a comb function with our function of interest looks like by just considering the convolution of a single delta function with our function. The relationship is straightforward:

$$F(f)*\delta(f - f)=F(f - f_0) \quad \textbf{(2.6)}$$

In other words, convolution with a translated delta function simply translates the original function by the same amount. This means that sampling in time leads to the following function in frequency space:

$$X(f) = F[f(t)g(t)] = \sum_{l=-\infty}^{+\infty} F(f - lf_s) \textbf{ (2.7)}$$

Thus, we get an infinite sum of periodic replicates (**figure 2.5)** of $F(f)$ with period $f_s = \frac{1}{T}$.

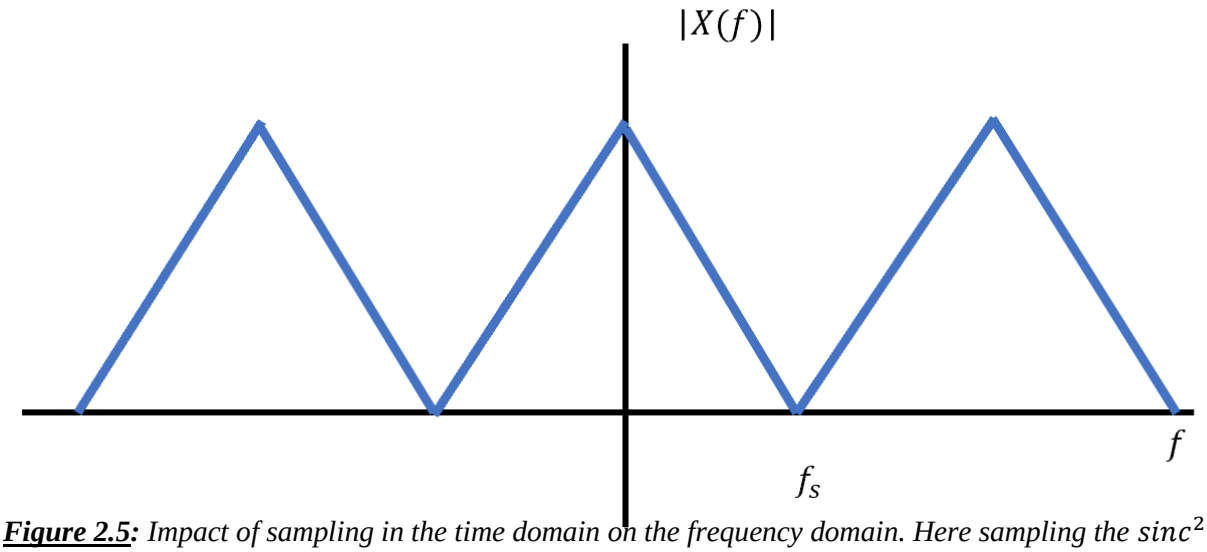


Figure 2.5: *Impact of sampling in the time domain on the frequency domain. Here sampling the* $sinc^2(t)$ *function with frequency* $2f_s$ *leads to periodic replicates spaced* $2f_s$ *apart (the Nyquist frequency) so there is no overlap in the periodic replicates. Sampling at a rate less than the maximum frequency in the triangle function leads to overlap of periodic replicates and aliasing.*

We can see in **figure 2.5** that the bandwidth ($2f_s$) of the original triangle function determines the minimal sampling rate needed to prevent interference between replicates. If we sample below this rate, periodic replicates overlap leading to coherent aliases.

To relate all this back to the MR acquisition, we are sampling in time (k-space) thus aliasing artifacts show up in (temporal frequency) image space. For Cartesian/equispaced imaging, signal energy contained in a given frequency above the Nyquist limit is always sent to exactly one frequency below the Nyquist rate leading to coherent artifacts a shown below:

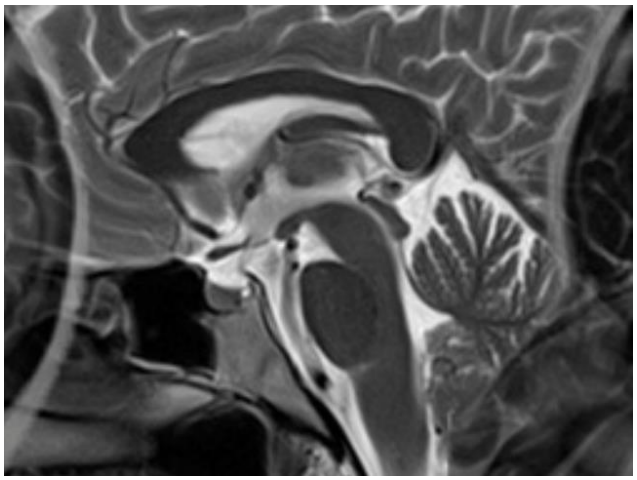


Figure 2.6*: Coherent Aliasing in Undersampled Cartesian Acquisition. In this image, wrap-around can be seen in the phase encoding direction. Image courtesy of Allen D. Elster, MRIquestions.com*

For non-Cartesian acquisition where sampling is non-equispaced, simple application of Nyquist's theorem is not possible. This is because there may be local parts of the signal that are sampled above the Nyquist limit and other local parts of the signal sampled below the Nyquist limit. This leads to samples that cannot be fit by a single frequency.

As samples require multiple frequencies to be accurately fit, signal energy that cannot be captured at the given sampled rate is smeared across many temporal frequencies leading to noise-like aliases **(figure 2.7):**

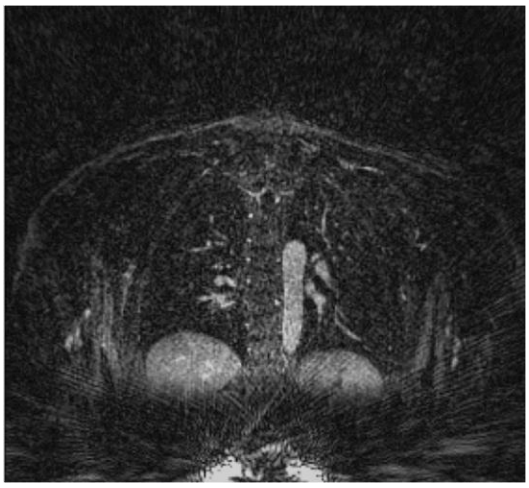


Figure 2.7*: Noise-like undersampling Artifacts in 3D Radial UTE Imaging. Here, k-space with 50k spokes was retrospectively undersampled to 5k spokes, and then a gridded reconstruction was performed. Notice the lack of coherent aliases.*

As mentioned in **section 2.3**, undersampling allows us to relax this spatial vs temporal resolution tradeoff, but its easy to see that the images (**figure 2.6 and 2.7**) are far from diagnostic quality. A great deal of work over the last two decades has gone into developing reconstruction techniques that produce diagnostic quality images from undersampled acquisitions. In the following sub-sections, I discuss parallel imaging and compressed sensing like techniques, two foundational methods that allow high quality reconstructions from undersampled acquisitions.

2.4. Parallel Imaging

The intuition behind parallel imaging is straightforward. Coils can be thought of as providing different perspectives of the object to be imaged. Suppose two 2X accelerated acquisitions of the same object were acquired with each scan using a different coil. Here are these scans (**figure 2.8**):

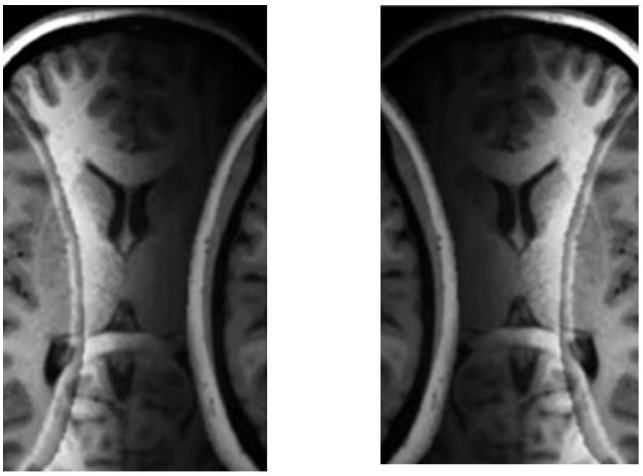


Figure 2.8: *2X Undersampled Cartesian Acquisitions collecting using two different coils. Image courtesy of Allen D. Elster, MRIquestions.com*

Note that parts of the object closer to the coil have higher intensity than objects further away. The image space output from the coil can be written as:

$$O = CI \quad \textbf{(2.8)}$$

This expression is saying that the coil weights the aliased image I with unique pixel-wise weights C, also known as SENSE maps.

Our goal is to fuse the aliased images in **figure 2.8** to give a single unaliased image (**Figure 2.9**).

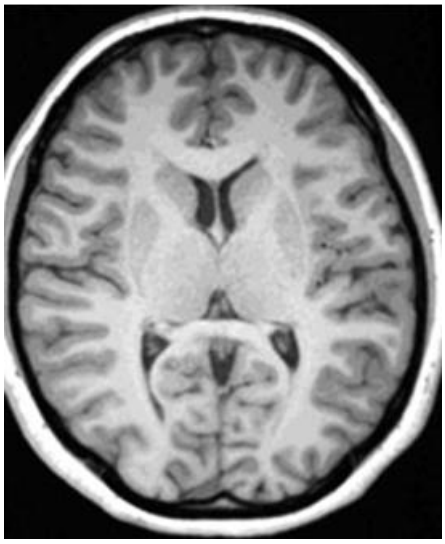


Figure 2.9*: Coil Combined Reconstruction without Aliasing Artifact*

Lets compare these figures. In the fully sampled acquisition, there is no aliasing. This means there is a one-to-one correspondence between the position of an object feature and its pixel position in image-space. Under-sampling disrupts this correspondence. For cartesian acquisitions, 2-fold under-sampling means the intensity at each pixel position in **figure 2.8** is a linear combination of 2 intensities that would be found in **differen**t pixel positions in the fully sampled image:

$$O_i = c_{1j} I_j + c_{1k} I_k \textbf{ (2.9)}$$

where c_{1j} and c_{1k}represent weights from the first coil at different pixel positions in the ideal image

But this is just an underdetermined linear system: one equation, two unknowns which in fact is what leads to the aliasing. Notice though, we can add the information from the second coil (**figure 2.8**) to our system of equations which gives us:

$$O_i = c_{1j} I_j + c_{1k} I_k \textbf{ (2.10)}$$

$$A_i = c_{2j} I_j + c_{2k} I_k \textbf{ (2.11)}$$

This is now a fully determined linear system meaning we can solve the true pixel intensities at positions j and k through (in this case) matrix inversion.

What we described above is the intuition behind parallel imaging, but we acquired the undersampled acquisition in series rather than in *parallel* so there was no speed up in scan time. Parallel imaging simply uses the coils simultaneously during a single scan allowing significant speed up in scan time.

Now let us put this in matrix form for the L-fold undersampled case where $A \in \mathbb{R}^M$ represents the aliased pixels in one location in the image weighted by each of the coils, $C \in \mathbb{R}^{MxL}$ has size corresponding to M coil weights by L aliased pixels, and $I \in \mathbb{R}^L$ is the underlying unaliased pixel intensities. I note that if less than M coils are used, the system is underdetermined. The full expression is:

$$A = CI \text{ (2.12)}$$

This expression still only applies parallel imaging to one pixel. To account for all pixels in the image, let $P \in \mathbb{R}^{MxK/L}$, $C \in \mathbb{R}^{Mk/LxLN}$, and $v \in \mathbb{R}^{LN}$ where:

$$P = Cv \text{ (2.13)}$$

The reconstruction problem can be directly incorporated into this formulation by replacing the matrix of coil-wise images P with a matrix Y of coil-wise k-space data:

$$Y = FFT[Cv] \text{ (2.14)}$$

Notice though that both the FFT and coil-wise multiplication are linear operations, and can thus be represented as a single linear operator E:

$$Y = Ex \text{ (2.15)}$$

In this thesis, E represents whatever transforms are needed to move from image-space back to k-space. For non-Cartesian imaging, E is often the combination of the adjoint NUFFT,sense maps and density compensation although more operations can be added as needed, for instance motion fields.

The question now is how to solve for x. E is not necessarily a square matrix, and thus cannot be inverted. To get around this, let E^H be defined as the adjoint of E such that $E^H E$ is a square matrix. To solve for x, first multiply both sides of equation X by the adjoint operator:

$$E^H Y = E^H Ex \text{ (2.16)}$$

$E^H E$ is now a square matrix which can be inverted:

$$(E^H E)^{-1} E^H Y = x \text{ (2.17)}$$

It turns out that this relation is exactly the normal equation derived from a least squares minimization.

The problem is that computing the pseudoinverse: $(E^H E)^{-1} E^H$ is computationally expensive for large matrices. MRI physicists, however, are far from the first researchers to run into the computational issues associated with the analytic solution to least squares minimization. There is a VAST literature that replaces analytic solutions of least squares with various methods that attempt to **iteratively** approach solutions to least squares problems. Iterative methods are used throughout this thesis and in much of modern MR reconstruction research. To get a sense of why iterative methods considerably reduce the computational burden of least squares minimization, lets perform gradient descent on our parallel imaging problem.

First, let's modify our reconstruction model slightly. Above we expressed it as:

$$Y = Ex \textbf{ (2.18)}$$

For least squares minimization, we express our reconstruction problem as:

$$\underset{x}{\operatorname{argmin}} \|Ex - y\|^2 \textbf{ (2.19)}$$

Here we seek to find x that minimizes the function $f = \|Ex - y\|^2$

Gradient descent will allow us to iteratively approach the minimizer of this function. For simplicity, we assume f f is convex (meaning a global minimizer exists). Gradient descent involves:

1. Starting with an initial guess x_0
2. Taking the gradient of f
3. Updating our guess as follows: $x_1 = x_0 - \alpha \nabla f$
4. Iterating until convergence

The gradient of our function is:

$$\nabla f = 2E^H(Ex - y) \textbf{ (2.20)}$$

Notice that no computationally expensive inverses are needed. Simply repeated application of our forward and adjoint linear operators. In this thesis, $f = \|Ex - y\|^2$ is often combined with regularizer terms that constrain the optimization toward certain preferred solution. In this context, $f = \|Ex - y\|^2$ is referred to as the data-consistency step as iterative methods that use regularizers often alternate between minimizing the regularizer step and minimizing this least square loss term enforcing similarity between the acquired k-space data y and our guess x transformed into k-space.

There are a host of other methods ((with faster convergence) available to minimize least squares problems. For example, the landmark paper by Preussman [8] uses conjugate gradient descent to minimize this least squares problem . I use gradient descent steps in **Chapter 3** and conjugate gradient iterations in **Chapter 4** to enforce data-consistency in the context of using neural network based regularizers. **In Chapter 5**, I use *ADAM*, a first order stochastic descent method commonly used for neural network training. In this case, however, I use it for iterative optimization of motion fields.

The undersampling problem is largely solved if we wish to say within the bounds of acceleration allowed by parallel imaging. We are still left with undersampling artifact, however, when we really try to push acquisition acceleration. **Chapter 5** is all about trying to reconstruct extremely undersampled dynamic volumetric imaging time series (100ms-500ms temporal resolution). To reconstruct such highly accelerated acquisitions, other methods are needed during reconstruction. In the next sub-section, I discuss compressed sensing-like regularization methods that allow us to bring prior assumptions about the behavior of the data to allow us to undersample even further than parallel imaging alone allows.

2.5. CS-like methods

In the context of parallel imaging, I described how coil sensitivities can be used to shift from an undetermined system of equation to a fully determined system of equations. Compressed-sensing methods take a different approach. In place of trying to achieve a fully determined system which in many cases is not possible (**see Chapter 5**), these methods try to develop priors that when incorporated into reconstruction can constrain the undetermined system to unique (or close to unique) solutions.

Classical compressed sensing focuses on recovering a signal x such that in some transform domain: Ψx is sparse (Wavelet-based transforms are often used) when we are handed **randomly acquired** (in MRI, psuedorandom) samples Ex where E is an operator, for instance the Fourier transform.

A nice way to motivate why trying to find sparse representations makes sense is to consider JPEG image compression. Roughly, JPEG compression takes an image in pixel-space and transforms it onto cosine basis functions using the discrete cosine transform (DCT). Many of the coefficients of the cosine bases are near zero so are simply thrown out (in other words the signal vector is sparsified). Only a subset of these coefficients are then used to reconstruct the image. The reconstructed image is often a close approximation to the original image even though information has been thrown away. Compressed sensing starts with the question: why bother acquiring all this data if we are going to throw it anyway. Lets sample the data we need.

One way of motivating why random sampling is required is to think back to our discussion on aliasing in Cartesian vs non-Cartesian acquisitions. Aliasing in cartesian imaging where samples are uniformly

acquired is coherent. If we took the DCT of an aliased Cartesian image, and threw away small coefficients, the aliasing artifact wouldn't be removed because it is primarily contained in the lower frequency terms (where all the image content is found!). So with uniform sampling, it is difficult to separate out image content from coherent aliases. Aliasing in Non-cartesian acquisitions though are noise-like artifacts. Noise-like undersampling artifacts tends to distribute across many frequencies thus throwing out DCT coefficients DOES remove undersampling artifact. This also means non-Cartesian trajectories closely adhere to the requirements for compressed sensing-like acquisitions making these trajectories very useful for highly undersampled acquisitions reconstructed using CS-like methods.

Although compressed sensing methods that use sparsifying transforming are not used in this thesis, a related problem known as matrix completion IS very relevant. Whereas compressed sensing seeks to reconstruct an unknown vector x, the matrix completion problem tries to determine missing entries from a matrix.

Where does the matrix completion problem come up in MR imaging? Consider an acquisition during free breathing (**Chapter 4**/**Chapter 5**). If we collect T undersampled k-space data/images we can place them as columns in a matrix. As k-space at the given temporal resolution is undersampled, values are missing in each column. This is exactly the matrix completion problem, that is, given a matrix with missing values, apply some prior to fill in these missing values. It is important to note that filling in missing values in k-space is equivalent to removing undersampling artifact in image-space.

A very powerful prior on matrices is low rankness. Intuitively, the rank of a matrix can be thought of as a measure of how correlated columns are to one another. In MR, these columns are often time frames that are very similar to one another up to deformations between them. By minimizing rank, we allow for information sharing across columns to effectively fill in missing data points. Directly minimizing rank though is computationally challenging. Significant work [9], [10], however, has been done to find accurate proxies for rank minimization. A commonly used proxy is minimization of the nuclear norm:

$$\|X\|_* = \sum \sigma_i \quad \textbf{(2.21)}$$

Where σ_i are singular value from the singular value decomposition of X

One thing we have left out of this discussion is how we can incorporate these priors into MR reconstruction. Analytic solutions incorporating priors into reconstructions are rarely available and if available are often computationally intractable. This leads us right back to the iterative methods discussed in **section 2.4.**

We now move on to the final piece of background needed for my thesis: neural networks.

2.6 Deep Learning

Depending on who you ask, deep learning is either the technology that will put radiologists out of business (it won't any time soon) or is overhyped. From my perspective, deep learning is just another useful tool to add to our toolbox in medical physics and radiology. It has benefits and limitations but should be viewed as something worth learning. In my thesis, I use deep learning in the context of image reconstruction. Here I provide a brief overview of what neural networks are (in the context of their use in regression) and how they can be used for reconstruction.

Suppose I have collected a dataset: $\{(x_i, y_i)\}$ where x_i and y_i are vectors in $\mathbb{R}^N$. If I assume there is a linear relationship between inputs x_i and outputs y_i then I am trying to trying to fit a model:$y = Ax + b$ where $A \in \mathbb{R}^{NxN}$ and $b \in \mathbb{R}^N$. I can estimate these parameters A and b by minimizing a least squares function:

$$f = \|y_i - (Ax + b)\|^2 \textbf{ (2.22)}$$

This model can be viewed another way. Entries in A weight the input x and thus are called weights. B translates the vector output: Ax and thus its entries are known as biases. Let a neuron be defined as a function that takes in some vector input, multiplies the vector input with weights and then translates this vector output with biases. Both weights and biases are learned. To learn, we randomly initialize the weights, and stochastically train the model using samples from take our training data $\{(x_i, y_i)\}$. After training, the model can take an unseen input x_i and predict an output y_i. This model is a single layer neuron which is of course just linear regression.

Now, if these neurons are placed in series, then we get a composition of linear functions which is of course still linear (and still equivalent to a single layer neuron). Suppose though element-wise non-linearities are inserted between each neuron in series. An example of a non-linearity (also known as an activation function) is the rectified linear unit (ReLU) where for values with $x > 0$, inputs are passed through a ReLU unchanged. For values less than or equal to 0, the ReLu zeroes inputs out. It has been shown [11] that such a neural network with such non-linearities is a universal function approximator. In other words, neural network with non-linearities can learn to model potentially complicated relationships in data without significant prior assumptions on the model.

I will now briefly discuss how these neural networks are trained with the goal of updating weights and biases each training iteration. As networks involve many matrix operations, graphical processing units (GPUs) which are optimized for these operations significantly accelerate training.

Neural network training is complicated by the fact that weights associated with a given intermediate layer i are dependent on the weights from all the previous layers as these weights determine the input to layer i **and** weights from layers after i as these weights play a role in determining the ultimate output sent to the loss function. This means intermediate layer weights are coupled to all other layers.

The algorithm that simplifies some of the math behind this training is known as backpropagation. I will very briefly sketch out the algorithm here. If the reader wishes to really dig in to the mathematics, I highly recommend [12].

Like any iterative optimization, we must compute an output, compare this output to the actual data through a loss function and then use some descent type algorithm to update weights for each iteration. The process of computing the neural network output is the forward pass. The process of passing gradient information/updating weights back through the network is called backpropagation. Where the forward pass involves passing an input through from first to last layer, the backward pass involves updating weights from last layer to first sequentially. Pytorch and Tensorflow, the most used platforms to train neural networks speed up these computations by saving intermediates produced during the forward pass in memory. This is because the partial derivative of the loss function with respect to a weight in layer i requires the output from the previous layer. If these outputs are not saved in memory, multiple forward passes are necessary to produce these intermediate outputs to allow for gradient computation slowing down backpropagation

For problems where saving intermediate outputs can fit in GPU memory , it is optimal to use backpropagation as implemented in Pytorch and Tensorflow. For problems where intermediates do not fit in memory, it is more efficient to recompute these intermediate losses during backprop. This technique is known as gradient checkpointing [13] . **Chapter 3** is all about using gradient checkpointing to allow neural network based reconstruction methods to be applied to volumetric non-Cartesian data.

As there are often many training examples, most neural network training leverages some form of stochastic gradient descent (SGD) to reduce memory and computational costs associated with updates over the entire dataset at once. In SGD, a single sample (batches of samples are often used as well) is randomly drawn from the training set, neural network weights are updated, and then another sample is drawn. This is repeated over many training iterations. We use SGD-like methods (*ADAM*) throughout this thesis.

The model we have introduced above is based on layers with neurons with number of weights equal to number of input pixels. For images, this model requires very large numbers of parameters making it both difficult to train, susceptible to overfitting, and for large enough images, difficult to fit into GPU memory.

In classic computer vision, hand-crafted convolutional filters (also known as kernels) are regularly used to find edges, take derivatives and smooth images. For instance, the matrices below represent a Sobel filter used to approximate image gradients along the x and y directions:

$$G_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \text{ and } G_y = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \textbf{(2.23)}$$

Convolutional filters are represented as a MxN matrices. Convolutions over images are taken by pointwise multiplying filter weights with image pixels, sliding the filter over, repeating the pointwise multiplication, and iterating. Convolutional neural networks (CNNs) replace hand-crafted convolutions with fixed weights with weights that are learned directly from the data itself. Notice that compared to fully connected layers, convolutions require significantly fewer parameters regardless of image size. Not only does using fewer parameters significantly reduce memory requirements for the model, but also helps to regularize the model and prevent overfitting. We use CNNs throughout this thesis.

2.7: Deep Learning and Image Reconstruction

In the context of using deep learning in MRI image reconstruction, two approaches have dominated. The first (somewhat older) approach is purely data driven [14]. It involves training a CNN on pairs of undersampled and ground truth images, and then applying the trained CNN to unseen data during inference. The primary issue with this approach is that it requires the CNN to learn to both remove undersampling artifact AND enforce data-consistency. This requires a large amount of training data which often isn't available. The second approach known as model based deep learning (MBDL) is very similar to the iterative reconstruction techniques we discussed in section 2.5 that alternate between data-consistency and regularizer steps [3]. In place of fixed regularizers like nuclear norm minimization that starts with prior assumptions about the data, MBDL use a CNN that learns to remove undersampling artifact (often through a supervised loss) directly from the data. MBDL is trained end to end by unrolling data-consistency and neural network steps for a fixed number of iterations. MBDL requires significantly

less training data then purely data driven approaches as the neural networks only have to learn to remove undersampling artifact. MBDL type architectures are used throughout this thesis.

All the ideas discussed in the background are used throughout this thesis. In **chapter 3** and **4**, I use MBDL to efficiently reconstruct highly undersampled 3D non-cartesian data sets. In **chapter 5**, I use a compressed representation originally developed in ([1]) for motion-compensated, large scale, time-resolved reconstructions.

Chapter 3: Memory Efficient Model Based Deep Learning Reconstructions for High Spatial Resolution 3D Non-Cartesian Acquisitions

In this chapter, I tackle the extreme GPU memory requirements seen when trying to apply model based deep learning to 3D non-Cartesian acquisitions.

3.1. Introduction

Fully non-Cartesian 3D trajectories offer many benefits over Cartesian methods. This includes enabling efficient acquisition in all three spatial directions, offering intrinsic motion and flow robustness, and allowing for ultrashort echo time imaging (1,2). For these reasons, acquisitions using 3D non-Cartesian trajectories are being developed and commercialized for highly accelerated imaging during free breathing, among other applications. However, one barrier to the clinical adoption of non-Cartesian imaging is the need for lengthy iterative reconstructions for parallel imaging and constrained reconstruction. Reconstruction times often remain clinically impractical even when run on graphical processing units (GPUs) (3).

Model based Deep Learning (MBDL) offers a principled technique for faster and higher quality 3D non-Cartesian reconstructions (4–6)**.** MBDL is similar to iterative reconstructions employed in compressed sensing (CS) that alternate between data-consistency steps that enforce the physical model of data acquisition and regularization steps that constrain image solutions to have certain assumed properties (7) (e.g. low rankness, sparsity). MBDL, however, uses a fixed number of iterations (unrolls), and in place of fixed regularizers, MBDL learns the regularizer from prior data using convolutional neural networks (CNNs). MBDL has consistently been faster and outperformed conventional compressed sensing reconstructions primarily in the context of 2D Cartesian acquisitions (4–6)**.**

Unfortunately, the application of MBDL to 3D non-Cartesian trajectories is challenging, in part, due to GPU memory limitations. For this reason, MBDL applied to the non-Cartesian setting has focused on reconstruction of relatively low resolution images and single channel data (8,9)**.** Unlike volumetric cartesian acquisitions and hybrid trajectories like stack of stars/spirals that can be decoupled into smaller 2D or 3D sub-problems, 3D non-cartesian acquisitions require solving the 3D reconstruction problem over the full volume at once. This requires that the entire 3D volume be passed through the deep learning regularizer prior to enforcing data-consistency when using MBDL. GPU memory requirements for a single unroll using networks routinely used for reconstruction (e.g., 32 or 64 channel residual networks) can easily be greater than 50 GB per unroll. This means realistic DL implementations with multiple unrolls can push the limits of even state of the art GPU clusters.

Gradient checkpointing is a memory efficient modification to traditional backpropagation that has been increasingly used to reduce memory requirements for neural network training(10–13). Unlike

traditional backpropagation where all intermediate features across unrolls are saved in memory for gradient computation, gradient checkpointing saves only a subset of these intermediates in memory, and then during backpropagation, recomputes missing intermediates to allow for gradient flow with only intermediates between checkpoints transiently held in memory. By balancing the number of saved intermediate with those recomputed, gradient checkpointing can be used to trade-off between computation and memory. In the context of MBDL, gradient checkpointing like methods has been used to increase the number of unrolls for 2D/3D Cartesian reconstructions (13)**.** This work assumes though that a single checkpointed unroll can fit in GPU memory. For high resolution imaging, passing the full volume through a single checkpointed unroll can still lead to prohibitively high memory usage. Thus, gradient checkpointing alone may not allow for high resolution, 3D non-Cartesian reconstructions with MBDL.

In other applications requiring 3D networks, patch-based methods are often used to reduce memory load. In such cases, input/supervision image pairs are broken into patches during pre-processing, and the neural network is trained directly on these patches. This application of patch-wise methods will not work for 3D non-Cartesian MBDL reconstructions because the full volume is required for each data-consistency step. However, if we decompose the volume during training into smaller patches, apply gradient checkpointing when pushing each patch through the network, and then recompose the full volume from the output patches for data-consistency such a method would combine the memory reducing benefits of patch-based trained while allowing for full-volume data consistency. We call this combination of gradient checkpointing and patch-wise CNN regularization allowing for full volume data-consistency: block-wise learning.

In this work, I explore the use of MBDL with block-wise learning to reconstruct highly undersampled, high resolution, fully non-Cartesian volumetric acquisitions on a single GPU. Specifically, I train an MBDL architecture using supervised learning with residual networks (14) alternating with multi-channel NUFFT data-consistency gradient steps. I investigate this network architecture for the reconstruction of 1.25mm isotropic, 3D pulmonary MRI radial acquisitions. MBDL with block-wise learning is then compared to L1 Wavelet Compressed Sensing in terms of image quality and reconstruction time.

3.2 Theory

3.2.1 Model Based DL

Consider the problem of reconstructing an image from under-sampled data y. For highly accelerated acquisitions, this problem is ill-posed and is often solved using minimization of a regularized least squares objective function:

$$\arg\min_{x} \|Ex - y\|^2 + \lambda R(x) \quad \textbf{(1)}$$

where x is the image to be reconstructed, E is the forward non-uniform Fourier transform (NUFFT) operator including sensitivity maps and density compensation, y is the acquired k-space data, and R is the regularizer with weight λ. The first term/data-consistency term ensures that solutions remain consistent with the acquired data. The second term/regularizer term constrains x to satisfy certain properties to encourage removal of under-sampling artifacts. Common regularizers include L1-sparsity in a given transform domain or low rankness for dynamic acquisitions. **Eq. 1** is generally solved for iteratively, often using gradient methods alternating between the data-consistency and regularizer steps.

MBDL leverages this model but replaces hand-crafted regularizers with a CNN, and alternates between the CNN regularizer in image-space and data-consistency steps for a fixed number of iterations also called unrolls. Given this unrolled model, the network weights can be trained end-to-end in a supervised fashion with outputs compared against ground-truth data using some pixel-wise distance metric (commonly an L2 norm). Such algorithms have been successful in achieving high quality images primarily for 2D Cartesian sampling problems (4,5).

MBDL methods for fully 3D non-Cartesian sampling are limited. Due to the GPU memory limitations discussed earlier, the only approaches to apply DL to fully volumetric non-Cartesian data up to this point have been to either 1) rely on patch-wise image-space training without iterative data-consistency enforcement, 2) pre-train a neural network regularizer (again using patch-wise image space training), and integrate this fixed regularizer into an unrolled framework, 3) use MBDL with lower resolution data. Below I present MBDL with block-wise learning that overcomes these constraints allowing end to end training of MBDL reconstructions.

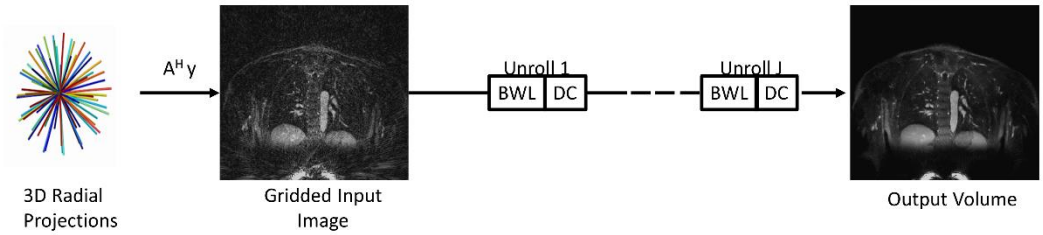
3.2.2 Block-wise Learning Algorithm

The block-wise learning algorithm applied to a single unroll of MBDL is as follows:

1. A $N_x \times N_y \times N_z$ zero-padded, undersampled image is decomposed into user-selected $P_x \times P_y \times P_z$ patches.
2. Individual patches are sequentially passed through the CNN regularizer with each patch gradient checkpointed.
3. The output blocks are then recomposed into the full volume with correction for edge artifacts due to zero-padding at internal edges (see appendix 1 for more detail).
4. The full volume is then passed to the data-consistency step. A standard gradient descent data consistency step is taken using 3D NUFFT operations. For multi-channel k-space data, k-space data-consistency is enforced iteratively one channel at a time. To fit this in memory, gradient checkpointing for each channel-wise data-consistency step is applied.
5. This technique is then applied to the next unroll

Figure 3.1a demonstrates the unrolled MBDL model with block-wise learning. **Figure 3.1b** demonstrates block-wise learning for a single unroll.

(a)



(b)

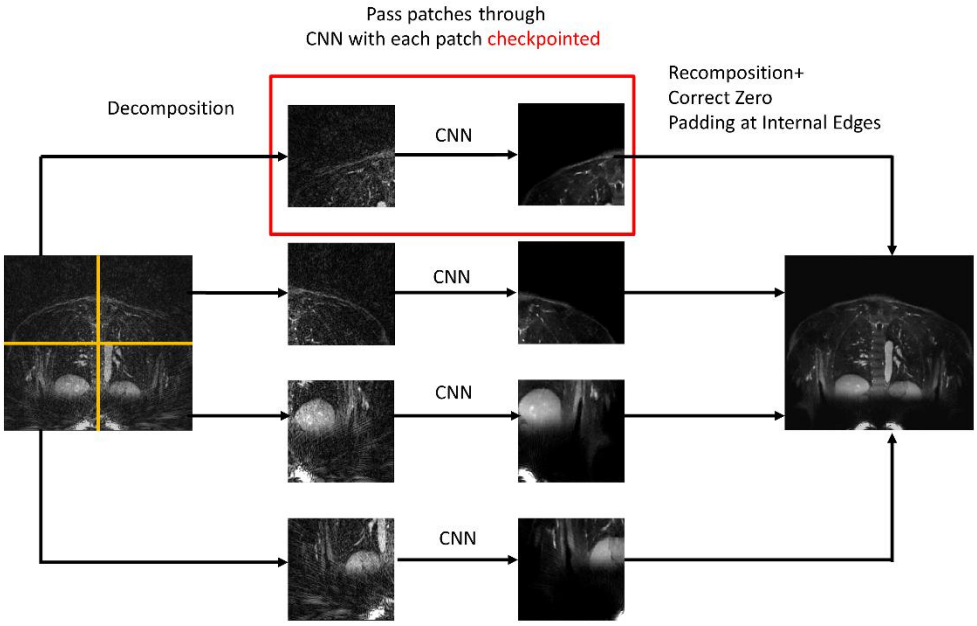


Figure 3.1: *MBDL with block-wise learning model.* ***(a)*** *The MBDL architecture with block-wise learning (BWL) is shown for all unrolls. A gridded image is reconstructed from an undersampled 3D radial acquisition and used as input to the MBDL architecture.* ***(b)*** *This input image is decomposed into smaller patches with each patch checkpointed. These patches are then iteratively passed through the CNN. The output patches are then recomposed into the full volume and zero-padding error correction is applied. The full volume is then passed to the data-consistency step. This process can be repeated for all unrolls.*

For a single unroll, block-wise learning with gradient checkpointing reduces memory by at most $\frac{N_x N_y N_z}{P_x P_y P_z}$ fold. For instance, a 300 x 300 x 300 volume broken into patches of size 150 x 150 x 150 results in an eight-fold reduction in memory use compared to pushing the entire volume through the unroll. As each unroll is effectively checkpointed, memory use scales across the entire architecture as $(N+1)x$ during the forward training pass where x is input array memory use. During the backward pass, memory use is $(N+1)x + \sum I_k$ where I_k are CNN intermediates proportional to patch size transiently saved in memory between checkpointed patches.

3.3 Methods

3.3.1 Non-Cartesian Data

Data acquired in 15 volunteers from a previously described study (15) was used for training and testing. In this study, post-Ferumoxytol (4mg/kg) contrast enhanced, pulmonary magnetic resonance angiography (MRA) UTE images were acquired during free breathing with respiratory positions recorded using a respiratory belt on a 3T MRI (MR750, GE Healthcare, Waukesha, WI, USA). Scan parameters included use of a 32-channel coil (Neocoil, Pewaukee, WI, USA), scan time of 5:45 minutes, TE=0.25ms, TR=3.6ms, and 1.25mm isotropic resolution. Four acquisitions were acquired per volunteer with flip angles of 6°, 12°, 18°, and 24°. A total of 94,957 projections were acquired using 3D pseudorandom bit-reversed view ordering (2). Data was coil compressed to 20-channels using PCA coil compression (16). The acquisition provided whole chest coverage with matrix sizes varying between 300-450 x 200-300 x 300-450 based on automatic field of view determination. Density compensation was normalized using the max eigenvalue of the NUFFT operator, and k-space was rescaled based on this (17).

Fully sampled data is difficult to obtain for pulmonary UTE acquisitions so a proxy for fully sampled data was used. The 50,000 spokes closest to the end-expiratory phase were reconstructed using 30 iterations of conjugate gradient SENSE and used for supervision. Coil sensitivity maps were determined using JSENSE (18).

From the 15 volunteers imaged, 8 cases were used for training and 1 case for validation. For training, only acquisitions with the highest flip angle (24°) were used. The remaining 6 cases were used for testing. Performance was evaluated between images with the same contrast as the training data (flip angle 24°) and in images collected with a lower flip angle of 6°. For training, retrospectively undersampled images were generated by randomly selecting 5,000 spokes from the ground truth data. Radial projections were selected at random during each training iteration to mitigate the effect of differing motion states between the ground truth and subsampled data. Separate coil sensitivity maps using JSENSE (18) were then generated. Gridded images from this retrospectively undersampled k-space data were used as input to the model.

3.3.2 MBDL Architecture

MBDL with block-wise learning was implemented using PyTorch (Open Source, https://pytorch.org/) with an Adam optimizer and NUFFTs from SigPy (Open Source, https://github.com/mikgroup/sigpy) on Intel Xeon workstations using one 40 GB A100 GPU. MBDL with block-wise learning has several tunable parameters including number of unrolls, choice of neural network architecture, and choice of block size during the neural network step. Architecture choices were guided by prior literature on MBDL models (network choice and number of training cases (14)), required GPU memory and ease of padding correction (choice of block-size), and a small-scale experiment was run to investigate optimal unroll number. For this experiment, I used lower resolution data (readout length 300 points, spatial resolution 1.91 mm isotropic) to reduce the substantial training time.

Similar to (14), a residual network (32 channels/conv, 3D conv with 3 x 3 x 3 kernels, no bias) with Leaky-ReLU activations (using in place activation) was used. Input to the architecture consists of complex-valued volumes converted to 2-channel images representing real and imaginary components. The architecture was then trained to minimize the mean square error between model output and ground truth 2-channel supervision data. For data-consistency, I used multi-channel non-uniform fast Fourier transform (NUFFT) gradient descent steps with a learnable step size. To fit this into memory, gradient checkpointing was applied along the NUFFT channel dimension.

Choice of block size is a trade-off between memory savings and number of internal volume edges that must be corrected due to padding artifacts. For this work, each volume dimension was divided in two, yielding eight blocks and 12 edges that required padding correction. Matrix sizes up to 500 x 500 x 500 were capable of being processed using this choice of block size on the A100 GPU which is sufficient for use in this work. Smaller block sizes could be utilized to reduce memory but require additional padding correction steps.

The model was trained for 4,000 iterations using a learning rate of 1e-3. In the low-resolution experiment, several models were trained including a model with no data-consistency term (residual network alone), and MBDL models with 1,3 and 5 unrolls respectively. Finally, MBDL was trained at full resolution using 5 unrolls.

3.3.3 Evaluation

The performance of MBDL with block-wise learning was evaluating by comparing reconstructions to proxy ground truth images obtained by taking the first 50,000 spokes closest to end-expiration and then reconstructed using CG-SENSE and to L1 Wavelet Compressed Sensing (CS) reconstructions (100 iterations, regularization weight: .0001). The primary goal of this evaluation was to demonstrate that

MBDL with block-wise learning can reconstruct very large matrix arrays in a memory and time efficient manner while simultaneously out-performing CS reconstructions.

For the low-resolution experiment investigating the impact of number of unrolls on image quality, test data was generated by retrospectively and randomly undersampling radial projections from the ground truth k-space data to 5,000 spokes. For each unroll number, PSNR and SSIM relative differences were computed. PSNR and SSIM relative difference is defined as the difference between the PSNR and SSIM of the model output from the PSNR and SSIM of the gridded undersampled image all compared to the proxy ground truth data.

Test data for reconstructions at full resolution was generated by retrospectively and randomly undersampling radial projections from the proxy ground truth k-space data to 10,000 spokes. We first compared test data reconstructed using MBDL with the same contrast (flip angle 24°) as the training data to L1 wavelet CS reconstructions run on the same GPU. We then investigated the ability of MBDL to reconstruct the same underlying patient anatomy, but with a different contrast (flip angle 6°). Image quality was evaluated quantitatively using PSNR/SSIM relative difference from the gridded image (as defined earlier) against L1 Wavelet CS methods and qualitatively with a radiology reader study. In the reader study, a radiologist blinded to reconstruction type was asked to choose the reconstruction preferred between L1-wavelet and MBDL reconstructions across test cases. We then investigated how image quality and PSNR/SSIM change as a function of number of radial projections by reconstructing full resolution images at 15k, 10k, 17.5k and 5k spokes on a representative case.

All statistical comparisons between reconstructions were run using paired t-tests. Differences between reconstructions were considered significant if $P < 0.05$.

3.4 Results

3.4.1 Hyperparameter Choices: Number of Unrolls

Figure 3.2 demonstrates that both PSNR and SSIM relative difference increase as a function of number of unrolls except for the MBDL architecture with one unroll for retrospectively undersampled data with 5k projections. SSIM and PSNR relative difference for reconstructions with five unrolls are significantly higher ($P < 0.01$) when compared against the purely data-driven architecture (0 unrolls) and MBDL architectures with 1 and 3 unrolls respectively.

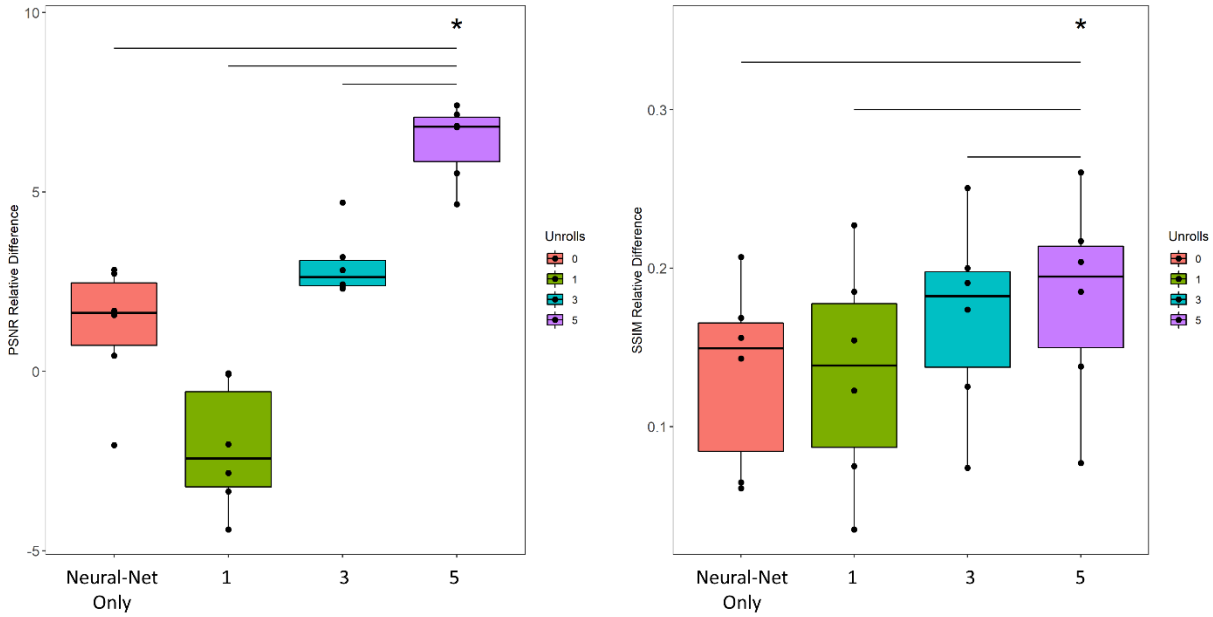


Figure 3.2: *MBDL performance vs. number of unrolls. The impact of number of unrolls on the MBDL architecture with block-wise learning was investigated by training and testing four different models on lower resolution (~2 mm isotropic), highly accelerated data (5k spokes). The models trained included a residual network without data-consistency (0 unrolls), and MBDL models with 1,3 and 5 unrolls respectively. Image quality was evaluated using PSNR and SSIM relative differences across six test cases with identical contrast to the training data (flip angle 24°). The model with five unrolls had significantly greater PSNR relative difference (P<.001) and SSIM relative difference (P<.001) than all other models as shown by the asterisk. Statistical comparisons between other models were not computed.*

Figure 3.3 demonstrates a representative example of how image quality improves with unroll number. In this coronal section, increasing unroll number is associated with improved ability to resolve vascular features. This is particularly striking when moving from the neural network only model in column 1 to the MBDL-based methods that have data-consistency terms in columns 2-5. Although PSNR and SSIM relative difference are lowest for the architecture with one unroll, visually, small vascular features are seen (orange arrow) that are not observed in the neural network only model. These small vascular features are sharpened further in the three and five unroll architecture (orange arrow). Notice though at this acceleration there is significant drop out and some blurring of vascular features across reconstructions compared to the proxy ground truth.

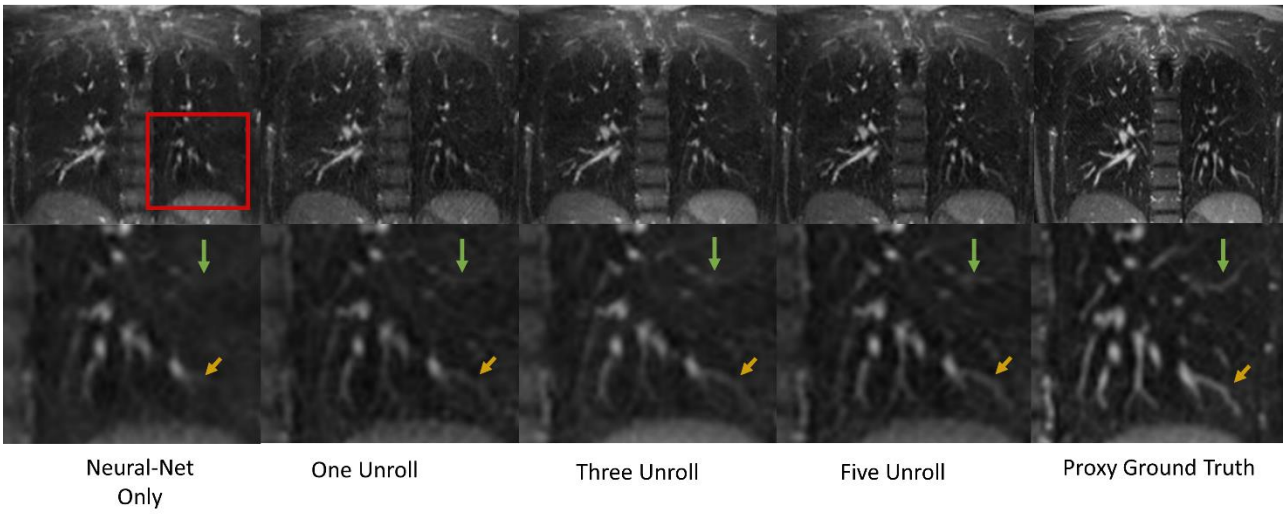


Figure 3.3: *Reconstruction quality versus number of MBDL unrolls. Coronal slices from volumetric reconstructions of gridded images with 5k spokes are shown here compared to each other and the proxy ground truth (50k spokes). The neural network only reconstruction compared to any of the MBDL architectures is less able to resolve small vascular features (yellow arrow). As unroll number increases, the ability to resolve these small vascular features improves. It is important to note though that relative to the proxy ground truth, there is feature loss (green arrow) across the neural network reconstructions independent of unroll number*

Total training time is around 8 days for the five unroll architecture trained on full resolution data. Based on run-time for a single forward and backward pass, training with seven unrolls would take 12-13 days, and training with 9 unrolls would take ~20 days. To keep training times reasonable, the five unroll architecture was chosen.

3.4.2 Full Resolution Results

Figure 3.4,3. 5 and 3.6 shows PSNR (left) and SSIM relative difference (right) in the test subjects by reconstruction and contrast type (flip angle 24° and flip angle 6°) for data retrospectively undersampled to 10k projections.

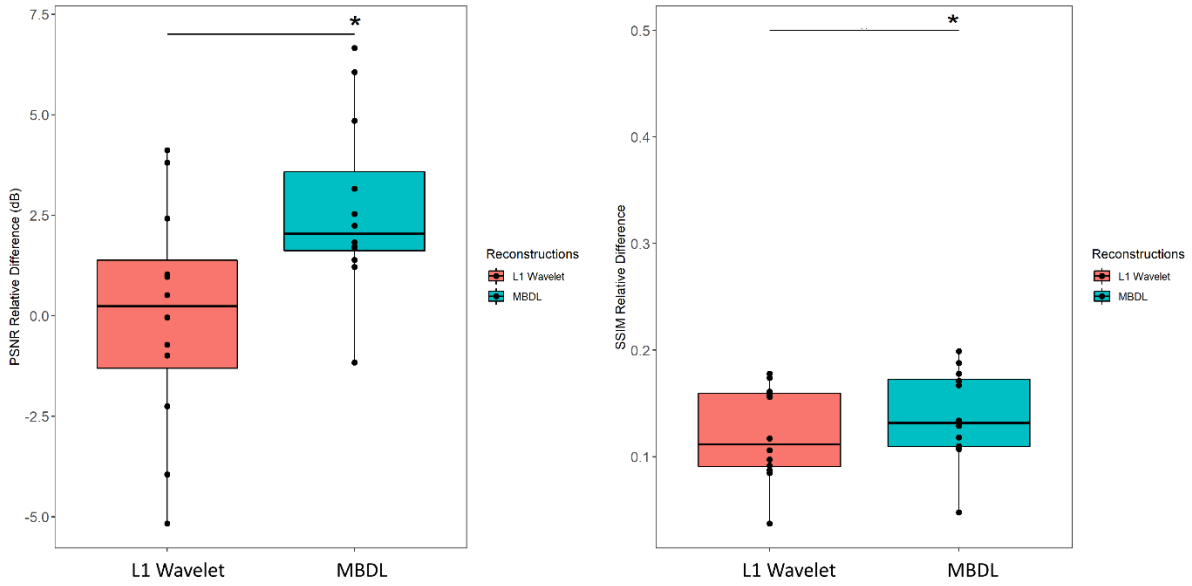


Figure 3.4*: L1 Wavelet vs. MBDL performance for all test data (both flip angles). High resolution (1.25 mm isotropic) volumes were reconstructed using an MBDL architecture with five unrolls using block-wise learning and compared to L1 Wavelet reconstructions. MBDL had significantly higher PSNR relative difference (P<0.005) and SSIM relative difference (P<1e-5) than L1 wavelet reconstructions.*

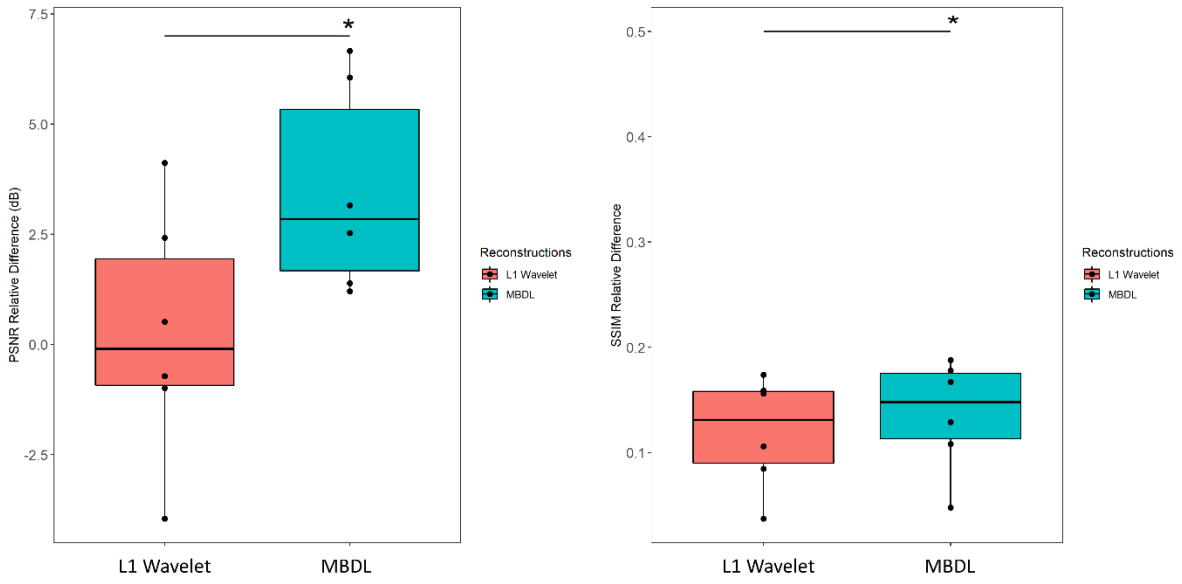


Figure 3.5*: L1 Wavelet vs. MBDL performance for flip angle 24°. These box plots compare test data with similar contrast (flip angle: 24°) to that seen by MBDL during training. MBDL significantly outperformed L1 wavelet in terms of both PSNR relative difference (P<0.05) and SSIM relative difference (P<1e-3)*

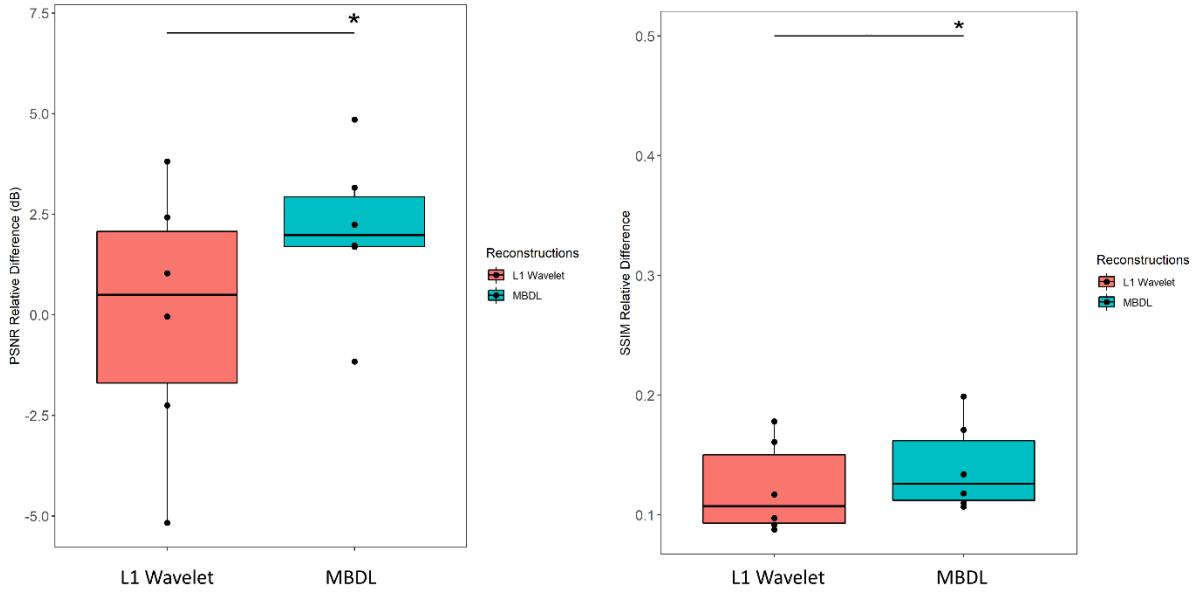


Figure 3.6*: L1 Wavelet vs. MBDL performance for flip angle 6°. These box plots compare data with different contrast (flip angle: 6°) to that seen by MBDL during training. MBDL significantly outperformed L1 wavelet in terms of both PSNR relative difference (P<1e-3) and SSIM relative difference (P<1e-4)*

MBDL with block-wise learning significantly outperforms L1 wavelet CS reconstructions ($P < 0.01$) across all comparisons. This includes significantly outperforming L1 wavelet CS reconstructions across both contrast types. The reader study blinded to reconstruction method validated these findings

with the radiologist preferring MBDL reconstructions in 12/12 comparisons primarily due to the sharpness of the vasculature in MBDL.

Figure 3.7 shows representative coronal slices for gridded, L1 wavelet, MBDL, and proxy ground truth reconstructions from acquisitions with a 24° flip angle.

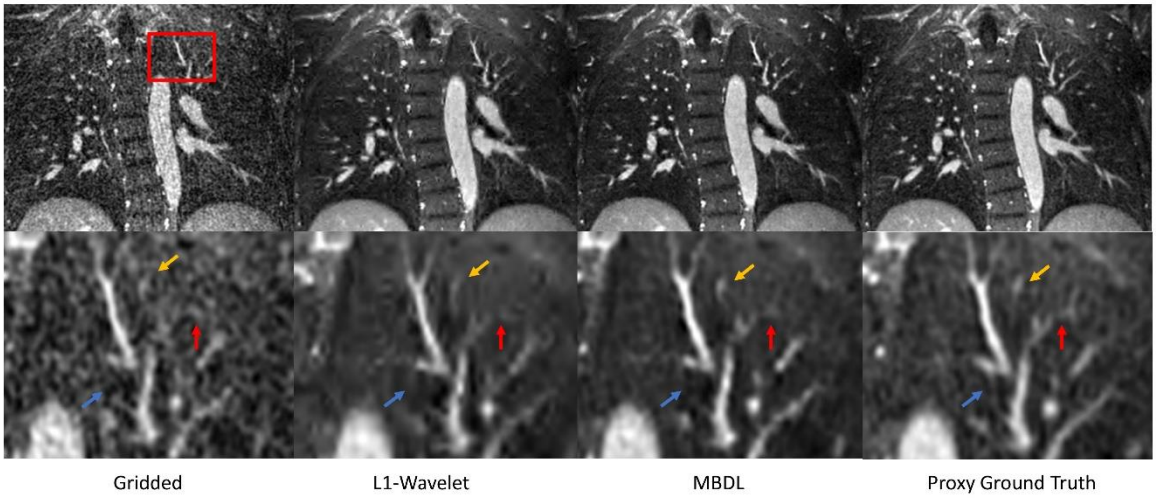


Figure 3.7*: Coronal slices from high resolution volumetric images using different reconstruction strategies. L1-wavelet and MBDL images were reconstructed from retrospectively undersampled data with 10k spokes. Proxy ground truth data had 50k spokes. The gridded image (column 1) has significant undersampling artifact present in the zoomed-out and zoomed-in images. This undersampled artifact obscures small vascular structures. L1 wavelet, MBDL and the proxy ground truth have significantly reduced undersampling artifact. The zoomed-in images though show significant blurring in the L1 wavelet reconstruction that obscures structures (red arrow) that can be seen in both MBDL and the proxy ground truth reconstructions. The proxy ground truth zoomed-in image has smoother vascular structures than MBDL and resolved some features (blue arrow) not seen in MBDL. Interestingly though, MBDL does resolve a feature (orange arrow) that cannot clearly be seen in the proxy ground truth image*

The gridded image has a significant amount of undersampling artifact that obscures vascular structures. Both L1-wavelet and MBDL significantly reduce this undersampling artifact as can be seen in row 1. In the zoomed-in slices in row 2, L1-wavelet reconstructions blur both small vascular features (orange arrow) and lung parenchyma. These features in the MBDL reconstructions are sharper and closer to the proxy ground truth although blockier in appearance. There is some drop-out of small vascular features (blue arrow) and blurring of features (red arrow) relative to the proxy ground truth in the MBDL reconstruction. However, some features are resolved in the MBDL reconstruction that are not visible in the proxy ground truth (orange arrow).

Figure 3.8 shows PSNR (left) and SSIM (right) relative difference comparisons between MBDL reconstructions on the same patient but with different contrasts (flip angle 24° and flip angle 6°). No significant differences in quantitative difference in PSNR ($P < .349$) or SSIM ($P < .214$) relative differences were observed.

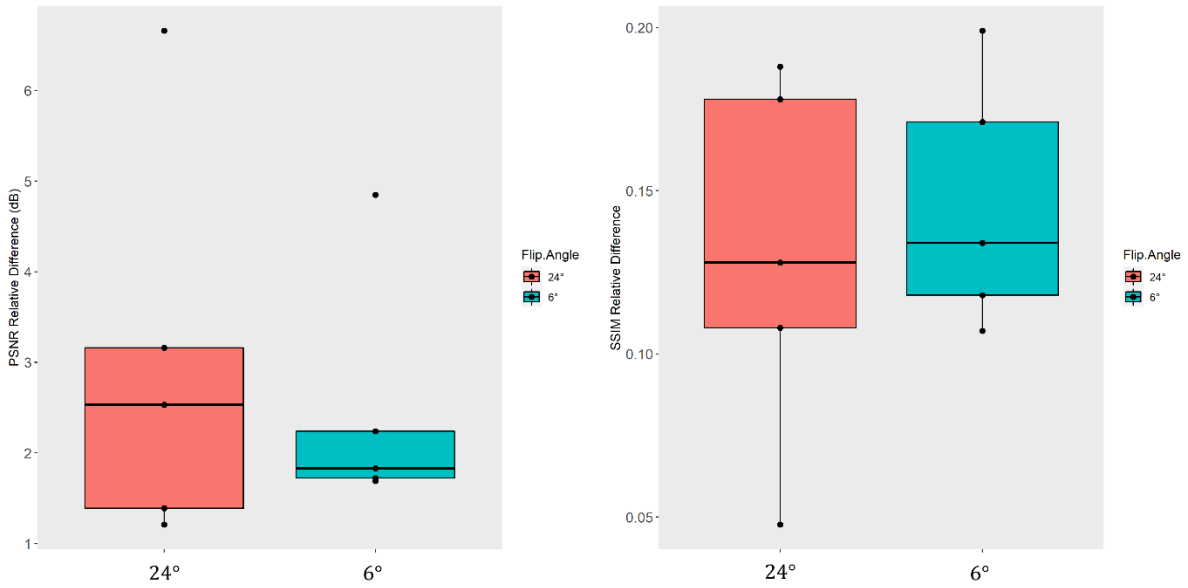


Figure 3.8: *Performance of MBDL on different flip angles. MBDL reconstructions from acquisitions on the same volunteer, but with different flip angle were compared using PSNR and SSIM relative difference. No statistically significant differences in performance were observed. Note only five paired contrasts were used for this comparison as the data from one acquisition in the sixth pair was corrupted*

Figure 3.9 shows matched representative sagittal slices for ground truth, L1 wavelet, MBDL and proxy ground truth reconstructions for both flip angles.

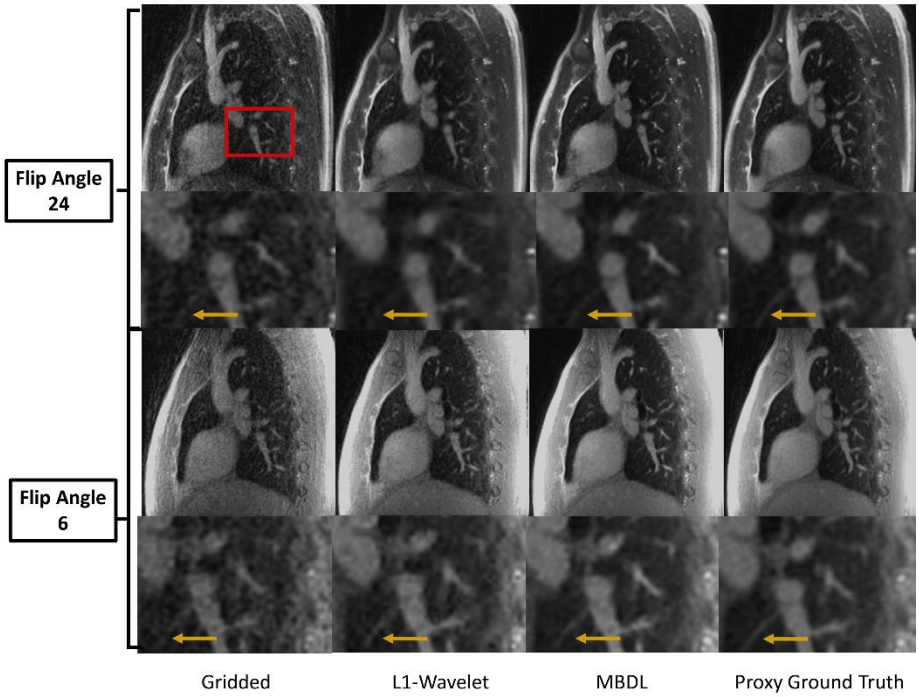


Figure 3.9: *Reconstructions of acquisitions from the same volunteer, but with different flip angles (flip angle 24° and flip angle 6°). Both gridded images have significantly more undersampling artifact than all other reconstructions. MBDL reconstructions for both flip angles were sharper and visualized smaller features (yellow arrow) better than L1 wavelet reconstructions. The proxy ground truth had higher quality images for both contrasts than all other reconstructions.*

For both flip angles, MBDL reconstructions are sharper than the L1-wavelet reconstructions. This can be most clearly seen in the zoomed-in view. There is minimal visual deterioration in quality between the contrast (flip angle 24°) the neural network was trained on and the reconstruction with different contrast (flip angle 6°).

Figure 3.10 shows a representative axial slice from MBDL reconstructions with 5k, 7.5k, 10k, and 15k radial spokes and proxy ground truth with 50k spokes.

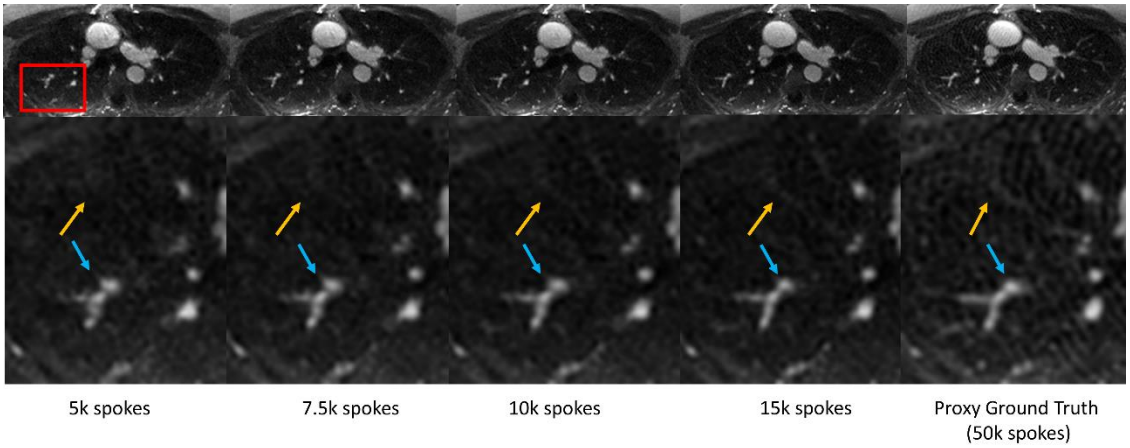


Figure 3.10: *MBDL Reconstruction with varying numbers of spokes. Similar image quality can be seen across a relatively wide range of accelerations (row 1). No significant differences in streaking or undersampling artifact are seen although features appear sharper as the number of spokes increases. . Interestingly, there is artifact present in the proxy ground truth (wave-like streaks across the field of view) not seen in the MBDL reconstructions. In row 2, the ability of the reconstruction to resolve subtle vascular features (yellow arrow) improves with increased number of spokes. Further, the blockiness of the y-shaped vascular structure is reduced with increased number of spokes. PSNR/SSIM values were 5k spokes: 47.6/.984, 7.5k spokes: 48.1/.985,10k spokes: 47.3/.983, 15k spokes: 46.6/.983*

The ability to capture small vascular features (yellow arrow) improves with increasing number of spokes, however, overall, the reconstructions did not differ significantly in image quality. There is artifact present in the proxy ground truth reconstruction not seen in the other reconstructions.

The average reconstruction time for L1 wavelet CS reconstruction (100 iterations) was 872 ± 32 seconds versus 23 ± 4 seconds for MBDL with block-wise learning on the same A100 GPU representing a ~38X speed-up in reconstruction time.

3.5. Discussion

In this work, I demonstrate a block-wise training approach that allows MBDL to be applied to the reconstruction of accelerated, high resolution, fully non-Cartesian volumetric acquisitions. For a single unroll, this approach in combination with gradient checkpointing takes an input volume, decomposes this volume into a series of smaller patches, passes each patch iteratively through the CNN, recomposes the patch output into the full volume, performs padding artifact correction, and then sends the full volume to the data-consistency step. This algorithm on a 40 GB GPU enabled the training and reconstruction of volumes with matrix sizes up to 500 x 500 x 500 from 3D Radial MRI acquisitions. Memory use over all unrolls during the forward pass scales as $(N+1)x$ where N is the number of unrolls and x is the memory required to store the input array. The backward pass scales as $(N+1)x + \sum I_k$ where I_k are CNN intermediates proportional to patch-size held in memory between checkpointed patches. MBDL with

block-wise learning demonstrated significantly reduced reconstruction time (~38X faster) and improved image quality over L1 Wavelet Compressed Sensing run on the same GPU. The architecture was further shown to generalize to acquisitions with different contrast and different levels of undersampling.

This work specifically addresses the GPU memory constraints seen when trying to apply MBDL to high spatial resolution, volumetric non-Cartesian data. In this context, passing the full volume through single unrolls even with gradient checkpointing may not fit in available GPU memory. Block wise learning in large part removes these GPU memory constraints for high dimensional problems because in place of the full-volume, GPU memory per unroll is instead tied to user-selected patch size. This approach not only applies to reconstruction of single frames as seen in this work, but with smaller block sizes, could be applied to dynamic volumetric reconstructions. Compared to prior MBDL work on 3D non-Cartesian reconstructions which has been limited to single channel, low spatial resolution data (8,9) due to GPU memory constraints, block-wise learning extends MBDL to multi-channel, high spatial resolution, 3D non-Cartesian acquisitions.

While this study was aimed at demonstrating the feasibility of 3D non-Cartesian deep learning, a step is taken toward development of high-resolution, breath held acquisitions by reconstructing retrospectively undersampled, highly accelerated acquisitions (5000-10000 spokes, approximately a 15-30 second breath-hold). MBDL significantly outperforms L1-wavelet methods in terms of image quality both quantitatively and through a reader study while simultaneously shortening reconstruction time from minutes to seconds. Image quality, however, is not yet comparable to state-of-the-art motion resolved reconstructions (19) or the proxy ground truth images. The primary issues observed are blocky vascular structures and drop-out of small vascular features.

There are several potential issues that may have limited performance in this context. First, a limited number of unrolls (five unrolls) was used primarily to maintain reasonable training times as five unrolls over 4000 iterations corresponded to around 8 days of training. Moving to seven or nine unrolls would extend training time to weeks. It is clear from **figure 3** that increasing number of unrolls improves ability to capture small vascular features Recent work by (12,13) also demonstrates improved ability to resolve small features with increasing number of unrolls. Given the limited number of data-consistency steps used, the model was probably not taking full advantage of parallel imaging which may account for the loss of small vascular features seen across the MBDL reconstructions.

In general, increased training time is a drawback to the use of gradient checkpointing as intermediates need to be recalculated during the backward pass. This increased training time not only impacted the number of unrolls used, but also limited the number of training iterations that could reasonably be run. The network is likely highly underfit to the underlying data and would be more so if additional training data was used.

There are several potential ways to address these issues. Replacing gradient descent with conjugate gradient steps computed as in (4) would likely allow the architecture to take advantage of parallel imaging and improve convergence without requiring more unrolls. Use of computationally efficient alternatives to gradient checkpointing like those suggested in (12) may also reduce training time. Further, the code used for training had not been optimized for speed.

Another issue is our proxy ground truth is a composition of several motion states meaning there are features present in subsets of the ground truth data that are blurred out in the ground truth. In addition to removing undersampling artifacts then, MBDL was being asked to learn to blur and remove features. This effect, however, should have been mitigated somewhat during training by randomly selected spokes each MBDL pass. A potential solution to this issue is to use self-supervised learning so that reliance on ground truth proxies is no longer necessary.

3.6.Conclusion:

Model based deep learning with block-wise training allows for reconstruction of high resolution, volumetric, non-cartesian acquisitions on a single GPU. This work lays the foundation for future development for MBDL reconstruction of volumetric breath-held, respiratory binned and time resolved data.

3.7.Supplemental Section

3.7.1 Zero Padding Correction

Passing patches through a CNN is not equivalent to passing the full volume through a CNN due to zero padding at internal edges. To see this, consider a 1D convolution of a length 8 array vs a 1D convolution of the same length 8 array broken into two length 4 arrays and then concatenated together. Let $f(x) = [1,1,1]$ be the convolution kernel and $g(x) = [1,1,1,1,1,1,1,1]$ be the length 8 array. Entry 0 and Entry 7 are zero padded. The zero-padding convolution then is

$$f(x) * g(x) = [2,3,3,3,3,3,3,2] \quad \textbf{(2)}$$

Consider the second convolution where the convolution kernel is as before, but $g(x) = [g_1(x), g_2(x)]$ where $g_1(x) = [1,1,1,1]$ and $g_2(x) = [1,1,1,1]$. We preserve the original numbering from the length 8 array. Notice now that zero padding is applied not only to entry 0 and entry 7, but also to entry 3 and entry 4 as these are new edges created by splitting $g(x)$ that will be zero padded. The convolution of each length 4 array is:

$$f(x) * g_i(x) = [2,3,3,2] \text{ for } i = \{1,2\} \quad \textbf{(3)}$$

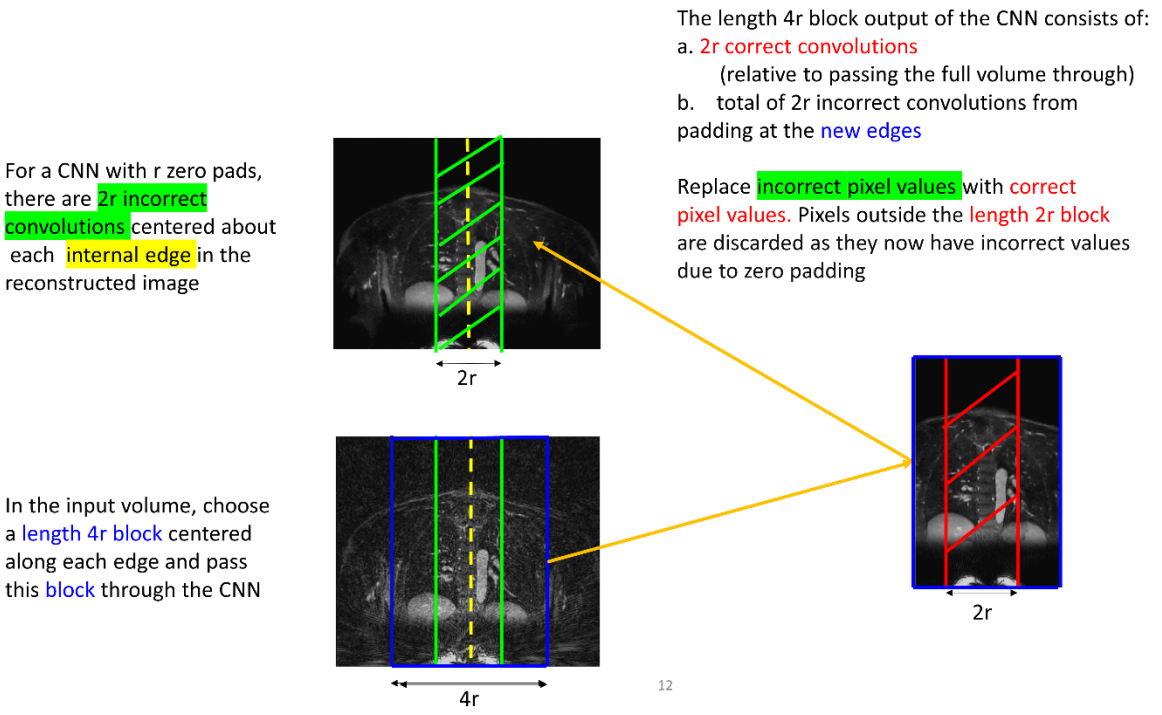
Concatenating these individual convolutions back together yields:

$$(f(x) * g_1(x))\ concat\ (f(x) * g_2(x)) = [2,3,3,2,2,3,3,2] \quad \textbf{(4)}$$

Notice that **(2)** differs from **(4)** only at entry 3 and 4 where the new edges were created.

To correct this after convolving the two length 4 arrays, simply choose a new subset κ of $g(x)$ that **1)** contains entry 3 and entry 4 and **2)** when convolved with $f(x)$, entries 3 and 4 are not zero padded. For instance, choose a length 4 block centered on entries 3 and 4: [1,1,1,1], convolve $f(x) * k(x) = [2,3,3,2]$. The middle two entries in this array correspond exactly to the incorrect convolutions in the concatenated array. Throw out the new zero padded entries in $f(x) * k(x)$ and replace the incorrect entries in **(2)** with [3,3].

This intuition can be used to derive a general algorithm for zero padding correction as shown in supplemental figure 3.1:



Supplemental Figure 3.1: *General Method for zero-padding correction. Any edge in a patch not also seen in the full volume has padding errors. Padding correction is applied after rebuilding the volume. If a CNN regularizer (NN) has r zero pads, in general there are 2r incorrect convolutions centered about new edges created by decomposing the volume into patches. This can be corrected by choosing a length 4r block centered on the edge dimension. This new block is then passed through CNN regularizer. The convolution errors are now clustered along the outer edges generated from this new block while the inner 2r convolutions are correct. This correct inner 2r convolutions in this patch in blue can then replace the incorrect 2r convolutions originally.*

4.1.Introduction

3D non-Cartesian trajectories are increasingly used for free breathing acquisitions as they are motion robust and allow for retrospective respiratory binning. Such acquisitions have the potential to be applied to a range of clinical applications including lung [15], 4D-Flow [27], and dynamic contrast enhanced imaging [1]**.** These scans are particularly powerful imaging patients with difficulty breath-holding (e.g. pediatric lung imaging [28], [29]), and when high spatial resolution imaging is needed [15]. These free breathing acquisitions have benefited substantially from advanced image reconstruction techniques such as temporal compressed sensing (e.g. XD-GRASP [30]) and motion compensated reconstruction approaches (e.g. iMoCo [26]). These reconstruction methods, however, are computationally demanding making these acquisitions difficult to integrate into clinical practice.

Model based deep learning (MBDL) applied to these acquisitions has the potential to simultaneously reduce reconstruction time while improving image quality relative to compressed sensing [3], [4]**.** In MBDL, a fixed number of iterations are unrolled alternating between data-consistency steps that enforce the physical model of acquisition and convolutional neural network (CNN) steps that learn directly from the data to remove under-sampling artifact. For Cartesian acquisitions, MBDL has been shown to significantly outperform compressed sensing [3]. The application of MBDL to high spatial resolution, 3D non-Cartesian imaging has been challenged by GPU memory constraints; however, recent progress combining gradient checkpointing with patch-based methods has enabled memory efficient 3D non-Cartesian MBDL-based reconstructions.

MBDL is typically trained using supervised methods limiting this technique to situations where ground truth images can be obtained. Fully sampled ground truth images, however, are essentially impossible to acquire for free breathing 3D non-Cartesian acquisitions as k-space sampling is limited by respiratory motion. For instance, in retrospectively respiratory gated 3D pulmonary UTE imaging, end inspiratory frames contain very few projections (e.g. 2-7k projections), even for long (5-10 minute) scans [2].

Self-supervised MBDL is a promising method that allows training without ground truth data [6] that could be applied to address the difficulty obtaining fully sampling ground truth data for 3D non-Cartesian acquisitions. In this approach, acquired k-space data is partitioned into two subsets. One k-space subset is used as input to the MBDL architecture, and the second k-space subset is used only in the self-supervised loss term during training. Self-supervised MBDL is then trained to start with data from one subset of k-space data and solve for the other. This is similar to Noise2Noise approaches found in the computer vision literature [31].

A limitation, however, of self-supervised MBDL in its current form is it only leverages spatial correlations in the data. I refer to this technique as spatial self-supervised MBDL. 3D non-Cartesian acquisitions even when respiratory binned are often highly undersampled, particularly phases close to end-inspiration. Reconstruction methods that leverage spatial correlations alone are often unable to recover sufficient quality images from these highly accelerated acquisitions.

Iterative methods used to reconstruct 3D non-Cartesian data often use the fact these acquisitions are acquired dynamically; that is, they take advantage of correlations across frames**.** These methods can reconstruct more highly under-sampled data than methods that rely on spatial regularization alone. Examples include temporal difference (XD-Grasp) and nuclear norm regularized reconstructions [30], [32]**.** Recent work incorporating non-rigid motion field estimation into these reconstructions has demonstrated even higher quality results as aligning data improves correlations across frames [26], [33] .

In this work, I investigate the combination of a self-supervised MBDL architecture called dynamic MBDL that takes advantage of correlations across frames with efficient GPU-based motion correction to reconstruct a single respiratory phase from free breathing 3D non-Cartesian acquisitions. This technique consists of three steps: 1) motion resolved reconstructions using dynamic MBDL trained on unregistered data, 2) motion field estimation by registering the motion resolved reconstructions, 3) reconstruction of a single respiratory phase using a final dynamic MBDL architecture with both training and inference on registered data. This is similar to the steps proposed in the iterative motion compensated (iMoCo [26]) technique but replacing iterative methods and CPU-based image registration with MBDL and GPU-based registration. As proof of concept, I apply this technique to reconstruct the end-inspiratory phase from high resolution (1.25 mm isotropic) respiratory binned 3D pulmonary UTE acquisitions. I compare image quality and reconstruction time to spatial self-supervised MBDL, XD-grasp and iMoCo reconstructions.

4.2 Theory

In this section, I review the original self-supervised MBDL method, and introduce the dynamic MBDL architecture and GPU-based motion correction used in our motion compensation technique.

4.2.1 Self-supervised MBDL

MBDL is based on classic iterative reconstructions that alternate between regularizer steps that constrain the space of possible solutions and data-consistency steps that enforce that physical model of acquisition. MBDL unrolls a fixed number of iterations and in place of regularizers that assume certain properties of the input data, MBDL uses CNNs that learn to regularize directly from the data itself. To accelerate convergence due to the limited number of unrolls that can fit in GPU memory, MBDL architectures often use conjugate gradient iterations for data consistency [3] .The majority of work using

MBDL has focused on 2D Cartesian reconstructions [3]–[5] trained using fully sampled ground truth data minimizing:

$$\min_{\theta} \frac{1}{\mathrm{N}} \sum_{i=1}^{N} \mathcal{L}_I(x_{ref}^i, f(y_\Omega^i, E_\Omega^i; \theta) \qquad [1]$$

Where $\mathcal{L}()$ is the supervised loss (often an L1 or L2 norm enforced in image-space), x_{ref}^i is an example ground truth image from the training set, y_Ω^i is an example gridded training image reconstructed from retrospectively undersampled k-space data Ω, $f(y_\Omega^i, E_\Omega^i; \theta)$ is the MBDL image-space output architecture with E_Ω^i representing the combination of the Fourier transform operator and coils sensitivity maps, and θ is the set of learnable network weights. I use the same notation found in Yaman Et al. [6].

Noise2Noise (N2N) methods offer an alternative approach to supervised methods when ground truth data is unavailable. In place of supervised training that uses pairs of corrupted and ground truth images to learn to remove various artifacts, Lehitenen Et al. show that simply by training on pairs of differentially corrupted images, the neural network will learn the average of the distribution of these corrupted images i.e. the clean image.

In the original N2N paper [31], this method was applied to MRI image reconstruction as proof of concept. Fully sampled Cartesian k-space brain data was subsampled to generate undersampled image pairs from the same volume. A purely data driven network architecture without data-consistency steps was trained with 5,000 image pairs by enforcing the L2 norm between the Fourier transformed neural network output and the k-space data unseen by the network. On retrospectively undersampled test data (up to 10X acceleration), N2N reconstructions had comparable PSNR and visual quality improvements to images reconstructed by networks trained using supervision. Although this example demonstrates that N2N can be applied to MRI reconstruction, it assumes ground truth data to subsample from is available. For many acquisitions, ground truth data to subsample from is not available even during the training phase. Training a network on subsampled data that is already accelerated to start with may not, on average, yield an image close to fully sampled ground truth.

Recent work by Yaman Et. al explored the performance of N2N approaches trained by subsampling accelerated k-space data. In addition, they integrated the N2N framework into MBDL. In their work [6], Yaman Et al. demonstrated that this approach could still reconstruct high quality images comparable to supervised methods when trained on undersampled Cartesian data.

In this approach, undersampled k-space Ω is divided into disjoint subsets Θ and Λ such that $\Omega = \Theta \cup \Lambda$. The supervised loss in Eq. 1 is replaced with:

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_k(y_\Lambda^i, f(y_\Theta^i, E_\Theta^i; \theta)) \qquad [2]$$

Where $\mathcal{L}_k$ is a self-supervised loss enforced in k-space, y_Λ^i represents the vector of k-space entries associated with k-space subset Λ, y_Θ^i represents the vector of k-space entries from k-space subset θ that is transformed into a gridded image and used as input to the MBDL architecture. The loss is enforced in k-space between the image space output of MBDL transformed back to k-space: $f(y_\Theta^i, E_\Theta^i; \theta)$ and y_Λ^i,.

As discussed in the introduction though, the reliance of this architecture on spatial correlations alone can result in lower performance for dynamic applications where high levels of undersampling are required.

4.2.2 Dynamic MBDL Architecture

To overcome the limitations of spatial self-supervised learning, I propose dynamic MBDL, a self-supervised MBDL architecture that leverages correlations across frames to boost image quality. Consider spatial self-supervised MBDL as discussed above. This framework consists of N unrolls where each unroll alternates between residual networks that remove undersampling artifact on a single image and conjugate gradient data consistency steps. To incorporate the ability to leverage correlations across frames into MBDL, I propose a small modification to this model. For the first unroll, I replace the spatial residual network with an encoder-like residual network that takes in N data-frames along the channel dimension as input and outputs a single frame (out_1) . This output is then passed to conjugate gradient data-consistency steps and spatial residual networks in downstream unrolls. All data-consistency steps and the self-supervised loss are enforced on only a single target frame **(figure 1a).**

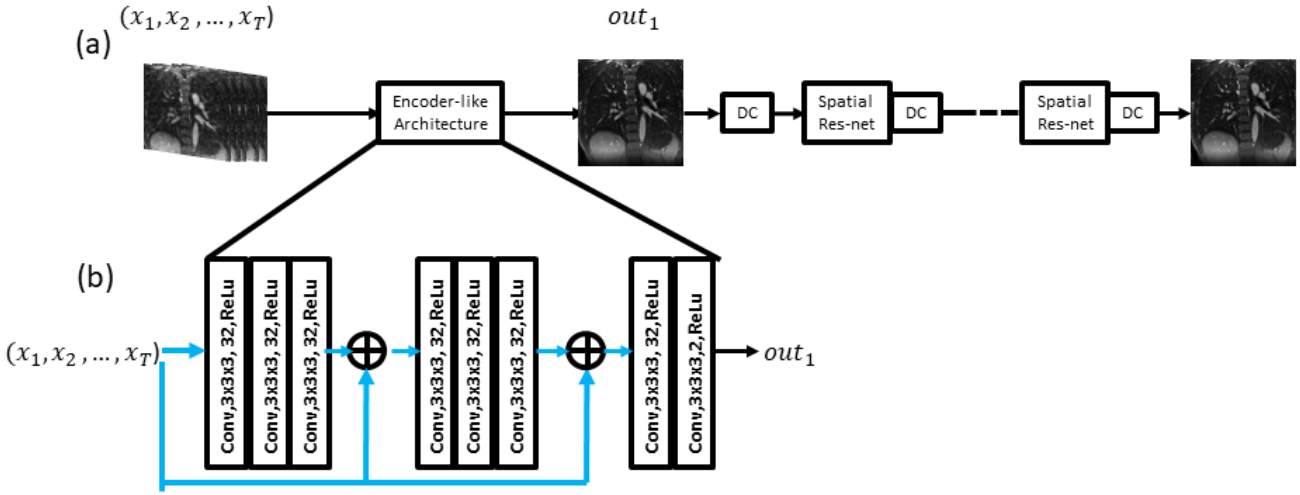


Figure 4.1*: Dynamic MBDL Architecture. An array of (registered or unregistered) gridded images $(x_1, x_2, \dots, x_T)$ is used as input to the architecture **(a)** with the first frame x_1 chosen as the target frame for reconstruction. All data-consistency steps and self-supervised loss are enforced on this target frame. The $target\ frame\ x_1$ is reconstructed from data corresponding to a subset Θ_1 of k-space Ω_1*

*where $\Theta_1 \cup \Lambda_1 = \Omega_1$ and Λ_1 is used solely to enforce self-supervised loss. All other gridded images x_i use all available k-space data corresponding to that frame. This array is passed through the encoder-like architecture **(b)** compressing T frames to 1 frame. This output frame out_1 is then passed to subsequent conjugate gradient data-consistency steps and spatial residual networks (Spatial Res-net) to produce a single reconstructed image corresponding to the target image.*

This encoder-like architecture (**figure 1b**) has several benefits. First, passing from N frames to one frame act as an information bottleneck that forces compression along the temporal dimension. This is like the encoder part of autoencoder architectures that takes in N features as input and solves for a low dimensional representation of these features. Second, this network allows MBDL to leverage correlations across frames while avoiding the memory and computational burden associated with fully 4D MBDL reconstructions as data-consistency steps only need to be applied to a single frame.

In the dynamic MBDL approach, T frames of k-space data vectors with maximum length N are organized into an array Y^i. Under-sampled k-space Ω_1 of frame 1 only is partitioned into disjoint subsets Θ_1 and Λ_1 such that $\Omega_1 = \Theta_1 \cup \Lambda_1$. This first entry in Y^i is the target frame which the self-supervised loss is enforced on during training. The undersampled k-space associated with all other frames is not partitioned. It follows then that $Y^i = [y^i_{\Theta_1}, y^i_{\Omega_2}, \dots, y^i_{\Omega_T}]$. The k-space vectors in Y^i are transformed into gridded images and used as input to the dynamic MBDL architecture. The self-supervised loss for the dynamic MBDL architecture is only slightly modified from [6] to allow for these multi-frame inputs:

$$\min_{\theta} \frac{1}{\mathrm{N}} \sum_{i=1}^{N} \mathcal{L}_k(y^i_{\Lambda_1}, f(Y^i, E^i_{\Theta_{d_1}}; \theta) \qquad [3]$$

Where $\mathcal{L}_k$ is the self-supervised loss in k-space, $y^i_{\Lambda_1}$ is the vector of k-space data associated with Λ_1 from frame 1, and Y^i is as defined above. The loss is enforced between the image space neural network output transformed back to k-space $f(Y^i, E^i_{\Theta_{d_1}}; \theta)$ and $y^i_{\Lambda_1}$ both from frame 1.

4.2.3 Motion Correction Method

Incorporating motion compensation into reconstructions has been previously shown to improve image quality, and I hypothesized that the same would be true for the dynamic MBDL architecture [26], [33]. For motion compensation, I apply a method inspired by [1], [34]. Motion fields are estimated directly as multi-scale low rank (MSLR) components.

Let $\phi_t \in \mathbb{R}^{3xN}$ represent a dense 3-channel deformation field of size N with each voxel assigned a displacement: $\mathrm{Id} + \mathrm{r}(\mathrm{x}, \mathrm{y}, \mathrm{z})$ that warps a given motion state at time t to a reference image. For T frames, deformation fields are stacked into a spatiotemporal matrix $\Phi \in \mathbb{R}^{3xNxT}$. This matrix can be decomposed into the sum of 3-channel rank 1 block-wise matrices across varying block scales. If J is the

number of block scales for the MSLR decomposition then for a given block scale $j \in J$, there are B_j blocks of size $3xN_j\,x\,T$ which are then factored into a block-wise left spatial deformation field bases $\Phi_j \in \mathbb{R}^{3xN_jx1}$ and right temporal bases: $\Psi_j \in \mathbb{R}^{3xTx1}$. The sum of this decomposition across block-sizes for ϕ_t is:

$$\Phi_t = \sum_{j=1}^{J} M_j(\Phi_j \Psi_j^H) \quad [4]$$

Where M_j is the block to 3 channel deformation field operator. This representation can then be incorporated into the classic registration problem:

$$\min_{\substack{\Phi_j,\Psi_j \\ \forall j \in J}} \mathcal{L}\left(I_{ref}, I_t(\phi_t)\right) \quad [5]$$

Where $\mathcal{L}()$ is restricted to be a pair-wise loss in this work, I_{ref} is the selected reference image, I_t is the motion state to be warped, and Φ_t is the motion field such that $\Phi_t = \sum_{j=1}^{J} M_j(\Phi_j \Psi_j^H)$. Frames are optimized using stochastic updates

4.3 Methods

4.3.1 Overview

I apply this motion compensation method to reconstruction of the end-inspiratory phase of free breathing, retrospectively gated 3D pulmonary UTE acquisitions. The overall workflow can be seen in **figure 2**. 3D contrast enhanced UTE acquisitions from a prior study [23] were binned into six respiratory phases based on respiratory belt signal. Unregistered gridded respiratory phase images were used to train a dynamic-MBDL architecture that was used during testing to generate motion resolved reconstructions of both training and test data. Motion fields from all respiratory phases to end-inspiration were then estimated using the motion correction algorithm. These deformation fields were then applied to motion correct both the training and test data to train a motion compensated dynamic-MBDL architecture used ultimately to reconstruct end-inspiratory phase images.

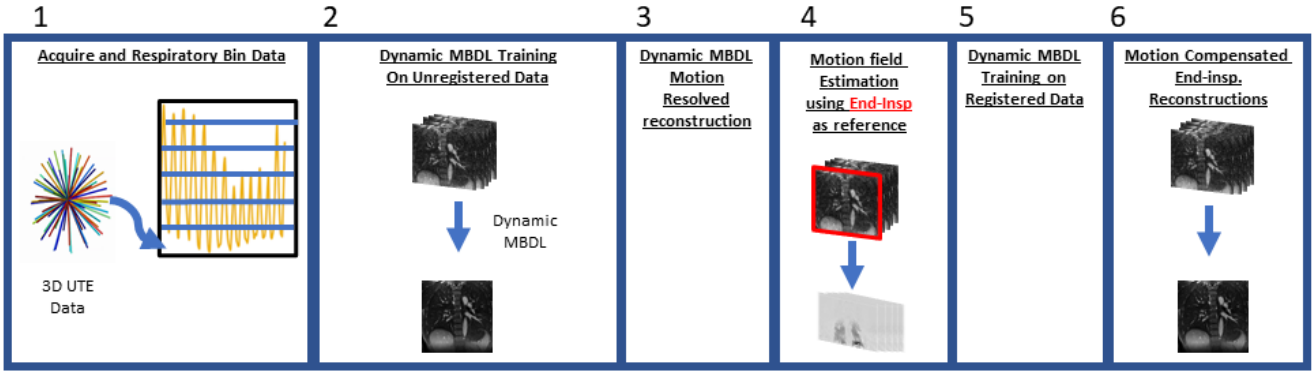


Figure 4.2: *Motion Compensated Workflow.* ***1.*** *3D Pulmonary UTE data is acquired and binned based on respiratory belt signal into N different respiratory phases.* ***2.*** *A dynamic MBDL architecture is then trained on this unregistered gridded data with the end-inspiratory phase chosen as the target frame.* ***3.*** *This trained dynamic MBDL architecture is then used to generate motion resolved reconstructions on both test and training sets.* ***4.*** *Motion fields are then estimated using GPU-based non-rigid registration with the end-inspiratory phase as reference.* ***5.*** *The registered training data is then used to train a motion compensated dynamic MBDL architecture that is then used during testing* ***6.*** *to reconstruct end-inspiratory images from registered test data.*

The implementation of the dynamic MBDL architecture is first described which is identical across both the model trained on unregistered data and the model trained on registered data The overall motion compensation workflow is then discussed in detail.

4.3.2 Dynamic MBDL Implementation

The dynamic MBDL architecture was unrolled for five iterations alternating between residual convolution networks (32 channels/conv, 3D conv with 3 x 3 x 3 kernels, no bias) with ReLU activations and conjugate gradient data-consistency steps with five inner iterations. Conjugate gradient data-consistency is applied similar to that found in [3] with learnable parameter α. The encoder-like network discussed in section 4.2.2 was used in the first unroll while spatial residual networks were used in remaining unrolls. For the encoder-like network, N complex valued respiratory phase images were converted to 2N channel data as input. The output from the data-sharing network was a 2-channel image that was then passed to subsequent data-consistency and spatial residual networks. The complex valued volume output from each data-consistency step was converted to 2-channel data as input for each spatial residual network. The self-supervised loss was implemented as an L2 norm in k-space summed over channels.

Block-wise learning with gradient checkpointing was used for memory efficient dynamic MBDL training. Without this method, reconstruction of these high resolution, volumetric datasets would be difficult even on GPU clusters. For a single unroll, this technique decomposes input volume(s) into patches, checkpoints each patch, iteratively passes these patches through the network, and then recombines the output patches into the full volume for data-consistency. Each input volume was decomposed into eight patches. Gradient checkpointing was also applied to the multi-channel data-consistency step to reduce memory use.

4.3.3 Motion Compensation Workflow

Non-Cartesian Data Acquisition and Retrospective Respiratory Binning

Post-Ferumoxytol (4mg/kg) contrast enhanced, pulmonary magnetic resonance angiography (MRA), ultrashort echo time (UTE) imaging acquired in a previously described study during free breathing in healthy volunteers [23] were used in this study. 11 cases were included for training and

testing with 5 cases used for training, 1 case for validation, and 5 cases for testing. Datasets were acquired with a 3T GE Scanner with a 32-channel coil. Scan parameters were scan time of 5:45minutes, TE=0.25ms, TR=3.6ms, and 1.25mm isotropic resolution. A total of 94,957 projections were acquired using 3D pseudorandom bit-reversed view ordering with readout length of 636 points [15]. Respiratory positions were recorded with a respiratory belt. Data was coil compressed to 20-channels using PCA coil compression [24]. The acquisition provided whole chest coverage with matrix sizes varying between 300-450 x 200-300 x 300-450 based on automatic field of view determination at full resolution. Density compensation was normalized using the max eigenvalue of the NUFFT operator, and k-space was then rescaled based on this value [1]. Acquired data was then sorted into different respiratory motion states similar to XD-Grasp [30]. In brief, the respiratory belt signal was divided into six respiratory phases from end-inspiration to end-expiration. Acquired data that fell within a given bin was then assigned to that respiratory phase.

Dynamic MBDL Motion Resolved Reconstruction

Training: The dynamic MBDL model used for motion resolved reconstructions during inference was trained with the end-inspiratory phase as the target image on unregistered data. Training input was generated as follows:

1. Prior to training, gridded respiratory phase images were generated from all phases except end-inspiration (the target frame) without partitioning k-space. This data does not change over iterations, so it only needs to be computed once.

For each training iteration:

2. End-inspiratory phase k-space data was randomly partitioned along the radial dimension into two disjoint subsets such that $\Omega_{end\ insp} = \Theta_{0.4} \cup \Lambda_{0.6}$ where $\Theta_{0.4}$ represents 40% of the radial spokes, $\Lambda_{0.6}$ represents 60% of the radial spokes. This partition was chosen based on the results from Cartesian data in [6].
3. Gridded end-inspiratory images were generated from the k-space data corresponding to $\Theta_{0.4}$.
4. Images created in step 1 were stacked with the gridded end-inspiratory image into an array and uses as input into MBDL.
5. $\Lambda_{0.6}$ was used only in the self-supervised loss

This architecture was trained in Pytorch (open source, pytorch.org) for 2000 iterations using an Adam optimizer with learning rate of 1e-3 and NUFFTs from SigPy (Open Source, https://github.com/mikgroup/sigpy) on Intel Xeon workstations using one 40 GB A100 GPU.

Testing: During inference, motion resolved reconstructions at full resolution were generated for both training and test data for motion field estimation. Although dynamic MBDL only outputs a single target frame, motion-resolved reconstructions were generated by iteratively rolling the input respiratory binned data along the phase dimension such that the target frame was always the first frame and the phases most similar in motion state were always nearest that frame. To make this concrete, let the end-inspiratory frame be frame 0 and end-expiratory frame be frame 5. To reconstruct end-inspiration, the input was 0,1,2,3,4,5. To reconstruct frame 1, data was loaded as 1,2,3,4,5,0. To reconstruct frame 2, data was loaded as 2,3,4,5,1,0 because motion state 1 is closer to motion state 5 than motion state 0. This pattern continues for frame 3: 3,4,5,2,1,0, and so on. The target frame data during inference was not partitioned into subsets.

Motion field estimation

The end-inspiratory phase was chosen as reference for motion correction as this would eliminate the need to estimate a second set of motion fields that warp registered data back to the motion state for data-consistency. The full resolution motion-resolved images were down-sampled two-fold in each dimension to reduce registration time. The left $\{\Phi_j\}$ components were initialized using gaussian noise, and the right $\{\Psi_j\}$ components were initialized with all zeros. L2 norm was used for the similarity metric with no explicit regularization applied. 30 epochs of the registration algorithm at lower resolution were run with each epoch consisting of stochastic updates over all frames. Once initial lower resolution fields were estimated, a second registration problem was run where the field estimates were refined by interpolating to full resolution and enforcing loss on the full resolution data over 5 epochs. This method was implemented using auto-differentiation in Pytorch with an Adam optimizer with learning rate of .01. Motion fields were estimated for both training and test data.

Motion Compensated Dynamic MBDL

Training: Training was essentially identical to the technique described for the architecture trained on unregistered data for motion resolved reconstructions. The only difference was registered gridded respiratory training data was used in place of unregistered gridded respiratory training data. The end-inspiratory phase was selected as the target frame.

Testing: Registered gridded images from the test data were used as input. Target frame data during inference was not partitioned into subsets.

4.3.4 Evaluation

For image quality assessment, a similar approach was taken to [26]. Apparent signal to noise (aSNR), defined as the signal in a region of interest divided by the standard deviation of signal outside the body, was measured in the aorta, parenchyma, and airway. Contrast to noise ratio (CNR), defined as the contrast difference between selected regions of interest versus the standard deviation of signal outside of the body, was measured between the aorta and airway and parenchyma and airway. Liver edge sharpness was computed by fitting a logistic curve to image intensities along the liver edge and computing the maximum gradient of this curve. All quantitative metrics across reconstructions were compared using paired t-testing. Differences between reconstructions were considered significant if $P < .05$.

The impact of motion correction on reconstructed image quality was investigated by training separate dynamic MBDL architectures on unregistered and registered data. Both architectures were then subsequently tested on unregistered and registered data.

Motion resolved reconstructions generated using dynamic MBDL and XD-Grasp were then compared to investigate whether 1) motion dynamics were similar between the two reconstructions, 2) whether final motion compensated dynamic MBDL reconstruction quality differed between using motion fields estimates derived from motion resolved reconstruction using dynamic MBDL vs. XD-Grasp. XD-Grasp has previously been used as part of motion compensation workflows [26] and was treated as the gold standard. Motion dynamics were compared by manually segmenting end-inspiratory and end-expiratory volumes on test cases and then taking the difference between these measures to compare tidal volumes.

Finally, motion compensated dynamic MBDL end-inspiratory phase image quality and run-time were compared to spatial self-supervised MBDL, XD-Grasp, and iterative motion compensated reconstructions (iMoCo). CG-SENSE was used as baseline. Details on the implementation of these reconstructions can be found in supplement **4.7.1** associated with this chapter.

4.4 Results

4.4.1 Impact of Motion Correction during training and testing on dynamic MBDL Image quality

Figure 4.3 demonstrates sagittal end-inspiratory slices for all dynamic MBDL architectures combinations with training/inference on registered/unregistered data with CG-SENSE as baseline.

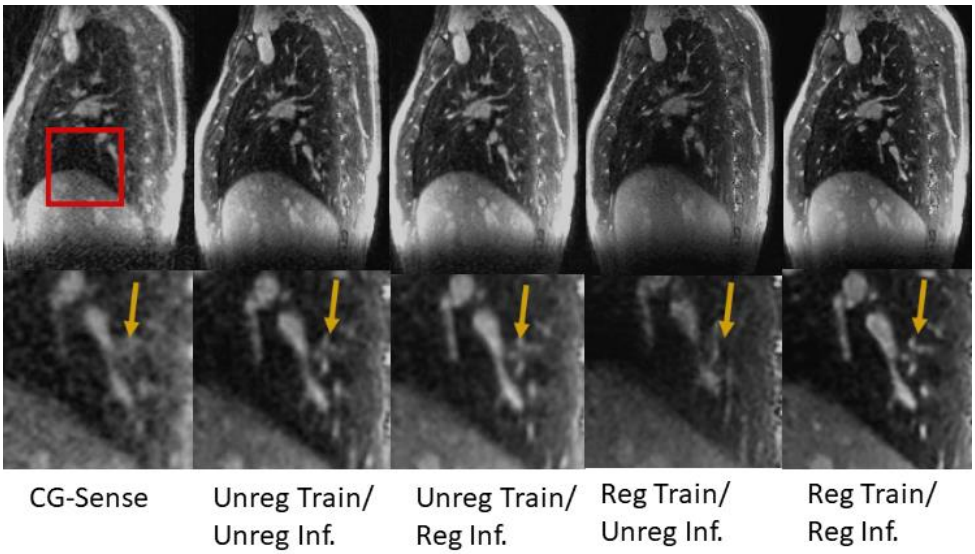


Figure 4.3: *Impact of Registration on Reconstruction Results. Two dynamic MBDL architectures were trained, one on registered data, the other on unregistered data. Reconstructions using registered and unregistered test data were then run on both architectures. Displayed here are sagittal slices from reconstructions on the same case using these different strategies. The red bounding box shows the location where the image has been zoomed in on row 2. Motion compensated dynamic MBDL (Reg Train/Reg Inf.) was significantly sharper and resolved more features (yellow) than all other reconstructions suggesting that motion correction significantly improves reconstruction quality relative to unregistered reconstructions. Motion compensated dynamic MBDL also remains much sharper than the architecture trained on unregistered data, but with inference on registered data (Unreg Train/Reg Inf.). Interestingly, the architecture trained on unregistered data preserved the features when performing inference on unregistered data significantly better than the architecture trained on registered data.*

Motion compensated dynamic MBDL had minimal streaking artifact, and sharply resolved even small vascular features (orange arrow). Image quality was improved over all other dynamic MBDL reconstructions and CG-SENSE. This included dynamic MBDL trained on unregistered data, but with inference on either unregistered data or registered data. Training and inference on registered data clearly improves image quality.

Although dynamic MBDL trained on registered data with inference on unregistered data had similar aSNR (aorta arch: P<.258, parenchyma: P<.356, airway: P<.062) and CNR (aortic arch: P< .2577,

parenchyma: P<.232) to motion compensated dynamic MBDL, many of the features present in the image (orange arrow) are not seen in any of the other reconstructions suggesting motion state was not preserved. All other reconstructions maintained visual alignment of features.

Motion-compensated dynamic MBDL had significantly higher aortic arch, parenchyma, and airway aSNR values than CG-SENSE and dynamic MBDL models trained with unregistered data (Figure 4a, aorta/parenchyma: P<.001 , airway: P<.05). Although airway aSNR should be close to zero, the CNR (Figure 4b, P<1e-3) for the parenchyma and aorta arch (relative to the airway) remained higher for motion-compensated dynamic MBDL than these other reconstructions suggesting that aortic arch, parenchyma, and airway can be better distinguished using this method.

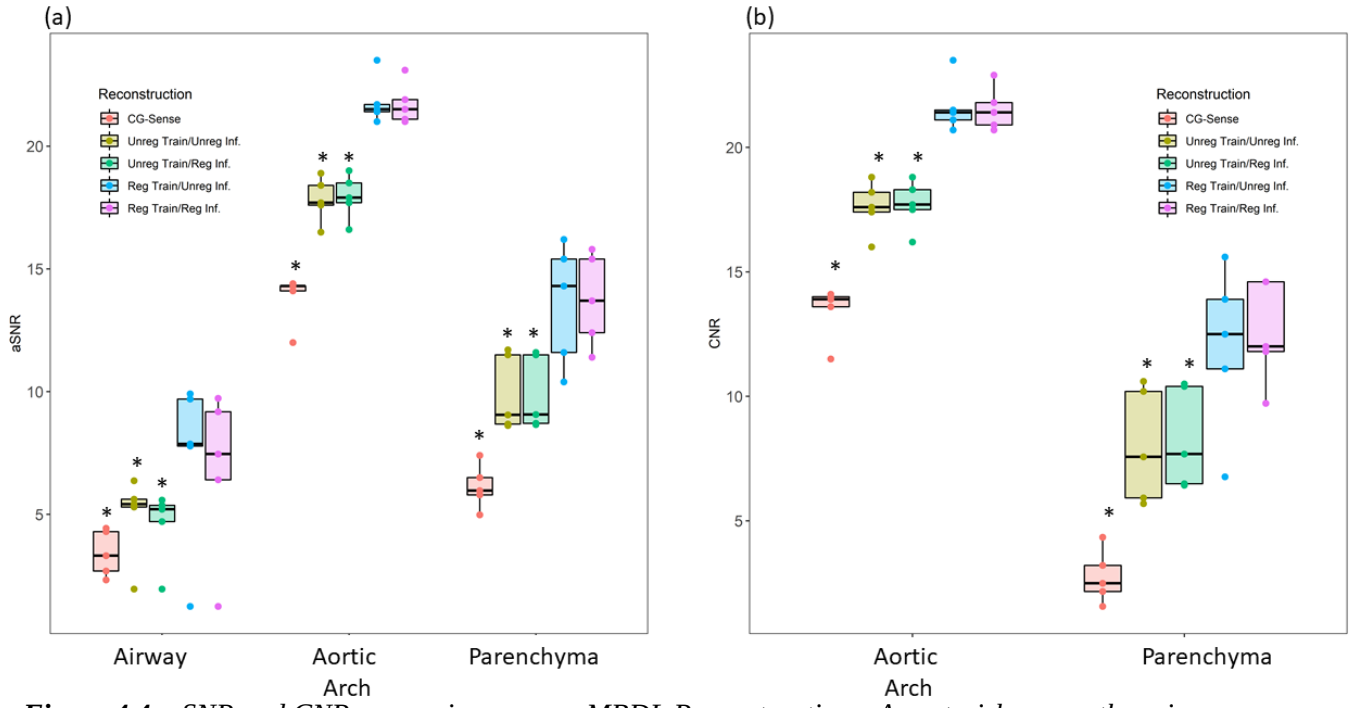


Figure 4.4: *aSNR and CNR comparison across MBDL Reconstructions. An asterisk means there is a significant difference in the metric between that architecture and motion compensated MBDL (Reg Train/Reg Inf). Following (*[26]*), a major airway, the aortic arch, and a section of lung parenchyma were segmented in all test cases. In* ***(a),*** *motion-compensated MBDL. had significantly higher aortic arch and parenchyma aSNR than all other reconstructions besides the reconstruction from the same architecture, but with inference on unregistered data. Airway aSNR should be close to 0. Here, airway aSNR was significantly higher for motion-compensated MBDL than the architecture trained on unregistered data and CG-sense. However, CNR* ***(b)*** *of both the aorta and parenchyma were significantly higher for motion-compensated MBDL suggesting that aorta, parenchyma and airway could be best distinguished in this reconstruction.*

Sharpness of the liver edge was not significantly different for reconstructions on registered data independent of whether dynamic MBDL was trained on registered or unregistered data (**Figure 4.5**).

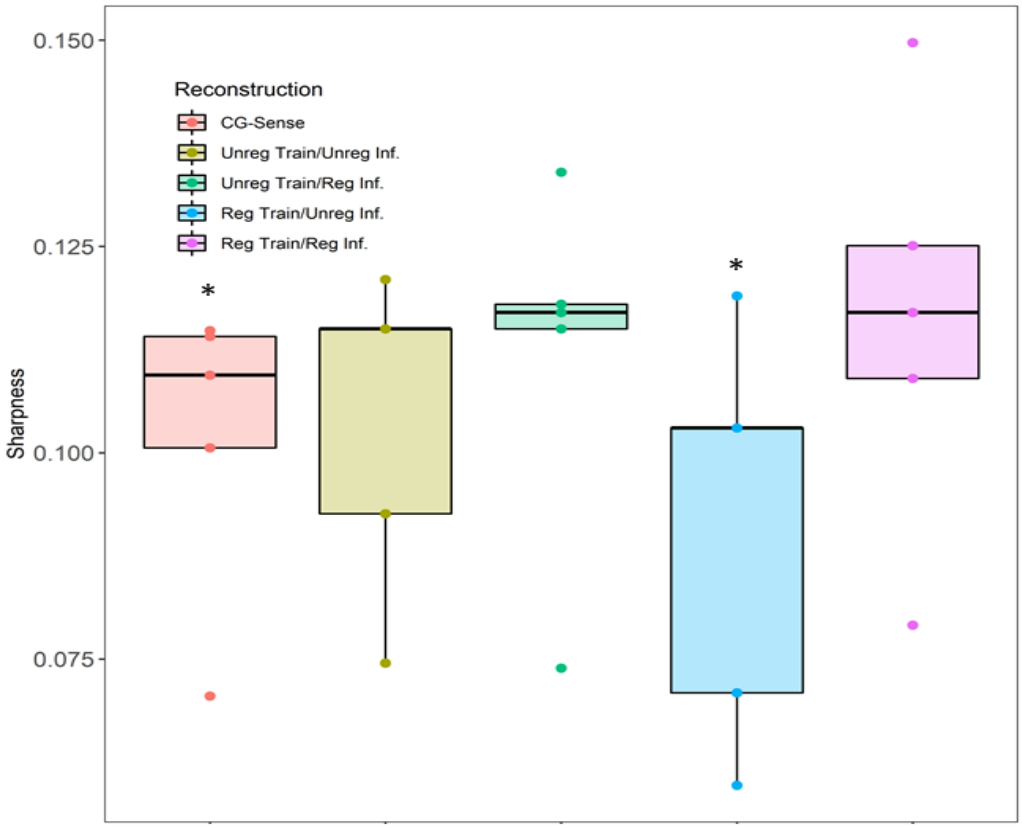


Figure 4.5: *MBDL Reconstructions Liver Edge Sharpness. An asterisk means there is a significant difference in the metric between that architecture and motion compensated MBDL (Reg Train/Reg Inf). Liver edge sharpness was computed by fitting a logistic curve to normalized image intensities starting from the lung down to the liver parenchyma and then taking the maximum gradient of this curve. No statistically significant differences were seen between reconstructions performed on registered data-sets (Reg Train/Reg Inf. and Unreg Train/Reg Inf.). Liver edge was significantly sharper for Reg Train/Reg Inf. than any reconstruction performed on unregistered data.*

4.4.2 Motion Resolved Reconstruction Comparison

Supplemental video 4.1 (https://doi.org/10.6084/m9.figshare.19584067.v1**)** shows XD-grasp and dynamic MBDL motion resolved reconstructions. Dynamics across respiratory phases are similar for both reconstructions. No significant differences in tidal volume were seen between these two reconstructions (P<.3). **Figure 4.6** shows motion compensated reconstructions using motion fields estimated from motion resolved XD-grasp and dynamic MBDL reconstructions respectively.

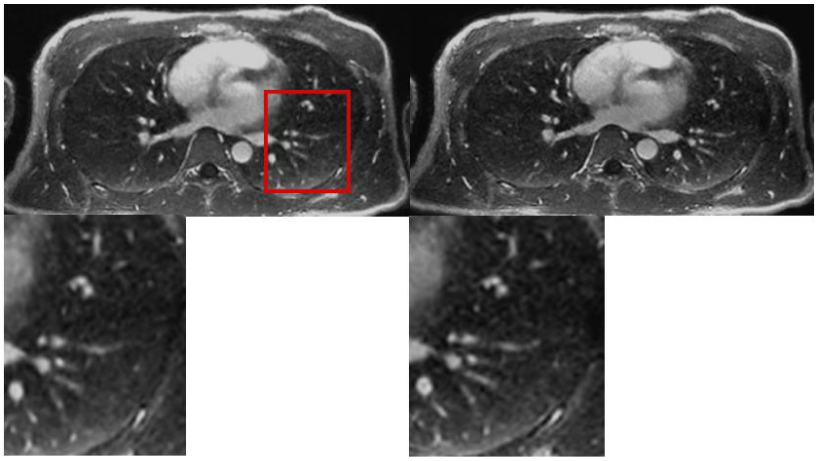


Figure 4.6: *Motion Compensated Dynamic MBDL from Motion fields estimated using Motion Resolved reconstructions from Dynamic MBDL (MBDL MR) vs XD. Grasp (XD-Grasp MR). These representative axial slices demonstrate no significant visual quality differences between the two reconstruction strategies.*

No difference in image quality between these reconstructions was observed visually or quantitatively. and based on aSNR (aortic arc: P<.223, parenchyma: P<.066, airway: P<.365) , CNR (aortic arch: P<.153,parenchyma: P<.171) and liver edge sharpness (P<.47). These metrics can be found in **figure 4.9** and **4.10** below.

4.4.3 Motion Compensated Dynamic MBDL Image quality Comparison

Figure 4.7 show coronal slices for motion compensated dynamic MBDL, iMoCo, XD-Grasp, Spatial Self-supervised MBDL, and CG-sense reconstructions.

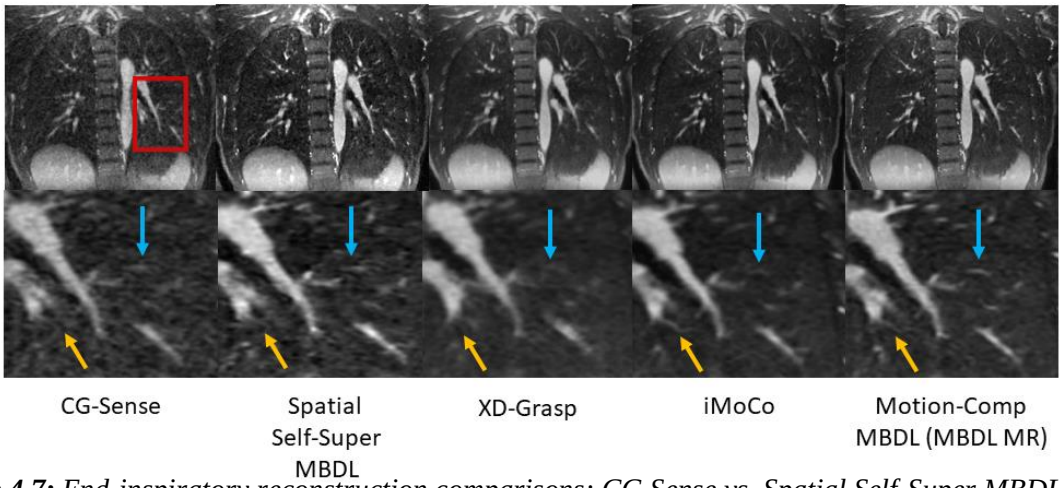


Figure 4.7: *End-inspiratory reconstruction comparisons: CG Sense vs. Spatial Self-Super MBDL vs XD-grasp vs iMoCo vs. Motion compensated Dynamic MBDL. Displayed here are representative coronal slices from different reconstructions on the same case. The red bounding box shows the location where the image has been zoomed in on row 2. Motion compensated Dynamic MBDL was sharper than all other reconstructions including spatial self-supervised dynamic MBDL which had significant remaining undersampling artifact. iMoCo and motion compensated Dynamic MBDL and iMoCo both resolved small vascular features that could not clearly be seen in the other reconstructions. iMoCo resolved some of these features (yellow arrow) more clearly than Motion compensated Dynamic MBDL. The reverse was also true (blue arrow).*

From **figure 4.7**, motion compensated dynamic MBDL has clearly improved image quality over XD-Grasp, spatial self-supervised MBDL, and CG-SENSE with improved ability to resolve small vascular features. Relative to iMoCo, motion compensated dynamic MBDL does have sharper features; however, there are some features (orange arrow) that are better resolved with iMoCo. **Figure 4.8** show coronal maximum intensity projections taken over 30 slices for these same reconstruction methods

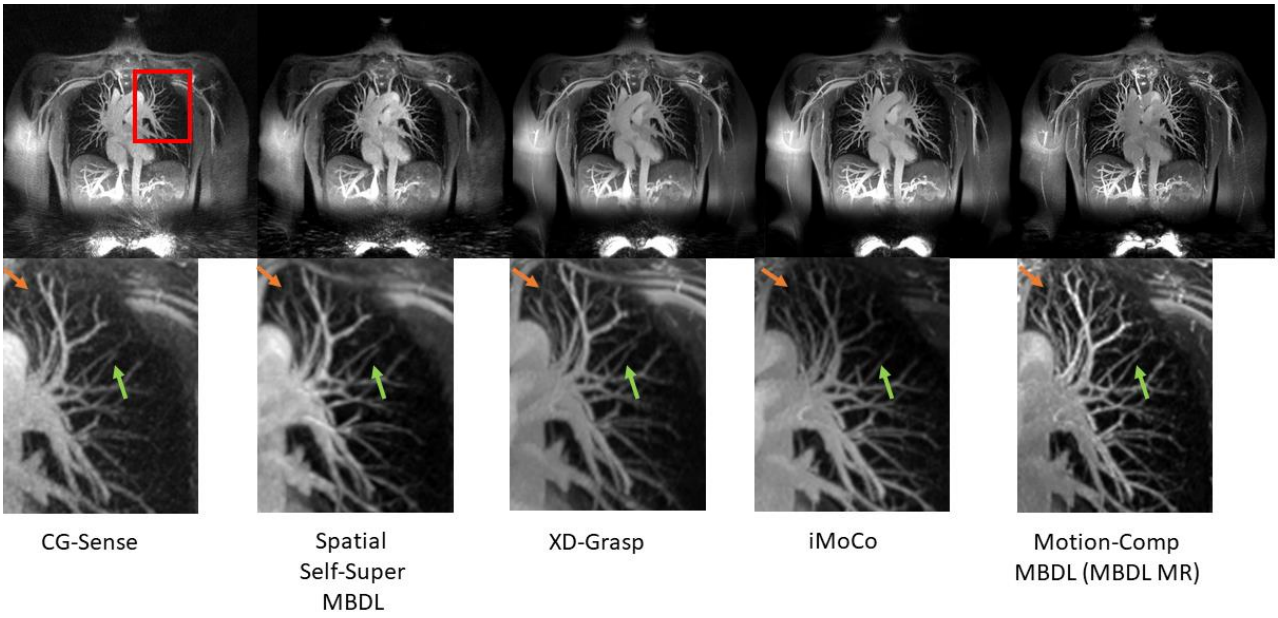


Figure 4.8: *Maximum Intensity Projection comparisons: CG Sense vs. Spatial Self-Super MBDL vs XD-grasp vs iMoCo vs. Motion compensated Dynamic MBDL. Maximum intensity projections of 30 slices in the AP direction centered around the lung hilum were generated from varying reconstructions on a single case. iMoCo smoothly resolves vascular features compared to motion compensated Dynamic MBDL. Motion compensated Dynamic MBDL, however, was sharper and resolved more subtle vascular features (orange arrow, yellow arrow) than all other reconstructions. Note only motion compensated dynamic MBDL with motion fields estimated from dynamic MBDL motion resolved reconstructions (MBDL MR) is shown here.*

In the MIP images in **figure 4.8**, motion compensated dynamic MBDL clearly captures more vascular structure than all other reconstructions including iMoCo.

Motion compensated dynamic MBDL had significantly higher aorta and parenchyma aSNR (P<.001) than all other reconstructions. Motion compensated dynamic MBDL airway aSNR was significantly higher than CG-SENSE (P<.05), however, it did not differ significantly from spatial self-supervised learning (P<.26), iMoCo (P<.07) or XD-grasp (P<.434) **(Figure 4.9a).** Both aorta (P<.01) and parenchymal CNR (P<.01) were significantly higher for motion compensated dynamic MBDL than all other reconstructions **(Figure 4.9b).**

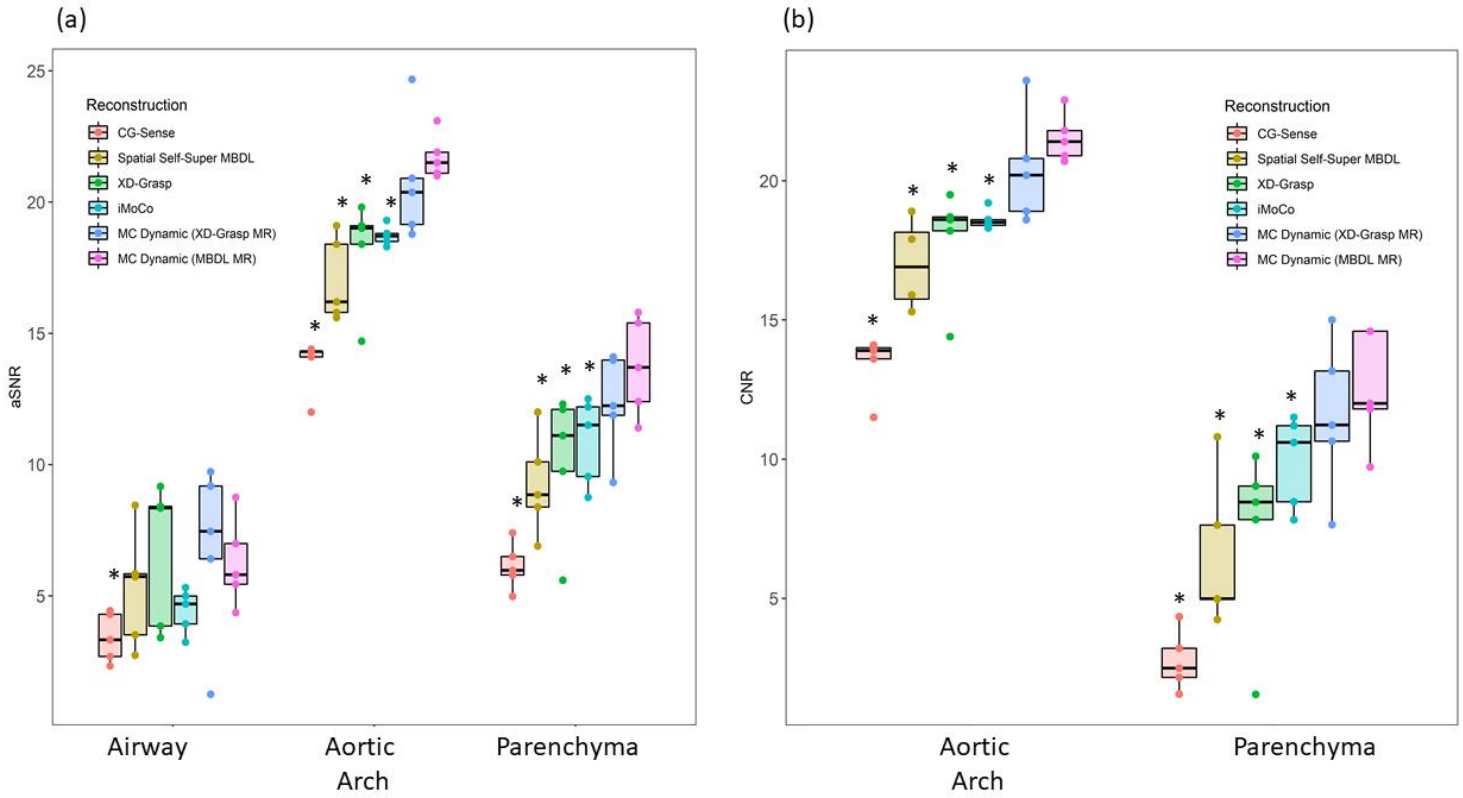


Figure 4.9: *aSNR and CNR comparison across CG Sense vs. Spatial Self-Super MBDL vs XD-grasp vs iMoCo vs. motion compensated dynamic MBDL from motion fields estimated using XD-Grasp motion resolved reconstructions (**XD-Grasp MR**) vs. motion compensated dynamic MBDL from motion fields estimated using dynamic MBDL motion resolved reconstructions (**MBDL MR**). An asterisk means there is a significant difference in the metric between a given reconstruction approach and motion compensated dynamic MBDL (**MBDL MR**). Motion compensated dynamic MBDL **(MBDL MR)** had significantly higher aortic arch and parenchymal aSNR as well as CNR than all other reconstructions.*

Liver edge sharpness **(figure 4.10)** for reconstructions using registered data (motion compensated dynamic MBDL /iMoCo) did not significantly differ.

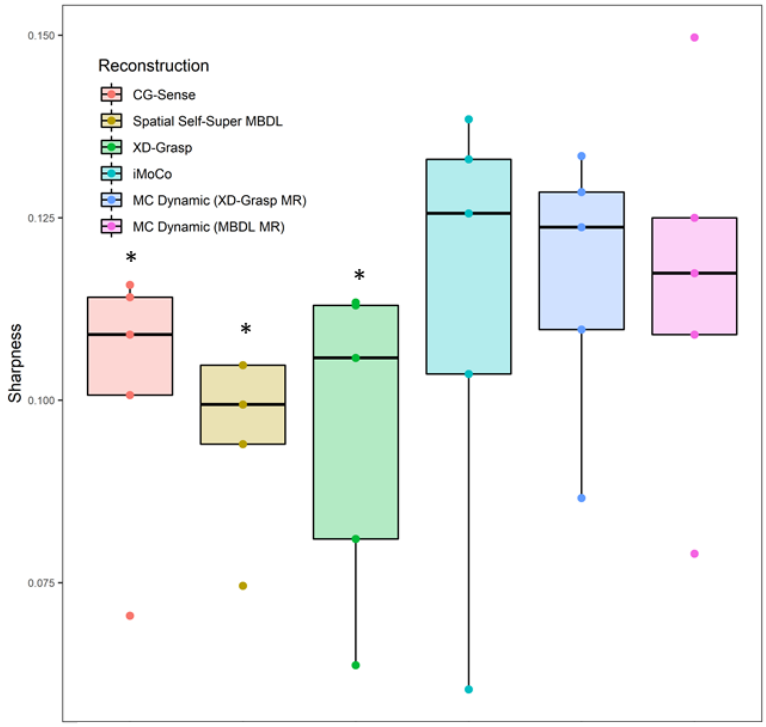


Figure 4.10: *Liver edge sharpness comparison across CG Sense vs. Spatial Self-Super MBDL vs XD-grasp vs iMoCo vs. motion compensated models with motion fields estimated using different motion resolved reconstruction strategies. An asterisk means there is a significant difference in the metric between that architecture and motion compensated MBDL (Reg Train/Reg Inf). Techniques that ran reconstructions on registered data had higher liver sharpness than techniques that used either unregistered data or a single image.*

Run times for implementations of the various reconstruction methods can be found in **supplement section 4.7.2**.

4.5 Discussion

In this work, I combine dynamic MBDL, a self-supervised MBDL method that efficiently leverages correlation across frames with a GPU based registration technique to develop a motion compensated DL framework. This method was applied to reconstruction of highly undersampled, end-inspiratory images from respiratory binned, free breathing, 3D Pulmonary UTE acquisitions. This technique (during inference) consists of a respiratory binned acquisition, motion resolved reconstruction using dynamic MBDL trained on unregistered data, estimation of motion fields from the motion resolved reconstruction, and a final motion compensated dynamic MBDL step. I first showed that incorporating

motion compensation into dynamic MBDL improved the quality of deep learning reconstruction (**figure 4.3**). I then demonstrated that the motion compensated method resulted in higher quality images than spatial self-supervised deep learning, XD-Grasp, and iMoCo (**figure 4.7** and **figure 4.8**) while significantly reducing reconstruction time relative to both XD-grasp and iMoCo (run-times:17, 32, and 62 minutes respectively, see supplement).

The fundamental goal of this work was to address the difficulty obtaining fully sampled 3D non-Cartesian data for supervised training of MBDL architectures. Although spatial self-supervised MBDL does allow for reconstruction of these datasets, reconstruction quality was not comparable to state-of-the-art methods like iMoCo. This can be seen in **figure 4.3** where there is drop-out of small vascular features and undersampling artifact. The key distinction between the spatial self-supervised MBDL and iMoCo approaches that drives differences in reconstruction quality is the amount of data used during reconstruction. iMoCo leverages all data acquired during the scan (95,000 spokes) while spatial self-supervised MBDL uses only a single frame often with less than five thousand spokes. Spatial self-supervised MBDL then is data-starved relative to iMoCo.

The dynamic MBDL architecture proposed here addresses the data starvation seen in spatial self-supervised MBDL by leveraging correlations across frames. Like XD-grasp, dynamic MBDL with training and inference on unregistered data has motion blur and is not close to the image quality of iMoCo. It is the combination of motion correction with dynamic MBDL through increased correlation across frames that significantly boosts image quality.

Like iMoCo, motion compensated DL reconstruction quality is dependent on high quality motion resolved reconstructions that preserve dynamics in addition to accurate motion field estimation. Although the dynamic MBDL motion resolved reconstructions had comparable dynamics to XD-grasp, the dynamic MBDL motion resolved method proposed here was likely more a test of model generalization then a method to be followed in the future. This is because this model was trained with self-supervised loss and data-consistency enforced only on the end-inspiratory phase. This is likely suboptimal for reconstructing all motion states as network weights in the encoder-like network were tuned toward outputs close to end-inspiration meaning motion states closer to end-expiration may not be accurately preserved.

A better technique could be to randomly choose a respiratory phase each training iteration to enforce self-supervised loss on so that network weights learn to account for all motion states. Another approach might be to develop fully 4D MBDL architectures for motion resolved reconstructions, however, this is computationally challenging due the higher GPU memory requirements and associated lengthy training time.

A potential problem with using DL in general though for motion resolved reconstruction is the architecture's ability to appropriately model dynamics during inference is dependent on the similarity between motion dynamics in the test and training data. The training and test data used in this work was from healthy volunteers with periodic breathing where the number of radial spokes binned from end-inspiration to end-expiration increased monotonically. Respiratory patterns can be highly irregular, particularly in patients with diffuse lung disease, and each phase may be assigned widely varying numbers of spokes. It is an open question how well DL methods would generalize to such challenging datasets.

There are several limitations then to the present study that warrant further investigation. First and foremost, the image quality evaluation was limited by lack of a ground truth, as is the case for any self-supervised method. As in past applications of deep learning, there is potential for the neural network to create images which do not represent underlying anatomy. A primary focus of work moving forward should be on evaluating reconstructions on both greater numbers of patients and patients with a wider variety of respiratory dynamics particularly in clinical cases where gold standard images may be available e.g. CT. Second, the architecture was likely undertrained due to the limited number of training iterations used to keep training times reasonable. Additionally, parameters such as the number of unrolls, optimal θ and Λ splitting ratios, and the network architecture were not investigated at this stage. Finally, more work is needed to investigate the generalization of this technique to other dynamic applications, including 4D-Flow and Dynamic Contrast Enhanced MRI.

4.6 Conclusions

In this study, I developed a motion-compensated self-supervised MBDL reconstruction method that combines motion estimation with an MBDL architecture that leverages correlations across frames. I demonstrate on healthy volunteers that this approach allows for fast and high-quality 3D pulmonary UTE reconstructions.

4.7 Supporting Information

4.7.1 Reconstruction Training Methods

Spatial Self-Supervised MBDL: Spatial self-supervised MBDL was trained on end-inspiratory phase data using the same training data used for dynamic MBDL. To train the network, end inspiratory phase k-space data Ω_{end_insp} was divided into two disjoint subsets along the spoke dimension such that subset Θ was randomly assigned 40% of the projections while subset Λ was randomly assigned 60% of the projections.K-space subset Θ was used to reconstructed gridded input images to spatial self-supervised

MBDL while k-space subset Λ was used solely in the self-supervised loss. The architecture was trained for 2000 iterations with an Adam optimizer with learning rate of 1e-3.

XD-Grasp: XD-grasp reconstructions were implemented by modifying the code available at (https://github.com/mikgroup/extreme_mri/blob/master/motion_resolved_recon.py) to accept previously binned data as input. Reconstructions were run for 200 iterations with temporal differences regularization weight of 1e-6.

iMoCo: iMoCo reconstructions were performed similar to [26] using modified code from (https://github.com/PulmonaryMRI/imoco_recon) to allow bins with varying number of spokes. The ANTS CPU based image registration used in [26] was replaced with the GPU-based motion correction method used in this work to reduce computation time. As iMOCO reconstructs a template image data-consistent with all respiratory phases, I aligned this template image with the end-inspiratory phase for comparison to Dynamic MBDL. Eighteen iterations of iMoCo were run.

CG-Sense: Thirty iterations of CG sense on the end-inspiratory phase data was performed.

4.7.2: Run Time Comparisons:

iMOCO:

Total Run-Time: 62 minutes with XD-Grasp run at full-resolution

Run-Time Breakdown

a. Motion Resolved Reconstruction full res: 32 min

b. GPU based Motion Correction: 3 minutes

c. Final motion compensated iterative reconstruction: 29 minutes

XD-Grasp: 32 minutes at full resolution

Motion Compensated Dynamic MBDL

Total Run-time: 17 minutes

Run-Time Breakdown:

a. Dynamic MBDL Motion Resolved Reconstruction full res: 14 min

b. GPU based Motion Correction: 3 minutes

c. Final motion compensated iterative reconstruction: 1 minute

Run-Time during Training/Testing:

Forward pass: ~60 seconds

Backward pass: 205 seconds

Inference: ~60 seconds

Spatial Self-Supervised MBDL:

Run-Time during Training/Testing:

Forward pass: ~60 seconds

Backward pass: 202 seconds

Inference: ~60 seconds

Chapter 5: Motion Compensated High Spatiotemporal Resolution MRI[1]

In chapter 5, I move away from deep learning reconstruction approaches. In this work, I tackle the challenge of integrating motion compensation into high spatiotemporal reconstruction.

5.1 Introduction

In recent years, significant work has gone towards development of free-breathing, high spatiotemporal resolution 4D acquisitions [36],[2]. These acquisitions combined with robust reconstruction methods have the potential to reduce the challenge of imaging pediatric [28] and neonatal subjects and allow patients with severe cardiopulmonary disease to feel more comfortable during scanning by removing the need for breath-holds ([37]. These methods can also provide improved spatiotemporal resolution for dynamic contrast-enhanced acquisitions with implications for the visualization and quantification of functional measures of hemodynamics and contrast uptake. In addition, these methods provide significant advantages for thoracic imaging, where motion corruption is common and difficult to manage [26] .

These dynamic acquisitions are often acquired using non-Cartesian methods with pseudorandom view ordering. One of the benefits of this approach is that acquired data can be flexibly re-binned after the acquisition. This allows reconstructions across multiple dimensions in order to, for instance, resolve respiratory and cardiac motion. These binning methods are often performed using surrogate motion signals derived using respiratory belts, pilot tone modulation, or center of k-space based navigators [1] [38] [30]. Using the motion surrogate, k-space data is typically binned prior to image reconstruction into a small number of motion states with the assumption these motion states recur periodically through the acquisition. In acquisitions with irregular respiratory or bulk motion, however, reconstruction performance using these binning techniques can be significantly degraded due to artifact from intraframe motion [1]

One approach to solving this problem is to bin data through time with sufficient temporal resolution (e.g. for respiratory motion ~500ms) to reduce intraframe motion. Reconstructing such data, however, is challenging due to the extreme degree of undersampling of individual frames and sheer amount of data generated by binning at sub-second intervals in minutes long scans. For smaller scale problems (i.e. lower spatiotemporal resolution), techniques that leverage correlations across frames via nuclear norm minimization are often used to reconstruct highly undersampled data [32]. However, with increased matrix size and frame count, nuclear norm minimization quickly become infeasible with respect to memory and computation time [1] .

[1] in collaboration with Luis Torres

Ong et al [1] proposed a way to overcome this memory and computational bottleneck by directly optimizing for a highly compressed multi-scale low rank (MSLR) representation of the 4D time series. This method, –dubbed "Extreme MRI", is not only able to capture irregular and bulk motion in free breathing high spatiotemporal ultrashort echo time (UTE) pulmonary and DCE MRI acquisitions [28], but is able to further reduce the rank of the data set directly in the compressed space using the variational definition of the nuclear norm -[10], [39].

Like all low rank methods though, Extreme MRI is dependent on correlations across frames. Bulk and irregular motion disrupts these correlations and erodes image quality. I hypothesized that incorporating motion compensation into Extreme MRI would improve image quality as it improves these correlations. This hypothesis is supported by a large body of work showing that incorporating motion compensation into reconstruction significantly improves reconstruction quality [26], [33]

Much of this work, however, relies on motion field estimation through retrospective registration of low-resolution navigator images. This is problematic if the initial low-resolution reconstruction is unable to capture all motion dynamics. In the case of Extreme MRI at high temporal resolution (<500ms per frame), the accurate reconstruction of low resolution images themselves is challenging due to high levels of undersampling. Furthermore, many of these motion correction algorithms operate on relatively small-scale problems where memory constraints are less of a concern. For the scale of the problems Extreme MRI is attempting to reconstruct, use of dense motion fields can easily triple the memory footprint of reconstruction.

In recent work, Huttinga Et al. [34] have overcome these constraints by developing memory efficient methods to estimate motion fields directly from k-space data binned through time. This method warps a reference image-template according to loss enforced in k-space, and directly solves for a cubic B-spline parameterization of low rank representations of the motion fields. Using this method, Huttinga Et al. can recover respiratory motion up to 100ms temporal resolution. As they use a k-space representation of motion fields relative to one static frame, they do not need prior dynamic reconstructions to accurately model motion.

In this work, motivated by the developments in [34] and [1], I integrate a memory efficient representation of the motion fields estimated in k-space with Extreme MRI reconstructions. My proposed method involves estimating low resolution motion fields directly as multi-scale low rank components by enforcing k-space loss between a warped template image and acquired k-space data, interpolating these fields directly in the compressed space, updating these higher resolution fields through k-space based loss, and then integrating these high resolution fields into Extreme MRI. I apply this Motion Corrected MSLR

technique (MoCo-MSLR) to 3D free breathing radial acquisitions and compare MoCo-MSLR to Extreme MRI reconstructions at temporal resolutions required to resolve respiratory (500ms) and cardiac dynamics (100ms).

5.2 Theory

5.2.1 Extreme MRI: Multi-scale Low Rank Reconstruction Review

The MSLR model ([1], [40] stacks a time series with T frames and image size N into a spatiotemporal matrix $\boldsymbol{X}$ of size T x N. This spatiotemporal matrix is then represented as the sum of rank 1 block-wise matrices across varying block size scales. If J is the number of block scales for the MSLR decomposition then for a given block scale $j \in J$, there are Bl_j blocks of size $N_j \, x \, T$ which are then factored into a block-wise left spatial basis $L_j \in \mathbb{C}^{N_j \, x \, 1}$ and a right temporal basis $R_j \in \mathbb{C}^{Tx1}$. The sum of this decomposition across block-sizes for a frame X_t is:

$$X_t = \sum_{j=1}^{J} \boldsymbol{B_j}(L_j R_{j,t}^H) \quad \textbf{(5.1)}$$

Where $\boldsymbol{B_j}$ is a block-to-image operator.

The forward model for the reconstruction problem then with acquired multi-channel k-space data stacked into a matrix $Y \in \mathbb{C}^{CM \, x \, T}$ where C is the number of coils, M is the number of measurements and T is number of frames is:

$$Y_t = \mathcal{A}\left(\sum_{j=1}^{J} \boldsymbol{B_j}\left(L_j R_{j,t}^H\right)\right) \quad \textbf{(5.2)}$$

Where $\mathcal{A}$ is a linear operator incorporating sensitivity maps and the non-uniform fast Fourier transform operator. To regularize the problem, Ong et al. applies block-wise low rank constraints by using the variational form of nuclear norm minimization:

$$\min_{X=\sum_{j=1}^{J} M_j (L_j R_j)} \|X\|_* = \sum_{j=1}^{J} \left(\left\|L_j\right\|_F^2 + \left\|R_j\right\|_F^2\right) \quad \textbf{(5.3)}$$

This formulation allows for block-wise rank reduction directly in the compressed space significantly reducing the memory and computational requirements associated with computing the nuclear norm. The full MSLR reconstruction objective function to be minimized is:

$$f(L,R) = \frac{1}{2}\left\|Y - \mathcal{A}\left(\sum_{j=1}^{J} \boldsymbol{B_j}\left(L_j R_j^H\right)\right)\right\|^2 + \frac{\lambda_j}{2}\sum_{j=1}^{J}\left(\left\|L_j\right\|_F^2 + \left\|R_j\right\|_F^2\right) \quad \textbf{(5.4)}$$

To further reduce reconstruction run-time , stochastic optimization is used to solve for the right and left vectors, taking gradient steps frame by frame rather than averaging across all frames.

In this formulation, the MSLR factorization attempts to models all dynamics including motion and contrast change contained in the time series. The greater the complexity of dynamics contained in this decomposition, the higher the rank must be to appropriately model these dynamics. As the decomposition intrinsically constrains rank, complex dynamics that cannot be modeled in this setting can be lost resulting in artifacts, blurring, and/or misrepresentation of the dynamics. Irregular respiratory and bulk motion is particularly challenging to model as it is usually associated with high rank.

5.2.2 MoCo-MSLR Reconstruction

Let **forward** motion fields be defined as warps from a fixed template image to a given motion state and **adjoint** motion fields be warps from a given motion state back to the image template. Here I develop a multi-resolution reconstruction scheme that first solves for forward and adjoint low resolution motion fields and interpolates these motion fields to the desired resolution all as MSLR components. These interpolated motion fields at the desired resolution can then further refined through k-space based template warping. These fields are then used in a final motion-compensated Extreme MRI reconstruction.

Low Resolution Forward Motion Field Formulation

Where applicable I follow the notation introduced in the MSLR reconstruction review above. Let acquired k-space data be stacked into a matrix $Y \in \mathbb{C}^{CM\,x\,T}$. I model a bin of this time series in k-space as

$$Y_t = \mathcal{A}(I_{temp}(\Omega_{for,t})) \textbf{ (5.5)}$$

where I_{temp} is a template image, $\Omega_{for,t} \in \mathbb{R}^{3XN}$ represent 3 channel dense deformation fields of size N with each voxel assigned a displacement: $\mathrm{Id} + \mathrm{r}(\mathrm{x}, \mathrm{y}, \mathrm{z})$ that warp the template image to a given motion state at time t. $\mathcal{A}$ is an operator that transforms this warped template image into k-space.

To both regularize the problem and fit data on the GPU, I represent the deformation fields $\Omega_{for,t}$ in a MSLR representation. Let $\boldsymbol{\Omega_{for}} \in \mathbb{R}^{3xTxN}$ be the spatiotemporal matrix of the stacked three channel deformation fields over T frames. I decompose exactly as in [1] where:

$$\Omega_{for} = \sum_{j=1}^{J} \boldsymbol{B_j}(\Phi_{j,for}\Psi_{j,for}^{H}) \textbf{ (5.6)}$$

Where $\Phi_{j,for} \in \mathbb{R}^{3xNx1}, \Psi_{j,for}^{H} \in \mathbb{R}^{3xTx1}$ and $\boldsymbol{B_j}$ is the corresponding blocking operator.

Deformation fields are smoothed spatially using total variation regularization to allow for improved sliding motion at organ boundaries commonly found between the lung and chest wall [41]. Although the MSLR representation significantly regularizes the deformation fields along the time dimension there is still potential for under-sampling artifact to propagate into the fields leading to high frequency oscillations through time in the image. To help mitigate this issue, I minimize block-wise rank of the MSLR

deformation fields via the variational formulation of the nuclear norm. The regularization applied to the deformation field components at time t is:

$$f_{reg,t} = \sum_{j=1}^{J} \frac{\lambda_j}{2} \left(\frac{1}{T} \left\|\Phi_{j,for}\right\|_F^2 + \left\|D\Psi_{j,for}^H\right\|_F^2\right) + \gamma \left\|D\Omega_{for,t}\right\| \quad \textbf{(5.7)}$$

Note that the regularization on $\Psi_{j,for}^H$ enforces temporal smoothness through the finite difference operator D over time frames. The finite difference operator is also applied to compute approximate spatial gradients $\left(\frac{d\Omega}{dx}, \frac{d\Omega}{dy}, \frac{d\Omega}{dz}\right)$ for total variation spatial smoothing of the deformation fields. The deformations fields are solved stochastically as in Ong et al [1]. The complete objective function then to solve for forward motion fields at time t in the MSLR basis is

$$\underset{\substack{\Phi_{j,for}, \Psi_{j,for}^H \\ \forall\, j \in J}}{\operatorname{argmin}} \left\|Y_t - \mathcal{A}(I_{temp}(\Omega_{for,t}))\right\| + \sum_{j=1}^{J} \frac{\lambda_j}{2} \left(\frac{1}{T} \left\|\Phi_{j,for}\right\|_F^2 + \left\|D\Psi_{j,for}^H\right\|_F^2\right) + \gamma \left\|D\Omega_{for,t}\right\| \quad \textbf{(5.8)}$$

Low Resolution Adjoint Motion Field Formulation

After solving for the forward motion fields, I solve for the adjoint motion fields that relate a motion state at time t back to the template image. Forward motion fields are fixed and then applied to warp the chosen template frame to the motion state at time t. The MSLR representation of the adjoint deformation fields is then estimated by learning to warp this motion state back to the template. The algorithm then is:

for iterations

1. Randomly select time point $t \in \{t_1, t_2, \ldots, t_T\}$
2. Forward warp I_{temp} to this motion state $I_{temp}(\Omega_{for,t})$
3. Optimize $\underset{\substack{\Phi_{j,adj}, \Psi_{j,adj}^H \\ \forall j \in J}}{\operatorname{argmin}} \left\|I_{temp} - I_{temp}(\Omega_{adj,t}(\Omega_{for,t}))\right\|^2 + \sum_{j=1}^{J} \frac{\lambda_j}{2} \left(\left\|\Phi_{j,adj}\right\|_F^2 + \left\|D\Psi_{j,adj}^H\right\|_F^2\right) + \gamma \left\|D\Omega_{adj,t}\right\|$

MSLR Interpolation

I then interpolate the MSLR representation of the low resolution forward and adjoint deformation fields to the desired resolution used for the final reconstruction. I first initialize $\Phi_{j,desired\ res}$ and $\Psi_{j,desired\ res}$ for the forward and adjoint fields that warp the time series at the desired resolution. The algorithm then is as follows

for iterations:

1. Randomly select time point $t \in \{t_1, t_2, \ldots, t_T\}$

2. Interpolate $\Omega_{low\ res,t} = \sum_{j=1}^{J} B_j\left(\Phi_{j,low\ res}\Psi_{j,low\ res}^H\right)$ to $\Omega_{desired\ res,t} = \sum_{j=1}^{J} B_j\left(\Phi_{j,desired\ res}\Psi_{j,desired\ res}^H\right)$ by applying a cubic B-spline interpolation operator
3. Optimize $\underset{\substack{\Phi_{j,desired\ res},\Psi_{j,desired\ res} \\ \forall j \in J}}{\operatorname{argmin}} \left\|\Omega_{desired\ res,t} - \Omega_{low\ res,t}\right\|^2$

The interpolated motion fields at the desired resolution can then be further refined by the same k-space based motion field estimation introduced earlier.

Motion Compensated Extreme MRI

I then integrate the MSLR representation of the forward and adjoint motion fields that warp the time series at the desired resolution into Extreme MRI.

$$\min_{L_j,R_j\ \forall j \in J} \left\|Y_t - \mathcal{A}[I_t(\Omega_{for,t}]\right\|^2 + \sum_{j=1}^{J} \frac{\lambda_i}{2}\left(\frac{1}{T}\left\|L_j\right\|_F^2 + \left\|R_j\right\|_F^2\right) \textbf{(5.9)}$$

Where $I_t = \sum_{j=1}^{J} M_j\left(L_j R^H{}_{j,t}\right)$ and $\Omega_{for,t} = \sum_{j=1}^{J} M_j(\Phi_{j,t}\Psi_{j,t}^H)$

The algorithm using stochastic gradient descent proceeds as follows:

Initialize $\left\{L_j\right\}_{j=1}^{J}$ and $\left\{R_j\right\}_{j=1}^{J}$ as in [1] then

for iterations:

1. Randomly choose a time frame t and reconstruct its image: $I_t = \sum_{j=1}^{J} M_j\left(L_j R_{j,t}\right)$, and associated forward and adjoint fields: $\Omega_{for,t} = \sum_{j=1}^{J} M_j\left(\Phi_{for,j,t}\Psi_{for,j,t}^H\right)$, $\Omega_{adj,t} = \sum_{j=1}^{J} M_j(\Phi_{adj,j,t}\Psi_{adj,j,t}^H)$. I_t should be aligned with all other time frames.
2. Warp this image to its appropriate motion state: $I(\Omega_{for,t})$
3. Take the gradients of the data-consistency term with respect to $\{L_j\}$ and$\{R_{j,t}\}$. By the chain rule first take the gradient of the data-consistency term: DC_{grad} with respect to $I_t(\Omega_{for,t})$, warp this gradient back to the aligned space using the adjoint deformation field: $DC_{grad}(\Omega_{adj,t})$, and finally take the gradient with respect to $\{L_j\}$ and$\{R_{j,t}\}$.
4. Take the gradients of $f_{reg} = \sum_{j=1}^{J} \frac{\lambda_i}{2}\left(\left\|L_j\right\|_F^2 + \left\|R_{j,t}\right\|_F^2\right)$ with respect to $\{L_j\}$ and$\{R_{j,t}\}$.
5. Update L and R as follows: $L_j = L_j - \alpha T[\nabla_{\mathrm{L_j}} f_{reg} - \nabla_{L_j}(DC_{grad}\left(\Omega_{adj}\right)]$ and $R_{j,t} = R_{j,t} - \alpha[\nabla_{\mathrm{R_j}} f_{reg} - \nabla_{R_j}(DC_{grad}\left(\Omega_{adj}\right)]$

5.3.Methods

I applied MoCo-MSLR to free breathing 3D radial imaging acquisitions in the lung and placenta from previously acquired datasets. Lung data was acquired in one healthy volunteer and 2 patients with diffuse lung disease [cystic fibrosis (CF) and idiopathic pulmonary fibrosis (IPF)]. Placental data was acquired in one healthy pregnant patient in the third trimester. All subjects were asked to breath normally during the acquisition. For all subjects, I performed reconstructions at ~500ms to resolve respiratory motion. For subjects with sufficient contrast between the ventricular wall and blood (healthy volunteer and CF case), I performed a second reconstruction at ~100ms to resolve both cardiac and respiratory motion.

5.3.1 Reconstruction Implementation

K-space data was coil compressed to 20 channels if greater than 20 channels were used during acquisition, otherwise data was not coil compressed. Similar to [1], the 3D radial data used an oversampled field of view (FOV) and was adjusted automatically to include all areas producing MRI signal. Signal outside the reconstructed FOV can lead to artifacts from data-inconsistencies between the acquired k-space data and the NUFFT transformed image data. Further, modeling motion that falls in and out of the FOV is difficult and leads to non-topology preserving deformation fields. To counter this, I followed the steps in [1] by reconstructing a gridded image at twice the prescribed FOV, thresholding the image at 0.1 of the maximum amplitude to estimate the FOV. Density compensation was used to improve convergence. Sensitivity maps were estimated using J-sense from all data binned together [25]. For the motion correction steps that require k-space data and the final MSLR reconstruction, k-space data was binned in time with number of projections per bin determined by dividing the total number of projections by the number of required frames for reconstruction.

Low resolution template images (~3.5 mm isotropic) were reconstructed by running an Extreme MRI reconstruction with all projections binned together. The reconstruction was run for 200 iterations to ensure data-consistency. Block sizes of [8,16,32] with regularization weight of 1e-8 were used across all cases, however, these choices do not substantially impact the template reconstruction as only a single frame was reconstructed.

Spatial deformation field bases $\{\Phi_j\}_{j=1}^{J}$ were initialized using Gaussian noise and temporal deformation field bases $\{\Psi_j^H\}_{j=1}^{J}$ were initialized with all 0s.

In place of explicitly computing gradients for the low-resolution motion estimation and interpolation steps, I used auto differentiation in Pytorch using an Adam optimizer. For low resolution steps, a learning rate of .01 across all block scale was chosen. For interpolation, a learning rate of .001

across all block scales was chosen. To fit the spatial deformation field bases used in the full resolution reconstruction with matrix size $P_x \, x \, P_y \, x \, P_z$ on the GPU, I created blocks corresponding to a matrix of control points of size $\frac{Px}{3} \, x \frac{Py}{3} \, x \frac{Pz}{3}$ that was then trilinearly interpolated to the full deformation field size during reconstruction.

The final motion compensated reconstruction used the code found at https://github.com/mikgroup/extreme_mri as a foundation. This code was modified to allow for forward and adjoint warping of time frames. For all MoCO-MSLR reconstructions I represented the time series using 2 block scales with sizes [64,128] to allow the reconstructions to fit on the GPU.

For all MoCO-MSLR reconstructions, I compared image quality and motion dynamics against Extreme MRI. For all Extreme MRI reconstruction, three block scales with block-sizes of [32,64,128] with regularization weight of 1e-8 were used. These reconstructions were run for 60 iterations. For reconstructions with targeted temporal resolution ~500ms, respiratory dynamics was tracked by fixing a volumetric window about the liver-lung interface, and then auto-correlating this fixed window with a sliding window through time. For reconstructions with targeted temporal resolution near ~100ms, both cardiac and respiratory dynamics were tracked if the motion was resolved on visual inspection of CINEs. Cardiac dynamics was tracked by fixing a volumetric window about the left ventricle, autocorrelating as above, Fourier transforming this signal, and then filtering the signal in a .05 hz pass band about the presumed cardiac cycle rate.

Respiratory dynamics was tracked as above and then gaussian smoothed using $\sigma = 3$ pixels in Scipy.

5.3.2 Healthy Volunteer 1

One healthy volunteer UTE lung dataset [23] was acquired with a 32 channel coil, scan time of 5 minutes and 45 seconds, TE=0.25ms, TR=3.6ms, flip angle=24° and 1.25mm isotropic resolution, Ferumoxytol (4mg/kg) was given prior to the scan. The number of projections was 94,957 with 636 readout length acquired using 3D pseudorandom bit-reversed view ordering. Two reconstructions were performed. The first reconstruction targeted a spatial and temporal resolution of 1.25mm isotropic and 690ms with the goal of resolving respiratory motion. The second reconstruction targeted a spatial and temporal resolution of 1.67mm isotropic and 115ms respectively with the goal of resolving both cardiac and respiratory motion.

5.3.3 Cystic Fibrosis Patient

One UTE lung dataset of a cystic fibrosis (CF) patient was acquired with an 8-channel coil array, an overall scan time of 4 minutes 18 seconds, TE=80μs, TR=3.48ms, flip angle 4 degrees and 1.25 mm

isotropic resolution. The number of projections was 75,768 and 654 readout length. This dataset is publicly available and was included in the original Extreme MRI work ([1]). Two reconstructions were performed. The first reconstruction targeted a spatial and temporal resolution of 1.25mm isotropic and 515ms respectively with the goal of resolving respiratory motion. The second reconstruction targeted a spatial and temporal resolution of 1.67 mm isotropic and 83ms temporal respectively with the goal of resolving both cardiac and respiratory motion.

5.3.4 IPF Patient

One UTE lung dataset of a patient with idiopathic pulmonary fibrosis (IPF) was acquired with an 8-channel coil array, an overall scan time of 4 minutes 54 seconds, TE=80μs, TR=3.27ms, flip angle 4 degrees and 1.25 mm isotropic resolution. The number of projections was 89964 and 654 samples per projection. One reconstruction was performed. The targeted spatial and temporal resolution for this reconstruction was 1.25mm isotropic and 588ms respectively with the goal of resolving respiratory motion.

5.3.5 Third Trimester Pregnant Patient

One placental dataset of a healthy pregnant patient in the third trimester was acquired with GE Air Coil, an overall scan time of 4 minutes, 2 seconds, TE=1.3ms,TR=5.0ms, flip angle of 25 degrees, 1mm isotropic resolution. One reconstruction was performed. The targeted spatial and temporal resolution for this reconstruction was 1.8 mm isotropic and 605ms with the goal of resolving respiratory motion.

The healthy volunteer and CF datasets were acquired on a 3 Tesla GE scanner. The IPF and placental datasets acquired on a 1.5 Tesla GE scanner

5.4 Results

Figure 5.1 shows extracted respiratory signals for ~500 ms reconstructions across all cases.

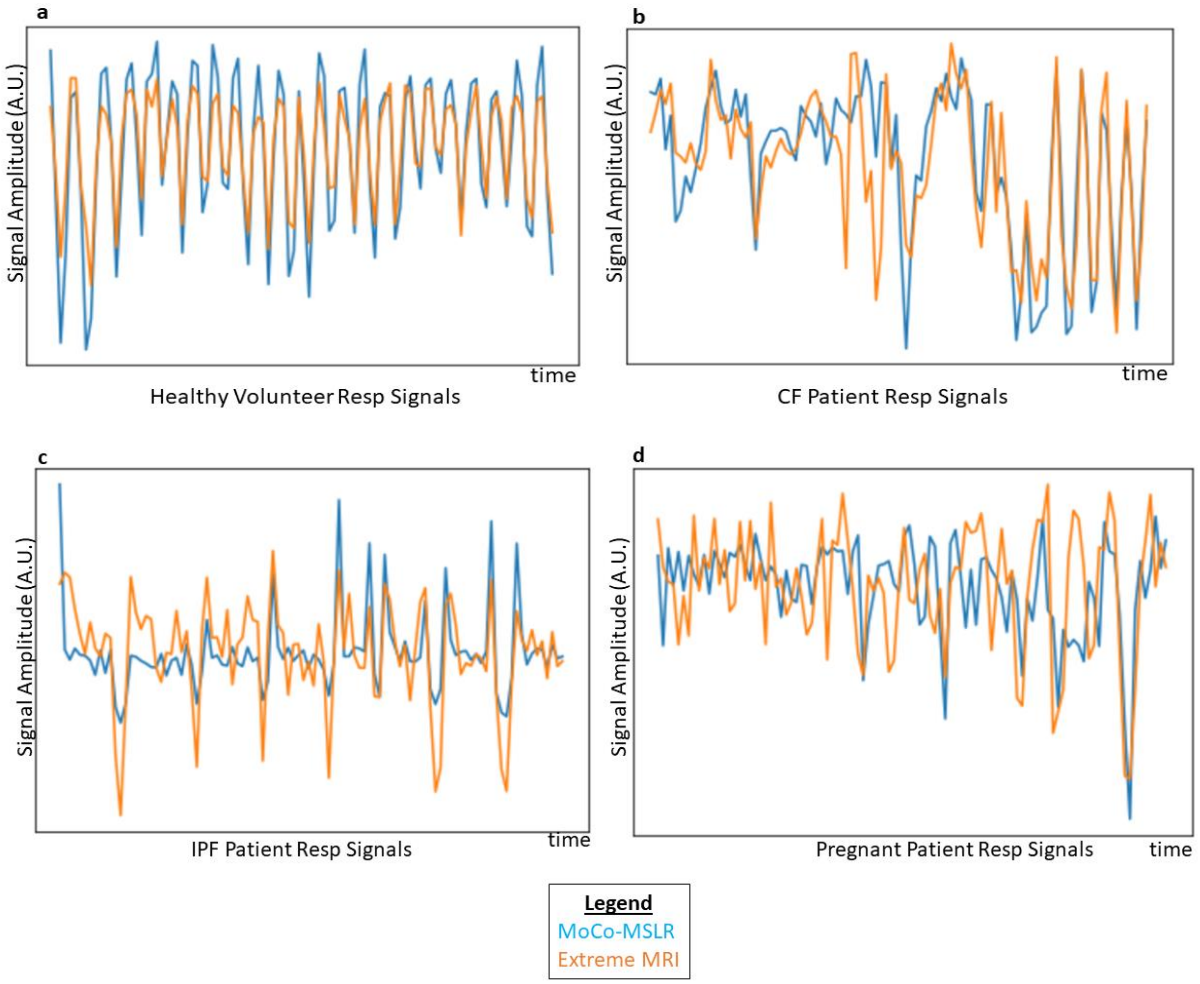


***Figure 5.1**: Respiratory Signal Tracking. For reconstructions near 500 ms that visualized the diaphragm, a fixed volumetric window was placed on the right hemidiaphragm and autocorrelated with a sliding window through time. The respiratory dynamics in the healthy volunteer is nearly periodic (a). Both MoCo-MSLR and Extreme-MRI are in phase. The respiratory dynamics in the CF patient were much more variable (b). In general though, MoCo-MSLR and Extreme-MRI are roughly in phase. In the IPF patient, respiratory dynamics between MoCo-MSLR and Extreme-MRI are generally in phase (c). In the pregnant patient, a fixed volumetric window was placed on the edge between the uterine wall and placenta and autocorrelated with a sliding window through time. Although respiratory dynamics are a little harder to extract here, overall, both MoCo-MSLR and Extreme-MRI remain roughly in phase.*

5.4.1 Healthy Volunteer Dataset

Figure 2 and **supplemental Video 5.1** (https://doi.org/10.6084/m9.figshare.19583887.v2) compare MoCo-MSLR versus Extreme MRI for the reconstruction targeting 690ms temporal resolution.

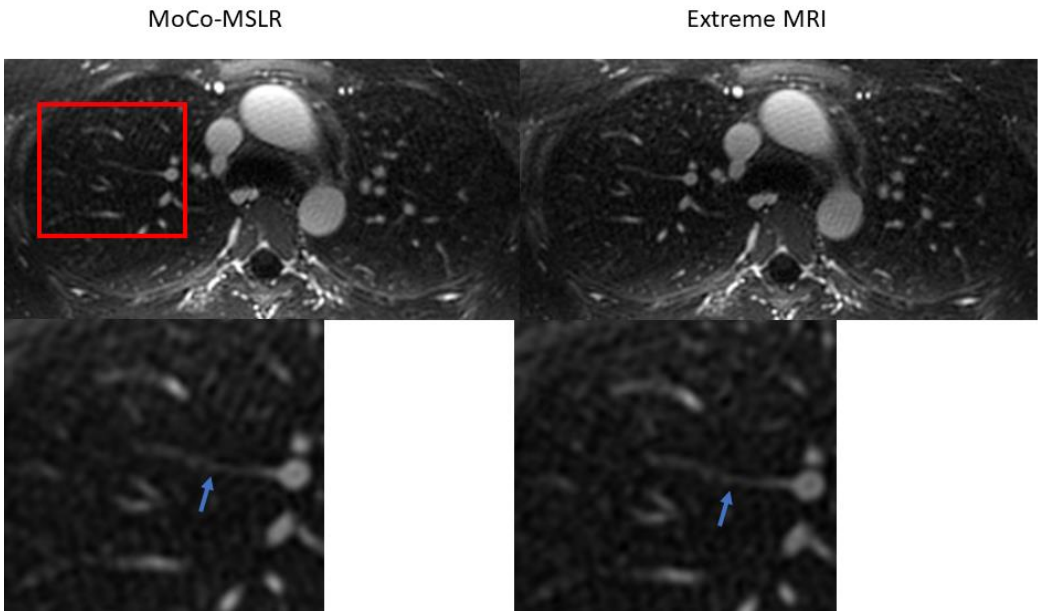


Figure 5.2: *Reconstruction Results on Healthy Volunteer. Displayed here are representative axial slices from MoCo-MSLR (left) and Extreme MRI (right) reconstructions with targeted temporal resolution: 690 ms and spatial resolution: 1.25 mm isotropic. The red bounding box represents the portion of the image zoomed in on row 2. In this healthy volunteer with nearly periodic respiratory motion, no significant differences in image quality can be seen. Both reconstructions resolve small vascular features equally well (blue arrow)*

Image quality is similar between the reconstruction methods with minimal flickering artifact; however, the liver edge appears sharper for MoCo-MSLR during motion **(supplemental video 1).** Vascular structures are resolved similarly by both methods **(figure 5.2, blue arrow, row 2).** Both reconstructions resolve similar motion dynamics as seen from the video and the extracted respiratory signal **(figure 5.1a).**

Figure 5.3 and **supplemental Video 5.2 (https://doi.org/10.6084/m9.figshare.19583914.v1**) compare MoCo-MSLR versus Extreme MRI for the reconstruction targeting 115ms temporal resolution. From **supplemental video 2**, MoCo-MSLR resolves cardiac and respiratory dynamics. Respiratory dynamics and some degree of left ventricular wall motion are resolved by Extreme MRI. Significant

blurring though at both the diaphragm and left lateral ventricular wall is observed. MoCo-MSLR shows limited blurring of these structures. Similar findings can be seen in **figure 5.3a and 5.3b**.

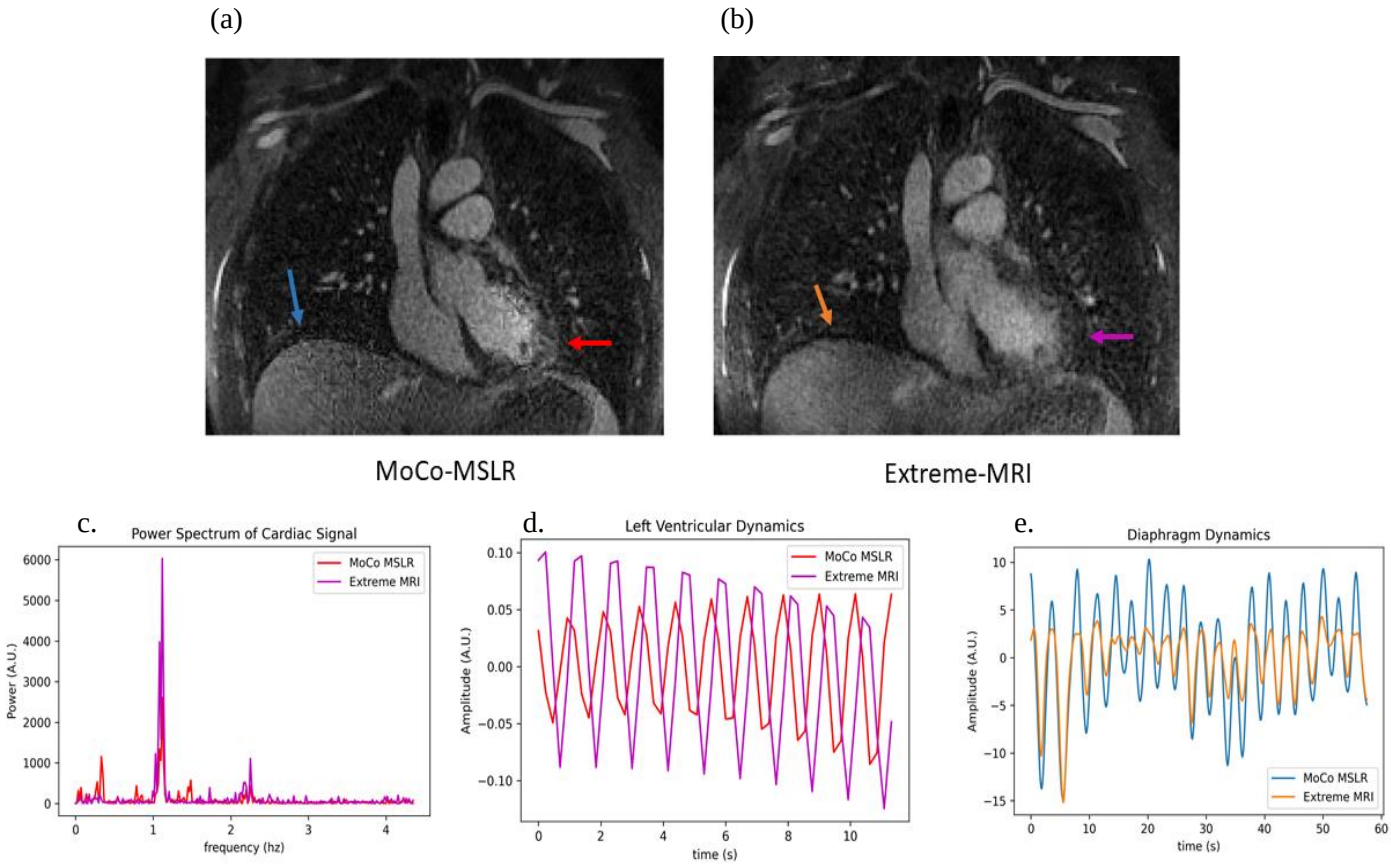


Figure 5.3: *Cardiac and Respiratory Dynamics at High temporal resolution. MoCo-MSLR and Extreme MRI reconstructions were run on the healthy volunteer at a targeted temporal resolution of 115ms and spatial resolution of 1.67 mm isotropic. Two volumetric windows were fixed about the lateral left ventricular wall (red/purple arrows) and the right hemidiaphragm (blue/orange arrows), and autocorrelated with a sliding window at the same spatial location through time to extract cardiac and respiratory dynamics respectively. The power spectrum (c) of the autocorrelation about the lateral left ventricular wall was then computed demonstrating a strong frequency peak around 1.11 hz corresponding to a physiologically reasonable 68 beats per minute. Cardiac signal (d) was then extracted by filtering a .05 hz passband around the peak signal in frequency space. Both MoCo-MSLR and Extreme MRI cardiac signals maintain the same phase relationship through time. The autocorrelation around the right hemidiaphragm was Gaussian smoothed to show respiratory dynamics (e). Both reconstructions remain in the same respiratory phase through multiple respiratory cycles. It is important to note from supplemental video X that the cardiac motion resolved in MoCo-MSLR is more realistic than that resolved by Extreme MRI. Evidence for this can be seen comparing the sharpness of (a) and (b) about the diaphragm (blue/orange arrow) and lateral left ventricle (red/purple arrow). In both locations, MoCo-MSLR is significantly sharper than Extreme MRI.*

Although the cardiac dynamics in **supplemental video 5.2** in the MoCo-MSLR reconstruction appear much more realistic than in Extreme MRI, both methods demonstrate strong peaks in their Fourier power spectra at 1.11 hz corresponding to a heart rate of 68 beats/min (**figure 5.3c**). Filtering this signal in a small passband around this frequency results in signals that resemble cardiac waveforms (**figure 5.3d)**. Diaphragm dynamics **(figure 5.3e)** also appear to be in phase.

Supplemental Video 5.3 (https://doi.org/10.6084/m9.figshare.19583932.v3) demonstrates axial, 2 chamber, 4 chamber, and short axis views of heart for the MoCo-MSLR reconstruction. Multiple cardiac phases in all views are clearly captured. **Figure 5.4,row 1** demonstrates left ventricular phases from late diastole to systole for the healthy volunteer (MRA)

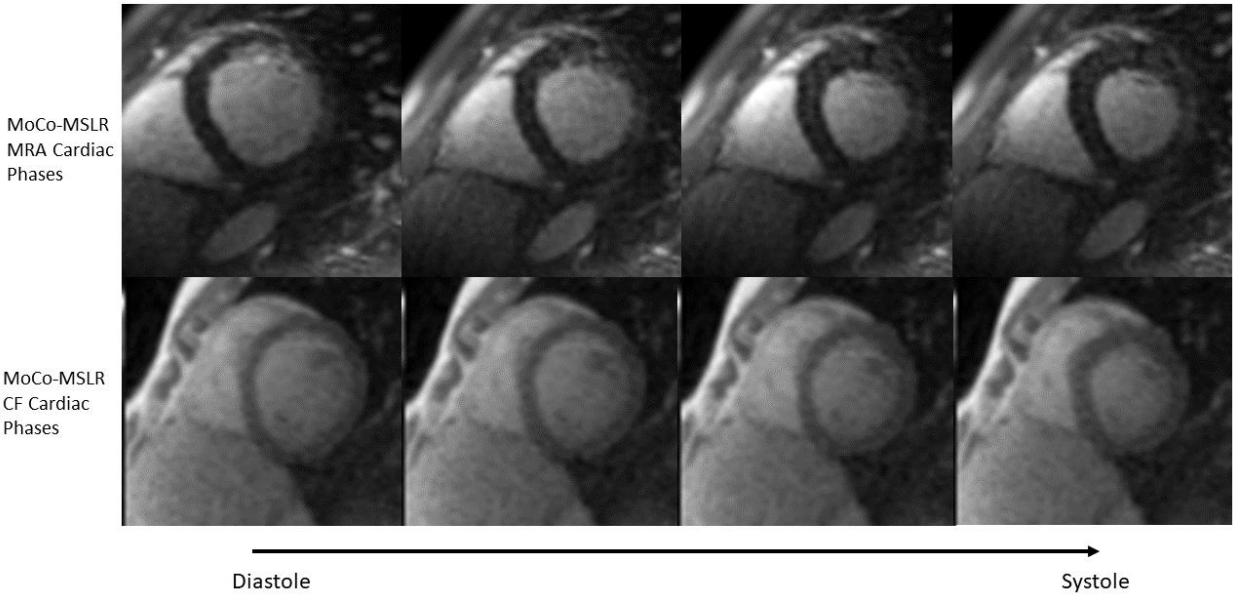


Figure 5.4: *Short axis Cardiac Phases. Cardiac dynamics from mid/late diastole through systole are shown from MoCo-MSLR on the healthy volunteer (targeted temporal resolution: 115 ms) and patient with cystic fibrosis (targeted temporal resolution: 83 ms)*

5.4.2 Cystic Fibrosis Lung Dataset

Figure 5.5 and **supplemental Video 5.4** (https://doi.org/10.6084/m9.figshare.19583938.v1) compares MoCo-MSLR versus Extreme MRI for the reconstruction targeting 515ms temporal resolution.

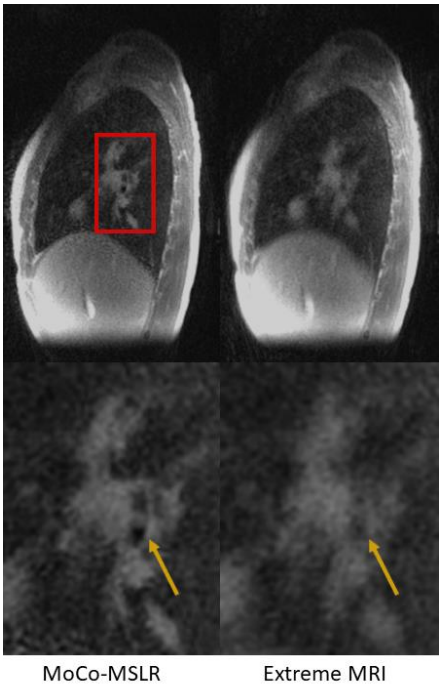


Figure 5.5: *Reconstructions Results on Patient with Cystic Fibrosis. Displayed here are representative sagittal slices from both reconstructions (targeted temporal resolution: 515 ms, spatial resolution: 1.25 mm isotropic). In the zoomed-out images in row 1, MoCo-MSLR sharply resolves the liver edge and larger airway structures compared to Extreme MRI. This can be seen even more clearly in the zoom-in images on row 2 (orange arrow).*

Figure 5. 5 shows that the MoCo-MSLR is significantly sharper than Extreme MRI demonstrating airway feature blurred out in Extreme MRI (yellow arrow). Similar findings are seen in **supplemental video 5.4** where significant blurring of the liver edge and small vascular structures are seen in the Extreme MRI reconstruction. These structures remain sharp for MoCo-MSLR. From the extracted respiratory signal alone **(figure 5.1b),** motion dynamics are similar. However, bulk motion and tracheal collapse seen in the MoCo-MSLR reconstruction are not observed in the Extreme-MRI reconstruction (**supplemental video 5.4)**.

Figure 5.6 and **Supplemental Video 5.5 (https://doi.org/10.6084/m9.figshare.19583944.v2)** compare MoCo-MSLR versus Extreme MRI for the reconstruction targeting 83ms temporal resolution. MoCo-MSLR does resolve cardiac and respiratory dynamics, however, high frequency oscillations through time are present. Further, significant flickering artifact is observed. No obvious left ventricular

wall motion is seen in the Extreme MRI reconstruction. Some small motions at the diaphragm are seen, however this is partly obscured by blur.

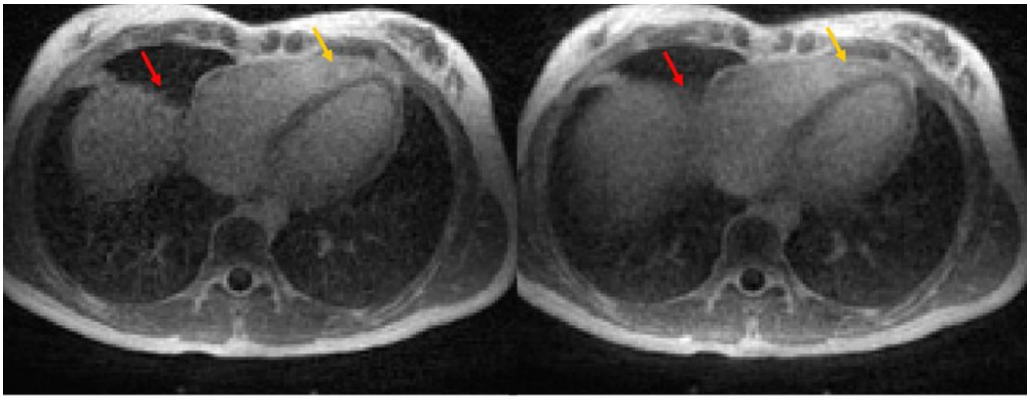


Figure 5.6*: Reconstruction Results on Patient with Cystic Fibrosis at high temporal resolution. Displayed here are representative axial slices from both reconstructions (targeted temporal resolution: 83ms, spatial resolution: 1.67 mm isotropic). Moco-MSLR is sharper particularly around structures that should be in motion like the liver due to respiratory motion (red arrow) and heart. Note that some subtle non-physiologic warping over the heart (yellow) can be seen in the MoCo-MSLR reconstruction.*

Figure 5. 6 shows that the MoCo-MSLR reconstruction has reduced blur around the heart relative to Extreme MRI (red arrow). However, some non-physiologic warping can be seen in the MoCo-MSLR reconstruction near the anterior part of the cardiac septum (yellow arrow). Comparisons between the dynamics for these two reconstructions were not performed as no cardiac dynamics and only subtle diaphragm motion was seen in Extreme MRI. **Supplemental Video 5.6 (https://doi.org/10.6084/m9.figshare.19583950.v2)** is a 15 frame CINE of axial, 2 chamber, 4 chamber, and short axis views of the heart again demonstrating realistic cardiac dynamics in all views. High frequency oscillations can clearly be seen. **Figure 5.4** (**row 2**) demonstrates left ventricular phases from late diastole to systole for the CF patient.

5.4.3 Idiopathic Pulmonary Fibrosis Dataset

Figure 5.7 and **supplemental Video 5.7(https://doi.org/10.6084/m9.figshare.19583953.v1)** compare MoCo-MSLR versus Extreme MRI for IPF reconstructions targeting 588ms temporal resolution. From **supplemental video 5.7**, structures around the lung hilum are sharp for MoCo-MSLR throughout respiration. These structures are blurred somewhat in Extreme MRI. Additionally, there is less flickering artifact in the MoCO-MSLR reconstruction than Extreme MRI. Notice that the blur around the liver edge in Extreme MRI is replaced by warping artifact in MoCo-MSLR.

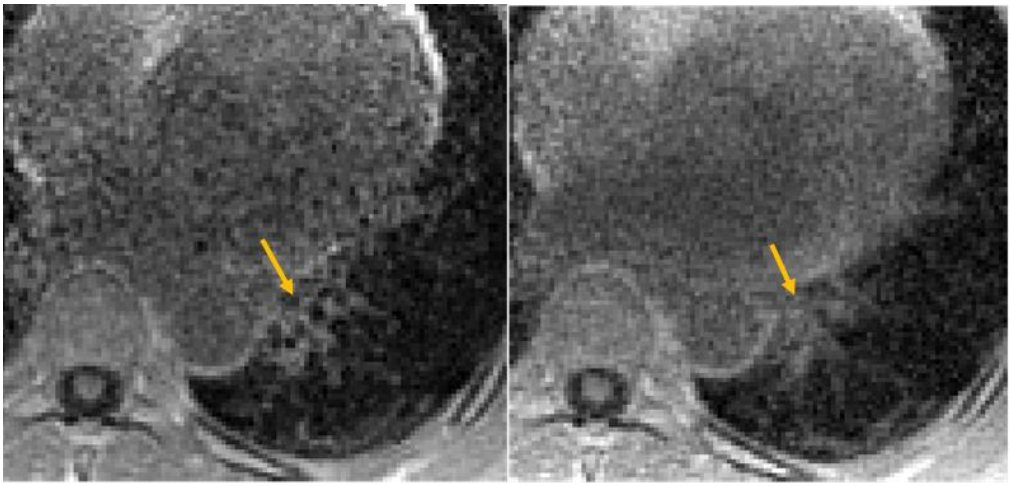


Figure 5.7: *Reconstructions Results on IPF Patient. Representative axial slices near the lung base is shown for both reconstructions. Fibrosis around the airway and the airways themselves are more clearly resolved in MoCo-MSLR than Extreme MRI.*

In **figure 5.7**, MoCo-MSLR clearly resolves small airways and associated fibrosis (orange arrow) not visualized in Extreme MRI. From both **supplemental video 5.7** and **figure 5.1c**, it appears that respiratory motion is similar between the two reconstructions, however, the MC-MSLR reconstruction does appear to miss a transient diaphragm excursion seen in Extreme MRI **Supplemental video 5.8 (https://doi.org/10.6084/m9.figshare.19583956.v1)** shows a sagittal slice paired with its associated motion field through time demonstrating how the displacement field changes throughout the respiratory cycle.

5.4.4 Third Trimester Pregnant Patient Dataset

Figure 5.8 and **Supplemental Video 5.9** **(https://doi.org/10.6084/m9.figshare.19583959.v1)** compare MoCo-MSLR and Extreme MRI for reconstructions targeting 605ms temporal resolution

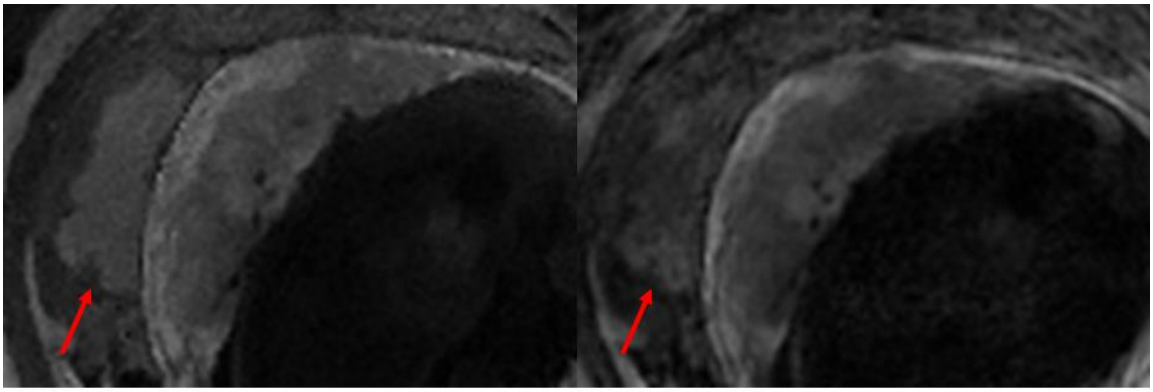


Figure 5.8: *Reconstructions Results on Healthy Pregnant Patient in Third Trimester. View showing the placenta and uterine layers. Significant artifact obstructs the uterus not seen in MoCo-MSLR (red arrow).*

Figure 5.8 shows that MoCo-MSLR results in sharper delineation between uterine layers than Extreme MRI where these layers are obscured by artifact. Motion dynamics appear to be similar between reconstructions both in **supplemental video 5.9** and from the respiratory signals in **figure 5.1d**. A uterine contraction is observed from 3.38 to 5.93 time units.

5.5 Discussion

In this work, I developed a method to derive and then integrate memory efficient representations of forward and adjoint motion deformation fields into Extreme MRI reconstructions. In this method, MoCo-MSLR, low resolution motion fields are first learned directly as multiscale low rank components by enforcing k-space based loss between a deformed template and acquired k-space data. These fields are then interpolated in the MSLR space to match the desired full resolution reconstruction. Finally, the deformation fields and their adjoint are incorporated into Extreme MRI in the forward model. By using compact representations for both motion fields and the time series, motion compensation high spatiotemporal reconstructions are made possible with very low memory footprint.

MoCO-MSLR results in improved image quality compared to Extreme MRI at ~500 ms temporal resolution. Image quality improvements seen with our method include reduced undersampling and

flickering artifacts, sharper image features, the ability to resolve small vascular and airway features, and resolve certain dynamics not seen in Extreme MRI reconstruction. MoCo-MSLR at higher temporal resolutions (~100 ms) realistically captures cardiac dynamics. Extreme MRI incompletely resolved cardiac dynamics in the healthy volunteer with high blood pool to myocardium contrast. In the CF case with lower blood pool to myocardium contrast, Extreme MRI completely failed to resolve cardiac dynamics.

This work demonstrates similar image quality improvement seen with past strategies incorporating motion fields directly into reconstructions. This improvement was expected to some extent because the time series modeled by the left spatial and right temporal bases in MoCo-MSLR is aligned meaning maximal correlations exist across frames. Image quality improvements can be seen in the work of [26], [33] when aligning data during reconstruction. Our model simply extends this notion of improved reconstruction through alignment to a much larger scale problem. Without motion correction, the left spatial and right temporal bases in MSLR model all dynamics in the time series which reduces the degree of correlation across frames ultimately reducing image quality.

There were, however, significant variations in the degree of image quality improvement across cases. This appeared to be, in part, related to the complexity of motion. These differences can be seen particularly well when comparing the healthy volunteer with nearly periodic motion (**supplemental video 5.1**) to the CF patient with both irregular respiratory and bulk motions. In the healthy volunteer, MoCo-MSLR and Extreme MRI are comparable with respect to image quality (**supplemental video 5.1** and **figure 5.2**) at temporal resolution targeting respiratory motion (~500ms). Minimal flickering and streaking artifact are seen, and small vascular features are resolved well by both reconstruction methods. On the other hand, MoCo-MSLR demonstrated significantly higher image quality (**supplemental video 5.4** and **figure 5.5**) then the Extreme MRI reconstruction for the CF patient. The liver edge is sharper in MoCo-MSLR even during irregular respiratory motion. Additionally, airway/vascular features blurred out in Extreme MRI are clearly resolved in the MoCo-MSLR reconstruction (**figure 5.5**). One possible explanation for this is Extreme MRI is not actually resolving all motion at the targeted temporal resolution which would lead to blur. For instance, in **supplemental video 5.4**, bulk motions and tracheal collapse seen in the MoCo-MSLR reconstruction are not observed in Extreme MRI. Although there is no way to validate if these motions are real, the quality of the MoCo-MSLR reconstructions suggests they are. Further, tracheomalacia which can lead to tracheal collapse especially when there are large fluctuations in thoracic pressures e.g. during a cough is common in patients with cystic fibrosis.

In general, MoCo-MSLR does appear to resolve irregular respirations and bulk motion with minimal blurring better than Extreme MRI. This makes sense because as mentioned in [1], irregular

respirations and bulk motion are not necessarily low rank even for small block sizes. By explicitly modeling these motions, MoCo-MSLR significantly reduces blur while capturing these motions. A counterargument to this is motion fields represented using multi-scale low rank components may suffer from the same issue. Although to some extent this is true, deformation fields only have to model motion, not the background plus dynamics and thus may admit more compressible representations allowing MoCo-MSLR to reconstruct even more undersampled data with high fideliy than the original Extreme MRI approach. The ability of MoCo-MSLR to capture cardiac dynamics at ~100 ms temporal resolution while Extreme MRI struggles lends experimental evidence to this hypothesis.

At high temporal resolutions (~100 ms) significant differences in reconstruction quality remain both when comparing MoCo-MSLR to Extreme MRI and when comparing each reconstruction to itself across different cases. Although complexity of motion may still play a role here, it appears that higher SNR results in improved ability to capture high temporal resolution dynamics. This can be seen when comparing the higher SNR contrast enhanced healthy volunteer acquisition to the lower SNR CF acquisition. Extreme MRI captures some cardiac motion in the healthy volunteer, but no cardiac motion can be seen in the lower SNR CF acquisition. Although MoCo-MSLR captures cardiac dynamics in both the healthy volunteer and CF patient, the CF reconstruction has significantly more high frequency oscillations present **(supplemental video 5.5)** suggesting the deformation fields are also modeling noise in addition to signal. This preliminary finding suggests that at high temporal resolution, contrast-enhanced acquisitions may be preferred.

There are a number of limitations to this work. There are several image artifacts that arise because the deformation fields are not topology preserving (i.e., non-diffeomorphic). In the IPF case (**supplemental video 5.7**), a sandpaper like texture can be seen in and around the liver edge. In L.T's experience using other motion correction algorithms like iMoCo, these same artifacts arise when the deformation fields are not topology preserving i.e. non-diffeomorphic. Use of algorithms that ensure the fields are diffeomorphic removes these artifacts in the context of iMoCo. A related warping artifact can be seen in the high temporal resolution reconstructions. This artifact occurs when tissues that locally should be moving together, displace with different velocities essentially tearing the tissue apart. The result is a kind of blurring. A potential direction for this work is to develop multi-scale compressed representations for diffeomorphic fields. It is not immediately clear though how to develop such a method with theoretical guarantees.

Another artifact unrelated to non-diffeomorphic fields seen primarily in the ~100ms resolution is high frequency oscillations. This is significantly worse in the CF case then the healthy volunteer with the same regularization weights. Although the regularization on both spatial smoothing and rank

minimization can be increased to attempt to remove this artifact, the higher the regularization weight, the more difficult it becomes to capture motion. Exploring the hypothesis that ability to resolve high temporal resolution dynamics may be dependent on SNR may be fruitful to better define acquisition parameters to generate optimal high temporal resolution reconstructions.

Similar to [1], it is unknown whether the prescribed temporal resolution matched the true dynamics at that temporal resolution. Validation is a major challenge for this work. Few real time imaging modalities can scan simultaneously with MR to provide ground truth data, however, recent progress in simultaneous MRI/Ultrasound systems [42] may be a promising future approach for validation.

Finally, in its current form, MoCo-MSLR only works for images without contrast dynamics as it relies on warping a fixed template. The ability to incorporate motion estimation for high spatiotemporal reconstruction of acquisitions with contrast dynamics is an interesting avenue for future work.

5.6 Conclusion:

In this work I improve on a state-of-the-art image reconstruction algorithm (Extreme MRI) by incorporating motion. I demonstrate that MoCo-MSLR makes it possible to reconstruct motion compensated 3D dynamic acquisitions at high spatiotemporal resolutions in a computationally efficient manner. My method shows improved image sharpness and motion robustness when compared to Extreme MRI at the same temporal resolution. Additionally, when pushed to temporal resolutions of ~100ms, MoCo-MSLR can depict cardiac and respiratory dynamics beyond the capabilities of Extreme MRI.

Chapter 6: Summary and Future Directions

In this thesis, I have developed techniques that significantly reduce reconstruction time for 3D non-Cartesian acquisitions using model based deep learning (**chapter 3 and 4**). Further, I have developed methods that improve upon state-of-the-art techniques for reconstructing high spatiotemporal resolution data by integrating motion compensation into these large scale reconstructions (**chapter 5)**. Below, I summarize my contributions:

6.1: Summary of Contributions

Memory Efficient MBDL Reconstructions for High Spatial Resolution 3D Non-Cartesian Acquisitions

I have developed a method termed block-wise learning with gradient checkpointing that allows MBDL to be applied to 3D non-Cartesian reconstructions on a single GPU. Prior to this work, high spatial resolution reconstructions using MBDL would have required state of the art GPU clusters during training. I show that this technique significantly improved image quality over compressed sensing techniques while significantly reducing reconstruction time from minutes to seconds.

Self Supervised Deep Learning for Highly Spatial Resolution 3D Non-Cartesian Acquisitions

In this work, I addressed the challenge obtaining fully sampled 3D Non-Cartesian ground truth data for supervised training of MBDL. I extend the self-supervised learning model proposed in [6] to leverage correlations across frames without significantly extending training time. I then combine this model with GPU-based motion correction to further improve reconstruction quality. I show that this motion compensation method is competitive with state-of-the-art iterative techniques like iMoCo while significantly reducing reconstruction time.

Motion Compensated High Spatiotemporal Resolution MRI (MoCo-MSLR)

I have presented a method for integrating motion compensation into high spatiotemporal resolution reconstructions. I represent motion fields directly in a compressed multi-scale low rank space, and estimate these motion fields at low resolution using loss enforced in k-space. These interpolated motion fields are then integrated into the Extreme MRI model for final reconstruction. I show that MoCo-MLR significantly improves reconstruction quality over Extreme MRI at ~500 ms temporal resolution. Further, I demonstrate that MoCo-MSLR captures realistic cardiac dynamics at ~100 ms temporal resolution.

6.2: Future Directions

Here I discuss future work I plan to complete over the course of the fourth year of medical school.

6.2.1 Pulmonary Lesion Study

A major limitation to the model based deep learning work in chapters 3 and 4 is both training and testing of these models was done using data from healthy volunteers. The goal of the pulmonary lesion project is to simultaneously address questions regarding the generalizability of these models to patients with disease, and to apply this work to an unmet clinical need.

PET/MR systems are increasingly used for clinical staging, radiation planning, and surveillance for cancer patients. Use of these systems is attractive because it significantly reduces ionizing radiation dose and has improved soft tissue contrast over CT. Imaging the lung, a key part of cancer surveillance, however, is challenging with conventional Cartesian MR sequences. Recent work [43] demonstrated that 3D radial UTE sequences significantly improved pulmonary lesion detection rate over Cartesian sequences with CT as a gold standard. These scans were acquired during free breathing and reconstructed by binning data and applying traditional compressed sensing methods. The end-expiratory phase in these reconstructions tends to have the highest image quality as the majority of the respiratory phase is spent close to end-expiration. Use of the end-expiratory phase for lesion detection, however, may be suboptimal as the lung is maximally compressed during this phase potentially reducing lesion identification and distorting lesion shape.

To address this issue, I will build upon the block-wise learning methods developed in chapter 3 to reconstruct end-inspiratory breath held scans in patients with known pulmonary lesions. Additionally, I will use the motion compensation techniques developed in chapter 4 to reconstruct end-inspiratory images from free breathing UTE acquisitions in patients with known pulmonary lesions. Specifically, in collaboration with Ali Pirasteh, MD and Kevin Johnson, PhD, end-inspiratory and end-expiratory breath hold and free breathing pulmonary UTE scans will be acquired in (at least) 10 patients with known lung lesions as add on to previously scheduled PET/MR acquisitions. To be recruited for the study, patients will have recently acquired pulmonary CT scans for use as a gold standard. Breath-held end-inspiratory, breath-held end-expiratory and free breathing reconstructions will be compared both in terms of image quality and pulmonary lesions identified against the CT gold standard.

I will make several modifications to the block-wise learning model proposed in chapter 3 that may potentially improve reconstruction quality and ability to identify pulmonary lesions. First, in place of supervised learning, I will use self-supervised learning as in (X) to train the model to remove the need to rely on proxy ground truth images. Second, in place of gradient descent data-consistency steps, I will use conjugate gradient iterations to allow faster convergence with fewer unrolls. Third, I will train the model for a greater number of training iterations than was used in chapter 3.

Within a given breath held acquisition, I will compare reconstruction quality and ability to identify pulmonary lesions across several different architectures including the model described above trained on pulmonary lesion data only, the same model trained on healthy volunteers, and the original model in chapter 3 trained on solely on healthy volunteers. Reconstruction quality will be assessed through aSNR, CNR, and sharpness metrics and a radiology reader study. Pulmonary lesion identification rates will be assessed also by radiology reader study using the same approach described in (X).

The best breath held end-inspiratory and end-expiratory reconstructions as found by the analysis above will be compared with respect to both image quality and pulmonary lesion identification rates to the end-inspiratory and end-expiratory phase of motion compensated MBDL reconstructions and iterative motion compensated reconstructions.

There is limited work applying DL reconstruction models to pathology. This work is an opportunity to better understand how generalizable and to be frank useful the models I have built are. Further, this work, may help start identifying where these architectures fail and promising directions for future development.

In the most optimistic case where the reconstructions work well, there is an entirely separate question of how to integrate these reconstructions into clinical workflows. Although these models do significantly reduce reconstruction time compared to traditional methods, all comparisons I made were on state-of-the-art GPUs. A much longer-term question that will not be addressed by this project is how to integrate DL architectures that require state of the art GPUs for both training and inference cleanly into clinical practice.

With respect to timelines, I will likely start this project as a part of the MSTP 902 class in October 2022. I have significant clinical commitments (i.e.,relearning how to be a medical student) prior to this point. Given the scope of this work, I expect the project including data acquisition, data processing and data analysis to take the remainder of fourth year of medical school.

6.2.2. Validating Motion Dynamics in MoCo-MSLR and Extreme MR

A major limitation of both MoCo-MSLR and Extreme MRI for clinical implementation is the lack of validation of the dynamics captured by these reconstructions. For low rank reconstructions that leverage correlations across frames, the targeted temporal resolution and actual temporal resolution resolved by these reconstructions may be different. For instance, there were several points in time where motion resolved by MoCo-MSLR was not observed in Extreme MRI reconstructions. A separate but equally important question for clinical implementation is determining when these computationally

expensive reconstructions are useful to apply to acquisitions versus less burdensome respiratory binned reconstructions.

In this project, I will compare respiratory dynamics resolved by MoCo-MSLR and Extreme MRI along with other center of k-space and respiratory belt signals against respiratory signal from MRI compatible 4D ultrasound acquired simultaneously with the MRI scan. 4D ultrasound provides near real-time, volumetric imaging independent of the MRI acquisition, and thus acts as a gold standard for assessing respiratory dynamics.

I will then compare reconstruction quality between iterative motion compensated reconstructions (iMoCo) using motion fields estimated from k-space data binned based on center of k-space, respiratory belt, and ultrasound signals. Multi-scale low rank motion fields will then be estimated from low resolution Extreme MRI reconstructions (3D Dynamic Navigators). These motion fields will then be integrated into an time-resolved iMoCo reconstruction. Multi-scale low rank motion fields estimated from MoCo-MSLR will also be integrated into time-resolved iMoCo reconstructions. Reconstruction quality will be compared similarly to the work in Chapter 4 (also in [26]) using aSNR, CNR and sharpness.

Specifically, in ten healthy volunteers, two simultaneous ultrasound and 3D pulmonary UTE acquisitions will be acquired. The ultrasound probe will be placed with an intercostal window over the dome of the window. In the first acquisition, the patient will be asked to breath normally to capture close to periodic breathing when both MoCo-MSLR and Extreme MRI should perform optimally. In the second acquisition, the patient will be asked to perform an end-inspiratory breath-hold one minute into the scan. Regular ventilation disrupted by an end-inspiratory breath hold simulates highly irregular breathing and should push the limits of both reconstruction methods.

Respiratory dynamics will be compared between MoCo-MSLR, Extreme MRI, center of k-space respiratory navigators, and the respiratory belt against the ultrasound gold standard. This is similar to our work proposed in [44] for 4 healthy volunteers. **Figure 6.1** shows an example of this comparison:

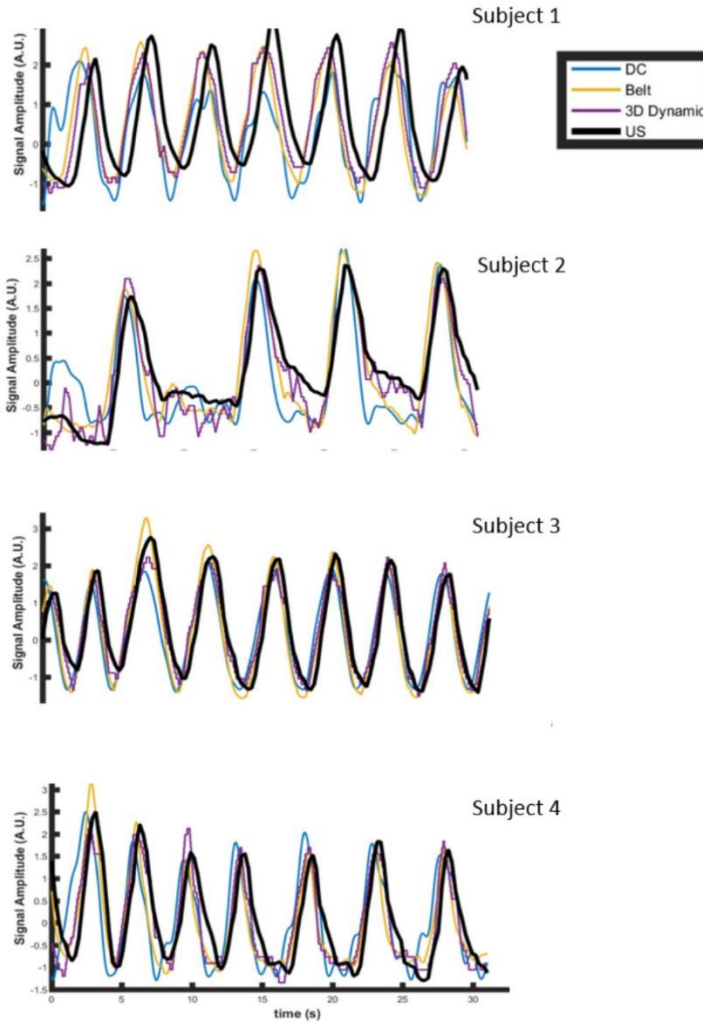


Figure 6.1*: Representative Respiratory Signals using Different Navigation Strategies*

With respect to timelines, acquisitions in four healthy volunteers have already been acquired. It is possible that prior to October, a couple more healthy volunteers may be scanned. The goal, however, will be to finish remaining acquisitions in October, and start data processing and analysis by November 2022.

6.2.3. Motion Compensated High spatiotemporal resolution Dynamic contrast Enhanced Reconstructions

A very interesting direction for MoCo-MSLR is reconstructing acquisitions with **both** motion and contrast dynamics. There are two potential directions for this work. The first method uses the following steps:

1. reconstruct low spatial resolution navigator images at ~500ms temporal resolution.
2. Estimate multi-scale low rank motion fields by leveraging a group-wise nuclear norm based loss. Unlike ([35]) that requires taking an SVD of all the data to compute the nuclear norm, I propose taking a randomly chosen batch of frames and minimizing the nuclear norm over this batch of frames, and iterating. I caution though that there is zero theoretical support (and I have done very little empirical testing of this idea) for this approach as the singular values from a subset of frames may vary significantly across subsets.
3. Integrate this motion field estimates into a final Extreme MRI reconstruction.

The second potential method is an extension of MoCo-MSLR. In the current version of MoCo-MSLR, I solve the low-resolution k-space based motion estimation problem as minimization of:

$$f(\phi,\psi) = \left\|Y_t - E\left((I_{aligned}(\Omega_{for,t}))\right)\right\|^2 + \sum_{j=1}^{J} \frac{\lambda_j}{2}\left(\left\|\Phi_{j,for}\right\|_F^2 + \left\|D\Psi_{j,for}^H\right\|_F^2\right) + \gamma\left\|D\Omega_{for}\right\| \quad \textbf{(6.1)}$$

Where $I_{aligned}$ is a fixed template image. Here I propose incorporating contrast dynamics by replaced the fixed template with a fixed data-consistent frame I_{fixed} plus a multi-scale low rank representation of contrast dynamics:

$$I_t = I_{fixed} + \sum_{j=1}^{J} M_j(L_j R_j^H) \quad \textbf{(6.2)}$$

Where $I_c = \sum_{j=1}^{J} M_j(L_j R_j^H)$ represents contrast dynamics through time. Contrast dynamics can be learned by minimizing:

$$f(L,R) = \frac{1}{2}\left\|Y_t - E\left((I_{aligned}(\Omega_{for,t}))\right)\right\|^2 + \frac{\lambda_j}{2}\sum_{j=1}^{J}\left(\frac{\|L_j\|_F^2}{T} + \|R_j\|_F^2\right) \quad \textbf{(6.3)}$$

To learn both motion field and contrast dynamics during reconstruction, I propose alternating between motion field updates and contrast dynamic updates. I have found while this approach *can* work, it requires significant tuning as motion tends to leak into the MSLR model of contrast dynamics and the motion fields attempt to model contrast dynamics.

Starting and finishing this project during fourth year of medical school will be very much dependent on how the pulmonary lesion and MoCo-MSLR ultrasound validation work go.

References

[1] F. Ong *et al.*, "Extreme MRI: Large-scale volumetric dynamic imaging from continuous non-gated acquisitions," *Magn. Reson. Med.*, vol. 84, no. 4, pp. 1763–1780, 2020, doi: 10.1002/mrm.28235.

[2] X. Zhu, F. Tan, K. Johnson, and P. Larson, "Optimizing trajectory ordering for fast radial ultra-short TE (UTE) acquisitions," *J. Magn. Reson.*, vol. 327, p. 106977, Jun. 2021, doi: 10.1016/j.jmr.2021.106977.

[3] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model-Based Deep Learning Architecture for Inverse Problems," *IEEE Trans. Med. Imaging*, vol. 38, no. 2, pp. 394–405, Feb. 2019, doi: 10.1109/TMI.2018.2865356.

[4] K. Hammernik *et al.*, "Learning a Variational Network for Reconstruction of Accelerated MRI Data," *Magn. Reson. Med.*, vol. 79, no. 6, pp. 3055–3071, Jun. 2018, doi: 10.1002/mrm.26977.

[5] C. M. Sandino, P. Lai, S. S. Vasanawala, and J. Y. Cheng, "Accelerating cardiac cine MRI using a deep learning-based ESPIRiT reconstruction," *Magn. Reson. Med.*, vol. 85, no. 1, pp. 152–167, 2021, doi: 10.1002/mrm.28420.

[6] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya, "Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data," *Magn. Reson. Med.*, vol. 84, no. 6, pp. 3172–3191, 2020, doi: 10.1002/mrm.28378.

[7] J. A. Fessler and B. P. Sutton, "Nonuniform fast Fourier transforms using min-max interpolation," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 560–574, Feb. 2003, doi: 10.1109/TSP.2002.807005.

[8] K. P. Pruessmann, M. Weiger, P. Börnert, and P. Boesiger, "Advances in sensitivity encoding with arbitrary k-space trajectories," *Magn. Reson. Med.*, vol. 46, no. 4, pp. 638–651, Oct. 2001, doi: 10.1002/mrm.1241.

[9] E. J. Candes and B. Recht, "Exact Matrix Completion via Convex Optimization," p. 49.

[10] "Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization." https://epubs.siam.org/doi/epdf/10.1137/070697835 (accessed Apr. 12, 2022).

[11] "Universal Approximation Using Feedforward Neural Networks: A Survey of Some Existing Methods, and Some New Results - ScienceDirect." https://www.sciencedirect.com/science/article/pii/S089360809700097X?casa_token=1ip-Vi4oaCcAAAAA:YSGjh3fBt5_xbu7vIdvpF1dscfWQYdbUtr73cvIxtI1QXkshPjDHAV3j0VhWGghe23aPqmjvIQ (accessed Apr. 13, 2022).

[12] "Backpropagation | Brilliant Math & Science Wiki." https://brilliant.org/wiki/backpropagation/ (accessed Apr. 12, 2022).

[13] T. Chen, B. Xu, C. Zhang, and C. Guestrin, "Training Deep Nets with Sublinear Memory Cost," *ArXiv160406174 Cs*, Apr. 2016, Accessed: Apr. 11, 2022. [Online]. Available: http://arxiv.org/abs/1604.06174

[14] G. Zeng *et al.*, "A review on deep learning MRI reconstruction without fully sampled k-space," *BMC Med. Imaging*, vol. 21, no. 1, p. 195, Dec. 2021, doi: 10.1186/s12880-021-00727-9.

[15] K. M. Johnson, S. B. Fain, M. L. Schiebler, and S. Nagle, "Optimized 3D ultrashort echo time pulmonary MRI," *Magn. Reson. Med.*, vol. 70, no. 5, pp. 1241–1250, 2013, doi: 10.1002/mrm.24570.

[16] F. Ong, M. Uecker, and M. Lustig, “Accelerating Non-Cartesian MRI Reconstruction Convergence Using k-Space Preconditioning,” *IEEE Trans. Med. Imaging*, vol. 39, no. 5, pp. 1646–1654, May 2020, doi: 10.1109/TMI.2019.2954121.
[17] J. C. Ye, “Compressed sensing MRI: a review from signal processing perspective,” *BMC Biomed. Eng.*, vol. 1, no. 1, p. 8, Mar. 2019, doi: 10.1186/s42490-019-0006-z.
[18] Z. Ramzi, G. R. Chaithya, J.-L. Starck, and P. Ciuciu, “NC-PDNet: a Density-Compensated Unrolled Network for 2D and 3D non-Cartesian MRI Reconstruction,” *IEEE Trans. Med. Imaging*, pp. 1–1, 2022, doi: 10.1109/TMI.2022.3144619.
[19] M. O. Malavé *et al.*, “Reconstruction of undersampled 3D non-Cartesian image-based navigators for coronary MRA using an unrolled deep learning model,” *Magn. Reson. Med.*, vol. 84, no. 2, pp. 800–812, Aug. 2020, doi: 10.1002/mrm.28177.
[20] N. S. Sohoni, C. R. Aberger, M. Leszczynski, J. Zhang, and C. Ré, “Low-Memory Neural Network Training: A Technical Report,” *ArXiv190410631 Cs Stat*, Apr. 2019, Accessed: Apr. 11, 2022. [Online]. Available: http://arxiv.org/abs/1904.10631
[21] M. Kellman *et al.*, “Memory-Efficient Learning for Large-Scale Computational Imaging,” *IEEE Trans. Comput. Imaging*, vol. 6, pp. 1403–1414, 2020, doi: 10.1109/TCI.2020.3025735.
[22] K. Wang *et al.*, “Memory-efficient Learning for High-Dimensional MRI Reconstruction,” *ArXiv210304003 Cs Eess*, Mar. 2021, Accessed: Apr. 11, 2022. [Online]. Available: http://arxiv.org/abs/2103.04003
[23] G. Knobloch *et al.*, “Comparison of gadolinium-enhanced and ferumoxytol-enhanced conventional and UTE-MRA for the depiction of the pulmonary vasculature,” *Magn. Reson. Med.*, vol. 82, no. 5, pp. 1660–1670, Nov. 2019, doi: 10.1002/mrm.27853.
[24] M. Buehrer, K. P. Pruessmann, P. Boesiger, and S. Kozerke, “Array compression for MRI with large coil arrays,” *Magn. Reson. Med.*, vol. 57, no. 6, pp. 1131–1139, 2007, doi: 10.1002/mrm.21237.
[25] L. Ying and J. Sheng, “Joint image reconstruction and sensitivity estimation in SENSE (JSENSE),” *Magn. Reson. Med.*, vol. 57, no. 6, pp. 1196–1202, 2007, doi: 10.1002/mrm.21245.
[26] X. Zhu, M. Chan, M. Lustig, K. M. Johnson, and P. E. Z. Larson, “Iterative motion-compensation reconstruction ultra-short TE (iMoCo UTE) for high-resolution free-breathing pulmonary MRI,” *Magn. Reson. Med.*, vol. 83, no. 4, pp. 1208–1221, 2020, doi: 10.1002/mrm.27998.
[27] M. Markl, A. Frydrychowicz, S. Kozerke, M. Hope, and O. Wieben, “4D flow MRI,” *J. Magn. Reson. Imaging JMRI*, vol. 36, no. 5, pp. 1015–1036, Nov. 2012, doi: 10.1002/jmri.23632.
[28] T. Zhang *et al.*, “Fast Pediatric 3D Free-breathing Abdominal Dynamic Contrast Enhanced MRI with High Spatiotemporal Resolution,” *J. Magn. Reson. Imaging JMRI*, vol. 41, no. 2, pp. 460–473, Feb. 2015, doi: 10.1002/jmri.24551.
[29] T. Zhang *et al.*, “Clinical Performance of Contrast Enhanced Abdominal Pediatric MRI with Fast Combined Parallel Imaging Compressed Sensing Reconstruction,” *J. Magn. Reson. Imaging JMRI*, vol. 40, no. 1, pp. 13–25, Jul. 2014, doi: 10.1002/jmri.24333.
[30] L. Feng, L. Axel, H. Chandarana, K. T. Block, D. K. Sodickson, and R. Otazo, “XD-GRASP: Golden-Angle Radial MRI with Reconstruction of Extra Motion-State Dimensions Using Compressed Sensing,” *Magn. Reson. Med.*, vol. 75, no. 2, pp. 775–788, Feb. 2016, doi: 10.1002/mrm.25665.
[31] J. Lehtinen *et al.*, “Noise2Noise: Learning Image Restoration without Clean Data,” *ArXiv180304189 Cs Stat*, Oct. 2018, Accessed: Apr. 11, 2022. [Online]. Available: http://arxiv.org/abs/1803.04189
[32] R. Otazo, E. Candès, and D. K. Sodickson, “Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components,” *Magn. Reson. Med.*, vol. 73, no. 3, pp. 1125–1136, 2015, doi: 10.1002/mrm.25240.
[33] R. Otazo, T. Koesters, E. Candes, and D. K. Sodickson, “Motion-guided low-rank plus sparse (L+S) reconstruction for free-breathing dynamic MRI,” in *Proc Intl Soc Mag Reson Med*, 2014, vol. 22, p. 0742.

[34] N. R. F. Huttinga, T. Bruijnen, C. A. T. van den Berg, and A. Sbrizzi, “Nonrigid 3D motion estimation at high temporal resolution from prospectively undersampled k-space data using low-rank MR-MOTUS,” *Magn. Reson. Med.*, vol. 85, no. 4, pp. 2309–2326, 2021, doi: 10.1002/mrm.28562.
[35] W. Huizinga *et al.*, “PCA-based groupwise image registration for quantitative MRI,” *Med. Image Anal.*, vol. 29, pp. 65–78, Apr. 2016, doi: 10.1016/j.media.2015.12.004.
[36] C. Jaimes, J. E. Kirsch, and M. S. Gee, “Fast, free-breathing and motion-minimized techniques for pediatric body magnetic resonance imaging,” *Pediatr. Radiol.*, vol. 48, no. 9, pp. 1197–1208, Aug. 2018, doi: 10.1007/s00247-018-4116-x.
[37] L. Torres *et al.*, “‘Structure-Function Imaging of Lung Disease Using Ultra-Short Echo Time MRI,’” *Acad. Radiol.*, vol. 26, no. 3, pp. 431–441, Mar. 2019, doi: 10.1016/j.acra.2018.12.007.
[38] J. Ludwig, P. Speier, F. Seifert, T. Schaeffter, and C. Kolbitsch, “Pilot tone–based motion correction for prospective respiratory compensated cardiac cine MRI,” *Magn. Reson. Med.*, vol. 85, no. 5, pp. 2403–2416, 2021, doi: 10.1002/mrm.28580.
[39] S. Burer and R. D. C. Monteiro, “A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization,” *Math. Program.*, vol. 95, no. 2, pp. 329–357, Feb. 2003, doi: 10.1007/s10107-002-0352-8.
[40] F. Ong and M. Lustig, “Beyond low rank + sparse: Multi-scale low rank matrix decomposition,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2016, pp. 4663–4667. doi: 10.1109/ICASSP.2016.7472561.
[41] V. Vishnevskiy, T. Gass, G. Székely, and O. Goksel, “Total Variation Regularization of Displacements in Parametric Image Registration,” in *Abdominal Imaging. Computational and Clinical Applications*, vol. 8676, H. Yoshida, J. J. Näppi, and S. Saini, Eds. Cham: Springer International Publishing, 2014, pp. 211–220. doi: 10.1007/978-3-319-13692-9_20.
[42] W. Lee *et al.*, “A Magnetic Resonance Compatible E4D Ultrasound Probe for Motion Management of Radiation Therapy,” *IEEE Netw.*, vol. 2017, p. 10.1109/ULTSYM.2017.8092223, Sep. 2017, doi: 10.1109/ULTSYM.2017.8092223.
[43] N. S. Burris *et al.*, “Detection of Small Pulmonary Nodules with Ultrashort Echo Time Sequences in Oncology Patients by Using a PET/MR System,” *Radiology*, vol. 278, no. 1, pp. 239–246, Jan. 2016, doi: 10.1148/radiol.2015150489.
[44] Miller, Zachary, “MR compatible 4D Ultrasound based validation of respiratory motion compensation strategies”.