DISS. ETH NO. 27376

BAYESIAN METHODS FOR UNDERSAMPLED MR IMAGE RECONSTRUCTION WITH A LEARNED PRIOR

A dissertation submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

KEREM CAN TEZCAN

M.Sc., Heidelberg University

born on 18 March 1989
citizen of Turkey

accepted on the recommendation of

Prof. Dr. Ender Konukoglu, examiner
Prof. Dr. Michaël Unser, co-examiner
Prof. Dr. Yi Wang, co-examiner

2021

DOI: 10.3929/ethz-a-

To friends

ABSTRACT

Undersampling the k-space during acquisition saves time, however results in an underdetermined system, leading to an ill-posed inverse problem and raising the question of how to obtain images from these measurements. In this thesis we explore different solutions to this question.

Firstly, to establish a background for the proposed methods, we briefly provide the mathematical model of the image generation process and describe the undersampled reconstruction problem as well as motivate the use of a Bayesian framework as a generic setting to express the forward and inverse problems. We then build on top of this and describe our contributions.

The key component of the proposed approaches in the thesis is specifying the posterior distribution of the images given the measured k-space data, using a likelihood and a prior distribution. Though the former is more straightforward to model, the latter is not. To this end we rely on recent methods on density estimation from the deep learning literature and learn the distribution of fully sampled MR images using a variational autoencoder (VAE) and use this as the prior. Combining the likelihood and the learned prior specifies the posterior, which can then be further characterized.

The first characterization is to find the maximum-a-posteriori (MAP) estimate, which yields the image with the highest probability given the measured data. In the thesis we propose and evaluate a method that achieves this using the projection onto convex sets algorithm.

Afterwards we investigate and address the effect of the bias field in the images which causes a domain shift between the image to be reconstructed and the images used to train the prior. To mitigate its negative effects on MAP estimation, we propose and evaluate a method that estimates and models the bias field in the image generation process.

As the MAP estimation yields a single image, it cannot characterize the posterior comprehensively. From the perspective of the underdetermined inverse problem, this means it yields only one of the infinitely many solutions that are plausible under the measured data and the prior model. To obtain multiple of these plausible images we propose and evaluate a sampling scheme that implements a random walk in the latent space of the VAE using the Metropolis adjusted Langevin algorithm.

ZUSAMMENFASSUNG

Die unterabgetastete Aufnahme des k-Raums ist zeitsparend, jedoch ergibt sich daraus ein schlecht gestelltes inverses Problem und wirft so die Frage auf, wie man aus einer solchen Aufnahme Bilder gewinnen kann. In dieser Dissertation möchten wir verschiedenen Lösungsansätzen zu dieser Frage nachgehen.

Als Erstes stellen wir kurz das mathematische Modell für den Prozess der Bilderzeugung vor und beschreiben das unterabgetastete Rekonstruktionsproblem um so eine Grundlage für die vorgeschlagenen Methoden zu schaffen. Zudem führen wir bayessische Statistik ein und zeigen, wie die direkten und inversen Probleme in diesem Rahmen formuliert werden können. Auf dieser Basis beschreiben wir dann unsere Beiträge.

Der zentrale Bestandteil der in dieser Arbeit vorgeschlagenen Herangehensweisen ist die Bestimmung der a-posteriori Verteilung der Bilder mit den gemessenen k-Raum Daten mittels einer Likelihood-Funktion und einer a-priori Verteilung. Die Erste ist einfacher zu modellieren, im Gegensatz zu der Zweiten, für welche wir neuere Methoden für unüberwachte Dichteschätzung aus der Literatur zu tiefgehendem Lernen benutzen. Insbesondere verwenden wir ein Variational Autoencoder (VAE) als eine Annäherung zu der a-priori Verteilung, also der Verteilung der vollabgetasteten MR Bilder. Das Kombinieren der Likelihood-Funktion und der gelernten a-priori Verteilung ergibt dann die a-posteriori Verteilung, die weiter charakterisiert werden kann.

Die erste Charakterisierung ist die Maximum-a-posteriori-Schätzung (MAP), die das Bild mit der höchsten a-posteriori Wahrscheinlichkeit ergibt. In dieser Dissertation schlagen wir eine Methode vor, die dies anhand des Algorithmus Projection onto Convex Sets (POCS) ermöglicht.

Anschliessend untersuchen wir den Effekt des Bias Fields in den Bildern, der einen Domain Shift zwischen den zu rekonstruierenden und den bei der Optimierung des Prior-Modells verwendeten Bildern verursacht. Um dessen negative Effekte auf die MAP-Schätzung zu reduzieren, schlagen wir eine Methode vor, die das Bias Field einschätzt und es explizit im Modell des bildgebenden Verfahrens einsetzt.

Da die MAP-Schätzung nur ein Bild als Lösung ergibt, kann es die a-posteriori Verteilung nicht umfassend charakterisieren. Aus der Perspektive des schlecht gestellten inversen Problems gesehen, ist dieses Bild nur eine

von unendlich vielen Lösungen, welche durch die gemessenen Daten und das Prior-Modell definiert werden. Um mehrere dieser möglichen Bilder zu erhalten, schlagen wir ein Stichprobenverfahren vor, in welchem eine zufällige Irrfahrt im latenten Raum des VAEs nach Prinzipien des Langevin-Algorithmus mit Metropolis-Anpassung stattfindet.

ACKNOWLEDGEMENTS

First and foremost, I am obliged to thank Ender, not only because writing this thesis would obviously have been impossible if he had not given me the opportunity but due to much more. I am grateful to him for sharing his vast technical knowledge and creativity with me, while maintaining a high regard for my well-being as his PhD student. He has been more than just a Doktorvater – as one would say in German – he has been a Doktor-friend, brother, psychologist, mentor, role model... I liken our relationship to the one of master and apprentice, where he trained me in many aspects of research, such as showing attention to details while not forgetting the bigger picture, thinking intuitively but writing rigorously, not compromising on the quality or ethics and many more. I will always – very pleasantly – remember the meetings spent debugging code together or late evening discussions on diverse topics. I imagine it is very demanding to train an apprentice, so thank you!

Secondly, I must say thank you to all my colleagues at the ETH. I had a laugh with all of them and shared a lot, both in research and in life. I unfortunately cannot list all the names but can mention just a few starting with the initial team of Christian, Lisa, Xiaoran, Krishna, Neerav, Anna, Katarína, later with Cécile, Jayden and Alexander, then with the addition of Gustav, Ertunç and Kyriakos... It has been a pleasure working with Mélanie, Georg and Christina as well. The lunch breaks with Ece, Fırat, Matija and Fabién will also be happily remembered. Many thanks to all my colleagues who shared the fussball table with me as well as my colleagues from the IBT. I must also express my gratitude to the administrative staff, Kristine, Christina and Christine, who run things so well and smoothly that you may even forget they exist, as well as the IT staff, especially Harry, without whom none of the research we do would be possible. Furthermore, I thank our collaborators Klaas Pruessmann and Roger Luechinger for their support in acquiring the data used throughout the thesis as well as Prof. Unser and Prof. Wang for agreeing to be co-examiners.

Next I must express my gratitude to my parents, Emel and Rüştü, who have never hesitated to sacrifice their own comfort to make sure that I got a good education. They have provided me with tremendous support at certain turning points during the PhD, without which I would not have

had the courage to continue. I hope this thesis gives them a bit of a feeling that it has not all been in vain.

A great thank you goes to all my friends. I derive so much power and energy from them... Again, just to mention a small subset of the names I should, I want to say thanks to my two dear flat mates of six years, Peter and Carlo, then Serhan, Ayşe, Türker, Melis and many more for being so close despite being geographically far away, then Cihan, Lynne, Shane and all my friends in Switzerland who made life here so much fun and finally, of course, Nina for bringing harmony and consonance into my life and being an admirably understanding companion during the PhD.

Finally, I feel an enormous gratitude towards all the friends of my parents, maybe Mustafa the most, who have not left my parents alone in their time of need and, perhaps, taught me more about friendships than what I learned on technical topics during the PhD. It is natural, in that regard, that I dedicate this thesis to them, to all my friends, to friendships and chosen families.

Dixi et salvavi animam meam.

CONTENTS

NOTATION

ABBREVIATIONS

MR	magnetic resonance
MRI	magnetic resonance imaging
CT	computer tomography
CS	compressed sensing
TV	total variation
SNR	signal to noise ratio
CN	contrast
CNR	contrast to noise ratio
DNN	deep neural network
CNN	convolutional neural network
GAN	generative adversarial network
cWGAN	conditional Wasserstein generative adversarial network
VAE	variational autoencoder
MRF	MR fingerprinting
MAP	maximum a posteriori
DL	deep learning
POCS	projection onto convex sets
SENSE	sensitivity encoding
ADNI	Alzheimer's Disease Neuroimaging Initiative
HCP	Human Connectome project
FOV	field of view
NUFFT	nununiform Fourier transform
GM	brain gray matter
WM	brain white matter
RMSE	root mean square error
FS	fully sampled image
ZF	zero filled image

T1w	T1 weighted image
T2w	T2 weighted image
GPU	graphical processing unit
MCMC	Markov chain Monte Carlo
HMC	Hamiltonian Monte Carlo
MALA	Metropolis adjusted Langevin Algorithm
CG	conjugate gradient
FFT	fast Fourier transform
DAE	denoising autoencoder

1

INTRODUCTION

Magnetic Resonance (MR) Imaging has become a prominent diagnostic and research tool in medicine in the last decades and has found numerous applications both in the clinics and research. It can be used for imaging many organs, especially where the soft tissue contrast is of importance. Examples include the brain, for instance, for diagnosis in case of brain trauma [1] or to investigate the structural [2] or functional connectivity of the brain [3]. Similarly, one can employ it as a as a diagnostic tool for cardiovascualar diseases [4]. Knee/cartilage imaging is another avenue in MRI [5]. Imaging of the lung [6] or of the abdominal region [7] is also possible. Furthermore, spectroscopic methods can yield information about the chemical composition of the tissue [8]. MRI provides a soft tissue contrast at a very high signal to noise ratio which is not attainable by other medical imaging methods such as computer tomography, ultrasound or x-ray imaging. Furthermore, it does not use ionizing radiation, making its use safer than methods relying on ionizing radiation.

This wide spread usability of MRI is reflected in the high number of machines available for human imaging. There are about 360.000 machines globally in total [9]. Japan, one of the leading countries in MR availability has about 50 machines per million population, followed by the US with 40 machines per million. There are annually 2500 new machines being produced by companies like Siemens, GE, Philips and Toshiba. Furthermore companies such as Bruker specialize on production of more research oriented machines for animal imaging or spectroscopy rather than medical human imaging. The market for MR systems was valued in total 5.9 billion USD [10] in 2017. This was 7.24 billion in 2019 and is projected to be 11.36 billion USD in 2027 [11]. This shows that the MR imaging is here to stay and to grow in use.

On the other hand MR imaging has not yet reached the end of its potential and there is still room for improvement in the field itself. One such aspect of MRI that is disadvantageous in comparison to other widely used medical imaging devices, which can be improved, is the acquisition time. A typical computer tomography (CT) or x-ray acquisition is in the order of seconds, ultrasound images can be taken as real-time measurements. In contrast, a regular MR image acquisition is in the order of minutes and

the whole examination can take between 30 minutes to two hours. This is a limiting factor for the applicability of MRI. In settings such as traumatic brain injury a CT scan is preferred due to its short acquisition time, even though the tissue contrast is better with MRI [12]. Furthermore, the long scan duration can be challenging for the patient as well. Due to the narrow structure of the scanner some patients experience claustrophobic discomfort. Similarly, patients can start moving in the scanner due to either impatience or due to pre-existing conditions which leads to motion artifacts and hence degradation in the image quality. Decreasing the scan duration has a positive effect on all the aspects above. Moreover, from an economic perspective the long scan duration results in a lower number of patients measured in a unit time. Decreasing the scan time increases the throughput of the scanner and hence decreases the cost of the device per patient to the benefit of the medical institution as well as reducing the waiting times for the patients. As a matter of fact, however, the resulting image quality in MR imaging is tightly connected to the duration of acquisition. This means, when employing any scan time reduction technique, good care must be taken not to compromise the image quality, since a degraded image quality would make the result unusable, rendering the mentioned economic and medical benefits irrelevant.

The central topic of this thesis is exactly the question of how to reduce scan times while compromising the image quality as little as possible. This question has already found many attempted solutions in the literature, where most of the answers propose a certain way of only acquiring a subset of the required data together with a method to recover an image from this subset without compromising image quality. In the MR literature the acquisition space is referred to as the k-space and to acquire only a subset of the k-space is referred to as undersampling. The general problem is then stated as an image recovery problem called MR image reconstruction from undersampled k-space measurements. Most approaches in MR image reconstruction can be categorized into hardware related solutions, compressed sensing and more conventional signal recovery techniques. More recently, deep learning has been employed in development of methods for this purpose as well. The thesis at hand contributes to these attempts by formulating the undersampled acquisition and providing solutions for it in a Bayesian framework.

The structure of the thesis is as follows:

Chapter 2 provides a more formal description of the MR image reconstruction problem. The aforementioned approaches from the literature are

summarized and put into a Bayesian perspective, when applicable. While doing that, we also motivate the development of our proposed approaches in relation to the existing methods. In this chapter we do not provide a literature survey for methods utilizing deep learning, rather provide this in each chapter seperately, where relevant.

Chapter 3 describes a method for reconstructing an image from the undersampled k-space acquisition in the Bayesian framework using learned priors based on deep learning. Here we evaluate and compare our proposed method to multiple existing methods.

In Chapter 4, the method from the previous chapter is extended to incorporate the effects of the bias field. As the bias field is different for each acquisition, it leads to a degradation in reconstruction quality. Here we show that by modeling its effects explicitly, we can reduce this degradation.

In Chapter 5 we describe a different approach to obtaining images from the undersampled k-space, where we build on the formulations from Chapter 3 and its extension in Chapter 4. Here, instead of obtaining a single image, we obtain multiple images that reflect the possibility of multiple images due to the missing data.

Finally, Chapter 6 provides a summary of the described methods from a critical point of view, identifies methodological or fundamental shortcomings or strengths of the used methods and discusses future directions for research.

2

BACKGROUND

In this chapter we describe the basics of MR imaging in a very simplified manner, introduce the notations, lay the basics of the undersampled MR image reconstruction problem and talk about several different pre-deep learning approaches.

2.1 IMAGE GENERATION IN MR

The content of this section is summarized from [13] and [14].

MR imaging works by placing the magnetic spin of different nuclei in the tissue in a high static magnetic field and then manipulating and measuring the signal generated by these. Though different nuclei can be targeted, the most commonly used nucleus is of the hydrogen atom due to its abundance in the human body and its high gyromagnetic ratio, where when combined, these two factors yield the highest signal among all nuclei [13, 15]. In this thesis we only focus on hydrogen imaging, though the developed methods could be partially used for different nuclei as well.

The contrast in MR imaging comes from the interactions of these magnetic spins with each other and with their surroundings. In both cases, the behavior of the spins can be modeled by the Bloch equations [16]. The Bloch equations describe how the spins behave in a static or changing magnetic field, where the behavior of the spin is dependant on its surroundings. Hence, when one perturbs the spins and observes how they behave, this provides information about their surroundings. When this information is discretized and quantized in space, the image is formed.

Formally, Bloch equations are given as

$$\frac{dM_x(t)}{dt} = \gamma_L(M(t) \times B(t))_x - \frac{M_x(t)}{T_2} \tag{2.1}$$

$$\frac{dM_y(t)}{dt} = \gamma_L(M(t) \times B(t))_y - \frac{M_y(t)}{T_2} \tag{2.2}$$

$$\frac{dM_z(t)}{dt} = \gamma_L(M(t) \times B(t))_z - \frac{M_z(t) - M_0}{T_1}, \tag{2.3}$$

where t is time, M is the magnetization vector of the spins, B is the static and varying magnetic field, subscripts denote the corresponding spatial dimension, T_1 is the spin-lattice relaxation time, T_2 is the spin-spin relaxation time [13], M_0 is the equilibrium magnetization proportional to the spin density (number of spins per unit volume), i.e. $M_0 \propto \rho_0$ and γ_L is the gyromagnetic ratio. The first parts of the equations define a cross product between M and B and the second parts define the relaxation for the particular spatial dimension.

In MR imaging the most basic idea is as follows: Once the space is discretized into pixels, each pixel has a different composition of tissue. As different tissue types in the body have different ρ_0, T1 and T2 values, this results in different M_0 and relaxation speeds for the magnetization vector M in each pixel. Then by measuring a signal from M in each pixel, one obtains different measurement values for different pixels corresponding to different tissue types. When put together, these form an image with a contrast dependent on the ρ_0 and relaxation times. In reality, there are many other processes which determine the contrast such as diffusion or $T2^*$ relaxation, however we omit these in our description for simplicity.

In MR the measurement devices are simple antennae placed close to the imaged body part. However the antennae do not capture a localized signal, but rather collect the signal from the whole body part at once. This raises the question of how to localize the signal source. The way this is achieved is referred to as Fourier encoding and works as follows.

The nuclei in a magnetic field B undergo a motion called precession with the so-called Larmor frequency ω

$$\omega = 2\pi\gamma_L B. \tag{2.4}$$

As such, the frequency of precession and similarly the phase of the precession (given by the time integral $\phi = \int \omega dt$) for a given nucleus are both determined by the effective magnetic field experienced by this nucleus. The idea is then to introduce spatially varying magnetic fields $G = (G_1, G_2, G_3)$ for the three spatial dimensions, called gradients, which encode the space information into the frequency and phase of the precession of the nuclei. Denoting spatial coordinates with $s = (s_1, s_2, s_3)$, the phase accumulated by a nucleus at position s during time τ is written as

$$\varphi(\tau) = 2\pi\gamma_L \int_0^\tau G(s)dt. \tag{2.5}$$

Assuming a linear gradient $G(s) = s \cdot G$, the expression can be rewritten as

$$\varphi(\tau) = 2\pi(s \cdot \gamma_L \int_0^\tau G dt) = 2\pi(s \cdot k), \tag{2.6}$$

where we defined the vectoral quantity k as

$$k := \gamma_L \int_0^\tau G dt \tag{2.7}$$

using $\cdot$ to denote the dot product. Notice that the units of G is Tesla per meter (T/m), units of the gyromagnetic ratio is Hertz per Tesla (Hz/T) and hence multiplying and taking the time integral gives the unit $1/m$ for the k vector in each dimension. Such a unit lends itself to an interpretation as a "spatial frequency", which we will use in the following.

Now we imagine a scenario where the spins were perturbed equally in space using a radio-frequency pulse. Due to this pulse, the magnetization vector will be moved away from its equilibrium and then start to undergo the previously described relaxations. As the relaxation is spatially dependent, the resulting signal will also be dependent on space. Let us denote this as I(s). However, the signal ξ acquired by the coil is the sum of the signal from all spatial positions. Considering this, the acquired signal for a given k vector can be written as a spatial integral

$$\xi(k) = \int I(s) e^{i2\pi(s \cdot k)} ds, \tag{2.8}$$

which we recognize as a spatial Fourier transform with the k vector as the spatial frequency, i.e. $\xi(k) = \mathcal{F}(I(s))$.

The advantage here is that it is possible to generate the desired k vectors by adjusting the amplitude of the gradient G and switching it on/off for the necessary duration. The rest is taken care of by the fact that Fourier transform is an orthogonal transform. That means, if we sample the signal ξ in the k-space at a rate that adheres to the Nyquist criterion for the desired spatial resolution, we can uniquely reconstruct the spatial signal I(s) from its frequency components, i.e.

$$I(s) = \mathcal{F}^{-1}(\xi(k)). \tag{2.9}$$

This idea lies at the heart of Fourier imaging in MRI.

Conventionally, the space s and consequently the spatial frequencies k are discretized to allow a finite representation of the imaged object. As the Fourier transform is a linear operation and can be written as a discrete

convolution matrix, the above relationship then can be written in the linear algebraic form as

$$x = F^{-1}y, \tag{2.10}$$

with $x \in \mathbb{C}^N$ as the discretized complex image I, $y \in \mathbb{C}^N$ the discretized measured k-space signal ξ and $\mathrm{F}^{-1} : \mathbb{C}^N \to \mathbb{C}^N$ as the inverse of the Fourier transform operation. Notice that F is a unitary matrix and its inverse is equal to its Hermitian transpose: $F^{-1} = F^H$.

Now, looking at the imaging procedure more closely reveals the challenges of Fourier imaging with respect to the total acquisition time. For one of the spatial dimensions, usually referred to as the frequency encoding direction, it is possible to acquire the signal while the gradient is switched on. That is, as the gradient is on and the k values are being swept during the signal acquisition, the sampling of a full line of k-space values is possible in this acquisition. However, this idea only works for one of the spatial dimensions as it needs to sample throughout the time dimension, of which there is only one. For the other two spatial dimensions, the so-called phase encoding idea is used. Here, the gradients are switched on for a duration to produce the prescribed k value and then switched off again prior to the acquisition. This allows the nuclei to acquire the necessary phase before the acquisition and their signal to be placed into the prescribed k-space position. Then, this is repeated as many times necessary to sample the signal for all of the necessary k-space positions. However, prescribing a k value and acquiring the signal specific to this k-space position is a procedure that takes a certain amount of time, the calculation of which we will skip in this thesis. Repeating this many times, then, increases the total time of the acquisition of the full k-space and becomes the limiting factor of speed in MR imaging.

2.2 UNDERSAMPLED MR IMAGING

Many approaches have been proposed in order to overcome this limitation. Some of these rely on the properties of the k-space, such as Partial Fourier/echo encoding [13]. Here the idea is that for an image that is real, e.g. a structural MR image that only has a magnitude and a negligible phase, the k-space is conjugate symmetric about its center. This means the information in the k-space is redundant. Following this partial Fourier/echo imaging acquires only the half of the k-space in the phase encoding/frequency encoding directions and obtains the missing parts from the measured portions. In reality more than the half is acquired due

the imperfections in the imaging, but applying this idea ideally gives up to two fold acceleration in imaging time.

Another approach, in which we are interested in this thesis is the so-called undersampling of the k-space. Here, the idea is to acquire not the entire k-space, but rather leave out some of the phase encoding lines to save time. The acquired k-space can then be written mathematically by introducing the undersampling operator $U : \mathbb{C}^N \to \mathbb{C}^M$, with $M < N$ as

$$y = UFx. \tag{2.11}$$

Here the U is simply an operator that deletes some of the entries from the k-space representation of the object Fx, which, for example, can be implemented as removing rows from the identity matrix. In this formulation, however, the full imaging operator becomes $UF : \mathbb{C}^N \to \mathbb{C}^M$, which is not full rank any more, implying it has a non-trivial null space. That means the simple inversion operation is not valid anymore to obtain the image x from the measured $y \in \mathbb{C}^M$ uniquely. Instead, under this imaging operator, there are infinitely many x that can lead to the same measured y. In the following we analyse this a bit more in depth and look at hardware related techniques or techniques from the signal recovery literature on this situation is tackled.

2.3 THE UNDERDETERMINED IMAGING OPERATION

One way of obtaining an x is to define the pseudo-inverse of $F_U := UF$, taken to be a matrix in $\mathbb{C}^{M\times N}$, which is given as

$$F_U^\dagger = F_U^H[F_U F_U^H]^{-1} \tag{2.12}$$

in the underdetermined case ($M < N$) [17]. In this case the recovered x is given as $\hat{x} = F_U^H[F_U F_U^H]^{-1}y$. Applying the imaging operation onto $\hat{x}$ yields

$$F_U\hat{x} = F_U F_U^\dagger y = F_U F_U^H[F_U F_U^H]^{-1}y = [F_U F_U^H][F_U F_U^H]^{-1}y = I_M y = y \tag{2.13}$$

with I_M as the identity matrix with m elements reveals that $\hat{x}$ is indeed a solution in accordance with the measurement y, though it is unfortunately not the only solution. However, $\hat{x}$ has the property that it is the solution with the smallest Euclidean norm among all possible solutions. To understand this better, we can look at different inverses as well, which can be generated using the formula

$$(F_U^\dagger)_R = F_U^\dagger + \left[I_M - F_U^\dagger F_U\right] R \tag{2.14}$$

for any arbitrary matrix R. In this formulation any arbitrary matrix generates a solution $\hat{x}_R = (F_U^\dagger)_R y$. One can verify this again by applying the imaging operation on this as

$$F_U \hat{x}_R = F_U F_U^\dagger y + F_U R y - F_U F_U^\dagger F_U R y = y + F_U R y - F_U R y = y. \quad (2.15)$$

This analysis shows that for the low-rank undersampled imaging operator there are infinitely many solutions and provides a way of generating these. However, most of these solutions, though purely mathematically valid solutions, are not necessarily valid MR images.

Here, one can propose multiple ways of tackling this problem. The first and simplest idea could be to generate multiple images using the procedure above and then assess the validity of the images as proper MR images. This approach, however, proves to be the most difficult one. Firstly, it would be very inefficient as the valid MR images cover only a comparably very small subset of this whole solution space. Secondly, assessing the validity of the generated images is not an easy task. An alternative way can be identified as using hardware related tricks to transform the problem into an (over-)determined problem, which we look into in the next section. The third way is to introduce some prior information regarding MR images into the inversion process in order to make generated solutions valid MR images.

2.4 MULTI-COIL IMAGING AND (OVER-)DETERMINED IMAGING

Now we turn our attention to a multi-coil acquisition framework, called sensitivity encoding [18] (SENSE). Here the idea is to acquire the signal with multiple coils instead of a single one, e.g. with 8 or 16 coils, where each of the coils has a different spatial sensitivity. The coil sensitivities are in general complex, meaning they manipulate the phase and the magnitude of the signal depending on the position of the signal source. The sensitivity of the γ'th coil can mathematically be written as an operator $S_\gamma : \mathbb{C}^N \to \mathbb{C}^N$ and the operator for all coils together as $S : \mathbb{C}^N \to \mathbb{C}^{N\gamma}$ which act on the image x. Incorporating the coil sensitivities into the imaging operation yields

$$y = UFSx \quad (2.16)$$

for the measured k-space $y \in \mathbb{C}^{M\gamma}$. The difference to Eqn. 2.11 is that, here, dimensions of y can be bigger than or equal to the dimensions of the image, i.e. $\mathbb{C}^{M\gamma} \geq N$. This means that the imaging operation can become fully or even over-determined. This requires, however, the coil sensitivities to be

sufficiently linearly independent. On the other extreme, for example, if the coil sensitivities are equal to each other everywhere, then using multiple coils would not add any information into the imaging system. In reality, the situation is somewhere in-between and is effected by multiple design considerations.

This form of the imaging operation is generally referred to as the encoding operator, given as $E = UFS$, whose rank ultimately determines the structure of the inverse problem.

After introducing the basics of the MR imaging, we now present different reconstruction techniques in the following.

2.5 PRE-DEEP LEARNING APPROACHES

Here we look at different approaches from the pre-deep learning era. We separate these approaches into two categories, conventional signal recovery approaches and compressed sensing based approaches. Though there are different conventional approaches exploiting spatio-temporal correlations in the signals [19, 20]. We do not try to give a long overview, rather focus on a certain example and relate it to the Bayesian approach. For the compressed sensing part we overview some of the sparsity constraints in the literature and again relate them to the Bayesian statistics.

2.5.1 *Compressed sensing*

Here we look at compressed sensing (CS), which has provided MR reconstruction with a framework for creating diverse reconstruction algorithms for many different imaging applications. The first application of the CS theory to the undersampled MRI came in 2007. Interestingly, the first paper to apply CS to MRI was by Block, Uecker and Frahm [21] even before the paper of Lustig, Donoho and Pauly [22].

Prior to the development of CS, the consensus was that one needed to adhere to the Nyquist rate for the sampling for a perfect recovery of the underlying signal. CS, on the other hand, revealed that given certain properties of the signal and signal acquisition, a perfect recovery is possible even when the Nyquist criterion is violated. First of these necessary properties is that the signal to recover has to be sparse in some domain. That means, when the signal is transformed into a new domain by some sparsity inducing linear transform, it can be represented with a low number of coefficients. Furthermore, the sparser the signal, the less measurements it requires for

the recovery. Second requirement is that the artifacts resulting from the undersampling have to be noise-like in the domain where the image is sparse. Given these, perfect recovery of the underlying signal is possible if a proper reconstruction algorithm is used, that enforces this sparsity. This enforcement is usually achieved in the form of a regularization of the l_1 norm of the sparse coefficients.

As can be identified from the previous paragraph, the compressed sensing approach has three avenues for research: i) design of the sparsity inducing transform to make the signal more sparse to either increase the undersampling or aid the recovery, ii) design of the imaging procedures to enforce more noise-like (incoherent) artifacts and iii) algorithm design for improved recovery of the signal and improved speeds for achieving this.

For the second of these avenues, the most straightforward approach is to use a random undersampling in the Cartesian case or use radial or spiral patterns, which already yield a high level of incoherency.

The more heavily researched aspects of CS MRI are the first and third. Designing more sparsity inducing transforms is an effort, which requires understanding the spatial or spatio-temporal nature of the images. Furthermore, improving the sparsity directly improves the quality of the recovered signal or allows for increased undersampling. Similarly algorithm design is also an essential part as the algorithms realize the theoretical potential of the methods as well as determine the usability due to factors such as reconstruction time.

Here we look at the basic idea of the simpler forms of CS image reconstruction, namely using the l_1 and total variation (TV) as the sparsity scores of the images. In the former, the assumption is that the image itself is actually sparse in the image domain. This holds true for instance in angiography. For more typical images, the image itself is not sparse, however its gradients exhibit sparsity. TV [23] is a way of quantizing the gradients given as [21]

$$TV(x) = \sum_s \sum_i |D_{s_i} x(s)| \tag{2.17}$$

with $|.|$ as the absolute value, the outer sum over all pixel positions, the inner sum over the spatial dimensions and D_{s_i} the finite difference operator in the s_i direction, calculating the gradient of the image in that direction (here we think of the image x not as a vector, but a function $\mathbb{R}^3 \to \mathbb{C}$).

One observation is that the inner sum is essentially the l_1 norm of the finite differences, i.e. $\sum_i |D_{s_i} x(s)| = ||D_s x(s)||_1$ with D_s as the full finite

differences operator. Then a more general form of a function for expressing preference on images $R(x)$ similar to Eq. 2.17 can be written as

$$R(x) = ||\Phi x||_1. \tag{2.18}$$

The Φ operator can be any sparsifying operator, with the condition that the undersampling artifacts are incoherent in the transformed domain. In the case of TV, for instance, Φ is the finite differences operator as above.

The idea is that the function $R(x)$ defines a mathematical preference on the images and that one can use this in a constrained optimizations setting. The problem then is written as

$$\arg\min_x R(x) \;\; \text{s.t.} \;\; ||Ex - y||_2^2 < \epsilon. \tag{2.19}$$

The second term is the so-called data consistency, ensuring that the solution does not deviate from the measurements more than a prescribed ϵ.

Then one can write this using a Lagrange multiplier as

$$\arg\min_x ||Ex - y||_2^2 + \lambda R(x), \tag{2.20}$$

which gives the conventional form for the regularized inversion.

This problem can be solved using different optimization algorithms. The data consistency term is a linear convex function of x, where simple algorithms are usually sufficient. On the other hand, the regularization term R can be non-linear or non-convex, mostly dictating the choice of the optimization method. In case the problem also possesses these nice properties, simpler methods such as gradient descent or conjugate gradients can be used. If the regularization is non-linear as in the case of l_1 norm, for instance, one can use methods such as non-linear conjugate gradients or l-BFGS [24], where the latter improves convergence speed by conditioning the optimization. For applications involving a time dimension as in dynamic imaging, one can also introduce different regularizations for space and time [25]. In such cases, where the optimization of the data consistency and regularization terms together is difficult, one can employ variable splitting to yield an augmented Lagrange formulation and convert the complex problem into simpler sub-problems, which can be solved in an iterative manner as in the case of ADMM [26] or split Bregman methods [27], for instance.

Now, again, following our program of motivating Bayesian statistics, we look at an alternative derivation of the above formulation. Firstly we assume

the data is distributed around the true measurement with Gaussian noise as

$$y = Ex + \eta, \tag{2.21}$$

where $\eta \sim N(0, \Psi)$ i.e. both its real and imaginary part follow the Gaussian distribution. This is a valid modeling assumption as the noise on the coils is mostly from the thermal noise which can be modeled with a Gaussian [13]. For simplicity purposes, we assume isotropic noise, i.e. $\Psi = \sigma_\eta I$. In reality, the noise might have some correlation especially in multi-coil acquisitions, which can be corrected by pre-whitening. Here, in order to keep things simple, we neglect these. Using this, the probability distribution of y given x, termed the data likelihood, can be written as $p(y|x) = N(Ex, \sigma_\eta)$.

Under this noise model the data likelihood becomes

$$p(y|x) = \mathcal{N}(y|Ex, \sigma_\eta) = \frac{1}{(2\pi\sigma_\eta^2)^{M/2}} e^{-\frac{1}{2\sigma_\eta^2}(Ex-y)^H(Ex-y)}, \tag{2.22}$$

where H denotes the Hermitian transpose and σ_η is the standard deviation of the noise. In reconstruction, the quantity of interest is the posterior distribution $p(x|y)$, i.e. the probability of the image being x given the k-space measurements. Then we can write the Bayes theorem for the log-probabilities as

$$\log p(x|y) = \log p(y|x) + \log p(x) - \log p(y). \tag{2.23}$$

Here, $p(x)$ is the prior term and represents the information one has about the fully sampled image before the data acquisition.

The approach we use here for the reconstruction problem is to the maximum a posteriori (MAP) estimation given as

$$\arg\max_x \log p(x|y) = \arg\max_x \left[\log p(y|x) + \log p(x)\right]. \tag{2.24}$$

This expression is closely related to the conventional regularized formulation in Eqn. 2.20. To see this, first we take the log of the Gaussian likelihood term yielding

$$\arg\max_x \log p(x|y) = \arg\max_x \left[-\frac{1}{2\sigma_\eta}\|Ex - y\|_2^2 + \log p(x)\right]. \tag{2.25}$$

Finally, taking the maximum (or equivalently taking the minimum of the negative of the expression), defining the constant $\lambda \triangleq 2\sigma_\eta$ and multiplying both terms with it recovers the conventional formulation in Eq. 2.20 with

the data consistency and the regularization term that is weighted by the trade-off parameter λ

$$\hat{x} = \arg\min_{x} \left[\|Ex - y\|_2^2 - \lambda \log p(x) \right]. \tag{2.26}$$

Furthermore, we show that we can choose a log prior term to recover the CS regularization as well. To demonstrate this, we choose the prior term as follows

$$p(\Psi x) = Lap(0, 1/\lambda) = 2\lambda e^{-\lambda ||\Psi x||_1}, \tag{2.27}$$

where Lap denotes the Laplace distribution and $||.||_1$ is the l_1 norm. Then taking the logarithm yields

$$\log p(\Psi x) = -\lambda ||\Psi x||_1 + C, \tag{2.28}$$

with C as some constant term w.r.t. x, recovering the regularization term in Eq. 2.18. This completes the connection between the two approaches, namely the regularization and the Bayesian.

2.5.2 *k-t SENSE reconstruction from the Bayesian framework*

As previously stated, two avenues in pre-deep learning accelerated imaging are the CS approaches with random undersampling and approaches that use regular undersampling and exploit the spatio-temporal correlations in the signal for reconstruction. After having demonstrated how the CS based approaches fit into a Bayesian setting in the previous section, now we attempt the same here for a method from the latter category.

To this end we turn our attention to a particular SENSE based reconstruction method for dynamic MRI, called k-t SENSE [28]. The problem setting is to accelerate imaging by acquiring only a portion of the k-space in each temporal frame (i.e. the k-t space) using regular undersampling, which leads to a point spread function in the image-frequency space (i.e. the x-f space) causing the aliasing. The idea is to use some estimations about the voxel intensities, which informs the de-aliasing algorithm how much signal in proportion it should expect for each of the aliased pixels. This can be achieved by modeling the covariance matrix for the pixels M. These estimated values in M are taken from a low resolution version of the image, which takes less time to acquire. Furthermore the images in the x-f space can be assumed to be distributed around a mean image $\overline{x_f}$. Then the

authors describe a reconstruction formula to recover the image in the x-f space x_f as

$$x_f = \overline{x_f} + (E^H \Psi E + M^{-1})^{-1} E^H \Psi (y - E\overline{x_f}), \tag{2.29}$$

where Ψ is the noise covariance matrix and E is the encoding matrix. The authors derive this equation for the general case where the coil sensitivities can be time dependent, however, here we assume these to be constant.

As the authors also make a remark, this reconstruction equation can be derived by formulating the problem in the Bayesian framework. Here we shortly do this in accordance with our general program of motivating the use of Bayesian framework. We are interested in the posterior $p(x_f|y_t)$ of the image x_f in the x-f space given the measurements y_t in the k-t space. We incorporate the prior information into the formulation as a Normal distribution $p(x_f) = N(\overline{x_f}, M^{-1})$. We also model the data likelihood as a Normal distribution $p(y_t|x_f) = N(Ex_f, \Psi)$ where the encoding matrix E incorporates an additional temporal Fourier transform. Using Bayes' theorem as $p(x_f|y_t) \propto p(y_t|x_f)p(x_f)$ and the fact that the Normal distributions are conjugate to each other yielding another Normal for the posterior, we arrive at an expression for the mean of x_f as

$$\mu_{x_f} = (E^H \Psi^{-1} E + M^{-1})^{-1} (E^H \Psi^{-1} y + M^{-1} \overline{x_f}) \tag{2.30}$$

$$= (E^H \Psi^{-1} E + M^{-1})^{-1} E^H \Psi^{-1} y + (E^H \Psi^{-1} E + M^{-1})^{-1} M^{-1} \overline{x_f} \tag{2.31}$$

Rewriting the second term in Eqn. 2.31 as

$$\left[I - (E^H \Psi^{-1} E + M^{-1})^{-1} E^H \Psi^{-1} E \right] \overline{x_f} \tag{2.32}$$

and rearranging the terms yields

$$\mu_{x_f} = \overline{x_f} + (E^H \Psi^{-1} E + M^{-1})^{-1} E^H \Psi^{-1} (y - E\overline{x_f}), \tag{2.33}$$

which recovers the derived expression in Eq. 2.29 (Eq. 11 of the paper).

2.6 SUMMARY

In this chapter we gave the basics of the MR imaging such as signal generation, the concept of k-space and the mathematical formulation of the imaging and defined the ill-posed imaging problem. We then showed how some conventional approaches can be also recovered under the Bayesian

framework and motivated the use of this framework. The next chapters will build on this and show how one can solve the inverse problem by characterizing the posterior, first by finding its mode as the MAP estimation and second by taking samples from it for a more thorough characterization.

3

RECONSTRUCTION AS MAP ESTIMATION

This chapter is based on the publication "MR Image Reconstruction Using Deep Density Priors" [29] where the related works section has been extended for a more comprehensive summary of the literature including articles published after the preparation of the original manuscript. The author of the thesis is the main contributor to this project.

As discussed in Chapter 2 acquisition time in magnetic resonance (MR) imaging is directly related to the number of samples acquired in k-space. For high quality images, a large number of samples, and therefore long acquisition times are necessary. Reducing acquisition time and still reconstructing a high quality image in a reliable manner is an important question in MR imaging. In this chapter we tackle the problem of reconstruction as a MAP estimation problem.

3.1 RELATED WORKS

First we extend the discussion in Chapter 2 and provide a more complete literature review focusing on the deep learning approaches.

Many methods exploiting different properties of the k-space and hardware design for sampling a smaller portion of the k-space have found wide use in clinical practice. For instance, partial Fourier encoding [30] in the first category can save up to two fold acceleration. However the conjugate symmetry in the k-space it relies on is a fixed property and the idea cannot be pushed further for higher acceleration. Similarly the SENSE based acceleration methods also have intrinsic limitations. Firstly, the ultimate achievable SNR in case of ideal coil design is bounded by the principles of electrodynamics [31]. On top of that, using a finite number of non-ideal coils decreases the bound further [32]. Secondly, the noise in reconstructed images from undersampled acquisitions becomes amplified in a spatially dependent manner compared to the the full acquisitions. This amplification ratio is given by the so-called geometry factor, which heavily depends on the design of the coil array [18, 33].

Due to these limitations to improvement for the above mentioned methods, the question also received considerable attention from the image and

signal processing research communities to push the realistic undersampling factors even higher. The main focus here was reconstruction methods from randomly or regularly undersampled k-space acquisitions. The random undersampling approach was primarily motivated by the compressed sensing framework [22, 25, 34–36], where the incoherence between k-space sampling and some sparsifying transform was exploited as discussed in Chapter 2. Some works used hand-crafted sparsifying transformations, such as gradient in total variation [23] (TV) and wavelet decomposition [37, 38], while others used dictionary learning to determine the transformation from data [39, 40]. An alternative approach in [41] used block matching and 3D filtering, based on the BM3D denoising method [42], to tackle reconstruction as a denoising problem and exploit similarity of image patches to determine a sparsifying transform (BM3D-MRI). On the other hand, regular undersampling schemes and corresponding reconstruction algorithms were also extensively investigated [28, 43–46], one of which we described in Chapter 2.

The common aspect in all of these approaches, notably except BM3D-MRI, is that they invert underdetermined and ill-posed systems of equations by using good regularizers, which can be viewed as introducing explicit prior information on the expected structure that helps eliminate artifacts. This falls nicely into the Bayesian framework and we have shown this connection for some of the methods explicitly in Chapter 2.

Recently, researchers started to employ deep neural networks (DNN) [47] for MR reconstruction [48]. There are multiple main approaches which are being explored. The first is to learn a mapping from undersampled images to the fully sampled versions. For a given undersampling pattern, [49–51] proposes to learn a feed-forward network and during test time feeds the undersampled image to the network. Similarly, authors in [52] propose to train a network to remove spiking artifacts in computed tomography images and modify the mapping for MR using domain adaptation. The feed-forward network strategy, however, cannot guarantee data consistency at test time. To address this, authors in [53] added explicit data-consistency. Here, a mapping block, based on Convolutional Neural Networks (CNN), is followed by a data consistency block and the dual block structure is cascaded. This strategy guarantees consistency with measured data during test time while the CNNs perform de-aliasing. Alternatively, authors in [54] propose a model where a feed-forward mapping is learned and during test time its output is used as a regularization term alongside the

data consistency and other regularizers. The aim is to make sure the final reconstruction does not deviate too much from the mapping output.

The second DL approach leverages networks to improve existing iterative reconstruction algorithms in terms of reconstruction quality, called the unrolling approach. In [55] the authors showed that the iterations of the alternating direction method of multipliers (ADMM) algorithm can be unrolled as a multi-layer CNN. Instead of being fixed as in the original method, kernels and non-linear functions are parameterized and learned by the network, which improved reconstruction accuracy. A similar method, namely the Variational network was presented in [56] applying the idea for a gradient descent based optimization scheme. The idea was developed in [57], which used the same principle as [55] for improving diffusion filtering. The key idea here is that diffusion on the image, which is a widely used process for denoising, can be written as successive applications of fixed kernels and non-linearities. Similar to the ADMM-Net, the Variational Network approach also parameterizes and learns these throughout training. From this perspective these methods can be seen as generalized denoising methods. The same strategy of expressing an existing iterative method as a convolutional network was also used in [58–60] and in [61] using recurrent neural networks. Algorithms taking either of the DL approaches also integrate prior information on the structure of fully sampled images to account for missing k-space measurements but in an implicit manner, embedded in the trained weights of the feed-forward networks.

A third approach connected to the previous two is using an adversarial loss to improve the reconstruction quality in addition to the more conventional l_1 or l_2 types [62, 63]. This is achieved by formulating the reconstructing network as the generator of a generative adversarial network (GAN), which usually takes the undersampled data as input, and training the discriminator to differentiate between real examples and the generated ones. This way, the discriminator loss becomes a measure of how realistic the reconstructed image is and leads to improvements. The idea can be extended to use cyclic-GANs as well [64].

A fourth avenue to recognize here is the unsupervised learning based reconstruction, which developed concurrently with the manuscript on which this chapter is based. Here the idea is to use a generative model to learn the distribution of fully sampled images, instead of a mapping from the undersampled image to the fully sampled image. This allows then to have a reconstruction module that is independent of the undersampling. One example of this using generative adversarial networks (GAN) is [65],

where the authors describe a method to learn a generator that can generate MR images starting from simply distributed latent vectors. The idea is that, after training, the span of this generator is the space of valid MR images. Then the reconstruction is done in the latent space as looking for the latent vector which minimizes the l_2 distance between the measured data and the generated image from this latent vector in the undersampled k-space. As, however, the generator does not perfectly approximate the space of MR images, the generator output for this minimizer latent vector does not recover the proper MR image. Then the authors introduce a further test time optimization of the generator parameters for the particular image at hand, which recovers MR image closer to the ground truth. This extra optimization, hovewer, requires early stopping to prevent overfitting to the noise in the image, becoming instable and in the extreme case becoming the identity mapping. This issue of constraining the span of the generator to the space of MR images (or the "manifold") has been also investigated in [63], where the authors explicitly project the output of the generator to the null space of the imaging operation. This, however, requires knowing the imaging operation at training time.

A different avenue investigates another relevant problem setting, that is the case where one does not have access to a big dataset of images, on which to train the algorithms. Such settings cover, for instance, cardiac imaging, where due to the motion of the heart it is impossible to obtain fully sampled images. Another example can be the dynamic contrast enhanced imaging, for which it can be impossible to obtain fully sampled images at the desired time resolution. To solve this problem, the authors in [66] propose a Wasserstein GAN [67] based approach. They feed the undersampled images into a generator which outputs an image. This image is afterwards transformed into k-space using a different undersampling pattern and back to image space. The discriminator then tries to discriminate between a real measurement and this image. By using multiple different undersampling patterns after the generator, the generator is forced to learn to generate a realistic image. Another approach developed for this setting uses a subset of the measured k-space data to train a network which is then used for reconstructing the whole measured k-space [68, 69]. Similarly, [70] learns an unrolled optimization network from only undersampled images with the "hope" to learn a regularizer from a large-enough dataset which has enough diverse samples.

All the works mentioned above use the convolutions in the image space with limited kernel sizes. This is because convolutions are good in capturing

local correlations and images have such local correlations. On the other hand, due to the nature of the Fourier transform these local correlations become global correlations in the k-space, which render convolutional operations with small kernel sizes less useful. However, in the case of multi-coil imaging, such local correlations between the k-spaces of the coils is induced. This has been investigated, for instance, in [71], where the authors show that when the multi-coil k-space data is re-organized into a particular matrix structure, it yields a Hankel matrix and this matrix has a low rank structure due to the correlations data points from the same coils. Then reconstruction can be written as a low-rank matrix completion in k-space and the method revokes the need for a fully sampled k-space center for estimating the coil sensitivities. Interestingly, in [72] it was shown that the local k-space neighborhoods also admit a low rank structure when the image has limited spatial support or a slowly varying phase even for a single coil, which can be leveraged for reconstruction. Then [73] analysed this idea further and offered a unification of multi-coil imaging with compressed sensing based MR (see also [74] for an extended discussion relating the use of convolutional neural networks in inverse problems to representations with Hankel matrices). This approach has also been investigated using deep learning. For instance [75] uses convolutional neural networks to capture the correlations in multi-coil k-space data acquired with a fully sampled center. Another work, building on [73] and applying the convolutional neural networks in the k-space is [76].

Another approach has been to adapt the deep image priors to MRI reconstruction, which have been shown to yield good results for similar tasks such as denoising or in-painting [77]. Here, for instance, [78] applied this idea to cardiac imaging with an extension to encode temporal information, which allowed temporal interpolation between frames for cine images as well. However, with this approach early stopping becomes an important aspect to avoid overfitting to the measured data [77, 79] (an interpretation of the deep image prior is that it acts as a spectral low pass filter when used with early stopping [80]). Alternatively, in [81] the authors propose to fine tune a network in a deep image prior manner, which was initially trained on healthy subjects. The idea is then the unsupervised fine tuning adapts the network to previously unseen structures, improving reconstruction quality.

Finally, for completeness, we mention reconstruction approaches for MR fingerprinting (MRF) [82, 83]. MRF is a technique that randomly changes the acquisition parameters during the measurement and matches the measured signal to the entries in a pre-computed dictionary of signal evaluations for

given acquisition and tissue parameters. By matching the measurements with given acquisition parameters, it recovers multiple tissue parameters at the same time. Thus one obtains multiple maps of different tissue parameters instead of an image with a single weighting (e.g. T1, T2 weighting). The authors of [84] show, for instance, that by learning a mapping from pixel values acquired with different parameters to the tissue parameters, they can speed up the reconstruction of the tissue parameter maps. Furthermore, as the mapping can be compactly represented, it reduces the size requirements compared to the dictionary approach. However this methods does not take the known signal model based on the Bloch equations into account. Other similar methods have been proposed [85], for instance, based on recurrent neural networks [86] or methods that take the Bloch equations into account as well [87].

3.1.1 *Proposed approach*

In this work, we employ a neural network as an explicit prior similar in essence to the non-DL based reconstruction methods proposed in the signal processing communities, the difference being the power of the prior model. Unsupervised learning with DNNs has been very successful in approximating probability distributions of high dimensional data, including images, from a set of samples. One such approach, of particular interest to the method proposed here, is the variational auto encoder (VAE) algorithm [88, 89]. Using VAEs, it is possible to approximate the distribution of MR image patches and likelihood of a previously unseen image. Furthermore, the approximate likelihood function is a network and therefore differentiable. These two aspects allow using a VAE as a prior model that can approximate distributions of large image patches, e.g. patches of 28x28 pixels, for iterative reconstruction.

We propose a novel probabilistic reconstruction method that uses priors learned via VAEs, which we term as Deep Density Prior (DDP) based reconstruction. We formulate a Bayesian model of the imaging process, including the prior and a data consistency term that embeds the encoding operation, and express DDP as the Maximum-A-Posteriori (MAP) estimation. Compared to non-DL based methods, the main difference of DDP is the powerful prior that can capture distribution of large image patches. Compared to DL approaches, the fundamental differences of DDP are: (i) the explicit prior, which is trained to capture the structure of the fully sampled images as opposed to the implicit prior in feed-forward networks that is learned to

reduce artifacts seen during training; and (ii) the decoupling of the prior from the data consistency term. The latter difference leads to two theoretical advantages: (i) while previously proposed DL-based approaches required paired datasets for training, including undersampled and corresponding fully sampled images, DPP does not; and (ii) decoupling prior and data consistency terms eliminates possible sensitivities in accuracy to deviations in acquisition specifications between undersampled images used for training and encountered during test time, such as sampling parameters, coil settings and k-space trajectories, which for instance has been demonstrated to be the case in [49].

In the rest of the chapter, we first present the method and then show reconstruction results and comparisons with conventional approaches as well as recent DNN based methods.

3.2 METHODS

In the first two parts of this section, we summarize the a brief background on Bayesian formulation of the MR reconstruction problem and the VAE algorithm. We present our main technical contribution, learning a prior for MR patches and integrating it in the reconstruction problem, starting from Section 3.2.3.

3.2.1 *Bayesian formulation of the MR reconstruction problem*

Firstly we recite the notation from Chapter 2 introduce some new ones and reformulate the problem.

An MR image is denoted as $x \in \mathbb{C}^N$, where N is the number of pixels[1]. An imaging operation is given by an undersampling encoding operation $E = UFS$, where $S : \mathbb{C}^N \to \mathbb{C}^{N\gamma}$ is a sensitivity encoding operator. γ is the number of coils, $F : \mathbb{C}^{N\gamma} \to \mathbb{C}^{N\gamma}$ is the Fourier operator and $U : \mathbb{C}^{N\gamma} \to \mathbb{C}^{M\gamma}$ is an undersampling operator, with $M < N$. Let us also define $m \in \mathbb{C}^P$ as an image patch of P pixels extracted from x.

Assuming complex-valued, zero mean, normal distributed and uncorrelated additive noise, denoted as η, the acquired data $y \in \mathbb{C}^{M\gamma}$ can be

1 In this work we focus on 2D imaging, however, the same techniques can be applied to 3D imaging and this extension will be a part of future work.

modelled as $y = Ex + \eta$. Under this noise model the MAP estimate is given as

$$\hat{x} = \arg\min_x \left[\|Ex - y\|_2^2 - \log p(x) \right]. \tag{3.1}$$

In this work, we propose to estimate the prior term from examples of fully sampled images and approximate $-\log p(|m|)$, i.e. the negative log prior of magnitude of image patches, with a neural network model. We train a VAE on patches extracted from fully sampled MR images to capture the distribution and use this prior for reconstruction. This allows us to utilize the prior independent of the sampling operation in contrast to the feed-forward mapping approach.

3.2.2 *Learning the data distribution with VAEs*

VAE is an unsupervised learning algorithm proposed to approximate high-dimensional data distributions [88, 89]. We introduce VAEs very briefly and refer the reader to [88] for further details. VAE is a generic algorithm that can be applied to any signal but we focus on the magnitude image patches in our description.

The main goal of the VAE algorithm is to approximate the data distribution using a latent variable model and optimize its parameters for a given set of examples using variational approximation. The model is given as

$$p(|m|) = \int_Z p(|m|, z)dz = \int_Z p(|m||z)p(z)dz, \tag{3.2}$$

where $z \in \mathbb{R}^L$ denotes the latent variable, $p(z)$ the prior over the z's and $L << P$. A known distribution is assumed for $p(z)$, e.g. unit Gaussian, and a parameterized $p(|m||z)$ is optimized to maximize $\log p(|m|)$ of observed samples. This modeling strategy is also taken in other probabilistic latent variable models, such as probabilistic principal component analysis [90]. The VAE model parameterizes $p(|m||z)$ as a neural network whose set of parameters we denote with φ. To optimize $\log p(|m|)$ for the given samples, the integral over z needs to be evaluated and this is not feasible even for moderate L. Variational approximation uses an approximate distribution for the posterior $q(z||m|) \approx p(z||m|)$ to address this problem. Using $q(z||m|)$, $\log p(|m|)$ can be decomposed into two terms [90]

$$\log p(|m|) \;=\; \mathbb{E}_{q(z||m|)} \left[\log \frac{p(|m|, z)}{q(z||m|)} \right] + \mathrm{D}_{\mathrm{KL}} \left[q(z||m|) || p(z||m|) \right]. \tag{3.3}$$

The first term is referred to as the evidence lower bound (ELBO) and the second term is the Kullback-Leibler divergence (KLD) between the approximate and true posteriors. The KLD term is intractable because the true posterior $p(z||m|)$ is intractable. It is, however, always larger than or equal to zero, which makes ELBO a lower bound for $\log p(|m|)$. The strategy of VAE is to maximize the ELBO as a proxy to $\log p(|m|)$.

Similar to $p(|m||z)$, the VAE algorithm models $q(z||m|)$ as a separate neural network with parameters θ and during training optimizes both θ and φ to maximize the ELBO of the training samples. Rewriting the ELBO with $p(|m||z)$, $p(z)$ and $q(z||m|)$, the optimization for training can be written as

$$\max_{\theta,\varphi} \sum_{n=1}^{N} \text{ELBO}(|m^n|) = \max_{\theta,\varphi} \left[\sum_{n=1}^{N} \mathbb{E}_{q_\theta(z||m^n|)} \left[\log p_\varphi(|m^n||z) \right] - \text{D}_{\text{KL}} \left[q_\theta(z||m^n|) || p(z) \right] \right], \quad (3.4)$$

where m^n is the n^{th} training sample and we added the network parameters as subscript at the corresponding terms to indicate the dependence. Notice that the KLD term in Equation 3.4 is distinct from the one in Equation 3.3, it includes $p(z)$ instead of $p(z||m|)$, which makes it tractable.

The networks $q_\theta(z||m|)$ and $p_\varphi(|m||z)$ are typically called the encoder and the decoder, respectively. The former takes a data sample $|m|$ and encodes it into a posterior distribution in the latent space with network parameters θ. If the posterior distribution $q_\theta(z||m|)$ is modelled as a Gaussian, then the encoder outputs a mean and a covariance matrix for z depending on $|m|$. The decoder network on the other hand, takes a latent vector z and maps it to a conditional distribution of the data given z. During training, z vectors are sampled from $q_\theta(z||m|)$ to evaluate the expectations in 3.4. In this work, we use the original VAE design [88] except for the data likelihood, for which we use a multi-modal Gaussian $p_\varphi(|m||z) = N(|m||\mu_\varphi(z), \Sigma_\varphi(z))$ with a diagonal covariance matrix, similar to [91]. We note that the Gaussian distribution here is different from the data likelihood given in Equation 2.22. The Gaussian in Equation 2.22 models the complex valued observation noise whereas the one here models conditional distribution of $|m|$ given z in the essence of compound probability distributions. We provide further network design details in Section 3.2.5.

3.2.3 *Deep density prior (DDP) reconstruction model*

Once the VAE model is trained we can integrate the prior within a Bayesian formulation of the reconstruction problem as given in Equation 2.25. We make two key observations to achieve this. First, given by the theory, the ELBO($|m|$) can be used as a proxy to the true distribution $\log p(|m|)$. So, an approximate log likelihood of a magnitude image patch $|m|$ can be obtained by evaluating ELBO($|m|$)

$$\text{ELBO}(|m|) = \mathbb{E}_{q_{\theta^*}(z||m|)}\left[\log p_{\varphi^*}(|m||z) + \log \frac{p(z)}{q_{\theta^*}(z||m|)}\right], \tag{3.5}$$

where θ^* and ϕ^* are the optimal VAE parameters learned during training. The approximate log-likelihood allows us to formulate the proposed reconstruction model as the following MAP estimation problem

$$\arg\min_x \left[\|Ex - y\|_2^2 - \sum_{m_r \in \Omega(x)} \text{ELBO}\left(|m_r|\right)\right], \tag{3.6}$$

where $\Omega(x)$ denotes a set of (overlapping) patches covering the image x and $|m_r|$ is the magnitude of the r^{th} image patch. Note that this approach assumes independence between different patches, ignoring statistical dependencies between them though it would be possible to extend the model to achieve this.

Since an exact computation of the ELBO term requires evaluating the expectation with respect to $q(z||m|)$, which is computationally not feasible, we use a Monte Carlo sampling approach to calculate the ELBO as follows

$$\text{ELBO}(|m|) \approx \frac{1}{J}\sum_{j=1}^{J} \log p(|m||z^j) + \log \frac{p(z^j)}{q(z^j||m|)}, z^j \sim q(z||m|). \tag{3.7}$$

Here J represents the number of Monte-Carlo samples.

Plugging the ELBO approximation into Equation 3.6, we obtain the formulation of the proposed DDP reconstruction problem

$$\arg\min_x \|Ex - y\|_2^2 - \sum_{m_r \in \Omega(x)} \left[\frac{1}{J}\sum_{j=1}^{J} \log p(|m_r||z^j) + \log \frac{p(z^j)}{q(z^j||m_r|)}\right], \tag{3.8}$$

where $z^j \sim q(z||m_r|)$, the first term is the usual data term and the second term within the summation is the regularization term that arises from the learned prior.

Our second key observation is that the approximation in Equation 3.7 is differentiable since each term is defined through networks that are themselves differentiable. This is the critical aspect that allows integrating the trained VAE as a prior into an iterative reconstruction algorithm. We can compute the total derivative of the prior term with respect to each image patch as follows

$$\mathcal{R}(|m|, z^j) \triangleq \log p(|m||z^j) + \log \frac{p(z^j)}{q(z^j||m|)}$$

$$\frac{d}{dm}\left[\frac{1}{J}\sum_{j=1}^{J}\mathcal{R}(|m|, z^j)\right] = \frac{1}{J}\sum_{j=1}^{J}\frac{d}{dm}\mathcal{R}(|m|, z^j)$$

$$= \frac{m}{|m|}\left[\frac{1}{J}\sum_{j=1}^{J}\frac{\partial}{\partial|m|}\mathcal{R}(|m|, z^j) + \frac{\partial}{\partial z^j}\mathcal{R}(|m|, z^j)\frac{dz^j}{d|m|}\right], \quad (3.9)$$

where we defined $\mathcal{R}(|m|, z^j)$ for notational simplicity. The second term in the last line is due to the dependency of the samples z^j on m and $m/|m|$ is due to taking the derivative of the magnitude with respect to the image patch.

3.2.4 *Optimization using projection onto convex sets*

We solve the DDP optimization problem given in Equation 3.8 using the projection onto convex sets (POCS) algorithm [92], specifically using the formulation in [93]. POCS is an iterative minimization process, where the solution variable is projected sequentially onto different convex sets, each defined by one of the constraints in the problem.

The projection for the data consistency term is implemented using the method proposed in [92], which is $\mathcal{P}_{DC}x = x - E^H(Ex - y)$. When there are multiple coils, this projection implements SENSE reconstruction[2]. Since we do not have a projection operator for the prior term, we approximate it by several gradient ascent steps with a small step size α as in [93]. We use the final image at the end of the ascent steps as the projected image patch. We define the prior projection with the following steps: i) create a set of patches $\{m_r^t\} = \Omega(x^t)$ from the image x^t at iteration t, ii) obtain the derivatives for each of these patches using Equation 3.9, which have the same size as the patches themselves, iii) combine the derivatives of

2 For optimal signal-to-noise ratio, one needs to account for the noise covariance as well, which we ignore in this work for simplicity.

Algorithm 1 Deep density prior (DDP) reconstruction using POCS. See text for a more detailed explanation.

y: undersampled k-space data
E: undersampling encoding operator
VAE: the trained VAE
procedure DDPRECON(y, E, VAE)
    $x^0 \leftarrow E^H y$ ▷ initialize with the zero-filled image
    for $t = 0 : T - 1$ **do** ▷ main loop: POCS iterations
        $x^{t,0} \leftarrow x^t$
        for $k = 0 : K - 1$ **do** ▷ inner loop: iterations for the prior projection $\mathcal{P}_{prior}$
            $\{m_r^{t,k}\} \leftarrow$ image2patches$(x^{t,k})$ ▷ creates a set of patches covering the image
            for $r = 1 :$ no of patches **do** ▷ loop over all the patches in $\{m_r^t\}$
                $g_r \leftarrow \frac{d}{dm}\left[\frac{1}{J}\sum_{j=1}^{J}\mathcal{R}(|m|, z^j)\right]_{m=m_r^{t,k}}$ ▷ calculate the derivative acc. to Eq. 3.9
            $g \leftarrow$ patches2image$(\{g_r\})$ ▷ $\mathcal{P}_{prior}x^t$
            $x^{t,k+1} \leftarrow x^{t,k} + \alpha \cdot g$
            $x^{t,k+1} \leftarrow \mathcal{P}_{phase}x^{t,k+1}$ ▷ (optional)
        $x^{t+1} \leftarrow x^{t,K} - E^H(Ex^{t,K} - y)$ ▷ $\mathcal{P}_{DC}x^t$
    return x^T ▷ Resulting reconstruction

the patches to form a derivative image by averaging the values where the patches overlap, iv) update the image using the derivative image, v) repeat this K times. Notice that the set defined by the prior projection is not necessarily convex in $|m|$, however we have not encountered any problems in convergence during our experiments. To reduce edge effects resulting from patchwise projections, we use four sets of overlapping patches.

With the data consistency and prior projections defined as above, one step of reconstruction within the POCS framework becomes

$$x^{t+1} = \mathcal{P}_{DC}\mathcal{P}_{prior}x^t. \tag{3.10}$$

The prior term in the DDP method does not explicitly provide information on the phase, therefore, reconstruction of the phase is driven by the data consistency projection in the update equation above. For acquisition with multiple coils, the reconstruction method can recover a phase close to the true phase without any modification. For single coil acquisitions, to account for the less amount of information we use an additional projection $\mathcal{P}_{phase}$ in the update equation. A reconstruction iteration is then given as $x^{t+1} = \mathcal{P}_{DC}\mathcal{P}_{phase}\mathcal{P}_{prior}x^t$. We use the $||C\exp(i\angle x)||$ term [94] as the regularization with C as the finite difference operator, to prefer smooth phase images. We implement $\mathcal{P}_{phase}$ as taking 10 steps with a step size of 0.1 in the negative gradient direction of the regularization term. It is however possible to change this to other constraints on the phase depending on the application, such as a zero-divergence constraint for phase contrast flow imaging reconstruction [95]. It is also possible to extend the VAE model and train a prior for complex image patches with an appropriate training set.

We apply T POCS steps to complete the reconstruction. Algorithm 1 provides a summary of the reconstruction procedure.

3.2.5 *Details on VAE network architecture and training*

Both $p_\varphi(|m||z)$ and $q_\theta(z||m|)$ networks are depicted in Figure 3.1. For the encoding network, the input is an image patch and the output is the mean and the covariance of the corresponding posterior. The decoding network takes a vector z as input and outputs the mean and the covariance of the corresponding likelihood for the patch. Both networks are mostly convolutional with single fully connected layers. All convolutional layers for both networks use 3x3 kernels and all layers also have additive bias

terms throughout the network. We use rectified linear units (ReLU) as the non-linear activation function.

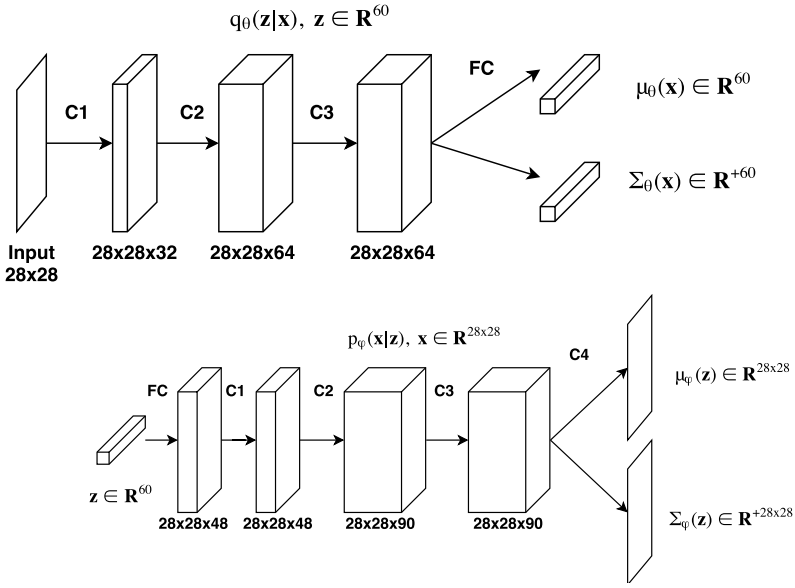


FIGURE 3.1: Architecture of the encoding (top) and decoding (bottom) networks of our VAE. Arrows indicated with *C#* are convolutional layers followed by ReLU non-linear activation except *C4* of decoding network, which is not followed by non-linear activation. Arrows indicated by FC are fully connected layers. FC of the decoding network is followed by a ReLU activation but not of the encoding network.

To avoid numerical stability issues we use the log of the variance values throughout the network. We initialize the network weights with a truncated normal initializer with standard deviation 0.05. We use Adam [96] for optimization (learning rate of 5e-4, default momentum values in Tensorflow).

As the base model, we used patches of size 28×28 and a 60 dimensional latent space. However, we also experimented with different patch size and latent dimensions.

3.2.6 *Experimental setup*

3.2.6.1 *MR image data*

We used structural images from three different data sources to demonstrate the proposed algorithm.

First, we used images from the Human Connectome Project (HCP) data set [97] (see https://www.humanconnectome.org/study/hcp-young-adult/document/500-subjects-data-release). The HCP images were acquired at a 3T Siemens device with 2400 ms, 2.14 ms and 1000 ms for TR, TE and TI, respectively, flip angle of 8 degrees, a standard field of view for all subjects of 224x224x224 mm^3 with 0.7 mm isotropic resolution using a bandwidth of 210 Hz/voxel. Fat suppression was used during acquisition. We used the minimally preprocessed images in order to also have the corresponding FreeSurfer [98] segmentations, which we use in our evaluation to compute contrast-to-noise ratio. The preprocessing steps consisted of resampling to a FOV matrix 260x311x260 and rigid alignment. Despite the rigid alignment, there were substantial local orientation differences between the images from different subjects. We used five central slices of the volumes, skipping four slices between each. We cropped the images to a size of 252x308, removing only background, to reduce computational load. The high quality and large number of images from the HCP dataset are ideal for learning priors with the VAE model. We took 2D slices from the T1 weighted 3D MPRAGE images from 158 subjects (790 images in total) to train the prior VAE model. We normalized the training images by mapping their 99^{th} intensity percentile to 1 per image slice. The VAE model was trained for 200k iterations with a batch size of 50.

Second, to verify that the proposed reconstruction method can be used on a domain that is different from the one the prior is trained on, we experimented with two slices from the Alzheimer's Disease Neuroimaging Initiative (ADNI) data set (for up-to-date information, see www.adni-info.org). The acquisition parameters of the ADNI images were different than HCP: TR, TE, TI values were 7.34 ms, 3.03 ms, 400 ms, flip angle was 11 degrees, FOV matrix was 196x256x256 with resolution 1.2 mm x 1 mm x 1 mm. The images were acquired with a 3T GE scanner with no fat suppression and were bias-field corrected with the N3 algorithm [99]. The images were selected from subjects with Alzheimer's disease and who have visible white matter lesions. Images with lesions allowed us to also test whether the proposed method will be able to faithfully reconstruct such lesions. We extracted the central slices that showed the largest lesions from these images

and further cropped the FOV to 168x224 to remove the empty regions in the images to accelerate computations.

Lastly, we acquired images of 8 healthy volunteers after written informed consent and according to the applicable ethics approval, using a similar protocol to the HCP and ADNI datasets with a 3T Philips Ingenia scanner. To acquire the images, we used a turbo field echo sequence with 1 mm isotropic resolution, flip angle of 8 degrees, TR and TE set to 8.1 ms and 3.7 ms, an inversion pulse with 1 s delay and a linear sampling order. Images were acquired at 3T with a Philips Ingenia device. The FOV was planned according to the volunteer and a head coil array with 16 elements for receiving was used. We acquired fully sampled complex k-space data for retrospective undersampling study, measuring the coil sensitivities using the standard Philips SENSE reference scan (for only 6 volunteers, due to technical issues) and also estimating them using ESPIRiT [100] (Code: http://people.eecs.berkeley.edu/∼mlustig/Software.html) with default parameters to obtain the autocalibrated coil maps.

3.2.6.2 *Setup and evaluation*

We used images from 17 HCP subjects (separate from training subjects), the ADNI images and the acquired data from 8 subjects in our evaluation. We retrospectively undersampled the test images in k-space, reconstructed them back and compared the results with the original images. For the acquired data, the fully sampled k-space data was present while for HCP and ADNI images, we used Fourier transform to compute the fully sampled k-space data. We experimented with varying undersampling (US) ratios, which we denote with R when presenting results.

We used Cartesian US in one dimension while fully sampling in the other dimension, corresponding to phase-encoding and readout directions, respectively. We present an example of US patterns in Figure 3.2. We generated the patterns by randomly sampling a 1D Gaussian distribution along the phase-encoding dimension. We randomly drew many sampling patterns from the Gaussian distribution and selected the ones with the best peak-to-side ratio. In addition, we added the central 15 profiles to these selected patterns to fully sample the low-frequency components. We used 2, 3, 4 and 5 for net US ratios (including the fully sampled center). In addition, we used 4-fold radial US with non-uniform fast Fourier transform (NUFFT) [101]. We used the implementation given in [102] with a conjugate gradient solver. For reconstruction in this case, we only changed the fast

Fourier transform (FFT) to NUFFT in the data consistency term and did not modify the prior projection.

At reconstruction time the undersampled images need to be in the intensity range for which the DDP prior is trained. To meet this requirement, we normalized the undersampled images to their 99^{th} percentile before reconstruction with DDP. Notice that this may not be the same value as the 99^{th} percentile of the fully sampled images, which would not be available in a real setting for test images.

While assessing our proposed DDP method, we generated a new random US pattern for each test image to make sure our empirical analyses assess the effects of the variability of the US patterns. We reconstructed the test images from HCP and ADNI using 30 POCS iterations (T=30) for R=2 and 3 and 60 for R=4 and 5, 10 iterations for the prior projection (K=10) and the step size was set to α=1e-4, where convergence was observed. We initialize the POCS algorithm using the zero-filled image. When reconstructing images from volunteers, we run 10 iterations of only data consistency projections initially and then turn on the prior projection. We also use lower number of iterations (T=5 for both R=2 and R=3) to prevent divergence due to discrepancy between the true and used coil sensitivities. We implemented the coil combination as given in [103] at each data projection step.

We compared the reconstruction with the fully sampled "ground truth" images using Root-Mean-Squared-Error (RMSE), Contrast-to-Noise-Ratio (CNR) and Contrast difference (CN) computed at the gray and white matter boundary in the brain. We use the FreeSurfer [98] to obtain the segmentations required for CNR and CN. Details are given in Section 3.2.6.4. We present these results in Section 3.3.

Additionally, we performed three experiments to test sensitivity of ADMM-Net to deviations in acquisition specifications between undersampled images used in training and test. Experiments in [49] provide evidence to this end for basic feed-forward networks. Authors showed that when k-space trajectories between training and test images differ, the feed-forward network's performance decreased. ADMM-Net is notably different than a basic feed-forward network as it integrates explicit data consistency in the feed-forward architecture. In these experiments we test the null hypothesis that using a different US pattern in reconstruction than in training does not result in a decrease in the performance in the RMSE sense. When measuring RMSE, we used a mask to only measure the reconstruction errors in the brain tissue and skull ignoring artifacts in the background. In each experiment, we trained two networks using different US patterns and applied

them on the same test images. In Experiment I, we used R=2 and R=4 for training and tested on R=2. In Experiment II, we used the same networks as in I but tested on R=4. Lastly, we trained networks with R=3 Cartesian and R=3 pseudo-radial patterns and tested on R=3 Cartesian. We used the 17 test images for the evaluation and performed paired Wilcoxon signed-rank tests to assess the null hypothesis.

We further experimented with different latent space dimensions, patch sizes and signal-to-noise ratio in the measurements to better characterize the proposed algorithm.

3.2.6.3 *Compared methods*

We implemented several methods to compare against our proposed approach. These are zero-filling reconstruction, total variation (TV) [22], dictionary learning (DLMRI) [39], ADMM-Net [55] ,BM3D-MRI [41], SIDWT [37], FDLCP [40] and PBDW [38]. Available implementations for the last three methods only ran on cropped FOV images (square image slices). To compare we used a corresponding DDP reconstruction on cropped FOV.

As the baseline, we evaluated zero-filling (ZF) reconstructions. This baseline represents a lower-bound of the reconstruction accuracy. The first method we used is the total variation (TV) reconstruction as described in [22]. We used the BART toolbox implementation that is publicly available [104], specifically the "pics" tool with TV regularization (regularization strength 0.075) and the ADMM parameter ρ as 1. We used 20 conjugate gradient steps and 4500 total iterations ("bart pics -R T:3:0:0.0075 -u1 -C20 -i4500"). The parameters above are chosen by a grid search for the best parameter setting in the RMSE sense.

As the second method, we used reconstruction using dictionary learning (DLMRI) as proposed in [39][3]. We used 200 iterations, a patch size of 36 voxels and 36 dictionary atoms. Furthermore, we set number of signals used for training to 7200 and the overlap stride to 2. K-SVD learning was used with both sparsity and error threshold. The sparsity level was set to 7. The error threshold was set to 0.046 for the first four iterations, then to 0.0322 for the rest of the iterations. We used 15 K-SVD iterations. We chose the parameters as suggested by the authors in the code, but increased the number of iterations.

Third, we used the ADMM-Net[4] [55] algorithm, a feed-forward neural network with data consistency term that unrolls ADMM iterations, designed

3 Code available at http://www.ifp.illinois.edu/~yoram/DLMRI-Lab/Documentation.html
4 https://github.com/yangyan92/Deep-ADMM-Net

for square images with radial undersampling. We modified the code to work with non-square images and Cartesian undersampling patterns. To train the model, we used the same 790 images that were used to train the VAE model. In order to correctly evaluate the method, we used the same undersampling pattern to train and test the method, except while performing sensitivity tests described in the main text. We normalized the image intensities in the same way as for DDP. We used 15 stages with 8 filters (filter size 3x3), and set padding as 1. We did not use weight decay during training. We set the maximum iteration number to 25 for the L-BFGS algorithm. It trained for maximum number of iterations for R=2 (45 hours), 13 iterations before convergence for R=3 (42 hours), 18 iterations before running out of time (120 hours) for R=4 and maximum number iterations for R=5 on a GPU (GeForce GTX TITAN X). The normalized mean squared errors were 0.078 and 0.035 for R=2, 0.11 and 0.071 for R=3, 0.15 and 0.11 for R=4, and 0.19 and 0.13 for R=5, before and after training, respectively, on the training set. For the radial undersampling patterns (R=3) used in the sensitivity tests, the network was trained for 36 hours before termination after 11 iterations. We took the parameter setting for which the best results were reported in the paper.

We also compared with SIDWT [37], FDLCP [40] and PBDW [38], which were reported to achieve highly accurate reconstructions. For these methods we had to crop the images to 256x256 and generate new undersampling patterns for this size since the authors implementations worked only with that size[5]. We did not modify the code for these methods, as only the binaries were available, and took the parameters as set by the authors in the code. We ran our proposed method on these images as well.

Lastly, we compared to BM3D-MRI[6] [41], a powerful reconstruction method that leverages redundancy in the images, and used the parameters as set by the author in the code.

3.2.6.4 *Performance Metrics*

Here we supply the expressions for the performance metrics used in the presented evaluation.

Normalized Root-Mean-Squared-Error (RMSE) is defined as:

$$\mathrm{RMSE}(gt, rec) = 100 \cdot \sqrt{\frac{\sum_i^N (|gt_i|^2 - |rec_i|)^2}{\sum_i^N |gt_i|^2}}, \tag{3.11}$$

5 Implementations from http://csrc.xmu.edu.cn/csg_publications_en.html
6 Code available at http://web.itu.edu.tr/eksioglue/pubs/BM3D_MRI.htm

where *gt* and *rec* are the fully sampled and reconstructed images. |.| denotes the magnitude operator. The subscript i goes through all the N voxels in the image. If a mask is used, then only the voxels in the masked region are used. The denominator provides the normalization factor.

Contrast-to-Noise Ratio (CNR) is given as:

$$\mathrm{CNR}(rec) = \frac{|mean_{GM}(rec) - mean_{WM}(rec)|}{std_{GM}(rec) + std_{WM}(rec)}, \tag{3.12}$$

where $mean_{GM}$ and $mean_{WM}$ denote the mean values of the voxels in the gray matter and white matter, and std_{xx} denotes the standard deviation in the respective tissue type. To compute these statistics, we use the whole GM but only the edge of the WM. In order to obtain white matter boundaries we applied binary erosion (with a square structuring element of size 7x7) to the white matter segmentations, as computed by FreeSurfer, and took the difference of the original and eroded segmentations. This gives us the voxels at the boundary of WM and GM but within WM.

Contrast-to-noise is similar, but without the division by the standard deviation:

$$\mathrm{CNR}(rec) = |mean_{GM}(rec) - mean_{WM}(rec)|. \tag{3.13}$$

3.3 RESULTS

We start by showing visual reconstruction results. Figure 3.2 shows results for one of the test images from the HCP data for R=3. The sampling pattern is also shown in the figure. Visual quality of reconstructions for different methods varied. TV and PBDW reconstructions had problems in restoring the structure and texture. Respective reconstructed images in Figure 3.2 appear cartoon-like and error images show higher error in regions where gray (GM) and white matter (WM) structures are intertwined. SIDWT did not complete dealiasing and ADMM-Net did not reconstruct the small GM island in the zoomed image. DDP, DLMRI and FDLCP perform well. We show four more randomly selected images from the test set in Figures 3.3, 3.4 for R=2 and 3.5, 3.6 for R=3 as well as reconstruction results for R=2,4,5 in Figures 3.7, 3.8 and 3.9.

We present the quantitative results for reconstruction accuracy in Table 3.1. We show results for full FOV and cropped images separately as the numbers are not comparable. In terms of RMSE our DDP method performed best for the full FOV reconstructions (except for R=2, where BM3D-MRI performed better) and second best against FDLCP for the cropped FOV setting. In

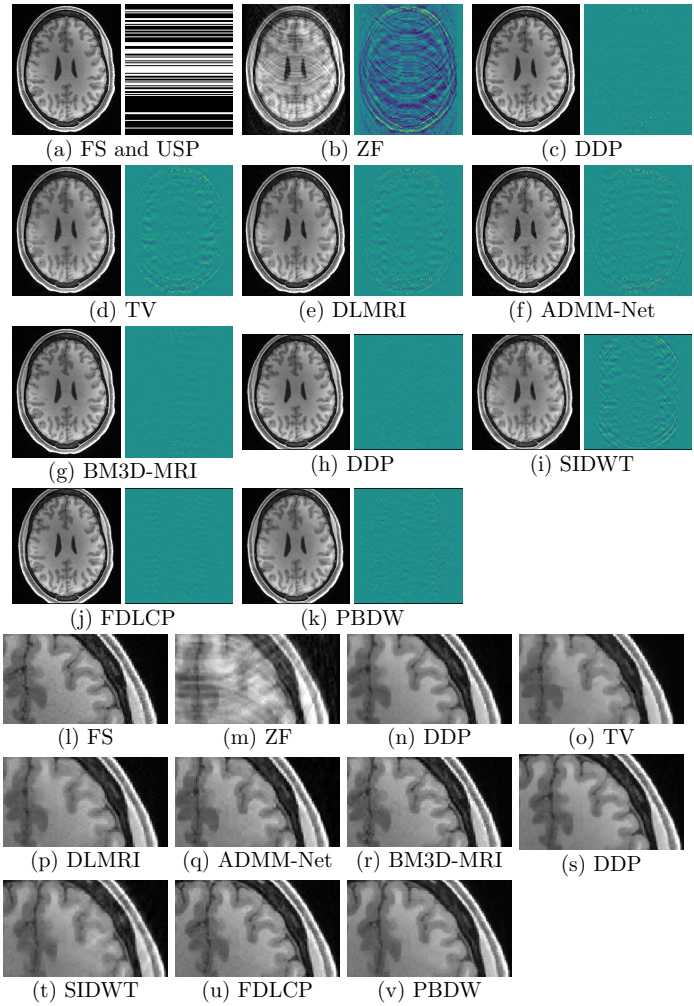


FIGURE 3.2: Reconstruction results for R=3. (a) shows the fully sampled image (FS) and the undersampling pattern (USP). (b-g) show reconstruction results and the error maps (intensities clipped to (-0.3, 0.3)) for the full FOV. (h-k) show the results for the cropped FOV. (l-v) show a zoomed region from the images above. (b-g) are produced with the undersampling pattern that was used to train the ADMM-Net for comparability. In (h-k), we also used the same pattern for undersampling for all methods for comparability.

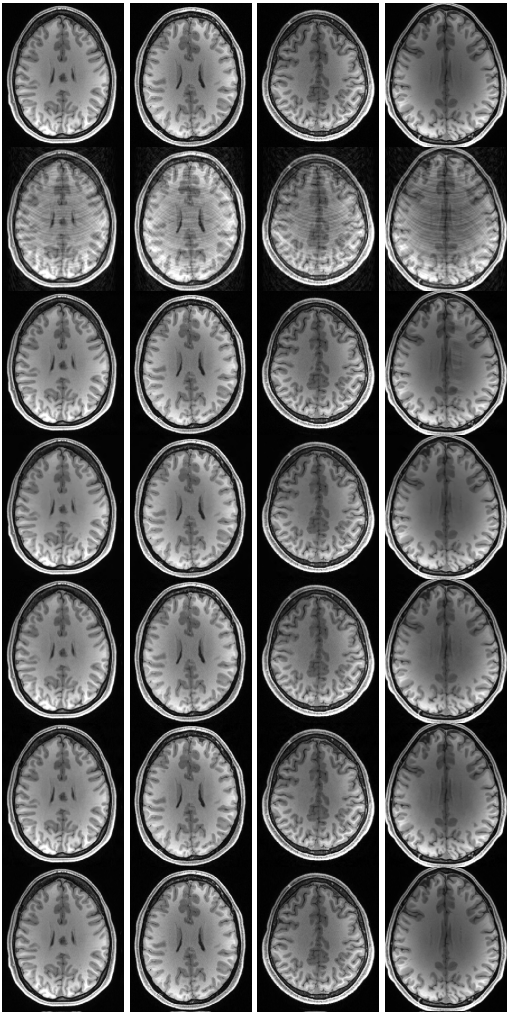


FIGURE 3.3: Reconstruction results from 4 different subjects with R=2. Rows show in top-down direction: the fully sampled images, zero-filling images, DDP, TV, DLMRI and ADMM-Net, BM3D-MR reconstructions for full FOV.

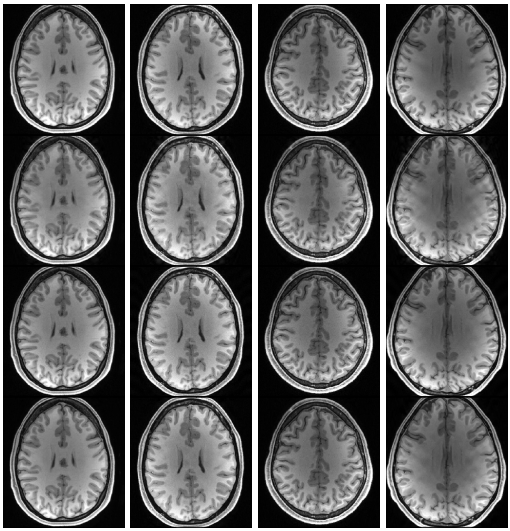


FIGURE 3.4: Similar display as in as Figure 3.3 with R=2, reconstructions with the cropped FOV for DDP, SIDWT, FDLCP, PBDW.

terms of CNR, the proposed method performed equally or better than all the other methods for both full and cropped FOVs.

In Figure 3.12 we show reconstruction results from retrospectively undersampled k-space data acquired within this study. First example is reconstruction of 2-fold Cartesian undersampling and the second is of 4-fold radial undersampling using NUFFT in the encoding operation. Our DDP method can reconstruct the original magnitude and phase fairly well, though the magnitude image was smoother than the fully sampled magnitude image in both cases. Quantitatively, the mean (std) RMSE for the Cartesian case for all subjects was 6.97% (0.37) using ESPIRiT coil maps (N=8) and 6.92% (0.58) using measured coil maps (N=6) for R=2. Mean (std) value for R=3 was 10.35% (1.53) with ESPIRiT coil maps and 9.93% (1.82) with measured sensitivities. These values were higher than those for HCP data. In order to see the portion of the increase in RMSE due to the inaccuracies in the encoding operator and phase, we also reconstructed the images by setting the phase of the fully sampled image to zero before undersampling and expanding the image by known coil maps. In this case the mean RMSE reduced to 5.76% (0.56) (R=2, N=8). Furthermore, we also reconstructed the images using only data consistency projections, corresponding to SENSE reconstruction, to see the added value of the DDP projections. In this case

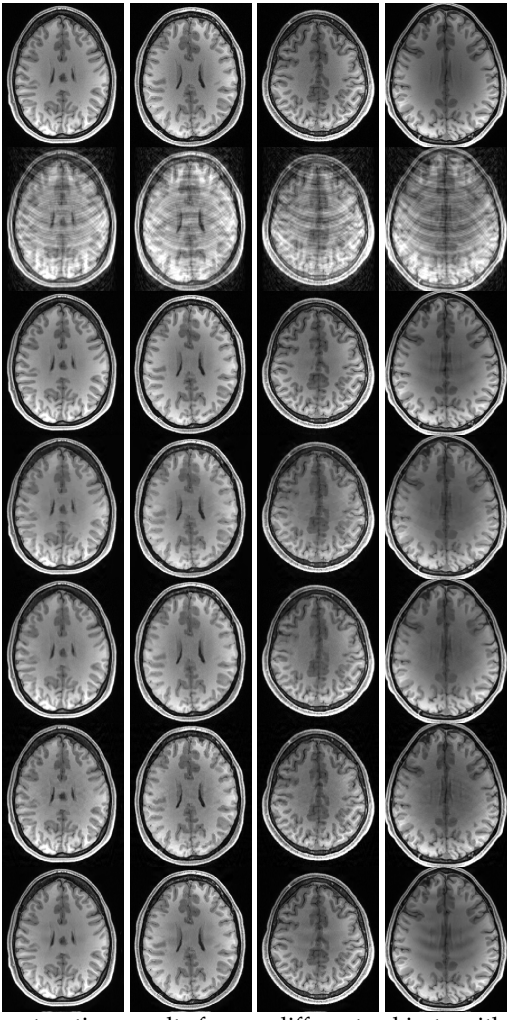


FIGURE 3.5: Reconstruction results from 4 different subjects with R=3. Rows show in top-down direction: the fully sampled images, zero-filling images, DDP, TV, DLMRI and ADMM-Net, BM3D-MR reconstructions for full FOV.

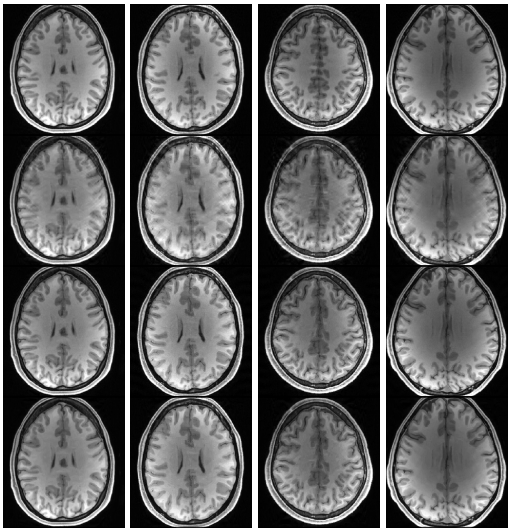


FIGURE 3.6: Similar display as in as Figure 3.5 with R=3, reconstructions with the cropped FOV for DDP, SIDWT, FDLCP, PBDW.

the mean RMSE was 7.95% and 11.42% for (R=2,3, respectively, N=8, ESPIRiT maps). Figure 3.10 shows the improvement achieved in RMSE by using DDP with SENSE in contrast to performing only SENSE reconstruction for one image. As described, we do the reconstruction with only data consistency projections for the first 10 iterations, then switch on the DDP projection. We show only 40 iterations. The drop in the RMSE value after iteration 10, i.e. right after the DDP projection is applied, shows the added value of doing the DDP projection. We also observe increase in RMSE for both reconstruction methods with increasing iterations. This increase is attributed to the discrepancy between the true and used coil sensitivities. Reconstructed images for this subject are given in Figure 3.11. We observe larger aliasing artifacts in only SENSE reconstruction.

In Figure 3.13, we show DDP reconstructions for the ADNI images for R=2. We used the VAE model that was trained on the HCP dataset, which had only healthy subjects, to reconstruct the images here. The reconstructed images recover GM and WM structures and edges faithfully. The WM lesions were also well reconstructed. The error maps do not indicate a specific increase in the lesion regions.

In addition to these results, we show the convergence of the POCS algorithm, and results for patch size, latent dimension and noise analyses.

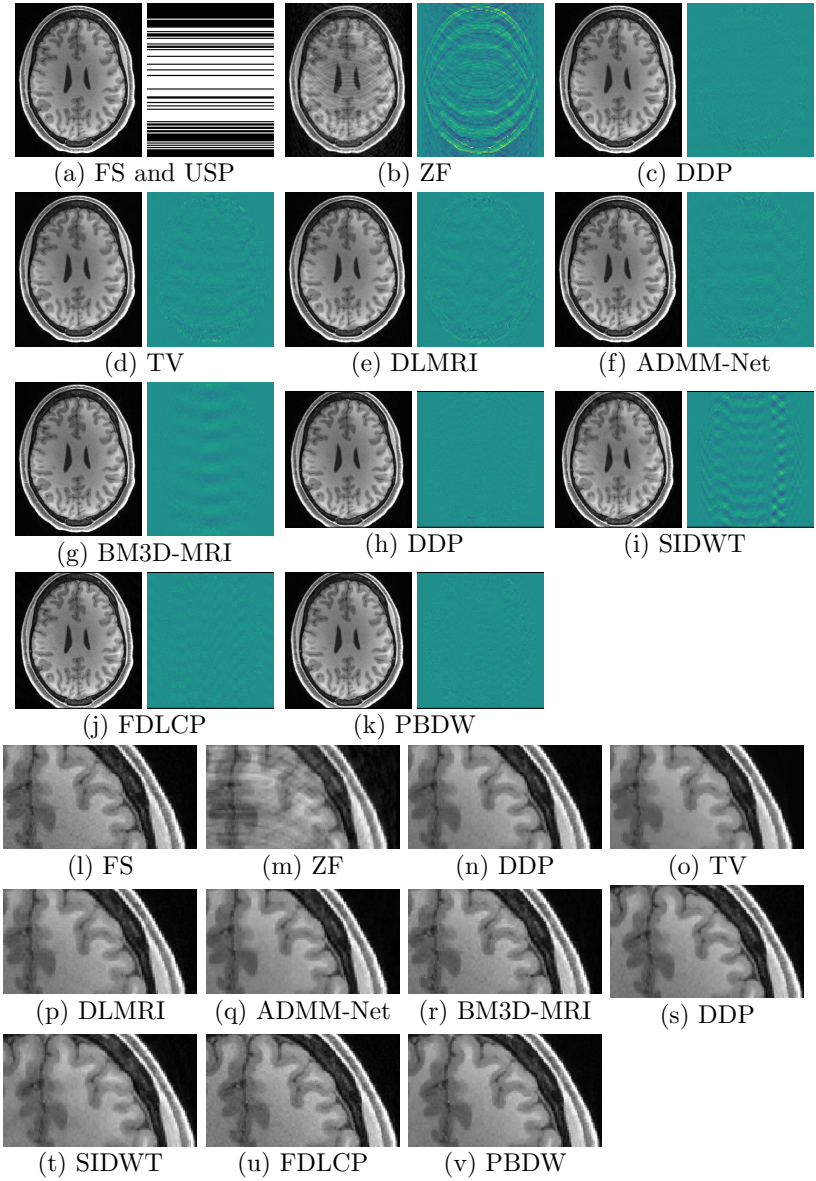
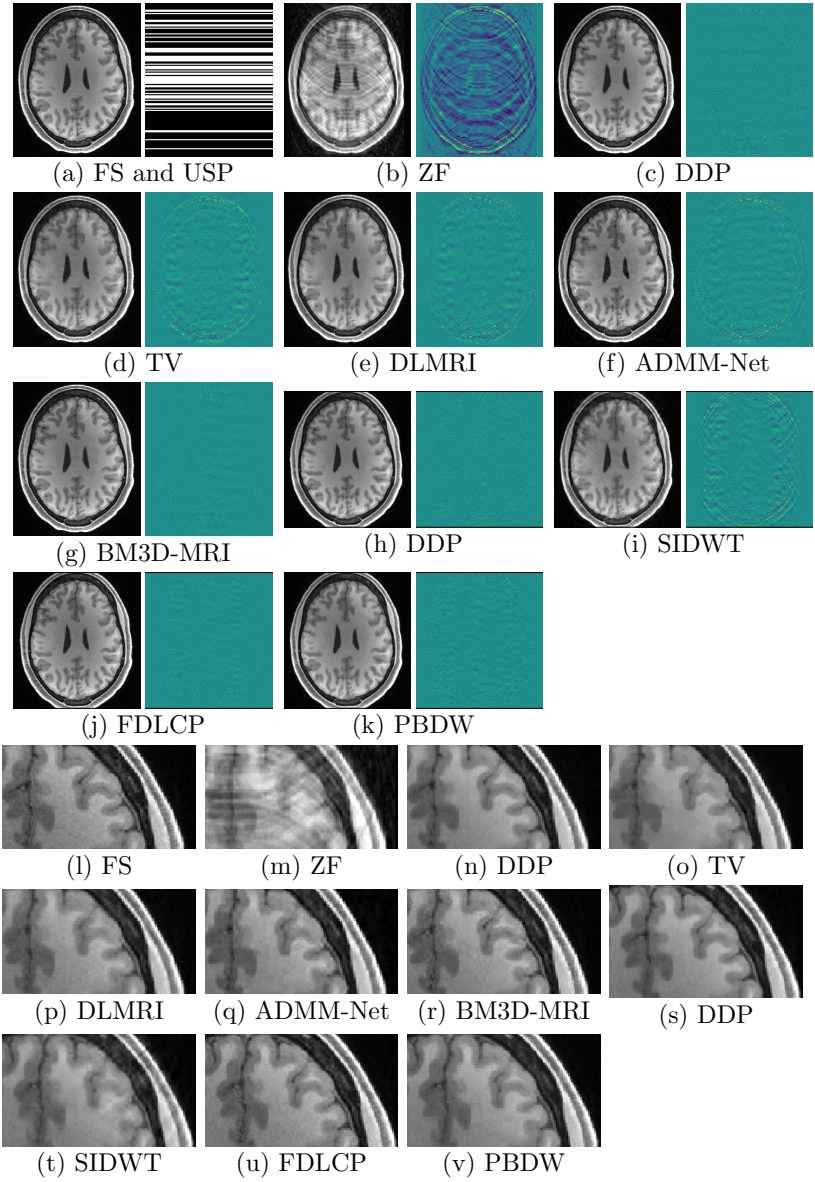


FIGURE 3.7: Similar display as in Figure 3 with R=2.



(a) FS and USP (b) ZF (c) DDP

(d) TV (e) DLMRI (f) ADMM-Net

(g) BM3D-MRI (h) DDP (i) SIDWT

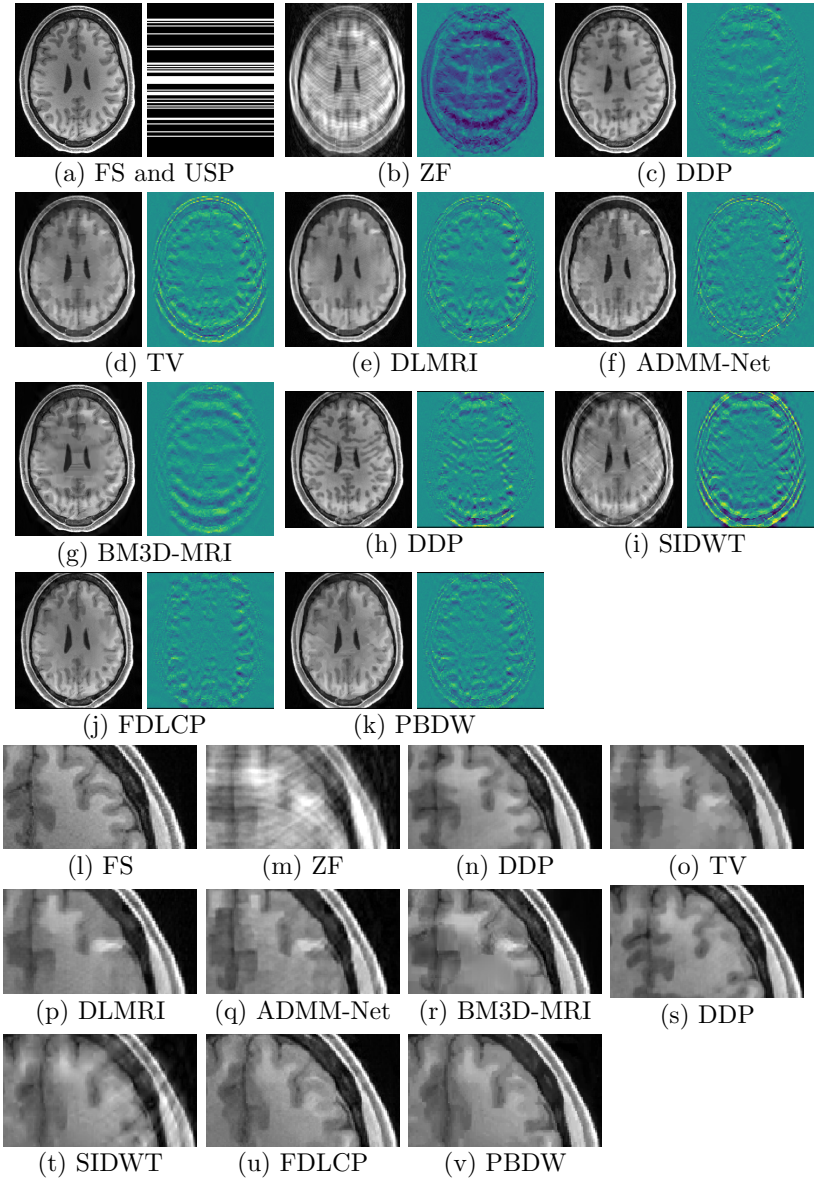
(j) FDLCP (k) PBDW

(l) FS (m) ZF (n) DDP (o) TV

(p) DLMRI (q) ADMM-Net (r) BM3D-MRI (s) DDP

(t) SIDWT (u) FDLCP (v) PBDW

FIGURE 3.8: Similar display as in Figure 3 with R=4.



(a) FS and USP (b) ZF (c) DDP

(d) TV (e) DLMRI (f) ADMM-Net

(g) BM3D-MRI (h) DDP (i) SIDWT

(j) FDLCP (k) PBDW

(l) FS (m) ZF (n) DDP (o) TV

(p) DLMRI (q) ADMM-Net (r) BM3D-MRI (s) DDP

(t) SIDWT (u) FDLCP (v) PBDW

FIGURE 3.9: Similar display as in Figure 3 with R=5.

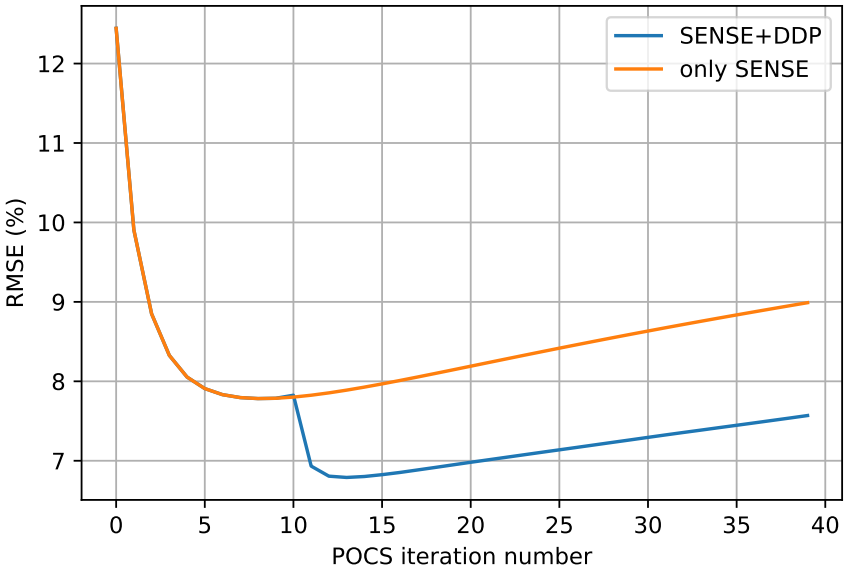
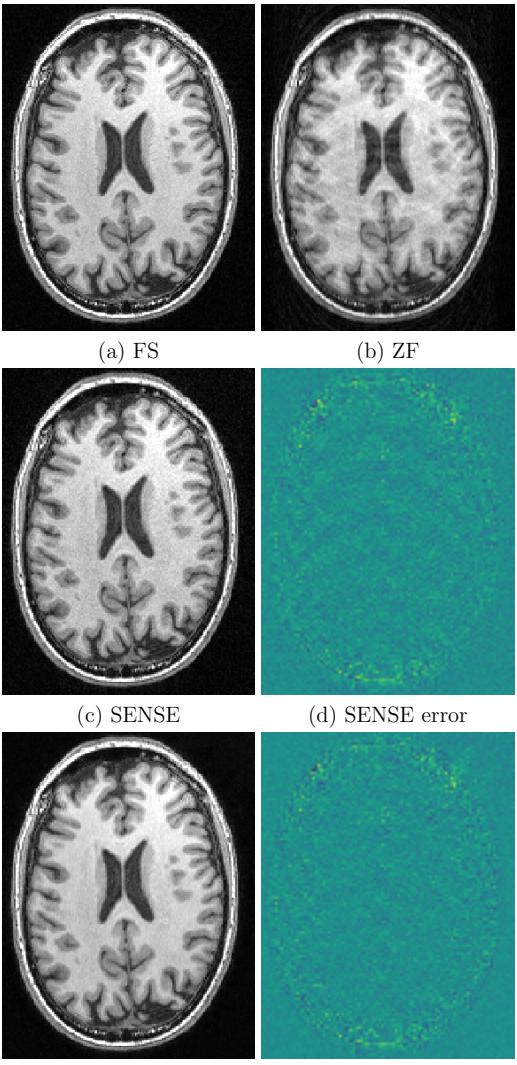


FIGURE 3.10: RMSE values for using doing SENSE reconstruction (i.e. only data consistency projections) vs doing combined SENSE and DDP reconstruction. Notice the DDP projection is switched on after the 10th iteration. The drop right after the first DDP projection demonstrates the added value due to the DDP projection.



(a) FS (b) ZF

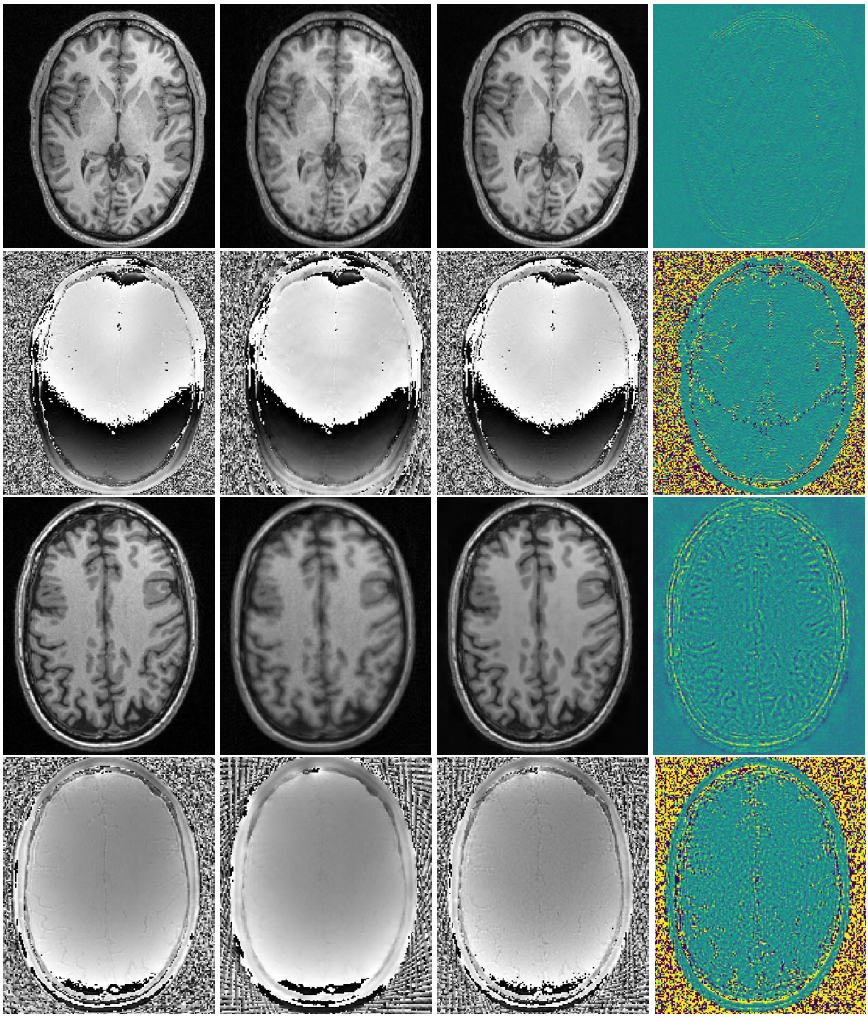
(c) SENSE (d) SENSE error

(e) DDP+SENSE (f) DDP+SENSE error

FIGURE 3.11: Reconstruction results for a measured subject with using only SENSE and the combination of SENSE and DDP. Error maps are clipped to (-0.3,0.3)

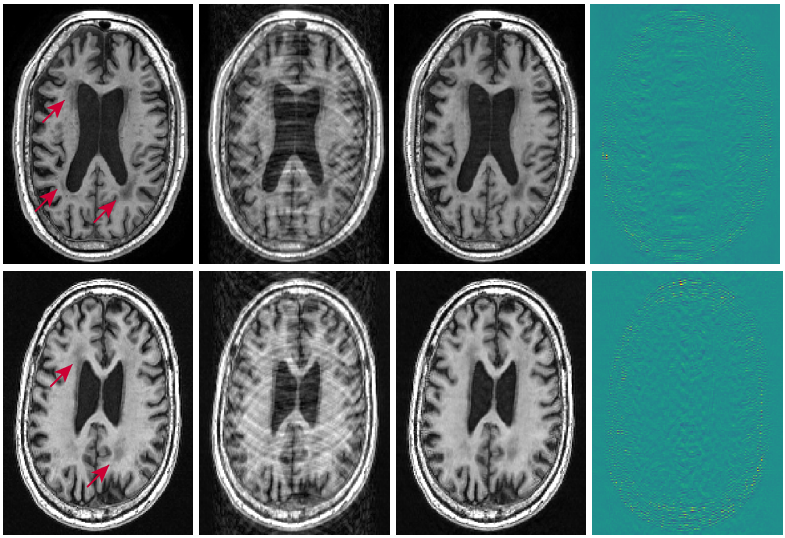
	R=2			R=3			R=4			R=5		
	RMSE	CNR	CN	RMSE	CNR	CN	RMSE	CNR	CN	RMSE	CNR	CN
FS	-	0.48(0.10)	0.12(0.02)	-	0.48(0.10)	0.12(0.02)	-	0.48(0.10)	0.12(0.02)	-	0.48(0.10)	0.12(0.02)
Zero-fill	13.03(1.13)	0.40(0.09)	0.12(0.02)	21.15(1.36)	0.33(0.07)	0.09(0.02)	24.92(1.91)	0.31(0.06)	0.08(0.02)	27.36(1.79)	0.30(0.06)	0.08(0.02)
DDP	2.76(0.53)	0.48(0.11)	0.12(0.02)	4.25(0.61)	0.48(0.10)	0.12(0.02)	6.46(1.57)	0.46(0.11)	0.11(0.02)	11.13(2.39)	0.41(0.10)	0.10(0.02)
TV	3.87(0.47)	0.46(0.11)	0.12(0.02)	7.56(0.83)	0.40(0.10)	0.09(0.02)	11.40(1.39)	0.35(0.09)	0.08(0.02)	14.56(1.23)	0.31(0.08)	0.07(0.02)
DLMRI	4.48(0.52)	0.46(0.11)	0.12(0.02)	7.25(0.83)	0.40(0.10)	0.10(0.02)	10.72(1.31)	0.33(0.09)	0.08(0.02)	13.87(1.25)	0.30(0.08)	0.07(0.02)
ADMMNet	3.55(0.40)	0.48(0.11)	0.12(0.02)	7.06(0.52)	0.45(0.11)	0.11(0.02)	11.26(0.72)	0.36(0.09)	0.09(0.02)	13.05(0.70)	0.32(0.08)	0.08(0.02)
BM3D-MRI	1.92(0.36)	0.48(0.10)	0.12(0.02)	4.23(1.05)	0.46(0.10)	0.11(0.02)	8.08(2.48)	0.43(0.10)	0.10(0.02)	11.70(2.76)	0.38(0.09)	0.09(0.02)
FS	-	0.48(0.10)	0.12(0.02)	-	0.48(0.10)	0.12(0.02)	-	0.48(0.10)	0.12(0.02)	-	0.48(0.10)	0.12(0.02)
DDP	2.68(0.38)	0.48(0.10)	0.12(0.02)	4.61(1.12)	0.47(0.10)	0.11(0.02)	7.39(1.47)	0.45(0.10)	0.11(0.02)	13.00(3.01)	0.39(0.08)	0.10(0.02)
SIDWT	4.49(0.98)	0.45(0.11)	0.12(0.02)	9.42(1.62)	0.39(0.09)	0.12(0.02)	14.57(1.96)	0.33(0.08)	0.10(0.02)	18.76(2.80)	0.32(0.07)	0.10(0.02)
FDLCP	2.63(0.35)	0.48(0.10)	0.12(0.02)	4.35(0.87)	0.45(0.10)	0.14(0.03)	6.72(0.89)	0.41(0.10)	0.13(0.03)	9.62(1.48)	0.35(0.08)	0.11(0.02)
PBDW	3.24(0.38)	0.47(0.11)	0.12(0.02)	5.59(0.94)	0.44(0.10)	0.13(0.02)	8.51(0.98)	0.38(0.09)	0.12(0.02)	11.38(1.39)	0.34(0.08)	0.10(0.02)

TABLE 3.1: Table summarizing results for different reconstruction quality metrics. Numbers indicate the mean (and standard deviation) of the error metric for N=17 test images. Top group are the results for experiments with full FOV images and bottom group are for cropped FOV images.



(a) FS (b) ZF (c) DDP (d) Error

FIGURE 3.12: Reconstruction results for measured data. Rows show the magnitude and phase of (from left to right) the fully sampled image, (FS), the zero-filling image (ZF), the DDP reconstruction (using ESPIRiT coil maps) and the error (clipped to (-0.3,0.3)). Upper two rows: A Cartesian US pattern (R=2) was used to retrospectively undersample the k-space data. Lower two rows: A radial US pattern (R=4) was used. In both cases the prior projection is the same, only the data consistency projection differs (using FFT and NUFFT, respectively).



(a) FS (b) ZF (c) DDP (d) Error

FIGURE 3.13: DDP reconstruction results for two images with white matter lesions due to Alzheimer's disease from the ADNI data set for R=2. Images show the original (FS), zero-filling (ZF), reconstructed (DDP) and the error maps from left to right. Lesions are clearly visible in the reconstructed images as well. Error map values are clipped to (-0.3,0.3). Arrows denote lesions.

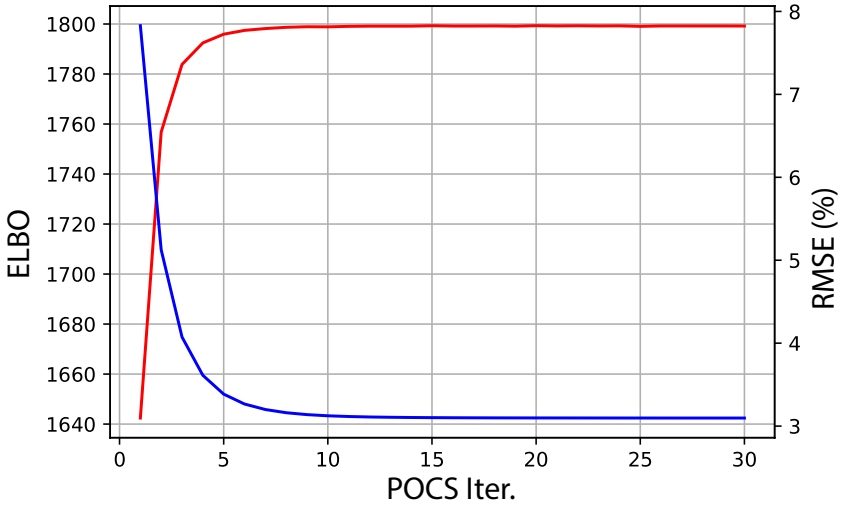


FIGURE 3.14: ELBO (red) and RMSE (blue) values between an original and reconstructed HCP image during DDP reconstruction from an undersampled image with R=2. Convergence is achieved approximately after 15 iterations for R=2. For higher R values, more iterations are necessary. The RMSE starts with a high value as the iterations are initialized with the zero-filled image. Notice that ELBO also increases stably.

Figure 3.14 demonstrates the convergence of RMSE and ELBO values during iterations of the POCS algorithm for a random image from the test set for R=2. We observe the decrease in RMSE and increase in ELBO, and convergence in both measures. Figure 3.15 shows results for different configurations of the model. We vary the dimensions of the latent space and the patch size to see how sensitive the method is to these design parameters. We do the experiments for 5 test subjects at R=3. We observe that patch-size has a much larger impact than latent-dimension. Reconstruction accuracy values do not change substantially with changing latent-dimension. On the other hand, we observe decrease in reconstruction error with decreasing patch-size, especially going from 12 to 20. This suggests the value of building explicit priors for larger image patches. Figure 3.16 shows the

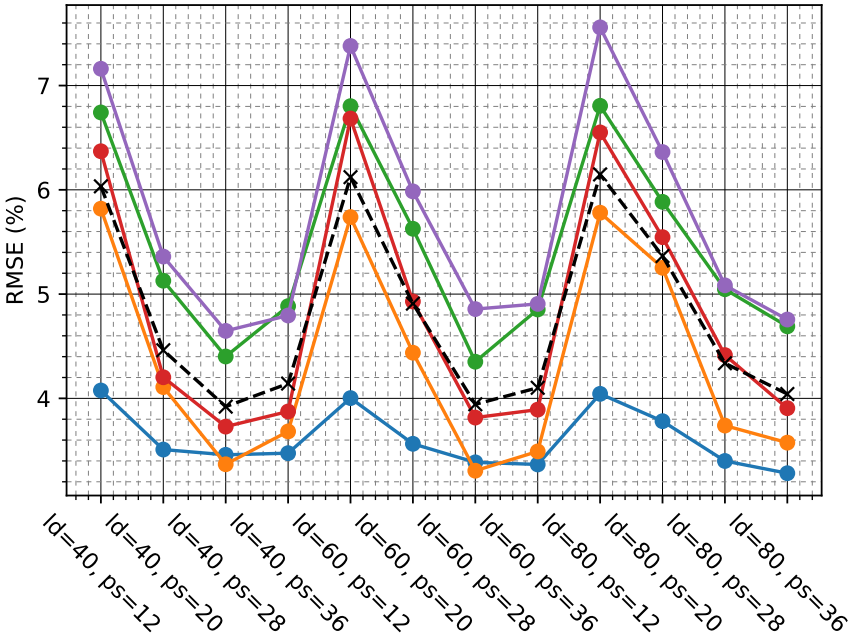


FIGURE 3.15: RMSE values for varying the latent space dimension (ld) and patch size (ps) (5 subjects, R=3). The method performs similarly for different configurations indicating robustness to parameter selection to some extend. Each color indicates a different subject. The dashed line is the mean value for all subjects.

decrease in performance with decreasing SNR in the undersampled image for 5 subjects for R=2 to 5. We calculate the RMSE between the noisy fully sampled image and the reconstruction. Increase in RMSE is observed, as expected, however, the rate of increase was slow for all the cases. Figure 3.17 shows examples of reconstructions with varying patch size and SNR for R=3, and latent dimension 60.

Next, we show results for the experiments assessing sensitivity to deviations in acquisition specifications between training and test images of ADMM-Net in Table 3.2. All the differences were statistically significant at the 0.05 level according to the paired test, performance of the method decreased significantly when the training and test patterns/ratios differed. Performance differences in the Experiment I were particularly high. In

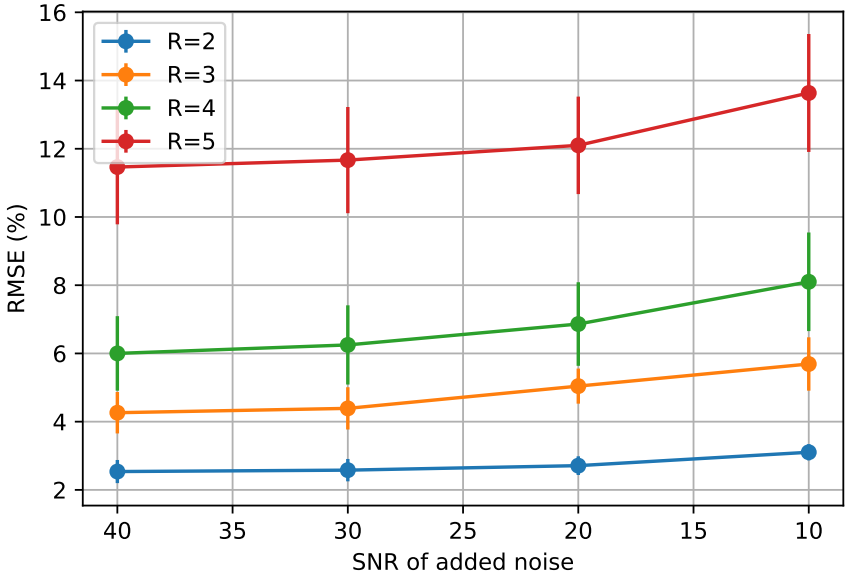
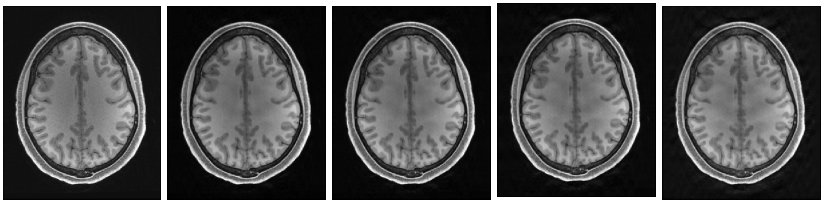


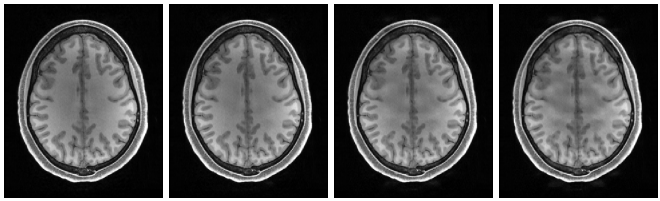
FIGURE 3.16: RMSE vs SNR of added noise (5 subjects). As expected increasing noise increases the error in reconstruction. Notice the net SNR of the images is around half of the added value, due to the base noise in the original images.

	Experiment I		Experiment II		Experiment III	
R_train	2	4	4	2	3 cart	3 rad
R_US	2	2	4	4	3 cart	3 cart
RMSE mean	3.17	4.11	10.55	10.70	6.49	6.67
RMSE std	0.39	0.44	0.68	0.66	0.49	0.51
RMSE diff. 25%	0.82		0.09		0.14	
RMSE diff. median	0.86		0.15		0.16	
RMSE diff. 75%	1.02		0.22		0.2	
p-value	0.0002		0.0009		0.0002	
RMSE DDP mean (std)	2.47 (0.44)		6.08 (1.45)		3.93 (0.56)	

TABLE 3.2: Reconstruction results for ADMM-Net, when using different US patterns at training and test times. Statistics were derived from N=17. R_{train} and R_{test} denote the US factor (pattern) used in training and testing. If not stated, a Cartesian pattern (cart) was used and "rad" denotes pseudo-radial pattern. RMSE diff. refers to the n^{th} percentile or median value of the distribution of pairwise RMSE differences. p-values are calculated with paired Wilcoxon signed-rank test. Last row presents the DDP RMSE values for comparison purposes. These RMSE values were calculated with disregarding the artifacts in the background using a brain mask.



(a) Orig. (b) SNR=40 (c) SNR=30 (d) SNR=20 (e) SNR=10



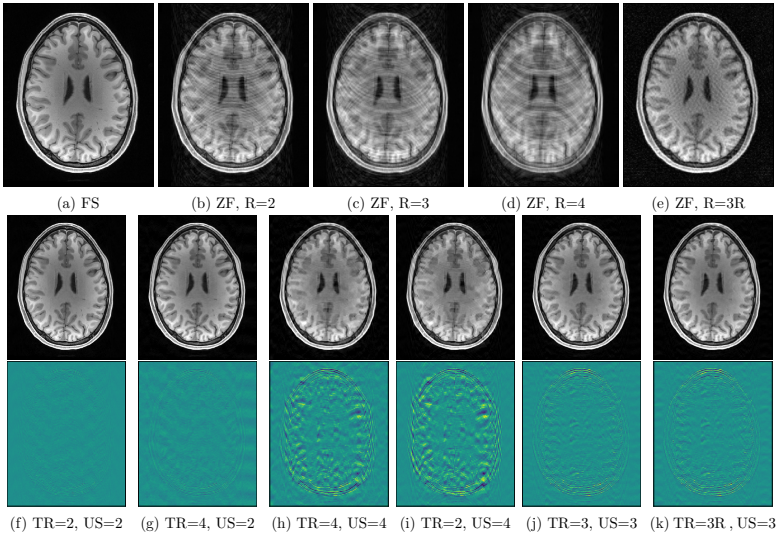
(f) PS=36 (g) PS=28 (h) PS=20 (i) PS=12

FIGURE 3.17: Example reconstructions for varying SNR and patch size (PS).

Figure 3.18 we show ADMM-Net visual reconstruction results for this experiment.

In Figure 3.19 we show results for two slices from the inferior and superior regions of the brain, for which the prior model was not trained. Visual inspection shows that the superior slices are reconstructed fairly well, whereas the model struggles more in the inferior slices, especially in the cerebellum. RMSE values for N=17 subjects are 4.28% (0.68) and 6.90% (1.23) for R=2 and 3, respectively. These values show an overall decrease in performance, as expected, when the prior is used for reconstructing slices from regions which are not represented in the training set. This decrease might also be partly due to the fact that the inferior slices generally comprise more complex structures than central slices, for instance the cerebellum among others. Furthermore, the RMSE values are higher for the inferior slices, supporting the visual observations: mean (std) RMSE for superior slices are 3.66% (0.44) and 5.43% (0.54) (N=6) in contrast to 4.44% (0.71) and 7.41% (1.06) (N=11) for inferior slices for R=2 and 3, respectively. This difference is highly likely due to the difference in the structural complexity between the regions.

Lastly, we show patches sampled from the prior model trained for patch-size of 28x28 and latent dimension of 60 in Figure 3.20. These patches were

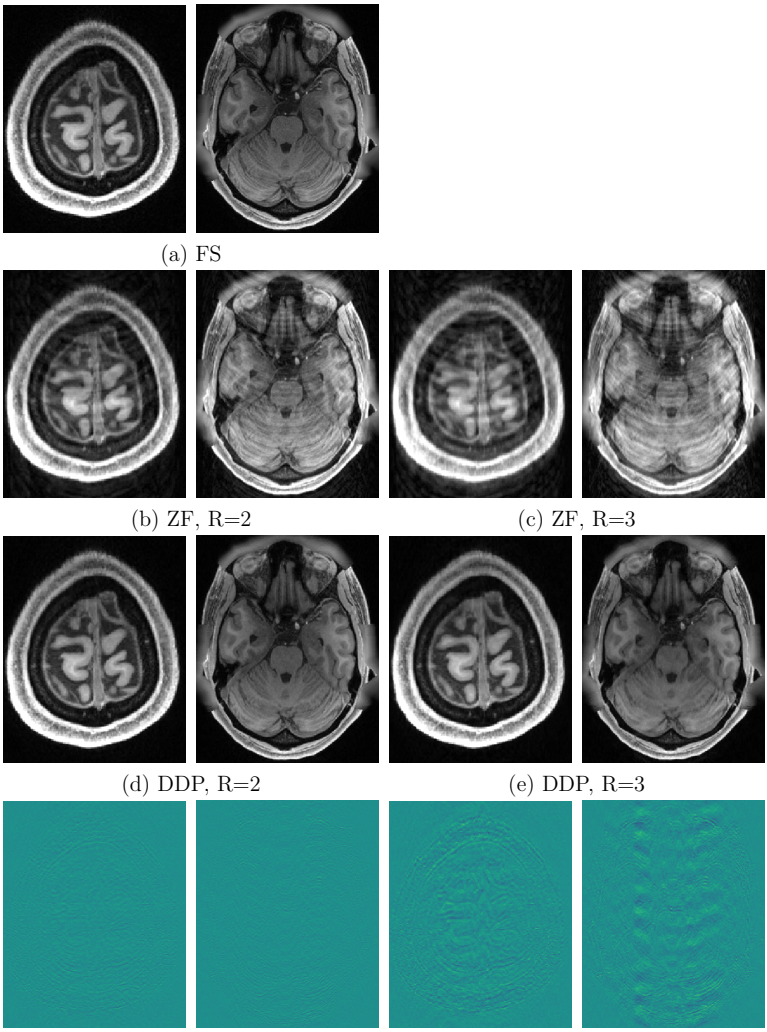


(a) FS (b) ZF, R=2 (c) ZF, R=3 (d) ZF, R=4 (e) ZF, R=3R

(f) TR=2, US=2 (g) TR=4, US=2 (h) TR=4, US=4 (i) TR=2, US=4 (j) TR=3, US=3 (k) TR=3R , US=3

FIGURE 3.18: ADMM-Net reconstruction results with different training and test undersampling patterns. First row shows the fully sampled image and zero-filled images. R denotes the Cartesian undersampling ratio used for generating the zero-filled images. Second and third rows show the reconstruction results and the error maps, respectively. TR and US stand for Cartesian undersampling factors for training and testing, respectively. "3R" denotes pseudo-radial undersampling pattern with factor 3. Error maps are clipped to (-0.3,0.3)

generated by simply feeding 16 random z vectors drawn from unit Gaussian to the decoder network. The decoder network outputs the mean images, i.e. $\mu_\varphi(z)$, and the corresponding variance images, i.e. $\Sigma_\varphi(z)$. We like to note that we have not cherry-picked these examples. The sampled mean patches, shown on the left, look realistic where gray matter (GM), white matter (WM) and gyri/sulci structures are clearly defined and properly positioned. These generated samples suggest that the VAE algorithm is able to approximate the underlying distribution of MR patches. The variance images on the right show that VAE places high variance in the cerebrospinal fluid (CSF) areas as well as boundaries between structures. We observe that isolated GM or CSF islands receive high variance. The samples support our hypothesis that the VAE model can learn to approximate the distribution of MR patches. The variance maps show that the sulci-like generated structures filled with CSF



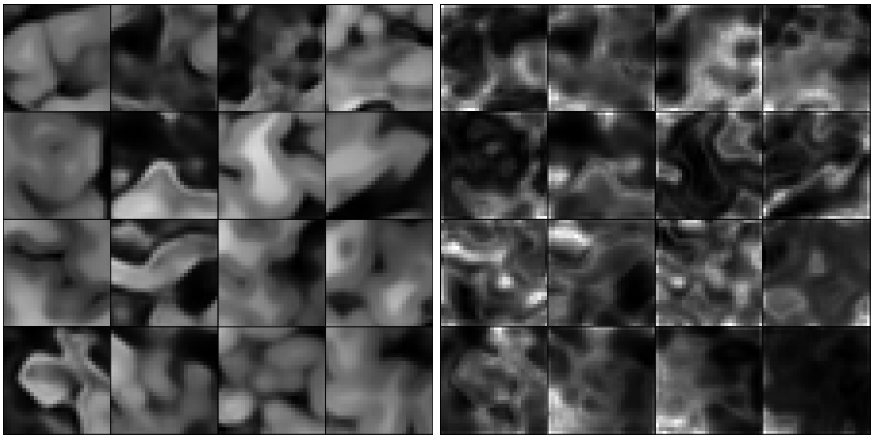
(a) FS

(b) ZF, R=2 (c) ZF, R=3

(d) DDP, R=2 (e) DDP, R=3

(f) Error, R=2 (g) Error, R=3

FIGURE 3.19: DDP reconstruction results (R=2 and 3) for two slices from regions, for which the prior was not trained. Left and right columns show slices from superior and inferior regions, respectively. Error maps are clipped to (-0.3,0.3)



(a) Mean ($\mu_\varphi(\mathbf{z})$) (b) Variance ($\Sigma_\varphi(\mathbf{z})$)

FIGURE 3.20: 16 image patches sampled from the prior learned by the VAE algorithm. Images on the left are the mean predictions and the images on the right are the corresponding variance maps. The patches are 28x28.

have higher variance, which is in accordance with the works that quantified uncertainty in image synthesis [105].

3.4 DISCUSSION

The reconstruction examples in Figure 3.2 and the quantitative results in Table 3.1 show that the proposed deep density prior reconstruction method produced highly accurate reconstructions both in terms of RMSE and CNR compared to the other methods. While BM3D-MRI and FDLCP could achieve slightly lower RMSE's, they both decreased CNR further than DDP, indicating that both methods sacrificed contrast to reduce RMSE, generating blurrier images. Leveraging the powerful prior over image patches, DDP was able to restore structures in the image faithfully while removing aliasing artifacts. Comparing methods that use explicit priors, DDP's performance demonstrate the huge potential in DL-based priors for MRI reconstruction. Compared to ADMM-Net, DDP's higher accuracy show that it is a highly attractive alternative to feed-forward network approaches.

Experiments with images from the ADNI dataset and acquired k-space data showed further properties of DDP and also prompted some key questions for further research. Results in Figure 3.12 show that the DDP method yielded high quality reconstructions for both the magnitude and phase images, though with higher mean RMSE values than those in the HCP test images. The fact that the RMSE decreased when the correct coil sensitivities and phase were used, suggest that some part of the error is due to the discrepancies between real and measured coil sensitivities. Hence, increasing accuracy of the coil maps would potentially decrease the error rate. Similarly, the method currently uses theoretical or no prior for the phase but it would benefit from better priors for this component as well.

The ADNI reconstruction results are encouraging in two respects. First, they show that the learned model does not blur out the lesions during the prior projection, which could have been expected since the training images did not contain any examples of lesions. We believe the lesions could be reconstructed because they are structurally not very different from healthy brain structures. In addition, the data consistency term makes sure the sampled information regarding the lesions is kept in the final image. However, the proposed method, similar to other reconstruction methods, requires further investigation as to where its limits lie, such as with bigger, brighter lesions. We believe for an optimal treatment of lesions, the training data set should include such images to provide the prior the capability to represent them.

Second, the proposed method performed reasonably well despite the domain difference between the training and test sets. Although the two data sets were acquired at the same field strength, they still differed in some acquisition protocol and imaging parameters. Their FOV and voxel resolution are different, and HCP was acquired with fat suppression while ADNI was not, which affects the image contrast. Our method is invariant to changes in FOV but not in scale. The lack of fat suppression also makes the dealiasing more challenging in the ADNI images since artifacts become more prominent. This results in very faint artifacts in the ventricles in the reconstructed images. Despite these differences, the quality of the reconstruction results indicate that the learned prior model could generalize to slightly different scales and similar but not identical imaging protocols. However, these experiments raised two key questions and exciting research directions regarding the sensitivity of the proposed method.

Even though the proposed method is not sensitive to variations in undersampling patterns or coil settings, when contrast or resolution of the

acquired images differ substantially from training images used to learn the prior, we believe reconstruction quality will decrease. There are two interesting directions to remedy this issue. First is to use a different prior for different contrast and resolutions. Second is to further improve the reconstruction quality using appropriate domain adaptation methods. Integrating invariance to domain differences in the prior and even building a joint prior for multiple contrasts are interesting research directions. Similar sensitivity issues can also be expected when the underlying anatomy differs between the acquired image and the prior. In this scenario, the safest approach would be to train a prior for each anatomy but building a prior that is capable of representing multiple anatomies is also an interesting direction. We also note that it is possible that all learning based methods suffer from the mentioned sensitivities [106]. On the other hand, methods such as [81] may offer a remedy without training multiple models.

An underlying assumption of our model is the unit Gaussian prior for the latent space, for which different extensions such as Gaussian mixtures or graphical models were investigated in [107, 108], which could in principle improve the representation capacity of the prior. Similarly, different density estimation methods are investigated, for instance using generative adversarial networks (GAN) [109, 110]. In [111], the authors proposed to use the discriminator of a GAN to superresolve images. The authors in [112] explain the drawbacks of this approach and similar to [113] suggest projecting images onto the span of the generator. A similar approach is given in [114]. The advantage of using VAEs is the simplicity of the inverse mapping. Determining the latent space representation of a given image is much more straightforward with VAEs, not requiring a separate optimization as GANs.

Another alternative approach to the VAE would be to use denoising autoencoders (DAE) [115]. Even though these might offer faster algorithms, the advantage of VAEs compared to DAEs are i) a more principled way of approximating the target distribution and ii) ease and rigor in approximating the likelihood of an image patch due to the variational inference mechanism that underlies VAEs.

One limitation of our method is the requirement of training in contrast to methods using fixed bases for regularization. Training of the prior relies on the availability of high quality data. Such data is available for commonly used sequences, such as structural T1w and T2w MRI. For other sequences, such as functional or diffusion MRI, construction of appropriate training set is of interest for future research.

Experiments with different configurations showed that the DDP reconstruction is not sensitive to patch size and latent space dimensions for a reasonable range. Furthermore, reconstruction quality is higher when patches larger than 12x12 are used. In addition, as expected, the performance degrades with decreasing SNR, however, with a low rate indicating robustness of the method to noise. This means the method is likely to perform fairly well in the regions with higher g-factor that have lower SNR due to parallel imaging.

In our current implementation a single evaluation of the derivative through the network takes around 9 seconds for 360 28x28 patches (0.6 seconds per batch of 25 patches in parallel). The total run time is given by this multiplied by number of iterations and all the other operations including mainly the phase projection, data projection and application of the derivatives for these operations. This time can be reduced by optimizing the code, increasing the parallelization and changing the network to work with images directly rather than patches. Furthermore, we demonstrate our method on 2D slices with a single phase encoding/undersampling direction as a proof of concept. However the method can be extended to a 3D setting, either by doing slice-by-slice reconstruction with a VAE trained on the whole brain or by training a VAE with 3D patches.

Results presented in Table 3.2 demonstrate that ADMM Net model is also sensitive to deviations in undersampling patterns used in training and test images, similar to the method proposed in [49]. Other feed-forward networks may also show similar sensitivities however, to the best of our knowledge, such an analysis is rarely performed. In contrast, we emphasize that the proposed method, due to decoupling of the prior and the data consistency term, does not share the same sensitivity. It learns the prior distribution on fully-sampled images and can be used to reconstruct any sampling scheme faithful to the measured data without the need of retraining as long as the images are from the same domain.

3.5 CONCLUSION

In this chapter we proposed a novel method termed DDP for MR reconstruction from undersampled k-space acquisitions. The method uses the VAE algorithm to learn the distribution of MR patches from fully sampled images, removing the sensitivity of the model to the sampling pattern. The model then uses this learned distribution as a probabilistic prior in a Bayesian reconstruction framework. We have shown that the reconstruction

with the DDP approach yielded promising results for HCP and ADNI data sets as well as multi-channel k-space measurements in terms of visual quality and quantitative measures.

4

JOINT RECONSTRUCTION AND BIAS FIELD ESTIMATION

4.1 INTRODUCTION

This chapter is based on the publication "Joint reconstruction and bias field correction for undersampled MR imaging" [116]. The author of the thesis contributed to all aspects of the project. The experiments/analysis related to bias field estimation were performed by the co-author M. Gaillochet and the preliminary experiments regrading the modified phase regularization were performed by the author of the thesis.

In the previous chapter we presented the reconstruction problem, gave a review of different approaches and presented our method DDP. In this chapter we investigate and address the effects of the bias field as a special case of domain shift in reconstruction.

One important difference between conventional approaches such as compressed sensing [22, 35] and parallel imaging [18, 100] and learning based approaches to reconstruction is their generalization capabilities. The latter approach, relying on deep learning, has achieved state-of-the-art performance, surpassing the conventional approaches. This is due to the higher expressiveness of deep learning based methods, such as the capability of learning/representing much more complex non-linear mappings compactly in contrast to the more "shallow" and linear mappings in the conventional approaches. On the other, another negative aspect of neural networks is their capability of overfitting to a data distribution. If the test image does not match the training distribution, this results in a drop in the performance. This susceptibility to discrepancies between the training set and the test image, i.e. the domain shift, is characterized by the generalization capabilities of the learned model. In that regard, the conventional methods, due to their hand-crafted and fixed nature do not suffer from the lack of such generalization properties. Various methods have been developed to tackle the domain shift problem in the deep learning community. One approach is to augment the training data to hopefully resemble the test data better [117] [118]. Although this approach increases the generalization capabilities of the learned model, it does not optimize the solution specifically for the test image at hand. Another approach is to modify the network

parameters to reduce the domain gap [119], which indirectly optimizes the solution by modifying the network rather than the solution and tries to address a more general domain gap rather than the more specific bias field. On the other hand, the effect of the bias field on the image encoding is easy to model and there are readily available methods that can estimate it well, hence, utilizing these, one can devise methods to address the domain gap specific to the bias field.

In this work we propose a joint reconstruction algorithm, which estimates and explicitly models the bias field throughout the reconstruction process. By doing so, we remove one degree of variation between the training set and the image to be reconstructed. The training is done on images without bias field, and the bias field itself is modeled as a multiplicative term in the image encoding process, linking the training and test domains. In order to be able to do this we use a reconstruction algorithm which decouples the learned prior information from the image generation process, namely the DDP algorithm [29]. Using the N4 algorithm, we iteratively estimate the bias field in the test image throughout the reconstruction, and this estimation improves as the reconstructed image becomes better. We compare the proposed method to reconstruction without bias field estimation, using a publicly available dataset as well as in-house measured images, and show improvement in performance.

4.2 METHODS

The notation is the same as in previous chapters. We recite it here for completeness.

The measured k-space data, $y \in \mathbb{C}^M$ and the underlying true MR image, $x \in \mathbb{C}^N$ are related through the encoding operation $E : \mathbb{C}^N \to \mathbb{C}^M$ (which incorporates the coil sensitivites, Fourier transformation and the undersampling operation) as $y = Ex + \eta$, where η is complex Gaussian noise.

4.2.1 *DDP reconstruction with a modified phase prior*

The deep density prior reconstruction from Chapter 3 is given as

$$\min_x ||Ex - y|| - ELBO(x). \tag{4.1}$$

The VAE is trained on patches from fully sampled images and the ELBO term operates on a set of overlapping patches that cover the whole image.

As evident in the equation above, the ELBO term is independent of the image encoding. The DDP method solves the problem using the projection onto convex sets (POCS) algorithm. In this scheme the optimization is implemented as successive applications of projection operations for prior $\mathcal{P}_{prior}$, data consistency $\mathcal{P}_{DC}$ and the phase of the image $\mathcal{P}_{phase}$, i.e. $x^{t+1} = \mathcal{P}_{DC}\mathcal{P}_{phase}\mathcal{P}_{prior}x^t$. The prior projection is defined as a gradient ascent for a fixed number of steps for the ELBO term w.r.t. the image magnitude, i.e. $\mathcal{P}_{prior}x = x^N$, where $x^{n+1} = x^n + \alpha\frac{d}{dx}ELBO(|x|)|_{x=x^n}$ for $n = 0...N$. The data consistency projection is given as $\mathcal{P}_{DC}x = x - E^H(Ex - y)$.

For the phase projection, we use the one defined in the previous chapter [94], to which we add the minimization of the data consistency with respect to the image phase. Separating the image into its magnitude and phase as $x = |x| \circ \exp(i\varphi)$, where $\circ$ is the element-wise product and $\varphi \in \mathbb{R}^N$ is the phase, we write the whole regularization as

$$R_{phase} = ||E(|x| \circ e^{i\varphi}) - y||_2^2 + \lambda_{phase}||Ce^{i\varphi}||_2^2, \tag{4.2}$$

with C as the finite difference operator.

As the projection is done using gradient descent, we provide also the gradient here as

$$\frac{\partial R_{phase}}{\partial \varphi} = -2\mathrm{Re}\Big\{i(|x| \circ e^{-i\varphi})E^H[E(|x| \circ e^{-i\varphi}) - y] \tag{4.3}$$

$$+ \lambda_{phase} ie^{-i\varphi}C^H Ce^{i\varphi}\Big\}, \tag{4.4}$$

where C^H is implemented as the divergence operator. We use $\lambda_{phase} = 1$.

4.2.2 *Modeling and estimating the bias field*

Signal intensities are often not constant across the MR images, even inside the same tissue. Instead, they usually vary smoothly, with fluctuations of 10%-20%, across the measured 3D volume [99]. These variations, collectively called the bias field, can be attributed to factors like patient anatomy, differences in coil sensitivity or standing wave effects [99, 120], which are difficult or impossible to control during an acquisition. Hence they can introduce a degree of variation between images used for training reconstruction models and an image to be reconstructed at test time. In order to prevent a loss of performance, this variation has to be taken into account.

To this end, we model the bias field explicitly and incorporate it into Equation 4.1 as a multiplicative term B before the encoding operation, modifying the image intensities pixelwise. The bias field is an additional unknown in the reconstruction process that is estimated alongside x as

$$\min_{x,B} ||EBx - y|| - ELBO(x). \tag{4.5}$$

Here, the measured k-space y carries the effect of a bias field, while x is bias field free as the bias field in the image is explicitly modeled using the B term. This setting allows us to learn the VAE model on images without bias field. The advantage of this idea is two fold. Firstly, since training can be performed on bias field free images, it is easier for the VAE to learn the distribution as there is less spurious variation in the data. Secondly, we make the reconstruction problem easier by explicitly providing the bias field information, which otherwise would have to be reconstructed from the undersampled k-space as well. We solve Equation 4.5 as a joint iterative reconstruction problem by minimizing alternatively two sub-problems:

$$1. \quad x^t = \min_x ||EB^{t-1}x - y|| - ELBO(x) \tag{4.6}$$

$$2. \quad B^t = \mathrm{N4}(B^{t-1}x^t), \tag{4.7}$$

where N4 denotes the bias field estimation algorithm, which we will explain below. To account for the bias field, the data consistency projection $\mathcal{P}_{DC}$ needs to be adapted and becomes $\mathcal{P}^B_{DC}x = B^{-1}[Bx - E^H(EBx - y)]$. In this case the reconstructed image corresponding to y is given as Bx. This modification can be interpreted as doing a forward-backward projection with $\mathcal{P}_{bias} = B$ before and after the data consistency projection to move the image between the "normalized" bias field free domain and the bias field corrupted acquisition domain, i.e. $x^{t+1} = \mathcal{P}^{-1}_{bias}\mathcal{P}_{DC}\mathcal{P}_{bias}\mathcal{P}_{phase}\mathcal{P}_{prior}x^t$. The pseudocode for the described joint optimization scheme is presented in Algorithm 2.

4.2.2.1 *N4 bias field estimation*

N4 is a variant of the widely used N3 algorithm [121], a non-parametric iterative method that approximates intensity non-uniformity fields in 3-dimensional images. Given an image, its objective is to find a smooth, slowly varying, multiplicative field [99]. N4 improves upon the N3 algorithm by modifying the B-spline smoothing strategy and the iterative optimization scheme used in the original framework. We use the implementation available as N4ITK [122]. We denote this method as $N4(\cdot)$ in our formulations.

Algorithm 2 Joint reconstruction

1: y: undersampled k-space data
2: E: undersampling encoding operator
3: VAE: trained VAE
4: NumIter, BiasEstimFreq, DCProjFreq
5: **procedure** JOINTRECON(y, E, VAE)
6: $B \leftarrow \mathrm{N4}(E^H y)$
7: $x^0 \leftarrow (B)^{-1} E^H y$
8: **for** t: 0 to NumIter - 1 **do**
9: $x^{t+1} \leftarrow \mathcal{P}_{prior} x^t$
10: $x^{t+1} \leftarrow \mathcal{P}_{phase} x^{t+1}$ ▷ Optional
11: **if** t% DCProjFreq == 0 and $t \neq 0$ **then**
12: $x^{t+1} \leftarrow B^{-1}\left[Bx^{t+1} - E^H(EBx^{t+1} - y)\right]$ ▷ $\mathcal{P}^B_{DC}$
13: **if** t% BiasEstimFreq == 0 and $t \neq 0$ **then**
14: $B \leftarrow \mathrm{N4}(Bx^{t+1})$
return x^{t+1}, B

4.2.3 *Datasets used*

To train the VAE we used 5 central, non-adjacent slices with 0.7mm isotropic resolution from T1 weighted images of 360 subjects from the Human Connectome Project (HCP) preprocessed dataset [123, 124], which by default have a bias field. We used N4 on the images to also create a bias field free training set.

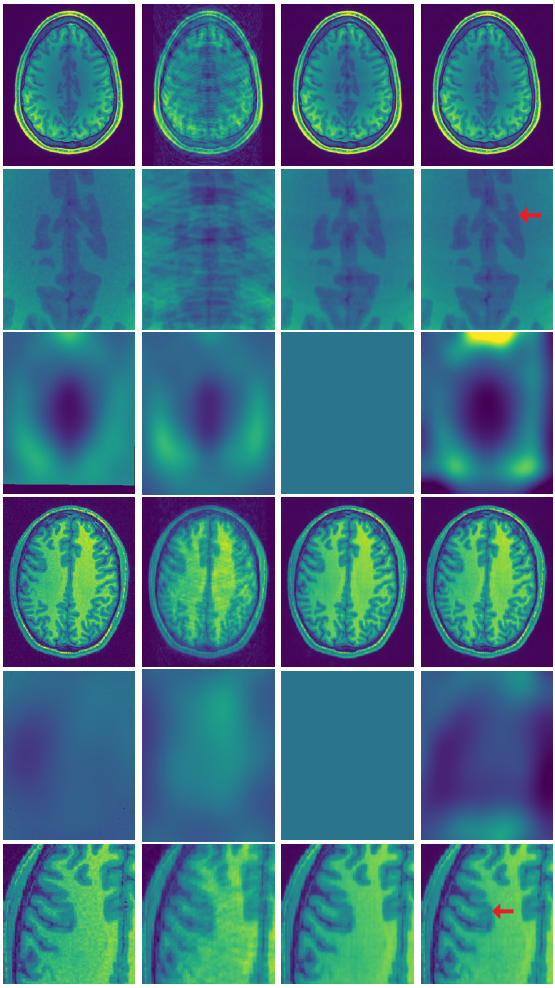
For test images, we took a central slice from 20 different test subjects from the HCP data. As the HCP images tend to have a similar bias field, we additionally created a modified test set where we estimated the bias fields with N4, took their inverse and multiplied them with the bias field free images. In addition to HCP data, we also tested the proposed method with central slices from 9 in-house measured subjects. These images were acquired using a 16 element head coil and have similar acquisition parameters as the HCP dataset with a 1mm isotropic resolution. We used ESPIRiT [100] to estimate the coil sensitivity maps for these images.

4.2.4 *Training the VAE*

We trained four patch-wise VAEs - for two different resolution levels to match the datasets, each with and without bias field. We used patches of size 28x28 with a batch size of 50 and ran the training for 500,000 iterations. The patches were extracted randomly from the training images with replacement. All the VAEs were trained with the same training images extracted from the HCP dataset, as described above.

4.2.5 *Experimental Setup*

We used random Cartesian undersampling patterns with 15 fully sampled central profiles. We generated a different pattern for each subject, and applied the same pattern for a given subject throughout all experiments for comparability of results. When reconstructing test images from the HCP dataset, we used 302 iterations (NumIter) for R=2, 602 for R=3, 1002 for R=4 and 1502 for R=5, to allow for convergence of reconstructed images. Since the in-house measured images have multiple coils, the successive applications of data consistency projections coincide with a POCS-SENSE reconstruction [92], speeding up convergence and requiring less iterations. Hence, when performing reconstruction on images from the in-house measured dataset, we ran the first 10 iterations without prior projection, applying only data consistency projections. Additionally, the discrepancy between the actual coil sensitivities and the ESPIRiT [100] estimations may lead to divergence after too many iterations. Hence, a lower number of iterations was used for reconstruction experiments on the in-house measure dataset: 32 iters for R=2, 102 for R=3 and R=4, and 202 for R=5. For all test datasets, the parameters were set as α = 1e-4, BiasEstimFreq=10, DCProjFreq=10. As for the N4 bias field estimation algorithm, the default parameters were used.



(a) FS (b) Zero-filled (c) Baseline (d) Joint rec.

FIGURE 4.1: Reconstruction results for R = 3, with (a) the fully sampled image, (b) the zero-filled image, (c) the reconstruction with no bias field estimation, (d) the joint reconstruction with bias field estimation using N4. The first three rows show reconstruction results for an HCP image, its zoomed-in version and the corresponding bias field. The next three rows show results for an in-house measured image. For visualization purposes, MR images are clipped to [0, 1.2] and bias fields, to [0.5, 1.8].

To evaluate the reconstruction quality of the different models, we computed, for each image, the percentage RMSE. More specifically, the evaluation metric was given as $\text{RMSE} = 100 \times \sqrt{\frac{\sum_i (|Bx| - |\hat{x}|)_i^2}{\sum_i |\hat{x}|_i^2}}$ where $\hat{x}$ is the original, fully sampled test image, and where the summation was applied pixel-wise. The error was computed on the skull-stripped images only in the brain area.

To evaluate the statistical significance of our results, we performed a permutation test [125] with 10,000 permutations to assess the null hypothesis that the RMSE's for the reconstruction with and without bias field estimation are from the same distribution. From these tests, we reported the p-values.

4.3 EXPERIMENTS AND RESULTS

To assess the hypothesis that correcting the bias field improves the overall reconstruction quality, we performed reconstructions on the original and modified HCP test sets as well as on the in-house measured images.

When applying our proposed method, we reconstructed undersampled images with a bias field, and used a VAE trained on images without bias field as well as the N4 bias field estimation algorithm. To evaluate our approach, we also ran baseline experiments on the same images, where we used a VAE trained on images with a bias field and did not apply bias field estimation during reconstruction. Given that the ground truth (fully sampled) images naturally have a bias field in them, we utilized the last bias field estimate multiplied with the reconstructed image, i.e. Bx, for visualisations and RMSE calculations.

Method	HCP dataset				Modified HCP dataset				In-house measured dataset			
	R= 2	R = 3	R = 4	R = 5	R= 2	R = 3	R = 4	R = 5	R= 2	R = 3	R = 4	R = 5
Baseline	2.24 (0.31)	3.39 (0.45)	4.42 (0.51)	5.72 (1.05)	2.24 (0.38)	3.45 (0.60)	4.26 (0.46)	5.58 (1.20)	4.64 (0.391)	6.852 (0.726)	8.593 (1.344)	11.046 (1.727)
Joint recon.	2.27 (0.34)	3.34 (0.40)	4.35* (0.47)	5.52 (0.66)	2.20* (0.39)	3.33* (0.53)	4.16 (0.54)	5.12* (0.57)	4.62 (0.418)	6.714* (0.821)	8.218* (1.266)	10.567* (1.632)

TABLE 4.1: Table of RMSE values. R is the undersampling factor. Numbers indicate the mean (std). The * indicates a p-value of less than 0.05. The baseline method is the DDP algorithm described in [29] that does not explicitly model the bias field. The proposed joint reconstruction method estimates the bias field using N4 and explicitly models it during reconstruction.

The results in Table 4.1 indicate that the proposed joint reconstruction method with bias field estimation improves the reconstruction quality in terms of RMSE when the bias field of the test image is different from those in the training set. In these cases, namely the experiments with the modified HCP and in-house measured images, the improvement is statistically significant with a p-value of less than 0.05 for nearly all undersampling factors. For the unmodified HCP dataset, where the bias field in the test images matches those in the training set, we do not expect a big difference in the performance, which is reflected in the results.

The quantitative improvement is also supported by the visual inspection of the images given in Figure 4.1. From the HCP image, one can observe that the level of artifacts is reduced with the proposed method. This becomes more evident in the zoomed-in images. The red arrow points to a part of the image where the proposed method can reconstruct the structures faithfully, whereas the baseline method struggles. Aliasing artifacts are globally suppressed better with the joint reconstruction method. Similarly, for the in-house measured image, the grey matter structure that the red arrow points to is not reconstructed in the baseline method, whereas it again appears with the proposed method.

In this work, we demonstrate the performance loss due to the bias field for a specific algorithm, into which we also integrate our proposed solution. However, the problem may not be specific to the algorithm used as it arises from the domain gap, which is a fundamental problem affecting machine learning based methods in general. Furthermore, the proposed method of estimating and explicitly modeling the bias field in reconstruction is also a generic approach, which can be integrated into different algorithms.

4.4 CONCLUSION

In this chapter we proposed and evaluated a method for joint reconstruction and bias field estimation to address variations due to the bias field when reconstructing undersampled MR images. The results indicate that the proposed method improves the baseline method (unsupervised learning based reconstruction algorithm), both in RMSE and visually. The improvements can be attributed to two factors. First, the proposed method allows the VAE prior to learn a simpler distribution, and second, providing the bias field explicitly makes the reconstruction problem easier. In essence, estimating the bias field during reconstruction makes the model less sensitive to dif-

ferences between the data used to train the model and the test data used during reconstruction.

5

SAMPLING FROM THE POSTERIOR

5.1 INTRODUCTION

This chapter is based on the pre-print "Sampling possible reconstructions of undersampled acquisitions in MR imaging" [126], under review for publication. The author of the thesis is the main contributor to this project.

As discussed in Chapter 2, undersampling the k-space in MR imaging reduces scan time by speeding up acquisition, allowing a higher throughput as well as higher comfort for patients. However, contrary to a fully acquired k-space, where an inverse Fourier transform is mostly enough to uniquely determine the underlying image up to measurement noise, the undersampled acquisition leads to an underdetermined system of equations. Mathematically, this means that there are infinitely many images that match the acquired portion of the k-space data and it is impossible to know which one is the underlying image in a general setting.

Traditionally, this problem of infinitely many solutions has been tackled as a deterministic reconstruction problem, where different methods were proposed to choose a single "best" image out of the set of possible images. This was achieved by introducing some prior information term that implicitly prefers a solution according to its certain properties, such as smoothness with Tikhonov regularization [127], sparsity of its gradients with total variation regularization [22] or how well it matches with some lower resolution version of the image [28], which we presented in Chapter 2. This converts the problem into a well-posed regularized inverse problem and allows to obtain a single solution as the reconstructed image. Similarly, in the Bayesian framework a suitable prior along with the data likelihood allows to write the corresponding posterior probability which then can be maximized to obtain the maximum-a-posteriori (MAP) image as the single "best" reconstruction. We presented a method implementing this idea in Chapter 3.

By providing a single reconstructed image as their output, however, these approaches miss the uncertainty in the solution that arises due to the missing portion of the k-space data. The reconstructed image is formed using information from the measured data and the prior. Measured data matches the underlying true image up to the measurement noise while the

prior complements the unmeasured information. However, it is important to remember that the prior is independent of the underlying true image and hence, the image that best satisfies the prior need not be the same as the true image. There is an inherent ambiguity. When a single reconstructed image is output, information coming from the prior is treated as definite. This treatment implicitly sacrifices the opportunity to reveal the inherent ambiguity in the reconstructed image.

An alternative approach, which we will pursue in this chapter is producing multiple images as solutions to the ill-posed inverse problem, where the images match the measured data while being highly likely according to the prior at the same time. Such an approach is capable of capturing the uncertainty in the inversion due to the missing data. The idea is that in this case the information from the prior is not treated as definite, but rather possibility of different images is explored under the specific prior used.

As the proposed approach provides multiple images, clinical, research or further analysis tasks, such as segmentation, can be applied separately on each image to propagate the uncertainty to any given task's output. Alternatively the uncertainty can be quantified at the image level and passed on the following tasks, such as in the form of a mean image along with a standard deviation map. Inspecting the images or the quantified uncertainty for different regions can also be indicative of which parts of the images might be more prone to differ from the underlying true image. To the best of our knowledge, this is the first time such an approach is being proposed for undersampled MRI acquisition.

In the recent years, though not directly related to the uncertainty due to missing data in undersampled MRI, there has been research efforts to quantify uncertainty in different aspects of the image reconstruction problem, especially with the deep learning based models [128]. One such aspect is the *epistemic* or model uncertainty, i.e. the uncertainty in the mapping learned by the neural network used in the reconstruction, which can be obtained using approaches such as drop-out [129]. However, these approaches quantify the uncertainty due to the ambiguity in the network parameters and do not capture the uncertainty due to the missing k-space data. Epistemic uncertainty tends to zero as the training set size goes to infinity [130], however, increasing training samples cannot be expected to diminish the ambiguity due to missing k-space data. Another aspect is the so called *aleatoric* uncertainty due to noise or other ambiguities in the images or the labels, which is more relevant to reconstruction from undersampled MRI. Quantifying aleatoric uncertainty can be approached using,

for instance, predictive heteroscedastic models, as in [131] for quantitative susceptibility mapping reconstruction. These models predict a different variance value for each output pixel and hence can in principle learn to predict high variance for pixels where the model expects to have incorrect mean predictions. There is one important limitation to these models. They generate only second order statistics of pixel-wise marginal distributions. Such a model may be useful in predicting pixels where predictions may be inaccurate, however, it cannot propose different possible reconstructions. For this, non-Gaussian pixel-wise distributions and non-trivial statistical dependencies across different pixels are crucial to capture. Hence these models are limited in the information they can provide [132]. Being able to generate samples makes uncertainty propagation trivial for any following task while only generating second order statistics makes this non-trivial.

One work that has similar aims as this one is by Adler et al. [133]. In this work, the authors train a modified conditional Wasserstein generative adversarial network (cWGAN) that generates high dose counterparts of CT images from low-dose measurements. However this approach does not explicitly model the known physics or the measurement noise of the imaging procedure, lacking an explicit data likelihood term. As such, there are no guarantees that the samples will be from the true posterior. Instead of using an explicit physics-based imaging model, the cWGAN requires supervised training with pairs of undersampled-fully sampled images, hence a separate cWGAN has to be trained for all different undersampling schemes and factors for best performance. Furthermore, though the authors modify their discriminator to reduce the mode collapse associated with the WGAN, they do not completely avoid it, leading to a possibly poorer implicit prior. Lastly, the model aims to minimize the Wasserstein distance between the predicted posterior and the true posterior. However, it is often impossible to have varying samples from the true posterior, in reality only one fully measured image is available for each low-dose measurement image. Therefore, the available training samples may not be able to support an accurate posterior approximation and minimizing the Wasserstein distance may converge to a degenerate version of the true posterior.

Another work in line with our purposes is by Pedemonte et al. [134], where the authors use Hamiltonian Monte Carlo (HMC) to sample from the posterior of emission rates given the photon counts for positron emission tomography. However, they use a uniform, i.e. a non-informative prior for the emission rates, reducing the strength of the model heavily in contrast to using more informative, data-driven priors. Furthermore the authors

use a Riemannian HMC scheme [135] to make sampling efficient despite the high dimensional posterior, where the use of the Riemannian metric speeds up the sampling by taking the geometry of the space of probability distributions into account. However this does not directly take into account the geometry of the space of the emission rates.

Here we identify two ideas that motivate us in proposing a new method that overcomes the limitations of the works mentioned above. The first one is regarding the geometry of the space of MR images. We make the assumption that the MR images actually live around a low dimensional subspace in the high dimensional image space and that we can learn a mapping from a low dimensional latent space to this subspace. This assumption has been demonstrated empirically in our prior work [29]. Then "walking around" and sampling in the subspace of MR images can be simply implemented as walking around and taking samples in the latent space. Secondly, deep learning based data-driven priors have shown great value in inverse problems in general in the recent years, as well as specifically in MR image reconstruction [29, 136, 137]. Such methodology allows learning a powerful mapping between the latent space and the image space, facilitating the sampling.

Embodying the ideas mentioned above, we propose a novel method based on a latent Bayesian model and Markov chain Monte Carlo (MCMC) sampling that addresses the issue of uncertainty in the inversion process. To this end, we use a variational autoencoder (VAE) [88, 89] trained on fully sampled MR images as our prior as in Chapter 3 and utilize its lower dimensional latent space to do the sampling instead of sampling in the high dimensional image space. For the target distribution of the MCMC, we use the posterior of the latent vectors given the measured k-space data. We obtain this distribution by marginalizing over the images. We use the Metropolis adjusted Langevin Algorithm (MALA) [135] as the MCMC method due to its effectiveness in high dimensional spaces. The latent samples coming from MALA are then guaranteed to be from the posterior and can be transformed to images using the decoder of the VAE. Although we use a VAE and MALA in our implementation, the framework is generic and can be used with other generative models as well as sampling schemes. We evaluate our method with data from the Human Connectome Project (HCP) [97] as well as in-house measured images [29] for changing settings of undersampling ratios and measurement noise levels and compare it to two other methods.

5.2 METHODS

We are interested in obtaining samples from the posterior distribution $p(x|y)$ of images $x \in \mathcal{C}^N$ given the observed undersampled noisy k-space data $y \in \mathcal{C}^{Mc}$ with c coils, (M≤N). We model the acquisition as $y = Ex + \eta$, where $E \in \mathcal{C}^{Mc \times N}$ is the extended MR encoding operation and η is complex Gaussian noise in the k-space with $\eta \sim N(0, \Sigma_{ns})$ with Σ_{ns} as the noise covariance matrix. Hence the data likelihood term is given as $p(y|x) = N(y; Ex, \sigma_{ns})$. The extended encoding operation comprises of the usual coil sensitivity encoding [44], Fourier transform and undersampling operations, and additionally of a scaling factor, a padding operator, an operation for combining the phase with the magnitude image and an operator for modeling the bias field [116, 120] in the image, which we explain further in Section 5.2.5.

Unfortunately, it is difficult to draw samples directly from their posterior $p(x|y)$ since the x's are very high dimensional. This renders simple sampling methods such as rejection sampling or vanilla MCMC [90] very inefficient since they would need too many samples to adequately travel around in the space to generate a good representation of the posterior. Moreover, the assumption that the high probability regions in the image space form a lower dimensional subspace would drastically lower the acceptance ratio for the samples. For instance, if the sampling is done in the image space using MCMC, the random walk will try to move in dimensions that will take it out of this lower dimensional subspace and sampling will become very inefficient. On the other hand, if we have a latent space connected to the lower dimensional subspace with a decoder, then we can implement the random walk in the latent space, and the images decoded from the latent samples will always be around this lower dimensional subspace.

The proposed model here samples directly in the lower dimensional latent space to address the difficulties mentioned above. To this end we use a VAE as the latent space model. This allows us to first sample latent vectors and then use the decoder of the VAE to obtain images from these.

In the following we describe the method in more detail.

5.2.1 *Sampling in the latent space*

Firstly, we need a prior term for the MR images with which we can evaluate the probability of a given image and also differentiate. To this end, we

use a VAE [88, 89] trained on fully sampled MR images. The trained VAE consists of an encoder $q(z|x) = N(\mu_z(x), \Sigma_z(x))$ and a decoder $p(x|z) = N(\mu_x(z), \Sigma_x)$ parameterized by neural networks with $z \in \mathcal{R}^D$ (D<<N) as latent vectors distributed according to a Gaussian prior $p(z) = N(\mu_{pr}, \Sigma_{pr})$. We drop the x and z dependencies in the rest of the text unless necessary. Here we use a diagonal non-isotropic covariance matrix for Σ_z, an isotropic diagonal matrix for Σ_x and a block diagonal for the Σ_{pr}. The VAE is trained to maximize the evidence lower bound (ELBO) which approximates the log likelihood $p(x)$. For the prior $p(z)$, we empirically estimate the parameters μ_{pr} and Σ_{pr} from training data [138], which is different than the case in the vanilla VAE, which we explain in Section 5.2.4.

Now, given the VAE, we can formalize our aim as to sample from the posterior of latent variables given the undersampled k-space data, namely $p(z|y)$. By doing so we obtain latent samples that match the measured k-space data and thus, the images associated with these latent samples will match the data too. However, before we describe how we obtain this distribution, we first introduce the sampling procedure.

Sampling from $p(z|y)$ can be implemented using different methods. In this work, we use the Metropolis adjusted Langevin algorithm (MALA) [135]. MALA is a variant of Markov chain Monte Carlo, consisting of a random walk given by Langevin dynamics and an acceptance scheme following the Metropolis-Hastings algorithm. The random walk for MALA with the target distribution $p(z|y)$ is written as

$$\hat{z}^{t+1} = z^t + \tau \nabla_z \log p(z|y)|_{z=z^t} + \sqrt{2\tau}\zeta, \;\; \zeta \sim N(0,1) \tag{5.1}$$

with step size τ. As can be observed, the update step for the random walk is composed of two terms. The second term models the randomness with a Gaussian distributed variable ζ, aiming to discover the space equally in all directions. As the latent space can be large, only sampling with the random part can be inefficient for sampling from $p(z|y)$. Here, the first term $\tau \nabla_z \log p(z|y)$ comes to aid by pulling the steps towards the high probability regions and preventing the random walk to move far away from such regions. However, the discrete nature of the walk requires a Metropolis-Hastings correction to be applied to ensure convergence. This means a sample is accepted as $z^{t+1} = \hat{z}^{t+1}$ with log probability $\alpha = \min\left\{0, \log\left[\frac{p(\hat{z}^{t+1}|y)q(z^t|\hat{z}^{t+1})}{p(z^t)q(\hat{z}^{t+1}|z^t)}\right]\right\}$, otherwise $z^{t+1} = z^t$. The proposal distribution for MALA is given as $q(z' \mid z) \propto \exp\left(-\frac{1}{4\tau}\|z' - z - \tau\nabla \log p(z|y)\|_2^2\right)$. Furthermore, in order to avoid a long burn-in period, we can initialize the

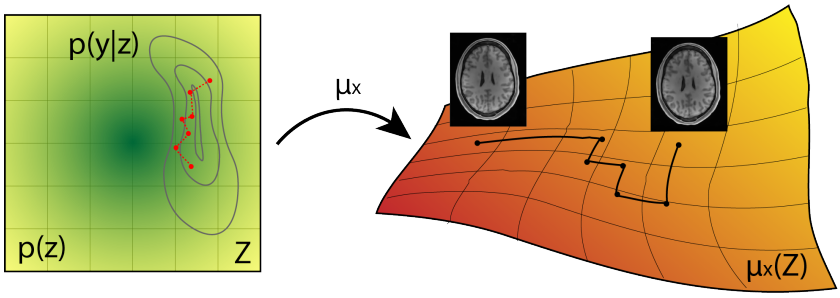


FIGURE 5.1: Illustration of sampling in the latent space. Left side shows the latent space equipped with a prior shown by the green color. The contours show the regions where z values result in high data likelihood values for the measured data. The random walk (red line) samples from the product of these two, i.e. the posterior $z^t \sim p(z|y) \propto (y|z)p(z)$. Right side shows the low dimensional subspace around which the MR images reside. Each sample in the Z space correspond to a distribution of images in the subspace, s.t. an image can be taken using the the mean of decoder of the VAE as $x^t = \mu_x(z^t)$.

chain close to the mode of the posterior instead of starting with a randomly chosen z^0. This can be achieved by encoding the MAP image into the latent space and using its mean as z^0.

Once the posterior samples $\{z^t\}$ are obtained, we then take samples from the posterior of the images as $x^t = \mu^t$, i.e. the mean of the decoder $p(x|z)$ when $z = z^t$. Notice that this procedure is quite similar to ancestral sampling due to the hierarchical structure of the model, however, without the sampling step on the image level. Now these image samples can be passed on to further tasks or used to calculate empirical statistics such as pixel-wise mean and variance.

5.2.2 *Obtaining the posterior $p(z|y)$*

The main component required for the random walk is the unnormalized posterior distribution of z, i.e. $p(y|z)p(z) \propto p(z|y)$, where we do not need the normalization constant $p(y)$ since it does not appear in the derivative nor the acceptance terms. To construct it, we use the trained VAE model with

its prior $p(z) = N(\mu_{pr}, \Sigma_{pr})$. We write the $p(y|z)$ term as a marginalization over the images as

$$p(y|z) = \int p(y, x|z)dx = \int p(y|x)p(x|z)dx, \tag{5.2}$$

where we use the decoder of the VAE as the $p(x|z)$ and the conditional independence assumption $p(y|x, z) = p(y|x)$. This expression can be interpreted as two terms glued together with the images functioning as the intermediate variables, connecting the latent space with the k-space. This integral can be evaluated analytically to yield another Gaussian distribution [139].

Here we do the derivation of this expression using conjugacy relations for Normal distributions instead of directly evaluating the integral. We begin by writing

$$p(y|x, z)p(x|z) = p(y|z)p(x|y, z). \tag{5.3}$$

Since $p(y|x, z)$ and $p(x|z)$ are Normal distributions, due to the conjugacy, the posterior $p(x|y, z)$ is also a Normal distribution given as $N(\mu_{post}, \Sigma_{post})$. Then

$$p(y|z) = \frac{p(y|x, z)p(x|z)}{N(\mu_{post}, \Sigma_{post})}, \quad \text{or} \quad p(y|z)N(\mu_{post}, \Sigma_{post}) = p(y|x, z)p(x|z). \tag{5.4}$$

Hence the posterior $p(y|z)$ acts as a normalizer to the product distribution to yield a Gaussian. We derive $p(y|z)$ using this relation in Eqn. 5.4. In the following we also use the conditional independence $p(y|x, z) = p(y|x)$ meaning that when the image is given, this posterior distribution in the k-space is determined without the need for the latent variable.

For the derivation we use this strategy: i) we first write the product of the two distributions $p(y|x)p(x|z)$, ii) then recognize the mean and covariance of the Normal posterior distribution $N(\mu_{post}, \Sigma_{post})$ in this, iii) and separate a the Gaussian with these parameters from the whole expression. What is left gives us the target distribution.

The product can be written as

$$p(y|x)p(x|z) = \tag{5.5}$$

$$\det(2\pi\Sigma_{ns})^{-1/2}\det(2\pi\Sigma_x)^{-1/2}\exp\left\{-\frac{1}{2}\left[(y-Ex)^H\Sigma_{ns}^{-1}(y-Ex)\right]\right\} \tag{5.6}$$

$$\cdot\exp\left\{-\frac{1}{2}\left[(x-\mu_x)^H\Sigma_x^{-1}(x-\mu_x)\right]\right\} \tag{5.7}$$

$$= \det(2\pi\Sigma_{ns})^{-1/2}\det(2\pi\Sigma_x)^{-1/2}\exp\left\{-\frac{1}{2}x^H\underbrace{(\Sigma_x^{-1}+E^H\Sigma_{ns}^{-1}E)}_{\Sigma_{post}^{-1}}x\right. \tag{5.8}$$

$$\left.+Re\{x^H\underbrace{(E^H\Sigma_{ns}^{-1}y+\Sigma_x^{-1}\mu_x)}_{\Sigma_{post}^{-1}\mu_{post}}\}-\frac{1}{2}y^H\Sigma_{ns}^{-1}y-\frac{1}{2}\mu_x^H\Sigma_x^{-1}\mu_x\right\}, \tag{5.9}$$

where we have recognized the parameters of the posterior. With these we have enough information to complete the posterior Gaussian. We can replace the terms with posterior parameters and add the missing term $\pm\frac{1}{2}\mu_{post}^H\Sigma_{post}^{-1}\mu_{post}$ to complete the quadratic form as well as the normalizing determinant $\det(2\pi\Sigma_{post})^{\pm 1/2}$.

$$= \det(2\pi\Sigma_{ns})^{-1/2}\det(2\pi\Sigma_x)^{-1/2}\det(2\pi\Sigma_{post})^{+1/2}\det(2\pi\Sigma_{post})^{-1/2} \tag{5.10}$$

$$\cdot\exp\left\{\underbrace{-\frac{1}{2}x^H\Sigma_{post}^{-1}x+Re\{x^H\Sigma_{post}^{-1}\mu_{post}\}-\frac{1}{2}\mu_{post}^H\Sigma_{post}^{-1}\mu_{post}}_{-\frac{1}{2}(x-\mu_{post})^H\Sigma_{post}^{-1}(x-\mu_{post})}\right. \tag{5.11}$$

$$\left.+\frac{1}{2}\mu_{post}^H\Sigma_{post}^{-1}\mu_{post}-\frac{1}{2}y^H\Sigma_{ns}^{-1}y-\frac{1}{2}\mu_x^H\Sigma_x^{-1}\mu_x\right\}. \tag{5.12}$$

We can combine the quadratic term in the exponent with the determinant term and obtain the complete posterior Gaussian. In this case the expression becomes

$$p(y|x)p(x|z) = N(\mu_{post},\Sigma_{post})\det(2\pi\Sigma_{ns})^{-1/2}\det(2\pi\Sigma_x)^{-1/2} \tag{5.13}$$

$$\cdot\det(2\pi\Sigma_{post})^{+1/2}\exp\left\{+\frac{1}{2}\mu_{post}^H\Sigma_{post}^{-1}\mu_{post}-\frac{1}{2}y^H\Sigma_{ns}^{-1}y-\frac{1}{2}\mu_x^H\Sigma_x^{-1}\mu_x\right\}. \tag{5.14}$$

Remembering Eqn. 5.4, we obtain

$$p(y|z) = \frac{\det(2\pi\Sigma_{post})^{+1/2}}{\det(2\pi\Sigma_{ns})^{1/2}\det(2\pi\Sigma_x)^{1/2}} \cdot \exp\left\{ -\frac{1}{2}y^H\Sigma_{ns}^{-1}y + \frac{1}{2}\mu_{post}^H\Sigma_{post}^{-1}\mu_{post} - \frac{1}{2}\mu_x^H\Sigma_x^{-1}\mu_x \right\}. \quad (5.15)$$

Now taking the logarithm and leaving out the terms that are independent of z we can arrive at the expression we use as

$$\log p(y|z) = +\frac{1}{2}\mu_{post}^H\Sigma_{post}^{-1}\mu_{post} - \frac{1}{2}\mu_x^H\Sigma_x^{-1}\mu_x + C, \quad (5.16)$$

where C denotes some constant with z. Notice that we could leave out the determinant term in the nominator due to our model choice of constant Σ_x.

Now we need the closed form expression for the first term in the above equation. Also we need to arrive at this using the terms we have access to from the above equations 5.8 and 5.9, namely $\Sigma_{post}^{-1}\mu_{post}$ and Σ_{post}^{-1}. First we write $\mu_{post} = (\Sigma_{post}^{-1})^{-1}\Sigma_{post}^{-1}\mu_{post}$ and rewrite the target term as $\mu_{post}^H\Sigma_{post}^{-1}\mu_{post} = (\Sigma_{post}^{-1}\mu_{post})^H\mu_{post}$. Combining the expressions and isoliting the terms constant with z as C then yields

$$\mu_{post}^H\Sigma_{post}^{-1}\mu_{post} = \mu_x^H\Sigma_x^{-1}(\Sigma_x^{-1} + E^H\Sigma_{ns}^{-1}E)^{-1}\Sigma_x^{-1}\mu_x \quad (5.17)$$

$$+ 2\mathrm{Re}\left\{ y^H\Sigma_{ns}^{-1}E(\Sigma_x^{-1} + E^H\Sigma_{ns}^{-1}E)^{-1}\Sigma_x^{-1}\mu_x \right\} + C \quad (5.18)$$

Applying the Woodburry identity on the term $(\Sigma_x^{-1} + E^H\Sigma_{ns}^{-1}E)$ followed by some algebraic manipulations reveals that this is equivalent to the expression given in [139].

Putting the terms together yields the distribution given as

$$\log p(y|z) = \mu_x^H\Sigma_x^{-1}(\Sigma_x^{-1} + E^H\Sigma_{ns}^{-1}E)^{-1}\Sigma_x^{-1}\mu_x + 2\mathrm{Re}\left\{ y^H\Sigma_{ns}^{-1}E(\Sigma_x^{-1} + E^H\Sigma_{ns}^{-1}E)^{-1}\Sigma_x^{-1}\mu_x \right\} - \frac{1}{2}\mu_x^H\Sigma_x^{-1}\mu_x + C. \quad (5.19)$$

where ^{H} denotes the conjugate transpose and C is some constant w.r.t. z.

For the random walk we also need the log gradient of the target distribution, which we write as $\nabla_z \log p(z|y) = \nabla_z \log p(y|z) + \nabla_z \log p(z)$. The $\nabla_z \log p(y|z)$ term can be easily obtained by automatic differentiation since

outputs of the decoder μ_x and Σ_x are modeled as neural networks and, thus, differentiable w.r.t z, given that we can implement $p(y|z)$ in a software package that allows such differentiation. Similarly, the prior $\nabla_z \log p(z)$ term is also straightforward and can be derived analytically. The more challenging part is the matrix inversion $(\Sigma_x^{-1} + E^H \Sigma_{ns}^{-1} E)^{-1}$ in Eq. 5.19, which we investigate in the next section.

Finally, looking at the two terms in the posterior $p(z|y) \propto p(y|z)p(z)$ reveals more insights regarding the method. The first term drives the chain to regions where the z^t values, when decoded as $p(x|z^t)$, lead to images which satisfy the data likelihood term $p(y|x)$ for x's coming form the $p(x|z^t)$. On the other hand, the second term $p(z)$ tries to pull the chain towards the middle of the empirical Gaussian in the latent space, discouraging the chain to move away from the meaningful regions of the latent space. As such, a random walk in the latent space with this target distribution explores the areas which satisfy the data likelihood and the prior terms simultaneously.

One natural alternative to the proposed method is to use the approximation posterior $q(z|x)$ modeled in VAE directly to do the sampling around a MAP estimate x_{MAP}, i.e. taking latent samples around the conditional distribution $z^t \sim q(z|x_{MAP})$ and then again decoding these to the image space. This approach is fundamentally different than the proposed method in that it takes the MAP image as the "true reconstruction" and samples only locally around it. Though these locally sampled images will still be in the subspace of MR images and show some structural variation, they do not explore the possibility of different images being the true underlying image. Hence the local sampling method is inherently very limited and cannot in general identify the regions where the reconstruction has failed. In contrast, the proposed method can "globally" explore the latent space as long as the data likelihood is satisfied. Furthermore, $q(z|x)$ is only an approximate posterior distribution for a given $p(x|z)$ and $p(z)$ while we extract samples from the exact posterior. We compare the two methods also experimentally and show results for the local sampling in Section 5.4 as well.

5.2.3 *Derivatives through the iterative matrix inversion*

For the sampling procedure, we need to evaluate the terms in Eq. 5.19 at each iteration. This means the inverse of $(\Sigma_x^{-1} + E^H \Sigma_{ns}^{-1} E)$ has to be recomputed at each iteration, if the Σ_x term depends on the z value. In this work we take Σ_x to be constant, however, even then, inverting the

matrix once and keeping it in memory is not an option, since it is a very big ($Nc \times Nc$) matrix. Approximating it as a diagonal matrix is also not an option as this would result in the loss of the aliasing information kept in the off-diagonals of the $E^H \Sigma_{ns}^{-1} E$ term.

Instead we propose to use an iterative matrix inversion that can also be applied when Σ_x changes with z. To this end, we write the common term in Eq. 5.19 in an approximation as

$$\gamma^* = \min_{\gamma} ||(\Sigma_x^{-1} + E^H \Sigma_{ns}^{-1} E)\gamma - \Sigma_x^{-1} \mu_x||_2^2. \tag{5.20}$$

We then solve this inversion as an optimization problem using conjugate gradients (CG) and obtain γ^*, which we plug-in to Eq. 5.19 to yield

$$\log p(y|z) = \mu_x^H \Sigma_x^{-1} \gamma^* + 2\mathrm{Re}\left\{ y^H \Sigma_{ns}^{-1} E \gamma^* \right\} - \frac{1}{2} \mu_x^H \Sigma_x^{-1} \mu_x, \tag{5.21}$$

where we dropped the constant C.

Though we use CG here, other gradient based methods can be used as well, since the gradients are well defined. Furthermore fast Fourier transform (FFT) can be used in the operations E and E^H, revoking the need to write the matrix explicitly and speeding up computations.

The key advantage here is that when the number of iterations (N_γ) for the CG is fixed, the whole inversion optimization also becomes a fixed operation, where the related vectors and matrices are added or multiplied with each other with fixed coefficients. Moreover, since automated differentiation is defined for all these operations seperately, the gradient through the whole inversion operation can also be taken using automated differentiation. This allows us to take the gradients of $p(y|z)$ according to z through this fixed optimization. Hence, before starting the Markov Chain we select the parameter N_γ for which we obtain a small L_2 error in Eq. 5.20 and keep this throughout the sampling. We observe that the error stays small throughout the sampling process for the chosen parameters for different μ_x values.

5.2.4 *2D latent space VAE with an empirical prior*

In the following we introduce the modified VAE model we use. Here we apply two modifications to the vanilla VAE model [88]. Firstly, the VAE is fully convolutional and we use a 2D latent space, i.e. L_1 x L_2 x D dimensions with two spatial and one channel dimension, effectively a latent image with D channels. Each spatial position in such a latent image has a receptive

field on the image when traced back and the architecture is designed using strided convolutions such that the receptive fields have only minimal overlap. This allows us to adhere to the independence assumption of the latent pixels reflected in the model by the use of a diagonal covariance matrix for $q(z|x)$. However, in reality, contents in receptive fields corresponding to different spatial locations in a latent image are not entirely independent from each other, as there are global correlations in the image. To be able to model these, we introduce an empirical prior as $p(z) = N(\mu_{pr}, \Sigma_{pr})$, similar to [138].

We obtain the parameters of $N(\mu_{pr}, \Sigma_{pr})$ empirically by sampling T samples $z_i \sim q(z|x_i)$ from T different training images x_i, after the VAE has been trained using a unit Gaussian prior. The mean μ_{pr} then is calculated as the mean of these samples. The estimation of a full covariance matrix is, however, difficult due to i) its size and ii) large number of samples required for the estimation. Instead we apply a Kolmogorov-Smirnov test [140] against the unit Gaussian seperately for each latent channel to find the channels that are the least unit Gaussian in terms of the p-values, i.e. approximately the most informative. Then we form a combined block diagonal covariance matrix, where we calculate the full covariance matrix for the K most informative channels (of size KL_1L_2 x KL_1L_2) and for the rest of the channels we assume they are independent from each other and calculate seperately for each channel only the spatial covariance matrix (of size L_1L_2 x L_1L_2). The proper combination of these block matrices yields the Σ_{pr}. In practice we do not form this matrix but implement the operations as sparse matrix-vector multiplications. This strategy also allows us to also reduce the number of samples T required for the estimation. We set $K = 10$ as preliminary experiments have shown this covers the informative channels sufficiently. We also set T high enough to make sure the estimated covariance matrix is full rank and found $T = 20000$ to be sufficient.

5.2.5 *The extended encoding matrix*

We extend the usual encoding operation in order to be able to model additional effects of the image acquisition. The aim while modeling these is to close the domain gap between a trained VAE and an observed k-space data by integrating acquisition specific knowledge as much as possible. Let us assume we have a trained VAE and observed a k-space data for the rest of this section.

The extended encoding matrix E is based on the usual MR encoding matrix $\tilde{E} = UFS$, with $S : \mathcal{C}^N \to \mathcal{C}^{Nc}$ the sensitivity encoding matrix [44] with c coils, $F : \mathcal{C}^{Nc} \to \mathcal{C}^{Nc}$ the coilwise Fourier transform, $U : \mathcal{C}^{Nc} \to \mathcal{C}^{Mc}$ the undersampling operation. Then the extended encoding operation is given as

$$E = \tilde{E}B\varphi Ps. \tag{5.22}$$

In the following we explain each term seperately.

Firstly, a discrepancy between the k-space data and the space of images on which the VAE operates can be due to differences in the field of view (FOV). Though our fully convolutional architecture is agnostic to the image size, the empirical prior is estimated for a specific resolution and FOV, and the k-space size can be different due to varying FOV during acquisition. To bridge this gap, we introduce a padding operation P that pads or crops the images to fit the required sizes for the VAE.

Secondly, for computational as well as implementation related purposes, we assume that the phase of structural images are highly independent of the magnitude image and smooth, and hence a single phase image can be used for all posterior samples we take. This allows us to separate the magnitude and the phase of the image and run the sampling only on the magnitude of the image. However, note that this assumption is not a requirement for the proposed idea as the phase could be sampled as well, but rather a methodological simplification motivated by empirical observations. Following this assumption, we write the phase as a diagonal matrix φ acting on the image.

Thirdly, we use a diagonal matrix B that explicitly models the bias field in the acquisition [116] as we did in Chapter 4. As discussed, the MR images unavoidably have a bias field due to several factors [120], and the bias field is difficult to avoid, however easy to estimate from the measured data. As the bias varies between different acquisitions, this is a potential source of discrepancy leading to a domain shift. In order to minimize this, we train the VAE on bias free images, which then gives us bias free image samples. However, as the measured data y has the bias field in it, we estimate this field and explicitly model it to bring the sampled images to the biased domain of the k-space data.

Finally, we introduce a scale factor to make the data likelihood invariant to any scaling difference between the samples and the k-space. During the random walk in the latent space, the corresponding images might get scaled at each step, meaning the image may be multiplied globally by a scale factor. If this scale factor moves away from 1, this causes the data likelihood value

to decrease, i.e. the data consistency error to increase, since the scales of the k-space data and the image samples do not match. However, from the perspective of sample quality, this does not pose a problem as long as the scaling factors are known. The sampled images can be brought to the same scale by multiplying them with the inverse of the scaling factor. Furthermore, allowing the scale factor to be different for each sample, allows more freedom to the random walk in the latent space, as it is less constrained by the increase in data likelihood due to scale changes. Hence such an invariance to this scaling is desirable. To this end we introduce a scalar s, that keeps the data likelihood at the lowest, inducing an invariance to scaling. We calculate its value by solving $s^* = \min_s ||Es\mu_x(z^t) - y||_2^2$, where we separated only the s term from the extended encoding and used the mean of the decoder as the image. Then we take the derivative of the expression according to s and set it to zero to obtain the minimizing s value, which is given analytically as $s^* = \frac{Re\{\mu_x(z^t)^H E^H y\}}{\mu_x(z^t)^H E^H E \mu_x(z^t)}$. We do this estimation seperately for each z^t sample at each step. Notice that the scaling allows images to move towards the pixel intensity range that the prior prefers, hence if the pixel intensity values of the initialization match these, even though there are no constraints on the s value, it will stay reasonably close to 1.

The complex conjugate of the extended encoding operation operation is given as $E^H = s^H P^H \varphi^H B^H \tilde{E}^H$, where we implement P^H as cropping if P is a padding operation and vice versa, φ^H is multiplication with the complex conjugate of the phase, $B^H = B$ since the bias field is real and $s^H = s$, again since the scale factor is real.

5.3 EXPERIMENTAL SETUP

5.3.1 *Data, training details and compared methods*

We used T1 weighted slices from the full 3D volumes of 780 subjects from the HCP dataset [97] for training of the VAE. There were in total 202800 slices of size 252x308, with an isotropic resolution of 0.7 mm. We ran the N4 bias field correction on the images and used the corrected images for training. The training ran for 2250000 iterations. We also trained another VAE after downsampling the images to 1mm isotropic resolution to work with lower resolution images for 1750000 iterations. For both, we augmented the images by translating them randomly (-4 to +4 pixels) in both directions and trained till convergence.

For testing we used 4 axial slices from subjects in the HCP dataset, different than those used in training, without bias field correction. We additionally tested with 3 axial slices from in-house measured T1 weighted brain images of different subjects [29]. These images have similar acquisition parameters as HCP and have an isotropic resolution of 1 *mm*. Furthermore these images are acquired with 13 coils and have non-zero phase. We used ESPIRiT [100] to obtain the coil sensitivity maps, which we used in the MAP estimation and sampling.

For the experiments, we retrospectively apply Cartesian undersampling to the test images with different patterns for each image. We obtain these patterns by generating 100 different patterns and choosing the one with the highest peak-to-side ratio of the associated point spread functions. In all patterns the 15 central profiles are always sampled.

For comparison purposes, we also modified the code by Adler et al. [133] to work with magnitude of MR images and evaluated in our experimental setting (the authors provided private access to their repository for the code). This method requires supervised training, i.e. pairs of zero-filled and fully sampled images. We generated such a training set with zero-filled images by undersampling the training images, which were also used for the VAE. We used the images with their bias field. We generated an undersampling pattern as described above, seperately for each of the images. We trained the cWGAN for 800000 iterations till convergence with the the decay ratio for the noisy linear cosine decay as 2000000 but otherwise with the default settings in the code provided by the authors and the augmentation used for the VAE.

We also trained a feed-forward heteroscedastic network as a baseline. This network has the same architecture as the VAE, without the KLD in the loss and outputs a pixelwise standard deviation alongside the mean prediction. It is trained for 4000000 iterations using the supervised training setup described for the cWGAN method.

Finally, we implement the local sampling method we described in Section 5.2.2 for comparison purposes. For these we use the same VAE that is used in the proposed method. We take the samples around the MAP reconstruction.

5.3.2 *Implementation Details*

We initialize the chain with the maximum-a-posteriori (MAP) images obtained by the deep density prior (DDP) reconstruction [29] to avoid a long

burn-in period and take 10000 samples in total. We empirically determine the step size $\tau = 4 \times 10^{-4}$ to obtain an acceptance ratio around 0.3-0.5 and use the same for both the HCP images the in-house measured images. We take a lower number of samples (1000) for the cWGAN and local sampling methods as the effective sample size is also lower for the MCMC chain due to correlated samples.

We used Tensorflow [141] for the implementation of VAE related parts of the proposed method. The VAE is fully convolutional with all padded convolutions and has a 2 dimensional latent space with D=60 channels. We describe the the details of the architecture in the next section (Section 5.3.3). For Σ_x we use a diagonal matrix with equal diagonal values set at 0.02. The same value was used for training of the VAE and sampling. Σ_{ns} was also taken to be a diagonal matrix, where the values were estimated by taking the variance of a small region at the upper 10 pixels of the fully sampled k-space center in the undersampled data, seperately for each coil. Though this approach neglects the covariance between coils, it allows for a different variance per coil, which is important as these can differ significantly. Alternatively, the data can be pre-whitened to decorrelate the coils. The number of iterations for the matrix inversion were determined empirically as $N_\gamma = 25$, which was enough to reduce the L2 error of the approximate inversion below 0.01%.

For padding we use simple zero padding and cropping. For bias field estimation we used the N4 method [121] on the magnitude of the MAP estimation with default parameters. For the phase of the samples we took the MAP phase estimate.

5.3.3 *Details of the VAE architecture*

All convolutions are padded and have a kernel size (3, 3) and stride (1, 1) and use a ReLU unless noted otherwise.

The encoder begins with four convolutional layers with 32, 64, 64, 64 output channels, respectively. Then a convolutional layer with kernel size (14, 14), stride (19, 19) and 60 output channels produces the mean of $q(z|x)$ from the fourth layer. Similarly another convolutional layer from the third layer produces the log standard deviation values for $q(z|x)$ with a kernel size of (14, 14), stride (19, 19), without ReLU and 60 output channels. The network is fully convolutional, hence can work with different image sizes. Assuming an input image size of 252x308 for demonstration, the latent space size becomes bx18x22x60, where b is the batch size. We use the usual

reparameterization trick to sample z's [88]. At the beginning of the decoder, we apply a scheme of increasing channel dimensions and using these to increase spatial dimensions. We do this in two steps, once for the first image dimension and once again for the second image dimension to obtain a proper reshaping while using the implementation of Tensorflow's reshaping function. First convolutional layer of the decoder does not use ReLU and has 64·19=1216 output channels, resulting in a tensor of size bx18x22x1216. The output of this layer is first transposed to bx18x1216x22 and reshaped to bx252x64x22. This layer then gets transposed to bx252x22x64, then goes through a convolutional layer with again 1216 output channels and without ReLU and becomes bx252x22x1216. This then gets reshaped again to yield a tensor size of bx252x308x64, which is the input image size. This tensor then goes through a ReLU. We then apply 6 convolutional layers with each 60 output channels. Finally another convolutional layer with 1 output channel yields the mean prediction.

5.4 RESULTS

We show most of the results at high undersampling factors on purpose to make sure that the uncertainty in the inversion is high and demonstrate that the model is able to capture it. We present the results with the bias field put back in for convention although the method provides the samples bias free. We also multiply the images with their corresponding scale values to bring them to the same scale as the observed k-space data. The scale values stay mostly in the 0.95-1.1 range throughout the sampling procedure. We note that it is quite difficult to see the variations in the samples presented in print as the pixels move together in the structures, rendering it difficult for the eye to see the subtle changes in the structures. Such changes, however are far easier to notice when contrasted with the same structure in the different samples. As such, the video format lends itself for inspection as the eye can focus on the same region of the image while the structures are changing as each frame of the video shows another sample. Hence we highly encourage the reader of the thesis to view the provided GIFs[1].

We start by showing three sample images from the latent MALA model in Fig. 5.2 for an image undersampled with factor R=5. The structures in the samples as well as in the mean image, obtained as the mean of the drawn samples, overlap well with the fully sampled image. On the

1 GIFs available for the shown results and additional images: `https://polybox.ethz.ch/index.php/s/3DPrRoYQnyzANAF`

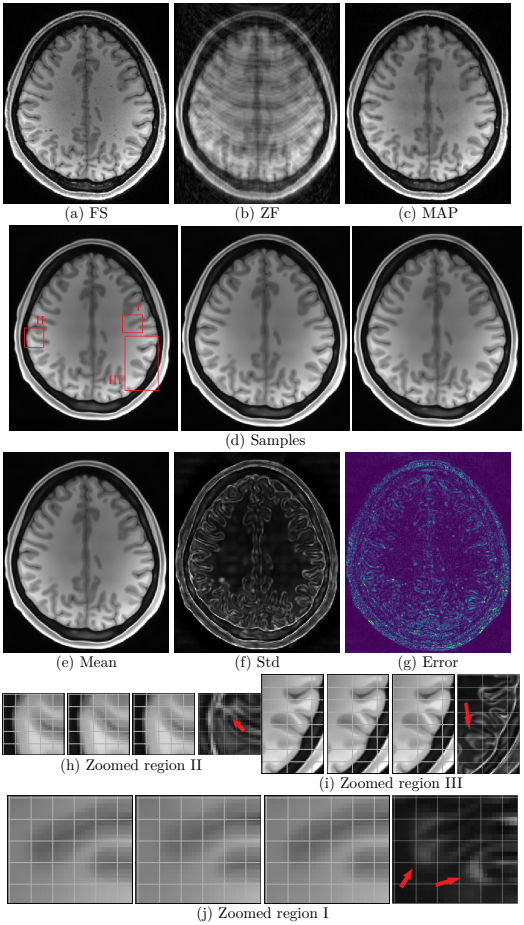


FIGURE 5.2: Results for the proposed latent MALA algorithm for R=5. FS, ZF and MAP denote the fully sampled, zero-filled and MAP estimation images. Second row presents three randomly chosen samples. (e-f) show the mean and pixelwise standard deviation (std) maps for all samples. (g) shows the absolute error map between the mean and the fully sampled image (clipped to 0-0.3). (h-j) show three zoomed-in regions indicated in (d) for three different samples as well as the pixelwise std maps. The grid lines are to aid visual inspection. As the variations are extremely difficult to see in this format, we strongly encourage the reader to look at the supplementary GIFs.

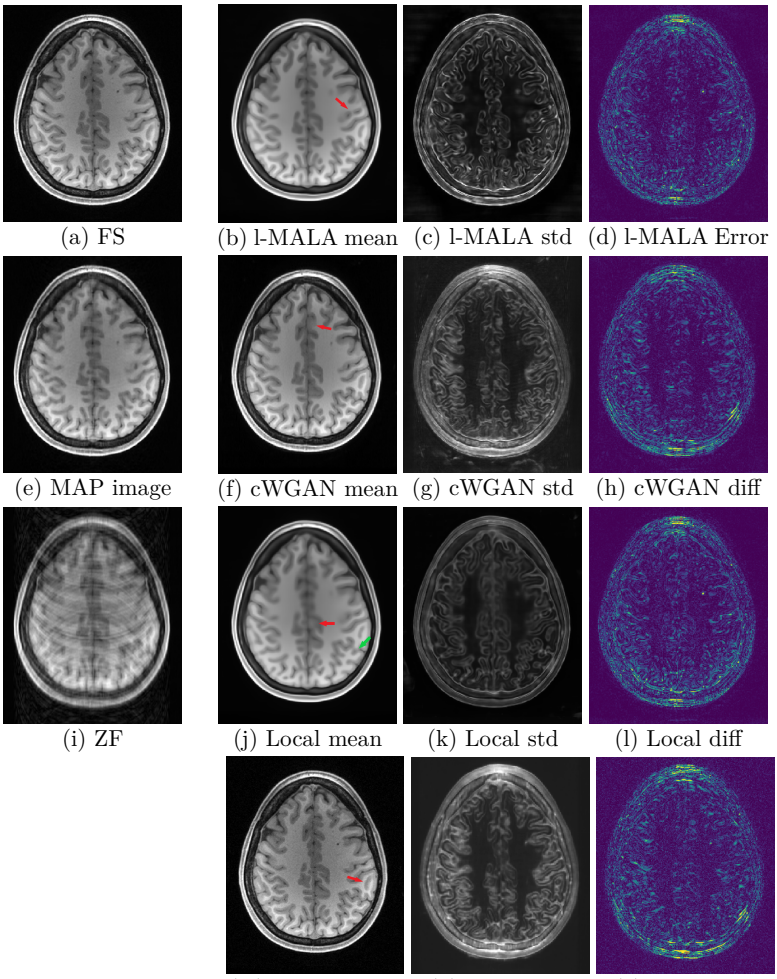
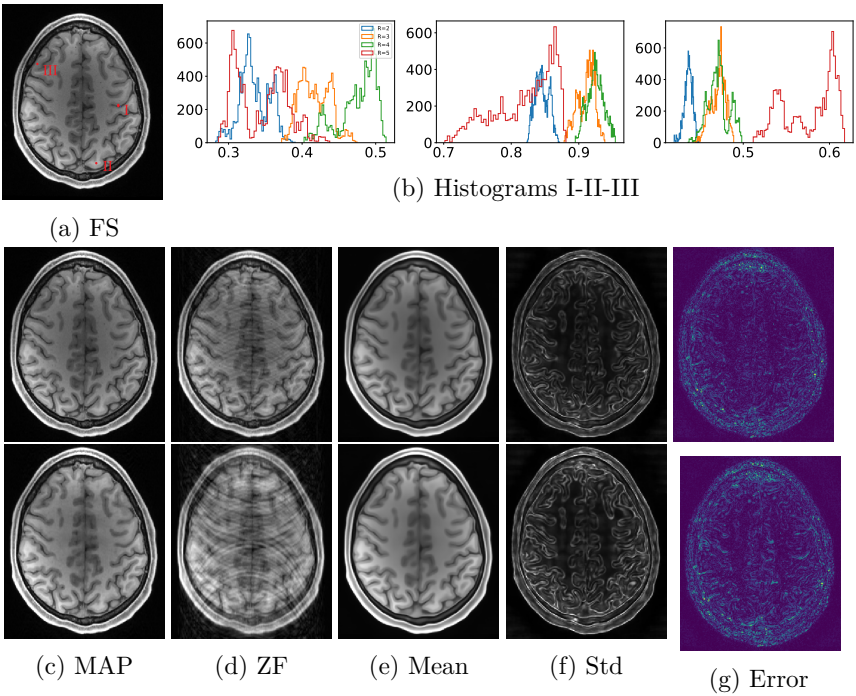


FIGURE 5.3: Sampling results for different methods at R=5. The left most column shows the fully sampled (FS), the MAP estimate and the zero-filled (ZF) images. In the rightmost three columns, the sample mean, pixel-wise standard deviations and the absolute error maps between the mean and fully sampled images from the respective method are given. The error maps are clipped to (0, 0.3), the std maps are clipped to (0, 0.06) and (0, 0.18) for the l-MALA and the other two methods, respectively.

other hand, structural variations between the samples are present, which can also be seen in the std map. Nearly all pixels corresponding to tissue edges have a high std value. This is expected since the missing data in the k-space is mostly in the high-frequency regions, whose contributions are more important for edge pixels. However, it is important to note that the std values on the edges are not homogeneous, indicating some parts of the edges have higher variability. Furthermore, the variations are not limited to edges but also structures in the white matter as well. Examples of these can be seen in the zoomed regions, indicated with arrows. Region II shows an example where the grey matter is connected and disconnected in different samples. In the lower part of region III one can see a gray matter structure inside the white matter becoming more and less visible in the samples. Similarly, in region I, one can see a gray matter structure showing variability in how deep it penetrates the white matter. The pixel-wise std maps are marginal maps, i.e. they present the variations in the pixels as if they were independent. In reality the variations are not pixel-wise, rather the structures as collections of multiple pixels move between different samples, which can be observed better in the GIFs.

Fig. 5.3 shows results for the different sampling methods for comparison purposes, namely the latent MALA, the cWGAN and the local VAE sampling. We use the same undersampling pattern for all the methods for comparability. Both the VAE cWGAN methods capture the underlying image fairly well in the mean of the samples, which is reflected in the difference images. The mean of the local VAE sampling is very blurry and cannot to capture the structures in the underlying image as well as the other methods. The heteroscedastic model performs worse in the mean prediction as expected due to the lack of a data consistency term. The pixelwise standard deviation maps for the VAE and cWGAN models are similar at first glance, in that both reflect the high uncertainty regions at the tissue edges. However, the cWGAN maps are quite noisier and blurrier in comparison. Latent MALA provides a much finer level distinction. This is expected since the proposed method generates samples based on examination of the given data instead of relying on a trained model to generalize and does not make assumptions about data availability from the joint distribution of fully and undersampled images as in cWGAN. The std maps from the heteroscedastic model yields even more blurry results. All methods except the local sampling are capable of indicating some of the regions where their mean maps differ from the ground truth image, by showing high diversity in the samples or high std values in those regions, as exemplified by the



(a) FS (b) Histograms I-II-III

(c) MAP (d) ZF (e) Mean (f) Std (g) Error

FIGURE 5.4: Results for changing undersampling ratios. First row shows the fully sampled image (FS) and histograms of pixels values in all samples for the pixels indicated on the FS image as I, II and III, respectively. Note the different bin positions for the histograms. Rows two and three show results for R=2 and R=4, respectively. Each row shows the MAP estimation, the zero filled image (ZF), the pixelwise mean and standard deviation maps and the absolute error map between the mean and the FS image (clipped to (0,0.3)).

arrows on the respective image. The local sampling method also captures variability on the edges. Furthermore, it can also indicate possible differences in its mean and the ground truth images, though only coincidentally. For example, in the region shown with the green arrow, it also assigns high std values, however this is rather due to the fact that there are two edges intersecting heavily in that region. In the region indicated by the red arrow, on the other hand, although the mean map differs from the ground truth, the local samples do not indicate a high variability in this region.

In Fig. 5.4, we show how the statistics from the samples change with changing undersampling ratios. Firstly, we show histograms from three pixels indicated on the FS image for R=2, 3, 4 and 5, from which one can observe that the pixel histograms become wider with increasing R, indicating higher uncertainty. This increase is also reflected in the std maps, which show an increase in std values for increasing R. This result shows that the proposed model is able to capture increasing ambiguity due to higher undersampling ratio.

Next we present results in Fig. 5.5 to show the methods sensitivity to the noise in the k-space. The quality of the MAP image degrades due to the high noise. This is reflected less in the mean maps, however the standard deviation values increase. This is how the model should behave since the added noise increases the values in Σ_{ns}, which then allows samples to move farther away from the measured data and show higher diversity. This is also reflected in the histograms of three pixel's intensities, which are indicated in the top std map, as the distributions become wider with increasing noise.

In Fig 5.6 we present results for an image from the multi-coil in-house measured dataset for R=2. The method yields similar results for this image as well. The mean map can capture the underlying structures, and most variation is concentrated on the edges. Similar to the HCP results, std map can indicate potential discrepancies between mean and FS image as well, for instance low intensity region indicated with the arrow, that is incorrectly represented in the mean image, has a high std value.

Finally we analyze the k-space error in the sampled images from three different methods, l-MALA, cWGAN and the local sampling and show the k-space error histograms in Figure 5.7 for a test slice at R=5. We calculate these as follows: we take 50 image samples $\{x_s\}_{s=1}^{50}$ from each method and apply the undersampled Fourier transform to transform each of them to k-space and take the measured voxels. Then we calculate the voxelwise difference between these and the measured data for all measured k-space voxels for all the samples together. The histogram then shows the distribution of the

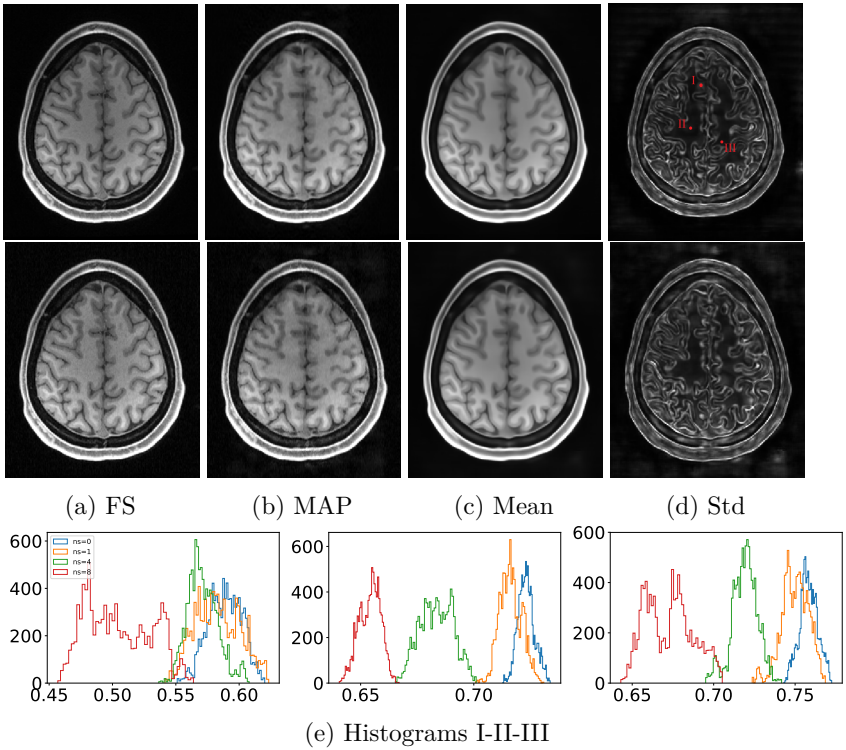
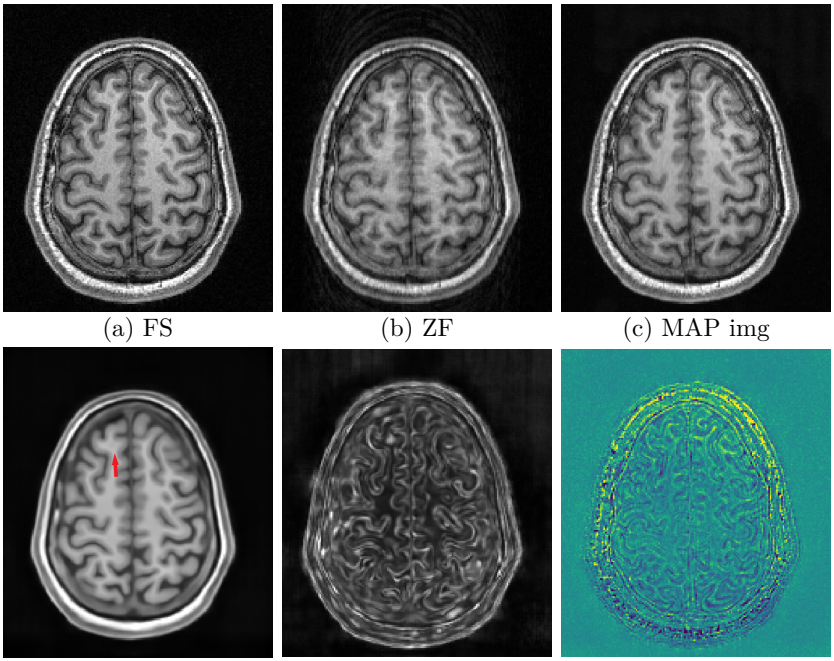


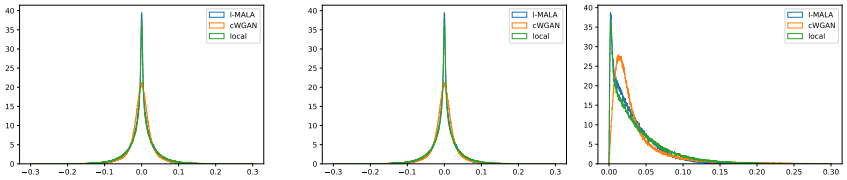
FIGURE 5.5: Results for changing the noise in k-space at R=5. First row shows the results with the basis HCP k-space noise. Second row shows the results with noise added on the k-space with 8 times the original noise standard deviation. Third row shows histograms of values of the pixels indicated on the std map (with added noise 1, 4 and 8 times of the basis noise). Note that the fully sampled (FS) image also changes due to the added noise.



(a) FS (b) ZF (c) MAP img

(d) Mean (e) Std (f) Error

FIGURE 5.6: Results for multicoil in-house measured images at R=2. Shown are the fulls sampled (FS), zero-filled (ZF), MAP estimate images, the mean and std maps for the latent-MALA samples as well as the difference image between the mean map and the FS (clipped to 0,0.3)).



(a) Real part (b) Imaginary part (c) Absolute error

FIGURE 5.7: Histograms of the voxelwise error in the measured voxels in the k-space for three different methods for a subject at R=5. As the error is complex, the real and imaginary parts as well as the magnitude of the error values are shown seperately.

error for all these k-space voxels from all 50 samples. As this difference is complex, we show two histograms seperately for the real and imaginary parts and also for the magnitude values. We can also look at the image-wise absolute error as

$$\text{abs. error}_s = \sum_{\text{all meas. voxels}} |Ex_{FS} - Ex_s|, \tag{5.23}$$

for a sample image x_s, the fully sampled image x_{FS} and the sum is over all measured k-space voxels. When calculated for all 50 samples, the mean (std) values for this slice are given as 484.23 (5.45), 584.71 (18.89) and 597.98 (15.69), for the l-MALA, cWGAN and local sampling methods, respectively.

To show how this generalizes, we do a similar analysis using slices from 9 test subjects. We undersample the slices with different undersampling patterns at R=5. Again, for each test subject we generate 50 samples for the three methods each. We then calculate the absolute errors and report the mean (std) values in Table 5.1. For all subjects (except subject 3) the l-MALA method yields significantly (p value lower than 0.001) lower absolute error for the 50 samples with the Wilcoxon signed-rank test. For subject 3, the cWGAN method yields significantly lower absolute error (p value lower than 0.001). Considering the mean absolute error for each subject, the l-MALA method yields significantly lower absolute error overall (p value lower than 0.011) We also calculate the root mean squared error (RMSE) between the 50 samples and the fully sampled image. Though achieving a

subject	Abs. error			RMSE		
	l-MALA	cWGAN	Local	l-MALA	cWGAN	Local
#1	639.55 (6.67)	679.18 (20.68)	808.54 (11.14)	9.29 (0.12)	11.29 (0.32)	11.67 (0.14)
#2	651.38 (7.02)	711.14 (22.23)	765.40 (7.42)	11.22 (0.10)	13.43 (0.44)	12.86 (0.12)
#3	518.36 (2.32)	497.98 (18.08)	570.67 (6.10)	13.18 (0.07)	15.51 (0.77)	14.51 (0.16)
#4	573.19 (3.24)	614.66 (18.35)	699.97 (7.15)	8.91 (0.08)	11.00 (0.57)	10.86 (0.13)
#5	482.08 (2.83)	584.71 (18.89)	597.98 (15.69)	8.50 (0.08)	11.38 (0.57)	10.34 (0.13)
#6	603.95 (5.45)	647.72 (16.12)	775.51 (11.20)	8.16 (0.05)	9.85 (0.37)	10.21 (0.13)
#7	595.17 (9.36)	694.03 (22.71)	754.85 (8.49)	9.08 (0.14)	12.48 (0.52)	11.01 (0.11)
#8	623.54 (7.65)	717.40 (19.04)	752.28 (7.06)	10.14 (0.19)	12.79 (0.41)	12.07 (0.09)
#9	528.14 (6.39)	556.63 (12.26)	646.52 (7.50)	8.11 (0.07)	10.11 (0.39)	9.70 (0.092)

TABLE 5.1: Absolute error and RMSE values for 9 test subjects at R=5.

low RMSE is not the purpose of any of the methods, we present these results as they still provide some insight into the performance of the methods. To calculate the RMSE we use the formula given in [29] and use a mask to disregard the background. The RMSE values are significantly lower for the 50 samples for each subject (p value lower than 0.001) and for the mean RMSE values for each subject (p value lower than 0.008) for the l-MALA method compared to the other methods.

5.5 DISCUSSION

The results show that the proposed method is able to capture the underlying ambiguity in undersampled MRI acquisitions in that it generates samples, such as those in Fig. 5.2, that show realistic structural diversity while retaining high fidelity to the fully sampled image. The variability also indicates potential discrepancies between the mean prediction and the FS image.

One observation is that the texture in the fully sampled image is not entirely captured in the samples. This is firstly because we do not add the noise on to the samples, of which the texture is partly composed of. Secondly, the lack of texture can be attributed to the VAE, which is known for preferring blurry images and ignore very high-frequency changes. We expect this aspect to improve with a better prior model.

More importantly, the variation in the samples summarized in the std maps are capable of highlighting the potential mistakes in their mean predictions as seen in Fig. 5.3 for the latent MALA, cWGAN and heteroscedastic feed-forward network approaches. This is very valuable information for

further decision making, as such regions where uncertainty is high should be approached with doubt. This information, when taken directly in the form of the samples or estimated standard deviations can be used for any decision making process for clinical or research purposes. The latent MALA and cWGAN are advantageous compared to heteroscedastic models in this respect as they can produce samples allowing quantification of uncertainty of any task performed on the images. The heteroscedastic modeling approach is limited in that it makes an independent Gaussian assumption for each pixel and cannot generate realistic samples.

The std maps from the local VAE sampling do not highlight the regions where its mean differs from the ground truth. This is expected, since it takes the MAP estimate as the underlying image and samples only around its latent representation, without exploring other areas of the latent space. Hence it can generate variations only on the tissue edges of the MAP image, but cannot explore possibility of different tissue structures.

We note that the cWGAN method was developed for CT images and our modified implementation for MR is very straightforward and limited. By feeding only the magnitude images, we set the phase to zero, which is the correct phase for the HCP dataset, giving the method an advantage. However this is not a generic situation and the cWGAN method needs to be extended from CT to work in a realistic MR setting.

As the undersampling ratio R increases the inversion problem becomes harder, hence the mean maps start diverging more from the ground truth, which is also reflected in the difference maps shown in Fig. 5.4. Proposed model successfully captures the higher ambiguity for higher undersampling ratios as can be seen in the higher values in the std maps. This increase is also visible in the increasing spread of the histograms. Results presented in Fig. 5.5, from the experiments with increasing k-space noise Σ_{ns}, show that the latent MALA model can incorporate the changing k-space noise. When the k-space noise increases ambiguity in the observations increase. Mathematically, the Σ_{ns} term in the Gaussian data likelihood increases, which in turn allows accepting samples farther away from the measured k-space data y. This results in higher sample diversity, which is reflected in the increase in the standard deviation maps, whereas the mean of the samples is not affected much by this.

Though we use here the VAE as the latent space model, the outlined method is generic. The integral in Eq. 5.2, which relates the latent space and the k-space is a generic formulation and can be used with other probabilistic models that provide a deterministic or probabilistic decoder structure. One

important necessary property, however, is that the decoder structure needs to be differentiable w.r.t. the latent variables for the Langevin walk. Utilizing another decoder structure here can also increase the quality of the final image samples, for instance models that suffer less from blurriness than VAEs.

Another factor to consider regarding the VAE is that the aggregate posterior of the VAE, given as $q(z) = \int q(z|x)p_{data}(x)dx$ does not necessarily overlap with the prior distribution $p(z)$. This can then cause the random walk to move towards regions of the latent space which are not in the aggregate posterior or similarly miss parts of the aggregate posterior which are zero in the prior. We corrected this discrepancy partially by introducing the empirical prior for $p(z)$ and have not observed problems regarding this issue.

Similarly, MALA is not the only way of doing the sampling. We choose it due to several factors, such as its efficiency in high dimensional spaces, theoretical guarantees on asymptotic convergence to the true posterior, not requiring normalized target distributions etc. Alternative pproaches such as Hamiltonian Monte Carlo [135] can be utilized so that the typical set can be traversed more quickly. Similarly, in cases of multi-modal distributions such as the one at hand, approaches tailored to such distributions, such as Stein variational gradient descent [142] can be considered. Furthermore natural gradient based methods, where the geometry of the target distribution is taken into account by introducing an associated Riemannian metric can be considered to speed up the MALA [135].

As discussed above, the proposed method has the advantage of having a modular structure as the prior is decoupled from the data acquisition model and the target posterior is decoupled from the sampling procedure. This, we believe, is quite advantageous in terms of future research and improvement opportunities. This is in contrast to the cWGAN approach, where the loss function and the architecture, by design, determine an implicit target distribution without explicitly modeling the prior or the acquisition. Lastly, the decoupling of the prior from the data acquisition model allows the latent MALA model to be used for different undersampling factors with the same prior without retraining and also incorporate additional details of the acquisition in a very straightforward way, such as the bias field without needing to retrain the prior.

Lastly, the proposed model is sampling from the true posterior for a given $p(x)$ and y. Predictive approaches for uncertainty quantification, such as cWGAN, relies on a trained network to generalize for a given sample

while having serious training data deficiency. They require training samples that show different fully sampled images for a given undersampled image, which is not readily available unless a method such as the one proposed here is used to construct them.

5.6 CONCLUSION

In this chapter we proposed and evaluated a method that can provide multiple possible images for the given undersampled k-space data. In contrast to reconstruction approaches, where a single image is output, the sampling approach can capture the uncertainty in the inversion process due to the missing data. The variation in the samples is indicative of the uncertainty, hence can open up new avenues for uncertainty quantification for following image analysis tasks and can point to potential mistakes in the mean prediction. Note that the existing methods for downstream tasks usually are not suitable for processing with multiple images, but designing such methods can be an interesting research direction in itself. The method we propose has a modular structure and can be improved by separately improving its components, such as the prior term or the sampling scheme.

6

DISCUSSION AND OUTLOOK

Here we discuss the methods introduced in the thesis, identify their advantages and disadvantages, shortcomings and possible improvements and further research directions.

The first and foremost of the discussion points is the choice of the prior used in the presented methods, namely the VAE. VAE is one of the three mainstream generative models in the literature for approximating unknown distributions with access only to samples from this distribution. The other two are GANs and flow-based models. Naturally, these methods have multiple extensions building on the vanilla approach, but regardless of extensions there are identifiable distinct characteristics to all three approaches. VAE and flow-based methods, for instance, are explicit models, in that they provide an unnormalized density function that one can evaluate for the samples to obtain probabilities for each sample. This is an exact probability in the case of flow models and the approximation ELBO for the VAE. In contrast GANs do not provide a density that can be evaluated, only the capability to sample from the distribution. This makes GANs not directly usable for our formulations where we explicitly need the density function $\log p(x)$. The other viable option in that regard is the flow based models. For these the gradients $\nabla_x p(x)$ are defined as well, so it is possible to incorporate the flow based models into the reconstruction scheme. Our preliminary investigations with the FFJORD model yielded also good results, which we do not report in this thesis. The drawback of the FFJORD model was that it took a long time to train, especially for full sized images and led to instabilities in the reconstruction algorithm and was not pursued further. Similarly, it is possible to incorporate the flow based methods into the sampling as well. This actually simplifies the sampling further as the marginalization necessary for the target distribution $p(y|z) = \int p(y|x)p(x|z)dx$ becomes simply an evaluation $p(y|z) = N(y|Ef(x), \Sigma_{ns})$ since $p(x|z) = \delta(f(z))$, where f denotes the flow function from the latent variable z to image x. This would also potentially speed up sampling a considerable amount as well, as it revokes the need for the iterative matrix inversion required by marginalization is a VAE is used. Furthermore, using flow based priors would also potentially benefit obtaining sharper images, as one of the main critiques to VAEs is that they produce blurry images.

Another very important topic is the domain gap problem, which results in a degraded performance if the statistics of the images in the training set and test images are different. In the case of MR images many factors induce a domain gap, such as device related factors (e.g. scanner vendor/model, receive coils, gradient coils), scan related factors (e.g. sequence parameters resulting in different tissue contrast or SNR) or factors that arise from the combination of these (e.g. image artifacts from eddy currents). We partially address this for the bias field in Chapter 4. Another relatively simple remedy could be to train the network with augmentations that can simulate these contrast changes, such as non-linearly transforming the histogram with piece-wise linear (and hence monotonic) transformations [143]. Beyond these specific cases, the domain gap is an inherent limitation of the learning based methods and there is considerable amount of work in the literature to increase the generalization capabilities of the learned networks [144]. On the other hand different learning based method suffer from this to varying degrees. We identify here a drawback of our unsupervised learning based approach for reconstruction in that regard compared to the unrolled denoising approaches. The denoising methods can be understood as learning to recognize the noise like artifacts in the images to remove these. However, for instance, the appearance of these artifacts are not very sensitive to the tissue contrast in the image, rendering the denoising method less sensitive to the contrast. On the other hand, by the same logic the denoising algorithms are sensitive to the SNR in the image as they expect similar statistics for the noise-like patterns to what they were trained on. These claims are also in accord with the experimental findings in [106]. Contrarily, in the unsupervised approach, the aim is to make the image similar to a learned image, not to remove the artifacts. That means, the algorithm tries to produce an image where the contrast matches the learned one and while doing this, it might diverge from the measured data if the contrast in the measured data does not match the learned contrast. In turn, this might lead to competing preferences for the prior and data likelihood and degrade performance. A similar logic applies to different sources of domain gap for both approaches as well. Note that this intuitive analysis is more difficult for the non-training and deep image prior based approaches, as these have less clear interpretations. One can argue that since these do not depend on any training set, per definition, they cannot suffer from the domain gap. On the other hand, there are implicit factors that affect the performance such as the choice of the architecture or the optimization scheme. If one takes, for instance, the spectral low pass filtering interpretation for the deep image

priors [80], then the structure of this filter is determined by the implicit factors mentioned above and a design that works well for one set of images may not be optimal for another.

The next point regarding the prior to be discussed is how it is utilized. For both reconstruction and sampling we actually not directly use the prior but rather its gradient. This relates to the common saying that one should actually approximate or learn the gradient rather than take the gradient of an approximation. Here, for instance, taking the gradient of the ELBO for reconstruction falls into the latter category. In actuality the ELBO is only a lower bound and its gradient does not necessarily overlap with the gradient of the true prior. One alternative is way to calculate the gradient is to make a local Gaussian assumption as in [145]. Here the authors assume a locally Gaussian distribution for the images $\log p(x) = \log N(x|\mu_{local}, \Sigma_{local})$ and take the gradient of this, where $\mu_{local} = \sum_l \mu_x(z^l)$ with $z^l \sim q(z|x)$ and Σ_{local} is either predicted by a network or calculated as $\sum_l (\mu_x(z^l) - \mu_{local})^2$. The idea can be seen as taking the derivative of the ELBO while assuming $q(z|x)/p(z) = 1$ and a homoscedastic decoder. Then the ELBO reduces to $\int q(z|x) \log p(x|z) dz$ with $p(x|z) = N(x|\mu_x(z), \Sigma_x(z))$ and to $(1/L) \sum \log p(x|z^l)$ with $z^l \sim q(z|x)$ after the Monte Carlo approximation with $l = 1...L$. Employing the homoscedastic decoder assumption we set $\Sigma_x(z) = \Sigma = \sigma I$. Inputting the log Gaussian density function into the expression yields $(1/L) \sum_l (x - \mu_x(z))^T \Sigma^{-1} (x - \mu_x(z))$. Here one can recognize that this expression can be written as a Gaussian with a single mean and covariance: $\Sigma_{new} = (1/L)[\sum_l \Sigma^{-1}]^{-1}$ and $\mu_{new} = (1/L)\Sigma_{new} \sum_l \Sigma^{-1} \mu_x(z^l)$. Then again using the isotropic $\Sigma = \sigma I$ as assumed for the decoder and cancelling with the 1/L terms yields $\Sigma_{new} = \Sigma = \sigma I$ and $\mu_{new} = \sum_l \mu_x(z^l)$. Then the authors use the two approach mentioned above for calculating the local covariance function. One natural alternative to the authors approach would be to instead use a heteroscedastic decoder and directly the learned $\Sigma_x(z)$. This would scale the means on an individual level instead of scaling after combining them and should yield a better approximation (as, for instance, μ/σ is a more "natural" parameterization of the mean for a single dimensional Gaussian in terms of distances in the space of Gaussian distributions, also realized by the Fisher information matrix [135]). However, the assumption $q(z|x)/p(z) = 1$, i.e. ignoring the KL term between the predicted $q(z|x)$ and its prior might be problematic for reconstruction due to reasons discussed below. In any case, as the authors show improvement for their task of anomaly detection, the approach would also be beneficial for the highly related reconstruction task.

Another aspect regarding the gradient of the ELBO for reconstruction is the following: the ELBO is trained for the fully sampled images, however itself and the gradients are evaluated for images, that are not fully sampled, especially in the beginning of the optimization when the image is closer to the zero-filled data. This obviously poses a problem for the highly non-convex optimization scheme that assumes the gradient directions are correct. Furthermore, the result becomes sensitive to initialization. On the other hand, despite these, we see that the reconstruction model indeed works, especially for small undersampling ratios. We explain this as follows: the VAE is trained with a regularization term that encourages the latent vectors corresponding to the training images to be distributed similar to the unit Gaussian prior. This, by extension, encourages an aggregate posterior similar to a unit Gaussian. Then, assuming a unit Gaussian aggregate distribution means that images not similar to training images will be mapped to outer regions of the unit Gaussian in the latent space. This, in turn, means the gradient from the $KL[q(z|x)||p(z)]$ term will be high and drive the latent variable towards the high probability regions of the latent space, bringing the the image closer to the training images, as desired. However the assumption that the aggregate distribution matching the prior does not hold perfectly. To see this we take the expectation of the KL term w.r.t. the distribution of training data: $\mathbb{E}_{p_{data}(x)}[KL[q(z|x)||p(z)]] = -\mathbb{E}_{p_{data}(x)}[H[q(z|x)]] + CE[q(z), p(z)]$, with H as entropy and CE as cross-entropy. Hence minimizing this term results in minimizing the cross-entropy between the aggregate distribution and the unit prior, however, this matches the mode but not the distributions (or for this specific unit prior: $CE[q(z), N(0,1)] = -C \int q(z) z^2 dz \propto -\mathbb{E}_{q(z)}[z]^2 - Var_{q(z)}[z]$, with some constant C, and as can be observed, minimizing this minimizes the mean and the variance of $q(z)$). This means that there are high probability regions in the latent space which do not map to the distribution of training images. This can then result in situations i) the zero filled image gets mapped to a high probability region in the latent space and the gradients cannot take them to a more meaningful region, or ii) the zero filled image gets mapped to an outer region in the latent space and the gradients take it closer to the high probability regions of the latent space, however these regions do not map to proper MR images. Considering these scenarios, having an aggregate distribution matching the prior would improve the performance of the optimization with initialization far away from the training images. In that regard another advantage of the flow based models would be to fulfill this condition. Another alternative to initializing the reconstruction

from the zero-filled image could be to start the optimization from a point, where the VAE is trained for, for instance the middle of the latent space while using the empirical prior described in Chapter 5. This way the VAE would operate in a domain where it is properly trained.

In our work we concentrate on reconstructing or sampling directly the images with a certain weighting and demonstrate the developed methods on T1 weighted images. An alternative to this would be to look into direct parameter mapping. For instance, for spoiled gradient echo sequences, such as the MPRAGE used in the HCP dataset, the signal model is quite simple [146], requiring the knowledge of only the T1 for the relaxation and the proton density for the net magnetization along with the acquisition parameters to obtain the signal in each pixel. Then, instead of learning the priors on the T1 weighted images, one can learn the tissue parameters and use this model to directly reconstruct the tissue parameters. The authors of [147], for instance, use a signal model for parameter map reconstruction for a fat/water separation sequence. This would be the ultimate solution to the contrast domain gap problem, since the tissue parameters are absolute measurements and any image characteristic related to acquisition can be explicitly modeled in the signal model. The problem, however, is that a single image would not be enough to reconstruct multiple parameters despite having the prior for these. A remedy for this is moving into the MR fingerprinting framework. Here already multiple images are acquired with the aim of obtaining tissue parameters and deep learning based methods have already been applied to accelerating the dictionary matching as discussed in Chapter 2. Similarly, datasets acquired with MR fingerprinting in the future could be used as training data with multiple tissue parameters to allow for reconstruction approaches directly estimating these.

Now we turn our attention to the posterior used in the latent MALA and search for different ways of obtaining samples from this distribution. The first observation is that, despite both $p(y|z)$ and $p(z)$ being Gaussians, the posterior $p(z|y)$ itself is not since the parameters of $p(y|z)$ are non-linear functions of z. This limits the options of finding an approximate distributions. A Laplace approximation misses the possibly multi modal structure or a mean field approximation assuming the pixel intensities factorize is again too simplistic as it ignores the covariances. One other option also mentioned in Chapter 5 is the Stein variational inference [142]. Here the idea is to initialize with multiple "particles" (in our case latent vectors) and to change the images in the direction of the target distribution (in our case the target posterior). At the same time a kernel based mechanism discourages

all the particles to become close to each other, effectively forcing to them to spread around in the domain of the target distribution and hence capture multiple modes. Using such a scheme with a low number of particles can provide a good approximation to the estimated statistics faster than MCMC.

A natural question to ask is what can be done to improve the reconstruction quality except for changing the reconstruction algorithm. As the undersampling (US) pattern has a big impact on the outcome, one approach here is to optimize the US pattern for a given algorithm and/or for the given image. There is some works in the literature that aim to do this for conventional or deep learning based reconstruction algorithms [148–150], optimizing the sampling pattern to reduce the root mean square error or peak SNR between the original image and the reconstructed image. Using such a cost, however, generally tends to yield optimal US patterns that sample more the low frequency components and mostly ignore the high frequency regions. This is an expected outcome as most of the information is concentrated in the central k-space and acquiring more of these results in lower RMSE error. On the other hand, this approach has two drawbacks. Firstly, these methods are generally not tested with abnormalities in the images, though they should be, as the general RMSE may be low even though the abnormalities may not be reconstructed well, especially if they consist of more high frequency components. Secondly, for image features that have high frequency content, the measured data may not be enough to determine these features which would lead to a high uncertainty for these regions. Considering these, it would make sense to look into minimizing cost functions that capture this uncertainty, such as the pixel-wise variance from the sampling method as obtained in Chapter 5. Different strategies for a quick solution to the optimization problem can be designed using approaches from optimal experimental design [151] such as Bayesian optimization [152] or active learning. Similarly, one can simultaneously optimize the reconstruction algorithm as well as the sampling pattern as in the works [153–155], which ideally should yield even better reconstruction quality. However, the US patterns from these methods also focus on the low resolution regions ignoring the high frequency information.

This brings us to our next remark, which is regarding how the developed reconstruction methods are tested. The usual way of testing the quality of reconstructions is to look at some metrics such as RMSE or peak SNR which depend on the l_2 norm between the original and reconstructed image. These usually yield high scores for images that have good structural correspondence, however mostly compare the low resolution features and

ignore the high frequency structures, such as texture. Another popular metric is the structural similarity measure [156] (SSIM), which in turn looks at different features such as texture as well. Measures such as contrast or contrast to noise ratio calculated as tissue edges yield information about the blurriness of the reconstructed images. Furthermore these evaluations are usually done on a set of images from healthy subjects and their performance is not necessarily evaluated for the downstream task, although the point of imaging is to assist these tasks. In a diagnosis scenario, for instance, in case of high frequency abnormalities such as multiple sclerosis lesions, algorithms optimized for an l_2 norm based evaluation metric may reduce the visibility of the lesions [157]. Similarly a tool utilized in a downstream task of segmentation for diagnostic or research purposes may suffer from changed high/low level image statistics. Again, as the aforementioned evaluation metrics (with the exception of contrast based metrics at the tissue edges) do not evaluate the quality of the tissue boundaries, the images from the resulting algorithms will be susceptible to be incorrectly segmented. This effect becomes more amplified if some deep learning based segmentation algorithm is used as these are quite sensitive to low/high level image statistics. Hence either the evaluations have to include such downstream tasks to measure actually the applicability of the reconstructions or evaluation metrics have to be used that are not insensitive to certain features in the images. One alternative such evaluation metric can be the perceptual loss, where the l_2 norm is not only calculated at the image level but also at feature level after being fed into a pre-trained network (such as the VGG [158]). Using such a loss can capture differences at both low levels from the features extracted in the beginning in the network and also high level information from the deeper parts of the network. In the literature the perceptual loss was already used for training networks [159, 160]. In [159] it was reported that training a simple reconstruction network using the perceptual loss yields better image quality as assessed by a radiologist than its counterpart trained on the l_1 or l_2 loss.

Finally the analysis in Chapter 3 shows that the reconstruction quality is adversely affected by the inaccurate coil map estimations from ESPIRiT. The methods that directly apply CNNs on the k-space, summarized in Chapter 3 show promising results which may revoke the need for the ESPIRiT estimation, improve the quality of the coil combined image as well as improve the match between the measured k-space and the k-space after the encoding operation, which, in turn, would improve the performance of both the reconstruction and the sampling.

BIBLIOGRAPHY

1. Shenton, M. E., Hamoda, H. M., Schneiderman, J. S., Bouix, S., Pasternak, O., Rathi, Y., Vu, M.-A., Purohit, M. P., Helmer, K., Koerte, I., Lin, A. P., Westin, C.-F., Kikinis, R., Kubicki, M., Stern, R. A. & Zafonte, R. A review of magnetic resonance imaging and diffusion tensor imaging findings in mild traumatic brain injury. *Brain Imaging and Behavior* **6**, 137 (2012).
2. Symms, M, Jäger, H. R., Schmierer, K & Yousry, T. A. A review of structural magnetic resonance neuroimaging. *Journal of Neurology, Neurosurgery & Psychiatry* **75**, 1235 (2004).
3. Clayden, J. D. Imaging connectivity: MRI and the structural networks of the brain. *Functional neurology* **28**, 197 (2013).
4. Saeed, M., Van, T. A., Krug, R., Hetts, S. W. & Wilson, M. W. Cardiac MR imaging: current status and future direction. *Cardiovascular diagnosis and therapy* **5**, 290 (2015).
5. Wang, Y., Wluka, A. E., Jones, G., Ding, C. & Cicuttini, F. M. Use magnetic resonance imaging to assess articular cartilage. *Therapeutic advances in musculoskeletal disease* **4**, 77 (2012).
6. Kumar, S., Liney, G., Rai, R., Holloway, L., Moses, D. & Vinod, S. K. Magnetic resonance imaging in lung: a review of its potential for radiotherapy. *The British journal of radiology* **89**, 20150431 (2016).
7. Noone, T. C., Semelka, R. C., Chaney, D. M. & Reinhold, C. Abdominal imaging studies: comparison of diagnostic accuracies resulting from ultrasound, computed tomography, and magnetic resonance imaging in the same individual. *Magnetic Resonance Imaging* **22**, 19 (2004).
8. Bogner, W., Otazo, R. & Henning, A. Accelerated MR spectroscopic imaging—a review of current and emerging techniques. *NMR in Biomedicine* **n/a**. e4314 NBM-19-0277.R2, e4314.
9. GI, O., AO, A., A, B.-P., Y, M. & DA., N. Survey of magnetic resonance imaging availability in West Africa. *Pan Afr Med J.* (2018).
10. *MRI Systems Market Overview* `https://www.psmarketresearch.com/market-analysis/mri-market`. Accessed: 2020-10-25.

11. *MRI Equipment Market* https://www.fortunebusinessinsights.com/industry-reports/magnetic-resonance-imaging-mri-equipment-market-100087. Accessed: 2020-10-25.
12. Lee, B. & Newberg, A. Neuroimaging in traumatic brain imaging. *NeuroRx : the journal of the American Society for Experimental NeuroTherapeutics* **2**, 372 (2005).
13. Brown, R. W., Cheng, Y. N., Haacke, E. M., Thompson, M. R. & Venkatesan, R. *Magnetic Resonance Imaging* (John Wiley and Sons, Ltd, 2014).
14. BERNSTEIN, M. A., KING, K. F. & ZHOU, X. J. *"Handbook of MRI Pulse Sequence* (Academic Press, Burlington, 2004).
15. BERNSTEIN, M. A., KING, K. F. & ZHOU, X. J. in *Handbook of MRI Pulse Sequences* (eds BERNSTEIN, M. A., KING, K. F. & ZHOU, X. J.) 67 (Academic Press, Burlington, 2004).
16. BERNSTEIN, M. A., KING, K. F. & ZHOU, X. J. in *Handbook of MRI Pulse Sequences* (eds BERNSTEIN, M. A., KING, K. F. & ZHOU, X. J.) 5 (Academic Press, Burlington, 2004).
17. Tarantola, A. *Inverse Problem Theory and Methods for Model Parameter Estimation* (Society for Industrial and Applied Mathematics, 2005).
18. Pruessmann, K. P., Weiger, M., Scheidegger, M. B. & Boesiger, P. SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine* **42**, 952 (1999).
19. Madore, B., Glover, G. H. & Pelc, N. J. Unaliasing by Fourier-encoding the overlaps using the temporal dimension (UNFOLD), applied to cardiac imaging and fMRI. *Magnetic Resonance in Medicine* **42**, 813 (1999).
20. Mistretta, C. A., Wieben, O., Velikina, J., Block, W., Perry, J., Wu, Y., Johnson, K. & Wu, Y. Highly constrained backprojection for time-resolved MRI. *Magnetic Resonance in Medicine* **55**, 30 (2006).
21. Block, K. T., Uecker, M. & Frahm, J. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. *Magnetic Resonance in Medicine* **57**, 1086 (2007).
22. Lustig, M., Donoho, D. & Pauly, J. M. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine* **58**, 1182 (2007).

23. Rudin, L. I., Osher, S. & Fatemi, E. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena* **60**, 259 (1992).

24. Kelley, C. T. *Iterative Methods for Optimization* (Society for Industrial and Applied Mathematics, 1999).

25. Lingala, S. G., Hu, Y., Dibella, E. & Jacob, M. Accelerated dynamic MRI exploiting sparsity and low-rank structure: K-t SLR. *IEEE Transactions on Medical Imaging* **30**, 1042 (2011).

26. Boyd, S., Parikh, N., Chu, E., Peleato, B. & Eckstein, J. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends in Machine Learning* **3**, 1 (2011).

27. Goldstein, T. & Osher, S. The Split Bregman Method for L1-Regularized Problems. *SIAM J. Imaging Sci.* **2**, 323 (2009).

28. Tsao, J., Boesiger, P. & Pruessmann, K. P. k-t BLAST and k-t SENSE: Dynamic MRI with high frame rate exploiting spatiotemporal correlations. *Magnetic Resonance in Medicine* **50**, 1031 (2003).

29. Tezcan, K. C., Baumgartner, C. F., Luechinger, R., Pruessmann, K. P. & Konukoglu, E. MR Image Reconstruction Using Deep Density Priors. IEEE Transactions on Medical Imaging. *IEEE Transactions on Medical Imaging* **38** (2019).

30. Feinberg, D. A., Hale, J. D., Watts, J. C., Kaufman, L & Mark, A. Halving MR imaging time by conjugation: demonstration at 3.5 kG. *Radiology* **161**. PMID: 3763926, 527 (1986).

31. Wiesinger, F., Boesiger, P. & Pruessmann, K. P. Electrodynamics and ultimate SNR in parallel MR imaging. *Magnetic Resonance in Medicine* **52**, 376 (2004).

32. Wiesinger, F, Zanche, N. D. & Pruessmann, K. *Approaching Ultimate SNR with Finite Coil Arrays* in *Proceedings of the 13th Annual Meeting of ISMRM* (2005).

33. Sodickson, D. K., Ohliger, M. A., Lattanzi, R. & Wiggins, G. C. in *eMagRes* (American Cancer Society).

34. Guerquin-kern, M., Häberlin, M., Pruessmann, K. P. & Unser, M. A fast wavelet-based reconstruction method for magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 1649 (2011).

35. Donoho, D. L. Compressed Sensing. *IEEE Transactions in Information Theory* **52**, 1289 (2006).

36. Candes, E. J., Romberg, J. K. & Tao, T. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics* **59**, 1207 (2006).

37. Ning, B., Qu, X., Guo, D., Hu, C. & Chen, Z. Magnetic resonance image reconstruction using trained geometric directions in 2D redundant wavelets domain and non-convex optimization. *Magnetic Resonance Imaging* **31**, 1611 (2013).

38. Qu, X., Guo, D., Ning, B., Hou, Y., Lin, Y., Cai, S. & Chen, Z. Undersampled MRI reconstruction with patch-based directional wavelets. *Magnetic Resonance Imaging* **30**, 964 (2012).

39. Ravishankar, S. & Bresler, Y. MR Image Reconstruction From Highly Undersampled k-Space Data by Dictionary Learning. *IEEE transactions on medical imaging* **30**, 1028 (2010).

40. Zhan, Z., Cai, J. F., Guo, D., Liu, Y., Chen, Z. & Qu, X. Fast Multiclass Dictionaries Learning With Geometrical Directions in MRI Reconstruction. *IEEE Transactions on Biomedical Engineering* **63**, 1850 (2016).

41. Eksioglu, E. M. Decoupled Algorithm for MRI Reconstruction Using Nonlocal Block Matching Model: BM3D-MRI. *Journal of Mathematical Imaging and Vision* **56**, 430 (2016).

42. Dabov, K., Foi, A., Katkovnik, V. & Egiazarian, K. Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering. *Trans. Img. Proc.* **16**, 2080 (2007).

43. Wissmann, L., Gotschy, A., Santelli, C., Tezcan, K. C., Hamada, S., Manka, R. & Kozerke, S. Analysis of spatiotemporal fidelity in quantitative 3D first-pass perfusion cardiovascular magnetic resonance. *Journal of Cardiovascular Magnetic Resonance* **19**, 11 (2017).

44. Pruessmann, K. P., Weiger, M., Börnert, P. & Boesiger, P. Advances in sensitivity encoding with arbitrary k-space trajectories. *Magnetic Resonance in Medicine* **46**, 638 (2001).

45. Pedersen, H., Kozerke, S., Ringgaard, S., Nehrke, K. & Won, Y. K. K-t PCA: Temporally constrained k-t BLAST reconstruction using principal component analysis. *Magnetic Resonance in Medicine* **62**, 706 (2009).

46. Schmidt, J. F. M., Wissmann, L., Manka, R. & Kozerke, S. Iterative k-t principal component analysis with nonrigid motion correction for dynamic three-dimensional cardiac perfusion imaging. *Magnetic Resonance in Medicine* **72**, 68 (2014).

47. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436 (2015).

48. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A., van Ginneken, B. & Sánchez, C. I. A survey on deep learning in medical image analysis. *Medical Image Analysis* **42**, 60 (2017).

49. Zhu, B., Liu, J. Z., Cauley, S. F., Rosen, B. R. & Rosen, M. S. Image reconstruction by domain-transform manifold learning. *Nature* **555** (2018).

50. Kwon, K., Kim, D. & Park, H. A parallel MR imaging method using multilayer perceptron. *Medical Physics* **44**, 6209 (2017).

51. Lee, D., Yoo, J. & Ye, J. C. *Deep residual learning for compressed sensing MRI* in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (2017), 15.

52. Han, Y. S., Yoo, J. & Ye, J. C. Deep Learning with Domain Adaptation for Accelerated Projection-Reconstruction MR. *ArXiv e-prints* (2017).

53. Schlemper, J., Caballero, J., Hajnal, J. V., Price, A. & Rueckert, D. in *Information Processing in Medical Imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings* (eds Niethammer, M., Styner, M., Aylward, S., Zhu, H., Oguz, I., Yap, P.-T. & Shen, D.) 647 (Springer International Publishing, Cham, 2017).

54. Wang, S., Su, Z., Ying, L., Peng, X., Zhu, S., Liang, F., Feng, D. & Liang, D. *Accelerating magnetic resonance imaging via deep learning* in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)* (2016), 514.

55. Yang, Y., Sun, J., Li, H. & Xu, Z. in *Advances in Neural Information Processing Systems 29* (eds Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I. & Garnett, R.) 10 (Curran Associates, Inc., 2016).

56. Hammernik, K., Klatzer, T., Kobler, E., Recht, M. P., Sodickson, D. K., Pock, T. & Knoll, F. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, n/a (2017).

57. Chen, Y., Yu, W. & Pock, T. On learning optimized reaction diffusion processes for effective image restoration. *CoRR* **abs/1503.05768** (2015).

58. Jin, K. H., McCann, M. T., Froustey, E. & Unser, M. Deep Convolutional Neural Network for Inverse Problems in Imaging. *IEEE Transactions on Image Processing* **26**, 4509 (2017).

59. Aggarwal, H. K., Mani, M. P. & Jacob, M. MoDL: Model Based Deep Learning Architecture for Inverse Problems. *CoRR* **abs/1712.02862** (2017).

60. Zhang, J. & Ghanem, B. *ISTA-Net: Interpretable Optimization-Inspired Deep Network for Image Compressive Sensing* in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), 1828.

61. Putzky, P. & Welling, M. Recurrent Inference Machines for Solving Inverse Problems. *CoRR* **abs/1706.04008** (2017).

62. Yang, G., Yu, S., Dong, H., Slabaugh, G., Dragotti, P. L., Ye, X., Liu, F., Arridge, S., Keegan, J., Guo, Y. & Firmin, D. DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction. *IEEE Transactions on Medical Imaging* **37**, 1310 (2018).

63. Mardani, M., Gong, E., Cheng, J. Y., Vasanawala, S. S., Zaharchuk, G., Xing, L. & Pauly, J. M. Deep Generative Adversarial Neural Networks for Compressive Sensing MRI. *IEEE Transactions on Medical Imaging* **38**, 167 (2019).

64. Quan, T. M., Nguyen-Duc, T. & Jeong, W. K. Compressed Sensing MRI Reconstruction Using a Generative Adversarial Network With a Cyclic Loss. *IEEE Transactions on Medical Imaging* **37**, 1488 (2018).

65. Narnhofer, D., Hammernik, K., Knoll, F. & Pock, T. *Inverse GANs for accelerated MRI reconstruction* in *Wavelets and Sparsity XVIII* (eds Ville, D. V. D., Papadakis, M. & Lu, Y. M.) **11138** (SPIE, 2019), 381.

66. Cole, E. K., Pauly, J. M., Vasanawala, S. S. & Ong, F. *Unsupervised MRI Reconstruction with Generative Adversarial Networks* 2020.

67. Arjovsky, M., Chintala, S. & Bottou, L. *Wasserstein Generative Adversarial Networks* in *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017* (2017), 214.

68. Yaman, B., Hosseini, S. A. H., Moeller, S., Ellermann, J., Urbil, K. & Akçakaya, M. *Self-Supervised Physics-Based Deep Learning MRI Reconstruction Without Fully-Sampled Data* in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* (2020), 921.

69. Liu, J., Sun, Y., Eldeniz, C., Gan, W., An, H. & Kamilov, U. S. RARE: Image Reconstruction Using Deep Priors Learned Without Groundtruth. *IEEE Journal of Selected Topics in Signal Processing* **14**, 1088 (2020).

70. Tamir, J. I., Yu, S. X. & Lustig, M. *Unsupervised Deep Basis Pursuit: Learning inverse problems without ground-truth data* 2020.

71. Shin, P. J., Larson, P. E. Z., Ohliger, M. A., Elad, M., Pauly, J. M., Vigneron, D. B. & Lustig, M. Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion. *Magnetic Resonance in Medicine* **72**, 959 (2014).

72. Haldar, J. P. Low-Rank Modeling of Local *k*-Space Neighborhoods (LORAKS) for Constrained MRI. *IEEE Transactions on Medical Imaging* **33**, 668 (2014).

73. Jin, K. H., Lee, D. & Ye, J. C. A General Framework for Compressed Sensing and Parallel MRI Using Annihilating Filter Based Low-Rank Hankel Matrix. *IEEE Transactions on Computational Imaging* **2**, 480 (2016).

74. Ye, J. C., Han, Y. & Cha, E. Deep Convolutional Framelets: A General Deep Learning Framework for Inverse Problems. *SIAM Journal on Imaging Sciences* **11**, 991 (2018).

75. Akçakaya, M., Moeller, S., Weingärtner, S. & Urbil, K. Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging. *Magnetic Resonance in Medicine* **81**, 439 (2019).

76. Han, Y., Sunwoo, L. & Ye, J. C. *k* -Space Deep Learning for Accelerated MRI. *IEEE Transactions on Medical Imaging* **39**, 377 (2020).

77. Lempitsky, V., Vedaldi, A. & Ulyanov, D. *Deep Image Prior* in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), 9446.

78. Jin, K. H., Gupta, H., Yerly, J., Stuber, M. & Unser, M. *Time-Dependent Deep Image Prior for Dynamic MRI* 2019.

79. Veen, D. V., Jalal, A., Soltanolkotabi, M., Price, E., Vishwanath, S. & Dimakis, A. G. *Compressed Sensing with Deep Image Prior and Learned Regularization* 2020.

80. Chakrabarty, P. & Maji, S. *The Spectral Bias of the Deep Image Prior* 2019.

81. Zhang, J., Liu, Z., Zhang, S., Zhang, H., Spincemaille, P., Nguyen, T. D., Sabuncu, M. R. & Wang, Y. Fidelity imposed network edit (FINE) for solving ill-posed image reconstruction. *NeuroImage* **211**, 116579 (2020).

82. Panda, A., Mehta, B. B., Coppo, S., Jiang, Y., Ma, D., Seiberlich, N., Griswold, M. A. & Gulani, V. Magnetic resonance fingerprinting – An overview. *Current Opinion in Biomedical Engineering* **3**. New Developments in Biomedical Imaging, 56 (2017).

83. Bipin Mehta, B., Coppo, S., Frances McGivney, D., Ian Hamilton, J., Chen, Y., Jiang, Y., Ma, D., Seiberlich, N., Gulani, V. & Alan Griswold, M. Magnetic resonance fingerprinting: a technical review. *Magnetic Resonance in Medicine* **81**, 25 (2019).

84. Cohen, O., Zhu, B. & Rosen, M. S. MR fingerprinting Deep RecOnstruction NEtwork (DRONE). *Magnetic Resonance in Medicine* **80**, 885 (2018).

85. Hamilton, J. & Seiberlich, N. Machine Learning for Rapid Magnetic Resonance Fingerprinting Tissue Property Quantification. *Proceedings of the IEEE* **PP**, 1 (2019).

86. Girardeau, S., Oksuz, I., Cruz, G., Vasquez, C. P., King, A. & Clough, J. *Deep Learning for Magnetic Resonance Fingerprinting* in *International Conference on Medical Imaging with Deep Learning – Extended Abstract Track* (London, United Kingdom, 2019).

87. Song, P., Eldar, Y. C., Mazor, G. & Rodrigues, M. R. D. HYDRA: Hybrid deep magnetic resonance fingerprinting. *Medical Physics* **46**, 4951 (2019).

88. Kingma D. P. Welling, M. Auto-Encoding Variational Bayes. *arXiv:1312.6114v10* (2014).

89. Rezende, D. J., Mohamed, S. & Wierstra, D. *Stochastic Backpropagation and Approximate Inference in Deep Generative Models* in *Proceedings of the 31st International Conference on Machine Learning* (eds Xing, E. P. & Jebara, T.) **32** (PMLR, Bejing, China, 2014), 1278.

90. Bishop, C. M. *Pattern Recognition and Machine Learning* (Springer, 2006).

91. Yeung, S., Kannan, A., Dauphin, Y. & Fei-Fei, L. Tackling Overpruning in Variational Autoencoders. *CoRR* **abs/1706.03643** (2017).

92. Samsonov, A. A., Kholmovski, E. G., Parker, D. L. & Johnson, C. R. POCSENSE: POCS-based reconstruction for sensitivity encoded magnetic resonance imaging. *Magnetic Resonance in Medicine* **52**, 1397 (2004).

93. De Pierro, A. R. & Helou Neto, E. S. From convex feasibility to convex constrained optimization using block action projection methods and underrelaxation. *International Transactions in Operational Research* **16**, 495 (2009).

94. Zhao, F., Noll, D. C., Nielsen, J. & Fessler, J. A. Separate Magnitude and Phase Regularization via Compressed Sensing. *IEEE Transactions on Medical Imaging* **31**, 1713 (2012).

95. Santelli, C., Loecher, M., Busch, J., Wieben, O., Schaeffter, T. & Kozerke, S. Accelerating 4D flow MRI by exploiting vector field divergence regularization. *Magnetic Resonance in Medicine* **75**, 115 (2016).

96. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *CoRR* **abs/1412.6980** (2014).

97. Essen, D. V., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S., Penna, S. D., Feinberg, D., Glasser, M., Harel, N., Heath, A., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S., Prior, F., Schlaggar, B., Smith, S., Snyder, A., Xu, J. & Yacoub, E. The Human Connectome Project: A data acquisition perspective. *NeuroImage* **62**. Connectivity, 2222 (2012).

98. Reuter, M., Schmansky, N. J., Rosas, H. D. & Fischl, B. Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage* **61**, 1402 (2012).

99. Sled, J., Zijdenbos, A. & Evans, A. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions in Medical Imaging* **17**, 87 (2002).

100. Uecker, M., Lai, P., Murphy, M. J., Virtue, P., Elad, M., Paul, y. J. M., Vasanawala, S. S. & Lustig, M. ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA. *Magnetic Resonance in Medicine* **71**, 990.

101. Fessler, J. A. & Sutton, B. P. Nonuniform fast Fourier transforms using min-max interpolation. *IEEE Transactions on Signal Processing* **51**, 560 (2003).

102. Lin, J.-M. & Chung, H.-W. *Pynufft: python non-uniform fast Fourier transform for MRI* in *Building Bridges in Medical Sciences 2017, St John's College, CB2 1TP Cambridge, UK* (2017).

103. M., B., D.J., L. & J.V., H. Combination of signals from array coils using image-based estimation of coil sensitivity profiles. *Magnetic Resonance in Medicine* **47**, 539.

104. Uecker, M., Tamir, J. I., Ong, F., Iyer, S., Cheng, J. Y. & Lustig, M. *BART: version 0.4.00* 2016.

105. Tanno, R., Worrall, D. E., Gosh, A., Kaden, E., Sotiropoulos, S. N., Criminisi, A. & Alexander, D. *Bayesian Image Quality Transfer with CNNs: Exploring Uncertainty in dMRI Super-Resolution* in *Intl Conf. Medical Image Computing and Computer Assisted Intervention (MICCAI)* (Springer, 2017).

106. Knoll, F., Hammernik, K., Kobler, E., Pock, T., Recht, M. P. & Sodickson, D. K. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. *Magnetic Resonance in Medicine* (2018).

107. Dilokthanakul, N., Mediano, P. A. M., Garnelo, M., Lee, M. C. H., Salimbeni, H., Arulkumaran, K. & Shanahan, M. Deep Unsupervised Clustering with Gaussian Mixture Variational Autoencoders. *CoRR* **abs/1611.02648** (2016).

108. Johnson, M., Duvenaud, D. K., Wiltschko, A., Adams, R. P. & Datta, S. R. in *Advances in Neural Information Processing Systems 29* (eds Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I. & Garnett, R.) 2946 (Curran Associates, Inc., 2016).

109. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. in *Advances in Neural Information Processing Systems 27* (eds Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D. & Weinberger, K. Q.) 2672 (Curran Associates, Inc., 2014).

110. Chang, J. H. R., Li, C. L., Póczos, B. & Kumar, B. V. K. V. *One Network to Solve Them All; Solving Linear Inverse Problems Using Deep Projection Models* in *2017 IEEE International Conference on Computer Vision (ICCV)* (2017), 5889.

111. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. & Shi, W. *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network* in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), 105.

112. Shah, V. & Hegde, C. Solving Linear Inverse Problems Using GAN Priors: An Algorithm with Provable Guarantees. *CoRR* **abs/1802.08406** (2018).

113. Yeh, R. A., Chen, C., Yian Lim, T., Schwing, A. G., Hasegawa-Johnson, M. & Do, M. N. Semantic Image Inpainting with Deep Generative Models. *ArXiv e-prints* (2016).

114. Mardani, M., Monajemi, H., Papyan, V., Vasanawala, S., Donoho, D. L. & Pauly, J. M. Recurrent Generative Adversarial Networks for Proximal Learning and Automated Compressive Image Recovery. *CoRR* **abs/1711.10046** (2017).

115. Vincent, P., Larochelle, H., Bengio, Y. & Manzagol, P.-A. *Extracting and Composing Robust Features with Denoising Autoencoders* in (2008), 1096.

116. Gaillochet, M., Tezcan, K. & Konukoglu, E. Joint reconstruction and bias field correction for undersampled MR imaging. *MICCAI* (2020).

117. Volpi, R., Namkoong, H., Sener, O., Duchi, J., Murino, V. & Savarese, S. Generalizing to Unseen Domains via Adversarial Data Augmentation. *NeurIPS* (2018).

118. K., C., N., K., C.F., B., Becker A., D. O. & E., K. Semi-supervised and Task-Driven Data Augmentation. *IPMI* (2019).

119. Sun B., S. K. Deep CORAL: Correlation Alignment for Deep Domain Adaptation. *ECCV* (2016).

120. Sled J.G., P. G. Understanding intensity non-uniformity in MRI. *MICCAI* (1998).

121. Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A. & C., G. J. N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging* **29** (2010).

122. SimpleITK: N4 Bias Field Correction. `https://simpleitk.readthedocs.io/en/master/link_N4BiasFieldCorrection_docs.html` (Last accessed 12 March 2020).

123. Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E. & Ugurbil, K. The WU-Minn Human Connectome Project: An Overview. *Neuroimage* **80** (2013).

124. Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E. & Ugurbil, K. Human Connectome Reference Manual. `https://www.humanconnectome.org/storage/app/media/documentation/s1200/HCP_S1200_Release_Reference_Manual.pdf` (Last accessed 4 March 2020).

125. Collingridge, D. A Primer on Quantitized Data Analysis and Permutation Testing. *Journal of Mixed Methods Research* **7** (2013).

126. Tezcan, K. C., Baumgartner, C. F. & Konukoglu, E. Sampling possible reconstructions of undersampled acquisitions in MR imaging. *arXiv:2010.00042* (2020).

127. Bilgic, B., Chatnuntawech, I., Fan, A. P., Setsompop, K., Cauley, S. F., Wald, L. L. & Adalsteinsson, E. Fast image reconstruction with L2-regularization. *Journal of Magnetic Resonance Imaging* **40**, 181 (2014).

128. Kendall, A. & Gal, Y. in *Advances in Neural Information Processing Systems 30* 5574 (2017).

129. Gal, Y. & Ghahramani, Z. Bayesian Convolutional Neural Networks with Bernoulli Approximate Variational Inference. *arXiv:1506.02158* (2015).

130. Tanno, R., Worrall, D. E., Enrico Kaden, A. G., Grussu, F., Bizzi, A., Sotiropoulos, S. N., Criminisi, A. & Alexander, D. C. Uncertainty Quantification in Deep Learning for Safer Neuroimage Enhancement. *arXiv:1907.13418v1* (2019).

131. Zhang, J., Zhang, H., Sabuncu, M., Spincemaille, P., Nguyen, T. & Wang, Y. *Probabilistic Dipole Inversion for Adaptive Quantitative Susceptibility Mapping* 2020.

132. Schlemper, J., Castro, D. C., Bai, W., Qin, C., Oktay, O., Duan, J., Price, A. N., Hajnal, J. & Rueckert, D. *Bayesian Deep Learning for Accelerated MR Image Reconstruction* in *Machine Learning for Medical Image Reconstruction* (Springer International Publishing, 2018).

133. Adler, J. & Öktem, O. Deep Bayesian Inversion. *arXiv:1811.05910* (2018).

134. Pedemonte, S., Catana, C. & Van Leemput, K. *Bayesian Tomographic Reconstruction Using Riemannian MCMC* in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (Springer International Publishing, 2015), 619.

135. Girolami, M. & Calderhead, B. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73** (2011).

136. Narnhofer, D., Hammernik, K., Knoll, F. & Pock, T. *Inverse GANs for accelerated MRI reconstruction* in *Wavelets and Sparsity XVIII* (eds Ville, D. V. D., Papadakis, M. & Lu, Y. M.) **11138** (SPIE, 2019), 381.

137. Schlemper, J., Caballero, J., Hajnal, J. V., Price, A. N. & Rueckert, D. A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction. *IEEE Transactions on Medical Imaging* **37**, 491 (2018).

138. Volokitin, A., Erdil, E., Karani, N., Tezcan, K. C., Chen, X., Gool, L. V. & Konukoglu, E. Modelling the Distribution of 3D Brain MRI using a 2D Slice VAE. *MICCAI* (2020).

139. Tóthová, K., Parisot, S., Lee, M. C. H., Puyol-Antón, E., Koch, L. M., King, A. P., Konukoglu, E. & Pollefeys, M. *Uncertainty Quantification in CNN-Based Surface Prediction Using Shape Priors* in *Shape in Medical Imaging* (Springer International Publishing, 2018), 300.

140. Daniel, W. W. *Applied Nonparametric Statistics* (Duxbury, 2000).

141. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y. & Zheng, X. *TensorFlow: A system for large-scale machine learning* in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* (2016), 265.

142. Liu, Q. & Wang, D. *Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm* 2016.

143. Lisani, J.-L., Petro, A.-B. & Sbert, C. Color and Contrast Enhancement by Controlled Piecewise Affine Histogram Equalization. *Image Processing On Line* **2**, 243 (2012).

144. Choudhary, A., Tong, L., Zhu, Y. & Wang, M. Advancing Medical Imaging Informatics by Deep Learning-Based Domain Adaptation. *Yearbook of Medical Informatics* **29**, 129 (2020).

145. Chen, X., Pawlowski, N., Glocker, B. & Konukoglu, E. *Unsupervised Lesion Detection with Locally Gaussian Approximation* in *Machine Learning in Medical Imaging* (Springer International Publishing, Cham, 2019), 355.

146. Wang, J., He, L., Zheng, H. & Lu, Z.-L. Optimizing the Magnetization-Prepared Rapid Gradient-Echo (MP-RAGE) Sequence. *PLOS ONE* **9**, 1 (2014).

147. Jafari, R., Spincemaille, P., Zhang, J., Nguyen, T. D., Luo, X., Cho, J., Margolis, D., Prince, M. R. & Wang, Y. Deep neural network for water/fat separation: Supervised training, unsupervised training, and no training. *Magnetic Resonance in Medicine* (2020).

148. Gözcü, B., Sanchez, T. & Cevher, V. *Rethinking Sampling in Parallel MRI: A Data-Driven Approach* in *2019 27th European Signal Processing Conference (EUSIPCO)* (2019), 1.

149. Gözcü, B., Mahabadi, R. K., Li, Y. H., Ilıcak, E., Çukur, T., Scarlett, J. & Cevher, V. Learning-Based Compressive MRI. *IEEE Transactions on Medical Imaging* **37**, 1394 (2018).

150. Sherry, F., Benning, M., los Reyes, J. C. D., Graves, M. J., Maierhofer, G., Williams, G., Schönlieb, C.-B. & Ehrhardt, M. J. *Learning the Sampling Pattern for MRI* 2020.

151. Atkinson, A., Donev, A. & Tobias, R. *Optimum Experimental Designs, with SAS* (Oxford University Press, 2007).

152. Mockus, J. *Bayesian Approach to Global Optimization: Theory and Applications* (Springer Netherlands, 2012).

153. Bahadir, C. D., Wang, A. Q., Dalca, A. V. & Sabuncu, M. R. Deep-Learning-Based Optimization of the Under-Sampling Pattern in MRI. *IEEE Transactions on Computational Imaging* **6**, 1139 (2020).

154. Zhang, J., Zhang, H., Wang, A., Zhang, Q., Sabuncu, M., Spincemaille, P., Nguyen, T. D. & Wang, Y. *Extending LOUPE for K-Space Under-Sampling Pattern Optimization in Multi-coil MRI* in *Machine Learning for Medical Image Reconstruction* (eds Deeba, F., Johnson, P., Würfl, T. & Ye, J. C.) (Springer International Publishing, Cham, 2020), 91.

155. Jin, K. H., Unser, M. & Yi, K. M. *Self-Supervised Deep Active Accelerated MRI* 2019.

156. Zhou Wang, Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* **13**, 600 (2004).

157. Antun, V., Renna, F., Poon, C., Adcock, B. & Hansen, A. C. On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proceedings of the National Academy of Sciences* **117**, 30088 (2020).

158. Simonyan, K. & Zisserman, A. *Very Deep Convolutional Networks for Large-Scale Image Recognition* 2015.

159. Ghodrati, V., Shao, J., Bydder, M., Zhou, Z., Yin, W., Nguyen, K.-L., Yang, Y. & Hu, P. MR image reconstruction using deep learning: evaluation of network structure and loss functions. *Quantitative Imaging in Medicine and Surgery* **9** (2019).

160. Seitzer, M., Yang, G., Schlemper, J., Oktay, O., Würfl, T., Christlein, V., Wong, T., Mohiaddin, R., Firmin, D., Keegan, J., Rueckert, D. & Maier, A. *Adversarial and Perceptual Refinement for Compressed Sensing MRI Reconstruction* in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* (eds Frangi, A. F., Schnabel, J. A., Davatzikos, C., Alberola-López, C. & Fichtinger, G.) (Springer International Publishing, Cham, 2018), 232.

CURRICULUM VITAE

PERSONAL DATA

Name	Kerem Can Tezcan
Date of Birth	March 18, 1989
Place of Birth	Istanbul, Turkey
Citizen of	Turkey

EDUCATION

2015– 2021	ETH Zürich, Switzerland *Doctoral studies*
2012 – 2014	Heidelberg University, Germany *Master of Science*
2008– 2012	Sabanci University, Turkey *Bachelor of Science*