Learning-based Algorithms for Inverse Problems in MR Image Reconstruction and Quantitative Perfusion Imaging

by

Anish Lahiri

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical and Computer Engineering)
in The University of Michigan
2021

Doctoral Committee:

Professor Jeffrey A Fessler, Co-chair
Professor Luis Hernandez-Garcia, Co-chair
Professor Douglas C. Noll
Professor Clayton Scott
Associate Professor Nicole Seiberlich

Anish Lahiri

anishl@umich.edu

ORCID iD: 0000-0002-5763-4106

ACKNOWLEDGEMENTS

First and foremost, I want to thank my advisors Prof. Jeff Fessler and Prof. Luis Hernandez-Garcia, without whose mentoring and constant support, I cannot imagine myself pursuing a doctoral degree. My directed study with Jeff and his group back in the summer of 2016 was such an amazing experience that I subsequently decided to stay on in Ann Arbor for the next several years. Jeff's patience and ability to get me out of difficult situations is probably unparalleled in my life. Deciding to be co-advised by Luis was the best decision I've made in my doctoral career. From teaching me how to use an MR scanner to helping me organize my research– his support and hands-on, jovial approach to mentoring has made my PhD a very enjoyable experience. I would also like to thank Profs. Douglas Noll, Nicole Seiberlich and Clayton Scott, it is an incredible honor to have them as part of my dissertation committee.

Having been part of two labs for the last several years, I have had the great fortune to be able to interact with some amazing labmates, who have inspired me and allowed me to grow both as a researcher and person. Melissa, Shouchang, Naveen, Michelle, Tianrui, Guanhua, Claire, Cameron, Jonas, Caroline, Mariama and others, both past and present– I hope that I continue to be good friends with them in the next stages of my life.

I cannot begin to thank my parents Jayanta and Leena, my sister Aatrika, and my uncle Sailendranath enough for the roles they have played in my life. Their unconditional love, support and care is the prime reason for my achieving this milestone. Finally, any acknowledgement of family would be incomplete without mention of my sizeable Ann Arbor family–Ari, Raghav, Jason, Charisma, Molly, Rachana, Mariana, Navid, Nayanika, Shreyas, Preeti, Bikash, Harsha, Zheng, Dinank, Sushmitha, Vikramjit, Partha, as well as my age-old friends Sayan, Manidip and Ankit–their love has got me past some of the most difficult times in my life. I am indebted to them forever.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

ABSTRACT

Medical imaging has become an integral part of the clinical pipeline through its widespread use in the diagnosis, prognosis and treatment planning of several diseases. Magnetic Resonance Imaging (MRI) is particularly useful because it is free from ionizing radiation and is able to provide excellent soft tissue contrast. However, MRI suffers from drawbacks like long scanning durations that increase the cost of imaging and render the acquired images vulnerable to artifacts like motion. In modalities like Arterial Spin Labeling (ASL), which is used for non-invasive and quantitative perfusion imaging, low signal-to-noise ratio and lack of precision in parameter estimates also present significant problems. In this thesis, we develop and present algorithms whose focus can be divided into two broad categories. First, we investigate the reconstruction of MR images from fewer measurements, using data-driven machine learning to fill in the gaps in acquisition, thereby reducing the scan duration. Specifically, we first combine a supervised and an unsupervised (blind) learned dictionary in a residual fashion as a spatial prior in MR image reconstruction, and then extend this framework to include deep supervised learning. The latter, called blind primed supervised (BLIPS) learning, proposes that there exists synergy between features learned using shallower dictionary-based methods or traditional prior-based image reconstruction and those learned using newer deep supervised learning-based approaches. We show that this synergy can be exploited to yield reconstructions that are $\approx 0.5 - 1$ dB better in PSNR (in avg. across undersampling patterns). We also observe that the BLIPS algorithm is more robust to a scarcity of available training data, yielding reconstructions that are ≈ 0.8 dB better (in terms of avg. PSNR) compared to strict supervised learning reconstruction when training data is very limited. Secondly, we aim to provide more precise estimates for multiple physiological parameters and tissue properties from ASL scans by estimation-theory-based optimization of ASL scan design, and combination with MR Fingerprinting. For this purpose, we use the Cramer-Rao Lower Bound (CRLB) for optimizing the scan design, and deep learning for regression-based estimation. We also show that regardless of the estimator used, optimization improves the precision in parameter estimates, and enables us to increase the available 'useful' information obtained in a fixed scanning duration. Specifically, we successfully improve the theoretical precision of perfusion estimates by 4.6% compared to a scan design where the repetition times are randomly chosen (a popular choice in literature) thereby yielding a 35.2% improvement in the corresponding RMSE in our *in-silico* experiments. This improvement is also visually evident in our *in-vivo* studies on healthy human subjects.

CHAPTER I

Introduction

Magnetic Resonance Imaging (MRI) uses strong magnetic fields and radio-frequency (RF) waves to generate high resolution, detailed images of tissues and other anatomy within the body. It has found widespread use in clinical practice, owing to advantages like good soft tissue contrast, absence of ionizing radiation, and its ability to capture a wide variety of physiological phenomenon through various techniques. It has found use in the diagnosis of several disorders like: aneurysms of cerebral vessels, disorders of the eye and inner ear, multiple sclerosis, spinal cord disorders, stroke, tumors, brain injury from trauma, etc. However, it does not come without challenges. One very prominent hurdle associated with its use is the amount of time that is required to acquire images. Some scanning procedures can last up to an hour, which involves confining the subject to a cramped space for these durations. Not only does this prove to be a nuisance in terms of the comfort associated with the process, the time costs associated with it also makes MRI an expensive procedure for imaging. Another consequence of this caveat is that it cannot be used in situations requiring immediate discernment.

In MRI, measurements are acquired by sampling the transversal spins in the object being imaged after it is excited by radiofrequencey waves. Due to the application of spatially varying magnetic field gradients, only spins at a certain resonant frequency can be sampled during acquisition– thereby enabling localization of spins based on their spatial frequency. A consequence of this is raw MR measurements are in the frequency domain, dubbed the k-space, unlike a lot of imaging modalities (X-Ray imaging, for example) where acquisition happens in the image domain. The scan duration in MRI is dependent on the number of such measurements obtained in the frequency domain. One way of reducing this scan duration is to acquire fewer k-space measurements than that would be traditionally deemed necessary, i.e., adopting sub-Nyquist level sampling. Furthermore, in MRI applications like cardiac imaging, that are aimed at capturing information about phenomena which are dynamic in nature, acquiring fully sampled measurements may be impossible. However, vigorous 'undersampling' entails losing out on information that is necessary for a 'perfect' reconstruction. Since reducing the scanning duration can lead to sacrificing the quality of the image being acquired, there is a need for developing algorithms that allow for the reconstruction of clean images from limited measurements.

The rising demand for information beyond qualitative assessment of medical conditions, to improve healthcare on an individual level, has also become very apparent in recent years. An increasing number of healthcare professionals are choosing to rely upon quantitative information about the health of their patients, in lieu of generic indicators, for making more informed case-specific decisions towards precise monitoring, diagnosis and treatment planning [14, 21, 33, 31].

As such, quantitative imaging is poised to become an important tool for diagnosing and subtyping diseases more precisely. One such category of quantitative imaging techniques targets brain hemodynamics, such as perfusion and blood volume. Quantitative perfusion imaging, for instance, can play a critical role in the diagnosis, prognosis, and assessment of treatment efficacy in cerebrovascular diseases. Some examples are stroke, Alzheimer's, multiple sclerosis, Moyamoya and other neurodegenerative disorders [12, 19, 38, 10, 17, 3, 16, 9, 2]. Traditional clinical practice involves the use of dynamic susceptibility-contrast MRI that requires gadolinium-based contrast agents for imaging perfusion (passage of blood flowing through a blood vessel into a tissue or region of interest), even though it introduces nuisances like: increased costs, inability to image repetitively or frequently due to the need for tracer clearance, and in certain cases, the risk of nephrogenic disorders.

Arterial Spin Labeling (ASL) MRI provides a viable alternative to such methods, owing to the fact that it is a non-invasive, quantitative perfusion imaging technique that magnetically labels blood itself using radio-frequency waves instead of relying upon contrast agents. But hemodynamic information only accounts for about $\approx 1\%$ of the total acquired signal in ASL imaging, which makes signal-to-noise ratio a significant obstacle in getting good quantitative estimates. In ASL, low signal-to-noise ratio is a challenge that is traditionally overcome by averaging several acquired images. Again, such a solution inevitably increases the time expended in generating ASL images, posing a problem. Another challenge is that parameter estimates obtained using ASL suffer from a bias-variance trade-off, wherein improving the accuracy of the estimates of perfusion or transit times through better modeling can lead to a loss of precision. This is because more parameters (nuisance, or otherwise) typically need to be estimated in an improved model.

Our work in this thesis focuses on addressing some of the problems in both the aforementioned areas by developing data-driven machine learning algorithms that allow for better reconstruction of MR images from limited or sub-Nyquist sampling of measurements. On the MR-based quantitative perfusion imaging front, we focus on optimization of ASL scan parameters for improved precision in estimates of perfusion and related physiological parameters and tissue properties. The organization and contributions of the thesis are as follows:

Chapter II briefly provides some background for the concepts that are pertinent to the algorithms that are formulated and developed in the subsequent chapters.

Chapter III focuses on using MR Fingerprinting (MRF) to overcome the challenges resulting from single-compartmental signal modeling in ASL, mainly in the form of inaccurate quantification of perfusion and other hemodynamic parameters [68]. Though MRF allows for the simultaneous estimation of multiple parameters through the use of an appropriate signal model, the two-compartment model for ASL signals accounts for 6-7 parameters. This complexity results in a significant loss of precision in estimates, and would greatly increase the estimation time using traditional dictionary-search based estimation in MRF. We use neural network-based regression as an alternative to reduce the estimation time, thereby improving the estimation time by three orders of magnitude. However, since the loss of precision in estimates is a hindrance that is completely independent of the estimator used, we adopt a Cramer-Rao Bound based optimization of the ASL labeling durations to increase the amount of information available in a fingerprint about the underlying parameters that generated it. We show that optimization makes a crucial difference in the viability of MRF-ASL as a clinically implementable technique.

Chapter IV expands upon our proof-of-concept work in the previous chapter, by developing a single multi-slice scan for MRF-ASL that, in a scan duration of 600s, allows for the simultaneous estimation of perfusion, transit time and other relevant parameters. For this purpose, we explore Velocity Selective ASL

(VSASL) pulses, which we show, using the CRB, has increased sensitivity to perfusion, when compared to traditional pseudo-Continuous ASL (pCASL) pulses. Proper modeling of VSASL MRF signals also removes interference from nuisance parameters like Magnetization Transfer Rate (MTR), but also requires inclusion of inversion efficiencies in the signal model. Our *in-silico* experiments indicate that VSASL and MRF, when combined, can provide both precise and unbiased estimates of hemodynamic parameters across the entire brain, while also maintaining clinically feasible scan times. Our preliminary *in-vivo* results show promising performance in perfusion estimation across several subjects using this method.

Chapter V introduces a novel dictionary learning method for combining supervised and blind learning in the presence of limited training data, dubbed Super-BReD Learning. Our technique uses two dictionaries: one learnt from uncorrupted data (supervised), and the other from corrupt data (blind), to jointly sparsify and represent signals such as image patches. The dictionaries represent the data in a residual fashion, thereby complementing each other. The algorithm shows promising performance in tasks such as denoising, as well as inverse problems such as single-coil compressed-sensing MRI. Using Super-BReD Learning, we also formulate a problem for denoising/quantifying ASL images from fewer acquired frames, in an attempt to reduce the number of required averages to form clean images, which in turn reduces the scan duration.

In Chapter VI, we propose a method to extend the Super-BReD Learning framework to be able to effectively utilize training data (consisting of fully sampled reconstructions and corresponding undersampled reconstructions) when available. Essentially, we examine a combined supervised unsupervised framework involving dictionary-based blind learning and deep supervised learning for MR image reconstruction from under-sampled k-space data. A major focus of the work is to investigate the possible synergy of learned features in traditional shallow reconstruction using sparsity-based priors and deep prior-based reconstruction. Specifically, we propose a framework that uses an unrolled network to refine a blind dictionary learning-based reconstruction. We compare the proposed method with strictly supervised deep learning-based reconstruction approaches on several datasets of varying sizes and anatomies. We also compare the proposed method to alternative approaches for combining dictionary-based methods with supervised learning in MR image reconstruction. The improvements yielded by the proposed framework suggest that the blind dictionary-based approach preserves fine image details that the supervised approach can iteratively refine, suggesting that the features learned using the two methods are complementary.

CHAPTER II

Background

This chapter discusses some of the applications and general concepts that will be used in this thesis. We briefly outline the concept of MRI, Compressive Sensing and its use in MRI, and Arterial Spin Labeling, an MRI-based technique to non-invasively quantify perfusion.

2.1 Magnetic Resonance Imaging

MRI involves using a strong magnetic field (B_0) to align the proton *spins* present in the hydrogen atoms in the body (mostly in the form of water), and *tipping* them with the help of Radio Frequency (RF) waves. From the perspective of classical mechanics, these spins then precess around the B_0 field, and produce a voltage/signal in a receiver coil which is recorded as a measurement. These measurements, however, are not in the image or spatial domain, as is the case with imaging techniques like X-Ray radiography or photography. Instead, these measurements are acquired in the frequency domain, known as *k-space* in MRI. Magnetic field gradients in the x and y direction (dubbed frequency and phase encoding) are used to localize the *k-space* frequency at which the measurement is being acquired. For 2D imaging, gradients in the z direction are used to selectively excite the spins in a slab of the object being imaged. Once a sufficient number of *k-space* measurements have been acquired, an image can be created by transforming the measurements to image domain using the Fourier transform. In 3D or volume imaging, spins in the entire volume is excited instead of a thin slab, and spatial encoding is performed by adding appropriate phase encoding in the z direction in addition to the usual phase and frequency encoding used in 2D acquisitions.

2.2 Compressed Sensing MRI

Acquiring every k-space measurement necessary to reconstruct an MR image can be a time-consuming process. Compressed sensing [61] allows for the reconstruction of a signal from sub-Nyquist level sampling by exploiting knowledge of the sparsity of the signal in some domain. When presented in the form of a linear inverse problem, this often involves inclusion of regularization of some form to the problem of solving for a single plausible image amongst uncountable possible candidates that fit the acquired data. Compressed Sensing MRI or CS-MRI formulates and solves the problem of reconstructing MR images from limited k-space measurements [69]. If $\boldsymbol{y}$ is a set of k-space measurements obtained for the object being imaged, and the MRI system matrix is $\boldsymbol{A}$, this regularized inverse problem to reconstruct an image $\hat{\boldsymbol{x}}$ from these measurements

is typically expressed as:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \nu \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \mathcal{R}(\boldsymbol{x}), \tag{2.1}$$

where $\mathcal{R}$ is a regularizer that reflects the prior information about the reconstructed image. Several works have investigated appropriate regularization for CS-MRI, ranging from TV regularization which penalizes the smoothness of the reconstructed image using pixel/patchwise difference operators (based on the assumption that naturally occurring images are smooth and not random), transform sparsity-based regularization of wavelet coefficients of the image which uses $\mathcal{R}(\boldsymbol{x}) = \|\Psi\boldsymbol{x}\|_1$, ($\Psi$ being a wavelet basis– essentially this focuses on recovery of significant wavelet coefficients by applying non-linear thresholding)) [69] to patch-based learned sparsifying dictionary or transforms [95, 78], where the regularizer exploits the learned transform domain sparsity of reconstructed image patches, or assumes that patches in the reconstructed image can be expressed as sparse linear combinations of the atoms of a learned dictionary:

$$\mathcal{R}(\boldsymbol{x}) = \min_{\boldsymbol{D},\boldsymbol{Z}} \|\mathcal{P}\boldsymbol{x} - \boldsymbol{D}\boldsymbol{Z}\|_2^2 + \lambda\|\boldsymbol{Z}\|_0, \tag{2.2}$$

$$\text{or } \mathcal{R}(\boldsymbol{x}) = \min_{\boldsymbol{W},\alpha} \|\boldsymbol{W}\mathcal{P}\boldsymbol{x} - \alpha\|_2^2 + \lambda\|\alpha\|_0. \tag{2.3}$$

Here, (2.2) and (2.3) correspond to dictionary and transform-based regularization respectively. $\mathcal{P}$ is a patch extraction operator, and $\boldsymbol{W}$ and $\boldsymbol{D}$ are the dictionary and transform matrices. These can either be learned or fixed beforehand. α and $\boldsymbol{Z}$ represent sparse representation coefficients. Eqn. (2.2) adopts the synthesis model that tries to express each patch in the image as the sum of a few fundamental components, while eqn. (2.3) adopts the analysis model that posits image patches can be decomposed into a few significant coefficients if an appropriate transform is applied.

The success of deep learning in domains like computer vision and image processing, and the availability of pairwise training data (consisting of fully sampled reconstructions and corresponding undersampled reconstructions) has ushered in the use of deep-neural networks in compressed-sensing MRI, where a majority of techniques rely upon the richness of CNNs and GANs in their ability to learn features from training data. Typically, in these algorithms, the output of a deep network is used to regularize the MRI reconstruction problem, i.e., $\mathcal{R}(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{V}_\theta(\boldsymbol{x}')\|_2^2$, where $\boldsymbol{V}$ is a deep CNN whose weights are denoted by θ, and $\boldsymbol{x}'$ is an initial estimate of the image being reconstructed, like a zero-filled reconstruction. These deep networks are typically trained in a supervised fashion using pairwise training data consisting of a fully sampled ground truth reconstruction as the target, and the corresponding undersampled reconstruction as an input to the network (often, a zero-filled reconstruction is chosen for this purpose). Where such pairwise training data is not available, generative adversarial networks are often used instead. In these settings, $\mathcal{R}(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{G}_\phi(\boldsymbol{x}')\|_2^2$, where $\boldsymbol{G}$ is a CNN trained using unpaired or partially paired training data, and an (additional) adversarial objective, and ϕ are its weights, and $\boldsymbol{x}'$ is an initial estimate of the image being reconstructed, like a zero-filled reconstruction. [49, 58] Other than the reduced demands for fully-sampled training data, an advantage of using GANs for regularized reconstruction is that they yield images that have more realistic texture.

The category of supervised algorithms that have found the most success in reconstructing MR images from limited measurements are a class of algorithms called *unrolled algorithms*[128]. A trademark of such algorithms is that they usually extend iterative approaches to image reconstruction to incorporate pairwise training data. Usually, this involves replacing one or multiple stages in a single iteration of an image reconstruction algorithm by a deep CNN [45, 76, 64]. While, unrolled loop algorithms have demonstrated

their superiority amongst supervised algorithms, and are often treated as the replacement to traditional prior-based iterative reconstruction algorithms, there has been little investigation into whether features learned by unrolled loop algorithms subsume those enforced in traditional priors like dictionary or transform learning priors, or even Total Variation (TV)-based methods. This investigation will be a focus of our work in chapter VI.

2.3 Arterial Spin Labeling

Brain perfusion (defined as the volume of blood delivered to a unit of tissue per unit of time) is a well known indicator of tissue metabolism and function. As such, it is proving to be a powerful workhorse to study brain function and gaining prominence as a clinical tool, but it is still an evolving technique. Arterial Spin Labeling (ASL) [47] is an MR based quantitative, yet non-invasive perfusion imaging technique. It does not require the use of tracer injections which have been known to cause complications in patients with nephrogenic disorders, and have low repeatability due to the time taken for the tracer to exit the circulatory system. The concept behind ASL is relatively simple: it is very similar to tracer injection perfusion measurements (e.g., bolus tracking MRI, autoradiography, PET, etc.) except that, instead of injecting a tracer into the blood stream and tracking its accumulation into the tissue, the tracer consists of the blood water in the arteries itself. The tracer is created by inverting the magnetization of the blood in the arteries that feed the organ of interest with a train of RF pulses. As this "labeled" blood flows into the tissue, it reduces the available magnetization in the tissue. As a result, images collected downstream of the labeling location appear slightly darker. By subtracting the labeled images from a set of control images, we can calculate the amount of blood that entered the organ since the beginning of the labeling period. For example, in brain perfusion imaging, the most common use for ASL, the blood is labeled just before it enters the brain through the carotid and vertebral arteries. After a brief delay (called post-labeling delay or PLD), which gives the labeled blood time to travel from the location of labeling to the tissue of interest, a set of "labeled" brain images is acquired. Next, a second set of "control" images is acquired identical to the first, except that this time, the labeling pulses do not invert the blood magnetization at all, but simply serve as a control for any side effects of the labeling pulses (namely magnetization transfer). The subtraction of these two images is roughly proportional to the perfusion rate (see Fig. 2.1). If desired, one can acquire a time series of such image pairs to examine the brain's activity during a stimulation paradigm. Once a difference image (or a set of difference images with multiple PLDs) has been acquired, it is usually fitted to an appropriate model to obtain images with physically meaningful units. Such models often vary on the basis of the desired application, but usually describe the ASL signal behaviour (difference, or otherwise) in terms of parameters like the perfusion, arterial transit time or bolus arrival time, and blood and tissue T_1 [66].

Most signal models used in ASL applications are either one or two compartments. In the single compartmental model, it is assumed that blood enters the voxel volume being imaged through an artery and the exchange of water from the artery to the tissue is comprehensive and instantaneous, i.e., assuming a well-mixed compartment at the imaging voxel. [5] Another (simpler and more popular) way of describing the ASL signal in a single compartment model is through the convolution of the arterial magnetization function with a product of the "residue" (also referred to as "retention") function, and a magnetization relaxation function

The problem with single-compartmental modeling is that it often leads to overestimation of perfusion, and fails when the bolus arrival times are short. This limitation is because it neglects the distinction between blood relaxing the tissue and blood relaxing the arteries, even though the T_1 corresponding to these are significantly

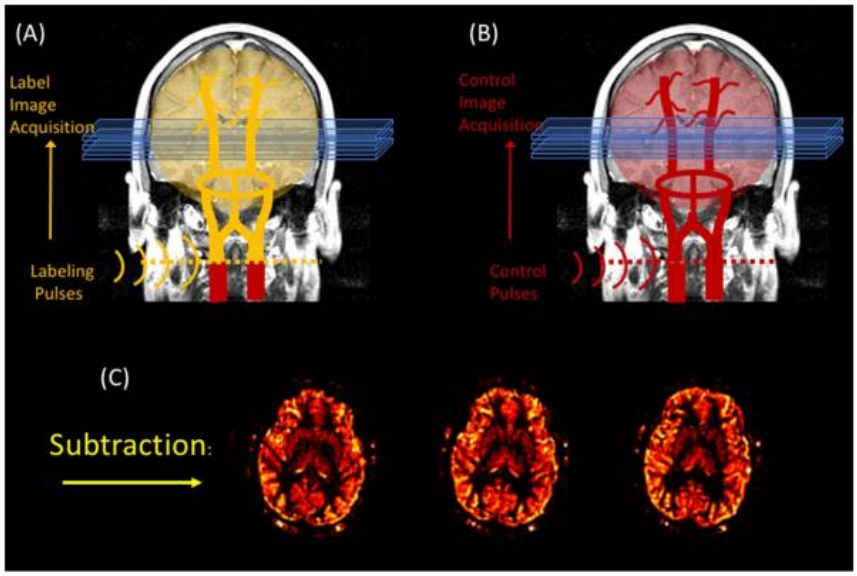


Figure 2.1: A basic pseudo-continuous ASL experiment. The arterial water is inverted as it crosses a plane through the neck prior to acquiring the "labeled" image. A second "control" image is acquired without the label. The subtraction of these two images yields a perfusion weighted image.

different.

Two compartmental modeling corrects for such issues by introducing an additional compartment in the model, where the arterial blood enters before perfusing into the tissue. The fraction of blood in the voxel that the arterial compartment accounts for is explained by a new parameter called the blood volume fraction [29, 44, 4, 6, 28]. The magnetization in the arterial and tissue compartment also relaxes separately. There are also other variations of the two-compartment model, where there is no exchange between the two compartments. Usually in such models the artery branches into two components: one component feeds blood into the tissue compartment and is absorbed in entirety, and a non-permeable pass-through arterial component, where there is no interaction with the tissue compartment. Fig. 2.2 depicts the one and two compartment models.

The main variations of ASL have to do with the labeling scheme. One could label a large segment of the neck region with a single pulse, as in the case of Pulsed ASL (PASL) techniques [52], or one could apply a long pulse (or a train of pulses) at a thin slice through the neck that labels the blood as it flows through it, as in the case of continuous and pseudo-Continuous ASL (pCASL) [55]. More recently, "velocity selective" techniques (VSASL) [53] have been developed such that only moving spins are labeled, regardless of their spatial position.

Regardless of the variations involved, a common theme that has plagued ASL techniques is the poor signal-to-noise ratio (SNR) that results from the difference image accounting for only about 1% of the total signal in a label or control image. To compensate for this shortcoming, the traditionally adopted solution is to acquire several label-control pairs for the same PLD, and averaging them to boost the SNR. This approach however, requires additional scanning time which can be inconvenient, and also renders this technique susceptible to artifacts such as patient motion, and necessitates additional strategies during acquisition or post-processing of these images. In a similar vein, another problem is associated with ASL is the accuracy-precision trade-off, which is rooted in the assumption that certain modeled parameters that influence the ASL signal can be held to literature values everywhere in a region of interest. Some of the work presented in this

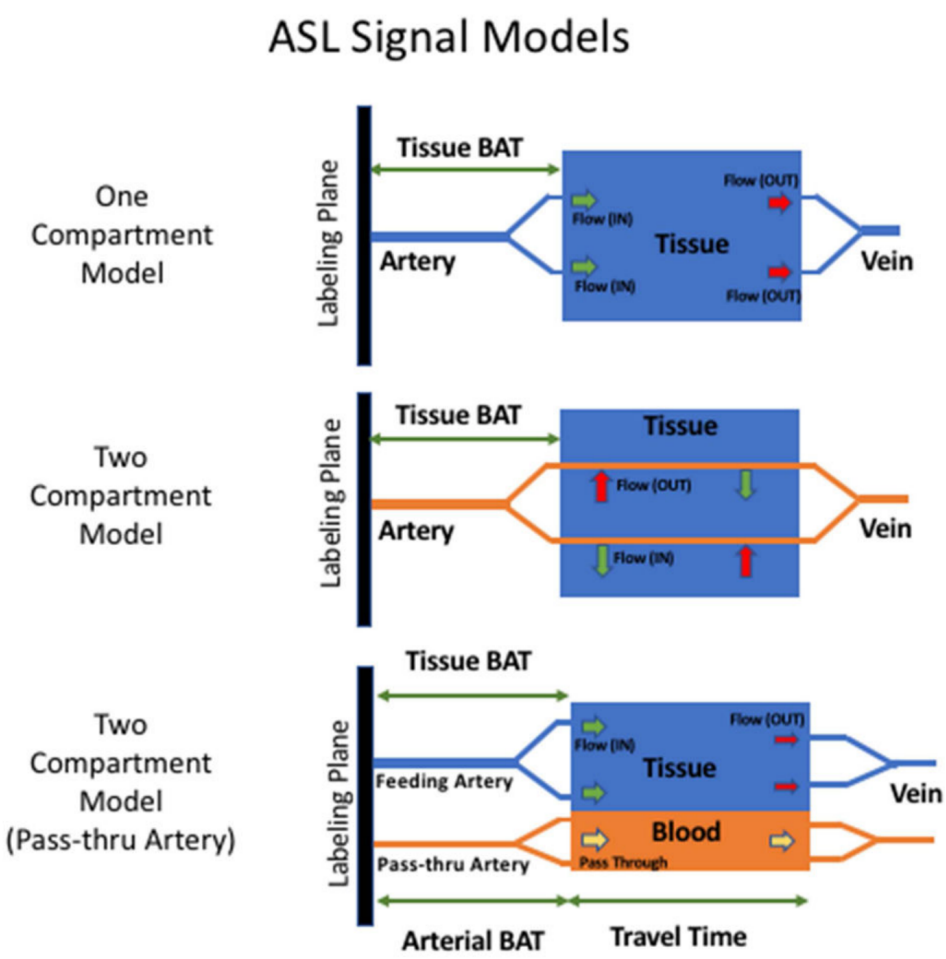


Figure 2.2: ASL signals are typically described using one or two compartmental models. While both the second and third figure in this model are two compartment, the first considers exchange between the arterial and tissue compartment, while the latter considers pass-through effects in the arterial compartment.

thesis aims at addressing this issue.

CHAPTER III

Optimizing MRF-ASL Scan Design for Precise Quantification of Brain Hemodynamics using Neural Network Regression[1]

3.1 Introduction

Quantitative imaging of tissue properties is gaining increasing prominence in the diagnosis, prognosis and treatment planning of several diseases, e.g., [14, 21, 33, 31]. Moving beyond the variations associated with qualitative intensity-based imaging allows gleaning information focused on the physiological phenomena being investigated. In the context of cerebrovascular disorders, quantitative perfusion imaging has found several applications [12, 19, 38, 10, 17, 3, 16, 9, 2]. Typically, quantitative imaging of perfusion involves gadolinium-based contrast enhanced MRI, which suffers from lack of fast repeatability and risks involved in cases of subjects with nephrogenic disorders [37, 32].

Arterial Spin Labeling (ASL) [11] provides an alternative to contrast agent based MRI by magnetically labeling blood flowing into organs or tissues of interest. ASL temporary inversion of the spins present in flowing blood upstream of the organ under scrutiny, by applying radiofrequency (RF) magnetic pulses. These inverted spins then behave like an *endogenous* tracer that is detectable after it perfuses into the tissue in the relevant organ shortly afterwards. ASL is non-invasive, non-toxic, quickly repeatable and has a much simpler workflow than contrast enhanced MRI. However, ASL images are limited by low spatial and temporal signal-to-noise ratio (SNR) [30, 39, 7]. This drawback is more pronounced in white matter, where traditional ASL methods perform poorly[36]. Estimating perfusion using ASL requires knowledge of a number of tissue properties or hemodynamic factors that are usually fixed to literature values. In reality, some of these (tissue T_1 for example) vary significantly from region to region. Fixing these parameter values can lead to significant biases in perfusion estimates and efforts to estimate such factors from separate scans can be undesirably time-consuming

MRF is a recently developed technique [24] that estimates multiple hemodynamic parameters and tissue properties simultaneously from a single acquisition. This approach improves accuracy at the possible expense of precision in estimates. Nevertheless, information accrued from the additional estimated parameter maps may aid understanding of physiological conditions. MRF utilizes transient signals obtained by varying imaging parameters such as the repetition time (TR) or flip angles as identifiers for the underlying physiological factors. If precise information about all the parameters that affect the ASL signal were available beforehand, such transient signals would not be necessary to obtain information about perfusion, and averaging multiple label and control images obtained using the same labeling duration followed by fitting the difference of the

[1] based on work in [68]

average images to a model would be sufficient to obtain estimates of perfusion. However, as mentioned before, obtaining such information separately can prove to be costly, and using MR Fingerprinting allows us to glean additional information about hemodynamic parameters and tissue properties from the same scan using multiparametric estimation.

A standard approach to multiparametric estimation using such a technique involves searching through 'dictionaries' consisting of signals generated by feasible combinations of parameters, in a Maximum Likelihood manner. In an ASL based fingerprinting [35, 41] setting, the observed signal depends on several parameters (typically 5-7), presenting a considerable challenge to precise estimation. For example, with more parameters to estimate, it becomes difficult to maintain and search a 'fine' dictionary. Specifically, [35] reports a 2 hr estimation time for a single slice, with a dictionary quantization of 6mL/100g/min for perfusion.

As an alternative, this work uses a regression-based estimator to generate predictions from fingerprint data. While the use of regressors for MRF-based estimation has become more prevalent recently [27, 8], our preliminary work [20] was the first to investigate neural network regression for ASL Fingerprinting, where there are considerably more parameters to estimate. Estimation using neural network regression allows for much faster estimation, and overcomes quantization error.

Regardless of the estimation technique used, if the ASL fingerprints themselves are insufficiently sensitive to the underlying parameters, then estimates obtained from them will lack precision. Thus, the first goal of this paper is to increase the information conveyed by fingerprints. This is done using Cramer-Rao bound based optimization of scan parameters in ASL. An example of such scan parameters are the labeling durations in the scan. While there has been some work on optimizing scan-design for MR sequences in quantitative imaging [26, 43], and even specifically in ASL [42, 40], our work is the first to investigate it in an ASL fingerprint setting. While most other pertinent methods focused on providing precise results in regions of gray matter, we use a cost function having a comprehensive uniform prior. This enables precise estimation over a wide range of feasible parameter values, including white matter or potential anomalies. We also constrain our optimization procedure to adhere to a fixed scan time for practicality. The primary focus of our work is to establish the need for scan design optimization regardless of the estimation technique involved. Through our work, we establish that optimized scan design coupled with regression-based estimation should further the transition of ASL Fingerprinting to clinical use.

The rest of the paper is organized as follows: Section 3.2 introduces the ASL signal model used in this work. This model, along with a Cramer-Rao bound based cost function, is used to optimize our scan design. Next, we design a neural network regressor for estimating hemodynamic parameters and tissue properties. The neural network is trained using fingerprints simulated with a combination of the optimized scan design and the described model. We also devise a post-processing technique to mitigate nuisance effects in our acquisitions. Thereafter, we describe the creation of in-silico datasets to test the performance of our methods, as well the methods we compare to in our work. We then describe the in-vivo experiments we performed in the validation of our designed methods. Section 3.3 shows the theoretical predictions of the performance of our scan design, as well as the results of comparisons in-silico and in-vivo with other methods, namely two other MRF ASL scans found in literature and multi-PLD ASL. Section 3.4 elaborates upon these results and the inferences we draw from them. Section 3.5 describes our conclusions.

3.2 Methods

3.2.1 ASL Signal Model

To describe the ASL signal in the brain for scan design optimization and parameter estimation, we used the two-compartment model depicted in Fig. 3.1. Although the single compartment model introduced in [5] has been the *de-facto* standard in ASL literature in the past, several works [29, 44, 4, 6, 28] have raised issues of oversimplification associated with single-compartment modeling, and have adopted two-compartment models. For ASL fingerprinting, such models have been highlighted in [41, 34]. Our chosen model for the ASL signal consisted of separate compartments for blood in tissue and arteries, as well as additional provisions to incorporate Magnetization Transfer [2] effects. In the model, magnetically labelled blood flows into the arterial compartment through the arterioles, and perfuses into the tissue compartment therein. The Cerebral Blood Volume fraction (CBVa) determines the portion of the acquired signal to which each compartment contributes, and T_1 relaxation of blood and tissue is accounted for in the signal description. The longitudinal magnetization of the tissue compartment thus evolved as:

$$\frac{dM_{\text{tis}}(t)}{dt} = -\frac{M^0_{\text{tis}} - M_{\text{tis}}(t)}{T_{1,\text{tis}}} + f \cdot M_{\text{art}}(t) - \frac{f}{\lambda} \cdot M_{\text{tis}}(t) - K_m(t) \cdot M_{\text{tis}}(t), \tag{3.1}$$

where M_{tis} and M_{art} represent the magnetization in the tissue-compartment and the arterial compartment respectively, λ is the blood-brain partition coefficient, K_m is the Magnetization Transfer rate (MTR), f is the rate of perfusion. Here, $T_{1,\text{art}}$ is the arterial relaxation time, and $T_{1,\text{tis}}$ (truncated to T_1 in later sections) the relaxation time of tissue. The input to the arterial compartment was determined by a labeling function, that is described in eqn. (3.2). The arterial magnetization was described using an input or labeling function as follows:

$$M_{\text{art}}(t) = 1 - 2 \cdot \alpha \cdot \text{inp}(t) \cdot e^{-\frac{(t-\delta)}{T_{1,\text{art}}}}, \tag{3.2}$$

where α is the inversion efficiency, δ is the Bolus Arrival Time (BAT), and

$$\text{inp}(t) = \begin{cases} 1, & \text{for labeling pulses} \\ 0, & \text{otherwise.} \end{cases}$$

The total longitudinal magnetization was thus described as the sum of the contributions of the two compartments:

$$M(t) = \text{CBV}_\text{a} \cdot M_{\text{art}} + (1 - \text{CBV}_\text{a}) \cdot M_{\text{tis}}, \tag{3.3}$$

where CBV_a is the Cerebral Blood Volume fraction described earlier. The observed ASL-MRF signal, $M(t) \cdot \sin(\beta)$, where β is the flip angle, was sampled at the time(s) of acquisition, which were dictated by the scan schedule. We used signals generated using this model for both optimization of scan design as well as for training the neural network estimators. For the purposes of our work, the values of λ and α were set to 0.9 and 85% respectively.

[2] "Magnetization Transfer (MT) is the physical process by which macromolecules and their closely associated water molecules cross-relax with protons in the free water pool." http://mriquestions.com/magnetization-transfer.html

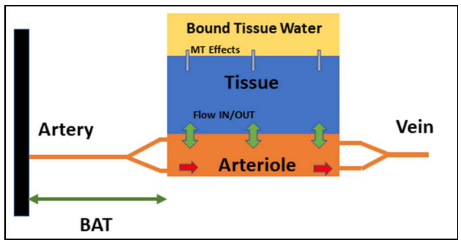


Figure 3.1: Two compartment ASL signal model used for both optimization and estimation. The unknowns in the model were: perfusion from the arteriole to the tissue, arrival time of the labeled blood bolus at the arteriole, the magnetization transfer rate, the arterial blood volume fraction, and the relaxation time of water in tissue.

3.2.2 Pulse Sequence

In ASL Fingerprinting, each repetition time (TR) in the sequence consists of a labeling period (T_{tag}), post labeling delay (T_{delay}), followed by a small period for signal acquisition (T_{aq}), (ideally instantaneous, but usually accounted for) and a period for adjustment (T_{adjust}) before the next label/control occurs. Every pulse in the sequence can either be a label, control, or 'silence' ('silence' describes acquisitions where there is no RF excitation at all in the ASL preparation phase). In this work, we vary the TR by changing the labeling durations, while holding all other parameters of the pulse sequence fixed ($T_{\text{delay}} = 55\text{ms}$, $T_{\text{aq}} = 32.4\text{ms}$, $T_{\text{adjust}} = 50\text{ms}$). The TRs in our sequence were varied to generate a signal that is informative of the underlying parameters. Section 3.2.3 describes how we optimized the aforementioned labeling durations by picking from a set of candidate schedules. The label-control order (also referred to as the 'label order' later) was pseudo-randomized, but had approximately equal numbers of label, control and silence pulses. We ensured that the total duration of the scan was fixed regardless of the number of pulses, or the duration of individual TRs. This fixed total duration was discretionary. Here, we acquired 700 images for our fingerprint, with a total scan duration of 600s for a single slice.

3.2.3 Optimization with CRLB

A major focus of our work was to investigate the benefits of scan design optimization in ASL Fingerprinting. From an information theoretic standpoint, the total information present in a signal about the underlying parameters that generated it is independent of the estimator used to quantify the parameters themselves. For example, in a regression-based estimation framework, if the signals (or 'fingerprints') themselves are too correlated, corresponding estimates will be imprecise. This is regardless of whether kernel methods or neural networks are used. In an effort to make our fingerprints more informative or sensitive to parameters like perfusion or BAT, we used the Cramer-Rao Lower Bound (CRLB) to optimize the scan design parameters (namely, the labeling durations).

The CRLB represents the minimum variance in estimates that any unbiased estimator can achieve, for a particular signal model and noise level. We focused on magnitude image data and modeled the noise as real additive white Gaussian noise (AWGN) with standard deviation σ (empirically calculated to be 0.01, which was low enough to justify the assumption of Gaussian noise in regions with sufficient SNR). Our signal model was $\underline{\mathbf{s}}(\underline{\theta};\underline{\nu}) : (\mathbb{R}^p \times \mathbb{R}^l) \rightarrow \mathbb{R}^t$, where $\underline{\theta} \in \mathbb{R}^p$ represents the p hemodynamic parameters of interest, and $\underline{\nu} \in \mathbb{R}^l$ are the l scan parameters for a scan with t time points. The CRLB is expressed as the inverse of the Fisher Information matrix. Fisher Information describes the amount of information conveyed by an

observable random variable about the parameters that generated it. It can be considered to be a measure of sensitivity of a signal to underlying parameters and is expressed as the $p \times p$ matrix:

$$\mathbf{F}(\underline{\theta};\underline{\nu}) = \frac{1}{\sigma^2} \cdot [\nabla_{\underline{\theta}}\underline{\mathbf{s}}(\underline{\theta};\underline{\nu})]^T [\nabla_{\underline{\theta}}\underline{\mathbf{s}}(\underline{\theta};\underline{\nu})], \tag{3.4}$$

where we calculated the signal gradient matrix, $\nabla_{\underline{\theta}}\underline{\mathbf{s}}(\underline{\theta};\underline{\nu}) \in \mathbb{R}^{t\times p}$ numerically at each time point using Newton's central difference method[3]. To design a 'good' fingerprinting sequence, we optimized over a set of feasible scan design parameters $\underline{\nu} \in \mathcal{V}$, which in our case were the labeling durations. For generating the fingerprints, we only varied the TRs, and fixed the flip angles nominally to 90° instead of varying it across acquisitions, which is more common in MRF literature. The reason for this is that the TRs used in our fingerprints are usually longer in duration, and as a result, the longitudinal magnetization– which is of interest in ASL, diminishes significantly, and the entirety of this magnetization needs to be sampled in the transverse plane during acquisition. In our optimization, we minimized our design cost function at a representative collection or set of true parameter values, $\mathbf{\Theta}$. These values were spread uniformly over a range. We picked our 'optimized' labeling schedule as the one that, among all others in the feasible set $\mathcal{V}$, minimized the following cost function:

$$\underline{\hat{\nu}} = \underset{\underline{\nu}\in\mathcal{V}}{\arg\min} \ \frac{1}{|\mathbf{\Theta}|}\sum_{\underline{\theta}\in\mathbf{\Theta}} \mathrm{Tr}\Big(\boldsymbol{W}\cdot\big(|\boldsymbol{F}^{-1}(\underline{\theta},\nu)|^{0.5} \oslash \boldsymbol{N}(\underline{\theta})\big)\cdot\boldsymbol{W}\Big), \tag{3.5}$$

where $\oslash$ denotes Hadamard division, $\boldsymbol{W}$ is a diagonal weighting matrix assigning priority to each hemodynamic parameter in the cost function, $|\mathbf{\Theta}|$ denotes the number of points at which the normalized standard deviation is evaluated in the cost, and $\boldsymbol{N}(\underline{\theta}) = (\underline{\theta}^{0.5})(\underline{\theta}^{0.5})^T$ is a normalization matrix that is divided element-wise into the inverse Fisher Information matrix.

We used exhaustive search to minimize the design cost function (3.5) to ensure that our optimized scan yields precise estimates over the set $\mathbf{\Theta}$ of 'ground truth' parameter values. Minimizing the above expression is tantamount to minimizing the average normalized standard deviation of parameter estimates, weighted appropriately, over a set of ground truth parameter values. Using normalized standard deviation bypasses having to combine variances of different values and units, and emphasizes cases when the standard deviation is comparable to the ground truth parameter value. For our experiments, we assigned twice the weight to perfusion precision as to all the other parameters in the cost function, which were weighted equally (this is done by manipulating $\boldsymbol{W}$). The cost was evaluated numerically. For the optimized scheme to function well even at values seen in pathological conditions, we used an adequately large feasible range for $\mathbf{\Theta}$. Specifically, the parameter ranges we used were: 12-90 ml/100g/min for perfusion, 0.002-0.03 for CBVa, 0.36-1.7 s for BAT, 0.01-0.03 s^{-1} for MTR, 0.3-3.3 s for T_1 and 54-112 degrees for flip angles. $|\mathbf{\Theta}|$ was picked to be 50 during labeling schedule optimization, and 75 during label order optimization and 250 for final evaluation of the designed scan. These points were picked using a uniform random distribution on the range for $\mathbf{\Theta}$ described earlier. $|\mathbf{\Theta}|$ was set to a lower number during optimization because increased number of parameter space evaluations would increase optimization time. We chose a uniform random distribution for the points to capture a wide variety of anomalous behaviour.

The set of feasible candidate labeling schedules, $\mathcal{V}$, consisted of ASL timing sequences with variable labeling times but fixed pre and post labeling delays. Thus, TR was allowed to vary depending on the labeling duration. As shown in Fig. 3.2, each labeling schedule (essentially a collection of labeling durations) in this

[3] $\nabla_{\theta}\underline{\mathbf{s}}(\theta;\underline{\nu}) = \frac{\underline{\mathbf{s}}(\theta+h)-\underline{\mathbf{s}}(\theta-h)}{2h} \in \mathbb{R}^t$ for a single parameter θ

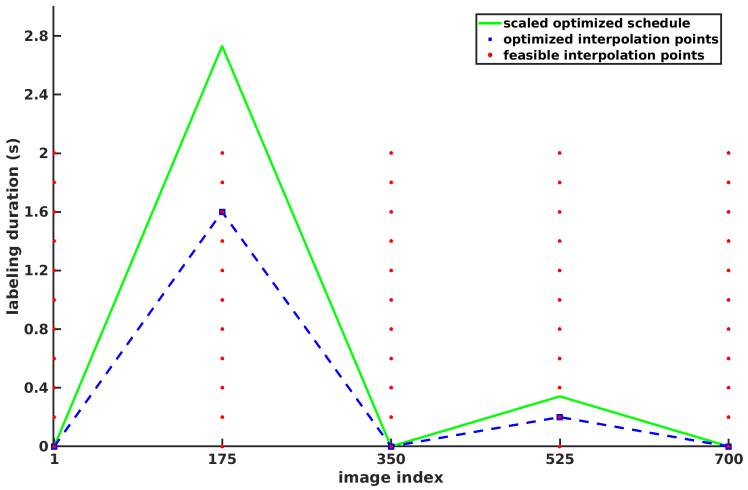


Figure 3.2: The red dots depict the feasible points for interpolating between or exploring the labeling space. The blue squares depict the label durations for the five interpolation points that, once scaled, leads to our optimized schedule. The green line depicts this scaled, best (among other candidates) schedule.

set was described using a linear interpolation between 5 points in the 'labeling space'. These points were spaced at a regular number of frame intervals (here, 175). Candidate schedules were created in a 'connect-the-dots' fashion from the resulting grid of points in the labeling space, as depicted in Fig. 3.2. The total number of feasible schedules in this case is 11^5. We chose linear interpolation as it allowed us to explore this labeling space effectively. It also allows for flexible finer sampling of the labeling space upon increasing the number of interpolation points or possible labeling durations. Albeit, this refinement would require more optimization time. Feasible schedules were scaled to be a fixed total duration (or scan duration), before the cost function was evaluated. Because there is a trade-off between sacrificing scan duration and sacrificing precision, the total duration may vary based upon the required precision. For our work, the scan duration is set to 600s. Acquiring a small number of images for this total scan duration would create labeling schedules with unreasonably long tagging times, while acquiring a large number of images increases memory and storage overhead. Under these considerations, we choose to acquire 700 images. Fig. 3.2 also depicts the optimized scan. Having obtained an optimized labeling schedule, we further minimized the predicted precision of flow estimates by trying several pseudo-random label-control-silence schedules while maintaining that the number of each are equal in our schedules. In section 3.3.1, we compare the theoretical performance of this labeling schedule to two others commonly encountered in MRF literature, with $|\boldsymbol{\Theta}|$ set to 250. While the neural network-based estimation framework described in section 3.2.4 may provide biased estimates, these CRLB predictions serve as a useful indicator for the performance of a schedule in terms of precision in estimates obtained from it. We also note that given the non-convexity of the optimization space, and our use of discrete search in solving the problem, the resulting scan cannot be dubbed 'optimal'. However, it should perform better than the other evaluated candidates. Because it is the result of our optimization, we call the output 'optimized' ASL-MRF scan.

3.2.4 Estimation with Neural Networks

We used a neural network based framework to estimate 5 hemodynamic properties of relevance in our model. Namely, these were the Perfusion f (CBF), the Bolus Arrival Time (BAT) or δ, the Cerebral Blood

Volume in artery (CBVa), Magnetization Transfer Rate (K), and the tissue relaxation time, henceforth called T_1 for simplicity. Additionally, we also estimated a field map of the Flip angles enacted by the scanner. Separate neural networks were used in the estimation of each parameter. The reason for moving away from the combined neural network framework used in our previous work [20] is to avoid the need for the relative weighting of targets during network training. Table 3.1 provides the architectural specifications of the networks used. The neural networks were chosen from a small set of candidate networks. When choosing, the performance of these was tested using three validation datasets similar to the simulated phantom described in section 3.2.6.1, but with different ground truth values. Some of the networks required fewer nodes/layers than the others. This may be because of the varying degree of non-linear dependence of the fingerprints on different parameters.

Parameter	Depth	Architecture (nodes per layer)	min value	max value
Perfusion	3	10-10-10	0 mL/100g/min	90 mL/100g/min
CBVa	3	10-10-10	0	0.015
BAT	2	10-5	0.3 s	3.0 s
MTR	4	10-10-5-5	0 s^{-1}	0.03 s^{-1}
T_1	1	20	0.33 s	3.33 s
Flip	1	20	48°	112°

Table 3.1: Description of the neural network architectures used in estimating hemodynamic parameters in our signal model, as well as the respective maximum and minimum values of the ranges used in the training data. The 'Architecture' column provides the number of nodes in every layer, separated by hyphens, starting from the input. Each node in the network learns a weight and a bias during training. The input to the networks are fingerprints generated from our designed optimized sequence, which has 700 frames.

To train our networks, we used 6×10^6 samples of synthetic fingerprints generated from the model described in section 3.2.1, with added real white Gaussian noise with standard deviation 0.01, along with the corresponding generating parameters (Fig. 3.3). The same training dataset was used across all neural networks. We used an independent uniform prior on values of each parameter for generating this data. The associated parameter ranges are also depicted in Table 3.1. We selected the ground truth parameter values for training signals from independent uniform distributions. We did this in hopes that the trained network can perform well at estimating possible anomalies in combinations of hemodynamic properties. (For example, elevated arterial transit time, but normative perfusion etc.) Each signal was normalized by the value of the first frame in the fingerprint. We applied the same process to signals obtained from the scanner, thereby ensuring consistency during testing and training. The cost function used to train the neural networks was Mean Square Error, the optimizer associated was ADAM [18], and the non-linearities were implemented as ReLU-s. Training times for the networks were roughly $15 - 20$ mins. Once trained, the network was tested on a gamut of test datasets described in section 3.3.

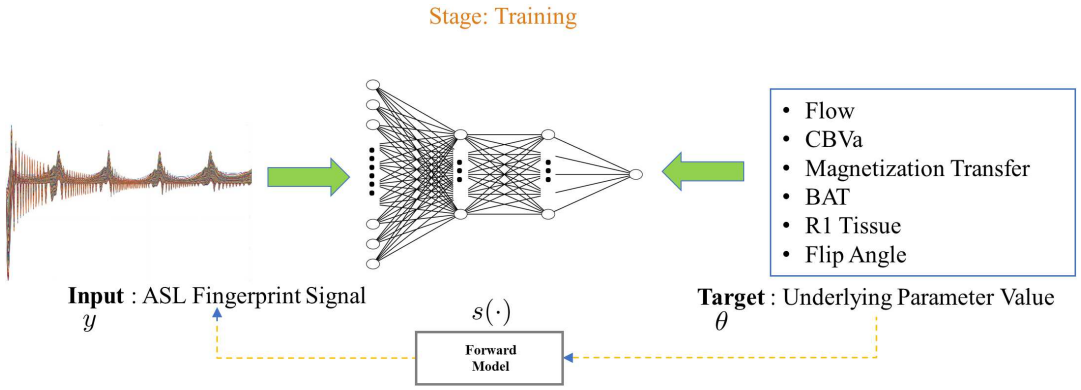


Figure 3.3: Diagram depicts a neural network regressor used as an estimator in our work, in the training stage. The targets for the training and the inputs are related through the forward model depicted in Fig. 3.1 with additive noise. Separate networks are trained for the different unknowns in the model. Once trained, the estimators can predict the generating parameters for a new fingerprint.

For training, we used simulated noisy fingerprints from the signal model instead of real (empirically obtained) fingerprints for two reasons: (i) ground truth estimates for real data are difficult to obtain. In a wide-scale multiparametric setting such as ours, they would suffer from granularity owing to the use of dictionary based methods in calculating the ground truth for training. They may also be biased because they were obtained from non-MRF techniques (ii) limited availability of real training data would pose a significant risk of overfitting neural networks (especially deeper ones).

3.2.5 Signal Preconditioning

Presence of scanner drift or cardio noise and breathing can cause severe distortions in the fingerprints from the hypothesized model [23, 15]. However, the labeling scheme modulates the perfusion information into the high frequency bands of the fingerprint signal similar to [22, 25]. This property, combined with the fact that the aforementioned nuisances generally manifest as low frequency components, motivated us to high-pass filter the fingerprints (both during training and testing) of the neural networks associated with perfusion, bolus arrival time, magnetization transfer rate and the cerebral blood volume fraction. We applied a 4-th order Butterworth filter with a cutoff at 0.05 Hz (when assuming the fingerprint was sampled at 1 Hz) for this purpose.

3.2.6 Data Collection

3.2.6.1 Simulated Anthropomorphic Pathological Phantoms

We synthesized a set of test data from hemodynamic parameter maps that closely reflected the corresponding spatial distributions in a digital phantom generated from standard gray and white matter maps (SPM12) [13]. We then introduced regions of abnormally elevated and reduced perfusion to it to quantify the performance of our methods on a range of normative and pathological parameter values. Real AWGN with standard deviation 0.01 was added to the fingerprints after generation. When estimating perfusion, CBVa, and BATs, the data was high-pass filtered as explained in section 3.2.5. Fig. 3.4 compares the predictions with the corresponding ground truths.

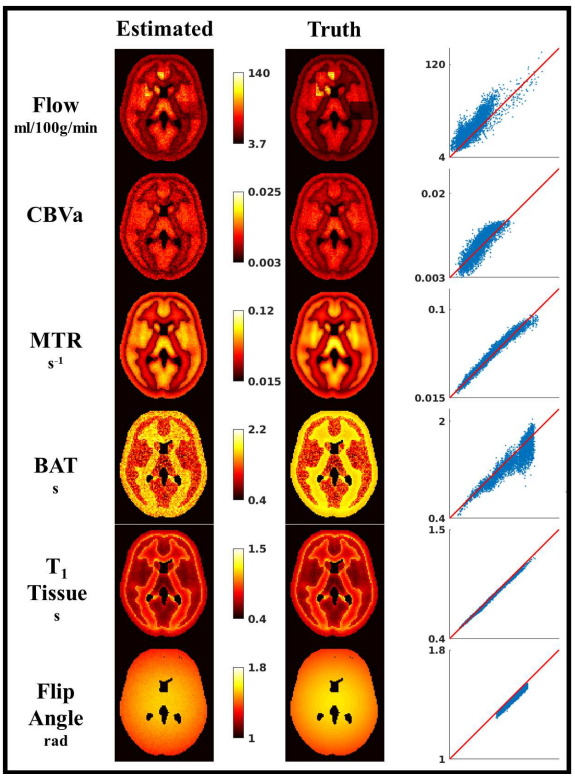


Figure 3.4: Performance of proposed neural network based estimation on the simulated dataset described in section 3.2.6.1. The first column depicts the predictions from the networks, while the second shows the corresponding ground truth parameter images. The third column are 'truth-vs-predicted' scatter plots of the former columns.

3.2.6.2 In vivo experiments

In-vivo data was acquired on a 3T General Electric MR750 scanner. The imaging parameters were: a single slice placed above the ventricles, single shot spiral readout, nominal resolution = 3.5×3.5×7 mm^3, matrix size = 64×64, bandwidth = 125 Hz and TE = 5ms, FOV = 240 mm. We tested our methods on data acquired from six human subjects. For four of these subjects, we also acquired data with two other MRF-ASL scan designs (similar to the ones used in [35, 41] and described in more detail later. See Fig. 3.5). We compared our optimized scan design to them using similar estimation techniques (same neural network architecture and training target distribution in Table 3.1). Our goal was to show the benefits of labeling sequence optimization in ASL fingerprinting. For 'random scan', we sampled the labeling durations from a uniform random distribution, while in 'decreasing affine scan', the durations decrease linearly with the image index. Both schemes were designed to be 700 frames and 600s long, and are collectively described as reference ASL-MRF scans later in this text. The metric for this comparison was the normalized standard deviation of parameter estimates obtained from the numerical CRLB evaluation described in section 3.2.3.

For all six subjects, we also performed a 409s multi-PLD ASL [40, 1] experiment with 40 PLDs, involving a single average over label-control pairs at each PLD. The post-labeling delays were chosen according to the protocol presented in [40]. No arterial suppression was used. The CBF, CBVa, BAT and T_1 maps obtained from these were compared to those from our methods. We fit the signals to a single compartment model in order to obtain CBF, BAT and T1. For fair comparison, we also fit these signals to a two-compartment model

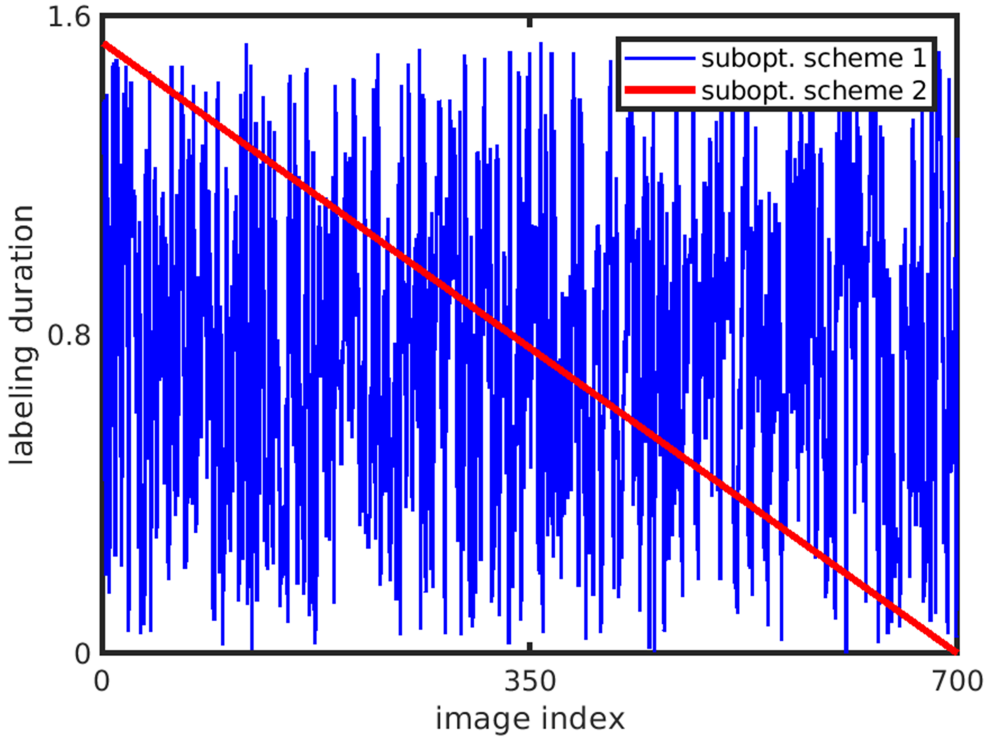


Figure 3.5: Two commonly used scan designs in MRF-ASL literature that we compare our optimized scan design to in this work. For 'random scan' (here, subopt. scheme 1), we sampled the labeling durations from a uniform random distribution, while in 'decreasing affine scan' (here, subopt. scheme 2), the durations decrease linearly with the image index.

to account for the arterial signal and also estimate CBVa, as in the proposed MRF method.

We used a two-stage estimation technique to generate quantitative parameter maps from the multi-PLD data. In the first stage, we estimated the tissue T_1 and M0 maps at every voxel by applying a least squares fit using the model in eqn. 3.1. For this stage, all other parameters in the equation were fixed to nominal, or where applicable, normative values. Next, the entire process was repeated for estimating the CBF, CBVa and BAT at each voxel (only CBF and BAT for single compartment estimation), but using M0 and T_1 values obtained in the previous stage. The MTR and flip angles were held fixed throughout.

3.3 Results

3.3.1 Optimized Scan Design

We compared the predicted performance of our optimized scan design against two reference ASL MRF scan designs described in section 3.2.6.2. Table 3.2 lists the predicted normalized standard deviation in estimates of each parameter for all three labeling schemes, and the total weighted design cost associated with each scheme. Of the 11^5 schedules evaluated, 371 were within 3% of the total cost. Of these, 150 schedules followed a similar pattern to the schedule with the minimum cost. While the other scans within the 3% margin were very different from the minimum-cost scan, these could also be grouped to be similar to $\approx 4-5$ different scan designs.

Parameter	normalized std deviation (%)		
	optimized scan	random scan	decreasing affine scan
Perfusion	46.4	51.0	46.1
CBVa	17.2	21.9	19.5
BAT	1.3	2.1	1.6
MTR	121.1	124.2	137.1
T_1	0.6	0.4	1.1
Flip	0.4	0.2	0.8
Cost	**233.4**	**250.8**	**252.3**

Table 3.2: Predicted normalized standard deviation of parameter estimates (in %) for ASL-MRF labeling schedules used in our comparisons. The last row shows the overall weighted design cost associated with each scheme based on eqn. (3.5).

3.3.2 Simulated Anthropomorphic Pathological Phantoms

Fig. 3.4 depicts the estimated maps from the Anthropomorphic Pathological phantom simulation, and the corresponding ground truth parameter images in the first two columns. From these images, we also generated 'truth vs predicted' scatter plots for each estimated parameter map to better visualize the accuracy and precision of our methods, shown in the third column of Fig. 3.4. For a more quantitative evaluation of the performance of our methods, Table 3.3 shows the correlations between the voxel values of the truth and estimated parameter maps, across multiple scan designs. The corresponding root mean squared errors (RMSEs) are provided as well.

Parameter	optimized scan		random scan		decreasing affine scan	
	Corr. (%)	RMSE	Corr. (%)	RMSE	Corr. (%)	RMSE
Perfusion (ml/100g/min)	**86.7**	**11.2**	71.9	17.3	68.2	15.48
CBVa	**86.7**	$\mathbf{1 \times 10^{-3}}$	80.0	2×10^{-3}	81.6	2.6×10^{-3}
MTR (s^{-1})	**98.4**	**0.003**	95.4	0.009	0	0.045
BAT (s)	**91.7**	**0.14**	90.0	0.18	89.1	0.19
T_1 (s)	98.6	**0.015**	**99.6**	0.016	98.0	0.029
Flip (rad)	**99.7**	0.029	98.6	**0.009**	84.7	0.062

Table 3.3: Correlation (in %) and RMSE (units in the parameter column) of each estimated parameter map with the corresponding ground truth map in the anthropomorphic digital phantom.

3.3.3 In vivo performance

It took approximately 1s to estimate each 64×64 parameter map for the MRF methods using the designated neural networks for the task on a 12GB NVIDIA Titan X Pascal GPU. The two stage fit for the multi-PLD data required approximately 1200s to estimate four 64×64 maps on a Intel Xeon E5-2650 with 40 cores. Fig. 3.6 compares the six estimated maps from a single human subject across all evaluated techniques. To gauge a sense of agreement between two-compartmental estimates from regression based ASL Fingerprinting and multi-PLD methods, Fig. 3.7 compares the mean gray matter CBF, BAT and CBVa across

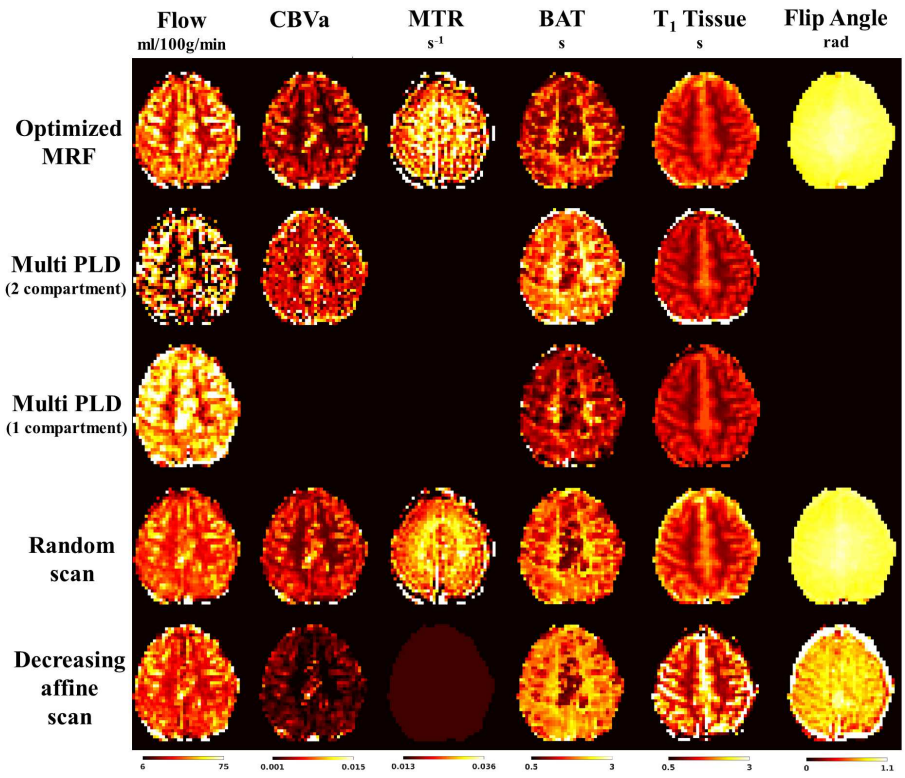


Figure 3.6: Comparison of the parameter estimates from various tested methods for a single subject. None of the methods used any spatial smoothing of the estimated maps or the ASL signal volume.

the six subjects, as well as the associated average T_1s for both gray and white matter using scatter plots. Fig. 3.8 shows a 'goodness-of-fit' comparison between the acquired signal and a 'synthetic' signal produced from the modeled equations (3.1)-(3.3) averaged over a region of interest. The synthetic signals were obtained by passing the parameters estimated from the neural networks through the ASL signal model.

3.4 Discussion

This work established a CRLB based optimization method for labeling durations for improving the information within an MRF-ASL scan, as a means to get more precise estimates from it for a fixed scan time. This enables us to 'get the most' out of available scan time, and is of particular importance because of the trade-off between total scan time and precision of estimates, regardless of the estimator. Of course, it would be possible to reduce the scanning duration at the expense of overall precision. We adopted a neural network regression based estimation framework to avoid the granularity/imprecision of dictionary-search based estimators for problems with many parameters like ASL-MRF. The methods provided estimates for six parameters in both gray and white matter regions in the brain. We validated our methods in silico using a simulated anthropomorphic phantom, and in vivo against a multi-PLD method as well as two reference ASL-MRF scans. For easy comparison with 'state-of-the-art' MRF ASL, 'decreasing affine scan' was very similar to the Perlin schedule used in [35], but comparatively yielded lower predictions of normalized standard deviation, while 'random scan' was similar to the one used in [41]. In most cases, the CRLB predictions were reflected in the performance of various methods in a relative sense. The following subsections elaborate on our observations

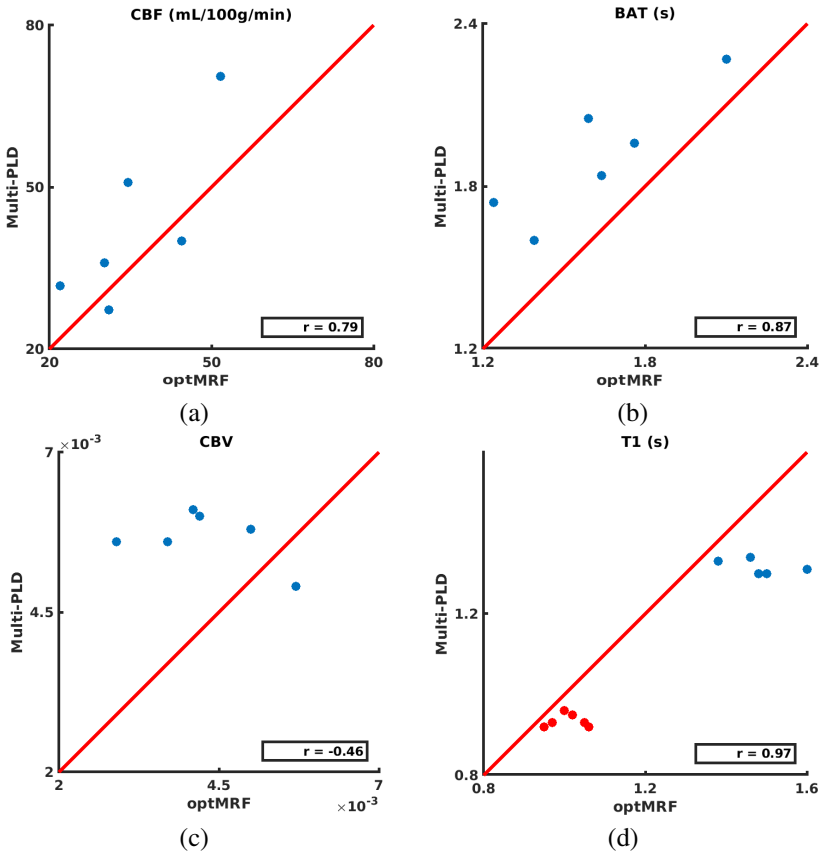


Figure 3.7: Scatter plots of slice-wide average estimates from optimized MRF vs multi-PLD of: (a) gray matter CBF, (b) gray matter BAT, (c) gray matter CBVa and (d) gray (blue dots) and white (red dots) matter T_1.

from section 3.3.

3.4.1 Optimized Scan Design

From the predicted standard deviations in Table 3.2, it is apparent that the optimized scan either at least performs comparably, or outperforms the other two at precisely estimating all relevant parameters. We note that while the improvements in the predicted standard deviations of parameter estimates using the CRLB are not astonishing, it might not be fair to translate these predicted values into indicators of performance. This is because the CRLB is evaluated at an ensemble of ground truth parameter values some of which may be very difficult to estimate. As explained in subsequent sections, the performance of our optimized scan is shown to be significantly better than reference scan designs across several metrics and datasets. In particular, the overall cost function for the optimized labeling schedule is significantly lower than that for the others. This hints its potential for improved precision at jointly estimating all the modeled dependencies in the ASL signal. The MTR parameter contributes significantly to the overall variance of estimates, but incorporating it into our model may provide additional information about tissue health and reduce bias in estimates of perfusion. We also find that while a lot of variation in the labeling durations can help in the estimation of T_1 and Flip angles, it can be detrimental when attempting to estimate Perfusion (CBF), Blood Volume Fraction (CBVa) or Arterial Transit Times (BAT). Importantly, some of the predicted normalized standard deviations

are very high even after optimization, but this is because the reported CRLBs are the average value over a very broad range of parameter combinations. Many of these yield little ASL signal under any acquisition scheme (consider a voxel with very low perfusion and long transit time); the method performs significantly better (normalized std. dev. of perfusion estimates decreases to 22.5% , while that of MTR is 13.6%) when evaluated over a smaller, more normative range. (This normative range was $\pm 2.5\%$ around ground truth perfusion 60 mL/100g/min, BAT 1.2s, CBVa 0.01, MTR $0.01 s^{-1}$ and T_1 1.4s)

3.4.2 Simulated Anthropomorphic Pathological Phantoms

Fig. 3.4 illustrates that there is good agreement between the estimated and the ground truth maps across all parameters. Additionally, our estimation is able to capture both the abnormally elevated and diminished regions of flow in the perfusion maps. This property is a consequence of: (i) optimizing the scan design over a large parameter range, (ii) training the neural network using a wide range of training parameters (Table 3.1).

Table 3.3 shows that for most estimated parameters, in comparison to the reference scans, our optimization leads to comparable or better correlation coefficients between truth and estimates. For perfusion in particular, we noted that the optimized scan yielded estimates that were significantly more aligned to the truth than the other scans. We note that 'decreasing affine scan' regressed the same value of MTR (≈ 0.015) for all inputs, thereby returning a correlation coefficient of 0%. Our conjecture is that this may be due to the neural network being unable to learn from the training data due to the insensitivity of the fingerprints to MTR. We also observed that the correlation coefficent associated with perfusion for this specific scan was lower than the others, even though its predicted normalized standard deviation was comparable to that of the optimized scan. We therefore hypothesize that predictions of low normalized standard deviation may not always translate to high correlation between truth and estimates in the case of a biased estimator. This is because variable estimation bias at different points in the parameter space may lead to low correlation, despite the precision in estimates. For each estimated parameter, we calculated the RMSE with respect to ground truth, and found that the optimized MRF-ASL scan estimates report lower values of RMSE than those from the reference MRF-ASL scans in a majority of parameters. We also noticed that some of the estimates show a consistent bias with the ground truth. This is possibly due to the difference in the distribution of ground truth parameter values in training and testing data. The training data had a uniform random prior, but the test data has the distribution of an SPM12 phantom.

3.4.3 In vivo performance

Fig. 3.6 illustrates that the performance of our designed method is relatively consistent with the predictions from the Cramer-Rao Bound (see Table 3.2). The map corresponding to the magnetization transfer rate looks the noisiest, while parameters like T_1 or BAT look much cleaner. The distinction between gray and white matter regions is also apparent across all relevant maps, even without any spatial smoothing or SNR boosting methods.

The comparisons between the optimized and reference MRF scans were also in accordance with our expectations based on Table 3.3. As evident in the depicted subject, 'decreasing affine scan' fails to estimate the MTRs, and the estimated T_1 maps have unreasonably high values in gray matter. Moreover, even the flip angle map for this scan shows significant artefactual contrast between gray and white matter, which are absent in the other MRF methods. The maps yielded by 'random scan' show agreement with the optimized scan, but we observe that the MTR maps from the former are less informative, and exhibit more artifacts.

These trends between the optimized and reference scans were noted to be consistent across all four subjects studied.

Fig. 3.6, shows that the MRF methods are able to estimate CBF, BAT and CBVa values in white matter, while the two-compartment multi-PLD method fails to do so: instead, the corresponding maps show many near-zero perfusion voxels in white matter regions, along with abnormally high transit times. The coefficient of variation (std. deviation/mean) for white matter perfusion averaged across all six subjects was 1.19 for the multi-PLD method, compared to 0.16 for optimized MRF-ASL. In gray matter, the avg. coefficient of variation was 0.54 for multi-PLD and 0.27 for optimized MRF-ASL. The obtained CBVa map from the latter method also appears extremely noisy, and is unable to pick out vasculature in the slice as effectively. We hypothesize that the poor performance in the multi-PLD method was due to the fact that its acquisition parameters were optimized for a single compartment model that does not account for CBVa, as well as the fact that no arterial suppression was used in its acquisition. We also noticed that the single compartment estimates reported higher values perfusion and lower BATs than two-compartment estimates. This is an expected observation due to the label present in the arteries seeping into the acquired tissue signal under the single compartment assumption, considering that no arterial suppression was used in our acquisition and the post-labeling delays weren't very long [1].

In Fig. 3.7 (a)-(b), we found the estimates from optimized MRF-ASL and multi-PLD to be fairly consistent, as observed from the high correlation coefficients between the two methods, for measurements of CBF as well as BAT. However, we note that the multi-PLD method consistently reports higher BATs in gray matter with a two-compartment fit. Fig. 3.7 (c) shows a similar scatter plot for CBVa, but there is little agreement between the methods. This may be due to the fact that the estimates of CBVa from the multi-PLD method have a lot of variance across the brain, for reasons explained above. The T_1 estimates for the two methods are compared in Fig. 3.7 (d), with the red dots indicating white matter T_1s, and the blue ones indicating gray matter T_1s. We see that while the optimized MRF method generally yields higher values of T_1, the measurements agree well.

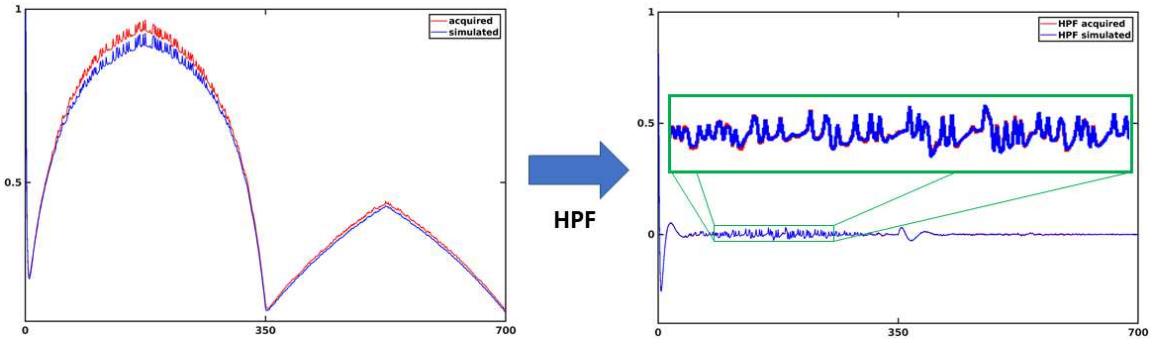


Figure 3.8: ROI-averaged comparison of acquired signals and simulated signals which were generated from our model, based on neural network estimates. The left and right panes show the signals before and after high pass filtering, respectively.

Fig. 3.8 reinforces that high pass filtering the fingerprint signals significantly improves the signal fidelity in the small, high frequency components, that correspond to manifestations of hemodynamic phenomenon. This increased agreement is because using a high pass filter removes low frequency components related to receiver drift, etc. that introduces discrepancies that were not accounted for in the model.

3.5 Conclusion

We have developed a framework for optimizing the scan design for MRF-ASL that yields more precise estimates in gray and white matter, than other scan designs for MRF-ASL used in literature, and has a lower coefficent of variation in white matter and gray matter perfusion than a reference multi-PLD method, while mainitaining agreement in gray matter perfusion and transit times. We also introduced a neural network regressor for fast precise estimates from ASL fingerprints. Combined with scan optimization, the neural networks can estimate six parameters from a single 600s ASL scan in a very short processing time and bypass quantization errors. While the work presented here is intended to serve as proof of concept, our comparisons with reference scans show that we are able to significantly improve upon the precision of the state-of-the-art in MR Fingerprinting ASL. Future work will focus on adapting the present scheme for 3D acquisition, including background suppression pulses and reducing the scan duration to make the technique practical in the clinical setting. It might be tempting to pursue avenues like the joint optimization of the labeling schedule and label-control order, however, such joint optimization of the label-control order maybe challenging due to the 'memoried' aspect of the ASL fingerprint, where the signal at each frame is affected by the signal in the acquisitions before it, as well as the fact that the label-control schedule is binary. However, one approach to tackle this problem is to use a sigmoid function to approximate the binary valued label-control order, which will enable gradient descent-based approaches to optimize the label-control schedule.

3.6 Supplementary Materials

3.6.1 Effect of parameter change on MRF-ASL signal

Here, we study how the ASL signal generated using the optimized schedule (Fig. 3.2) changes as the underlying parameter values are changed. For this purpose, use the model described in section 3.2.1, and vary each of the six parameters in our model slightly ($\pm 5\%$) while holding the others fixed. We observe, as also noted by Su et. al. in [35], that changing parameters like perfusion, BAT, CBV and MTR affect the high-frequency components of the signal, while changing the T_1 and flip angles cause the overall shape of the ASL signal (low-frequency) components to change. This is seen in Fig. 3.9.

Moreover, in Fig. 3.10 we study the change in the signal in a Multi-PLD scan due to change in the perfusion parameter, while assuming all other parameters are known, we see that even in this case we cannot ascertain if there is a significant change in the perfusion. This further emphasizes the necessity of CRLB based approaches to optimize the scan design in MRF-ASL.

3.6.2 Effect of changing labeling durations on MRF-ASL signals

Fig. 3.11 shows what happens when the labeling durations in the ASL fingerprint schedule are changed. For this purpose, we increased each labeling duration in our optimized schedule (Fig. 3.2) by 0.1 seconds. We find that increasing the labeling durations causes significant change in the low-frequency components of the signal (overall shape), while the high-frequency trends largely remain the same.

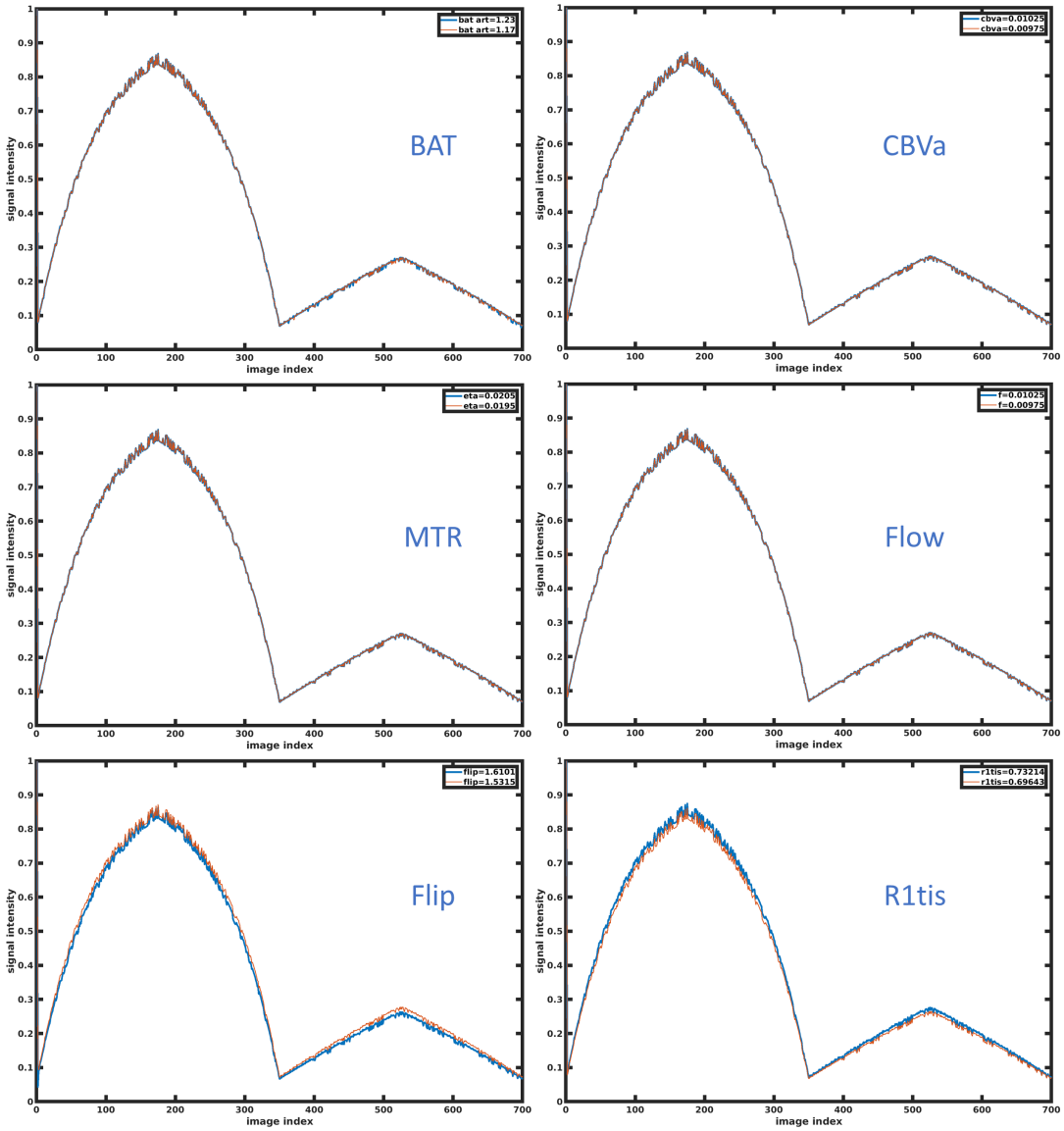


Figure 3.9: Effect of parameter changes on the ASL signal obtained using the optimized schedule in Fig. 3.2 and the model in section 3.2.1. Our observations were consistent with Su et. al.

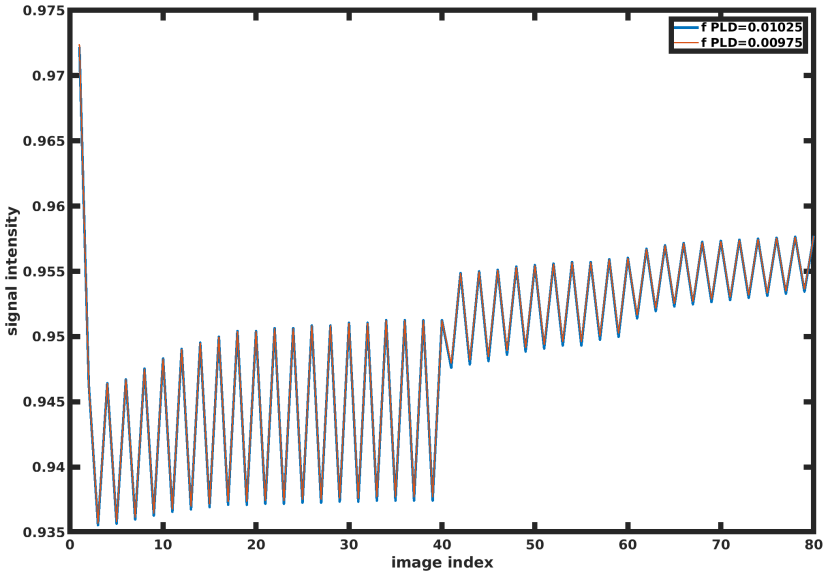


Figure 3.10: Effect of changing perfusion (f) on the ASL signal with all other parameters assumed to be known, obtained using the Multi-PLD schedule we validate our scan against in this work. Here, we see that the changes in the signal are still in the higher frequency details, and we cannot conclusively say which schedule is more sensitive to changes in perfusion. As a result, we adopt the CRLB approach to optimization.

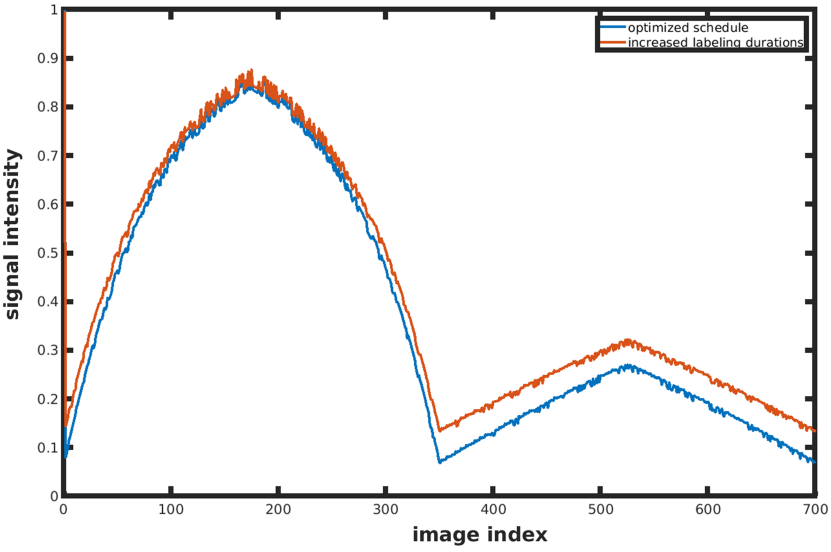


Figure 3.11: Effect of increasing the labeling durations in the optimized schedule by 0.1 seconds. The major changes are observed in the low-frequency components of the MRF signal.

CHAPTER IV

Magnetic Resonance Fingerprinting ASL using Velocity-Selective Inversion Pulses for Improved Precision in Perfusion Estimation[1]

4.1 Introduction

Chapter III explored the use of pseudo-continuous ASL pulses in combination with the MRF technique to reliably quantify perfusion and other physiological parameters in the brain. Our proof-of-concept work showed that it was possible to estimate 6 parameters, either hemodynamic properties or nuisance variables, in a single slice using a 600s scan. While it showed the importance of information theoretic optimization of ASL-MRF scan parameters like TR, the duration of scan itself may be ill-suited for clinical use considering it only provides estimates from a single slice. However, because we acquire relatively less information about the underlying parameters in a single slice when multiple slices or an entire volume is excited in the same duration, and because only about 1% of the ASL signal accounts for information regarding perfusion, trying to quantify the same factors for multiple slices (16 or 18) may pose a greater challenge. Put simply, the lack of sensitivity of the fingerprints to the underlying parameters, especially perfusion, is a major inconvenience that needs to be addressed for a clinically viable ASL-MRF scan to be implemented.

Velocity Selective Inversion (VSI) [54, 67, 72] pulses have recently gained attention in ASL perfusion imaging because of their improved sensitivity to perfusion in the resulting signal, leading to an improvement in the SNR over traditional pCASL pulses. Another potential advantage of these pulses is that they remove nuisance parameters like the Magnetization Transfer Rate from the model, and significanly reduce the variablity of other ones like BAT. While pCASL pulses invert in-flowing spins at a specific location (like a plane, as they cross it) in proximity to a target organ or tissue of interest, velocity selective labeling inverts spins regardless of their location, thereby reducing the time taken by the labeled bolus to arrive at the region of interest significantly, and yielding improved labeling efficiency.

In this work, we try to combine VSI-ASL with MRF to enable the precise estimation of hemodynamic parameters and tissue properties across multiple slices or an appropriate volume. Our preliminary experiments show that, from the perspective of Cramer-Rao Bounds, VSI pulses lead to lower predicted variances in relevant parameter estimates, compared to pCASL pulses. Initial *in-vivo* experiments with unoptimized scan designs also seem to corroborate this prediction. We also lay out, in the future work section, our plans for optimization and validation of a VSI-ASL-MRF schedule whose implementation is clinically feasible.

[1] based on preliminary work in [56]

4.2 Methods

In this section, we describe in more detail, the MR pulse sequence used in our work, both in terms of the ASL preparation as well as the image acquisition. We then describe how the ASL signal is modeled, i.e,, the contributions of various hemodynamic parameters to the signal, as well as the effect of VSIASL pulses on the observed signal. Finally, we also briefly describe our estimation framework.

4.2.1 Pulse Sequence

Unlike the more common pCASL-MRF sequences in Chapter III, in this work, the ASL preparation was done with a VSI pulse—either velocity selective (label) or non-selective (control), a variable postlabeling delay, and an optional velocity selective arterial saturation pulse for arterial suppression before acquisition. Fig. 4.1 shows a single repetition period for this method, similar to one used in prior work [67] by Hernandez et al. For example, the length of the acquisition changes according to the number of slices acquired in the volume. For the 18 slices in our work, it is 576ms. The delay between the vascular suppression pulses and acquisition is 100ms. Additionally, we also specify the duration between the end of one acquisition and the start of the next labeling as 1.5s. For acquisition, we use a 3D stack of spirals coupled with a Fast Spin Echo sequence with a echo train length of 18. The matrix size was 64×64. For fingerprinting, the TR is varied by varying the post-labeling delays from frame to frame while keeping other durations within the sequence constant.

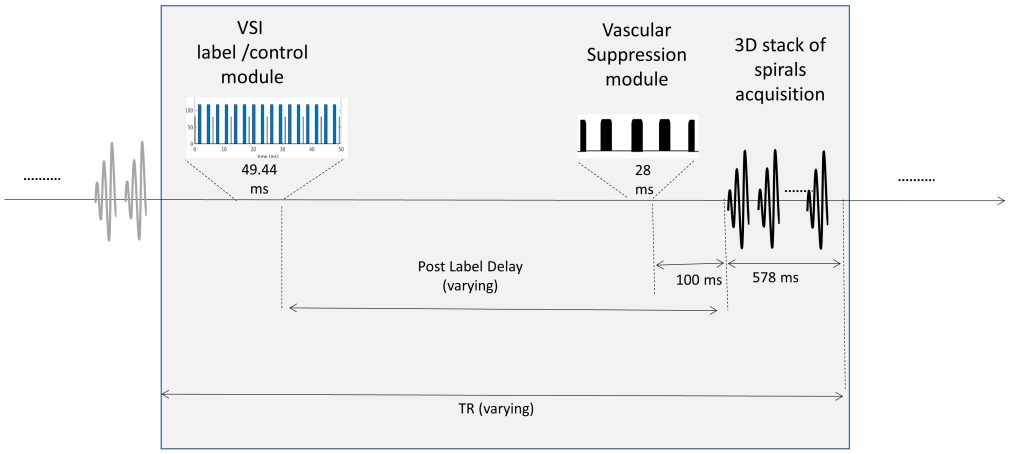


Figure 4.1: A single TR has been used to depict the details of the pulse sequence used in ASL preparation and acquisition. For the purposes of fingerprinting, the TR is varied by varying the post inversion delay across acquisitions.

The generated VSI pulses were based upon the method described by Qin et al. in [72]: the labeling module consisted of an inversion pulse that is split into multiple segments with gaps in between. These gaps are interleaved with pairs of 180° refocusing pulses each of which is surrounded by a pair of encoding gradient lobes with alternating polarity, such that a phase profile is induced in moving spins. At the end of the pulse, only stationary spins are inverted, while moving spins experience only a small tip.

In general, the objective of the control module in velocity selective ASL is to apply pulses that are non-selective, so that all spins are inverted regardless of their velocity. Though in traditional velocity selective

ASL, the control module is very similar to the labeling except the aforementioned gradient lobes are turned off, Qin et. al. propose the use of unipolar gradient lobes instead. This is to counter the overestimation of CBF due to diffusion attenuation on the static tissue at the end of the labeling pulse train. The vascular suppression module is composed of velocity selective saturation pulses that de-phase spins moving above a certain velocity, and slightly tip down other spins.

4.2.2 Modeling the VSI-ASL signal

The theorized model for our work effectively has two compartments, and is similar to the one in (3.1), with some adjustments. The arterial compartment is separated into two components, M_{art} and $M_{\text{art-ex}}$, to represent fast and slow-moving spins, respectively. While there is a continuum of spin velocities in an artery, we use these compartments to segregate the spins that are *fast enough* to not be inverted by the selective inversion pulses from the spins that move *slow enough* (especially when exchanging with tissue) to be inverted. (More about this assumption is detailed in [54]) Also, the term corresponding to MTR becomes unnecessary, as VSI-ASL signals are insensitive to it due to very little macromolecular saturation. We describe the signal in the tissue compartment as:

$$\frac{dM_{\text{tis}}(t)}{dt} = -\frac{M^0_{\text{tis}} - M_{\text{tis}}(t)}{T_{1,\text{tis}}} + f \cdot M_{\text{art-ex}}(t) - \frac{f}{\lambda} \cdot M_{\text{tis}}(t), \tag{4.1}$$

where $M_{\text{art-ex}}$ is an arterial exchange compartment that represents the blood in the artery at the exchange site, consisting of slow-moving spins. Section 3.2.1 described the other symbols used. To incorporate the effects of VSI pulses on the fast-moving spins in arterial blood, a separate compartment is included which represents the arterial blood M_{art}. Spins enter the arterial compartment at a 'fast' velocity, and slow down before exchanging into the tissue compartment after a delay (described in more detail in (4.4)). We track this behavior by modeling the fast and slow/exchanging spins separately:

$$\frac{dM_{\text{art}}(t)}{dt} = -\frac{M^0_{\text{art}} - M_{\text{art}}(t)}{T_{1,\text{art}}} \tag{4.2}$$

$$\frac{dM_{\text{art-ex}}(t)}{dt} = -\frac{M^0_{\text{art-ex}} - M_{\text{art-ex}}(t)}{T_{1,\text{art}}} \tag{4.3}$$

However, at the end of every BAT, the contents of M_{art} sub-compartment transfer over to $M_{\text{art-ex}}$ sub-compartment.

$$M_{\text{art-ex}}(t^* + \text{BAT}) = M_{\text{art}}(t^* + \text{BAT}), \tag{4.4}$$

where t^* denotes the time at which a velocity selective pulse is applied. These slow-moving spins are not affected by the velocity selective inversion pulses. In a general sense, velocity selective pulses affect the magnetization in the tissue or artery in the following manner:

$$M(t + \Delta t) = (1 - 2\alpha) \cdot M(t), \tag{4.5}$$

where M represents the magnetization (in tissue or artery), $\alpha \in [0, 1]$ is the inversion or labeling efficiency of the velocity selective pulse and Δt is the duration of the velocity selective pulse. Specifically, to model the effects of the various velocity selective pulses we introduce three new labeling efficiencies (αs) into our model, namely α_{ti}, α_{ai} and α_{ts}. α_{ti}, α_{ai} dictate the extent to which the VSI pulses invert spins in tissue and artery respectively, while α_{ts} dictates the extent of inversion in slow-moving tissue spins due to a velocity

selective saturation pulse. We express these effects mathematically as:

- **Velocity Selective Inversion (label):**

$$M_{\text{art}}(t + \Delta t) = \alpha_{\text{ai}} \cdot M_{\text{art}}(t) \tag{4.6}$$
$$M_{\text{tis}}(t + \Delta t) = (1 - 2\alpha_{\text{ti}}) \cdot M_{\text{tis}}(t) \tag{4.7}$$

- **Non-Selective Inversion (control):**

$$M_{\text{art}}(t + \Delta t) = (1 - 2\alpha_{\text{ai}}) \cdot M_{\text{art}}(t) \tag{4.8}$$
$$M_{\text{tis}}(t + \Delta t) = (1 - 2\alpha_{\text{ti}}) \cdot M_{\text{tis}}(t) \tag{4.9}$$

- **Velocity Selective Saturation (artsup):**

$$M_{\text{art}}(t + \Delta t) = 0 \tag{4.10}$$
$$M_{\text{tis}}(t + \Delta t) = (1 - 2\alpha_{\text{ts}}) \cdot M_{\text{tis}}(t) \tag{4.11}$$

Typical values for α_{ti}, α_{ai} and α_{ts} are obtained from separate Bloch simulations of the VSI pulses. Finally, the observed signal is given by:

$$s(t) = \left(\text{CBV}_{\text{a}} \cdot M_{\text{art}} + (1 - \text{CBV}_{\text{a}}) \cdot M_{\text{tis}}\right) \cdot \sin\beta, \tag{4.12}$$

where β is the flip angle.

4.2.3 Estimation via Neural Networks

We use the same regression-based estimation framework as in Chapter III, with minor changes, such as foregoing the high-pass filtering of the data and changing the architecture of the networks to have two hidden layers of 200 nodes each. The choice of 200 nodes per layer was arbitrary, and will need more investigation in the future. The new parameters α_{ti}, α_{ai} and α_{ts} were also varied in a $\pm 10\%$ range around nominal values for generating training data for preliminary experiments. Section 4.2.2 describes the model used for generating training data.

4.3 Experiments

Our preliminary experiments involve a Cramer-Rao Bound evaluation for a designed (but not optimized) VSI-MRF schedule, depicted in Fig. 4.2, and trying to estimate parameters from *in-vivo* data acquired using this schedule. This schedule determines the post-labeling delays used in our scan, which in turn determines how the TRs change across acquisitions. Section 4.2.1 describes the pulse sequence used in each acquisition.

Data for our *in-vivo* experiments was acquired on two healthy subjects using the pulse sequence described in 4.2.1.

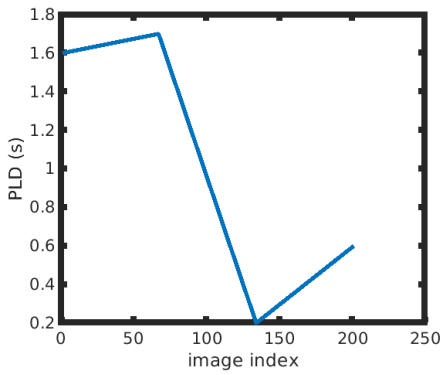


Figure 4.2: The varying post-labeling delays or PLDs (in seconds) across TRs for the schedule we use in our preliminary experiments is depicted above. The delays, however, have not been optimized.

4.3.1 Cramer-Rao Bound Evaluation

Due to our use of the pulse sequence in Fig. 4.1, which includes an arterial suppression pulse to crush the transverse magnetization in the arterial compartment, it becomes impossible to estimate CBV_a and α_{ai}. This is because the observed signal contains no information about either parameter. Hence, we remove these parameters from our Cramer-Rao Bound evaluation, which is done in a manner similar to the one described in section 3.2.3. We also hold the flip angle fixed to 90°. The results of our evaluation of the predicted normalized standard deviation of the parameter estimates from the scan are as follows:

Parameter	Normalized Std. Deviation (%)
Perfusion	41.4
BAT	27.0
T_1	1.07
α_{ti}	0.32
α_{ts}	1.55

Table 4.1: Predicted normalized standard deviation of parameter estimates (in %) for the VSIASL-MRF labeling schedule used in our experiments.

4.3.2 In-vivo Estimation

We acquired data from two healthy subjects using the delays depicted in Fig. 4.2. Parameter estimates for perfusion, BAT, and T_1 for one of these have been depicted in Fig. 4.3. We found the range of values of corresponding estimates for both subjects to be consistent.

We employed residue based masking of the parameter maps, where we used the voxel-wise parameter estimates from maps and the forward model in section 4.2.2 to generate synthetic signals which can be used calculate the residuals (norm difference) with respect to the corresponding signals obtained from the scanner. Other parameters that are not estimated are fixed to literature values. This allows us to place a quantitative measure of confidence in our fits and thus mask out regions in the estimated maps where the residue is too high. Fig. 4.4 shows the residues obtained for the slices depicted in Fig. 4.3

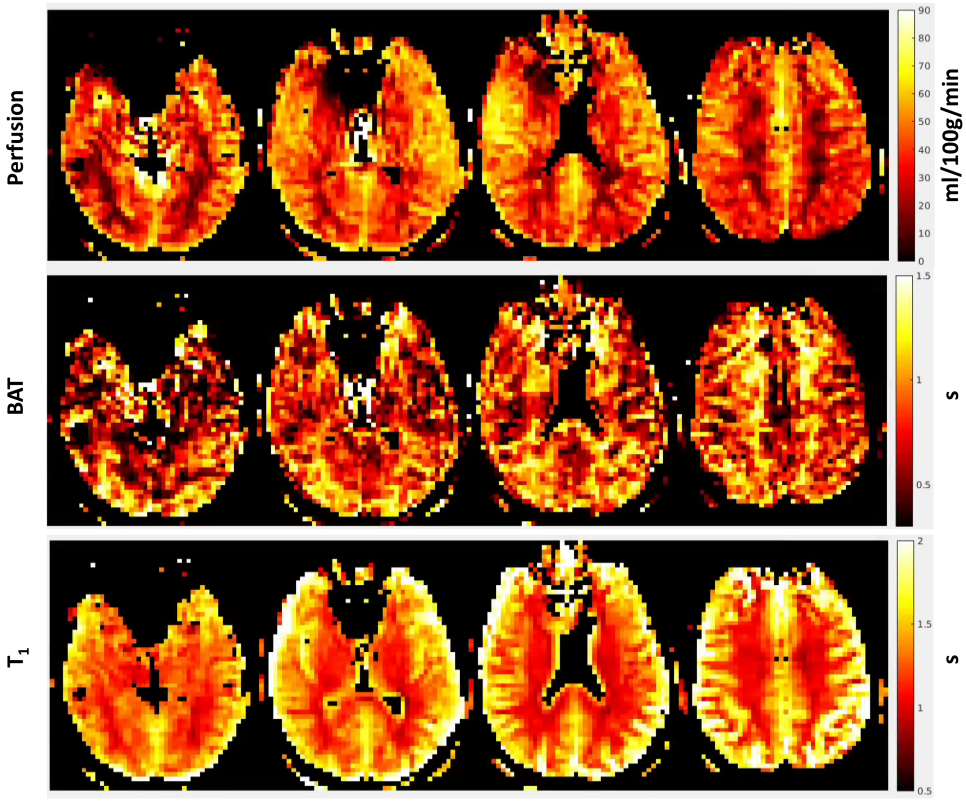


Figure 4.3: Estimates for Perfusion, BAT and tissue T1 obtained from a healthy human subject using our combined VSI and MRF-ASL framework (4 out of 18 slices have been shown).

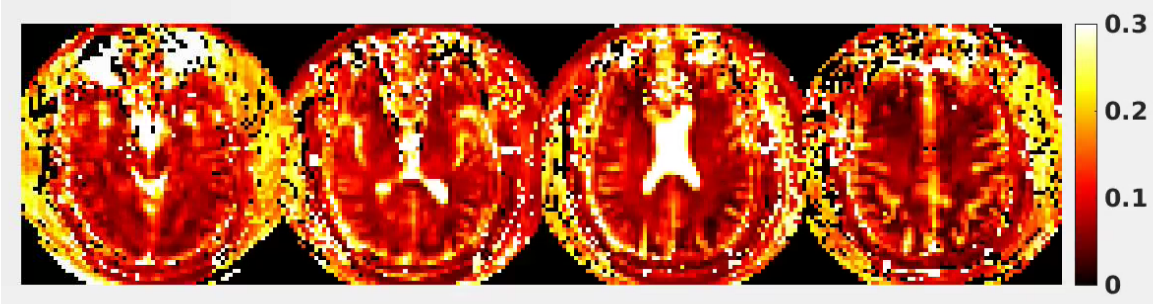


Figure 4.4: Residues maps for the slices in Fig. 4.3. These are obtained by calculating the norm difference between signals generated from the ASL signal model using obtained estimates for perfusion, BAT and T_1, and the corresponding acquired scanner signal for a specific voxel.

4.4 Discussion

From Table 4.1, it is evident that even an unoptimized MRF-ASL scan design using velocity selective ASL preparation pulses provides comparable predictions of normalized standard deviation in parameter estimates to the optimized pCASL-based MRF-ASL scan design in Fig. 3.2 (see Table 3.2). This is indicative of the potential of an optimized VSI-based MRF-ASL technique in precisely quantifying perfusion.

Furthermore, our preliminary *in-vivo* results are also encouraging, as they highlight the contrast between gray and white matter perfusion, BATs and T_1 estimates very well, and the range of values for both regions in all the estimated parameters seem reasonable. The residue maps (Fig. 4.4) seem to be high in regions of cerebro-spinal fluid (CSF). Since estimates from CSF are not of value in our applications, we have masked out these regions in Fig. 4.3.

4.5 Future Work

In our scan design for preliminary experiments, due to the presence of an arterial saturation pulse in every TR, all information about the $\mathrm{CBV_a}$ and α_{ai} was lost. Although the latter is of no clinical significance, the former is of critical importance in obtaining information about the vasculature of the tissue or organ of importance. So to be able to estimate CBV_a along with perfusion, BAT and T_1 from the same scan, we plan to design an new strategy that disables the arterial saturation pulses at certain TRs. Of course, since this design will result in losing some information about perfusion in the fingerprints, we also plan to choose an acceptable fraction of acquisitions where the saturation pulses are turned off, based on Cramer-Rao Bound evaluations of our model in section 4.2.2. We will then proceed to optimize for a suitable scan design using the same technique described in section 3.2.3. Upon fixing an optimized schedule for the post-labeling delays, we will optimize the positions of velocity selective and non-selective inversion pulses, as well as the turned-off arterial saturation pulses in the scanning schedule.

Having obtained a scan design that is optimized in terms of the delays, as well as the label/control and saturation pulses, we plan to perform *in-silico* validation tests on anthropomorphic digital phantoms with planted anomalies to gauge the performance of our method using neural-network based estimation. The next step is to acquire data using the optimized scan from healthy subjects and validate our results against other ASL techniques like multi-PLD VSIASL, and standard fixed delay pCASL. In the future, we will also acquire an additional anatomical scan in the same ASL protocol to use as additional data to aid our masking.

Finally, we plan to employ our designed and validated scan in a Precision Health study at the University of Michigan that involves studying the quantitative hemodynamics of patients with Alzheimer's disease. The mechanisms of such neuro-degenerative diseases have yet to be unraveled, and there are strong indications, based on studies like [82, 83, 84], that perfusion and other related physiological properties may serve as important biomarkers for studying them.

CHAPTER V

Combining Supervised and semi-Blind Residual Dictionary (Super-BReD) Learning for Inverse Problems[1]

5.1 Introduction

As mentioned in Chapter I, one of the major aims of this thesis is to develop algorithms to aid the reconstruction of MR images from limited measurements. Such algorithms typically involve solving under-determined linear inverse problems of some form. This chapter focuses on developing techniques that are useful in solving these problems. Some of the same techniques can also be applied to ASL image sequences to denoise them.

In solving under-determined inverse problems, regularization by expressing signals or patches in an image as a sparse linear combination of dictionary atoms has proven to be a powerful technique [46, 48]. Aside from its use in denoising or inpainting, dictionary-based regularization has also been used to model problems such as compressed sensing MRI reconstruction [95, 74]. In such applications, data-driven or adaptive dictionaries have often been of more benefit than fixed dictionaries. Using adaptive modeling involves learning sparsifying dictionaries from either uncorrupted or clean data and then applying these to corrupted data (supervised learning), or learning the dictionaries from the corrupted data itself without assuming knowledge of any other data (blind learning). In image reconstruction specifically, blind dictionary learning therefore involves leveraging information or patterns about the image being reconstructed present in its own measurements, whereas supervised dictionary learning involves exploiting information present in other 'clean' images that are similar to the image being reconstructed.

In this work, we look into developing a framework for combining these two methods of dictionary learning to represent data jointly. The supervised dictionary is learned from clean or uncorrupted image patches that are 'tailored' to the patches of the image being reconstructed. This 'tailoring' is achieved by block matching the patches of the image being reconstructed to the patches in the clean dataset. The unsupervised dictionary and its corresponding sparse contribution in representing the data is learnt from the residue of the representation of the supervised dictionary. We posit that such a model will be beneficial in settings where limited uncorrupted training data is available, and that the inclusion of tailored supervised learning will compensate for the lack of available training data in cases with fewer acquired measurements.

We experiment on denoising natural images with our algorithm, as well as single-coil undersampled MR image reconstruction. Preliminary results show promising performances in both applications. Furthermore, we also propose a deep generalization to the Super-BReD Learning algorithm that incorporates learning from

[1] based on work in [81]

pairwise training data when a sufficient/sizable dataset is available. Finally, we also propose a Super-BReD Learning based algorithm for denoising ASL images, with limited training data (ASL images with multiple averages), that allows for acquisition of fewer ASL frames.

5.2 Theory

In this section, we describe the general principle of dictionary learning, and briefly discuss how the SOUP-DIL [119] algorithm solves the dictionary learning problem. We then proceed to formulate our proposed approach for combining blind and supervised dictionary learning based on this framework.

5.2.1 Dictionary Learning

The synthesis dictionary model approximates a signal $\boldsymbol{s} \in \mathbb{C}^n$ as a linear combination of atoms (or columns) of a dictionary $\boldsymbol{D} \in \mathbb{C}^{n \times J}$, i.e., $\boldsymbol{s} \approx \boldsymbol{D}\boldsymbol{z}$, where $\boldsymbol{z} \in \mathbb{C}^J$ is sparse, or $\|\boldsymbol{z}\|_0 \leq k$, and k is the sparsity level. A dictionary is overcomplete when $J > n$, and such dictionaries are popular because owing to their richness, they can provide sparser representations of data. Given a fixed dictionary $\boldsymbol{D}$, (often an overcomplete DCT matrix is used as the fixed dictionary) finding a set of sparse coefficients $\boldsymbol{z}$ to represent $\boldsymbol{s}$ is called *sparse coding*, and is expressed mathematically as:

$$\arg\min_{\boldsymbol{z}} \|\boldsymbol{s} - \boldsymbol{D}\boldsymbol{z}\|_2^2 \qquad \text{s.t.} \qquad \|\boldsymbol{z}\|_0 \leq k. \tag{5.1}$$

The above problem is NP-hard, and while there are a number of algorithms that attempt to provide a solution to this problem [65, 70, 71], such methods are often computationally expensive, and any guarantees on the convergence of the solution are often based upon stringent conditions.

Often, the dictionary itself is learnt in a data-adaptive fashion from a collection of signals $\{\boldsymbol{s}_i\}_{i=1}^N$, which are the columns of a matrix $\boldsymbol{Y} \in \mathbb{C}^{n \times N}$, and the combined dictionary learning and sparse coding problem can be expressed as:

$$(P0) \qquad \arg\min_{\boldsymbol{D},\boldsymbol{Z}} \|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{Z}\|_F^2 \qquad \text{s.t.} \qquad \|\boldsymbol{z}_i\|_0 \leq k \; \forall \, i, \; \|\boldsymbol{d}_j\| = 1 \; \forall \, j, \tag{5.2}$$

where $\boldsymbol{z}_i$ is the ith column of $\boldsymbol{Z}$, and $\boldsymbol{d}_j$ is the jth column or 'atom' of $\boldsymbol{D}$. The atoms are constrained to be unit norm to avoid scaling ambiguities in the iterates. The above formulation is called the sparsity constrained dictionary learning formulation, and there exist several approaches to solving it [46, 75, 77, 79]. An alternative formulation, that relaxes the sparsity constraint to a sparsity penalty is given as:

$$\arg\min_{\boldsymbol{D},\boldsymbol{Z}} \|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{Z}\|_F^2 + \lambda^2 \|\boldsymbol{Z}\|_0, \qquad \text{s.t.} \qquad \|\boldsymbol{d}_j\| = 1 \; \forall \, j, \tag{5.3}$$

where $\lambda > 0$ is a parameter that regulates the sparsity. Algorithms for solving both (5.2) and (5.3) usually involve alternating between a dictionary update and a sparse code update. For our work, we will focus on one such algorithm, SOUP-DIL [74], which utilizes highly-efficient block coordinate descent to update sequentially each dictionary atom, and the corresponding sparse code rows.

5.2.2 Sum of Outer Products Dictionary Learning (SOUP-DIL)

The SOUP-DIL algorithm takes a computationally efficient block-coordinate descent (BCD) approach to solving the sparsity penalized dictionary learning problem (5.3) by expressing the representation term $\boldsymbol{DZ}$ as a sum of outer products of dictionary atoms and corresponding sparse code activations, i.e., $\boldsymbol{DZ} = \sum_{j=1}^{J} \boldsymbol{d}_j \boldsymbol{c}_j^H$ (where $\boldsymbol{c}_j$ is the jth column of $\boldsymbol{C} = \boldsymbol{Z}^H$), and then updating the $\boldsymbol{d}_j$s and $\boldsymbol{c}_j$s alternatingly. These alternating updates allow for every dictionary atom and corresponding sparse activation to residually learn to (learning by minimizing the residuals) express patterns/features in the image patches that were not captured by previous updates (or atoms and sparse representations). Essentially, the dictionary learning problem $(P0)$ is presented as:

$$(P1) \qquad \underset{\boldsymbol{d}_j \boldsymbol{c}_j}{\arg\min} \|\boldsymbol{Y} - \sum_{j=1}^{J} \boldsymbol{d}_j \boldsymbol{c}_j^H\|_F^2 + \lambda^2 \Sigma_{j=1}^{J} \|\boldsymbol{c}_j\|_0 \qquad \text{s.t.} \qquad \|\boldsymbol{d}_j\|_2^2 = 1,\ \|\boldsymbol{c}_j\|_\infty \leq L \ \forall\ j. \tag{5.4}$$

The ℓ_∞ constraint on the columns of $\boldsymbol{C}$ prevents problems due to the non-coercive nature of the objective in $(P1)$, and ensures convergence [74], but is typically unnecessary in practice. Defining $\boldsymbol{E}_j \triangleq \boldsymbol{Y} - \sum_{k \neq j} \boldsymbol{d}_k \boldsymbol{c}_k^H$, a solution to the above problem is to update the jth dictionary atom, followed by the jth sparse code row as in [74]:

1. **Sparse Code Update Step:**

$$\hat{\boldsymbol{c}}_j = \min\left(|H_\lambda(\boldsymbol{E}_j^H \boldsymbol{d}_j)|, L\boldsymbol{1}_N\right) \odot e^{j \angle \boldsymbol{E}_j^H \boldsymbol{d}_j}, \tag{5.5}$$

   where

$$H_\lambda(x) = \begin{cases} x, & \text{if } |x| \geq \lambda \\ 0, & \text{if } |x| < \lambda, \end{cases}$$

   is the hard-thresholding operator.

2. **Dictionary Update Step:**

$$\hat{\boldsymbol{d}}_j = \begin{cases} \frac{\boldsymbol{E}_j \boldsymbol{c}_j}{\|\boldsymbol{E}_j \boldsymbol{c}_j\|}, & \text{if } \boldsymbol{c}_j \neq 0 \\ \boldsymbol{v}_1, & \text{if } \boldsymbol{c}_j = 0, \end{cases} \tag{5.6}$$

   where $\boldsymbol{v}_1$ is any unit norm vector.

5.2.3 Super-BReD Learning

To incorporate both supervised and blind learning components in our algorithm, we use two separate dictionaries and corresponding sparse codes to represent data in tandem. In an inverse problem setting, given measurements $\boldsymbol{y} \in \mathbb{C}^p$, $\hat{\boldsymbol{x}} \in \mathbb{C}^q$ an image being reconstructed from the $\boldsymbol{y}$ ($\boldsymbol{x}$ being its current estimate), and $\boldsymbol{T} \in \mathbb{C}^{n \times N_1}$ a collection of N_1 vectorized $\sqrt{n} \times \sqrt{n}$ patches from clean images, every outer-iteration of our proposed algorithm can be decomposed into the following three stages:

1. **Block Matching:** For each patch in $\boldsymbol{x}$, we find the closest patch in the training dataset $\boldsymbol{T}$, to create our 'tailored' supervised training data $\boldsymbol{S}$. We formulate $\boldsymbol{S}$ as:

$$\boldsymbol{S} = \Big[\boldsymbol{s}_1, ..., \boldsymbol{s}_j, ..., \boldsymbol{s}_{N_2}\Big], \tag{5.7}$$

where

$$\boldsymbol{s}_j = \tau(\boldsymbol{T}, \mathcal{P}_j\hat{\boldsymbol{x}}) = \boldsymbol{T}_{\hat{i}(j)}. \qquad \hat{i}(j) = \underset{i\in\{1,\ldots,N_1\}}{\arg\min} \|H_{\lambda_0}\Big(\boldsymbol{W}\big(\boldsymbol{T}_i - \mathcal{P}_j\hat{x}\big)\Big)\|_2^2,$$

and $\mathcal{P}_j$ extracts the jth overlapping $\sqrt{n}\times\sqrt{n}$ patch in the image as a vector. $\boldsymbol{W}\in\mathbb{C}^{n\times n}$ is a DCT matrix. In effect, the block-matching is done in a transform domain after thresholding to mitigate noise effects.

2. **Supervised Dictionary Learning:** The supervised dictionary, $\hat{\boldsymbol{D}}_1$ is learned from the tailored patches $\boldsymbol{S}$ as

$$\hat{\boldsymbol{D}}_1 = \underset{\boldsymbol{D}_1}{\arg\min}\min_{\boldsymbol{Z}_3} \|\boldsymbol{S} - \boldsymbol{D}_1\boldsymbol{Z}_3\|_F^2 + \lambda_3^2\|\boldsymbol{Z}_3\|_0 \qquad \text{s.t.} \qquad \|(\boldsymbol{d}_1)_j\|_2 = 1\ \forall\ j. \tag{5.8}$$

The SOUP-DIL algorithm described in section 5.2.2 is used to solve for $\hat{\boldsymbol{D}}_1\in\mathbb{C}^{n\times J_1}$, where J_1 is the number of supervised dictionary atoms, chosen experimentally, and $\boldsymbol{Z}_3\in\mathbb{C}^{J_1\times N_2}$ is the sparse code matrix corresponding to $\boldsymbol{D}_1$ in this problem.

3. **Blind Dictionary Learning and Image Update:** The blind dictionary is learned from the residue of the supervised dictionary representation of the patches in the reconstructed image $\boldsymbol{x}$. The blind dictionary representation of the residue is then used in tandem with the supervised dictionary representation to regularize the image update step. The overall optimization problem is expressed as:

$$\begin{aligned}\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\arg\min}\min_{\boldsymbol{D}_2,\boldsymbol{Z}_2,\boldsymbol{Z}_1} \nu\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \|\mathcal{P}\boldsymbol{x} - \hat{\boldsymbol{D}}_1\boldsymbol{Z}_1 - \boldsymbol{D}_2\boldsymbol{Z}_2\|_F^2 + \lambda_1^2\|\boldsymbol{Z}_1\|_0 + \lambda_2^2\|\boldsymbol{Z}_2\|_0 \\ \text{s.t.} \qquad \|(\boldsymbol{d}_2)_j\|_2 = 1\ \forall\ j,\end{aligned} \tag{5.9}$$

where, $\mathcal{P}$ extracts all overlapping $\sqrt{n}\times\sqrt{n}$ patches of the image x and places them as columns of a $n\times N_2$ matrix, $\boldsymbol{D}_2\in\mathbb{C}^{n\times J_2}$ is the blind dictionary with J_2 columns, and $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are the sparse code matrices corresponding to $\hat{\boldsymbol{D}}_1$ and $\boldsymbol{D}_2$. Now, focusing on the inner minimization problem, where we alternate between updating $\boldsymbol{Z}_1$, $\boldsymbol{D}_2$ and $\boldsymbol{Z}_2$, we have:

$$\begin{aligned}\min_{\boldsymbol{D}_2,\boldsymbol{Z}_2,\boldsymbol{Z}_1} \|\mathcal{P}\boldsymbol{x} - \hat{\boldsymbol{D}}_1\boldsymbol{Z}_1 - \boldsymbol{D}_2\boldsymbol{Z}_2\|_F^2 + \lambda_1^2\|\boldsymbol{Z}_1\|_0 + \lambda_2^2\|\boldsymbol{Z}_2\|_0 \\ \text{s.t.} \qquad \|(\boldsymbol{d}_2)_j\|_2 = 1\ \forall\ j.\end{aligned} \tag{5.10}$$

Denoting $\boldsymbol{R}_2^k \triangleq \mathcal{P}\boldsymbol{x} - \boldsymbol{D}_2^k\boldsymbol{Z}_2^k$ and $\boldsymbol{R}_1^k \triangleq \mathcal{P}\boldsymbol{x} - \hat{\boldsymbol{D}}_1\boldsymbol{Z}_1^k$, we have

$$\boldsymbol{Z}_1^{k+1} = \underset{\boldsymbol{Z}_1}{\arg\min} \|\boldsymbol{R}_2^k - \hat{\boldsymbol{D}}_1\boldsymbol{Z}_1^k\|_F^2 + \lambda_1^2\|\boldsymbol{Z}_1\|_0 \tag{5.11}$$

$$\boldsymbol{D}_2^{k+1}, \boldsymbol{Z}_2^{k+1} = \underset{\boldsymbol{D}_2,\boldsymbol{Z}_2}{\arg\min} \|\boldsymbol{R}_1^{k+1} - \boldsymbol{D}_2\boldsymbol{Z}_2\|_F^2 + \lambda_2^2\|\boldsymbol{Z}_2\|_0 \qquad \text{s.t.} \qquad \|(\boldsymbol{d}_2)_j\|_2 = 1\ \forall\ j, \tag{5.12}$$

where k denotes the iteration number for the inner minimization problem. Thus, each of the updates for $\boldsymbol{Z}_1, \boldsymbol{D}_2$ and $\boldsymbol{Z}_2$ is solvable using the SOUP-DIL algorithm. The actual form of the image update varies based on the application due to the variations in the system matrix $\boldsymbol{A}$ associated. In our preliminary experiments, we focus on two applications: denoising and single-coil MRI reconstruction.

5.3 Experiments

5.3.1 Denoising

For a denoising setup, the forward system operator $\boldsymbol{A}$ is the identity matrix $\boldsymbol{I}$. The image update equation thus becomes:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \nu\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \sum_{j=1}^{N_2} \|\mathcal{P}_j\boldsymbol{x} - \hat{\boldsymbol{D}}_1\hat{\boldsymbol{Z}}_1[:,j] - \hat{\boldsymbol{D}}_2\hat{\boldsymbol{Z}}_2[:,j]\|_2^2, \tag{5.13}$$

where $\hat{\boldsymbol{Z}}_1[:,j]$ and $\hat{\boldsymbol{Z}}_2[:,j]$ are the jth columns of $\hat{\boldsymbol{Z}}_1$ and $\hat{\boldsymbol{Z}}_2$ respectively. The solution to the above optimization problem involves adding the denoised representation for the image patches $\hat{\boldsymbol{D}}_1\hat{\boldsymbol{Z}}_1[:,j]+\hat{\boldsymbol{D}}_2\hat{\boldsymbol{Z}}_1[:,j]_j$ back to their respective positions in the 2D image, and averaging between the resulting image and the original noisy one.

Our denoising dataset consisted of four natural images: *'fabric'*, *'hill'*, *'boat'*, and *'barbara'*. In each test case, noise is added (at $\sigma = 30$ and $\sigma = 60$) to one of the images, which is to be denoised, while patches from the other images serve as our 'clean' training dataset $\boldsymbol{T}$. We compared the performance of our algorithm against denoising using blind SOUP-DIL.

5.3.2 Single-Coil MR Image Reconstruction

For single-coil MRI, the system operator $\boldsymbol{A}$ is the undersampled Fourier encoding matrix $\boldsymbol{F}_u$, and the image update in (5.10) takes the form

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \nu\|\boldsymbol{F}_u\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \sum_{j=1}^{N_2} \|\mathcal{P}_j\boldsymbol{x} - \hat{\boldsymbol{D}}_1\hat{\boldsymbol{Z}}_1[:,j] - \hat{\boldsymbol{D}}_2\hat{\boldsymbol{Z}}_2[:,j]\|_2^2, \tag{5.14}$$

whose solution satisfies the normal equation

$$\sum_{j=1}^{N_2} \big(\mathcal{P}_j^H\mathcal{P}_j + \nu\boldsymbol{F}_u^H\boldsymbol{F}_u\big)\boldsymbol{x} = \sum_{j=1}^{N_2} \mathcal{P}_j^H\big(\hat{\boldsymbol{D}}_1\hat{\boldsymbol{Z}}_1[:,j] + \hat{\boldsymbol{D}}_2\hat{\boldsymbol{Z}}_2[:,j]\big) + \nu\boldsymbol{F}_u^H\boldsymbol{y}. \tag{5.15}$$

Since, we used overlapping patches with a patch-stride of 1, $\sum_{j=1}^{N_2}\mathcal{P}_j^H\mathcal{P}_j = n\boldsymbol{I}$, and defining $\boldsymbol{\Gamma}(k_x,k_y) \triangleq \boldsymbol{F}\Sigma_{i=1}^{N_2}\mathcal{P}_j^H\big(\hat{\boldsymbol{D}}_1(\hat{\boldsymbol{z}}_1)_j + \hat{\boldsymbol{D}}_2(\hat{\boldsymbol{z}}_2)_j\big)$ and $\boldsymbol{\Gamma}_0(k_x,k_y) \triangleq \boldsymbol{F}\boldsymbol{F}_u^H\boldsymbol{y}$, we have:

$$\boldsymbol{F}\boldsymbol{x}(k_1,k_2) = \begin{cases} \frac{\boldsymbol{\Gamma}(k_i,k_2)}{n}, & \text{if } (k_1,k_2) \notin \Omega \\ \frac{\boldsymbol{\Gamma}(k_i,k_2)+\nu\boldsymbol{\Gamma}_0(k_i,k_2)}{n+\nu}, & \text{if } (k_1,k_2) \in \Omega, \end{cases} \tag{5.16}$$

where (k_1,k_2) indexes *k-space* locations and Ω represents the subset of *k-space* locations that were sampled. The updated image is obtained by performing an inverse FFT on $\boldsymbol{F}\boldsymbol{x}$ from (5.16).

Our training dataset for single-coil MRI experiments consisted of patches from three complex-valued fully sampled images depicted in Fig. 5.1.

The test images consisted of two brain and two knee images at $4\times$ and $5\times$ acceleration, one of each is depicted in Fig. 5.3. For a comparison of performance, we used reconstructions for the same test data obtained from blind SOUP-DIL algorithm.

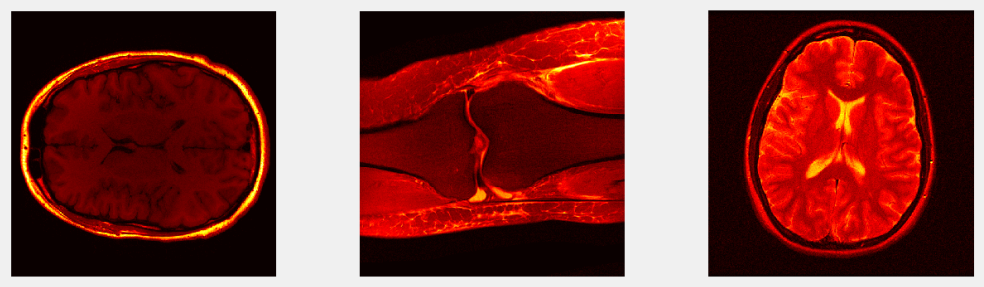


Figure 5.1: The training dataset for single-coil MRI consisted of patches from two brain and one knee images. The first image was rotated to include some measure of invariance to the orientation of features in patches during the block matching. All of the images were complex-valued. The images were obtained courtesy Prof Miki Lustig at UC Berkeley, the FastMRI knee dataset and Cameron Blocker at the University of Michigan.

PSNR (dB)								
Image	Noise Std. Dev. 30				Noise Std. Dev. 60			
	Noisy	Blind SOUP	Super-BReD	Δ	Noisy	Blind SOUP	Super-BReD	Δ
fabric	18.6	27.4	**27.6**	0.2	12.6	23.5	**23.9**	0.4
boat	18.6	28.4	**28.5**	0.1	12.6	24.9	**25.5**	0.6
barbara	18.6	**28.5**	**28.5**	0.0	12.6	24	**24.9**	0.9
hill	18.6	28.4	**28.7**	0.3	12.6	25.6	**26.2**	0.6

Table 5.1: Table depicting denoising performances of SOUP-DIL and Super-BReD Learning in PSNR (dB) at various noise standard deviations. $\sigma = 30$ corresponds to the moderate noise regime, while $\sigma = 60$ corresponds to high noise regime. The 'Noisy' column lists the PSNRs of the noisy images, the 'Blind SOUP' column lists PSNRs for the blind denoised images, while the 'Super-BReD' column depicts the PSNR performance of our proposed method. Δ denotes the improvement over blind SOUP-DIL that Super-BReD Learning produces.

5.4 Results

5.4.1 Denoising

Table 5.1 depicts the results for denoising. We observe that, while there are improvements of around 0.0-0.3 dB in the denoising performance of Super-BReD Learning and blind SOUP-DIL in the moderate noise regime of $\sigma = 30$, the real gains of incorporating supervised learning is observed in the high corruption regime of $\sigma = 60$. In the latter case, improvements obtained from using Super-BReD Learning are in the range of 0.4-0.9 dB. A possible reason for this improvement is that the supervised dictionary compensates for the incapability of the blind dictionary to learn from very corrupt data. Fig. 5.2 depicts an instance of the comparative results of denoising. It can be observed that despite the very high noise level, more details from the ground truth are preserved in the Super-BReD denoised image compared to the blind SOUP-DIL denoised image, and there is a visible improvement in the former compared to the latter. Overall, the denoised images from both methods appear very corrupted in the details. However, here we focus on the relative improvements brought about by our methods in preliminary denoising experiments, as our interest is in the performance of Super-BReD learning in MRI applications.

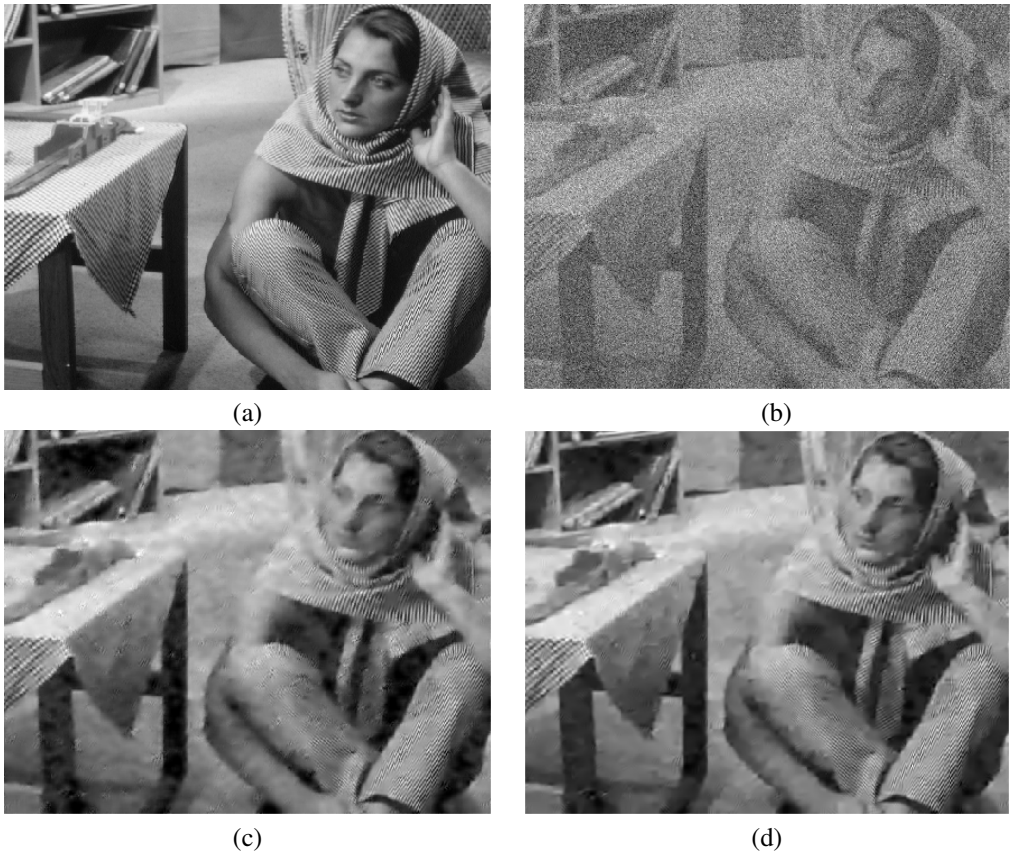


Figure 5.2: Comparison of denoising performance of blind SOUP-DIL and Super-BReD on *Barbara* at $\sigma = 60$. The performance metric used is PSNR (dB). The panels depict: (a) Ground truth image (b) Noisy image (12.6 dB) (c) SOUP-DIL denoised image (24.0 dB) and (d) Super-BReD denoised image (24.9 dB).

5.4.2 Single-Coil MR Image Reconstruction

Table 5.2 compares the performance of blind SOUP-DIL and Super-BReD at reconstructing single-coil MR images from undersampled reconstruction. For this experiment, we use two levels of undersampling– $4\times$ and $5\times$, where both the undersampling patterns were 1D phase encode. Once again, as in denoising, we notice increasing gains from incorporating a supervised learning component at higher undersampling/acceleration. Overall, in both cases, Super-BReD Learning produces an improvement over blind SOUP-DIL. Fig. 5.3 visualizes these improvements for one brain and one knee image. For the brain slice, we notice from the residue maps that a lot of the improvements are in regions near the skull, which accounts for a lot of the significant improvement in brain images and may not be clinically significant, but it possibly showcases the capability of Super-BReD Learning to identify (due to block matching) and learn from patterns that are common to both training and test data. Visualization of the knee image indicates that there are distinct differences in the preservation of details in the reconstructed image between the two techniques, and in the selected ROI, the Super-BReD reconstructed image recreates structures with higher fidelity.

However, we note that these improvements cannot be considered substantial, and our conjecture is that this is because of the limited availability of training data, and capability of the dictionary to learn compared to algorithms that leverage abundant pairwise training data. We will propose a deep generalization to the Super-BReD Learning algorithm in the next section that attempts to overcome this challenge.

UF	PSNR (dB)				
	Image	Zero-Filled	Blind SOUP-DIL	Super-BReD	Δ
4x	Brain 1	25.5	36.6	**37.7**	1.1
	Brain 2	25.9	36.6	**37.4**	0.8
	Knee 1	31.7	35.2	**35.4**	0.2
	Knee 2	34.0	37.6	**37.8**	0.2
5x	Brain 1	24.6	35.5	**36.3**	0.8
	Brain 2	24.9	35.5	**36.1**	0.6
	Knee 1	31.4	34.1	**34.4**	0.3
	Knee 2	33.7	36.3	**36.6**	0.3

Table 5.2: Table depicting performances (in PSNR) of blind SOUP-DIL and Super-BReD Learning at reconstructing MR images from undersampled single-coil measurements at 4x and 5x acceleration. Δ denotes the improvement over blind SOUP-DIL that Super-BReD Learning produces.

5.5 Future Work

We mainly focus on extending the Super-BReD Learning framework to denoising of ASL images, which ultimately allows for quantification of perfusion from fewer acquired ASL frames.

5.5.1 Denoising ASL time-series using Super-BReD Learning

In an application like ASL, where extensive clean or reliable datasets are unavailable, traditional Super-BReD Learning may be a compelling setting for combining blind and supervised learning. As explained in Chapter 2.3, ASL is a modality that suffers from low SNR, which is typically combated by acquiring more frames to average. This however, is a time consuming method, and we look at data-driven solutions to reduce the reliance on multiple averages.

Since the frames acquired in a single ASL scan can be considered to have the same ground truth information, and any discrepancies between the frames can be ascribed to motion and noise artifacts, we propose a spatio-temporal denoising framework, which strays away from the traditional, strictly model-based RETROICOR [62] or CompCor [59] processing popular in ASL literature.

RETROICOR uses low-order Fourier series to model the physiological noise component. This approach requires measurement of the physiological signals externally and removing them from an image time series through linear regression. The CompCor method improves upon the RETROICOR procedure in that it does not require external recording. This approach uses anatomical data in combination with variance maps of the ASL time course to identify white matter and cerebrospinal fluid. Oscillations in the perfusion signal of those regions are assumed to be dominated by physiological noise, rather than neuronal activation, which is assumed to be confined to the grey matter. Then, a Principal Component Analysis (PCA) from time series extracted from the white matter and the CSF regions yields time course fluctuations due to cardiac and respiratory fluctuations. The largest components can be regressed from the ASL time course, or used in tandem with a general linear model as nuisance variables. While RETROICOR requires external measurements to operate, CompCor ignores spatial information completely. We aim to exploit information about spatial features in our method, while still modeling inconsistent temporal fluctuations in the ASL time series as a nuisance variable.

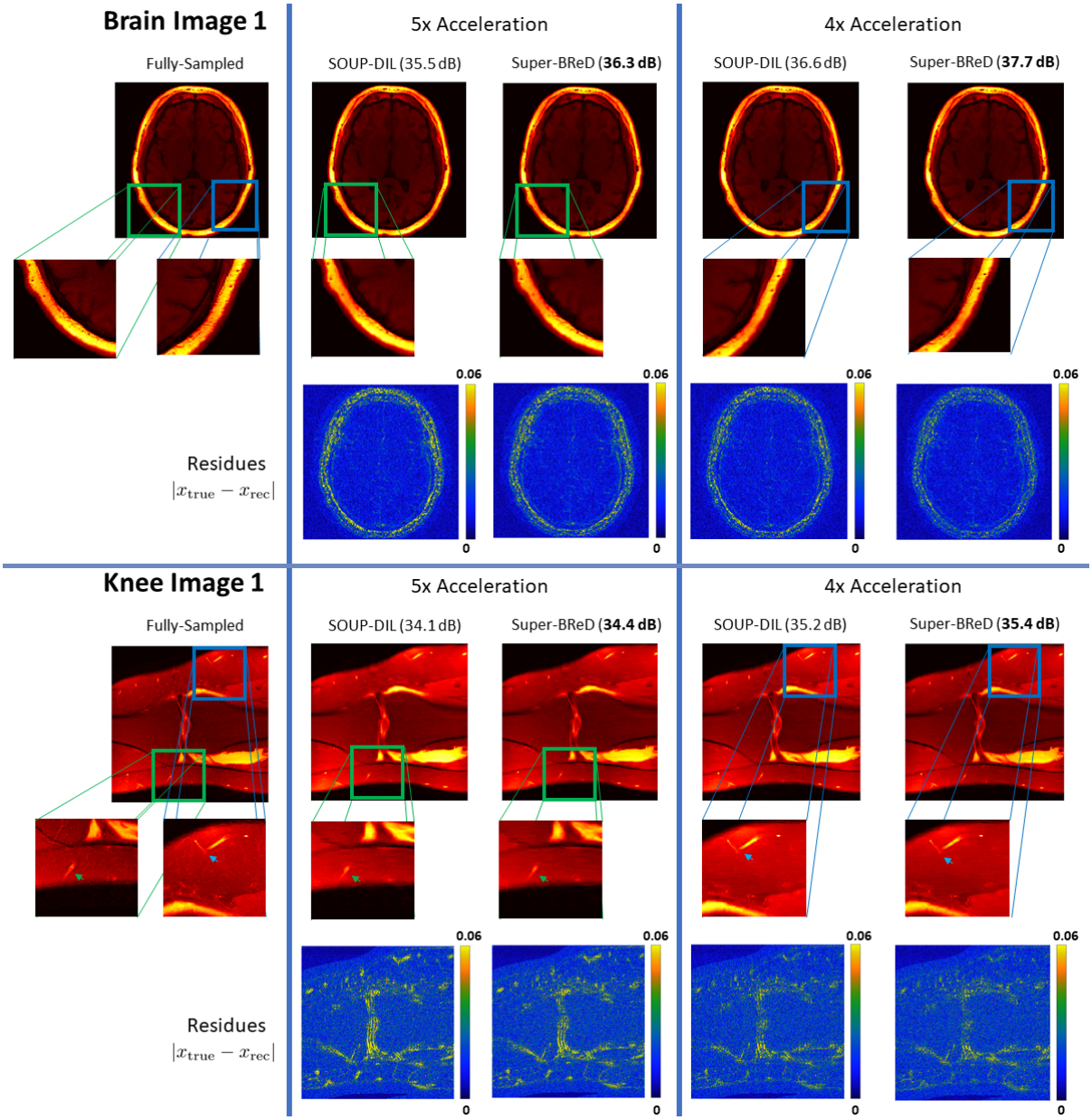


Figure 5.3: The test dataset for single-coil MRI consisted of one brain and one knee image. The left-most panel depicts the fully sampled reconstruction and corresponding zoomed-in ROIs used for comparisons. The center and right panels depict the reconstructions from blind SOUP-DIL and Super-BReD Learning at 4x and 5x accelerations respectively. Differences in reconstruction of details in aforementioned ROIs using the two algorithms are highlighted inset. The reconstruction quality is measured using PSNR (dB). The residues of the reconstructions w.r.t. the ground truth are also depicted.

In our proposed method, the spatial denoising uses a dictionary based regularization for the patches in the ASL perfusion image, while the temporal denoising and data consistency consists of a 'rank-one + sparse' decomposition of the acquired data. The spatial dictionary based denosing of the perfusion image patches is carried out using the Super-BReD learning framework, with two dictionaries: one learned in a supervised fashion from 'clean' perfusion image patches, and the other in a blind fashion from the noisy image patches.

We formulate this problem as:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \min_{\boldsymbol{S},\boldsymbol{D}_2,\boldsymbol{Z}_1,\boldsymbol{Z}_2} \|\boldsymbol{M} - \left(\boldsymbol{x}\mathbf{1}^T + \boldsymbol{S}\right)\|_F^2 + \lambda_{\boldsymbol{S}}\|\boldsymbol{S}\|_0$$
$$+ \sum_{j=1}^{N_2} \|\mathcal{P}_j\boldsymbol{x} - \hat{\boldsymbol{D}}_1\boldsymbol{Z}_1[:,j] - \boldsymbol{D}_2\boldsymbol{Z}_2[:,j]\|_F^2 + \lambda_1^2\|\boldsymbol{Z}_1\|_0 + \lambda_2^2\|\boldsymbol{Z}_2\|_0 \qquad \text{s.t.} \qquad \|(\boldsymbol{d}_2)_j\|_2 = 1 \;\forall\, j, \tag{5.17}$$

where, $\hat{\boldsymbol{x}} \in \mathbb{R}^q$ is the denoised ASL perfusion image, $\boldsymbol{M} \in \mathbb{R}^{q \times N_a}$ is a matrix whose columns are the N_a vectorized ASL difference image with same TR, delays etc., $\boldsymbol{S} \in \mathbb{R}^{q \times N_a}$ is a sparse matrix that accounts for discrepancies due to motion and other artifacts, and the other symbols have their usual meaning. The supervised dictionary $\hat{\boldsymbol{D}}_1$ is learned from patches obtained from temporally denoised ASL difference images with many frame-averages, and can thus be considered 'clean' or 'ground truth'.

The above problem can be separated as three sub-problems which can be solved alternatingly: (1) Temporal denoising, which involves solving for $\boldsymbol{S}$, which is essentially a hard-thresholding problem, that does not involve learning from data. (2) Sparse-coding and dictionary learning for $\boldsymbol{D}_2, \boldsymbol{Z}_1, \boldsymbol{Z}_2$, which involves the same technique discussed in section 5.2.3, and (3) Image update for $\boldsymbol{x}$, which is a least-squares problem.

We plan to test the efficacy of our algorithm against techniques like RETROICOR, CompCor as well as methods that rely on T1 priors to clean ASL difference images such as [63]. For our test data, we plan to use a section of the 'ground truth' datasets, wherein, we will use only a few of the acquisitions as the noisy input to our algorithm, while treating the average (temporally denoised) difference image created using the entire set of acquisitions as the ground truth. For this purpose, we have access to approximately 10 acquired datasets with 30 averages each, and will not need to acquire new data. Drawing from our observations in this chapter, we plan to use cerebral perfusion data from 3-4 of these subjects as training data, 2 subjects as our validation dataset for tuning parameters in our algorithm, and the rest as test data. As a metric for the performance of our method, we will use the error in CBF quantification with respect to the ground truth in regions of interest like gray and white matter. In [35], the choice for the granularity of CBF in the dictionary is 3mL/100g/min. Since this value is implicitly based upon a margin of significance, we will consider any improvement above this threshold, compared to other techniques, to be significant.

CHAPTER VI

Blind Primed Supervised (BLIPS) Learning for MR Image Reconstruction[1]

6.1 Introduction

Reconstruction of images from limited measurements requires solving an ill-posed inverse problem. In such problems, additional regularization is typically used. Often, such regularization reflects 'prior' knowledge about the class of images being reconstructed. Traditional regularizers exploit the sparsity of images in some domains [85, 86], or low-rankness [87, 88]. Compared to using a fixed regularizer, such as total variation (TV) or wavelet sparsity-based regularization, data-driven or adaptive regularization has proven to be very beneficial in several applications [89, 90, 91, 92, 93, 94]. In this form of reconstruction, one or more components of the regularizer, such as a dictionary or sparsifying transform, are learned from data adaptively, rather than being fixed to mathematical models like the discrete cosine transform (DCT) or wavelets. In particular, methods that exploit the sparsity of image patches in a learned transform domain or express image patches as a sparse linear combination of learned dictionary atoms have found widespread use in regularized MR image reconstruction [95, 96, 97, 98, 99].

A subset of this class of adaptive reconstruction algorithms relies only upon the measurements of the image being reconstructed to learn dictionaries or transforms, and uses no additional training data. These methods are dubbed *blind* learning-based reconstruction algorithms or blind compressed-sensing methods [100, 101]. One advantage of patch-based dictionary-blind reconstruction algorithms is that they do not require much (or any) training data to operate, and effectively leverage unique patterns present in the underlying data.

With the success of deep-learning-based methods for computer vision and natural language processing, there has also been a rise in methods that use neural networks to "regularize" (often in implicit manner) MRI reconstruction problems [102, 103, 104, 105, 106]. Some works treat reconstruction as a domain adaptation problem similar to style transfer and in-painting [107, 108, 109, 110]. Correspondingly, image refinement networks, such as the U-net [111], were adopted to correct the aliasing artifacts of the under-sampled input images. Although such CNN-based reconstruction methods achieved improved results compared to compressed sensing (CS) based reconstruction, the stability and interpretability of these models is a concern [112].

Besides improvements through algorithms, another driving force for supervised learning-based reconstruction is the curation of publicly available datasets for training. The availability of pairwise training data

[1] based on work in [129]

owing to initiatives like [113, 114] has further helped showcase the ability of deep learning-based algorithms for extracting or representing image features, and in learning richer models for image reconstruction in MR applications. These methods, due to their reliance on pixel-wise supervision perform exclusively *supervised* learning-based reconstruction, barring a few exceptions [115, 49].

Consequently, due to the popularity and computational efficiency of deep learning approaches across MRI applications, there has been a rising trend of favoring deep supervised methods over shallower dictionary-based methods—perhaps because the latter methods use "handcrafted" priors.

The rising popularity of supervised deep learning compared to shallow blind-dictionary learning may be based on an underlying assumption that the features learned using relatively unrestricted supervised deep models subsume those learned in a blind fashion, and other sparsity-based priors that are deemed "handcrafted". Though supervised deep-learned regularization may allow for the learning of richer models in reconstructing MR images, the aforementioned assumption is largely untested. Moreover, deep CNNs often require relatively large datasets to train well. This paper seeks to address these issues.

This work studies the processes of blind learning-based and supervised learning-based MRI reconstruction from under-sampled data, and highlights the complementarity of the two approaches by proposing a framework that combines the two in a residual fashion. We implement and compare multiple approaches for combining supervised and blind learning.

Our results indicate that supervised and dictionary-based blind learning may learn complementary features, and combining both frameworks using "BLInd Primed" Supervised (BLIPS) learning can significantly improve reconstruction quality. In particular, the combined reconstruction better preserves fine higher-frequency details that are very important in many clinical settings. We also find that this improvement from combining blind and supervised learning is relatively robust to changes in training dataset size, and across different imaging protocols.

The rest of this paper is organized as follows. Section 6.2 describes the blind and supervised learning-based approaches and the proposed strategies for combining them. Section 6.3 details the experiment settings, including datasets, hyper-parameters, and control methods. Section 6.4 presents the results and Section 6.5 provides related discussion. Finally, Section 6.6 explains our conclusions and plans for future work.

6.2 Problem Setup and Algorithms

This work combines two modern approaches to MR image reconstruction: dictionary-based blind learning reconstruction and CNN-based supervised learning reconstruction. The former approach capitalizes on the sparsity of natural images in an adaptive dictionary model. Usually, this method involves expressing patches in the MR image as a linear combination of a small subset of atoms or columns of a dictionary. Across several applications, including MR image reconstruction, learned or adaptive dictionaries often provide better representations of signals than fixed dictionaries. When these dictionaries are learned from the image being reconstructed, using no additional information, they are called *blind*, and can be considered to be 'tailored' specifically to the reconstruction at hand. Since individual image patches are approximated by different atoms, overcomplete dictionaries are often preferred for this approach because of their ability to provide richer representations of data.

For supervised learning reconstruction, this paper uses an unrolled network algorithm similar to the state-of-the-art method MoDL [104], whose variants have achieved top performance in recent open data-driven competitions in MR reconstruction [102, 116]. As 'unrolled' implies, the method consists of multiple

iterations or blocks. In each iteration, a CNN-based denoiser updates the image from the previous iteration. A subsequent data-consistency update ensures the reconstructed image is consistent with the acquired k-space measurements. By incorporating CNNs into iterative reconstruction, MoDL demonstrates improved reconstruction quality and stability compared to other direct inversion networks on large public datasets [102].

Given a set of k-space measurements $\boldsymbol{y}_c \in \mathbb{C}^p,\ c = 1, \ldots, N_c$, from N_c coils with corresponding system matrices $\boldsymbol{A}_c \in \mathbb{C}^{p \times q},\ c = 1, \ldots, N_c$, this section reviews the procedures of reconstruction using blind and supervised learning, and then proposes a method for combining them, along with a few special cases. We write the system matrix for the cth coil as $\boldsymbol{A}_c = P\mathcal{F}\boldsymbol{V}_c$, where $P \in \{0, 1\}^{p \times q}$ incorporates the mask that describes the sampling pattern, $\mathcal{F} \in \mathbb{C}^{q \times q}$ is the Fourier transform matrix and $\boldsymbol{V}_c \in \mathbb{C}^{q \times q}$ is the cth coil-sensitivity diagonal matrix, pre-computed from fully sampled k-space using the E-SPIRiT algorithm [117].

6.2.1 Reconstruction using Blind Dictionary Learning

Like most model-based regularized reconstruction approaches, the blind sparsifying dictionary learning-based reconstruction scheme solves for an image $\boldsymbol{x}$ that is consistent with acquired measurements, and possesses properties that are ascribed to the image (or a class of images). Mathematically, the approach optimizes a cost function that balances a data-fidelity term and a data-driven sparsity inspired regularization term as follows [118]:

$$\arg\min_{\boldsymbol{x}} \nu \sum_{c=1}^{N_c} \|\boldsymbol{A}_c\boldsymbol{x} - \boldsymbol{y}_c\|_2^2 + \mathcal{R}(\boldsymbol{x}), \tag{6.1}$$

where $\nu > 0$ reflects confidence in data fidelity and $\mathcal{R}(\boldsymbol{x})$ is a regularizer that, in the case of synthesis dictionary-based regularization, reflects the presumed sparsity of image patches as follows:

$$\begin{aligned} \mathcal{R}(\boldsymbol{x}) = \min_{\boldsymbol{D},\boldsymbol{Z}} & \sum_{j=1}^{N_1} \|\mathcal{P}_j\boldsymbol{x} - \boldsymbol{D}\boldsymbol{e}_j\|_2^2 + \lambda^2\|\boldsymbol{e}_j\|_0 \\ \text{s.t.} \quad & \|\boldsymbol{d}_u\|_2 = 1 \ \ \forall\, u, \end{aligned} \tag{6.2}$$

where $\mathcal{P}_j$ extracts the jth $\sqrt{r} \times \sqrt{r}$ overlapping patch of an image as a vector, $\boldsymbol{D} \in \mathbb{C}^{r \times U}$ denotes an overcomplete dicitionary, $\boldsymbol{d}_u$ its uth atom, $\boldsymbol{e}_j$ the sparse codes for the jth patch and the jth column of $\boldsymbol{Z}$, and λ is the sparsity penalty weight for dictionary learning, respectively.

A typical approach to solving this blind dictionary learning reconstruction problem is to alternate between updating the dictionary and sparse representation in (6.2) using the current estimate of the image $\boldsymbol{x}$, called *dictionary learning*, and then updating the reconstructed image itself (*image update*) through (6.1) using the current estimate of the regularizer parameters [119]. This alternation between dictionary learning and image update is repeated several times to obtain a *clean* reconstruction. Let $\boldsymbol{B}^i(\cdot)$ denote the function representing the ith iteration of this algorithm, and $\boldsymbol{x}_i \in \mathbb{C}^q$ be the reconstructed image at the start of the iteration, then we have

$$\boldsymbol{x}_{i+1} = \boldsymbol{B}^i(\boldsymbol{x}_i) = \boldsymbol{B}\big(\boldsymbol{x}_i; \nu_i, \lambda_i, \{\boldsymbol{A}_c, \boldsymbol{y}_c\}_{c=1}^{N_c}\big), \tag{6.3}$$

where ν_i, λ_i denote regularization parameters at the ith iteration for data fidelity and for dictionary learn-

ing, respectively. After K iterations, we have,

$$\boldsymbol{x}_{\text{blind}} = \boldsymbol{x}_K = \left(\bigcirc_{i=0}^{K-1} \boldsymbol{B}^i \right) (\boldsymbol{x}_0) = \boldsymbol{B}^{K-1} \left(\dots \boldsymbol{B}^2 \left(\boldsymbol{B}^0 (\boldsymbol{x}_0) \right) \right), \tag{6.4}$$

where $\bigcirc_{i=1}^{F}$ represents the composition of F functions $f_F \circ f_{F-1} \circ \dots \circ f_1$, and $\boldsymbol{x}_0$ is an inital image, possibly a zero-filled reconstruction.

In this work, we used a few iterations of the SOUP-DIL algorithm [119] for the dictionary and sparse representation update (or dictionary learning) in (6.2) and the conjugate gradient method for the image update step. (See next section for details.)

In our comparisons, we also investigated a similar iterative scheme as in (6.3), but the dictionary $\boldsymbol{D}$ in (6.2) is not learned from data, and is instead fixed (e.g., to a discrete cosine transform (DCT) or wavelet basis).

6.2.2 Reconstruction using Supervised Learning

The supervised learning module (MoDL [104]) also aims to solve (6.1). Often, in deep supervised regularized reconstruction, the reconstructed image is constrained to be in proximity of the output of a CNN. Introducing an auxiliary variable $\boldsymbol{z}$, which allows for the separation of the aforementioned constrained reconstruction problem into two components, (6.1) becomes:

$$\arg\min_{\boldsymbol{x},\boldsymbol{z}} \nu \sum_{c=1}^{N_c} \|\boldsymbol{A}_c \boldsymbol{x} - \boldsymbol{y}_c\|_2^2 + \mu \|\boldsymbol{x} - \boldsymbol{z}\|_2^2 + \mathcal{R}(\boldsymbol{z}), \tag{6.5}$$

where μ controls the consistency penalty between $\boldsymbol{x}$ and $\boldsymbol{z}$. MODL updates $\boldsymbol{x}$ and $\boldsymbol{z}$ in alternation. The $\boldsymbol{z}$ update is:

$$\boldsymbol{z}_{l+1} = \arg\min_{\boldsymbol{z}} \mathcal{R}(\boldsymbol{z}) + \mu \|\boldsymbol{x}_l - \boldsymbol{z}\|_2^2. \tag{6.6}$$

We replace the proximal operator in (6.6) with a residually connected denoiser $\boldsymbol{D}_\theta + \boldsymbol{I}$ applied to $\boldsymbol{x}_l$, where $\boldsymbol{I}$ is the identity mapping.

The $\boldsymbol{x}$ update involves a regularized least-squares minimization problem:

$$\boldsymbol{x}_l = \arg\min_{\boldsymbol{x}} \nu \sum_{c=1}^{N_c} \|\boldsymbol{A}_c \boldsymbol{x} - \boldsymbol{y}_c\|_2^2 + \mu \|\boldsymbol{x} - \boldsymbol{z}_l\|_2^2, \tag{6.7}$$

solved via conjugate gradient method.

Similar to blind learning, the lth iteration of supervised residual learning-based reconstruction algorithm can be written:

$$\begin{aligned} &\boldsymbol{x}_{l+1} = \boldsymbol{S}_l^\theta (\boldsymbol{x}_l) = \boldsymbol{S}\big(\boldsymbol{x}_l; \nu_l, \{\boldsymbol{A}_c, \boldsymbol{y}_c\}_{c=1}^{N_c}\big), \\ &\boldsymbol{S}\big(\bar{\boldsymbol{x}}; \nu, \{\boldsymbol{A}_c, \boldsymbol{y}_c\}_{c=1}^{N_c}\big) \triangleq \\ &\arg\min_{\boldsymbol{x}} \nu \sum_{c=1}^{N_c} \|\boldsymbol{A}_c \boldsymbol{x} - \boldsymbol{y}_c\|_2^2 + \|\boldsymbol{x} - \big(\boldsymbol{D}_\theta(\bar{\boldsymbol{x}}) + \bar{\boldsymbol{x}}\big)\|_2^2, \end{aligned} \tag{6.8}$$

where $\bar{\boldsymbol{x}}$ denotes the input image for the residual learning-based reconstruction algorithm. After L iterations,

we have

$$\boldsymbol{x}_{\text{supervised}} = \boldsymbol{x}_L = \left(\bigcirc_{l=0}^{L-1} \boldsymbol{S}_\theta^l \right) (\boldsymbol{x}_0). \tag{6.9}$$

The network parameters θ are learned in a supervised manner so that $\boldsymbol{x}_{\text{supervised}}$ matches known ground truths (e.g., in mean squared error or other metrics) on a training data set.

6.2.3 Combining Blind and Supervised Reconstruction

Fig. 6.1 (P1) depicts our proposed BLIPS approach to combining blind and supervised learning. The skipped connection in the deep network enables the addition of the previous iterate to the output of the denoiser during supervised reconstruction, and ensures separation (the output of the residual denoiser gets added to the blind image going into data consistency) of the blind learned image and the supervised learned image in the first iteration when the aforementioned algorithms are combined. In subsequent iterations, this skipped connection also causes the denoiser to learn residual features after the combination of blind and supervised learning in the previous iteration. The output of the full pipeline of our proposed method Fig. 6.1 (P1) is:

$$\text{(P1)} \qquad \hat{\boldsymbol{x}} = \left(\bigcirc_{l=0}^{L-1} \boldsymbol{S}_\theta^l \bigcirc_{i=0}^{K-1} \boldsymbol{B}^i \right) (\boldsymbol{x}_0) \triangleq \boldsymbol{\mathcal{M}}_\theta(\boldsymbol{x}_0). \tag{6.10}$$

This pipeline is also dubbed the (B+S) pipeline in our experiments.

6.2.4 Training the Denoiser Network

The denoiser $\boldsymbol{D}_\theta$ shares weights across iterations. To train it, we use the output of our proposed pipeline (P1) in a combined ℓ_1 and ℓ_2 norm training loss function as follows:

$$\hat{\theta} = \arg\min_\theta \sum_{n=1}^{N_2} C_\beta(\boldsymbol{\mathcal{M}}_\theta(\boldsymbol{x}_0^{(n)}); \boldsymbol{x}_{\text{true}}^{(n)}) = \arg\min_\theta$$
$$\sum_{n=1}^{N_2} \left(\left\| \boldsymbol{x}_{\text{true}}^{(n)} - \boldsymbol{\mathcal{M}}_\theta(\boldsymbol{x}_0^{(n)}) \right\|_2^2 + \beta \left\| \boldsymbol{x}_{\text{true}}^{(n)} - \boldsymbol{\mathcal{M}}_\theta(\boldsymbol{x}_0^{(n)}) \right\|_1 \right),$$

where n indexes the training data consisting of target images $\boldsymbol{x}_{\text{true}}^{(n)}$ reconstructed from fully sampled measurements and corresponding undersampled k-space measurements, and $C_\beta(\hat{\boldsymbol{x}}; \boldsymbol{x}_{\text{true}})$ denotes the training loss function. The initial $\boldsymbol{x}_0^{(n)}$ are obtained from the undersampled k-space measurements using a simple analytical reconstruction such as zero-filling inverse FFT reconstruction. Our implementation used $\beta = 0.01$ in (6.11), which was chosen empirically.

6.2.5 Direct Addition of Blind and Supervised Learning

A special case we investigate is when there is no residual connection in (P1), and we add the blind reconstruction output directly to the output of the supervised deep network during the data consistency update, as described in (6.11) below. Similar to (P1), the input to the supervised module is also the blind reconstruction

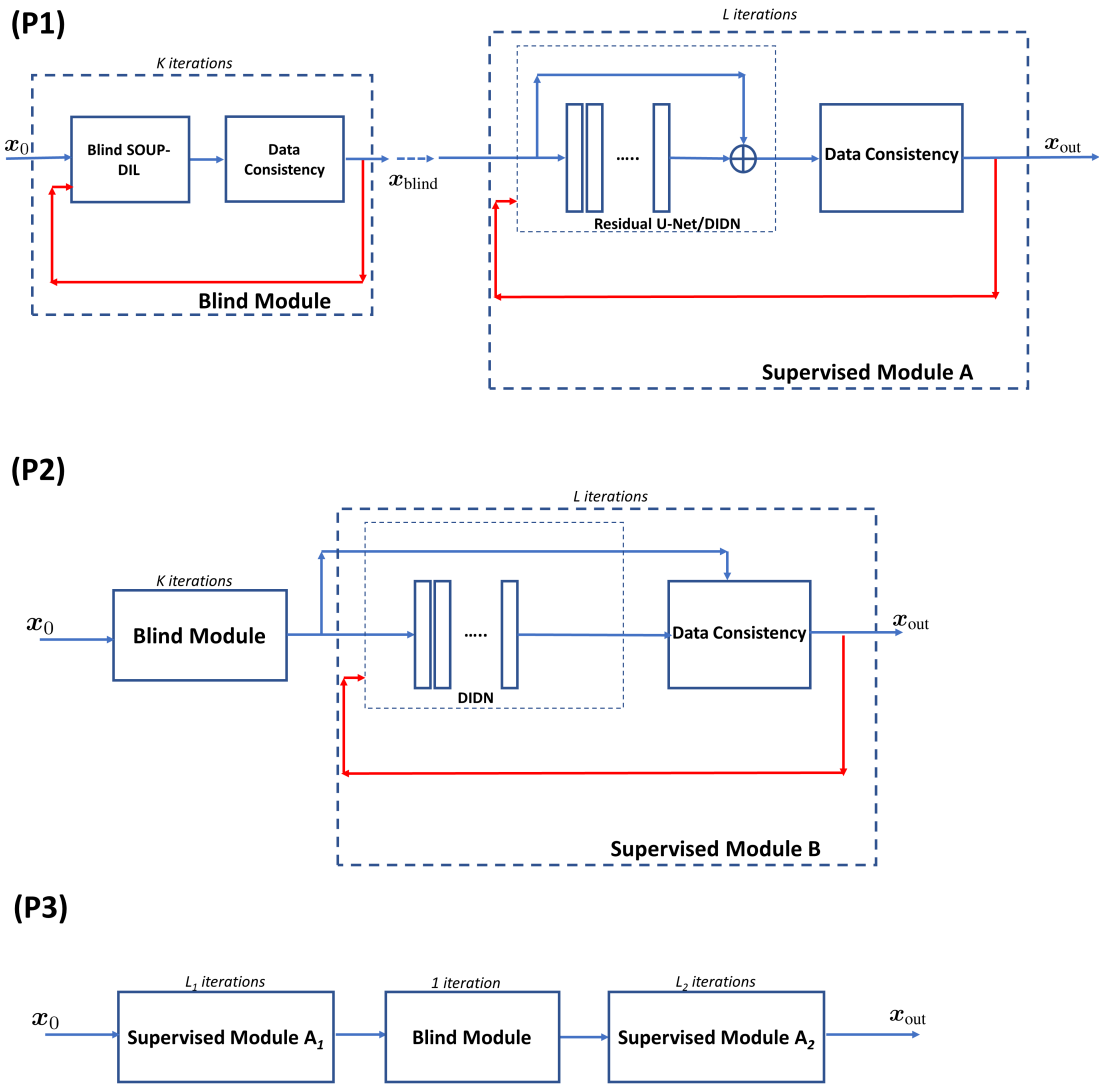


Figure 6.1: Proposed pipelines (P1), (P2) and (P3) for combining blind and supervised learning-based MR image reconstruction.

output. We express an iteration of such an algorithm as follows:

$$\begin{aligned}
&\boldsymbol{x}_{l+1} = \tilde{\boldsymbol{S}}_\theta^l(\boldsymbol{x}_l) = \tilde{\boldsymbol{S}}\big(\boldsymbol{x}_l; \boldsymbol{x}_{\text{blind}}, \nu_l, \{\boldsymbol{A}_c, \boldsymbol{y}_c\}_{c=1}^{N_c}\big), \\
&\tilde{\boldsymbol{S}}(\boldsymbol{x}; \boldsymbol{x}', \nu, \{\boldsymbol{A}_c, \boldsymbol{y}_c\}_{c=1}^{N_c}) = \arg\min_{\bar{\boldsymbol{x}}} \\
&\nu \sum_{c=1}^{N_c} \|\boldsymbol{A}_c \bar{\boldsymbol{x}} - \boldsymbol{y}_c\|_2^2 + \|\bar{\boldsymbol{x}} - \big(\boldsymbol{D}_\theta(\boldsymbol{x}) + \boldsymbol{x}'\big)\|_2^2,
\end{aligned} \tag{6.11}$$

where $\boldsymbol{x}_0 = \boldsymbol{x}'$ is the initial input to the supervised module. After L iterations, the reconstruction is:

$$\text{(P2)} \qquad \tilde{\boldsymbol{x}} = \left(\bigcirc_{l=0}^{L-1}\right) \tilde{\boldsymbol{S}}_\theta^l(\boldsymbol{x}_l) = \tilde{\boldsymbol{\mathfrak{M}}}_\theta(\boldsymbol{x}_{\text{blind}}), \tag{6.12}$$

where $\boldsymbol{x}_{\text{blind}} = \left(\bigcirc_{i=0}^{K-1} \boldsymbol{B}^i\right)(\boldsymbol{x}_0)$, as depicted in Fig. 6.1 (P2). The training loss for this variation is:

$$\hat{\theta} = \arg\min_{\theta} \sum_{n=1}^{N_2} C_\beta(\tilde{\mathfrak{M}}_\theta(\boldsymbol{x}_{\text{blind}}^{(n)}); \boldsymbol{x}_{\text{true}}^{(n)}). \tag{6.13}$$

6.2.6 Combined Supervised and Blind Learning with Feedback

Since iterations of blind learning-based reconstruction take significantly longer than propagating an image through a deep network, we investigated a feedback-based pipeline that reduces computation by only approximately optimizing the objective of blind learning reconstruction (using an outer single iteration of the blind learning module) that in turn is warm-started by a supervised learning reconstruction. The result of partial blind learning is then fed into a second stage with a supervised deep network similar to (P1), as depicted in Fig. 6.1 (P3), introducing image-adaptive features that may improve image quality. Essentially, the output for this pipeline can be expressed as:

$$\begin{aligned} \text{(P3)} \qquad \hat{\boldsymbol{x}} &= \left(\bigcirc_{l=0}^{L_2-1} \boldsymbol{S}_{\theta_2}^l \circ \boldsymbol{B}^1 \circ \bigcirc_{l=0}^{L_1-1} \boldsymbol{S}_{\theta_1}^l\right)(\boldsymbol{x}_0) \\ &= \bar{\mathfrak{M}}_{\theta_1,\theta_2}(\boldsymbol{x}_0), \end{aligned} \tag{6.14}$$

where θ_1 and θ_2 are the weights of the initial and second stage unrolled networks, respectively. The training losses for these unrolled networks are:

$$\hat{\theta}_1 = \arg\min_{\theta_1} \sum_{n=1}^{N_2} C_\beta(\bar{\mathfrak{M}}_{\theta_1}(\boldsymbol{x}_0^{(n)}); \boldsymbol{x}_{\text{true}}^{(n)}), \tag{6.15}$$

$$\hat{\theta}_2 = \arg\min_{\theta_2} \sum_{n=1}^{N_2} C_\beta(\bar{\mathfrak{M}}_{\theta_1,\theta_2}(\boldsymbol{x}_0^{(n)}); \boldsymbol{x}_{\text{true}}^{(n)}), \tag{6.16}$$

respectively, where $\bar{\mathfrak{M}}_{\theta_1}(\boldsymbol{x}_0) = \left(\bigcirc_{l=0}^{L_1-1} \boldsymbol{S}_{\theta_1}^l\right)(\boldsymbol{x}_0)$ and the other symbols are as explained above. We train θ_1 and θ_2 separately in two stages. The training of θ_2 starts after θ_1 converges. The combination of supervised and partial blind learning could be iterated. We worked with a two-stage network architecture and a single iteration of blind learning optimization to keep computations low. This pipeline is also dubbed the (S+B+S) pipeline in our experiments.

6.3 Experimental Framework

6.3.1 Training and Test Dataset

We trained and tested both our method and a strict supervised learning-based method with the same deep learning architecture (described below) on two datasets[2]. The first was a randomly selected subset from the fastMRI knee dataset, while the second consisted of the entire fastMRI brain dataset [113]. In the first case, our dataset for training and testing consisted of 8705 knee images, and were used in experiments involving

[2]The code will be publicly available on Github if accepted.

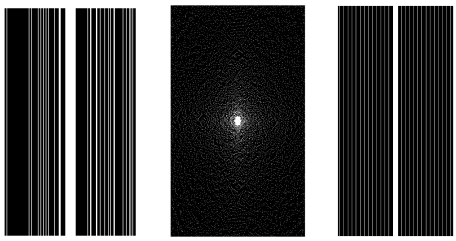
the proposed pipelines in (P1) and (P2). We used smaller and randomly-selected subsets for our various experiments, which is described in detail in section 6.4.

To test the pipeline proposed in (P3), we used the fastMRI Brain dataset, consisting of 23220 T1 weighted images, 42250 T2 weighted images and 5787 FLAIR slices. For each contrast, we reserved 500 images as the test data and the rest for training and validation.

All sensitivity maps were estimated using the ESPIRiT [117] method. The details of the algorithms in our work are explained below.

6.3.2 Undersampling Masks

For experiments with the pipeline (P1), we used three types of undersampling masks. First, we used the $5\times$ Cartesian phase encode undersampling mask shown in Fig. 6.2(a) that was held fixed across training and test images. This pattern had 29 fully sampled lines in the center of the k-space, and the remaining lines were sampled uniformly at random. We similarly tested (P1) on 2D Poisson-disk Cartesian undersampling at $20\times$ acceleration. Finally, we tested (P1) by varying the 1D phase encode undersampling mask Fig. 6.2(a) used across training and test images randomly, to further evaluate its generalizability across different sampling patterns. For this purpose, we used $\approx 4.5\times$ undersampling, and 24 fully sampled k-space lines. Pipeline (P2) was tested using only the sampling pattern in Fig. 6.2(a), while (P3) was tested using $8\times$ equidistant acceleration mask shown in Fig. 6.2(c), as well as the 1D phase encode mask in Fig. 6.2(a). This mask had 4% fully sampled lines at the center of k-space [113].



(a) (b) (c)

Figure 6.2: Undersampling masks used in experiments: (a) 5-fold undersampled 1D Cartesian phase-encoded; (b) 20-fold undersampled Cartesian Poisson-disk; and (c) $8\times$ equidistant.

6.3.3 Blind Dictionary Learning-based Reconstruction

We used the SOUP-DIL algorithm [119] to perform blind dictionary learning-based reconstruction initialized with a 'zero-filled' reconstruction of the data. In both (P1) and (P2), we set the number of outer iterations to be $K = 20$, and each outer iteration had 5 inner iterations of dictionary learning and sparse-coding. We set $\nu_i = 8 \times 10^{-4}$ and $\lambda_i = 0.2$ across iterations, respectively. The dictionary size was 36×144 and the initial dictionary was an overcomplete inverse DCT matrix, while the sparse code matrix was initialized with zeros. We used conjugate gradient method to perform the data consistent image update. It required ≈ 170 seconds to perform 20 iterations of SOUP-DIL reconstruction of a single 640×368 image slice, on an Intel(R) Xeon(R) E5-2698 with 40 cores. For (P3), we used only one ($K = 1$) iteration of SOUP-DIL reconstruction with $\nu = 0.5$ and $\lambda = 0.8$ on the fastMRI brain dataset (when used on the knee dataset, these were fixed to values mentioned earlier). For experiments involving the fastMRI brain dataset and pipeline (P1), we only use $K = 3$ outer iterations of SOUP-DIL reconstruction, due to the huge dataset size. A single

iteration of SOUP-DIL took ≈ 6.5 seconds to reconstruct a 640×320 image on the same server. (Table VIII in the Supplementary Materials compares reconstruction time for different methods.)

When performing non-adaptive dictionary-based reconstruction, we fixed the dictionary to its inverse DCT initialization across all iterations, while keeping all other algorithm parameters unchanged. The experiment and results are shown in Sec. 6.7.1 of Supplementary Materials.

An additional experiment compared the compressed sensing algorithm against blind dictionary learning. We used the MRI reconstruction instance included in the SigPy package[3], which uses the primal-dual hybrid gradient (PDHG) algorithm and 30 iterations. The sparsity penalty is the ℓ_1 norm of a orthogonal discrete wavelet transform, with a weight of 10^{-7} compared with the data-fidelity term.

6.3.4 Supervised Reconstruction

The denoiser $\boldsymbol{D}_\theta$ we used is the Deep Iterative Down-Up Network [120], which has been shown to be efficient on previous benchmark research with the same fastMRI dataset [102] and in an image denoising competition [121]. Real and imaginary component of the complex-valued images are formulated as two input channels of the network. The magnitude of the input image is normalized by the median absolute value. The batch size is set to 4. We set the data-fidelity weight $\nu = 2$ for the supervised learning.

In each iteration of (6.8), we used the conjugate gradient method to solve the least-squares minimization problem. Backpropagation of the least-squares problem (calculation of the Jacobian-vector product) is also performed using the conjugate gradient method. Here we set $L = 6$ to balance reconstruction quality and model dimension. In the inference phase, the time cost is around 1.2s for a 20-channel 640×320 slice on a single Nvidia(R) GTX1080Ti GPU. For a fair comparison, the denoiser training settings are the same between different scenarios in Section 6.4. The number of epochs is set to 40, with a linearly decaying learning rate from 1e-4 to 0. The optimizer was Adam [122], with parameter βs = $[0.5, 0.999]$.

6.3.5 Performance Metrics

For a quantitative comparison of the reconstruction quality, we used three common metrics: peak signal-to-noise ratio (PSNR, in dB), structural similarity index (SSIM) [123], and high-frequency error norm (HFEN) [96], to measure the similarity between reconstructions and ground truth. The HFEN was computed as the ℓ_2 norm of the difference of edges between the input and reference images. Laplacian of Gaussian (LoG) filter was used as the edge detector. The kernel size was set to 15×15, with a standard deviation of 1.5 pixels.

6.4 Results

6.4.1 Comparing Blind+Supervised vs Strictly Supervised Reconstruction

Table 6.1 compares the performance of combined blind and supervised learning versus strictly supervised learning on datasets of various sizes using (P1). We used 4 training dataset sizes: 1105, 2244, 4198, and 8205 slices. 10% for each training set was reserved for validation purposes. The test set consisted of 500 different slices. Training/validation set and test set are from different subjects to avoid data leakage between slices.

Our proposed method's improvements are fairly robust even when the total dataset size increases, as illustrated in Fig. 6.3 that depicts Table 6.1 as a bar chart. Moreover, for small-scale datasets, which are usually the case in medical imaging, our method still provides significant improvements over the strict supervised

[3] https://github.com/mikgroup/sigpy

scheme. We conjecture that the blind learning-based reconstruction provides an image where many artifacts have been resolved and details have been restored that the supervised learning reconstruction can further refine.

Tables 6.2 and 6.3 display the quantitative results with the 2D Poisson disk Cartesian sampling pattern and 1D variable density Cartesian sampling mask (changing randomly across training and test cases), respectively. The training/validation set consisted of 4198 slices and the test set consisted of 500 slices (same as the 4198/500 slices in the previous case). The improvement provided by our scheme (B+S) over strict supervised learning (S) holds for multiple sampling masks, and is significant under the paired t-test ($P < 0.005$).

To support the assertion that BLIPS can learn different features than supervised learning, Figs. 6.4, 6.5, and 6.6 also display example slices. Compared to supervised learning, the most obvious difference in the combined model is the better restoration of fine details. It can be seen that in the blind dictionary learning results, a fair amount of fine structure is already recovered from the aliasing artifacts. The dictionary learning results provide a foundation for supervised learning to then residually reduce aliasing artifacts while preserving these details. This is also strongly implied by our observations in Section VII B and accompanying Fig. 6.9.

Table 6.4 compares the proposed BLIPS techniques to strict supervised learning, and to supervised learning initialized with compressed sensing. The compared methods were trained and tested on identical datasets (4198 slices). The results indicate that the S+B+S BLIPS reconstruction yields the best performance, and the B+S reconstruction provides the second best performance. However, even compressed sensing reconstruction combined with supervised learning-based reconstruction performs better than strict supervised learning-based reconstruction.

Dataset Size	8205		4198		2244		1105	
Method	S	B+S	S	B+S	S	B+S	S	B+S
SSIM	0.944	**0.947**	0.942	**0.946**	0.939	**0.943**	0.930	**0.941**
PSNR (dB)	35.44	**35.70**	35.09	**35.53**	34.65	**35.05**	33.92	**34.82**
HFEN	0.450	**0.433**	0.470	**0.443**	0.494	**0.471**	0.538	**0.484**

Table 6.1: Comparison of supervised learning-based reconstruction (S) versus our proposed combined blind and supervised learning-based reconstruction (B+S) using (P1) at various knee training dataset sizes for 5$\times$ acceleration using 1D Cartesian undersampling. The undersampling mask in Fig. 6.2a was held fixed for training and testing. Bold digits indicate that B+S method performed significantly better than the S method under pairwise t-test ($P < 0.005$).

Recon. Method	Supervised	Blind	Blind+Supervised
SSIM	0.960	0.949	**0.963**
PSNR (dB)	35.33	33.38	**35.81**
HFEN	0.434	0.507	**0.403**

Table 6.2: Comparison of supervised learning-based reconstruction versus our proposed BLIPS and blind learning-based reconstruction using (P1) for 20 $\times$ acceleration using Cartesian 2D Poisson disk undersampling with mask shown in Fig. 6.2b. The fastMRI knee dataset was used for training and testing. Bold digits indicate that B+S method performed significantly better than the S method under paired t-test ($P < 0.005$).

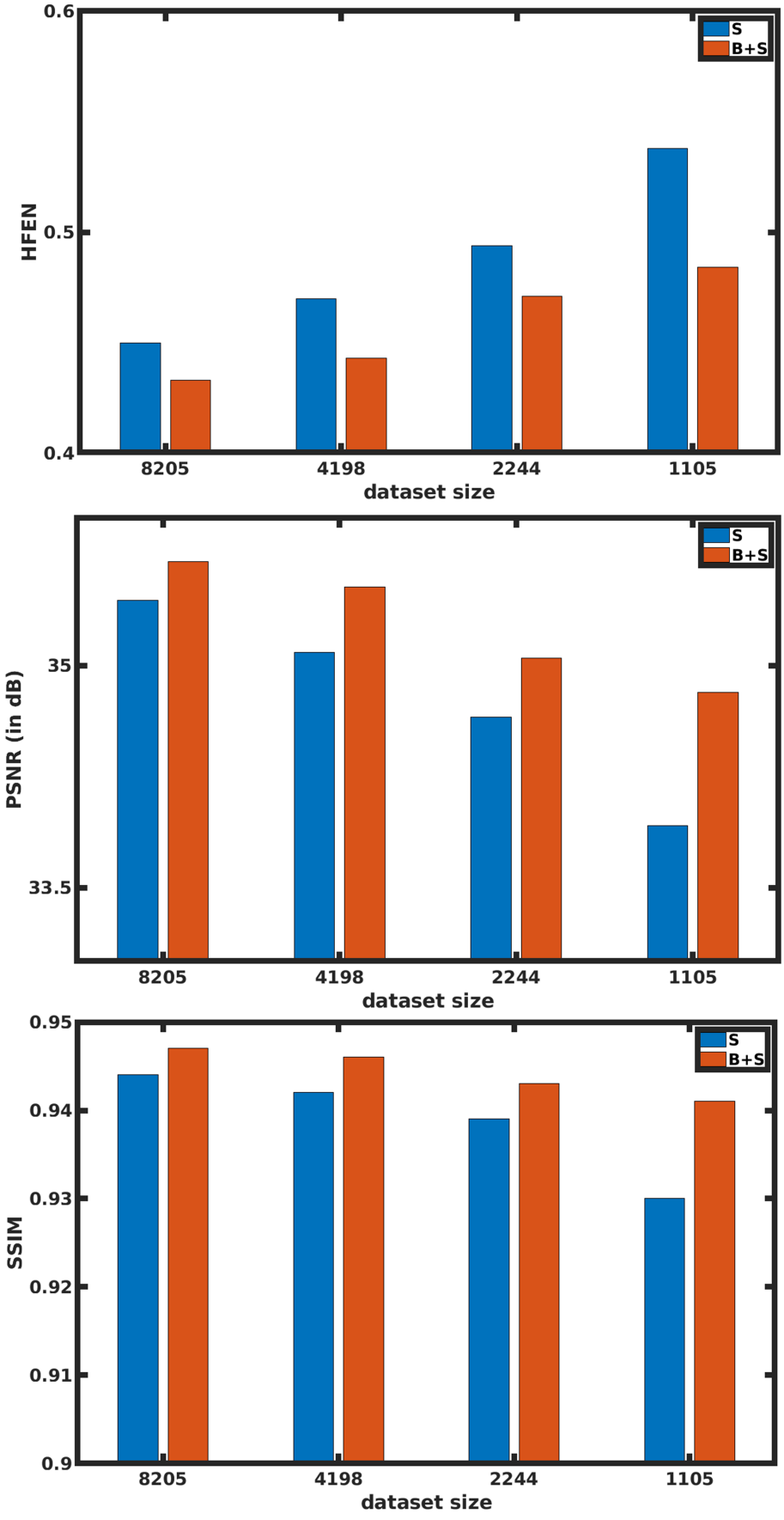


Figure 6.3: Comparison of strict supervised learning-based reconstruction with BLIPS reconstruction across various knee dataset sizes. Table 6.1 shows the corresponding quantitative values.

Recon. Method	Supervised	Blind	Blind+Supervised
SSIM	0.954	0.945	**0.957**
PSNR (dB)	34.34	32.79	**34.80**
HFEN	0.308	0.360	**0.284**

Table 6.3: Comparison of performance of supervised learning-based reconstruction against our proposed BLIPS and blind learning-based reconstruction using (P1) for $\approx 4.5\times$ acceleration using random variable density 1D sampling mask (changing randomly across training and test cases). The fastMRI knee dataset was used for training and testing. Bold digits indicate that B+S method performed significantly better than the S method under paired t-test ($P < 0.005$).

Recon. Method	S	B	CS+S	B+S	S+B+S
SSIM	0.942	0.906	0.943	0.946	**0.948**
PSNR (dB)	35.09	30.29	35.24	35.53	**35.83**
HFEN	0.470	0.648	0.464	0.443	**0.426**

Table 6.4: Comparison of supervised learning-based reconstruction versus various proposed BLIPS reconstruction approaches using (P1) and (P2), and CS-initialized supervised reconstruction for $5\times$ acceleration using 1D Cartesian undersampling with mask shown in Fig. 6.2a. Training was performed using 4198 knee slices from the fastMRI Knee dataset. Bold digits indicate that S+B+S method performed significantly better than the S method and CS+S method under paired t-test ($P < 0.005$).

6.4.2 Strict Separation of Blind and Supervised Learning Reconstruction

Table 6.5 compares explicitly combining blind and supervised learning using (P2) without residual learning against the proposed method for combining blind and supervised learning. The sampling pattern here is the same as in Fig. 6.2a. The dataset is the same as the 8205/500 case in Table 6.1. Compared to explicit consistency with blind learning results, our latent approach reaches a better result. The results demonstrate that rather than a fidelity prior, the blind learned reconstruction works better as an input to the deep residual network for further refinement.

Recon. Method	Explicit Blind + Supervised	Proposed Blind + Supervised
SSIM	0.938	**0.946**
PSNR (dB)	34.29	**35.53**
HFEN	0.495	**0.443**

Table 6.5: Comparison of combined blind and supervised learning using (P1) versus explicit addition of blind and supervised learning using (P2) for the mask in Fig. 6.2a. Training was performed using 4198 knee slices from the fastMRI Knee dataset. Bold digits indicate that B+S method performed significantly better than the explicit blind + supervised method under paired t-test ($P < 0.005$).

6.4.3 Combined Supervised and Blind Learning with Feedback

For the large-scale brain dataset, we tested the idea of using a supervised learning network's output as a potentially improved initialization for blind learning (6.14). The blind learning cost is then optimized for a single iteration with this improved initialization to incorporate additional details captured with blind learning to improve the first supervised network's reconstruction. The blind learning result is passed on to another (second) stage of supervised learning. The networks' parameters θ_1 and θ_2 are pre-trained on all

three contrasts and fine-tuned on individual contrast, including T1w, T2w and FLAIR. As a control method, we concatenated two supervised learned networks sequentially, which can also improve the reconstruction performance compared with a single unrolled supervised network, and demonstrates substantial improvements in PSNR, SSIM, and HFEN for S+B+S with a large dataset. We also compare to deep supervised reconstruction preceded by a few iterations of blind dictionary learning. (We used 3 iterations here due to the time constraints associated with generating data for the large fastMRI brain dataset.)

Table 6.6 summarizes the results of this comparison, showing that while S+B+S performs the best, even B+S (which on the brain dataset, only used 3 iterations of SOUP-DIL reconstruction in the blind module) manages to outperform strict supervised learning in most contrasts. Fig. 6.7 shows an example slice for this comparison. Again, combined blind and supervised learning using (P3) preserves finer details better than cascaded strict supervised learning.

Dataset	T1w			T2w			FLAIR		
Method	S+S	B+S	S+B+S	S+S	B+S	S+B+S	S+S	B+S	S+B+S
SSIM	0.965	0.966	**0.968**	0.964	0.966	**0.967**	0.944	0.945	**0.947**
PSNR (dB)	36.86	36.85	**37.27**	35.37	35.72	**35.88**	34.23	34.36	**34.62**
HFEN	0.388	0.384	**0.369**	0.371	0.353	**0.349**	0.481	0.470	**0.458**

Table 6.6: Comparison of strictly supervised learning-based reconstruction (S+S) versus the proposed combined blind and supervised learning-based reconstruction (S+B+S) in (P3) for the fastMRI brain dataset with $8\times$ undersampling with the mask in Fig. 6.2c

.

6.4.4 Performance in the Presence of Planted Features

To compare the ability of BLIPS reconstruction and strictly supervised reconstruction to faithfully reproduce image features that are not present in the training dataset (as is often the case with identifying pathologies, etc.), we planted some features in a knee image from the fastMRI dataset, from which raw k-space was simulated and undersampled, inspired by [112]. The undersampling pattern was 1D variable density $\approx 4.5\times$, and was chosen at random to further test robustness.

Fig. 6.8 shows the aforementioned comparison. The BLIPS reconstruction reproduces the planted features with significantly higher fidelity than strict supervised reconstruction, and has much fewer aliasing artifacts, as is evident from the residue maps (also pointed out by the blue arrows in the figure). The details or edges of the planted features are better preserved in the BLIPS reconstruction compared to strict supervised learning-based reconstruction. The phenomena are consistent across simulated attempts we have tried.

6.5 Discussion

This work investigated the combination of blind and supervised learning algorithms for MR image reconstruction. Specifically, we proposed a method that combines dictionary learning-based blind reconstruction with model-based supervised deep reconstruction in a residual fashion. Comparisons against strictly supervised learning-based reconstruction indicate that the proposed reconstruction method significantly improves reconstruction quality in terms of metrics including PSNR, SSIM, and HFEN, across a range of undersampling and acceleration factors. The robustness of these improvements to the training dataset size suggests that the features learned during blind learning-based reconstruction using a sparse dictionary adapted separately

for each training and testing image may differ significantly from features learned by deep networks trained on a large dataset with strictly pixel-wise supervision. While the latter showcases the potential for removing global aliasing artifacts, the former successfully leverages patterns in an image that are learned just from its measurements, thereby preserving the finer details of the image in the reconstruction. This claim is further supported by the error maps of regions of interest of reconstructed image slices. Moreover, the experiments using planted features suggest that BLIPS reconstruction can adapt to, and reproduce unfamiliar (absent from the training set) features better than strict supervised learning-based reconstruction. This ability may be a distinct benefit in the context of identifying pathology in MRI images. The combination of compressed sensing MRI and deep-supervised learning-based reconstruction also outperformed strict supervised learning-based reconstruction, reinforcing that features learned using supervision may not subsume traditional sparsity-based priors.

Past studies have shown that deep learning-based reconstruction is good at reducing aliasing artifacts compared with model-based iterative methods such as compressed sensing. The majority of supervised models are trained with pixel-wise ℓ_1/ℓ_2 norm loss. These approaches generally produce smooth images with high PSNR but can also introduce blurring. Other methods use GANs or perceptual loss to preserve details. However, these data-driven methods are often known to introduce realistic artifacts, which is very risky for medical imaging reconstruction. In our approach, the intrinsic sparsity of MR images is exploited in the dictionary learning phase to preserve fine structures. Thus, our method combines the advantages of both worlds: the representation ability of CNNs to resolve aliasing artifacts and dictionary-based signal modeling to recover high-frequency details. The superior performance in fine-detail recovery is reflected in the smaller HFEN values that quantify high-frequency features.

From the network training perspective, compared to the pure supervised model our network demonstrates improved stability and generalizability since it is powered and complemented by both model-based and adaptive dictionary learning-based components. First, on a relatively small dataset (1105/2244 images), the method still achieved similar results as with the full (8205 images) training dataset. This means that our method has clearly lower requirements on the amount of training data to work well compared to the massive amount of training data needed by typical deep learning-based reconstruction algorithms. Second, the improvements hold across different sampling patterns with very different PSFs. Third, although 40 training epochs were used in experiments, our approach requires only 5-8 epochs to converge (with no obvious over-fitting seen thereafter). In contrast, the supervised model required 20-30 epochs for the training loss to converge.

Due to the serial nature of the SOUP-DIL algorithm [119] used for dictionary learning here, our algorithm's reconstruction time is higher than that of strictly supervised reconstruction. The computational bottleneck is in the atom-wise block-coordinate descent approach to dictionary updating, which cannot be accelerated by simple vectorization. These alternating updates between each dictionary atom and the corresponding sparse codes [119] allow for the blind algorithm to residually learn and represent features in the reconstructed image. Further acceleration of the blind dictionary learning approach might be needed to use the approach in clinical settings that need a real-time imaging reconstruction workflow. However, it may be still acceptable for most conventional settings since the scanning itself is often the throughput bottleneck. The proposed S+B+S approach involves a much quicker (partial) dictionary learning-based step compared to the proposed vanilla B+S approach. Other fast blind learning approaches involving transform learning [96] could also make our schemes much more efficient.

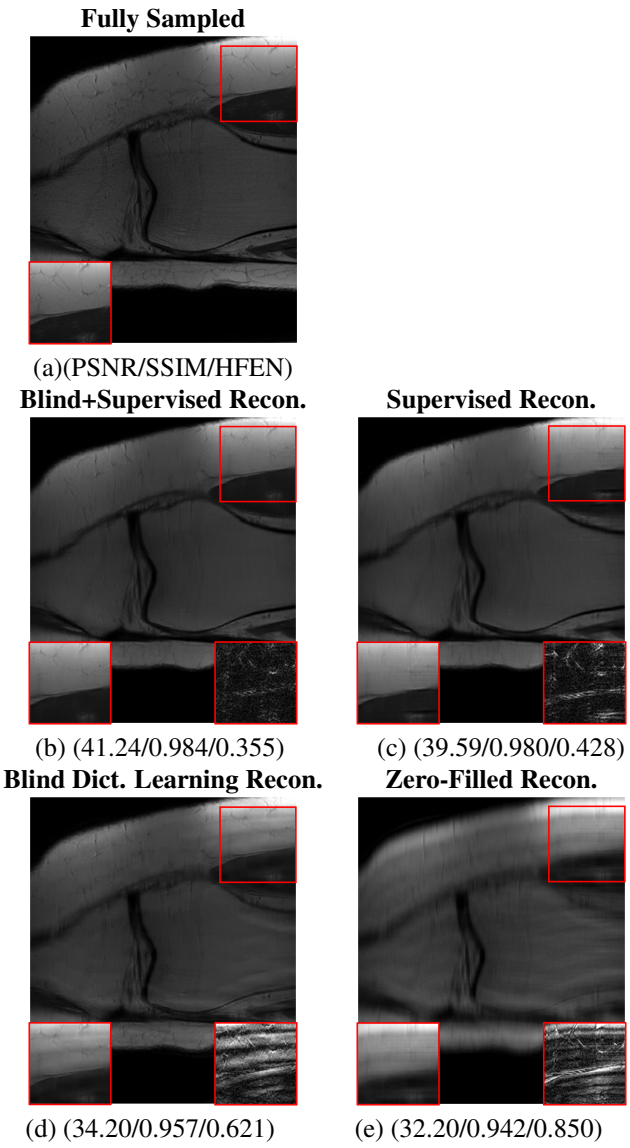


Figure 6.4: Comparison of reconstructions for a knee image using the proposed method versus strict supervised learning, blind dictionary learning, and zero-filled reconstruction for the 5× undersampling mask depicted in Fig. 6.2a. Metrics listed below each reconstruction correspond to PSNR/SSIM/HFEN respectively. The inset panel on the bottom left in each image corresponds to regions of interest (indicated by the red bounding box in the image) in the image that benefits significantly from BLIPS reconstruction, while the inset on the bottom right depicts the corresponding error map.

6.6 Conclusion and Future Work

This paper investigated a combination of shallow dictionary learning and deep supervised learning for MR image reconstruction that leverages the complementary nature of the two methods to bolster the quality of the reconstructed image. We verify this benefit by comparisons using a variety of metrics (including SSIM, PSNR, and HFEN) against strictly supervised learning-based reconstruction, reconstruction as initialization. We also investigate alternative approaches for combining the two forms of reconstruction. Our observations suggest that the primary benefits of including blind learning in the reconstruction pipeline are the preservation

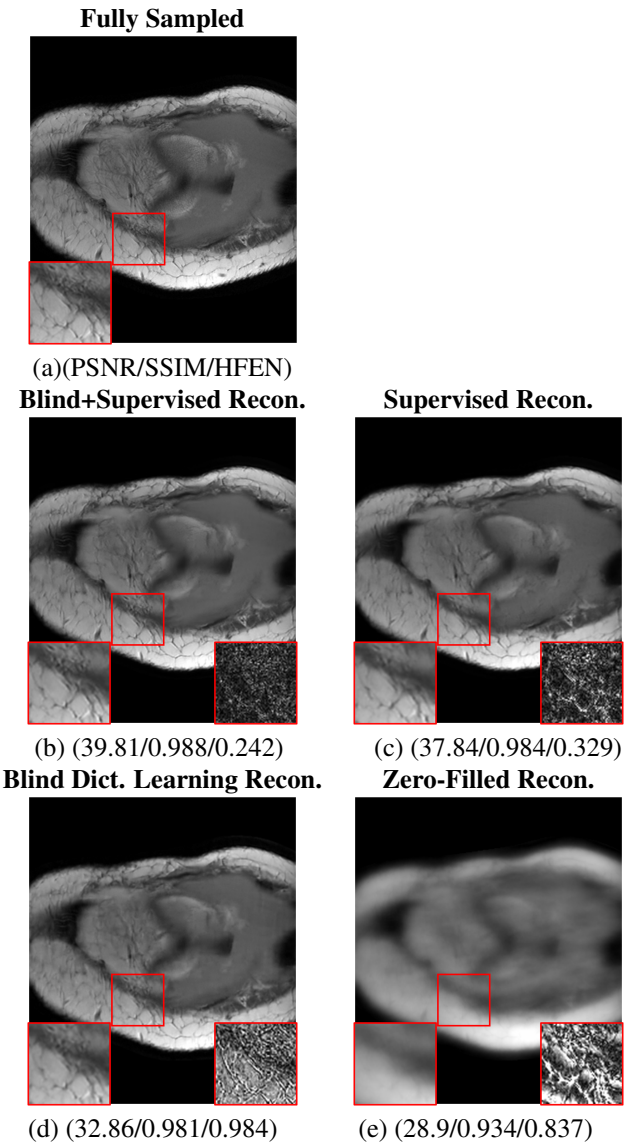


Figure 6.5: Comparison of reconstructions of a knee image using the proposed method versus strict supervised learning, blind dictionary learning, and zero-filled reconstruction for the 20$\times$ Poisson-disk undersampling mask depicted in Fig. 6.2b. Metrics listed below each reconstruction correspond to PSNR/SSIM/HFEN respectively. The inset panel on the bottom left in each image corresponds to regions of interest (indicated by the red bounding box in the image) in the image that benefits significantly from BLIPS reconstruction, while the inset on the bottom right depicts the corresponding error map.

of 'finer' details in the output image and robustness to the availability of training data.

In the future, we aim to apply our methods to non-Cartesian undersampling patterns such as radial and spiral patterns, and to other modalities. The generalizability of the method, especially with heterogeneous datasets, will be further explored. We observed some variation in the performance of our method to the imposed sparsity level in (6.2). More careful tuning of hyperparameters will be necessary to optimize the overall performance of such methods. Curiously, we also observed that using additional iterations of blind learning reconstruction in (6.14) adversely impacted the performance of our methods. The cause for this behavior is unknown (beyond oversmoothing), and needs further investigation. We also plan to investigate the benefits

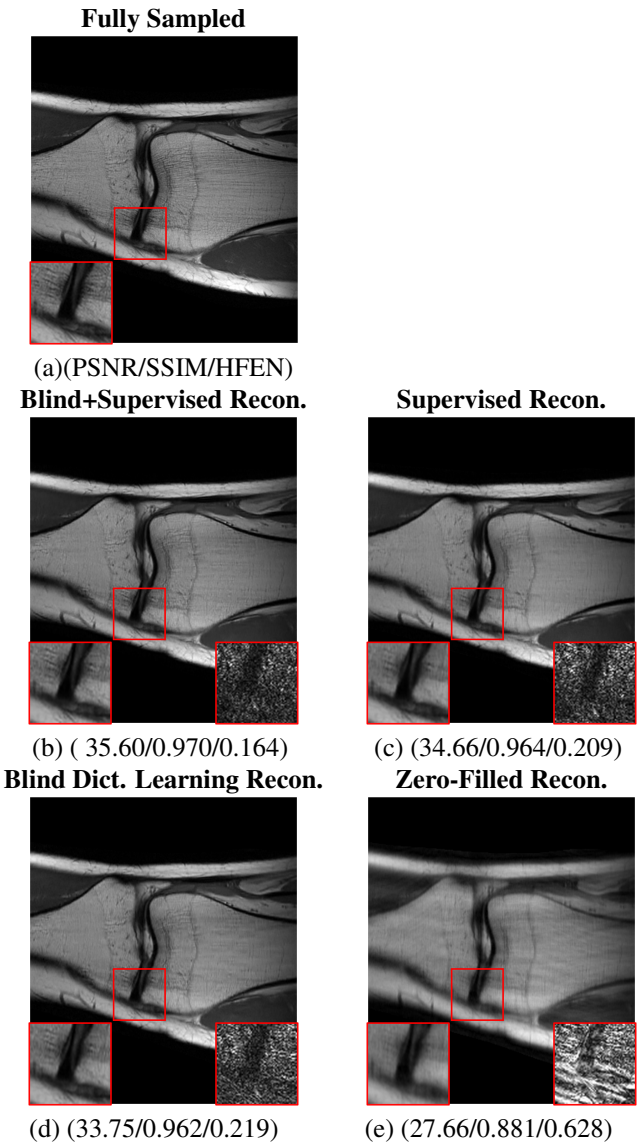


Figure 6.6: Comparison of reconstructions of a knee image using the proposed method versus strict supervised learning, blind dictionary learning, and zero-filled reconstruction for the random 1D undersampling masks ($\approx 4.5\times$). Metrics listed below each reconstruction correspond to PSNR/SSIM/HFEN respectively. The inset panel on the bottom left in each image corresponds to regions of interest (indicated by the red bounding box in the image) in the image that benefits significantly from BLIPS reconstruction, while the inset on the bottom right depicts the corresponding error map.

of multiple iterations of combined blind and supervised learning based reconstruction, extending the S+B+S approach considered here. Aside from the benefits of traditional 'handcrafted' priors in combination with supervised deep learning, from the perspective of learning only from measurements of the image being reconstructed, and then filling in the gaps with supervised data-driven learning, it would be interesting to study the combination of deep blind approaches[124, 125, 126, 127] with deep supervised learning.

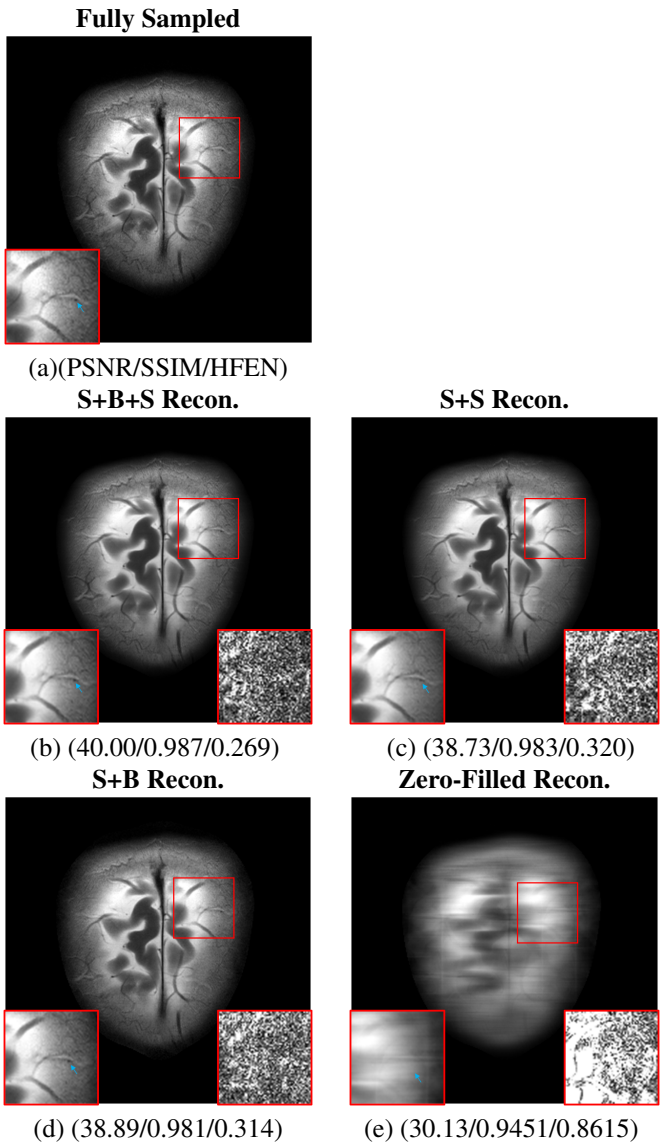


Figure 6.7: Comparison of reconstructions for two T2w brain images using the S+B+S learning reconstruction method proposed in (P3) versus cascaded S+S strict supervised learning-based reconstruction, S+B reconstruction, and zero-filled reconstruction for an $8\times$ equidistant undersampling mask. The S+B reconstruction depicts the output of one iteration of blind reconstruction initialized with a supervised reconstruction. Metrics listed below each reconstruction correspond to PSNR/SSIM/HFEN respectively.The inset panel on the bottom left in each image corresponds to regions of interest (indicated by the red bounding box in the image) in the image that benefits significantly from BLIPS reconstruction, while the inset on the bottom right depicts the corresponding error map. The blue arrows indicate the position of image detail that is present in the the BLIPS reconstruction, but not strict supervised learning-based reconstruction.

6.7 Supplementary Materials

6.7.1 Comparison with Non-Adaptive Dictionary-based Initialization for Supervised Learning

In this experiment, we fixed the dictionary $\boldsymbol{D}$ in (6.2) as an overcomplete inverse DCT matrix and did not update it. We then used the resulting reconstructed image to initialize the supervised reconstruction algorithm.

Table 6.6 compares the results of proposed blind+supervised learning versus non-adaptive dictionary-based initialization for supervised learning. The sampling pattern remains the same as in the previous case. 4198 slices were used for training, with 10% left for validation. The test set consisted of 500 slices.

Recon. Method	Fixed Dictionary+Supervised	Blind+Supervised
SSIM	0.945	0.946
PSNR (dB)	35.37	35.53
HFEN	0.452	0.443

Table 6.7: Comparison of performance of non-adaptive dictionary-based initialization for supervised learning-based reconstruction versus our proposed combined blind and supervised learning-based reconstruction, for the undersampling mask shown in Fig. 6.2b. The data set involved is the 4198/500 slices from the fastMRI knee dataset.

We surmise that the reason for relatively small improvements with blind learning over a fixed dictionary initialization in our proposed pipeline is due to the lack of proper parameter tuning during dictionary learning. One way to remedy this would be to vary the sparsity penalty weight, λ, across outer-iterations of dictionary learning-based reconstruction as is done in [119]. Furthermore, the initialization for blind dictionary learning was a zero-filled reconstruction, which can be detrimental to learning a 'good' dictionary. We expect that addressing these issues could further bolster the performance of BLIPS reconstruction.

6.7.2 Contribution of Residual Supervised Learning

To gain more insight into the mechanism of the proposed BLIPS reconstruction, we examined the residual component added to the blind dictionary learning-based reconstruction by the supervised learning-based reconstruction component. Essentially, we removed the blind learning output from the BLIPS reconstruction to study the contribution of the supervised module. Fig. 6.9 shows the contribution of the supervised learning component for Fig. 6.5.

We observe that the supervised learning module mainly contributes to removing left-over aliasing artifacts from the blind learning-based reconstruction, and also focuses on sharpening the details in the blind reconstruction. This observation reinforces the concept of complementarity of blind and supervised learning-based reconstruction.

6.7.3 Reconstruction Times

Table 6.8 lists the reconstruction times for the various methods proposed and compared to in this work. Strict supervised learning is the fastest, while the BLIPS approach in Fig. 6.1 (P1) is the slowest, because it requires several iterations of the SOUP-DIL algorithm [119], currently implemented in Matlab. This drawback may be remedied by providing a better initialization for dictionary learning and using GPUs for acceleration.

6.7.4 Failure Cases

We surmise that a primary limitation of our proposed BLIPS approach would occur when the blind reconstruction module would fail to reconstruct good quality images. This may happen, when there are not substantial patterns/features to learn from the undersampled data to aid in the blind reconstruction. In these settings, strict supervised reconstruction would fare better because of its ability to 'fill-in' gaps in acquisition

Recon Method	Recon Time (s)
S	1.2
B	170
CS+S	80.2
B+S	171.2
S+B+S	8.7

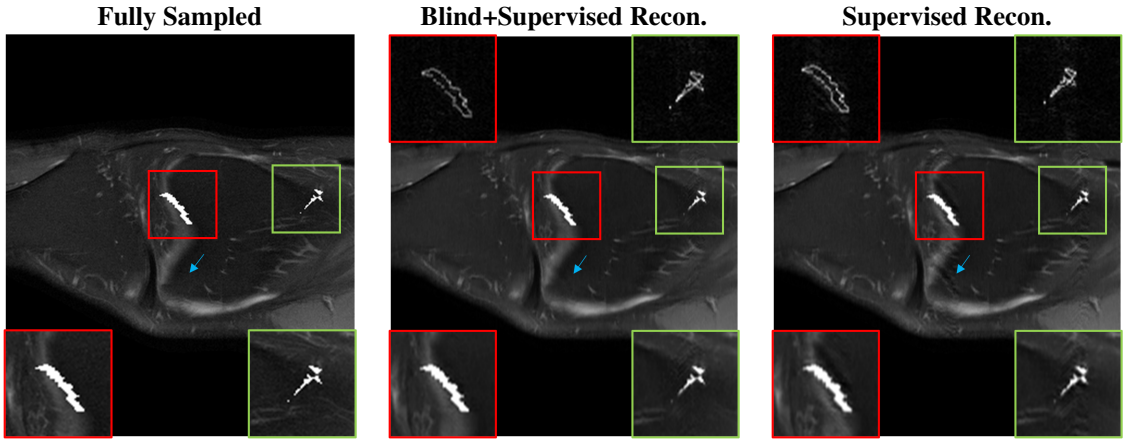
Table 6.8: Comparison of reconstruction times of various methods explored in our work

with learning from ensemble data, unlike in BLIPS where training datasets are used to learn features that refine the blind learning reconstruction. However, in the experiments that we have performed in this work, at various undersampling factors, we have not encountered a case in our test dataset where BLIPS learning under-performs strict supervised learning. We surmise that more vigorous undersampling is necessary to observe a failure case.

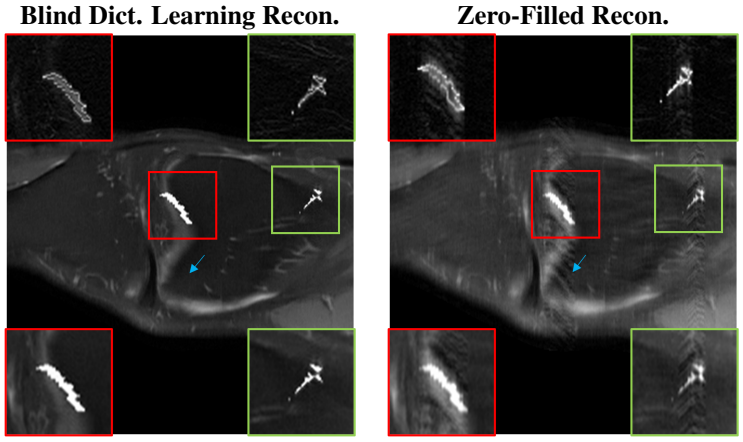
6.7.5 Effect of Dictionary Learning Compression on Fully-Sampled Images

The expression of signals as a sparse linear combination of dictionary atoms functions on the assumption that important details in the image can be preserved when using a sparse synthesis model, while the rest of the information that is not preserved is in effect noise/aliasing artifacts. Here, we study how the sparsity penalty for dictionary learning affects the features in a fully-sampled reconstruction when it is compressed using a blind-learned dictionary.

Fig. 6.10 shows a fully-sampled image and its synthesis using a blind-learned dictionary and sparse coefficients at $\lambda = 0.2$, which is the same sparsity penalty used for BLIPS dictionary learning in $(P1)$. We observe, from the comparison and the residues that most of the details from the ground truth are preserved in the dictionary-compressed image, though features in the latter appear to be smoothed out.



(a)(PSNR/SSIM/HFEN) (b) (35.00/0.960/0.2263) (c) (33.83/0.955/0.3020)



(d) (29.86/0.948/0.2985) (e) (30.07/0.915/0.6356)

Figure 6.8: Comparison of reconstructions of a knee image using the proposed method versus strict supervised learning for an image slice with artificially planted features. The undersampling mask was chosen to be random $\approx 4.5\times$. Metrics listed below each reconstruction correspond to PSNR/SSIM/HFEN respectively. The inset panels on the bottom in each image correspond to regions of interest (indicated by the red/green bounding boxes in the image) in the image that benefit significantly from BLIPS reconstruction, while the insets on the top depicts the corresponding error map. The blue arrows indicate the position of an aliasing artifact that is present in the zero-filled reconstruction and strict supervised learning, but not in the BLIPS reconstruction.

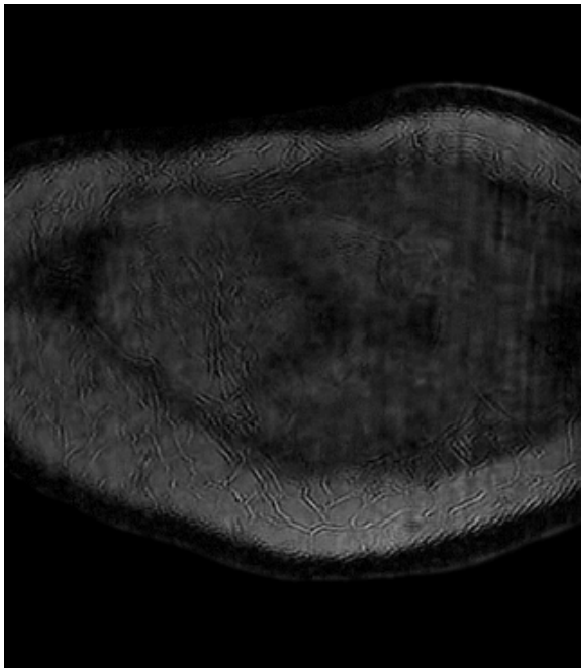
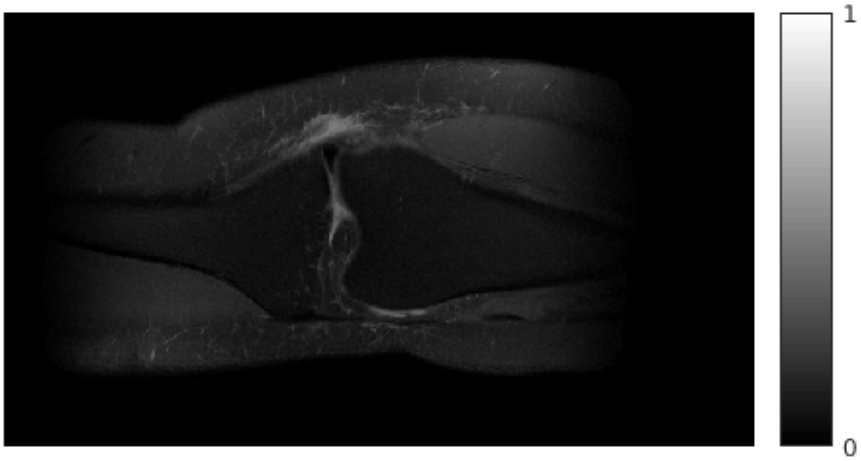
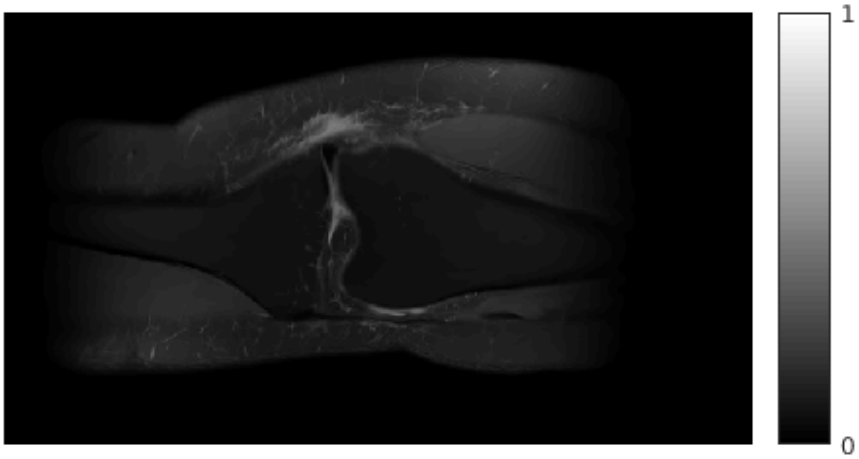


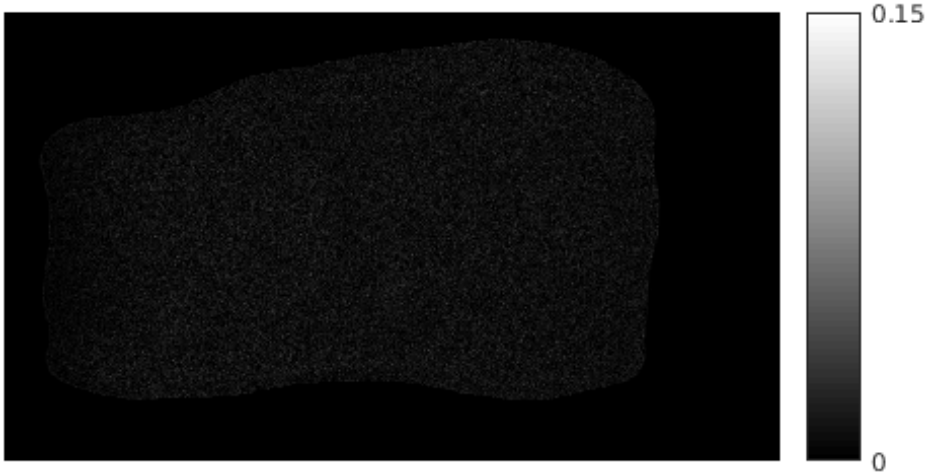
Figure 6.9: Residual contribution of the supervised learning module for the image in Fig. 6.5, obtained by removing the blind dictionary learning output from the BLIPS reconstructed image.



(a) Fully-Sampled Ground Truth



(b) Dictionary Representation (PSNR = 38.23 dB)



(c) Error Map

Figure 6.10: Comparision of (a) fully-sampled ground truth knee image, and (b) its synthesis from a blind learned dictionary and sparse coefficients at sparsity penalty $\lambda = 0.2$. We observe from the residues (c), that a majority of the features in the fully sampled image are preserved in the synthesis

CHAPTER VII

Future Work

This chapter lists some of the possible extensions to the work presented in the thesis.

- In Chapter III, we use a linear interpolation scheme to explore the space of labels to design an optimized labeling schedule. However, due to the time-costs associated with a grid search for the labeling schedule, as presented in our work, the space of labels is not explored thoroughly. This limitation would change if gradient-based methods were used to minimize our optimization cost. However, these methods have a tendency to converge to local minima. So, a better approach would be to first minimize our objective using a grid-search, and then use the optimized schedule to initialize a gradient-based minimization of our optimization objective.
- Also in Chapter III, the precision in perfusion estimates can be bolstered by attaching maximum weight to it during optimization. We need to study the extent to which this improves the perfusion estimates.
- In Chapter IV, we adopt VSI pulses to deal with the limitations of available perfusion information in multi-slice perfusion quantification using pCASL-MRF. An alternate strategy would be to use undersampling to reduce the acquisition time for every TR and use that time to acquire more information about using pCASL labeling pulses.
- Pertaining to the previous point, an avenue of further investigation would be to optimize how the sampling pattern varies across the acquired frames to aid the uniqueness of our fingerprints.
- In Chapter IV, we need to incorporate toggling of the arterial suppression module across TRs to obtain information about the blood volume fraction using the VSI pulses, and decide upon the fraction of acquisitions that have the arterial suppression turned off.
- In our current implementation of ASL-MRF optimization, we optimize the labeling duration/ post labeling delays first, and then proceed to optimize the label-control or arterial suppression order. We could attempt to do this jointly by optimizing for a $n \times 1$ vector at each TR instead of a single value. Since the latter are usually binary valued, we may need to approximate their values using a sigmoid function.
- We need to study the necessity of optimization of the VSI PLD schedule, i.e., whether the predicted improvement in precision due to optimization is significant.
- More *in-vivo* studies involving validation against other techniques like multi-PLD VSI-ASL or standard delay pCASL will allow us to ascertain the benefits of using VSI pulses in MRF-ASL.

- In chapter V, we discuss the possibility of using combined supervised and blind dictionary learning to denoise fixed-delay ASL time series, where there is a scarcity of available training data. This needs to be studied.
- An interesting avenue of research in Chapter V would be the undersampling/corruption level which would provide the highest benefits from the use of the Super-BReD algorithm. This would allow us to ascertain which applications would benefit the most from the Super-BReD framework.
- In Chapter VI, we discuss the possibility of exploring alternative algorithms for the blind learning module in BLIPS, including blind deep learning approaches. This needs to be investigated.
- A major setback of BLIPS learning in Chapter VI is the relatively high runtime of the SOUP-DIL algorithm used in the blind learning reconstruction module. Since the dictionary learning objective is only partially optimized (using a few iterations), we could adopt parallel dictionary learning algorithms to learn the dictionary and sparse coefficients to speed up this process. Parallel dictionary learning can also enable the blind reconstruction module to benefit from the use of GPUs.
- The selection of hyperparameters for regularization in dictionary learning and data-consistency update is an aspect that needs to be investigated more thoroughly. In our current work, we have adhered to values that to provide reasonable performance on a select few images. However, these regularization parameters can be learned during training or decided based on performance on a validation dataset.
- Another interesting alternative for the blind learning module would be to use transform learning as in [96] which would be significantly faster.
- Comparison and contrast of BLIPS reconstruction to self-distillation methods [130], which use the output yielded by a deep neural network as the targets for the subsequent training would make for an interesting study.
- Finally, a thoroughly planned study involving radiologist evaluation would help emphasize the improvements due to BLIPS reconstruction in a clinical setting.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] D. C. Alsop and J. A. Detre. Reduced Transit-Time Sensitivity in Noninvasive Magnetic Resonance Imaging of Human Cerebral Blood Flow. *Journal of Cerebral Blood Flow & Metabolism*, 16(6):1236–1249, 11 1996.

[2] D.S. Bolar, D.L. Levin, S.R. Hopkins, L.F. Frank, T.T. Liu, E.C. Wong, and R.B. Buxton. Quantification of regional pulmonary blood flow using ASL-FAIRER. *Magnetic Resonance in Medicine*, 55(6):1308–1317, 6 2006.

[3] J L Boxerman, K M Schmainda, R M Weisskoff, L.R. Schultz, J.P. Rock, J. Gutierrez, S.C. Patel, J. Ewing, and T. Mikkelsen. Relative cerebral blood volume maps corrected for contrast agent extravasation significantly correlate with glioma tumor grade, whereas uncorrected maps do not. *AJNR. American journal of neuroradiology*, 27(4):859–67, 4 2006.

[4] M. J. Brookes, P. G. Morris, P. a. Gowland, and S. T. Francis. Noninvasive measurement of arterial cerebral blood volume using Look-Locker EPI and arterial spin labeling. *Magnetic Resonance in Medicine*, 58(1):41–54, 2007.

[5] R B Buxton, L R Frank, E C Wong, B Siewert, S Warach, and R R Edelman. A general kinetic model for quantitative perfusion imaging with arterial spin labeling. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 40:383–396, 1998.

[6] Thibaut Capron, Thomas Troalen, Patrick J. Cozzone, Monique Bernard, and Frank Kober. Cine-ASL: A steady-pulsed arterial spin labeling method for myocardial perfusion mapping in mice. Part II. Theoretical model and sensitivity optimization. *Magnetic Resonance in Medicine*, 70(5):1399–1408, 2013.

[7] Yufen Chen, Danny J.J. Wang, and John A. Detre. Test-retest reliability of arterial spin labeling with common labeling strategies. *Journal of Magnetic Resonance Imaging*, 33(4):940–949, 4 2011.

[8] Ouri Cohen, Bo Zhu, and Matthew S. Rosen. MR fingerprinting Deep RecOnstruction NEtwork (DRONE). *Magnetic Resonance in Medicine*, 80(3):885–894, 9 2018.

[9] Cedric De Bazelaire, Neil M Rofsky, Guillaume Duhamel, M Dror Michaelson, Daniel George, and David C Alsop. Arterial spin labeling blood flow magnetic resonance imaging for the characterization of metastatic renal cell carcinoma(1). *Academic radiology*, 12(3):347–57, 3 2005.

[10] A R Deibler, J M Pollock, R A Kraft, H Tan, J H Burdette, and J A Maldjian. Arterial spin-labeling in routine clinical practice, part 1: technique and artifacts. *AJNR. American journal of neuroradiology*, 29(7):1228–34, 8 2008.

[11] John A. Detre, John S. Leigh, Donald S. Williams, and Alan P. Koretsky. Perfusion imaging. *Magnetic Resonance in Medicine*, 23(1):37–45, 1 1992.

[12] John A. Detre, Hengyi Rao, Danny J.J. Wang, Yu Fen Chen, and Ze Wang. Applications of arterial spin labeled MRI in the brain. *Journal of Magnetic Resonance Imaging*, 35(5):1026–1037, 5 2012.

[13] K. J. (Karl J.) Friston, John Ashburner, Stefan Kiebel, Thomas Nichols, and William D. Penny. *Statistical parametric mapping : the analysis of funtional brain images*. Elsevier/Academic Press, 2007.

[14] Robert A. Gatenby, Olya Grove, and Robert J. Gillies. Quantitative Imaging in Cancer Evolution and Ecology. *Radiology*, 269(1):8–14, 10 2013.

[15] G A R Y H Glover, T I E Iang L I, and D Avid R Ess. Image-Based Method for Retrospective Correction of Physiological Motion Effects in fMRI : RETROICOR. 167(March):162–167, 2000.

[16] Sven Haller, Greg Zaharchuk, David L. Thomas, Karl-Olof Lovblad, Frederik Barkhof, and Xavier Golay. Arterial Spin Labeling Perfusion of the Brain: Emerging Clinical Applications. *Radiology*, 281(2):337–356, 11 2016.

[17] V B Ho, J G Smirniotopoulos, F M Murphy, E J Rushing, K. Yamashita, E. Nagao, T. Shono, M. Mizoguchi, S. Nagata, T. Sasaki, S.O. Suzuki, T. Iwaki, K. Kobayashi, F. Mihara, and H. Honda. Radiologic-pathologic correlation: hemangioblastoma. *AJNR. American journal of neuroradiology*, 13(5):1343–52, 4 2008.

[18] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. 12 2014.

[19] John Kucharczyk, Jan Mintorovitch, Haleh S. Asgari, and Michael Moseley. Diffusion/perfusion MR imaging of acute cerebral ischemia. *Magnetic Resonance in Medicine*, 19(2):311–315, 6 1991.

[20] Anish Lahiri, Jeffrey A. Fessler, and Luis Hernandez-Garcia. Optimized Scan Design for ASL Fingerprinting and Multiparametric Estimation using Neural Network Regression. In *Proceedings of the 26th Annual Meeting of International Society for Magnetic Resonance in Medicine (ISMRM)*, page 309, 2018.

[21] Seung Soo Lee, Jae Ho Byun, Beom Jin Park, Seong Ho Park, Namkug Kim, Bumwoo Park, Jeong Kon Kim, and Moon-Gyu Lee. Quantitative analysis of diffusion-weighted magnetic resonance imaging of the pancreas: Usefulness in characterizing solid pancreatic masses. *Journal of Magnetic Resonance Imaging*, 28(4):928–936, 10 2008.

[22] Thomas T. Liu and Eric C. Wong. A signal processing model for arterial spin labeling functional MRI. *NeuroImage*, 24(1):207–215, 1 2005.

[23] Torben E. Lund, Kristoffer H. Madsen, Karam Sidaros, Wen-Lin Luo, and Thomas E. Nichols. Non-white noise in fMRI: Does modelling have an impact? *NeuroImage*, 29(1):54–66, 1 2006.

[24] Dan Ma, Vikas Gulani, Nicole Seiberlich, Kecheng Liu, Jeffrey L. Sunshine, Jeffrey L. Duerk, and Mark A. Griswold. Magnetic resonance fingerprinting. *Nature*, 495(7440):187–192, 3 2013.

[25] Jeanette A. Mumford, Luis Hernandez-Garcia, Gregory R. Lee, and Thomas E. Nichols. Estimation efficiency and statistical power in arterial spin labeling fMRI. *NeuroImage*, 33(1):103–114, 10 2006.

[26] Gopal Nataraj, Jon Fredrick Nielsen, and Jeffrey A. Fessler. Optimizing MR Scan Design for Model-Based T1, T2 Estimation from Steady-State Sequences. *IEEE Transactions on Medical Imaging*, PP(99):467–477, 2016.

[27] Gopal Nataraj, Jon-Fredrik Nielsen, Clayton Scott, and Jeffrey A. Fessler. Dictionary-Free MRI PERK: Parameter Estimation via Regression with Kernels. *IEEE Transactions on Medical Imaging*, 37(9):2103–2114, 9 2018.

[28] Tomoyuki Noguchi, Takashi Yoshiura, Akio Hiwatashi, Osamu Togao, Koji Yamashita, Eiki Nagao, and Hiroshi Honda. Arterial spin-labelin magnetic resonance imaging: The timing of regional maximal perfusion-related signal intensity revealed by a multiphase technique. *Japanese Journal of Radiology*, 30(2):137–145, 2 2012.

[29] Laura M. Parkes. Quantification of cerebral perfusion using arterial spin labeling: Two-compartment models. *Journal of Magnetic Resonance Imaging*, 22(6):732–736, 2005.

[30] Joanna E. Perthen, Mark Bydder, Khaled Restom, and Thomas T. Liu. SNR and functional sensitivity of BOLD and perfusion-based fMRI using arterial spin labeling with spiral SENSE at 3 T. *Magnetic Resonance Imaging*, 26(4):513–522, 5 2008.

[31] Anita Ramani, Jens H. Jensen, and Joseph A. Helpern. Quantitative MR Imaging in Alzheimer Disease. *Radiology*, 241(1):26–44, 10 2006.

[32] Giles Roditi, Jeffrey H. Maki, George Oliveira, and Henrik J. Michaely. Renovascular imaging in the NSF Era. *Journal of Magnetic Resonance Imaging*, 30(6):1323–1334, 12 2009.

[33] D I Rosenthal, N W Barton, K A McKusick, B R Rosen, S C Hill, F P Castronovo, R O Brady, S H Doppelt, and H J Mankin. Quantitative imaging of Gaucher disease. *Radiology*, 185(3):841–5, 12 1992.

[34] Pan Su, Deng Mao, Peiying Liu, Yang Li, Marco C. Pinho, Babu G. Welch, and Hanzhang Lu. Multiparametric estimation of brain hemodynamics with MR fingerprinting ASL. *Magnetic Resonance in Medicine*, 00(November):1–12, 2016.

[35] Pan Su, Deng Mao, Peiying Liu, Yang Li, Marco C Pinho, Babu G Welch, and Hanzhang Lu. Multiparametric Estimation of Brain Hemodynamics With MR Fingerprinting ASL. *Magn Reson Med*, 78:1812–1823, 2017.

[36] P. van Gelderen, J.A. de Zwart, and J.H. Duyn. Pittfalls of MRI measurement of white matter perfusion based on arterial spin labeling. *Magnetic Resonance in Medicine*, 59(4):788–795, 4 2008.

[37] Yingbing Wang, Tarik K. Alkasab, Ozden Narin, Rosalynn M. Nazarian, Rathachai Kaewlai, Jonathan Kay, and Hani H. Abujudeh. Incidence of Nephrogenic Systemic Fibrosis after Adoption of Restrictive Gadolinium-based Contrast Agent Guidelines. *Radiology*, 260(1):105–111, 7 2011.

[38] Jonathan M. Watts, Christopher T. Whitlow, and Joseph A. Maldjian. Clinical applications of arterial spin labeling. *NMR in Biomedicine*, 26(8):892–900, 8 2013.

[39] Eric C. Wong, Richard B. Buxton, and Lawrence R. Frank. A theoretical and experimental comparison of continuous and pulsed arterial spin labeling techniques for quantitative perfusion imaging. *Magnetic Resonance in Medicine*, 40(3):348–355, 9 1998.

[40] Joseph G. Woods, Michael A. Chappell, and Thomas W. Okell. A general framework for optimizing arterial spin labeling MRI experiments. *Magnetic Resonance in Medicine*, 81(4):2474–2488, 4 2019.

[41] Katherine L. Wright, Yun Jiang, Dan Ma, Douglas C. Noll, Mark A. Griswold, Vikas Gulani, and Luis Hernandez-Garcia. Estimation of perfusion properties with MR Fingerprinting Arterial Spin Labeling. *Magnetic Resonance Imaging*, 50:68–77, 7 2018.

[42] Jingyi Xie, Daniel Gallichan, Roger N. Gunn, and Peter Jezzard. Optimal design of pulsed arterial spin labeling MRI experiments. *Magnetic Resonance in Medicine*, 59(4):826–834, 2008.

[43] Bo Zhao, Justin P. Haldar, Congyu Liao, Dan Ma, Mark A. Griswold, Kawin Setsompop, and Lawrence L. Wald. Optimal Experiment Design for Magnetic Resonance Fingerprinting: Cram\'er-Rao Bound Meets Spin Dynamics. (Ml):1–15, 2017.

[44] Jinyuan Zhou, David A. Wilson, John A. Ulatowski, Richard J. Traystman, and Peter C. M. van Zijl. Two-Compartment Exchange Model for Perfusion Quantification Using Arterial Spin Tagging. *Journal of Cerebral Blood Flow & Metabolism*, 21(4):440–455, 2001.

[45] Hemant K. Aggarwal, Merry P. Mani, and Mathews Jacob. MoDL: Model-Based Deep Learning Architecture for Inverse Problems. *IEEE Transactions on Medical Imaging*, 38(2):394–405, feb 2019.

[46] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, nov 2006.

[47] David C. Alsop, John A. Detre, Xavier Golay, Matthias Günther, Jeroen Hendrikse, Luis Hernandez-Garcia, Hanzhang Lu, Bradley J. Macintosh, Laura M. Parkes, Marion Smits, Matthias J.P. Van Osch, Danny J.J. Wang, Eric C. Wong, and Greg Zaharchuk. Recommended implementation of arterial spin-labeled Perfusion mri for clinical applications: A consensus of the ISMRM Perfusion Study group and the European consortium for ASL in dementia. *Magnetic Resonance in Medicine*, 73(1):102–116, jan 2015.

[48] Chenglong Bao, Hui Ji, Yuhui Quan, and Zuowei Shen. norm based dictionary learning by proximal methods with global convergence. Technical report, 2014.

[49] Ke Lei, Morteza Mardani John M. Pauly, and Shreyas S. Vasanawala. Wasserstein GANs for MR imaging: From paired to unpaired training IEEE Trans. Med. Imaging, 2021 Jan;40(1):105-115.

[50] Richard B. Buxton, Lawrence R. Frank, Eric C. Wong, Bettina Siewert, Steven Warach, and Robert R. Edelman. A general kinetic model for quantitative perfusion imaging with arterial spin labeling Magnetic Resonance in Medicine, 1998 Sep;40(3):383-96.

[51] John A. Detre, Jonh S. Leigh, Donald S. Williams, and Alan P. Koretsky. Magnetic resonance imaging of perfusion using spin inversion of arterial water Proceedings of the National Academy of Sciences, 1992 Jan 1;89(1):212-6.

[52] Robert R. Edelman and Qun Chen. EPISTAR MRI: Mulitslice mapping of cerebral blood flow Magnetic Resonance in Medicine, 01 Dec 1998, 40(6):800-805

[53] Eric C. Wong, Matthew Cronin, Wen-Chau Wu, Ben Inglis, Lawrence R Frank and Thomas T Liu. Velocity-selective arterial spin labeling Magnetic Resonance in Medicine, 55:1334 –1341 (2006)

[54] Jia Guo,Shaurov Das and Luis Hernandez-Garcia. Comparison of velocity-selective arterial spin labeling schemes Magnetic Resonance in Medicine, 2021 Apr;85(4):2027-2039.

[55] Weiying Dai, Dairon Garcia, Cedric de Bazelaire, and David C. Alsop. Continuous flow driven inversion for arterial spin labeling using pulsed radiofrequency and gradient fields Magnetic Resonance in Medicine, 2008 Dec;60(6):1488-97.

[56] Anish Lahiri, Jeffrey A. Fessler and Luis Hernandez-Garcia. Perfusion Quantification using Velocity Selective Inversion pulses in a combined ASL-MRF Framework Proceedings of the 29th ISMRM Annual Meeting and Exhibition 2021, Abstract no. 1613

[57] Anish Lahiri, Guanhua Wang, Saiprasad Ravishankar and Jeffrey A. Fessler. Blind Primed Supervised (BLIPS) Learning for MR image reconstruction Proceedings of the 29th ISMRM Annual Meeting and Exhibition 2021, Abstract no. 1338

[58] Guang Yang, Simiao Yu, Hao Dong, Greg Slabaugh, Pier Luigi Dragotti, Xujiong Ye and others. DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction IEEE Trans. Med. Imaging, 2018

[59] Yashar Behzadi, Khaled Restom, Joy Liau, and Thomas T. Liu. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, 37(1):90–101, aug 2007.

[60] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal of Scientific Computing*, 20(1):33–61, jul 1998.

[61] David L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, apr 2006.

[62] Gary H. Glover, Tie-Qiang Li, and David Ress. Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magnetic Resonance in Medicine*, 44(1):162–167, jul 2000.

[63] Kuang Gong, Paul Han, Georges El Fakhri, Chao Ma, and Quanzheng Li. Arterial spin labeling MR image denoising and reconstruction using unsupervised deep learning. *NMR in Biomedicine*, 2019.

[64] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P. Recht, Daniel K. Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, 79(6):3055–3071, jun 2018.

[65] G. Harikumar, Christophe Couvreur, and Yoram Bresler. Fast optimal and suboptimal algorithms for sparse solutions to linear inverse problems. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, volume 3, pages 1877–1880, 1998.

[66] Luis Hernandez-Garcia, Anish Lahiri, and Jonas Schollenberger. Recent progress in ASL, 2019 Feb 15;187:3-16.

[67] Luis Hernandez-Garcia, Jon-Fredrik Nielsen, and Douglas C. Noll. Improved sensitivity and temporal resolution in perfusion FMRI using velocity selective inversion ASL. *Magnetic Resonance in Medicine*, 81(2):1004–1015, feb 2019.

[68] Anish Lahiri, Jeffrey A. Fessler, and Luis Hernandez-Garcia. Optimizing MRF-ASL scan design for precise quantification of brain hemodynamics using neural network regression. *Magnetic Resonance in Medicine*, 83(6):1979–1991, jun 2020.

[69] Michael Lustig, David Donoho, and John M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, dec 2007.

[70] Stephane G. Mallat and Zhifeng Zhang. Matching Pursuits With Time-Frequency Dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

[71] Y C Pati, R Rezaiifar, and P S Krishnaprasad. Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition. Technical report.

[72] Qin Qin and Peter C.M. van Zijl. Velocity-selective-inversion prepared arterial spin labeling. *Magnetic Resonance in Medicine*, 76(4):1136–1148, oct 2016.

[73] Saiprasad Ravishankar, Anish Lahiri, Cameron Blocker, and Jeffrey A. Fessler. Deep dictionary-transform learning for image reconstruction. In *Proceedings - International Symposium on Biomedical Imaging*, volume 2018-April, pages 1208–1212. IEEE Computer Society, may 2018.

[74] Saiprasad Ravishankar, Raj Rao Nadakuditi, and Jeffrey A. Fessler. Efficient Sum of Outer Products Dictionary Learning (SOUP-DIL) and Its Application to Inverse Problems. *IEEE Transactions on Computational Imaging*, 3(4):694–709, apr 2017.

[75] Ron Rubinstein, Student Member, Michael Zibulevsky, Michael Elad, and Senior Member. Double Sparsity: Learning Sparse Dictionaries for Sparse Signal Approximation. *IEEE Transactions On Signal Processing*, 58(3), 2010.

[76] Jo Schlemper, Jose Caballero, Joseph V. Hajnal, Anthony N. Price, and Daniel Rueckert. A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction. *IEEE Transactions on Medical Imaging*, 37(2):491–503, feb 2018.

[77] Abd-Krim Seghouane and Muhammad Hanif. A Sequential Dictionary Learning Algorithm With Enforced Sparsity. 2015 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 3876-3880

[78] Bihan Wen, Yanjun Li, and Yoram Bresler. Image recovery via transform learning and low-rank modeling: The power of complementary regularizers. *IEEE Transactions on Image Processing*, 29:5310–5323, 2020.

[79] Mehrdad Yaghoobi, Thomas Blumensath, and Mike E. Davies. Dictionary learning for sparse approximations with the majorization method. *IEEE Transactions on Signal Processing*, 57(6):2178–2191, 2009.

[80] Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J. Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, Marc Parente, Krzysztof J. Geras, Joe Katsnelson, Hersh Chandarana, Zizhao Zhang, Michal Drozdzal, Adriana Romero, Michael Rabbat, Pascal Vincent, Nafissa Yakubova, James Pinkerton, Duo Wang, Erich Owens, C. Lawrence Zitnick, Michael P. Recht, Daniel K. Sodickson, and Yvonne W. Lui. fastMRI: An Open Dataset and Benchmarks for Accelerated MRI. nov 2018.

[81] Anish Lahiri, Saiprasad Ravishankar and Jeffrey A. Fessler. Combining Supervised and semi-Blind Dictionary (Super-BReD) Learning for MRI Reconstruction *ISMRM 28th Annual Meeting and Exhibition*, Abstract no. 1989, 2020.

[82] Binnewijzend MA, Benedictus MR, Kuijer JP, et al. Cerebral perfusion in the predementia stages of Alzheimer's disease. *Eur Radiol. 2016*, 26(2):506-514. doi:10.1007/s00330-015-3834-9.

[83] Roquet D, Sourty M, Botzung A, Armspach JP, Blanc F. Brain perfusion in dementia with Lewy bodies and Alzheimer's disease: an arterial spin labeling MRI study on prodromal and mild dementia stages. Alzheimers Res Ther. 2016,8:29. Published 2016 Jul 12. doi:10.1186/s13195-016-0196-8

[84] Wolters FJ, Zonneveld HI, Hofman A, van der Lugt A, Koudstaal PJ, Vernooij MW and Ikram MA Cerebral Perfusion and the Risk of Dementia: A Population-Based Study Circulation. 2017,136:719–728

[85] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007.

[86] Y. Liu *et al.*, "Balanced sparse model for tight frames in compressed sensing magnetic resonance imaging," *PLOS ONE*, vol. 10, no. 4, p. e0119584, Mar. 2015.

[87] E. M. Eksioglu, "Decoupled algorithm for MRI reconstruction using nonlocal block matching model: BM3D-MRI," *J. Math. Imaging Vision*, vol. 56, no. 3, pp. 430–440, Nov. 2016.

[88] M. Jacob, M. P. Mani, and J. C. Ye, "Structured low-rank algorithms: Theory, magnetic resonance applications, and links to machine learning," *IEEE Sig. Process. Mag.*, vol. 37, no. 1, pp. 54–68, Feb. 2020.

[89] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmon. Anal.*, vol. 31, no. 1, pp. 59–73, 2011.

[90] B. Wen, Y. Li, and Y. Bresler, "Image recovery via transform learning and low-rank modeling: The power of complementary regularizers," *IEEE Trans. Image Proc.*, vol. 29, pp. 5310–5323, 2020.

[91] W. K. Pratt, H. C. Andrews, and J. Kane, "Hadamard transform image coding," *Proc. of the IEEE*, vol. 57, no. 1, pp. 58–68, 1969.

[92] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Probl.*, vol. 23, no. 3, p. 947, 2007.

[93] E. Kobler, A. Effland, K. Kunisch, and T. Pock, "Total deep variation: A stable regularizer for inverse problems," *arXiv preprint arXiv:2006.08789*, 2020.

[94] S. Roth and M. J. Black, "Fields of experts: A framework for learning image priors," in *2005 IEEE Comp. Soc. Conf. on Comp. Vis. and Pat. Recogn. (CVPR'05)*, vol. 2. IEEE, 2005, pp. 860–867.

[95] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Trans. Med. Imaging*, vol. 30, no. 5, pp. 1028–1041, May 2011.

[96] S. Ravishankar and Y. Bresler, "Learning sparsifying transforms," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1072–1086, 2012.

[97] S. Ravishankar and Y. Bresler, "Data-driven learning of a union of sparsifying transforms model for blind compressed sensing," *IEEE Trans. Comput. Imaging*, vol. 2, no. 3, pp. 294–309, Sep. 2016.

[98] J. Caballero, D. Rueckert, and J. V. Hajnal, "Dictionary learning and time sparsity in dynamic MRI," in *Int. Conf. on Med. Image Comput. and Comp.-Assist. Interv. (MICCAI 2012)*. Springer, 2012, pp. 256–263.

[99] D. Weller, "Reconstruction with dictionary learning for accelerated parallel magnetic resonance imaging," in *Proc. of the IEEE Southwest Symp. on Image Anal. and Interp.*, vol. 2016-April. Institute of Electrical and Electronics Engineers Inc., Apr. 2016, pp. 105–108.

[100] S. G. Lingala and M. Jacob, "Blind compressive sensing dynamic MRI," *IEEE Trans. Med. Imaging*, vol. 32, no. 6, pp. 1132–1145, 2013.

[101] S. Gleichman and Y. C. Eldar, "Blind compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6958–6975, 2011.

[102] J. Schlemper, C. Qin, J. Duan, R. M. Summers, and K. Hammernik, "Sigma-net: Ensembled iterative deep neural networks for accelerated parallel MR image reconstruction," *arXiv preprint arXiv:1912.05480*, 2019.

[103] S. Ravishankar, A. Lahiri, C. Blocker, and J. A. Fessler, "Deep dictionary-transform learning for image reconstruction," in *Proc.- 2018 IEEE 15th Int. Symp. on Biomed. Imag. (ISBI 2018)*, vol. 2018-April. IEEE Computer Society, May 2018, pp. 1208–1212.

[104] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model-based deep learning architecture for inverse problems," *IEEE Trans. Med. Imaging*, vol. 38, no. 2, pp. 394–405, Feb. 2019.

[105] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for dynamic MR image reconstruction," *IEEE Trans. Med. Imaging*, vol. 37, no. 2, pp. 491–503, Feb. 2018.

[106] K. Hammernik *et al.*, "Learning a variational network for reconstruction of accelerated MRI data," *Magn. Reson. Med.*, vol. 79, no. 6, pp. 3055–3071, Jun. 2018.

[107] G. Yang *et al.*, "DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction," *IEEE Trans. Med. Imaging*, vol. 37, no. 6, pp. 1310–1321, 2017.

[108] T. Eo, Y. Jun, T. Kim, J. Jang, H. J. Lee, and D. Hwang, "KIKI-net: Cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images," *Magn. Reson. Med.*, vol. 80, no. 5, pp. 2188–2201, Nov. 2018.

[109] D. Lee, J. Yoo, and J. C. Ye, "Deep residual learning for compressed sensing MRI," in *Proc.- 2017 IEEE 14th Int. Symp. on Biomed. Imag. (ISBI 2017)*. IEEE, 2017, pp. 15–18.

[110] M. Mardani *et al.*, "Deep generative adversarial neural networks for compressive sensing MRI," *IEEE Trans. on Med. Imaging*, vol. 38, no. 1, pp. 167–179, 2018.

[111] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Int. Conf. on Med. Image Comput. and Comp.-Assist. Intervent. (MICCAI 2015)*. Springer, 2015, pp. 234–241.

[112] V. Antun, F. Renna, C. Poon, B. Adcock, and A. C. Hansen, "On instabilities of deep learning in image reconstruction and the potential costs of AI," *PNAS*, vol. 117, no. 48, pp. 30 088–95, Dec. 2020.

[113] J. Zbontar *et al.*, "fastMRI: An open dataset and benchmarks for accelerated MRI," *arXiv preprint arXiv:1811.08839*, 2018.

[114] D. C. Van Essen *et al.*, "The WU-Minn human connectome project: An overview," *Neuroimage*, vol. 80, pp. 62–79, 2013.

[115] T. M. Quan, T. Nguyen-Duc, and W. K. Jeong, "Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss," *IEEE Trans. Med. Imaging*, vol. 37, no. 6, pp. 1488–1497, Jun. 2018.

[116] Z. Ramzi, P. Ciuciu, and J.-L. Starck, "XPDNet for MRI reconstruction: An application to the fastMRI 2020 brain challenge," *arXiv preprint arXiv:2010.07290*, 2020.

[117] M. Uecker *et al.*, "ESPIRiT - An eigenvalue approach to auto-calibrating parallel MRI: where SENSE meets GRAPPA," *Mag. Reson. Med.*, vol. 71, no. 3, pp. 990–1001, Mar. 2014.

[118] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Trans. Med. Imaging*, vol. 30, no. 5, pp. 1028–1041, May 2011.

[119] S. Ravishankar, R. R. Nadakuditi, and J. A. Fessler, "Efficient sum of outer products dictionary learning (SOUP-DIL) and its application to inverse problems," *IEEE Trans. Comput. Imaging*, vol. 3, no. 4, pp. 694–709, Apr. 2017.

[120] S. Yu, B. Park, and J. Jeong, "Deep iterative down-up CNN for image denoising," in *Proc. of the IEEE Conf. on Comp. Vis. and Pat. Recogn. (CVPR) Workshops*, 2019, pp. 0–0.

[121] A. Abdelhamed, R. Timofte, and M. S. Brown, "NTIRE 2019 challenge on real image denoising: Methods and results," in *Proc. of the IEEE Conf. on Comp. Vis. and Pat. Recogn. (CVPR) Workshops*, 2019, pp. 0–0.

[122] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[123] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[124] G. Oh, B. Sim, H. Chung, L. Sunwoo, and J. C. Ye, "Unpaired deep learning for accelerated MRI using optimal transport driven CycleGAN," *IEEE Trans. Comput. Imaging*, vol. 6, pp. 1285–1296, 2020.

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

[126] J. I. Tamir, S. X. Yu, and M. Lustig, "Unsupervised deep basis pursuit: Learning inverse problems without ground-truth data," *arXiv preprint arXiv:1910.13110*, 2019.

[127] Z. Ke, J. Cheng, L. Ying, H. Zheng, Y. Zhu, and D. Liang, "An unsupervised deep learning method for multi-coil cine MRI," *Phys. Med. Biol.*, vol. 65, no. 23, p. 235041, 2020.

[128] V. Monga, Y. Li and Y. C. Eldar, "Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing," in IEEE Sign. Proc. Maga., vol. 38, no. 2, pp. 18-44, March 2021, doi: 10.1109/MSP.2020.3016905.

[129] A. Lahiri, G. Wang, S. Ravishankar, J. A. Fessler,"Blind Primed Supervised (BLIPS) Learning for MR Image Reconstruction ," arXiv:2104.05028 [eess.IV], 2021

[130] H. Mobahi, M. Farajtabar, P. L. Bartlett, "Self-Distillation Amplifies Regularization in Hilbert Space," arXiv:2002.05715 [cs.LG],2021