Self-Supervised Physics-Guided Deep Learning for Solving Inverse Problems in Imaging

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Burhaneddin Yaman

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Prof. Mehmet Akçakaya, advisor

March, 2022

Acknowledgements

I would like to start by expressing my sincerest gratitude to my advisor Prof. Mehmet Akçakaya for his constant guidance and support in my Ph.D. journey and overall life! Prof. Akçakaya helped me to grow as a researcher by always helping me to develop innovative research directions to tackle main challenges in medical imaging using state-of-the-art machine learning tools. Prof. Akçakaya has guided me to think from a broader perspective while conducting research and given me the confidence to be an independent researcher. Without Prof. Akçakaya, it would not be possible to complete this thesis. Aside from being a great advisor, Prof. Akçakaya has always been very caring and helpful in any questions or problems I have faced. Dear Prof. Akçakaya, I am very grateful to have you as my advisor, and thank you so much for always supporting me!

I would like to extend my sincere thanks to Prof. Jeffrey Fessler, Prof. Georgios Giannakis, Prof. Mingyi Hong, and Prof. Youssef Saad for serving on my thesis committee. I would also like to thank my internship mentors Michael Hansen and Matthew Lungren as well as my co-intern friends Annie Zhang and Ruiyang Zhao for the amazing internship experience at Microsoft. Special thanks to my mentor Mariappan Nadar for the internship experience at Siemens Healthineers.

A big thanks to my labmates throughout these years Chi, Burak, Hongyi, Toygan, and Merve. I would like to thank my alumni-labmate and friend Seyed Amir Hossein Hosseini. Amir Hossein, you were the best labmate I could ever imagine of. I enjoyed conducting research, exchanging and developing ideas with you. I am so lucky to have you as a friend and appreciate your help and support in my life. Also, thanks a lot for all the delicious Persian meals you have cooked over the years! I would also like to thank my lab-sharing friend Adel Elmahdy. Adel, you have been a great friend who is very kind and helpful. You always helped me in any problem I have faced even though you were very busy. I am really grateful for your friendship and support over the years.

I would like to thank my dear friends Bosco Cheng, Selim Engin, and Fatih Evren. Bosco, you are the best roommate I have had in my life. I truly enjoyed living with you over the last 5 years and I will miss these days a lot. You were the perfect roommate and friend who shared a similar understanding of life, humor, and hobbies. You have always been a great friend. Whenever I had a problem, you were there to talk and help me. I am very happy to have such a friend and roommate throughout my Ph.D. You made Ph.D. more fun! Selim and Fatih, thank you guys so much for always being with me. I really enjoyed our talks, hiking trips, tea, and movie sessions. You were always there for me to share the problems I have faced and celebrate the happy moments together. Also thank you Fatih for cooking amazing Turkish foods for us over the years!

I would like to thank Yingxue. Before knowing you, my life was an ordinary Ph.D. life. Since the moment I met you, my life has become more meaningful and fun! We went through hard times together and overcame the problems together. Without you, Ph.D. would be much tougher with a lot of stress. You have enlightened my life. You always support and believe in me and help me to learn and change positively. I am so happy to share my life with you.

I would like to thank my best friend Enes Uyar, a great architect:). I have known Enes since the first day of my university life in Istanbul. Enes is the funniest, sincerest, and kindest friend I have met in my life. Enes, you have been a brother to me. I don't know how my life would be without you. You are always there for me in anything in my life. Any problem or happy moment I have in my life; you are there to share with me. You always helped me to be a better person in this life. I am very thankful to have known you! I would like to thank also my other friends Yan Luo, Mehmet Arslan, Mohamed Elshraiy, Mutaz, Ezgi, Ebtehal, Daisy, Xiangyi, Mucahit, Mustafa, Samed, Husrev, Mehran, Mishfad, Hadi, and many more I have met throughout my life.

Finally, I would like to express my deepest appreciation to my Family, Charde, to whom I owe everything. I would like to thank my Mum, Taybet, and Dad, Abdi. Daye and Yabo, thank you for your unbelievable love and support throughout all my life, I am forever grateful to you. I would like to thank my dear brother and sisters, Agit, Hikmet, Ayhan, Fikret, Ergin, Şükrü, Maruf, Beraat, Vedat, Sena and Üsame. I am so lucky to have each one of you. You are the reason I can smile, be happy and positive everyday. Any hardship I face, I know I am not alone and I can overcome it with your

constant help. Any success or happy things in my life makes me truly happy because I can share and celebrate together with you. You have always supported and believed in me. I wouldn't be where I am today without the love and support of you. No matter how much I write or what I say, it would not suffice my appreciation to have you as my family. My dear family, I love you so much...

Burhaneddin Yaman, Minneapolis, March 2022

Dedication

Dedicated to my dear family, Charde.

Abstract

Inverse problems in computational imaging seek to recover an unknown image of interest from observed measurements acquired using a known forward model. These inverse problems are often ill-conditioned, requiring some form of regularization. The corresponding objective function for inverse problems in computational imaging can often be solved using iterative optimization approaches that alternate between two sub-problems that enforce data consistency and promote the regularization approach, respectively. Such inverse problems arise in a multitude of imaging modalities, in particular in magnetic resonance imaging (MRI), which is the main application area for this thesis. Lengthy scan times remain a challenge in MRI, thus accelerating MRI scans has remained an open research problem over decades. Conventional accelerated MRI techniques acceleration rate is limited by noise amplification and residual artifacts. Recently, deep learning has emerged as an alternative approach for accelerated MRI. Among deep learning techniques, physics-guided deep learning (PG-DL) has drawn great interest, as it incorporates the known physical forward model into the network architecture. PG-DL unrolls a conventional iterative algorithm for solving a regularized least squares problem for a fixed number of iterations, and replaces the proximal operation corresponding to the regularizer implicitly with neural networks. These unrolled networks are trained end-to-end, with the goal of minimizing the difference between the network output and the corresponding reference data. Most of the existing deep learning approaches in MRI reconstruction are based on supervised learning, which requires ground-truth/ fully-sampled data for training. However, acquisition of fully-sampled data is infeasible in many applications due to physiological constraints, such as organ motion, or physical constraints, such as signal decay. In several other scenarios, such as high-resolution anatomical brain imaging, it is impractical to acquire fully-sampled datasets as the scan time becomes extremely lengthy. Therefore, enabling the training of PG-DL reconstruction without fully-sampled data is essential for the integration of deep learning reconstruction into clinical MRI practice.

The present thesis introduces novel frameworks to enable the training of deep learning reconstruction methods for inverse imaging problems in the absence of ground-truth/fully-sampled data. First, we introduce self-supervised learning via data undersampling (SSDU) approach to enable database training without fully-sampled data. Succinctly, SSDU partitions available measurements into two disjoint sets. One of these sets is used in the data consistency units of the unrolled network, while the other is used to define the loss in the measurement domain. Subsequently, we extend SSDU for processing 3D datasets and provide solutions for GPU memory constraints and data scarcity issues faced in 3D processing. To cope with potential performance degradation at very high acceleration rates, we develop a multi-mask self-supervised learning approach, which retrospectively splits available measurements into multiple 2-tuples of disjoint sets to perform training and define a loss function. Furthermore, we introduce a zero-shot self-supervised learning approach to enable training from a single scan/sample without any external training databases. ZS-SSL partitions the available measurements from a single scan into three disjoint sets. Two of these sets are used to enforce data consistency and define loss during training for self-supervision, while the last set serves to self-validate, establishing an early stopping criterion. Finally, we introduce a self-supervised learning algorithm for referenceless image denoising. Self-supervised deep learning algorithms split the pixels for each image into two disjoint sets to perform training and defining loss. In existent self-supervised denoising approaches which are purely data-driven, the set of pixels used as input to the network is not re-utilized in the end-to-end training since the network is only comprised of a neural network. Reusing the pixels within the network would promote consistency with acquired measurements, thus leading to a more robust and improved denoising performance. To tackle this challenge, we build upon existent self-supervised learning algorithms and recast the denoising problem into a regularized image inpainting framework which allows use of algorithm unrolling for denoising.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Inverse problems in computational imaging aim to recover an unknown image of interest from acquired measurements. These measurements are acquired based on a known forward imaging model characterizing the domain knowledge. Recovering the unknown image from the acquired data is often an ill-posed problem, requiring additional regularization. Such regularized objective functions for inverse imaging problems can often be solved using iterative optimization approaches that alternate between data consistency and regularization sub-problems. These inverse problems arise in a multitude of imaging modalities, such as magnetic resonance imaging (MRI), computed tomography, microscopy. The main application area for this thesis is MRI, but developed frameworks are applicable to a broad range of imaging modalities.

MRI is a non-invasive, radiation-free medical imaging modality that provides excellent soft tissue contrast for diagnostic purposes. However, the data acquisition process in MRI is inherently slow, as imaging in MRI is a sequential process that requires repeating data acquisition to image a volume. These lengthy acquisition times remain a main limitation in MRI and an open research problem, often requiring trade-offs between scan time, resolution and signal-to-noise ratio. However, improved resolutions are highly desirable to better delineate small structures not visible with current technology. Higher resolutions can be achieved by acquiring more data, which may lead to prohibitively long scan times or may be physically impossible in some cases such as signal decay or organ motion. Hence, accelerating MRI acquisitions has been a main motivation for numerous studies over the decades [1–10].

In clinical MRI systems, multi-coil receivers are used during data acquisition. Acquired MRI data is in Fourier domain, which is also known as k-space. In presence of fully-sampled k-space data, the desired image of interest for clinical evaluations is obtained by applying inverse Fourier transform, followed by coil combination. Accelerating MRI process requires performing data acquisition in a sub-sampled manner, often below the Nyquist rate. A direct image reconstruction approach using inverse Fourier transform on such sub-sampled data leads to aliasing artifacts in the reconstructed image. Therefore, accelerated MRI techniques use redundancies in the acquisition system or the images to remove the resulting aliasing artifacts during reconstruction. The image reconstruction from sub-sampled measurements is often an ill-posed problem, hence regularizers that induce prior information is utilized during reconstruction. Possible choices for the regularizer include total variation [11–13], ℓ_1-norm of wavelet coefficients [4,14,15], sparsity in adaptive transform domains [6,16], and more recently neural networks [7, 9, 17, 18].

Parallel imaging (PI) is the most clinically used method for accelerated MRI, and exploits the redundancies between these coils for reconstruction. PI uses linear methods to recover the image from the undersampled measurements. The clinical implementation of PI relies on a so-called uniform undersampling pattern, where acquired k-space lines are equispaced. This leads to coherent foldover artifacts in the aliased images, but also allows a non-iterative unaliasing solution in the image domain [1]. We note that the problem can also be solved in k-space via interpolation [2, 3]. PI typically suffers from noise amplification at high acceleration rates [1, 2].

Compressed sensing (CS) is another conventional accelerated MRI technique that exploits the compressibility of images in sparsifying transform domains [4–6, 14, 19, 20], and is commonly used in combination with PI. In CS reconstruction, a sparsity-inducing regularizer is used. A popular choice is the ℓ_1 norm of transform-domain coefficients, in a fixed linear sparsifying domain, such as wavelets [4]. CS-based approaches utilize random undersampling patterns to generate incoherent aliasing artifacts. One of the main limitations of CS approaches is blurring and residual artifacts seen at high acceleration rates [21].

Recently, deep learning (DL), which has shown great performance in various fields

such as computer vision and medical imaging, has recently emerged as an alternative approach for high-quality accelerated MRI. DL-based MRI reconstruction algorithms can be roughly divided into two categories, purely data-driven and physics-guided [18,22]. In purely data-driven approaches, a mapping between the undersampled k-space/aliased image to full k-space/artifact-free image is learned using neural networks [8, 23–26]. Hence, neural networks aims to learn the whole inverse problem solution without the forward operator, which models the physical process underlying the problem. This leads to very fast runtime for data-driven approaches, but it may face issues with generalizability especially when forward operator varies among test samples. Moreover, these approaches often require very large datasets, that can be challenging to acquire in many applications, to learn the underlying representation without overfitting [22].

In the physics-guided deep learning (PG-DL) methods, the forward encoding operator modeling the physical process of the problem is taken into account to solve an inverse problem based on a regularized least squares objective function [7, 9, 17, 27–32]. These techniques are based on the algorithm unrolling concept, which unrolls a conventional iterative reconstruction algorithm that alternates between data consistency and regularizer units for a fixed number of iterations. The regularizer units in the unrolled networks pare implicitly implemented with neural networks. The parameters of the unrolled network can be different [7, 29] or shared [9, 31] across the unrolled iterations. The unrolled networks are then trained end-to-end with a loss function that aims to minimize a reconstruction error with respect to an available reference [18]. Incorporating the forward operator for solving the inverse problem leads to a more robust network architecture and reasonably sized datasets suffice for achieving a good quality reconstruction [22].

The end-to-end training of physics-guided methods are typically performed in a supervised manner by using fully-sampled data as a reference. While supervised PG-DL provides state-of-the-art reconstruction results, it is infeasible to acquire fully-sampled datasets in many applications. One such impediment relates to imaging moving organs, such as in real-time MRI, where data acquisition needs to be performed in a short period of time [33]. In some other applications such as diffusion MRI with echo-planar imaging, the signal decays quickly hindering the acquisition of the fully sampled data [34]. In several other scenarios such as whole-heart coronary MRI or high-resolution

anatomical brain imaging, it is impractical to acquire fully-sampled datasets, as the scan time becomes extremely lengthy. In the absence of fully-sampled data, supervised deep learning approaches become inoperative. Hence, it is highly desirable to be able to train the state-of-the-art physics-guided deep learning reconstruction methods without fully-sampled data. In this thesis, we introduce novel methods detailed below to tackle such challenges.

In Chapter 2, we introduce a self-supervised learning approach that enables training of physics-guided neural networks without fully-sampled reference data [10, 35]. Our self-supervised learning approach, which is called Self-Supervised Learning via Data Undersamping (SSDU) splits the acquired k-space measurements for each scan into two disjoint sets. One of these sets is used during training to enforce data consistency within the network according to the MRI physics acquisition model, while the other set is used to define a k-space loss function to measure the reconstruction quality of the model. Thus, SSDU enables end-to-end training and evaluation of the network using only the acquired measurements. The extensive experiments on knee and brain datasets show that SSDU achieves state-of-the-art performance.

An extension of the SSDU for enabling training of 3D datasets is introduced in Chapter 3 [36]. In particular, we apply SSDU to late gadolinium enhancement (LGE) cardiac MRI (CMR) datasets. LGE is the clinical gold standard for identification of myocardial scar and fibrosis. The 3D LGE CMR imaging offers improved SNR and spatial resolution compared to 2D as 3D processing enables the capturing interdimensional interactions [37]. However, training of 3D large datasets is challenging due to several factors including the absence of fully-sampled datasets, lack of samples and GPU memory constraints. In order to enable the training in absence of fully-sampled data, we extend the self-supervised learning study introduced in Chapter 2 to 3D setting, using the same two-way splitting strategy to perform training and defining the loss. Fitting large volume of 3D datasets is challenging due to limitations in GPU memories. To tackle these challenges, we partition the large volumes for each subjects into smaller 3D sub-volumes. Moreover, the partitioning of a large volume into sub-volumes lead to an augmented dataset, which also helps to overcome the lack of datasets.

While SSDU enables training networks without fully-sampled data, its performance starts to degrade at very high acceleration rates due to scarcity of acquired data. Hence,

a more efficient usage of acquired undersampled data is essential for enhanced reconstruction quality at such high acceleration rates. In Chapter 4, we introduce a multi-mask SSDU technique to enhance reconstruction quality of self-supervised learning approaches, especially at high acceleration rates [38,39]. Multi-mask SSDU retrospectively splits available measurements into multiple 2-tuples of disjoint sets for each training sample, while using one of these sets for DC units and the other for defining loss. Thus, the multi-masking approach enables augmentation of the training dataset. Experiments on knee and brain datasets show that multi-mask SSDU provides enhanced and artifact free reconstruction quality.

SSDU and its extensions enable database training without fully-sampled data. However, they still require a database for training in order to learn the large number of parameters for the neural network. However, in some MRI applications involving time-varying physiological processes, dynamic information such as time courses of signal changes, contrast-related uptake or breathing patterns may differ substantially between subjects, making it difficult to generate high-quality databases of sufficient size for the aforementioned strategies. Furthermore, database training in general brings along concerns about generalization [40, 41]. Particularly, for MRI reconstruction, this may translate to training and test dataset mismatches on image contrast, sampling pattern, SNR, vendor, and anatomy. For instance, the fastMRI transfer track challenge shows that the performance of pretrained models degrades due to distribution shift or changes in acquisition parameters at inference time [42]. Moreover, bias due to datasets lacking examples of rare and/or subtle pathologies increases the risk of generalization failure [40,41]. Such challenges necessitate a new methodology to enable subject-specific DL MRI reconstruction without external training datasets, since it is clinically imperative to provide high-quality reconstructions that can be used to identify lesions/disease for every individual. In Chapter 5, we introduce a zero-shot self-supervised learning (ZS-SSL) approach to perform subject-specific accelerated DL MRI reconstruction to tackle these issues [43, 44]. The ZS-SSL partitions the available measurements from a single scan into three disjoint sets. Two of these sets are used to enforce data consistency and define loss during training for self-supervision, while the last set serves to self-validate, establishing an early stopping criterion. In the presence of models pre-trained on a database with different image characteristics, ZS-SSL can be combined with transfer

learning for faster convergence time and reduced computational complexity. Experimental results show that ZS-SSL methods perform similarly to database-trained learning methods despite being trained on a single sample. Moreover, ZS-SSL tackles the robustness and generalizability issues associated with database training such as domain shift and difference of acquisition parameters that can occur at inference time.

Finally, we show the utility of self-supervised physics-guided deep learning for referenceless image denoising. Deep learning based image denoising methods have been recently popular due to their improved performance [45–51]. Traditionally, these methods are trained in a supervised manner, requiring a set of noisy input and clean target image pairs [45, 52]. More recently, self-supervised approaches have been proposed to learn denoising from only noisy images [48, 50, 53]. These methods assume that noise across pixels is statistically independent, and the underlying image pixels show spatial correlations across neighborhoods. These methods rely on a masking approach that divides the image pixels into two disjoint sets, where one is used as input to the network while the other is used to define the loss. However, these previous self-supervised approaches rely on a purely data-driven regularization neural network without explicitly taking the masking model into account. Thus, the pixels used for training are not re-utilized in the end-to-end training since the network is only comprised of a neural network. Reusing of the pixels within the network may promote consistency with acquired measurements, thus leading to a more robust and improved denoising performance. In Chapter 6, building on these self-supervised approaches, we introduce Noise2Inpaint (N2I), a training approach that recasts the denoising problem into a regularized image inpainting framework [54]. We use algorithm unrolling to unroll an iterative optimization for solving this objective function and train the unrolled network end-to-end. The training paradigm follows the masking approach from previous works, splitting the pixels into two disjoint sets. Importantly, one of these is now used to impose data fidelity in the unrolled network, while the other still defines the loss. We demonstrate that N2I performs successful denoising on real-world datasets, while better preserving details compared to its purely data-driven self-supervised counterparts.

Chapter 2

Self-Supervised Learning of Physics-Guided Reconstruction Neural Networks without Fully Sampled Reference Data

2.1 Introduction

Data acquisition in MRI is inherently slow, necessitating the use of accelerated imaging techniques. In these approaches, data is acquired at sub-Nyquist rates, and reconstructed using additional information. Parallel imaging exploits the redundancies between receiver coils and is the most clinically used approach [1–3]. Compressed sensing is another method that utilizes the compressibility of images based on linear sparsifying transforms for a regularized reconstruction [4–6, 14, 19, 20], which can also be synergistically combined with multi-coil acquisitions [11, 55, 56]. At high acceleration rates, parallel imaging suffers from noise amplification [57–59], while compressed sensing may lead to residual artifacts [60, 61]. Furthermore, compressed sensing reconstruction is computationally lengthy in nature and typically requires empirical fine-tuning of regularization parameters, although recent approaches using rapid self-tuning show promise

This chapter is based on [10, 35].

for principled parameter selection [62, 63].

Recently, deep learning (DL) has gained interest for high-quality accelerated MRI. DL based MRI reconstruction algorithms can be roughly divided into two categories, purely data-driven and physics-guided [18, 22]. In purely data-driven approaches, a mapping between the undersampled k-space/aliased image to full k-space/artifact-free image is learned [8, 23–26]. In the so-called physics-guided methods, the knowledge of the forward encoding operator, which contains the undersampling pattern and typically the coil sensitivities, is taken into account to solve an inverse problem based on a regularized least squares objective function [7, 9, 17, 27–32]. Some other works have directly worked with multi-coil data without explicitly including the coil sensitivities [64, 65]. These techniques unroll an iterative reconstruction algorithm for solving this objective method for a fixed number of iterations. The unrolled network alternates between data consistency and regularization, where the regularization is implemented implicitly using a neural network. Subsequently, these unrolled networks are trained end-to-end with a loss function that characterizes similarity with a reference image obtained from fully-sampled data [18]. The parameters of the network can be different across the unrolled iterations [7, 29] or shared across them [9, 31].

The aforementioned physics-guided methods have been trained in a supervised manner, where fully-sampled data is used as a reference during the training. However, in many practical imaging scenarios, it is infeasible to acquire fully-sampled datasets. For instance, when imaging moving organs, such as the heart, there is often a short period of time during which the data needs to be acquired. Example acquisitions include real-time imaging, myocardial perfusion, and numerous contrast-enhanced scans [33, 66, 67]. Another hindrance for fully-sampled acquisitions in some applications include the signal decay. This is pronounced in acquisitions, such as diffusion MRI with echo-planar imaging, where the signal decays quickly with T_2^*, thus prohibiting use of fully-sampled acquisitions especially at high resolutions [34, 68]. In several other scenarios such as whole-heart coronary MRI or high-resolution anatomical brain imaging, it is impractical to acquire fully-sampled datasets as the scan time becomes extremely lengthy.

Furthermore, accelerated imaging methods are often used to improve acquisition resolution. When higher acceleration rates are achievable, these are not solely used for image time reduction, but rather a trade-off is made with improved resolution [20, 56, 69].

However, this newer resolution may necessitate re-training of the DL reconstruction, since neural networks do not necessarily generalize across different resolutions, as depicted in Supporting Information **Figure A.1**. Thus, if fully-sampled data is required for training at higher resolutions, this may lead to excessive scan times, even for anatomical imaging protocols, making it difficult to make protocol changes to fully utilize the benefits of accelerated imaging.

In this Chapter, we sought to develop a new self-supervised learning approach to train physics-guided DL-MRI reconstruction without fully-sampled reference data. The proposed self-supervised approach which we term as Self-Supervision via Data Undersampling (SSDU) splits the acquired k-space indices into two disjoint sets. One of these is used in the data consistency unit for the network, while the other set is used to define the loss function in k-space. Hence, end-to-end training and evaluation of the network is done through only the acquired measurements without making any other assumptions about image output or characteristics. We apply the proposed self-supervised training without fully-sampled data, on the fastMRI knee datasets and prospectively undersampled high-resolution brain MRI datasets. These are compared to parallel imaging, compressed sensing and a supervised training of a DL-MRI network when fully-sampled reference data is available. Our results indicate that the proposed self-supervised method performs similarly to the supervised approach trained on fully-sampled data, although it is trained only on undersampled data.

2.2 Theory

2.2.1 Physics-Guided Neural Networks for MRI Reconstruction

Let $\mathbf{x}$ denote the image to be recovered and $\mathbf{y}_\Omega$ represent acquired k-space measurements with undersampling pattern Ω. The forward model for the acquisition is given as

$$\mathbf{y}_\Omega = \mathbf{E}_\Omega \mathbf{x} + \mathbf{n}, \tag{2.1}$$

where $\mathbf{E}_\Omega : \mathbb{C}^{M_1 \times M_2} \rightarrow \mathbb{C}^P$ is the encoding operator including a partial Fourier matrix sampling the locations specified by Ω, and $\mathbf{n} \in \mathbb{C}^P$ is measurement noise. The forward model presented in Eq. (2.1) is usually ill-conditioned due to sub-Nyquist sampling and hence regularizers that induce prior information is incorporated into the objective

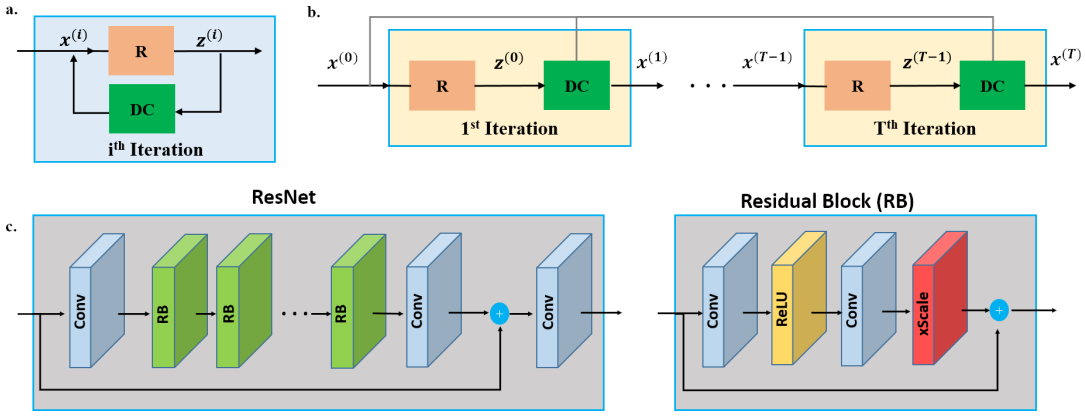

Figure 2.1: a) Depiction of a conventional iterative optimization algorithm for solving regularized inverse reconstruction problems. These algorithms alternate between regularization (R) and data consistency (DC). b) For neural networks, this iterative algorithm is unrolled for T steps, leading to a feed-forward structure alternating between R and DC units, where R is implemented by means of a neural network. c) The ResNet architecture used as regularizer (R) in this study consists of 15 residual blocks (RB), each of which contains two convolution layers with the first one followed by a ReLU and the second one followed by a constant multiplication layer.

function for the reconstruction. Possible choices for the regularizer include total variation [11–13], ℓ_1-norm of wavelet coefficients [4, 14, 15], sparsity in adaptive transform domains [6, 16], and more recently neural networks [7, 9, 17, 18]. The image recovery is then formulated as an optimization problem

$$\arg\min_{\mathbf{x}} \|\mathbf{y}_\Omega - \mathbf{E}_\Omega \mathbf{x}\|_2^2 + \mathcal{R}(\mathbf{x}), \tag{2.2}$$

where first term enforces data consistency with acquired measurements, while $\mathcal{R}(\cdot)$ is a regularization term. The optimization problem in Eq. 6.8 can be solved in numerous ways, including proximal gradient descent, variable splitting with quadratic penalty, alternating direction method of multipliers among others [7,28,30,70]. In this study, we will consider the variable splitting with quadratic penalty approach [70,71] for implementation, which has also been used in previous physics-guided DL-MRI approaches [9,30]. In this method, data consistency and regularization are decoupled as

$$\arg\min_{\mathbf{x},\mathbf{z}} \|\mathbf{y}_\Omega - \mathbf{E}_\Omega \mathbf{x}\|_2^2 + \mu\|\mathbf{x} - \mathbf{z}\|_2^2 + \mathcal{R}(\mathbf{z}), \tag{2.3}$$

where $\mathbf{z}$ is the auxiliary variable that is initially constrained to be equal to $\mathbf{x}$, and μ is the parameter for the quadratic penalty for relaxing this intermediate constrained problem to an unconstrained one. The optimization problem in Eq. 6.10 is then solved iteratively by alternating the minimization over the variables $\mathbf{x}$ and $\mathbf{z}$ as follows

$$\mathbf{z}^{(i-1)} = \arg\min_{\mathbf{z}} \mu\|\mathbf{x}^{(i-1)} - \mathbf{z}\|_2^2 + \mathcal{R}(\mathbf{z}) \tag{2.4a}$$

$$\mathbf{x}^{(i)} = \arg\min_{\mathbf{x}} \|\mathbf{y}_\Omega - \mathbf{E}_\Omega\mathbf{x}\|_2^2 + \mu\|\mathbf{x} - \mathbf{z}^{(i-1)}\|_2^2 \tag{2.4b}$$

where $\mathbf{x}^{(0)}$ is the initial image obtained from zero-filled under-sampled k-space data, $\mathbf{x}^{(i)}$ is the network output at iteration i and $\mathbf{z}^{(i)}$ is an intermediate variable. In compressed sensing methods, these problems are solved in an iterative manner by alternating between the regularizer and data consistency units until a stopping criterion met as shown in **Figure 2.1a**.

In physics-guided DL-MRI approaches, this iterative algorithm is unrolled for a fixed number of iterations, as depicted in **Figure 2.1b**. The regularization sub-problem in Eq. (6.12a) is implicitly solved using a neural network. The data consistency sub-problem in Eq. (6.12b) has a closed form solution

$$\mathbf{x}^{(i)} = (\mathbf{E}_\Omega^H\mathbf{E}_\Omega + \mu\mathbf{I})^{-1}(\mathbf{E}_\Omega^H\mathbf{y}_\Omega + \mu\mathbf{z}^{(i-1)}), \tag{2.5}$$

where $\mathbf{I}$ is the identity matrix and $(\cdot)^H$ is the conjugate transpose operator. Eq. (6.12b) can be solved using gradient descent or conjugate gradient, which itself is unrolled for a number of iterations [9].

2.2.2 Supervised Training with Fully-Sampled Reference Datasets

Supervised learning performs end-to-end training using ground truth images as the reference labels for the training loss function [7, 23]. Ground truth images are obtained through SENSE-1 coil combination [1], which is the sum across the coil dimension of the product of the conjugate of the coil sensitivity maps with the corresponding coil images [29, 30]. Suppose $\mathbf{x}_{\text{ref}}^i$ that is the ground truth image for subject i, and $f(\mathbf{y}_\Omega^i, \mathbf{E}_\Omega^i; \boldsymbol{\theta})$ denotes the output of the unrolled network that is parametrized by $\boldsymbol{\theta}$ for subsampled k-space data $\mathbf{y}_\Omega^i$, and corresponding encoding matrix $\mathbf{E}_\Omega^i$ of the same subject i. The supervised training of a physics-guided DL-MRI method can be performed by

minimizing the image domain loss

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(\mathbf{x}^i_{\text{ref}},\ f(\mathbf{y}^i_{\Omega}, \mathbf{E}^i_{\Omega}; \boldsymbol{\theta})), \tag{2.6}$$

where N is the number of fully-sampled training data in the database, and $\mathcal{L}(\cdot, \cdot)$ denotes the loss between the ground truth and network output image [7, 9, 29]. Alternatively, supervised training may be evaluated in k-space as

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(\mathbf{y}^i_{\text{ref}},\ \mathbf{E}^i_{full} f(\mathbf{y}^i_{\Omega}, \mathbf{E}^i_{\Omega}; \boldsymbol{\theta})), \tag{2.7}$$

where $\mathbf{y}^i_{\text{ref}}$ is the fully-sampled reference k-space and $\mathbf{E}^i_{\text{ref}}$ is the fully-sampled encoding operator that transforms network output to k-space across coils. Example loss functions include ℓ_1-norm, ℓ_2- norm, mixed norm and perception based loss [25, 30, 72–74].

2.2.3 Proposed Self-supervised Training without Fully-Sampled Reference Data

As discussed previously, acquiring fully sampled data is often difficult or impossible in many scenarios, due to constraints such as organ motion, signal decay or lengthy scan times. Such cases pose an important challenge for the practicality of DL-MRI reconstruction methods that rely on supervised training, since ground truth data is not available for training. To tackle this problem, we propose a self-supervised approach illustrated in **Figure 2.2**, where the acquired sub-sampled data indices, Ω from each scan is divided into two sets Θ and Λ as

$$\Omega = \Theta \cup \Lambda. \tag{2.8}$$

The set of k-space locations specified by Θ are used within the network during training in the data consistency units, while the set of k-space points in Λ are used to define the loss function. Thus, to enable training without using fully-sampled data, the following loss function is minimized

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\Big(\mathbf{y}^i_{\Lambda},\ \mathbf{E}^i_{\Lambda}\big(f(\mathbf{y}^i_{\Theta}, \mathbf{E}^i_{\Theta}; \boldsymbol{\theta})\big)\Big). \tag{2.9}$$

In other words, the unrolled network output image $f(\mathbf{y}^i_\Theta, \mathbf{E}^i_\Theta; \boldsymbol{\theta})$ which only uses the indices specified by Θ for data consistency is transformed to k-space using the encoding operator, $\mathbf{E}^i_\Lambda$ specified by the k-space indices in Λ. Then the loss is calculated in k-space with respect to the acquired k-space data at these locations. In the proposed SSDU approach, Θ was chosen as $\Omega\backslash\Lambda$.Thus, in our self-supervised training methodology, the unrolled network only sees the acquired k-space data at locations $\Theta = \Omega\backslash\Lambda$ to enforce data consistency. The quality of the final reconstruction, i.e. the network output image, is then checked by mapping to the individual coil k-spaces via $\mathbf{E}^i_\Lambda$, and checking the discrepancy to these acquired measurements at these remaining locations Λ. Thus, the network is trained to decrease the discrepancy between the network output transformed to all the coil k-spaces and the acquired measurements that it does not see within its unrolled data consistency units. After the network is trained with our proposed self-supervised approach, the reconstruction for unseen test data is performed by using all available measurements at locations Ω.

Our proposed self-supervised approach share similarities with the widely used concept of cross-validation. In machine learning, cross-validation is commonly used to evaluate how accurately a model will perform with robustness to bias and over-fitting issues. Cross-validation is performed by partitioning available data into two sets, one of which is used to train the model and the other for validation, i.e. check whether the trained model generalizes to unseen data. The key difference between our approach and cross-validation is that we perform partitioning per each slice in the dataset, whereas in cross-validation the whole dataset is partitioned only once. The key hyper-parameter for success of cross-validation is the number of folds, which should be well-designed [75]. Similarly, in our proposed self-supervised approach, subset selection mechanisms for Λ and Θ are critical, which are thoroughly studied in the next section.

2.3 Methods

2.3.1 Network and Training Details

The network for solving sub-problems (6.12a) and (6.12b) was unrolled for 10 iterations. The data consistency in the unrolled network was implemented with conjugate gradient method for solving Eq. 2.5, which itself was unrolled for 10 iterations. The neural

network for solving the sub-problem (6.12a) was implemented using a convolutional neural network (CNN) based on a ResNet structure, which has shown success in other regression problems [76]. This CNN, shown in **Figure 2.1c**, consisted of a layer of input and output convolution layers, and 15 residual blocks (RB) with skip connections that facilitate information flow during network training. Each RB comprised of two convolutional layers in which the first layer is followed by a rectified linear unit (ReLU) and second layer is followed by a constant multiplication layer, with factor $C = 0.1$ [76]. All layers had a kernel size of 3×3 and 64 channels. This ResNet CNN had a total of 592,129 trainable parameters, which were shared across the unrolled iterations. Coil sensitivity maps were generated from the 24×24 center of k-space using ESPIRiT (56) using a kernel size of 6×6, as well as thresholds of 0.02 and 0.95 for calibration-matrix and eigenvalue decomposition.

A normalized ℓ_1 - ℓ_2 loss, defined as

$$\mathcal{L}(\mathbf{u}, \mathbf{v}) = \frac{||\mathbf{u} - \mathbf{v}||_2}{||\mathbf{u}||_2} + \frac{||\mathbf{u} - \mathbf{v}||_1}{||\mathbf{u}||_1} \tag{2.10}$$

was used for both the supervised and the proposed self-supervised training. In the supervised setting, $\mathbf{u}$ and $\mathbf{v}$ correspond to the reference ground-truth image/fully-sampled k-space and network output image/network output k-space obtained by transforming network output images to k-space by applying a fully-sampled encoding operator, while for the proposed self-supervised training these correspond to the acquired k-space measurements at locations specified by Λ and the k-space corresponding to the network output image at the same locations. For supervised training, k-space loss was used throughout the study as it outperforms the image domain loss used in our preliminary results [35] (Supporting Information **Figure A.2**), while also matching our self-supervised framework. Prior to processing, maximum absolute value of the k-space datasets was normalized to 1 in all cases. The networks were trained using the Adam optimizer with a learning rate of 10^{-3} unless specified otherwise, by minimizing the corresponding loss function with a batch size of 1 over 100 epochs. All training was performed using Tensorflow in Python, and processed on a workstation with an Intel E5-2640V3 CPU (2.6GHz and 256 GB memory), and an NVIDIA Tesla V100 GPU with 32 GB memory.

2.3.2 Choice of the Loss Mask

The proposed SSDU approach divides the acquired sub-sampled data into two disjoint sets Θ and Λ. Furthermore, in our implementation, Λ is allowed to vary for each different slice in the training database, i.e. they can be indexed as $\{\Lambda_i\}_{i=1}^{N}$. The subset Λ is retrospectively selected from the acquired k-space points, Ω in order to define the loss function. Hence, unlike the data acquisition process for sampling k-space locations Ω, which is affected by concerns about contrast changes or eddy current artifacts [6], selection of Λ is not limited by any physical constraints. This is because Λ is selected after data acquisition and amounts to the selection of an index set from all possible acquired k-space locations. Thus, distribution and size of Λ were the two hyper-parameters that were studied. For the distribution of Λ, a uniformly random selection among elements of Ω, as well as a variable density selection based on Gaussian random weighting were investigated. For its size, the ratio $\rho = \Lambda/\Omega$ was varied among 0.05, 0.1, 0.2, ...,0.8, 0.9, where $|.|$ is the cardinality of the index set. A 5-fold cross-validation was also performed on training data for quantitative assessment of the distribution of Λ, as well as a subset of ρ values among 0.1, 0.2, 0.3, 0.4, 0.5, 0.6.

Additionally, the impact of the overlap between Θ and Λ on the reconstruction performance was also studied. The first scenario considered was the limiting case when $\Omega = \Theta = \Lambda$. Subsequently, we created three different partial overlap scenarios for the best performing ρ value as: 1) The first case, referred to as disjoint sets, in which there is no overlap between Θ and Λ (as originally proposed); 2) The second case, referred to as 50% overlap, where we included 50% of points from Λ in Θ as well. More formally, i.e. $|\Lambda \cap \Theta|/|\Lambda| = 0.5$; 3) Lastly, we have the 100% overlap case where all points in Λ is included in Θ as well (in this case $\Omega = \Theta$, but Λ is a subset of Ω).

2.3.3 Fully-Sampled Knee MRI

Knee datasets were obtained from the New York University (NYU) fastMRI initiative database, which was curated with an approval from the NYU School of Medicine Institutional Review Board [77]. Fully sampled raw data were acquired on a clinical 3T system (Magnetom Skyra, Siemens, Erlangen, Germany) with a 15-channel knee coil using 2D turbo spin-echo sequences. The imaging parameters used for the knee data

acquisitions are provided in the Supporting Information Table A.1.

The fully-sampled raw data were under-sampled retrospectively for both training and testing using equispaced sampling patterns provided in the fastMRI database with an acceleration rate (R) = 4 [7, 40, 77]. The center of k-space was fully-sampled with 24 lines of auto-calibrated signal (ACS). The training set consisted of 300 slices from 15 subjects for coronal PD, coronal PDFS, and 10 subjects for sagittal PD, sagittal T_2, axial T_2. Testing was performed on all slices from 10 different subjects for all knee sequences. Ground truth images for supervised training were generated with a SENSE-1 combination of the fully-sampled data [29, 30]. The proposed self-supervised approach was compared with supervised DL-MRI trained on fully-sampled dataset and conjugate gradient SENSE (CG-SENSE) [78]. Additionally, comparison to a multi-coil compressed sensing reconstruction incorporating coil sensitivities with total generalized variation (TGV) as regularizer [12] was carried out for illustration purposes. However, TGV was not performed on all test datasets since it is computationally expensive, and a comparison between supervised DL-MRI and TGV was already performed in [7]. For TGV, the MATLAB implementation provided by authors was utilized [12]. We note that TGV and CG-SENSE approaches are shown only for comparison purposes with more traditional methods, and are not considered as competitive baseline images, consistent with previously reported results in the literature [7].

2.3.4 Prospectively Accelerated Brain MRI

Brain imaging was performed on 19 healthy subjects at a 3T Siemens Magnetom Prisma (Siemens Healthcare, Erlangen, Germany) system using a 32-channel receiver head coil-array. The imaging protocols were approved by the local institutional review board, and written informed consent was obtained from all participants before each examination for this HIPAA-compliant study. Data acquisition was performed using a standard Siemens 3D-MPRAGE sequence with the following parameters: FOV = 224 × 224 × 157 mm^3, resolution = 0.7 × 0.7 × 0.7 mm^3, TR/TE = 2400 ms/2.2 ms, inversion time = 1000 ms, flip angle = 8°, band-width = 210 Hz/pixel, 3D matrix size = 320×320×224, prospective acceleration R = 2 (equispaced in k_y), ACS lines = 32, acquisition orientation = sagittal. The k-space data was inverse Fourier transformed along the read-out

(foot-head) direction, and these axial slices were processed individually. The prospectively undersampled brain datasets were further retrospectively undersampled to R = 4, 6, 8 using a sheared equispaced $k_y - k_z$ undersampling pattern [79], with a 32×32 ACS region in the $k_y - k_z$ plane. Sampling masks are provided in Supporting Information **Figure A.3**. We note that while in principle prospectively sub-sampled data can be acquired at all these different rates, we chose to utilize further retrospective sub-sampling of prospectively accelerated data since our focus is on the reconstruction quality and this approach avoids confounding factors between different scans, such as subject motion or variations from T1 recovery. We also note that when the self-supervised approach was used at one of these higher acceleration rates, it only had access to the k-space data corresponding to that acceleration rate, both during training and testing. The learning rate for training was set to 5×10^{-4}. The training set consisted of 300 slices from 10 subjects, formed by taking the central 30 slices from each subject. Testing was performed on all slices from 9 different subjects.

The proposed self-supervised DL-MRI results were compared to CG-SENSE method. We note that a comparison to supervised DL-MRI was not possible in this setting, since there was no fully-sampled ground truth data.

2.3.5 Image Evaluation

Experimental results were quantitatively evaluated using normalized mean square error (NMSE) and structural similarity index (SSIM). Additionally, qualitative assessment of the image quality was performed by an experienced radiologist. For knee MRI, the proposed self-supervised DL-MRI approach was compared to ground truth fully-sampled images, supervised DL-MRI trained on fully-sampled data and CG-SENSE at the same acceleration R = 4. As noted earlier, TGV was not included in the comparison due to its computational complexity and availability of a previous study comparing supervised DL-MRI and TGV [7]. For brain MRI, proposed self-supervised DL- MRI reconstructions at acceleration R = 4, 6 and 8 were compared with CG-SENSE approach at the acquisition acceleration R = 2. The reader was blinded to the reconstruction method, except for the knowledge of the reference image in knee MRI datasets. The order in which the methods were shown was also randomized. There were differences between the sequences used for the fastMRI database and our institutional sequences, thus this knowledge allowed

the radiologist to assess the baseline image quality. All five knee MRI weightings and brain dataset were evaluated on a 4-point ordinal scale, adopted from [7] for blurring (1: no blurring, 2: mild blurring, 3: moderate blurring, 4: severe blurring), SNR (1: excellent, 2: good, 3: fair, 4: poor), aliasing artifacts(1: none, 2:mild, 3: moderate, 4: severe) and overall image quality (1: excellent, 2: good, 3: fair, 4: poor). Wilcoxon signed-rank test was used to evaluate the scores with a significance level of $P < 0.05$.

2.4 Results

2.4.1 Choice of the Loss Mask

Figure 2.3 depicts the self-supervised network training using varying subsets across slices by uniformly random and variable-density Gaussian selection of $\Lambda \subset \Omega$ for $\rho = 0.1$. Uniformly random selection of Λ suffers from visible residual artifacts, marked by red arrows. These artifacts are further suppressed in the Gaussian-based approach and difference images align with these observations. The quantitative assessment from 5-fold cross-validation are consistent with these qualitative assessments. The median and interquartile range of SSIM values were 0.9380 [0.9197, 0.9527], 0.9457 [0.9293, 0.9575], and NMSE values were 0.0021 [0.0016, 0.0027], 0.0019 [0.0015, 0.0023] using uniform random selection and Gaussian selection, respectively. Supporting Information **Figure A.4** shows additional reconstructions for uniform random and Gaussian selection for different ρ values, which further highlights that Gaussian selection consistently outperforms uniform random selection across different ρ values. Thus, a variable-density Gaussian selection was used for Λ for the remainder of the study.

Figure 2.4 shows the impact of network training with varying $\rho \in \{0.05, 0.1, 0.2, \ldots, 0.8, 0.9\}$ using variable-density Gaussian selection. Red arrows show visible residual artifacts for low ρ values of 0.05, 0.1, 0.2. As cardinality of Λ increases towards $\rho = 0.4$, residual artifacts decrease. At $\rho = 0.4$, visible artifacts seen at lower ρ values are further suppressed. Residual artifacts start to reappear starting from $\rho = 0.5$, and these artifacts become more pronounced as ρ increases. The quantitative assessment from 5-fold cross-validation aligns with these qualitative assessments. The median and interquartile range of SSIM values were 0.9457 [0.9293, 0.9575], 0.9477 [0.9323, 0.9591],

0.9488 [0.9328, 0.9603], 0.9507 [0.9352, 0.9614], 0.9450 [0.9297, 0.9569], 0.9391 [0.9225, 0.9524], and NMSE values were 0.0019 [0.0015, 0.0023], 0.0018 [0.0013, 0.0023], 0.0018 [0.0014, 0.0022], 0.0017 [0.0013, 0.0021], 0.0020 [0.0015, 0.0024], 0.0022 [0.0016, 0.0028] using Gaussian selection for $\rho \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$, respectively. Hence, $\rho = 0.4$ was used for the remainder of the study.

Figure 2.5 shows the impact of different degrees of overlap between Λ and Θ for $\rho = |\Lambda|/|\Omega| = 0.4$, as well as the limiting case that uses all available data for both data consistency and loss (i.e. $\Omega = \Theta = \Lambda$). For the limiting case with $\Omega = \Theta = \Lambda$, the reconstruction results suffer from residual noise amplification. On the other hand, when Λ and Θ were disjoint as proposed, such noise amplifications are significantly suppressed. Quantitative SSIM and NMSE evaluation of these methods over the dataset are presented in Supporting Information Table A.2, indicating that for different rates of overlap between Λ and Θ with $\rho = 0.4$, the performance degrades as the amount of overlap increases. Thus disjoint sets were used for the remainder of the study.

2.4.2 Knee MRI

Figure 2.6 demonstrates the reconstruction results of coronal PD images using CG-SENSE, TGV, supervised DL-MRI and proposed self-supervised DL-MRI approach along with the ground truth reference, as well as difference images with respect to this reference. CG-SENSE and TGV suffer from visible residual artifacts, marked by red arrows, with the latter having fewer artifacts. The proposed self-supervised and supervised DL-MRI approaches successfully remove the residual artifacts, while achieving similar qualitative and quantitative performance. Quantitative metrics and difference images displayed in the figure are in agreement with these observations. Supporting Information **Figure A.5** shows the training loss curves for both approaches where loss decreases over epochs in a similar trend.

The same trends were observed for coronal PD-FS as depicted in **Figure 2.7**. Both proposed and supervised DL-MRI approaches show similar performance, while improving the suppression of residual artifacts that are visible in CG-SENSE and TGV methods. Quantitative evaluation and the residual artifacts apparent in the difference images also highlight these observations. Supporting Information **Figure A.6** show reconstruction results for axial T_2, sagittal T_2 and sagittal-PD weighted knee dataset which align

with observation from coronal weighted knee datasets.

Figure 2.8 shows a box-plot displaying the median and interquartile range (25^{th}-75^{th} percentile) of the quantitative metrics, SSIM and NMSE, across all test datasets for each knee sequence. In all sequences, supervised and self-supervised DL-MRI approaches achieve similar quantitative performance for both SSIM and NMSE, while significantly outperforming the CG-SENSE approach. We note again that TGV was not included in these comparisons, as it is computationally expensive, and a comparison between supervised DL-MRI and TGV was already performed in [7].

2.4.3 Prospectively Accelerated Brain MRI

Figure 2.9 depicts a sagittal slice of the 3D MPRAGE dataset at acquisition acceleration R = 2 and further retrospective acceleration R = 4, 6 and 8 reconstructed with CG-SENSE, as well as R = 4, 6 and 8 reconstructed with the proposed self-supervised DL-MRI on a representative test subject, following reformatting to the original acquisition (sagittal) plane. CG-SENSE suffers from significant noise amplification at higher acceleration rates. Self-supervised DL-MRI successfully performs reconstruction at these higher acceleration rates, while achieving lower noise level and similar overall image quality with CG-SENSE at R = 2. Results from another subject are depicted in Supporting Information **Figure A.7** and shows similar trends. TGV was not applied due to the high computational runtime across all axial slices, and supervised DL-MRI cannot be applied in this setting due to the lack of fully-sampled references.

2.4.4 Image Evaluation Scores

Figure 2.10 summarizes the results of the reader study for knee and brain datasets. For knee datasets, both supervised and self-supervised DL-MRI approaches get comparable scores to the reference image in terms of SNR, blurring, aliasing artifacts and overall image quality. There was no statistical difference between reference and DL-MRI approaches in terms of the evaluation criterions for all knee sequences, except for blurring between reference and DL-MRI approaches in coronal PD-FS. CG-SENSE was significantly outperformed by both DL-MRI approaches, while showing statistically significant differences to the reference and both DL-MRI approaches for all knee sequences,

except in blurring for coronal PD and PD-FS sequences. More comprehensive bar plots of the average scores including CG-SENSE and supervised training with image domain loss as in Eq. (2.6) are presented in Supporting Information **Figure A.8**.

For the 3D MPRAGE dataset, DL-MRI reconstructions trained using the proposed self-supervised approach at acceleration rates 4, 6 and 8 show similar statistical properties in terms of SNR and blurring with CG-SENSE at acquisition R = 2. However, in terms of aliasing artifacts and overall image quality, proposed self-supervised approach at all three acceleration rates (R = 4, 6 and 8) outperform CG-SENSE at R = 2. In terms of aliasing artifacts, proposed self-supervised approach for rates 4 and 6 show similar statistical behavior with each other, while significantly improving upon self-supervised DL-MRI at R = 8 and CG-SENSE at R = 2, which perform statistically similar among themselves. Proposed self-supervised approach at R = 4 shows the best overall image quality and shows statistically significant differences with self-supervision at R = 6, 8 and CG-SENSE at R = 2. As expected, the overall image quality decreases with higher acceleration rates using the proposed self-supervised DL-MRI approach, although these techniques still outperform CG-SENSE at R = 2.

2.5 Discussion

In this study, we developed a framework for self-supervised training of physics based DL-MRI reconstruction without fully sampled data. The proposed approach split the acquired under-sampled k-space indices into two disjoint sets Θ and Λ, where the former was used across the unrolled network to enforce data consistency, while the latter was used to define the loss function for the training. The results on retrospectively under-sampled knee datasets showed that our SSDU approach achieves comparable results with a supervised DL-MRI approach using the same neural network architecture, while outperforming conventional CG-SENSE and TGV approaches. Results on prospectively under-sampled brain datasets, for which supervised learning methods cannot be applied due to unavailability of fully-sampled data, further confirmed the effectiveness of the proposed self-supervised training approach for DL-MRI reconstruction. These reconstructions at higher acceleration rates of 4, 6 and 8, visually outperformed CG-SENSE at R = 2 according to the reader study. We note that CG-SENSE was implemented

without regularization, and its performance may be improved using Tikhonov regularization with the regularization parameter selected over a training set [7].

Most DL-MRI approaches use supervised learning for network training in order to provide improved accelerated MRI reconstruction [9, 17, 18, 27, 30, 31, 40]. However, acquiring fully-sampled data is challenging in many practical scenarios of interest. These may be due to constraints on timing, physiological constraints, signal decay or long scan times [33,34,66–68]. As an example, the fully-sampled acquisition for the 3D MPRAGE sequence with the resolution used in this study would be more than 15 minutes [34], which is impractical for large studies and may lead to patient discomfort. Furthermore, such long scan times increase susceptibility to motion artifacts, which would be more pronounced at these high resolutions. To further highlight the need for training data, we have also performed experiments on prospectively sub-sampled snapshot cardiac MRI, where it is infeasible to collect the ground truth data. Results from these experiments are shown in Supporting Information **Figure A.9**, showing the applicability of our method in this setting as well. Thus, being able to train DL-MRI reconstruction methods without fully-sampled data is imperative to broaden their application to settings in which such data is challenging to acquire, where supervised training are no longer practical. Furthermore, this may also facilitate the integration of DL-MRI methods to many clinical scans that readily include a form of accelerated imaging, most commonly in the form of parallel imaging, by enabling the use of prospectively undersampled raw k-space data for training.

Given the importance of training without fully sampled data, there have been several works which have tried to tackle this issue. For purely data-driven de-aliasing of single-coil data using image domain to image domain mapping without the encoding operator, a self-supervised approach has been proposed [80] using a mixture of measurement and k-space losses. Unlike our approach, it uses all available data for training and loss, i.e. identical sets. As a result, the reconstructions suffer from visible noise amplifications which also align with our observation about usage of identical sets in **Figure 2.5**. An alternative approach, which assumes the same data is acquired with two separate acquisitions using different undersampling patterns was also proposed [81, 82] extending on the Noise2Noise denoising framework [83]. In the same image-domain reconstruction setting, a self-supervised learning scheme using cycleGANs with optimal

transport cost minimization was proposed [84], although initial results exhibit blurring artifacts. Although purely data-driven image domain methods have been used for DL-MRI reconstruction, physics-guided DL-MRI techniques are more desirable as they offer a degree of interpretability by incorporating domain knowledge on the MRI encoding mechanism [7, 9, 18, 22, 29]. In this physics-guided setting, earlier work used the output of a regularized CG-SENSE algorithm based on compressed sensing as the reference for supervised training , showing that such training may outperform the compressed sensing output, as some images are over-regularized while others are under-regularized. However, this approach assumes that the compressed sensing algorithm output will be a reliable estimate of the image without residual aliasing artifacts, and thus is limited by sampling strategies and acquisition acceleration rates, as high acceleration rates or equispaced sampling may lead to degradation in the compressed sensing results. More recently, an unpaired learning approach using Wasserstein GANs was proposed [85], but this procedure still assumes the presence of high-quality images albeit not requiring pairwise matching with undersampled data. Another approach uses the so-called unsupervised basis pursuit [86], where the unrolled network consists of regularizer units followed by several consecutive DC units. This approach uses the current output of the DC unit as the training label, and iteratively updates both network parameters and this training label, in a method reminiscent of semi-supervised training. This method was investigated with random undersampling patterns, where intermediate outputs tend to suffer from noise amplification but without significant residual artifacts. In this setting, this approach was able to reduce noise further, even though noise amplification was observed when compared to supervised training [86]. However, this method was not investigated for equispaced undersampling, as is the focus of this study, where intermediate DC outputs are both noisy and likely to have residual aliasing artifacts. Thus, the utility of this method in equispaced undersampling is unclear and warrants further investigation. In contrast, our SSDU approach uses physics-guided DL-MRI reconstruction, while not making any explicit assumptions about the final output in image space. In particular, we do not enforce the output of our network to align with a generative model or consider intermediate estimates as reference output for training. The training in SSDU only considers the acquired k-space data to evaluate the reconstruction quality, in effect using a physics-guided self-supervision approach. Furthermore, SSDU works

for both equispaced undersampling patterns, as is the focus of the study, and random undersampling patterns (results not shown). Note the former was considered to be more challenging for physics-guided DL-MRI reconstruction in previous studies, as networks trained with equispaced sampling were shown to generalize well to random sampling, but not the vice versa [7, 87].

Our training method is also reminiscent of the broader and fundamental concept of cross-validation in machine learning and statistics [88]. When testing generalizability, the training database is partitioned into two sets of complementary datasets, one which is used for training the model (often called training set), and the other used to assess the performance in unseen data (often called validation/testing set). In our approach, we do a similar partitioning of the acquired data to two sets we denoted Θ and Λ. The main difference to typical cross-validation is that our partitioning is done for each subject in the training set from the database. But the intuition for partitioning within the network is similar, as the unrolled network only sees Θ for data consistency during training, while Λ is only used to establish the network loss. Indeed, as our experiments in **Figure 2.5** show that when Θ and Λ are taken to be the same as Ω, such training leads to poor image quality with insufficient removal of aliasing artifacts and noise amplification, as the DC unit operating on the full Ω, inherently matches well with the acquired data at these locations.

Selection of the loss mask, Λ plays an important role in the performance of the proposed self-supervised training. One major design advantage is that since it only exists in post-processing, it can be chosen freely among all the acquired measurements retrospectively, without physical constraints that are imposed during acquisition. Thus even though 40% of the acquired indices in Ω were included in Λ, this is not the equivalent to training with an $\sim$8-fold accelerated acquisition, especially for the 2D setting, since the points in Λ do not need to constitute fully-sampled readout encoding lines along k_x. This point is further illustrated in Supporting Information **Figure A.10**, in the context of supervised training. This advantage is not as clear in the training for the 3D brain dataset in this study, since the data had to be inverse Fourier transformed along the foot-head readout direction and axial slices had to be processed due to memory issues in the GPUs. In this case, the sheared equispaced $k_y - k_z$ undersampling pattern readily do not include any lines, thus the selection of Λ, may affect the DC units more

substantially than in the 2D knee MRI experiments. Accordingly, the self-supervised approach is expected to show more gains and better reconstruction quality at higher acceleration rates for 3D imaging if 3D neural networks can be used. Thus memory-efficient 3D neural network designs [89] may warrant further investigation, although it is beyond the scope of the current study.

The data reduction arising from data splitting between Θ and Λ poses more challenges for training and reconstruction at higher acceleration rates, even for 2D acquisitions. This was further investigated to check how the performance of self-supervised and supervised training would change at higher acceleration rates when all the training parameters and datasets are the same as described earlier. The results shown in Supporting Information **Figure A.11** indicate that both training methods perform similarly at R = 4 and 6 for knee MRI. However, at R = 8, where the supervised training is able to suppress artifacts albeit at the cost of blurring artifacts, the self-supervised approach starts suffering from additional residual aliasing artifacts. Thus, at higher acceleration rates, where reconstructions from the supervised training can operate without aliasing artifacts but with quality degradation, the self-supervised approach faces additional challenges including residual aliasing, due to the scarcity of data, especially after the splitting to two sets. The problem of data scarcity has been addressed by several important transfer learning methods when using supervised training with fully-sampled datasets [90, 91]. These approaches pre-train neural networks on fully-sampled large datasets and then fine-tune them on smaller datasets of interest. In such cases, if the smaller dataset of interest is additionally not fully-sampled, then the proposed self-supervised approach may be combined synergistically with transfer learning to tackle this challenging issue of both data scarcity and not having fully-sampled data, though this was beyond the scope of this study. We also note that there are differences between the weights of the networks from supervised and self-supervised training approaches. However, a quantitative difference, such as NMSE, between learned weights of these two training approaches does not directly translate to reconstruction performance, as shown by our results. Nonetheless, it is noteworthy that two different trained networks with differences among their weights have similar reconstruction performance during the testing stage, further alluding to the complexity of the parameter space for the neural network.

All experiments in this study were based on Cartesian acquisitions. The proposed self-supervised approach can be extended to non-Cartesian acquisitions. In non-Cartesian acquisitions such as radial or spiral acquisitions, one can choose the subsets for training and loss mask from the acquired radial spokes and spirals, similar to Cartesian acquisitions used in this study, since this amounts to selecting a subset of individual k-space points on the spokes or spirals. We also note that for non-Cartesian acquisitions, the encoding operator also contains the gridding/de-gridding operation to account for non-uniform Fourier transforms. These extensions were not investigated, as it was beyond the scope of the current work.

In this study, we compared uniformly random selection with a variable-density approach based on Gaussian weighting for selecting Λ. In our experiments, the latter selection was favored as it statistically outperformed and visibly improved upon the former. A self-supervised mask selection during the network training may further remove these hyper-parameters and potentially lead to further improvements in reconstruction. However, this is a difficult problem, which warrants further investigation, beyond the scope of the current study. Using different distributions for selecting a number of distinct Θ and Λ pairs per subject may further improve performance, but currently these distributions would need to be empirically chosen. Due to the ad-hoc nature of such a process and the wide range of available distributions, this was not explored in detail, but this idea also warrants more investigation in the context of self-supervised mask selection in future works. We also investigated the reconstruction performance using the same sets, Θ and Λ, across all training slices versus letting these vary across slices as Θ_i and Λ_i, as proposed. Although one can choose these sets to be same for all slices, such an approach bears the risk of a sub-optimal loss mask being used for all slices. Hence, having different sets for each slice in the dataset may provide additional robustness. Supporting Information **Figure A.12** shows that having different loss and training sets for each slice shows slight improvement over using the same sets across all the training dataset. Finally, a heuristic choice was made to keep 4×4 central k-space lines in the Θ set, as the DC units did not work well without these high-energy components. In our experience, use of larger (8×8 or 16×16) or smaller (2×2) regions deteriorated the overall performance.

The same residual network structure for regularizer and unrolled conjugate gradient

for data consistency units were used throughout the study. However, our approach is not restricted to these network and DC unit choices. Alternative approaches, such as a DenseNet, U-Net or variational neural network as a regularizer CNN [7, 92, 93], or gradient descent for the DC unit are also possible [7, 31]. However, these were not explored, since such network optimization was not the focus of our study. Instead we fixed one architecture, and used this for both supervised and self-supervised training. In this study, we also shared the regularizer CNN parameters across the unrolled network, similar to [9,31], in order to enable training with a smaller training dataset. However, it is possible to use different parameters for each unrolled regularizer unit, as in [7, 29], at the cost of a higher number of trainable parameters. A comparison between supervised training with shared and non-shared parameters in the unrolled network is provided in Supporting Information **Figure A.13**. The results indicate that the two approaches perform similarly in terms of qualitative and quantitative assessments.

Selection of proper loss functions also play a vital role for network training. The $\ell_2$loss is a frequently used metric in DL-MRI with promising results (20,28), but it is sensitive to outliers. On the other hand, ℓ_1 loss is more robust to outliers. Hence, we used a normalized ℓ_1-ℓ_2 loss to take advantage of the superior properties of each loss while minimizing their disadvantages [73]. Other choices of losses such as discriminative losses have also been popular for supervised training of DL-MRI methods [31,94]. There have also been works to incorporate the conventional loss functions such as $\ell_1$or $\ell_2$into adversarial losses [25, 95, 96]. To the best of our knowledge, there are no works that use an adversarial loss in k-space, but such an extension may benefit the reconstruction quality when using the proposed self-supervision approach.

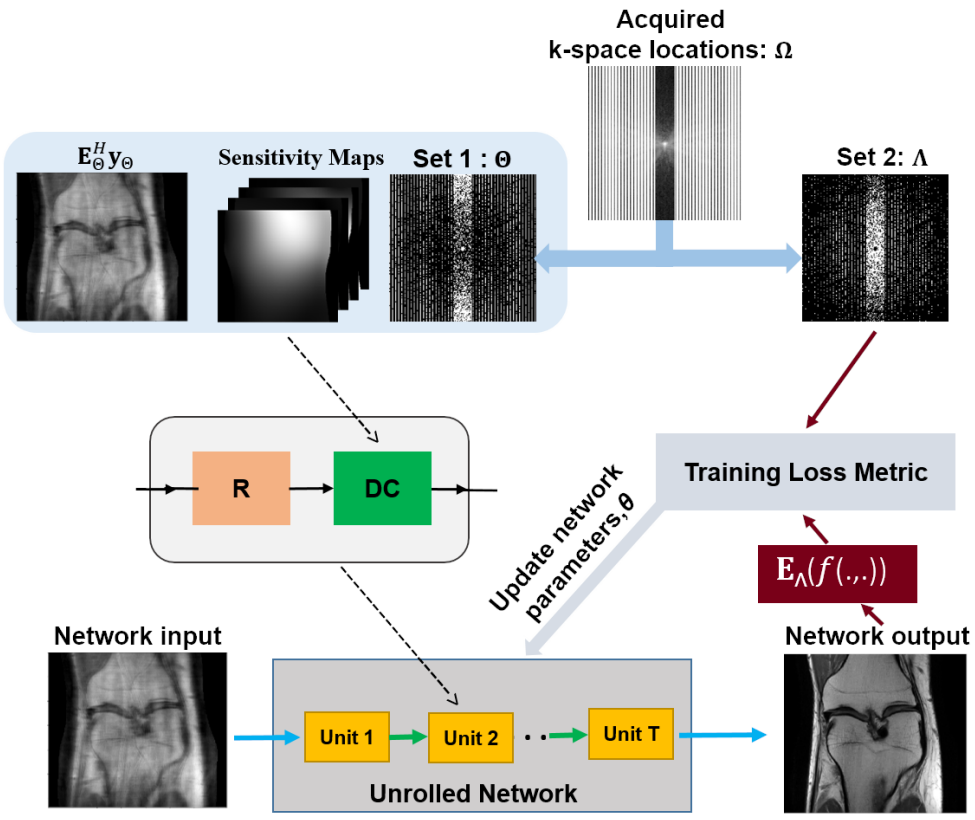

Figure 2.2: The self-supervised learning scheme to train physics-guided deep learning without fully-sampled data. The acquired sub-sampled k-space measurements, Ω, are split into two disjoint sets, Θ and Λ. The first set of indices, Θ, is used in the data consistency unit of the unrolled network, while the latter set, Λ is used to define the loss function for training. During training, the output of the network is transformed to k-space, and the available subset of measurements at Λ are compared with the corresponding reconstructed k-space values. Based on this training loss, the network parameters are subsequently updated.

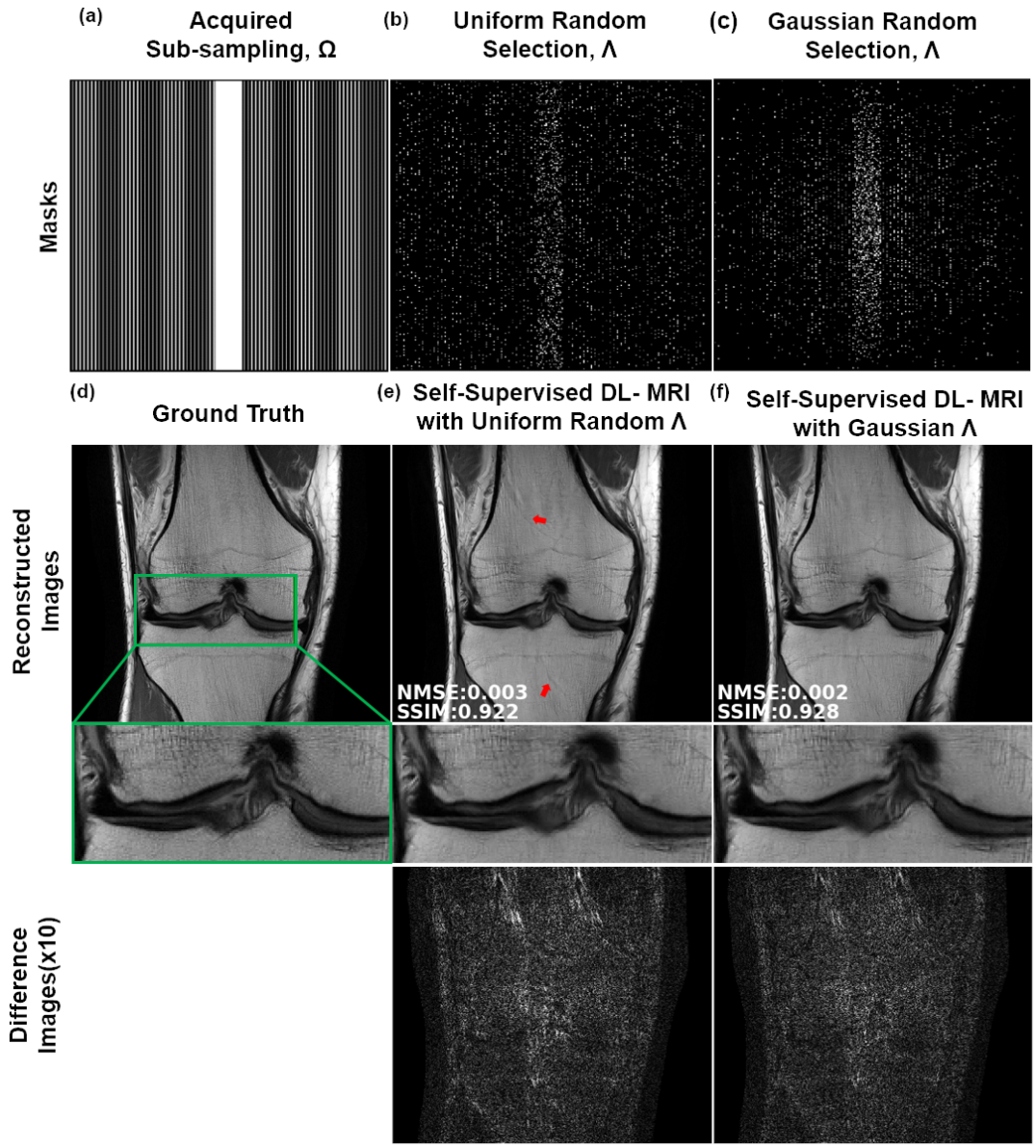

Figure 2.3: a) Acquired sub-sampling pattern, Ω; b) Example uniform random and c) variable-density Gaussian random selection for subset Λ (allowed to differ for each slice in the training dataset) that is used to define the training loss; d) Ground-truth reference data; e) and f) Self-supervised DL-MRI reconstruction and corresponding difference images with loss masks Λ as in b) and c), respectively. Red arrows mark residual artifacts in uniform random selection. These artifacts are further suppressed in the Gaussian random selection, which is used for the remainder of the study.

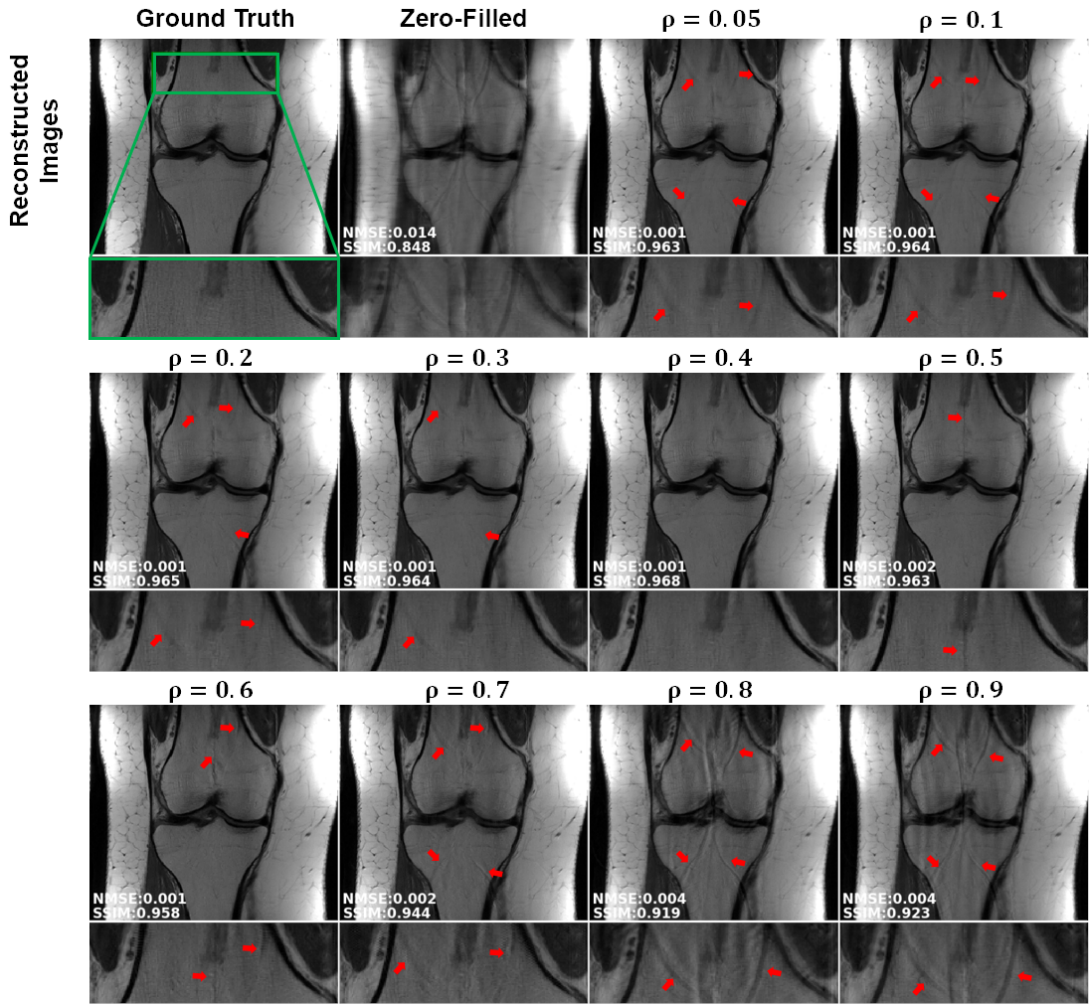

Figure 2.4: A representative test slice depicting the reconstruction results for different ratios of $\rho = \Lambda/\Omega$. Λ is used only for defining loss function, while $\Theta = \Omega \backslash \Lambda$ is only used within data consistency units. Red arrows mark visible residual artifacts for $\rho \leq 0.3$ and $\rho \geq 0.5$. These artifacts are suppressed at $\rho = 0.4$, which is used for the remainder of the study.

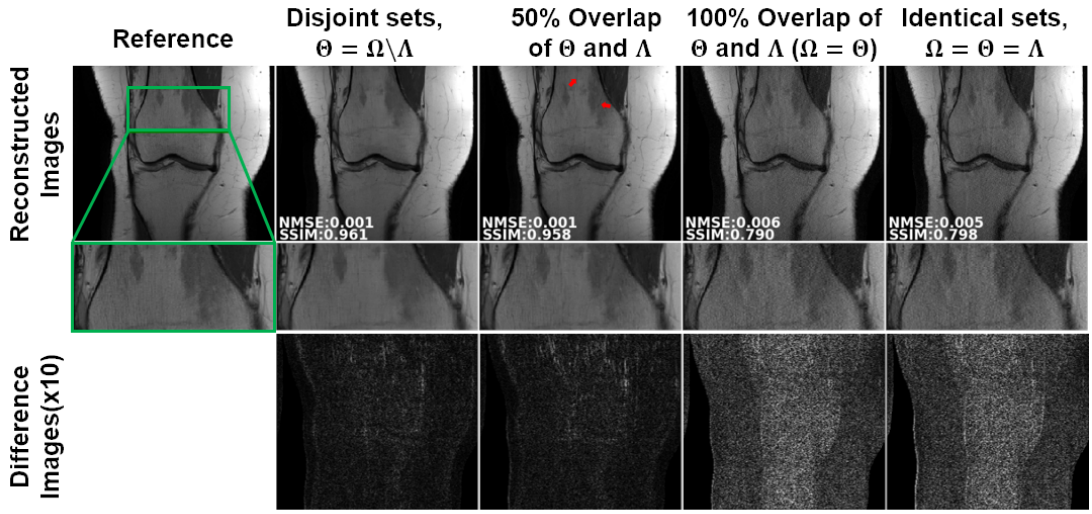

Figure 2.5: Reconstruction results for different degrees of overlap between Λ and Θ, i.e. $|\Lambda \cap \Theta|/|\Lambda|$, for $\rho = \Lambda/\Omega = 0.4$, as well as the limiting case that uses all available data for both data consistency and loss (i.e. $\Omega = \Theta = \Lambda$). For the limiting case with $\Omega = \Theta = \Lambda$, the reconstruction suffers from noise amplification, which is significantly suppressed for the proposed disjoint Λ and Θ. The performance of the self-supervised approach degrades as the amount of overlap increases.

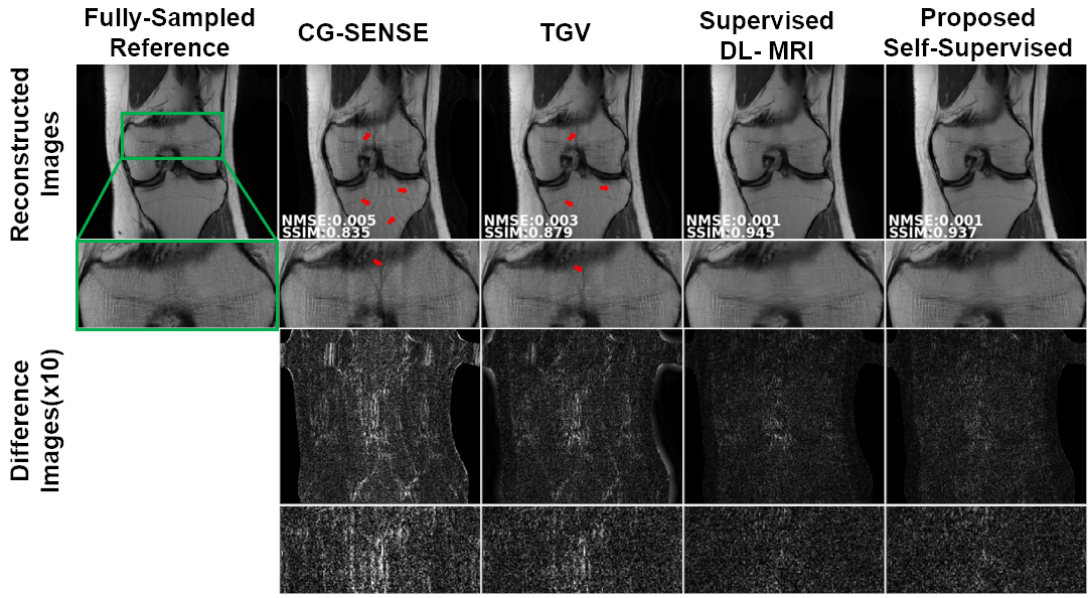

Figure 2.6: A representative test slice from fastMRI coronal PD knee MRI dataset depicting the reconstruction results for proposed self-supervised DL-MRI, supervised DL-MRI, CG-SENSE and TGV approaches for retrospective equispaced undersampling R = 4. Zoomed views and error images show the residual artifacts observed in CG-SENSE and TGV approaches. Both self-supervised and supervised DL-MRI approaches successfully suppress these artifacts, while showing similar quantitative performance.

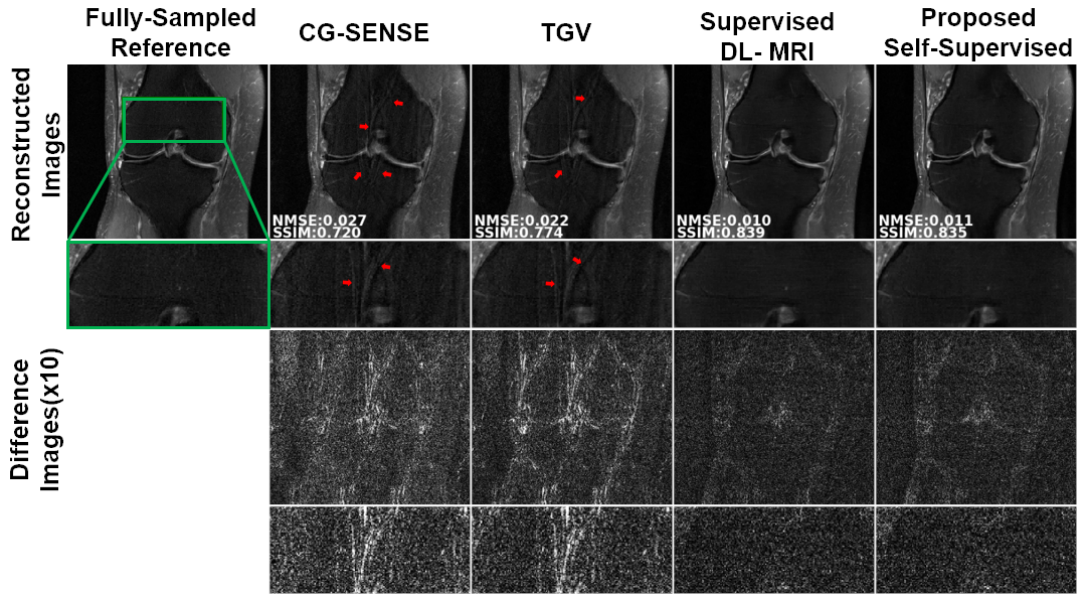

Figure 2.7: A reconstructed test slice showing reconstruction results from fastMRI coronal PD-FS datasets for retrospective equispaced undersampling R = 4. Red arrows indicate visible artifacts, especially apparent in the zoom views and error images for CG-SENSE and TGV techniques. Proposed self-supervised and supervised DL-MRI eliminate these artifacts, while showing similar quantitative and qualitative performance.

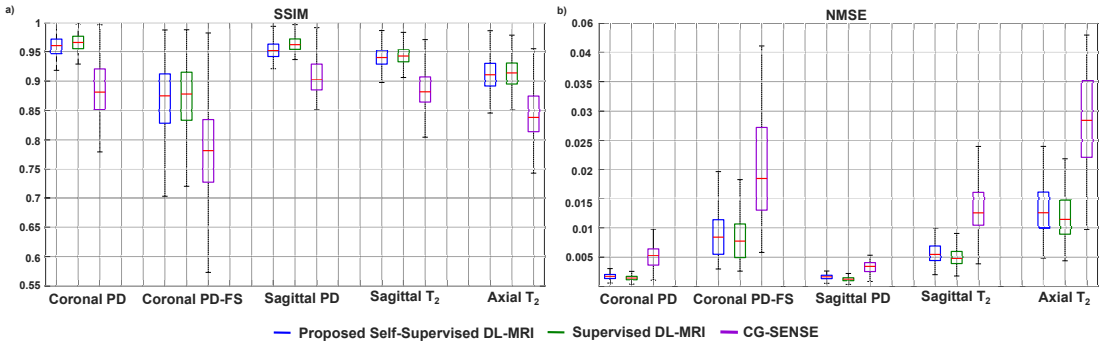

Figure 2.8: Boxplots showing the median and interquartile range (25^{th}-75^{th} percentile) of the quantitative metrics, (a) structural similarity index and (b) normalized mean squared error (NMSE) for all five knee MRI sequences. Both proposed self-supervised and supervised DL-MRI significantly outperform CG-SENSE in terms of SSIM and NMSE for all knee sequences, while showing similar quantitative performance.

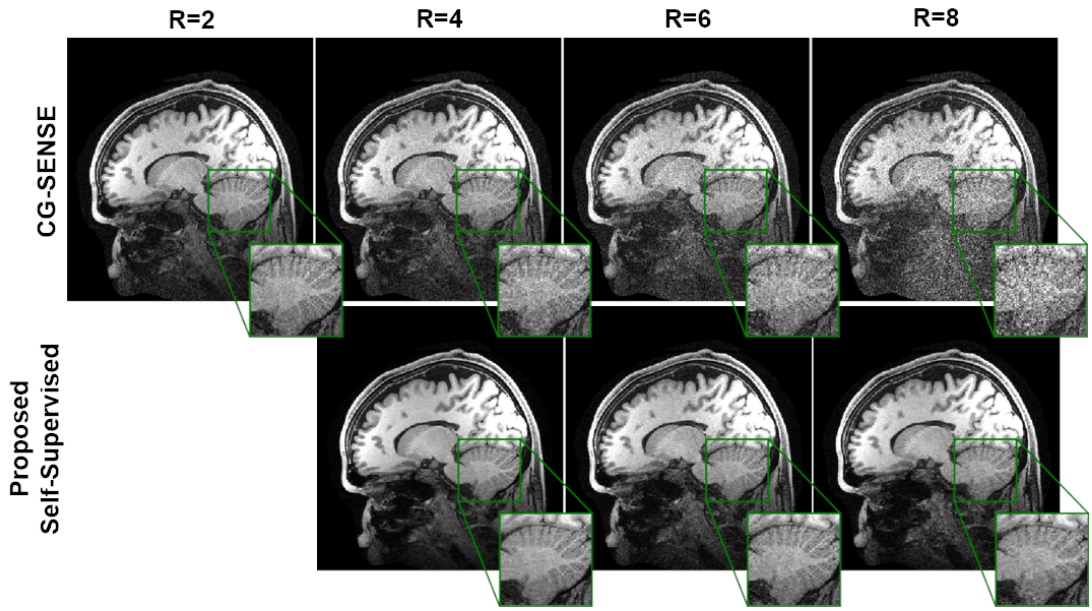

Figure 2.9: Reconstruction results from prospectively 2-fold equispaced undersampled brain MRI. CG-SENSE and the proposed self-supervised approach are applied at further retrospective acceleration rates of 4, 6 and 8 with equispaced sheared $k_y - k_z$ undersampling patterns, while CG-SENSE is also used at the acquisition rate of 2. CG-SENSE suffers from visibly higher noise amplification at high acceleration rates. The proposed approach successfully reconstructs brain MRI at these higher rates, achieving similar image quality to CG-SENSE at R = 2. Note the supervised DL-MRI cannot be applied here due to the lack of fully-sampled ground truth data for training.

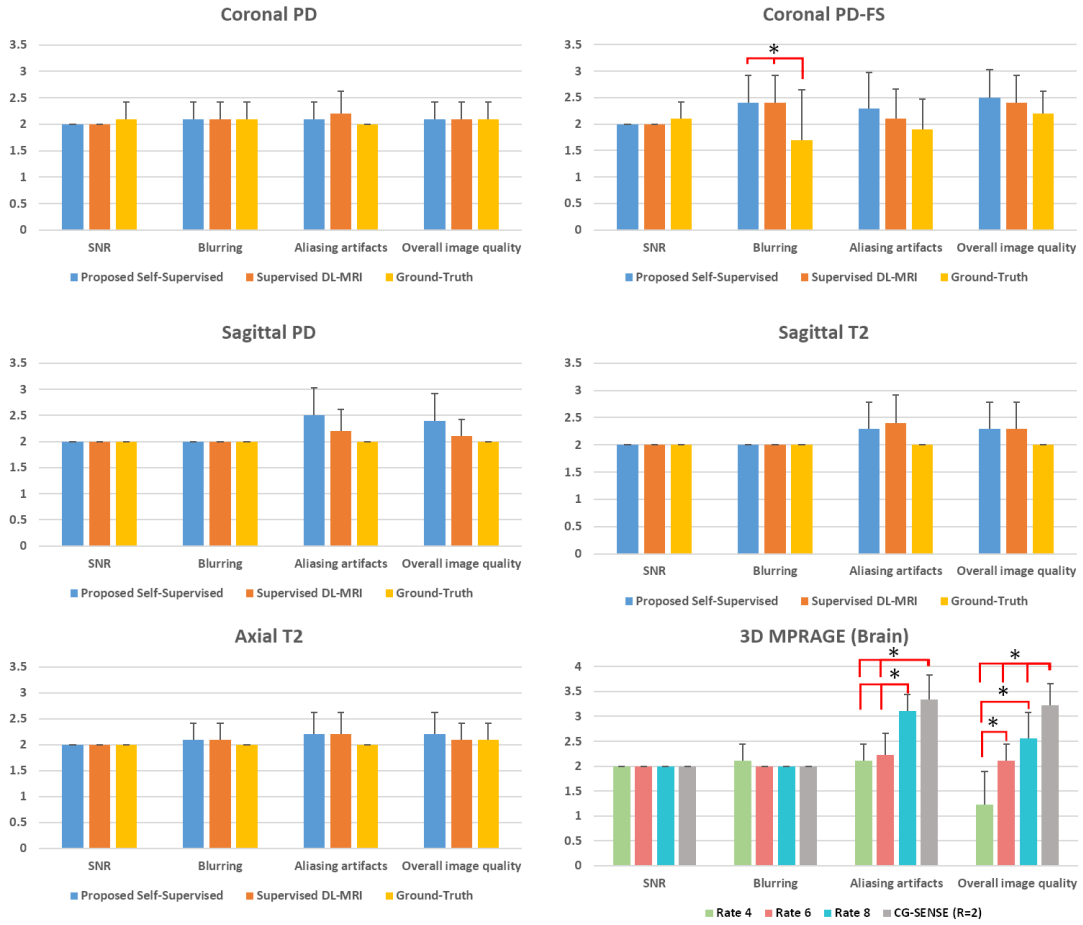

Figure 2.10: The image reading results from the clinical reader study for knee and brain datasets. Bar-plots show average reader scores and their standard deviation across the test subjects. Statistical testing was performed by one-sided Wilcoxon single-rank test, with * showing significant statistical difference with $P < 0.05$. For knee MRI, both supervised and self-supervised DL-MRI approaches get comparable scores to the reference image in terms of SNR, blurring, aliasing artifacts and overall image quality. There was no statistical difference between reference and DL-MRI approaches in terms of the evaluation criterions for the knee datasets, except for blurring between reference and DL-MRI approaches in coronal PD-FS. For brain MRI, CG-SENSE at R = 2 and self-supervision at R = 4, 6 and 8 do not show any significant differences in terms of SNR and blurring. Self-supervision at all rates were evaluated to be significantly improved compared to CG-SENSE in terms of aliasing artifacts and overall image quality. Additionally, self-supervision at R = 6 and 8 were also significantly worse than self-supervision at R = 4 in terms of overall image quality.

Chapter 3

Self-Supervised Physics-Guided Deep Learning Reconstruction For High-Resolution 3D LGE CMR

3.1 Introduction

Late gadolinium enhancement (LGE) cardiac MRI (CMR) is the clinical gold standard for identification of myocardial scar and fibrosis [97]. While 2D LGE CMR remains popular, 3D imaging offers improved SNR and spatial resolution [98, 99]. Isotropic high-resolution 3D LGE further enables better delineation of small structures, potentially improving the assessment of left ventricular scar heterogeneity, right ventricular scar and left atrial scar [100]. However, such high resolution 3D scans are prohibitively long, which is especially problematic in the presence of respiratory motion and contrast washout [99]. Thus, image acceleration by means of parallel imaging [1, 2] and compressed sensing (CS) [4, 6, 101–103], are necessary.

Recently, deep learning (DL) has been used for accelerated MRI due to its improved

This chapter is based on [36].

reconstruction quality over conventional approaches [7–9, 18, 24, 82]. Among such methods, physics-guided DL (PG-DL) techniques unroll conventional iterative algorithms consisting of data consistency (DC) and regularizer units for a fixed number of iterations [7, 9]. The DC units utilize conventional linear methods, while the regularizer units are implicitly implemented using convolutional neural networks (CNNs) [7, 9, 17]. PG-DL approaches are typically trained in a supervised manner using fully-sampled data as ground-truth reference. However, acquisition of high-quality fully-sampled data is not possible in high-resolution 3D LGE CMR due to contrast washout [99]. Thus, methods for training PG-DL reconstruction without fully-sampled data for improved LGE CMR are desirable.

Recently, several methods have been proposed for this goal [10, 82, 104, 105]. Among these, a recent approach named self-supervised learning via data undersampling (SSDU) [10, 35] trains neural networks without fully-sampled data by splitting available measurements into two disjoint sets. One of these is used for the DC units in the unrolled network and the other is used to define the loss function. SSDU was applied to 2D knee and 3D brain MRI [10], showing matching quality to supervised training and improved quality over conventional methods. But in the latter setting, an inverse Fourier transform was applied along the fully-sampled frequency encoding direction, and the slices in this direction were processed individually through a 2D unrolled network. However, a truly 3D processing may further improve reconstruction quality since: 1) Using 3D CNNs in regularizer units may capture higher-dimensional correlations than the 2D case, 2) For self-supervised learning, 3D acquisitions provide a higher degree of freedom along three dimensions for selecting the two subsets for loss and DC units.

In this Chapter, we sought to enable PG-DL reconstruction of 3D LGE CMR by extending the SSDU approach to 3D imaging. Results on prospectively 3-fold undersampled 3D isotropic high-resolution LGE CMR show that the proposed 3D self-supervised approach improves the reconstruction quality compared to the clinically-used CS approach [100], both at the acquisition acceleration rate, $R = 3$, and further retrospective acceleration $R = 6$.

3.2 Materials and Methods

3.2.1 Unrolling Iterative Algorithms

Let $\mathbf{y}_\Omega$ be the acquired k-space data with sub-sampling pattern Ω, and $\mathbf{x}$ be the image to be recovered. The regularized least squares problem for MRI reconstruction is given as

$$\arg\min_{\mathbf{x}} \|\mathbf{y}_\Omega - \mathbf{E}_\Omega \mathbf{x}\|_2^2 + \mathcal{R}(\mathbf{x}), \tag{3.1}$$

where $\|\mathbf{y}_\Omega - \mathbf{E}_\Omega \mathbf{x}\|_2^2$ is a data consistency term, $\mathbf{E}_\Omega : \mathbb{C}^M \rightarrow \mathbb{C}^P$ is the multi-coil encoding operator containing coil sensitivities and partial Fourier matrix sampling, and $\mathcal{R}(\cdot)$ is a regularizer. There exist several approaches to solve Eq. (6.8), such as proximal gradient descent or variable splitting with quadratic penalty (VSQP) [70].

In VSQP, DC and regularizer units are decoupled using an auxilliary variable, $\mathbf{z}$ that is constrained to be equal to $\mathbf{x}$. Then, Eq. (6.8) is reformulated as an unconstrained problem by imposing a quadratic penalty

$$\arg\min_{\mathbf{x}} \|\mathbf{y}_\Omega - \mathbf{E}_\Omega \mathbf{x}\|_2^2 + \mu\|\mathbf{x} - \mathbf{z}\|_2^2 + \mathcal{R}(\mathbf{z}), \tag{3.2}$$

where μ is the penalty parameter. Eq. (6.10) is solved via alternating minimization as

$$\mathbf{z}^{(i)} = \arg\min_{\mathbf{z}} \mu\|\mathbf{x}^{(i-1)} - \mathbf{z}\|_2^2 + \mathcal{R}(\mathbf{z}) \tag{3.3a}$$

$$\mathbf{x}^{(i)} = \arg\min_{\mathbf{x}} \|\mathbf{y}_\Omega - \mathbf{E}_\Omega \mathbf{x}\|_2^2 + \mu\|\mathbf{x} - \mathbf{z}^{(i)}\|_2^2 \tag{3.3b}$$

where $\mathbf{z}^{(i)}$ is an intermediate variable and $\mathbf{x}^{(i)}$ is the desired image at iteration i. In PG-DL, these conventional iterative algorithms are unrolled for a fixed number of iterations, in which each iteration contains a DC and a regularizer unit. In PG-DL, regularizer sub-problem in Eq. (6.12a) is solved with neural networks and DC sub-problem in Eq. (6.12b) is solved via conjugate gradient (CG) [9].

3.2.2 Supervised PG-DL Training

Unrolled networks are trained end-to-end by minimizing a cost function between the network output and a reference. In supervised PG-DL, fully-sampled data is used as reference. Supervised PG-DL performs end-to-end training by minimizing an objective

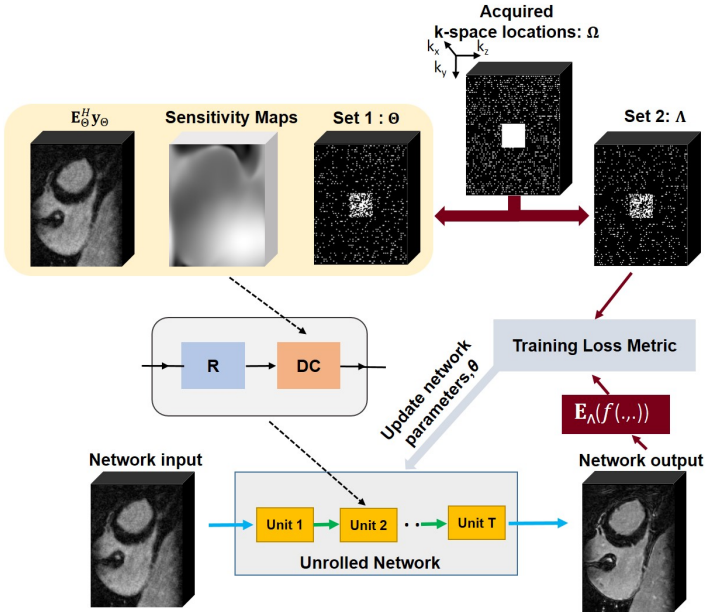

Figure 3.1: The self-supervised PG-DL training without fully-sampled data splits acquired sub-sampled k-space indices Ω, into two disjoint sets, Θ and Λ. The first set of indices, Θ, is used in the DC units of the unrolled network, while the latter set, Λ is used to define the loss function for training. During training, the output of the network is transformed to k-space, and the available subset of measurements at Λ are compared with the corresponding reconstructed k-space values. Based on this training loss, the network parameters are subsequently updated.

cost function given as

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(\mathbf{y}^i_{\text{ref}},\ \mathbf{E}^i_{\text{full}} f(\mathbf{y}^i_{\Omega}, \mathbf{E}^i_{\Omega}; \boldsymbol{\theta})), \tag{3.4}$$

where N is the number of samples in the training database, $\mathcal{L}(\cdot,\cdot)$ is a loss function, $\mathbf{y}^i_{\text{ref}}$ is the fully-sampled k-space for subject i, $f(\mathbf{y}^i_{\Omega}, \mathbf{E}^i_{\Omega}; \boldsymbol{\theta})$ is the output of the unrolled network for sub-sampled k-space data $\mathbf{y}^i_{\Omega}$ with the network being parameterized by $\boldsymbol{\theta}$, and $\mathbf{E}^i_{\text{full}}$ is the fully-sampled encoding operator that transforms network output to k-space.

3.2.3 Proposed 3D Self-Supervised PG-DL Training

When the acquisition of fully-sampled data is either challenging or impossible, hindering supervised training of PG-DL approaches, SSDU enables training without fully-sampled

data by splitting available undersampled measurements, Ω into two disjoint sets, Θ and Λ for DC and computing the loss. More formally, the cost function for training is modified to compute the loss only on indices belonging to Λ.

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\Big(\mathbf{y}^i_{\Lambda},\ \mathbf{E}^i_{\Lambda}\big(f(\mathbf{y}^i_{\Theta}, \mathbf{E}^i_{\Theta}; \boldsymbol{\theta})\big)\Big). \tag{3.5}$$

While the original SSDU was implemented with 2D networks [10], 3D processing is desirable as discussed in Section 5.1. However, in addition to the difficulty of acquiring fully-sampled data for 3D scans, it is challenging to generate large databases of 3D acquisitions to train neural networks with a high number of parameters. We proposed to tackle these issues by extending SSDU to 3D processing, as shown in **Figure 3.1**. In addition to the selection of Θ and Λ via Gaussian-weighted masking in three dimensions [10], we also tackle the issue of training data scarcity in databases by extracting multiple smaller 3D slabs from a whole heart acquisition of each subject. This is done by taking an inverse Fourier transform along the fully-sampled read-out direction direction, dividing the volume in this direction to smaller 3D sub-volumes, and processing the 3D k-space of these volumes.

End-to-end training was performed by unrolling iterative sub-problems in (6.12a)-(6.12b) for $T = 5$ iterations. DC units employed CG and regularizers used the same ResNet structure as in [10], except 2D kernels of size 3×3 were replaced with 3D kernels of $3 \times 3 \times 3$. Coil sensitivity maps were generated using ESPIRiT [106]. The proposed 3D self-supervised PG-DL algorithm was trained using an Adam optimizer with a learning rate of $5 \cdot 10^{-4}$ over 100 epochs by minimizing a normalized ℓ_1 - ℓ_2 loss [10]. All experiments for PG-DL approaches were performed using Tensorflow in Python.

3.2.4 Imaging Experiments and Evaluation

3D LGE CMR were acquired axially at 1.5T with a 32-channel coil array on 18 patients. The imaging protocols were approved by the local institutional review board, and written informed consent was obtained from all participants. Imaging parameters were: TR/TE = 5.2/2.6 ms, FOV = 320 $\times$ 320 $\times$ 100 mm^3, resolution = 1.2 $\times$ 1.2 $\times$ 1.2 mm^3, ACS = 40 $\times$ 24, prospective acceleration $R = 3$ with random $k_y - k_z$ sub-sampling [100].

The prospectively subsampled 3D LGE data was further retrospectively subsampled to $R = 6$ by keeping a 24×24 ACS region in the $k_y - k_z$ plane using a random uniform undersampling pattern. Due to the small training database size, smaller $20 \times 270 \times 102$ 3D slabs for training were generated as described in Section 3.2.3. Training was performed on 200 sub-volumes from 10 different subjects, and testing was performed on the whole volume of 8 different subjects.

The proposed 3D self-supervised PG-DL approach was performed at $R \in \{3, 6\}$. For $R = 6$ training, only data available at this rate were used. Results were compared with a clinically-used CS approach [100], LOST (Low-dimensional structure self-learning and thresholding) applied at $R = 3$. We note that results could not be compared with supervised learning due to a lack of fully-sampled data. Quantitative metrics such as PSNR or SSIM were also not available due to the lack of ground-truth data. Qualitative assessment of the reconstruction image quality was evaluated by an experienced cardiologist using evaluation criteria of perceived SNR, blurring and overall image quality. The reader was blinded to the reconstruction methods and R. Evaluations were based on a 4-point ordinal scale for blurring (1: no blurring, 2: mild blurring, 3: moderate blurring, 4: severe blurring), perceived SNR (1:excellent, 2: good, 3: fair, 4: poor), and overall image quality (1: excellent, 2: good, 3: fair, 4: poor). Wilcoxon signed-rank test was used to evaluate the scores with a significance level of $P < 0.05$.

3.3 Results

Figure 3.2 shows reconstruction results on a representative test slice with negative LGE. The proposed 3D self-supervised approach was applied at both prospective acceleration $R = 3$ and further retrospective acceleration $R = 6$, while LOST-CS was applied only at prospective acceleration $R = 3$. LOST-CS reconstruction shows a mixture of noise-like amplification and incoherent aliasing artifacts due to random undersampling, especially in the blood pool. The proposed 3D self-supervised approach at both $R = 3$ and 6 suppress these artifacts further and achieves improved reconstruction quality.

Figure 3.3 shows reconstruction results of LOST-CS and the proposed 3D self-supervised approach on a subject with positive LGE (marked by red arrows). Similar observations apply in this case, with the proposed method at $R = 3$ and 6 suppressing

the residual artifacts in LOST-CS at $R = 3$, while also allowing a sharper depiction of the myocardium-blood border. All approaches show the enhancement region clearly.

Figure 3.4 summarizes the reader assessment for the 3D LGE CMR dataset. Barplots show the average and standard deviation of reader scores across the test dataset. For blurring, all methods were statistically in good agreement, while the proposed 3D self-supervised at $R = 3$ and $R = 6$ was rated higher than LOST-CS at $R = 3$. For both perceived SNR and overall image quality, proposed 3D self-supervised at $R = 3$ and 6 were rated statistically better than LOST-CS at $R = 3$ which is used in clinical studies.

3.4 Discussion

The proposed 3D self-supervised PG-DL reconstruction enables training neural networks without fully-sampled data for 3D volumes, by splitting available measurements into two disjoint sets, and using one of these in the DC units and the other for computing the loss. Moreover, we proposed to tackle the issue of having a small number of training subjects in the database for 3D applications by splitting the whole volume for each subject into sub-volumes and training over these sub-volumes. Results on 3D LGE CMR show that the proposed approach significantly improves upon the state-of-the-art CS methods.

Training of 3D PG-DL methods has gained interest recently due to their ability to capture higher-dimensional interactions. Several supervised PG-DL approaches have been proposed for 3D training [89,107], either by using fully-sampled data or a surrogate reconstruction, such as CS or parallel imaging, as reference. However, the former is difficult in many CMR acquisitions due to scan length, contrast washout or motion; while the latter inherently limits the potential of PG-DL to existing reconstruction strategies. While self-supervised learning can tackle these issues by efficient utilization of acquired measurements, small training dataset size and GPU-memory issue of fitting large volumes [89] are common challenges for both supervised and self-supervised PG-DL approaches. The proposed approach of splitting large volumes into sub-volumes provides an alternative solution to both of these issues.

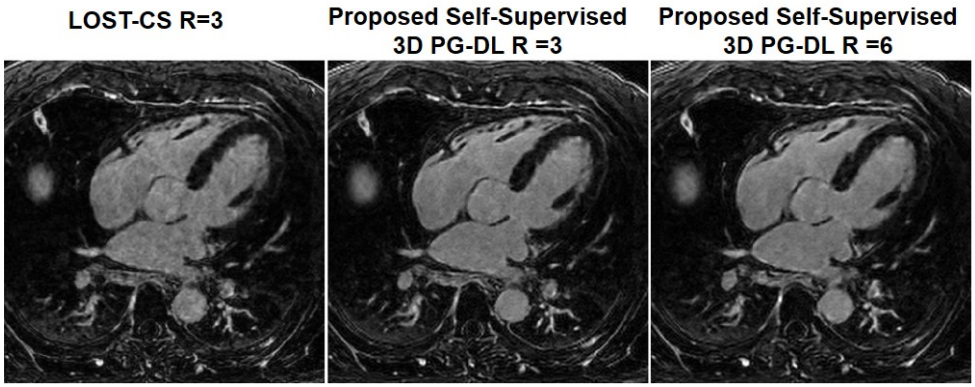

Figure 3.2: Reconstruction results from a representative test slice without enhancement. LOST-CS was applied at the acquisition rate of 3, while the proposed 3D self-supervised PG-DL approach was used at $R = 3$ and 6. LOST-CS suffers from visible noise-like and incoherent residual artifacts. The proposed approach provides improved reconstruction at both $R = 3$ and 6. We further note that the proposed approach at $R = 6$ only uses the data available at this rate for training, and does not have access to $R = 3$ data.

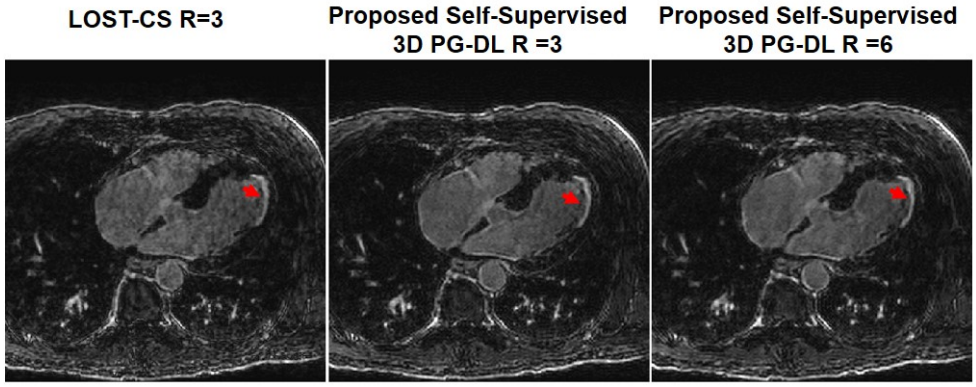

Figure 3.3: Reconstruction results from a representative test slice with positive LGE. The proposed self-supervised PG-DL approach at both $R = 3$ and 6 outperform LOST-CS reconstruction at $R = 3$ by suppressing noise and residual artifacts. All reconstruction methods successfully identify LGE shown with red arrows.

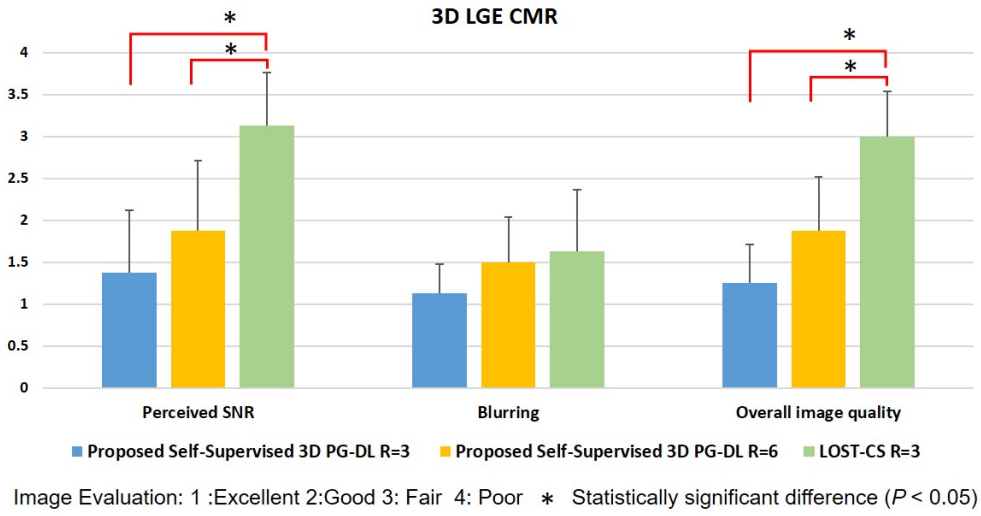

Figure 3.4: The image reading results from the clinical reader study for the 3D LGE CMR. Evaluations were based on a 4-point ordinal scale (1:best, 4: worst). Bar-plots depict average and standard deviation across test subjects, with * showing statistically significant differences. For blurring, proposed self-supervised 3D PG-DL at $R = 3$ and 6 were rated higher than LOST-CS at R=3, though the differences were not significant. For perceived SNR and overall image quality, proposed self-supervised 3D PG-DL at $R = 3$ and $R = 6$ were both rated statistically better than LOST-CS at $R = 3$.

Chapter 4

Multi-Mask Self-Supervised Learning for Physics-Guided Deep Learning in Highly Accelerated MRI

4.1 Introduction

Data acquisition is lengthy in many MRI exams, creating challenges for improving resolution and coverage, hence making accelerated MRI reconstruction an ongoing research topic. Parallel imaging [1–3] and compressed sensing [4–6, 14, 20, 108] are two commonly used acceleration methods, with the former being the clinical gold standard for fast MRI, and the latter providing additional acceleration in a number of scenarios. However, acceleration rates remain limited as reconstructed images may suffer from noise amplification [57] or residual artifacts [60,109] in parallel imaging and compressed sensing, respectively. Recently, deep learning (DL) has emerged as an alternative for accelerated MRI due to its improved reconstruction quality compared to conventional approaches, especially at higher acceleration rates [7, 8, 18, 23, 90, 96].

Among DL methods, physics-guided DL approaches, which incorporate the MRI

This chapter is based on [38, 39].

encoding operator to solve a regularized inverse problem, have gained interest due to its robustness [18, 22]. Physics-guided DL approaches unroll an iterative process that alternates between data consistency (DC) and regularization for certain number of iterations. They are trained end-to-end, typically in a supervised manner by minimizing the difference between network output and a ground-truth reference obtained from fully-sampled data [7, 9, 28, 30]. However, acquisition of fully-sampled data, especially on large patient populations, is either challenging or impossible in many practical scenarios [33, 34, 66–68].

As supervised training becomes inoperative in the absence of fully-sampled data, several methods have been proposed to train networks without fully-sampled data [10, 35, 80, 84, 86]. Among these approaches, Self-supervision via Data Undersampling (SSDU) trains physics-guided neural networks by utilizing only the acquired sub-sampled measurements [10]. In SSDU, the available measurements are split into two disjoint sets by a masking operation, which reduces the sensitivity to overfitting and is central for reliable performance. One of these sets is used in the DC units of the network, and the other is used to define the loss function in k-space. For moderately high acceleration rates, the networks trained using SSDU match the performance of those from supervised learning. While SSDU demonstrated that the splitting of acquired points into two sets was sufficient for training a neural network for reconstruction from undersampled data, a strategy that augments the use of the subsampled data to improve the performance is essential for higher acceleration rates.

In this Chapter, we sought to improve the performance of SSDU with multiple masks. The proposed multi-mask SSDU splits acquired measurements into multiple pairs of disjoint sets for each training slice, while using one of these sets in a pair for DC units and the other for defining loss, similar to the original SSDU. The proposed multi-mask SSDU approach is applied on fully-sampled 3D knee MRI datasets from mridata.org [110], as well as a prospectively undersampled high-resolution 3D brain MRI dataset, and compared to parallel imaging, SSDU with a single mask [10], and supervised DL-MRI when fully-sampled data is available. Results show that the proposed multi-mask SSDU approach at high acceleration rates significantly improves upon SSDU and closely performs with supervised DL-MRI, while the reader studies indicate that the proposed multi-mask approach also outperforms supervised DL-MRI approach in terms

of SNR improvement and aliasing artifact reduction.

4.2 Materials and Methods

4.2.1 Supervised Training of Physics-Guided DL-MRI Reconstruction

Let $\mathbf{y}_\Omega$ be the acquired subsampled measurements with Ω denoting the subsampling pattern and $\mathbf{x}$ the image to be recovered. The forward model for encoding is

$$\mathbf{y}_\Omega = \mathbf{E}_\Omega \mathbf{x} + \mathbf{n} \tag{4.1}$$

where $\mathbf{E}_\Omega : \mathbb{C}^M \rightarrow \mathbb{C}^P$ is the encoding operator including the coil sensitivities and a partial Fourier matrix sampling the locations specified by Ω, and $\mathbf{n} \in \mathbb{C}^P$ is measurement noise. For sub-Nyquist sampling at high rates, the forward model may be ill-conditioned, necessitating the use of regularization, leading to an inverse problem for image reconstruction:

$$\arg\min_{\mathbf{x}} \|\mathbf{y}_\Omega - \mathbf{E}_\Omega \mathbf{x}\|_2^2 + \mathcal{R}(\mathbf{x}), \tag{4.2}$$

where the first term represents DC and $\mathcal{R}(\cdot)$ is the regularizer. Several approaches may be used to iteratively solve the above optimization problem [70]. In this work, we use variable splitting via quadratic penalty method [10, 30, 70], which decouples DC and regularizer operations:

$$\mathbf{z}^{(i)} = \arg\min_{\mathbf{z}} \mu \|\mathbf{x}^{(i-1)} - \mathbf{z}\|_2^2 + \mathcal{R}(\mathbf{z}) \tag{4.3a}$$

$$\mathbf{x}^{(i)} = \arg\min_{\mathbf{x}} \|\mathbf{y}_\Omega - \mathbf{E}_\Omega \mathbf{x}\|_2^2 + \mu \|\mathbf{x} - \mathbf{z}^{(i)}\|_2^2 \tag{4.3b}$$

where μ is the quadratic penalty parameter, $\mathbf{x}^{(i)}$ is the network output at iteration i, $\mathbf{z}^{(i)}$ is an intermediate variable, and $\mathbf{x}^{(0)}$ is the initial image obtained from zero-filled undersampled k-space data. In physics-guided DL, this iterative optimization is unrolled for a fixed number of iterations. Eq. 6.12a corresponds to a regularizer, which is implicitly solved by a neural network, whereas Eq. 6.12b has a closed form solution [10] that can be solved by gradient descent methods such as conjugate gradient [9].

In traditional DL-MRI approaches, training datasets contain pairs of undersampled k-space/ image and fully-sampled k-space/ground-truth image [7, 9, 17, 30]. Let $\mathbf{y}^i_{\text{ref}}$ be the fully-sampled reference k-space data for subject i and $f(\mathbf{y}^i_\Omega, \mathbf{E}^i_\Omega; \boldsymbol{\theta})$ denote the output

of the unrolled network that is parametrized by $\boldsymbol{\theta}$ for subsampled k-space data $\mathbf{y}^i_\Omega$, and corresponding encoding matrix $\mathbf{E}^i_\Omega$. The supervised PG-DL training is performed by defining the loss function in image domain or k-space [10]. Training can be performed by minimizing a k-space loss function as

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(\mathbf{y}^i_{\text{ref}},\ \mathbf{E}^i_{full} f(\mathbf{y}^i_\Omega, \mathbf{E}^i_\Omega; \boldsymbol{\theta})), \tag{4.4}$$

where N is the number of fully-sampled training data in the database, $\mathbf{E}^i_{full}$ is the fully-sampled encoding operator that transforms network output to k-space, and $\mathcal{L}(\cdot,\cdot)$ is the loss between the fully-sampled and reconstructed k-spaces. The sampling locations, Ω may vary per subject in a more general setup, i.e. indexed by i. However, this was not included for simplicity of notation.

4.2.2 Self-Supervision via Data Undersampling (SSDU)

In order to enable training without fully-sampled datasets, SSDU has been proposed [10], where the acquired sub-sampled data indices, Ω, from each scan are divided into two disjoint sets Θ and Λ. Θ is used in DC units in the unrolled network and Λ is used to define the loss function, and the following self-supervised loss function is minimized

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\Big(\mathbf{y}^i_\Lambda,\ \mathbf{E}^i_\Lambda\big(f(\mathbf{y}^i_\Theta, \mathbf{E}^i_\Theta; \boldsymbol{\theta})\big)\Big). \tag{4.5}$$

Unlike the supervised approach, only a subset of measurements, Θ are used as the input to the unrolled network. The network output is transformed to k-space, where the loss is calculated only at unseen k-space indices, Λ. After training is completed, testing is performed on unseen data using all available measurements Ω.

4.2.3 Proposed Multi-Mask SSDU

SSDU reconstruction quality degrades at very high acceleration rates due to higher data scarcity, arising from the splitting into Θ and Λ. In order to tackle this issue, we propose a multi-mask SSDU approach, which retrospectively splits acquired indices Ω into disjoint sets Θ and Λ multiple times as shown in **Figure 4.1**. Formally, we split

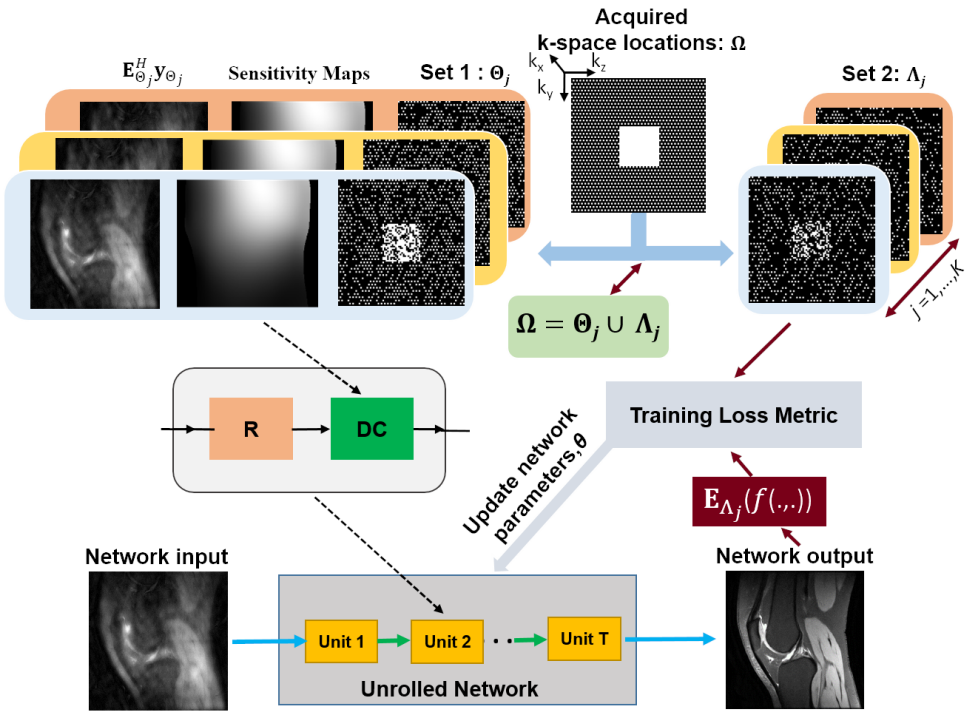

Figure 4.1: The proposed multi-mask self-supervised learning for PG-DL MRI reconstruction. Acquired k-space locations for each scan, Ω, are retrospectively sub-sampled into two disjoint sets of Θ_j and Λ_j for $j \in \{1, \ldots, K\}$. For each such partitioning, Θ_j is used for DC units and $\Lambda_j = \Omega/\Theta_j$ is used to define the loss function. Loss is performed in k-space by comparing acquired data with the multi-coil k-space of the network output at indices Λ_j. Based on this training loss, the network parameters are subsequently updated.

available measurements multiple times for each subject i such that for each partition partitioned K times such that

$$\Omega = \Theta_j \cup \Lambda_j, \quad \text{j} = 1, \ldots, K. \tag{4.6}$$

where K denotes the number of partitions for each scan. Similar to SSDU, each pair of sets in each scan were disjoint, i.e., $\Lambda_j = \Omega/\Theta_j$ for $\text{j} = 1, \ldots, K$. In other words, Ω is retrieved by the union of each pair of Λ_j and Θ_j for any $\text{j} = 1, \ldots, K$. Hence, the loss

function to minimize during training becomes

$$\min_{\boldsymbol{\theta}} \frac{1}{N \cdot K} \sum_{i=1}^{N} \sum_{j=1}^{K} \mathcal{L}\Big(\mathbf{y}^i_{\Lambda_j},\, \mathbf{E}^i_{\Lambda_j}\big(f(\mathbf{y}^i_{\Theta_j}, \mathbf{E}^i_{\Theta_j}; \boldsymbol{\theta})\big)\Big). \tag{4.7}$$

The proposed multi-mask approach enables efficient use of available data by ensuring a higher fraction of low and high frequency components are utilized in the training and loss masks. Such utilization was inherently limited in the original SSDU approach, since each acquired k-space point was either used in training or loss masks only once.

4.2.4 3D Imaging Datasets

Fully-sampled 3D knee dataset were obtained from mridata.org [110],which were acquired with approval from the local institutional review board on a 3T GE Discovery MR 750 system with an 8-channel knee coil array using a fast spin-echo (FSE) sequence. Relevant imaging parameters were: FOV = 160 × 160 × 154 mm^3 , resolution = 0.5 × 0.5 × 0.6 mm^3, matrix size = 320 × 320 × 256.

Brain imaging was performed on healthy subjects using a standard Siemens 3D-MPRAGE sequence at a 3T Siemens Magnetom Prisma (Siemens Healthcare, Erlangen, Germany) system using a 32-channel receiver head coil-array [10]. The imaging protocols were approved by the local institutional review board, and written informed consent was obtained from all participants before each examination for this HIPAA-compliant study. Relevant imaging parameters were: FOV = 224 × 224 × 157 mm^3, resolution = 0.7 × 0.7 × 0.7 mm^3, matrix size = 320 × 320, prospective acceleration R = 2 (uniform in k_y), ACS lines = 32 [10].

The 3D k-space datasets were inverse Fourier transformed along the read-out direction, and these slices were processed individually. The knee and brain datasets were retrospectively undersampled to R = 8 using a uniform sheared 2D undersampling pattern [79]. Additionally, for the knee datasets, where a fully-sampled reference is available, further undersampling was performed at R = 8 using uniform 1D and 2D ($k_y - k_z$) random, and 1D and 2D and Poisson undersampling masks. The undersampling masks are provided in Supporting Information **Figure A.14**. Finally, knee datasets were also undersampled to R = 12 using 2D random and Poisson undersampling masks. A 24×24 and 32×32 ACS regions in the $k_y - k_z$ plane were kept fully-sampled for knee and brain

datasets, respectively [10]. The training sets for both knee and brain datasets consisted of 300 slices from 10 subjects, formed by taking 30 slices from each subject. For knee MRI, 2 different subjects with 200 slices were used for validation in multi-mask hyperparameter tuning, and 8 other different subjects were used for testing of the final method. For brain dataset, the testing was performed on 9 different subjects. The proposed multi-mask SSDU approach was compared to SSDU and CG-SENSE for both datasets, as well as supervised DL-MRI for knee MRI.

4.2.5 Choice of Multi-Mask Hyperparameters

There are several tunable hyperparameters in multi-mask SSDU, including the number of partitions, K in Eq. 4.7, as well as the distribution and size of Λ as in SSDU. A variable-density Gaussian distribution was used for Λ in [10] for a single mask. In this study, we used a uniformly random distribution for the proposed approach, as the benefits of a variable density distribution diminish with multiple masks (Supporting Information **Figure A.15**). In [10], the size of Λ was optimized to $\rho = |\Lambda|/|\Omega| = 0.4$, which is also the optimal choice for the distribution considered here (Supporting Information **Figure A.16**). After these two hyperparameters were set, the number of partitions of each scan, K was varied among 3, 5, 6, 7, 8 and 10 to optimize the remaining distinct hyperparameter of the multi-mask SSDU.

4.2.6 Network and Training Details

The iterative optimization problem in (6.12a)-(6.12b) are unrolled for $T = 10$ iterations. Conjugate gradient descent was used in DC units of the unrolled network [9, 10].The proximal operator corresponding to the solution of Eq.6.12a) employs the ResNet structure used in SSDU [10]. It comprises input and output convolution layers and 15 residual blocks (RB) each containing two convolutional layers, where the first layer is followed by a rectified linear unit (ReLU) and the second layer is followed by a constant multiplication layer. All layers had a kernel size of 3×3, 64 channels. The unrolled network which shares parameters across the unrolled iterations had a total of 592,129 trainable parameters. As in SSDU, a ResNet structure as used for the regularizer in Eq. 6.12a,, where the network parameters were shared across the unrolled network [10]. Coil sensitivity

maps were generated from central 24×24 ACS using ESPIRiT [106].

As a pre-processing step, maximum absolute value of the k-space for each slice in the datasets was normalized to 1 in all cases. The real and imaginary parts of the complex MRI dataset were concatenated as two channels prior to inputting into the network. Separate networks for knee and brain datasets were trained using the Adam optimizer with a learning rate of 5×10^{-4}, by minimizing a normalized $\ell_1 - \ell_2$ loss function with a batch size of 1 over 100 epochs. All training was performed using Tensorflow in Python, and processed on a workstation with an NVIDIA Tesla V100 GPU with 32 GB memory.

4.2.7 Image Evaluation

Quantitative assessment of experimental results was performed using normalized mean square error (NMSE) and structural similarity index (SSIM) when fully-sampled data was available as reference. Moreover, qualitative assessment of the image quality from different reconstruction methods was performed by an experienced radiologist. For knee MRI, proposed multi-mask SSDU was compared with ground-truth obtained from fully-sampled data, SSDU and parallel imaging method CG-SENSE, all at R = 8 with 2D uniform undersampling. For brain MRI, the proposed multi-mask SSDU was compared with SSDU at R = 8. Additionally, CG-SENSE approach at the acquisition acceleration R = 2 was evaluated to serve as the clinical baseline. The reader, with 15 years of experience for musculoskeletal and neuro imaging, was blinded to the reconstruction method, which were shown in a randomized order to avoid bias except for the knowledge of the reference image in knee MRI dataset. Evaluations were based on a 4-point ordinal scale, adopted from [7] for blurring (1: no blurring, 2: mild blurring, 3: moderate blurring, 4: severe blurring), SNR (1: excellent, 2: good, 3: fair, 4: poor), aliasing artifacts(1: none, 2:mild, 3: moderate, 4: severe) and overall image quality (1: excellent, 2: good, 3: fair, 4: poor). Wilcoxon signed-rank test was used to evaluate the scores with a significance level of $P < 0.05$.

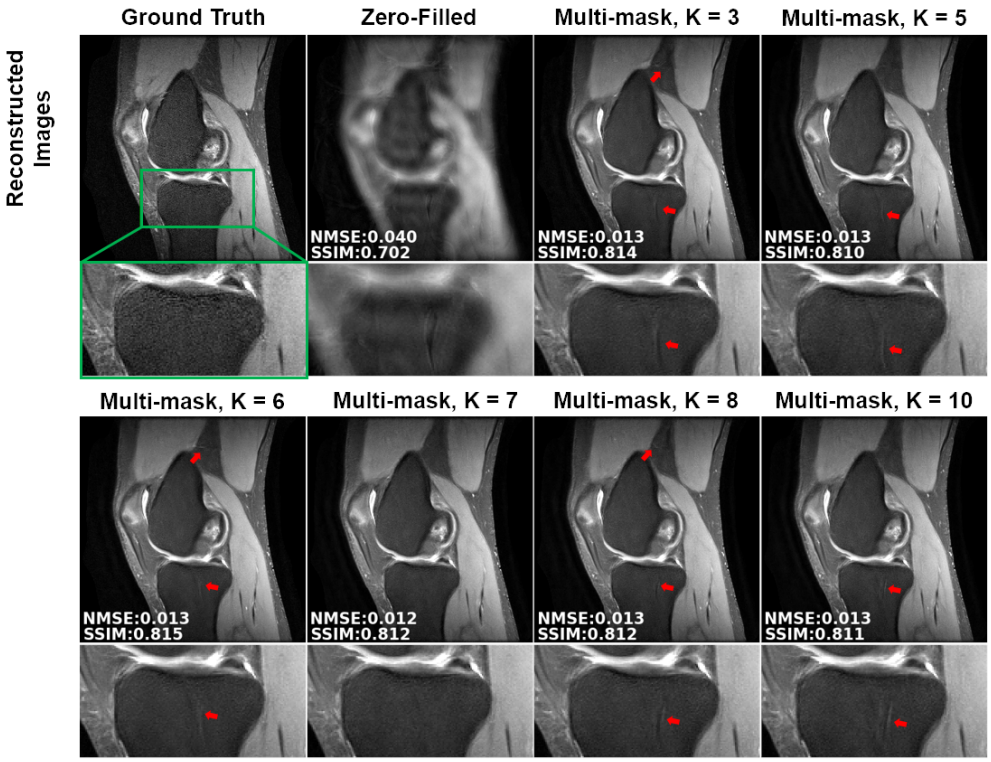

Figure 4.2: A representative test slice showing the reconstruction results for different number of partitions K. Red arrows mark residual artifacts for $K \leq 6$ and $K \geq 8$. These artifacts are suppressed at $K=7$, which is used for the remainder of the study.

4.3 Results

4.3.1 Number of Partitions for Multi-Mask SSDU

Figure 4.2 shows the effect of the proposed multi-mask self-supervised network training at R=8 using 2D uniform sheared undersampling with varying number of masks, K = 3, 5, 6, 7, 8, 10, as well as the ground-truth reference and the zerofilled undersampled data. Multi-mask SSDU approach suppresses residual artifacts as K increases from 3 to 6. At $K = 7$, the visible residual artifacts are removed completely. When K is further increased to 8 and 10, residual artifacts reappear. The quantitative assessment on validation dataset further confirms this qualitative assessment. The median and interquartile range of SSIM values on validation set were 0.8256 [0.7980, 0.8507], 0.8260

[0.8002, 0.8516], 0.8264 [0.8016, 0.8527], 0.8267 [0.8027, 0.8537], 0.8263 [0.8007, 0.8519], 0.8257 [0.7989, 0.8511], and NMSE values were 0.0138 [0.0121, 0.0158], 0.0135 [0.0119, 0.0158], 0.0135 [0.0119, 0.0157], 0.0134 [0.0118, 0.0156], 0.0135 [0.0119, 0.0158], 0.0137 [0.0121, 0.0159] using Gaussian selection for $K \in 3, 5, 6, 7, 8, 10$, respectively. Hence, $K = 7$ used for the remainder of the study.

4.3.2 3D Imaging Datasets

Figure 4.3 depicts representative reference and reconstruction results of the 3D knee dataset using CG-SENSE, supervised DL-MRI, SSDU and proposed multi-mask SSDU, as well as the difference images of these methods with respect to the reference for 2D uniform sheared R = 8 undersampling mask. Red arrows indicate that CG-SENSE suffers from highly-visible artifacts. SSDU alleviates these artifacts substantially, though residual artifacts remain. The proposed multi-mask SSDU further removes these artifacts for both slices shown, while achieving similar reconstruction quality to supervised DL-MRI for the first slice, and further reducing the residual aliasing artifacts visible in the supervised DL-MRI approach for the second slice. Quantitative metrics and difference images in the figure further confirm that multi-mask SSDU outperforms SSDU, while performing similarly to supervised DL-MRI. Additional experimental results on 3D knee dataset using different undersampling patterns at R=8 and R=12 are provided in Supporting Information Figures A.17,A.18, and A.19. In all these experiments, multi-mask SSDU visibly and quantitatively outperforms SSDU, further reducing the gap between self-supervised learning and supervised DL-MRI.

Table 4.1 summarizes the median and interquartile ranges for NMSE and SSIM values metrics for different undersampling masks and acceleration rates across the whole knee MRI test datasets. In all cases, CG-SENSE reconstruction is significantly outperformed by all the DL approaches. Among DL approaches, supervised DL-MRI outperforms self-supervised learning methods, while multi-mask SSDU quantitatively improves upon SSDU.

Figure 4.4 demonstrates CG-SENSE reconstruction of a slice of the 3D-MPRAGE dataset at prospective acceleration R = 2, as well as CG-SENSE, SSDU and the proposed multi-mask SSDU approach at retrospective acceleration R = 8 using 2D uniform

sheared undersampling mask. SSDU at high acceleration R = 8 achieves similar reconstruction quality as CG-SENSE at acquisition acceleration R=2. Multi-mask SSDU further improves reconstruction quality by suppressing the noise evident in SSDU and CG-SENSE.

4.3.3 Image evaluation scores

Figure 4.5a and **b** summarize the reader study results for knee and brain datasets for 2D uniform sheared R = 8 undersampling mask., respectively. For knee MRI, proposed multi-mask SSDU was rated highest in terms of SNR, with a statistically significant improvement over all methods except supervised DL-MRI. For blurring, ground truth data was rated better than all methods. In terms of aliasing artifacts and overall image quality, the proposed multi-mask SSDU approach was rated best compared to other methods and the ground truth. In terms of these two evaluation criteria, all DL-MRI approaches and the reference showed similar statistical behavior, except SSDU was statistically worse than proposed multi-mask SSDU and supervised approach in terms of aliasing artifacts. A more comprehensive comparison also containing reader scores for CG-SENSE is presented in Supporting Information **Figure A.20**.

For brain MRI, DL-MRI reconstructions trained using the proposed multi-mask SSDU and SSDU approach at acceleration rate of 8 performed similar with CG-SENSE at acquisition R = 2 in terms of SNR and blurring. However, in terms of aliasing artifacts, the proposed multi-mask SSDU significantly outperformed its counterparts. In terms of overall image quality, both SSDU methods at R = 8 showed statistically significant improvement over CG-SENSE at R = 2, while the proposed multi-mask SSDU achieved the best performance.

4.4 Discussion

In this work, we extended our earlier work on self-supervision via data undersampling, which trains physics-guided neural network without fully-sampled data, to a multi-mask setting where multiple pairs of disjoint sets were used for each training slice in the dataset. Training of physics-guided DL-MRI reconstruction without ground-truth data remains an important topic, since acquisition of fully-sampled data is either impossible

or challenging in a number of scenarios [33, 34, 66–68]. Among multiple methods proposed for this goal [10,80,84–86], self-supervision directly uses the acquired data without relying on generative models or intermediate estimates. Our work makes several contributions to these approaches. The main contribution of the proposed multi-mask self-supervised learning approach is to use the available undersampled data more efficiently to enable physics-guided DL training, by retrospectively splitting these data into multiple 2-tuple of sets for the DC units during training and for defining loss. Another resulting contribution of the proposed multi-mask SSDU is an alternative data augmentation strategy for DL-MRI reconstruction, via the retrospective hold-out masking of the acquired measurements multiple times, with potential applications even beyond self-supervised learning [111]. Finally, we applied the proposed multi-mask approach on knee and brain MRI datasets using different undersampling and acceleration rates, showing its improved reconstruction performance compared to single mask SSDU approach. Specifically, the extensive experimental results using different subsampling patterns on retrospectively subsampled 3D knee dataset at R = 8 and R = 12 show that the proposed multi-mask SSDU consistently outperforms SSDU, while performing similarly with the supervised DL-MRI approach. Similarly, on prospectively subsampled brain MRI, multi-mask SSDU at R = 8 enhances the reconstruction quality of SSDU, while achieving lower noise level compared to SSDU at R = 8 and CG-SENSE at the acquisition R = 2.

As mentioned earlier, the proposed multi-mask SSDU approach can be interpreted as an alternative technique for data augmentation in DL-MRI reconstruction. With the proposed multi-mask data augmentation, self-supervised training was rated higher than supervised training in the reader study for knee imaging by a musculoskeletal expert reader in terms of noise and aliasing artifacts. Furthermore, Figure 3 showed example slices, where multi-mask self-supervised learning showed better performance in handling artifacts compared to supervised DL-MRI despite having lower SSIM and NMSE values. While these observations may seem surprising at first, it is consistent with recent studies that show quantitative metrics may not always align with the reconstruction performance [41]. Additionally, there are recent studies that show self-supervised deep learning approaches outperforming its supervised counterparts in various applications [112, 113]. These and other studies suggest that supervised learning may preclude

discovery, hence it may not generalize well on unseen data or may not be as robust as self-supervised learning techniques [114]. Another interesting finding from the reader study on knee data was the worse scores given to the fully-sampled ground truth compared to DL-MRI methods. The expert reader noted the low SNR of the fully-sampled acquisition, due to the high acquisition resolution compared to conventional clinical scans, which was substantially improved visually using the inherent noise reduction of DL-MRI reconstruction.

While the proposed multi-mask SSDU approach enhances the SSDU performance, it also has a longer training time by a factor of K compared to SSDU due to the increased size of the training dataset. Due to these lengthy training times, holdout cross-validation was used for the hyperparameter selection sub-study for optimizing K instead of n-fold cross-validation. Furthermore, while the proposed multi-mask approach enables data augmentation, helping overcome data scarcity and enhance reconstruction quality, it also bears the risk of overfitting. In a broader context, it is understood that data augmentation can lead to massive datasets, but when this idea is applied to augment initially limited datasets, it may result in overfitting [115]. This phenomenon was also observed in our study as the reconstruction quality does not monotonically improve with increasing K, and residual artifacts reappear for $K \geq 8$. The problem of choosing the optimal size of the post-augmented dataset, which corresponds to K in our setup, remains an open problem in the broader machine learning community and this challenge has been highlighted in a recent survey on data augmentation as "There is no consensus as to which ratio of original to final dataset size will result in the best performing model" [115]. Hence, while we have optimized K, this was done using the same experimental settings of [10]. We note that the optimal value for K may differ based on the selection of hyperparameters, such as the number of epochs or the learning rate. Nonetheless, our results readily show that multi-mask SSDU improves upon the single-mask SSDU in terms of quantitative metrics for any choice of $K¿1$, while also suggesting that it is not advantageous to choose a very high value of K, both from a performance perspective, and from a practical viewpoint due to the increased training time.

A uniformly random selection of masks was used in the multi-mask SSDU. This was motivated by the issue that splitting Ω based on a Gaussian random selection

leads to selecting mostly low-frequency components from scarce data, especially at high acceleration rates. With a Gaussian selection of Λ, a multi-mask approach still tends to select low-frequency components for each mask. Supporting Information **Figure A.15** showed that using uniformly random selection in combination with multi-mask selection may circumvent this issue, since this will ensure both low and high frequency are contained in the loss masks of each scan.

The multi-masking approach proposed in this work may also be adapted to the supervised learning setting by using multiple random Θ_j in the DC units of the unrolled network, while calculating the loss on the fully-sampled k-space. This extension to supervised training, which introduces an additional degree of randomness to the training process, was recently shown to improve performance over conventional supervised DL-MRI approach [111]. However, we note that since this is a new extension arising from this work, comparisons in this study were made to the conventional DL-MRI approach that is used in the literature [9].

The proposed multi-mask SSDU approach also shares similarities with bootstrap aggregation mechanisms. In bootstrap approaches, multiple sub-datasets are generated by randomly sampling from a main dataset. The final prediction is performed by averaging outputs from each of these sub-dataset to reduce the variance among trained models. However, in multi-mask SSDU, each sample in the main dataset is sub-sampled multiple times by retrospectively splitting its measurements into disjoint sets. As a result, an aggregated large dataset which contains the measurements of each sample in the main dataset multiple times is obtained and used for training. Unlike bootstrapping approaches, the proposed approach performs final prediction by directly using the model trained on the aggregated large dataset.

The study has limitations. In the proposed multi-mask SSDU, we optimized the hyperparameters ρ and K independently for two main reasons: 1) If the joint optimization led to a different ρ value, then this would create a confounding variable for the direct comparison to the single-mask scenario, 2) Optimizing over the two hyperparameters jointly leads to a large number of trainings for marginal gain. Such large number of trainings, which does not show a substantial implication in terms of the perspective of the study may also come at an environmental cost, as training of DL models have been shown to lead to considerable amount of carbon emissions [116]. Thus, in the study, we

concentrated on the individual optimization of the K parameter for the fixed ρ value that works best for single-mask SSDU [10], as this provides a more fair comparison. Additionally, this study focused on methodological development to improve the performance on single-mask SSDU without a specific application that may leverage large stores of existing undersampled data. With the methodology in place, such applications are being pursued in subsequent studies, both in LGE cardiac and brain MRI applications [36, 117]. Moreover, the proposed multi-mask SSDU may also be synergistically combined with dynamic cardiac MRI applications, in which spatio-temporal information can be leveraged to boost the performance at high acceleration rates [107,118,119]. More recently zero-shot learning approaches based on generative networks such as deep image prior and self-supervision have gained interest as they enable training from a single slice [43, 120, 121]. While multi-mask SSDU is a database learning approach, the concept of multi-mask SSDU approach may also be beneficial for developing improved zero-shot learning approaches [43].

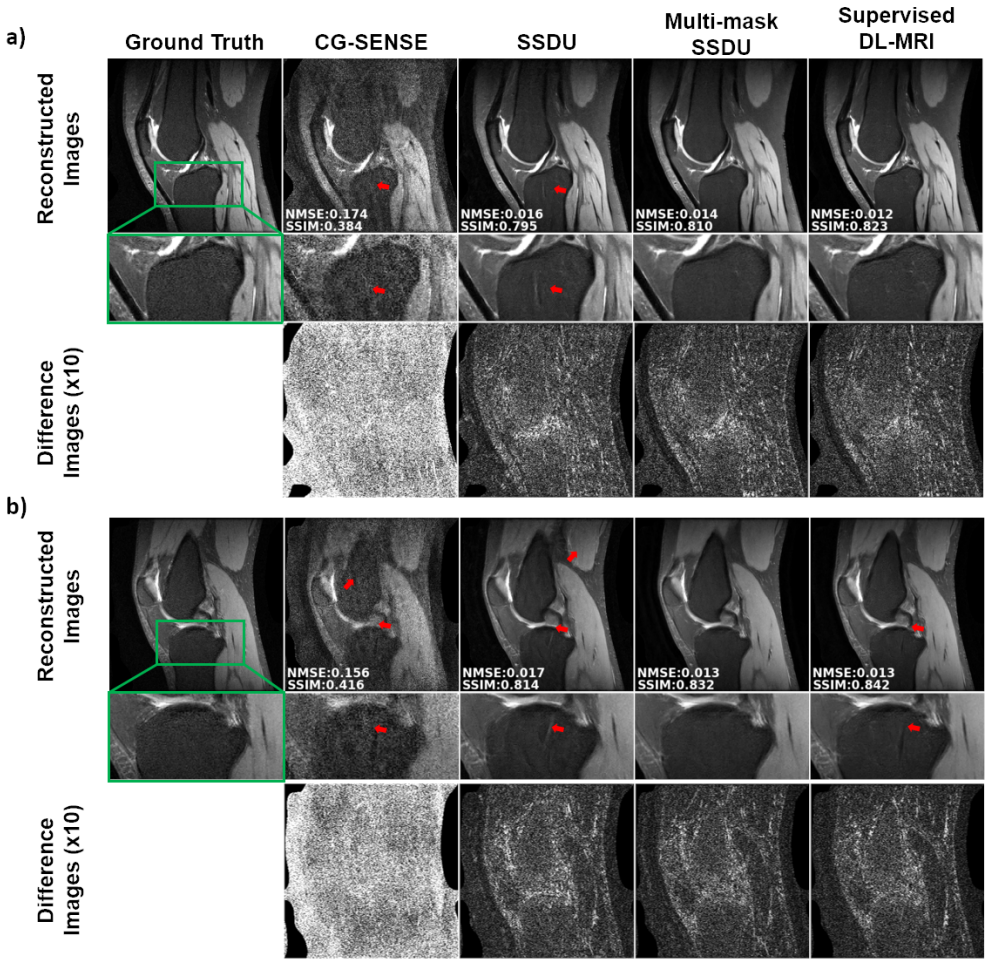

Figure 4.3: a) and b) Representative test slices from 3D FSE knee MRI dataset showing the reconstruction results for proposed multi-mask self-supervised DL-MRI (multi-mask SSDU), self-supervised DL-MRI (SSDU), supervised DL-MRI and CG-SENSE approaches for retrospective equispaced undersampling R = 8, as well as the error images with respect to the fully-sampled reference. CG-SENSE suffers from substantial residual artifacts that are shown with red arrows for both slices. DL-MRI with SSDU learning suppresses a large portion of these artifacts, but still exhibits visible residual artifacts in both scenarios. Proposed multi-mask SSDU successfully suppresses these artifacts further for both slices, in a) closely matches the performance of supervised DL-MRI and in b) reduces residual aliasing further compared to supervised DL-MRI.

		CG-SENSE	Supervised DL-MRI	SSDU	Multi-mask SSDU
Uniform 1D R=8	NMSE	0.0994 [0.0949, 0.1419]	0.0122 [0.0108, 0.0139]	0.0166 [0.0147, 0.0191]	0.0140 [0.0124, 0.0161]
	SSIM	0.4698 [0.4193, 0.5353]	0.8505 [0.8306, 0.8751]	0.8211 [0.7947, 0.8527]	0.8315 [0.8084, 0.8600]
Uniform 2D R=8	NMSE	0.1475 [0.1291, 0.1779]	0.0124 [0.0112, 0.0143]	0.0164 [0.0148, 0.0189]	0.0135 [0.0123, 0.0155]
	SSIM	0.4411 [0.3797, 0.4976]	0.8421 [0.8201, 0.8662]	0.8150 [0.7877, 0.8426]	0.8298 [0.8067, 0.8560]
Random 1D R=8	NMSE	0.0994 [0.0805, 0.1236]	0.0121 [0.0107, 0.0139]	0.0156 [0.0135, 0.0177]	0.0137 [0.0121, 0.0155]
	SSIM	0.4886 [0.4305, 0.5571]	0.8524 [0.8314, 0.8756]	0.8328 [0.8089, 0.8615]	0.8367 [0.8144, 0.8633]
Random 2D R=8	NMSE	0.1473 [0.1301, 0.1759]	0.0130 [0.0117, 0.0149]	0.0173 [0.0155, 0.0199]	0.0145 [0.0131, 0.0165]
	SSIM	0.4239 [0.3631, 0.4766]	0.8379 [0.8164, 0.8637]	0.8123 [0.7853, 0.8417]	0.8224 [0.8002, 0.8509]
Poisson R=8	NMSE	0.1035 [0.0937, 0.1206]	0.0101 [0.0091, 0.0112]	0.0131 [0.0118, 0.0149]	0.0108 [0.0098, 0.0121]
	SSIM	0.4885 [0.4394, 0.5397]	0.8554 [0.8365, 0.8793]	0.8312 [0.8066, 0.8585]	0.8421 [0.8212, 0.8679]
Random 2D R=12	NMSE	0.1331 [0.1207, 0.1556]	0.0157 [0.0141, 0.0179]	0.0221 [0.0198, 0.0254]	0.0185 [0.0167, 0.0208]
	SSIM	0.4325 [0.3756, 0.4796]	0.8148 [0.7916, 0.8431]	0.7809 [0.7517, 0.8151]	0.7982 [0.7722, 0.8288]
Poisson R=12	NMSE	0.0876 [0.0795, 0.1012]	0.0119 [0.0107, 0.0133]	0.0151 [0.0136, 0.0169]	0.0129 [0.0117, 0.0145]
	SSIM	0.5119 [0.4638, 0.5597]	0.8362 [0.8156, 0.8625]	0.8133 [0.7862, 0.8442]	0.8326 [0.8098, 0.8609]

Table 4.1: The median and interquartile ranges for NMSE and SSIM metrics for different undersampling masks and acceleration rates. Note that due to the different size of the ACS data, 1D masks correspond to an effective acceleration rate of 5.2, while the 2D masks yield an effective acceleration rate of 7.7.

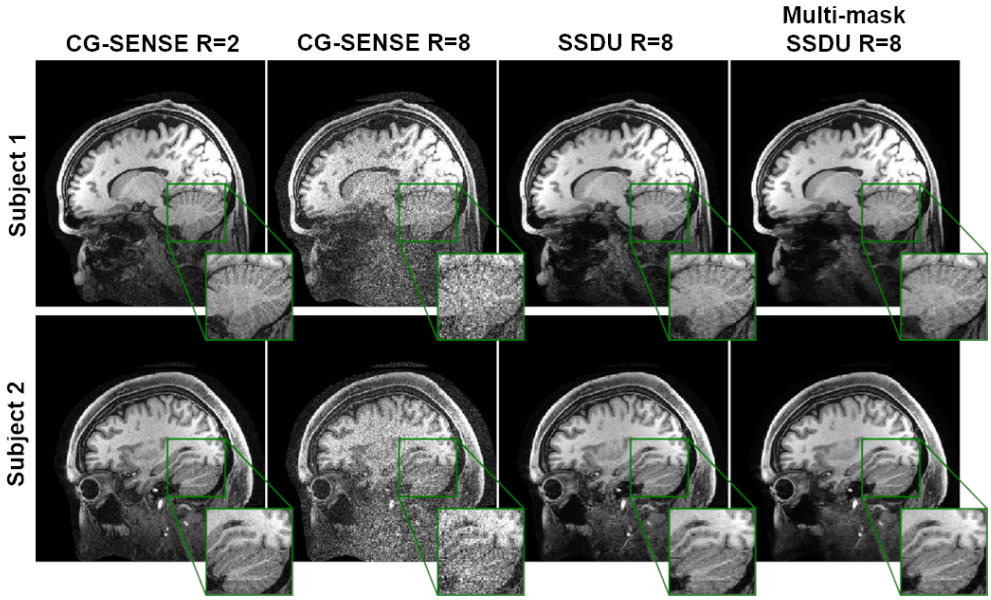

Figure 4.4: Reconstruction results from prospectively 2-fold equispaced undersampled brain MRI. SSDU, multi-mask SSDU and CG-SENSE are applied at further retrospective acceleration rates of 8 with equispaced sheared $k_y - k_z$ undersampling patterns, while CG-SENSE is also used at the acquisition rate of 2, which serves as the clinical baseline. CG-SENSE suffers from visibly higher noise amplification at R = 8. SSDU DL-MRI performs successful reconstruction at R = 8, while achieving similar image quality to CG-SENSE at R = 2. The proposed multi-mask SSDU DL-MRI further enhances the SSDU DL-MRI performance by achieving lower noise level in reconstruction results.

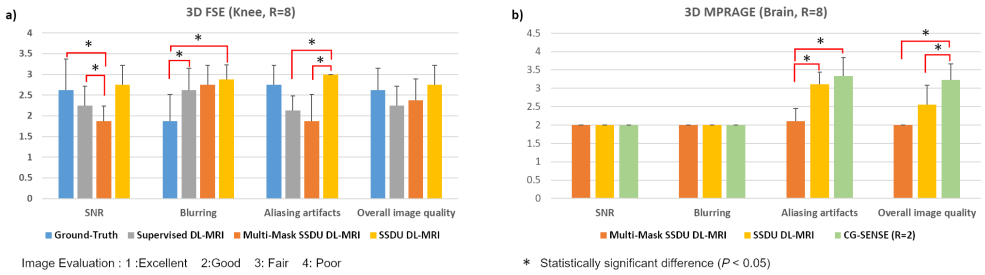

Figure 4.5: a) Reader study for knee MRI. Bar-plots show average reader scores and their standard deviation across the test subjects. Statistical testing was performed by one-sided Wilcoxon single-rank test, with * showing significant statistical difference with P ¡0.05. In terms of SNR, the proposed multi-mask SSDU was rated highest, and statistically better than all approaches except supervised DL-MRI. For blurring, ground truth data was rated statistically better than all methods except the proposed multi-mask SSDU. In terms of aliasing artifacts and overall image quality, the proposed multi-mask SSDU approach was rated best compared to other methods and ground truth. In terms of these two evaluation criteria, all DL-MRI approaches and the reference showed similar statistical behavior, except SSDU was statistically worse than proposed multi-mask SSDU and supervised approach in terms of aliasing artifacts. b) Reader study for brain MRI. CG-SENSE at R = 2, and proposed multi-mask SSDU and SSDU at R = 8 were in good agreement in terms of SNR and blurring. In terms of aliasing artifacts and overall image quality, the proposed multi-mask SSDU approach received the best scores, while CG-SENSE at R = 2 was rated lowest and showed significant statistical difference with proposed multi-mask SSDU in terms of both evaluation criteria and SSDU in terms of overall image quality. The proposed multi-mask SSDU was also rated statistically better than SSDU in terms of aliasing artifacts.

Chapter 5

Zero-Shot Self-Supervised Learning for MRI Reconstruction

5.1 Introduction

Magnetic resonance imaging (MRI) is a non-invasive, radiation-free medical imaging modality that provides excellent soft tissue contrast for diagnostic purposes. However, lengthy acquisition times in MRI remain a limitation. Accelerated MRI techniques acquire fewer measurements at a sub-Nyquist rate, and use redundancies in the acquisition system or the images to remove the resulting aliasing artifacts during reconstruction. In clinical MRI systems, multi-coil receivers are used during data acquisition. Parallel imaging (PI) is the most clinically used method for accelerated MRI, and exploits the redundancies between these coils for reconstruction [1, 2]. Compressed sensing (CS) is another conventional accelerated MRI technique that exploits the compressibility of images in sparsifying transform domains [4], and is commonly used in combination with PI. However, PI and CS may suffer from noise and residual artifacts at high acceleration rates [21, 57].

Recently, deep learning (DL) methods have emerged as an alternative for accelerated MRI due to their improved reconstruction quality compared to conventional approaches [7, 73]. Particularly, physics-guided deep learning reconstruction (PG-DLR) approaches

This chapter is based on [43, 44].

have gained interest due to their robustness and improved performance [7,9,17,29]. PG-DLR explicitly incorporates the physics of the data acquisition system into the neural network via a procedure known as algorithm unrolling [22]. This is done by unrolling iterative optimization algorithms that alternate between data consistency (DC) and regularization steps for a fixed number of iterations. Subsequently, PG-DLR methods are trained in a supervised manner using large databases of fully-sampled measurements [7, 9]. More recently, self-supervised learning has shown that reconstruction quality similar to supervised PG-DLR can be achieved while training on a database of only undersampled measurements [10].

While such database learning strategies offer improved reconstruction quality, acquisition of large datasets may often be infeasible. In some MRI applications involving time-varying physiological processes, dynamic information such as time courses of signal changes, contrast uptake or breathing patterns may differ substantially between subjects, making it difficult to generate high-quality databases of sufficient size for the aforementioned strategies. Furthermore, database training, in general, brings along concerns about robustness and generalization [41, 122]. In MRI reconstruction, this may exhibit itself when there are mismatches between training and test datasets in terms of image contrast, sampling pattern, SNR, vendor, and anatomy. While it is imperative to have high-quality reconstructions that can be used to correctly identify lesions/disease for *every individual*, the fastMRI transfer track challenge shows that pretrained models fail to generalize when applied to patients/scans with different distribution or acquisition parameters, with potential for misdiagnosis [42]. Finally, training datasets may lack examples of rare and/or subtle pathologies, increasing the risk of generalization failure [40, 41].

In this Chapter, we tackle these challenges associated with database training, and propose a zero-shot self-supervised learning (ZS-SSL) approach, which performs subject-specific training of PG-DLR without any external training database. Succinctly, ZS-SSL partitions the acquired measurements into three types of disjoint sets, which are respectively used only in the PG-DLR neural network, in defining the training loss, and in establishing a stopping strategy to avoid overfitting. Thus, our training is both self-supervised and self-validated. In cases where a database-pretrained network is available, ZS-SSL leverages transfer learning (TL) for improved reconstruction quality and reduced

computational complexity.

Our contributions can be summarized as follows:

- We propose a zero-shot self-supervised method for learning subject-specific DL MRI reconstruction from a single undersampled dataset without any external training database.
- We provide a well-defined methodology for determining stopping criterion to avoid over-fitting in contrast to other single-image training approaches [47].
- We apply the proposed zero-shot learning approach to knee and brain MRI datasets, and show its efficacy in removing residual aliasing and banding artifacts compared to supervised database learning.
- We show our ZS-SSL can be combined with with TL in cases when a database-pretrained model is available to reduce computational costs.
- We show that our zero-shot learning strategies address robustness and generalizability issues of trained supervised models in terms of changes in sampling pattern, acceleration rate, contrast, SNR, and anatomy at inference time.

5.2 Background and Related Work

5.2.1 Accelerated MRI Acquisition Model

In MRI, raw measurement data is acquired in the frequency domain, also known as k-space. In current clinical MRI systems, multiple receiver coils are used, where each is sensitive to different parts of the volume. In practice, MRI is accelerated by taking fewer measurements, which are characterized by an undersampling mask that specifies the acquired locations in k-space. For a multi-coil MRI acquisition, the forward model is given as

$$\mathbf{y}_i = \mathbf{P}_\Omega \mathcal{F} \mathbf{C}_i \mathbf{x} + \mathbf{n}_i, \;\; i \in \{1, \ldots, n_c\}, \tag{5.1}$$

where $\mathbf{x}$ is the underlying image, $\mathbf{y}_i$ is the acquired data for the i^{th} coil, $\mathbf{P}_\Omega$ is the masking operator for undersampling pattern Ω, $\mathcal{F}$ is the Fourier transform, $\mathbf{C}_i$ is a diagonal matrix characterizing the i^{th} coil sensitivity, $\mathbf{n}_i$ is measurement noise for i^{th}

coil, and n_c is the number of coils [1]. This system can be concatenated across the coil dimension for a compact representation

$$\mathbf{y}_\Omega = \mathbf{E}_\Omega \mathbf{x} + \mathbf{n}, \tag{5.2}$$

where $\mathbf{y}_\Omega$ is the acquired undersampled measurements across all coils, $\mathbf{E}_\Omega$ is the forward encoding operator that concatenates $\mathbf{P}_\Omega \mathcal{F} \mathbf{C}_i$ across $i \in \{1, \ldots, n_c\}$. The general inverse problem for accelerated MRI is given as

$$\arg\min_{\mathbf{x}} \|\mathbf{y}_\Omega - \mathbf{E}_\Omega \mathbf{x}\|_2^2 + \mathcal{R}(\mathbf{x}), \tag{5.3}$$

where the $\|\mathbf{y}_\Omega - \mathbf{E}_\Omega \mathbf{x}\|_2^2$ term enforces consistency with acquired data (DC) and $\mathcal{R}(\cdot)$ is a regularizer.

5.2.2 PG-DLR with Algorithm Unrolling

Several optimization methods are available for solving the inverse problem in (6.8) [70]. Variable-splitting via quadratic penalty is one such approach that decouples the DC and regularizer units. It introduces an auxiliary variable $\mathbf{z}$ that is constrained to be equal to $\mathbf{x}$, and (6.8) is reformulated as an unconstrained problem with a quadratic penalty

$$\arg\min_{\mathbf{x},\mathbf{z}} \|\mathbf{y}_\Omega - \mathbf{E}_\Omega \mathbf{x}\|_2^2 + \mu \|\mathbf{x} - \mathbf{z}\|_2^2 + \mathcal{R}(\mathbf{z}), \tag{5.4}$$

where μ is the penalty parameter. The optimization problem in (6.10) is then solved via alternating minimization as

$$\mathbf{z}^{(i)} = \arg\min_{\mathbf{z}} \mu \|\mathbf{x}^{(i-1)} - \mathbf{z}\|_2^2 + \mathcal{R}(\mathbf{z}), \tag{5.5a}$$

$$\mathbf{x}^{(i)} = \arg\min_{\mathbf{x}} \|\mathbf{y}_\Omega - \mathbf{E}_\Omega \mathbf{x}\|_2^2 + \mu \|\mathbf{x} - \mathbf{z}^{(i)}\|_2^2, \tag{5.5b}$$

where $\mathbf{z}^{(i)}$ is an intermediate variable and $\mathbf{x}^{(i)}$ is the desired image at iteration i. In PG-DLR, an iterative algorithm, as in (6.12a) and (6.12b) is unrolled for a fixed number of iterations [18]. The regularizer sub-problem in Eq. (6.12a) is implicitly solved with neural networks and the DC sub-problem in Eq. (6.12b) is solved via linear methods such as gradient descent [7] or conjugate gradient (CG) [9].

There have been numerous works on PG-DLR for accelerated MRI [7, 9, 10, 18, 29]. Most of these works vary from each other on the algorithms used for DC and neural networks employed in the regularizer units. However, all these works require a large database of training samples.

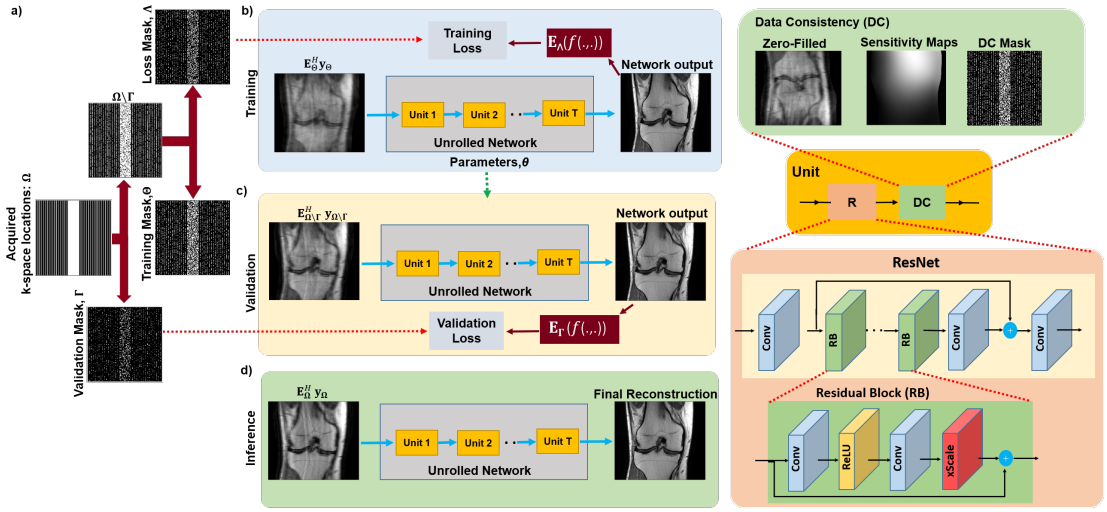

Figure 5.1: An overview of the proposed zero-shot self-supervised learning approach. a) Acquired measurements for the single scan are partitioned into three sets: a training (Θ) and loss mask (Λ) for self-supervision, and a self-validation mask for automated early stopping (Γ). b) The parameters, $\boldsymbol{\theta}$, of the unrolled MRI reconstruction network are updated using Θ and Λ in the data consistency (DC) units of the unrolled network and for defining loss, respectively. c) Concurrently, a k-space validation procedure is used to establish the stopping criterion by using $\Omega\backslash\Gamma$ in the DC units and Γ to measure a validation loss. d) Once the network training has been stopped due to an increasing trend in the k-space validation loss, the final reconstruction is performed using the relevant learned network parameters and all the acquired measurements in the DC unit.

5.2.3 Supervised Learning for PG-DLR

In supervised PG-DLR, training is performed using a database of fully-sampled reference data. Let $\mathbf{y}_{\text{ref}}^n$ be the fully-sampled k-space for subject n and $f(\mathbf{y}_\Omega^n, \mathbf{E}_\Omega^n; \boldsymbol{\theta})$ be the output of the unrolled network for under-sampled k-space $\mathbf{y}_\Omega^n$, where the network is parameterized by $\boldsymbol{\theta}$. End-to-end training minimizes [10, 73]

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{n=1}^{N} \mathcal{L}(\mathbf{y}_{\text{ref}}^n,\ \mathbf{E}_{\text{full}}^n f(\mathbf{y}_\Omega^n, \mathbf{E}_\Omega^n; \boldsymbol{\theta})), \tag{5.6}$$

where N is the number of samples in the training database, $\mathbf{E}_{\text{full}}^n$ is the fully-sampled encoding operator that transform network output to k-space and $\mathcal{L}(\cdot, \cdot)$ is a loss function.

5.2.4 Self-Supervised Learning for PG-DLR

Unlike supervised learning, self-supervised learning enables training without fully-sampled data by only utilizing acquired undersampled measurements [10]. A masking approach is used for self-supervision in this setting, where a subset $\Lambda \subset \Omega$ is set aside for checking prediction performance/loss calculation, while the remainder of points $\Theta = \Omega \backslash \Lambda$ are used in the DC units of the PG-DLR network. End-to-end training is performed using the loss function

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{n=1}^{N} \mathcal{L}\Big(\mathbf{y}_{\Lambda}^{n},\ \mathbf{E}_{\Lambda}^{n}\big(f(\mathbf{y}_{\Theta}^{n}, \mathbf{E}_{\Theta}^{n}; \boldsymbol{\theta})\big)\Big). \tag{5.7}$$

5.3 Zero-Shot Self-Supervised Learning for PG-DLR

As discussed in Section 5.1, lack of large datasets in numerous MRI applications, as well as robustness and generalizability issues of pretrained models pose a challenge for the clinical translation of DL reconstruction methods. Hence, subject-specific reconstruction is desirable in clinical practice, since it is critical to achieve a reconstruction quality that can be used for correctly diagnosing every patient. While the conventional self-supervised masking strategy, as in [10] can be applied for subject-specific learning, it leads to overfitting unless the training is stopped early [123]. This is similar to other single-image learning strategies, such as the deep image prior (DIP) or zero-shot super-resolution [47, 124]. DIP-type approaches shows that an untrained neural network can successfully perform instance-specific image restoration tasks such as denoising, super-resolution, inpainting without any training data. However, such DIP-type techniques requires an early stopping for avoiding over-fitting, which is typically done with a manual heuristic selection [47, 123, 125]. While this may work in a research setting, having a well-defined automated early stopping criterion is critical to fully harness the potential of subject-specific DL MRI reconstruction in practice.

Early stopping regularization in database-trained setting is conventionally motivated through the bias-variance trade-off, in which a validation set is used as a proxy for the generalization error to identify the stopping criterion. Using the same bias-variance trade-off motivation, having a validation set can aid in devising a stopping criterion, but this has not been feasible in existing zero-shot learning approaches, which either

use all acquired measurements [47, 80] or partition them into two sets for training and defining loss [123]. Hence, existing zero-shot learning techniques lack a validation set to identify the stopping criterion.

ZS-SSL Formulation and Training: We propose a new ZS-SSL partitioning framework to enable subject-specific self-supervised training and validation with a well-defined stopping criterion. We define the following partition for *the available measurement locations from a single scan*, Ω:

$$\Omega = \Theta \sqcup \Lambda \sqcup \Gamma, \tag{5.8}$$

where $\sqcup$ denotes a disjoint union, i.e. Θ, Λ and Γ are pairwise disjoint (**Figure 5.1**). Similar to Section 5.2.4, Θ is used in the DC units of the unrolled network, and Λ is used to define a k-space loss for the self-supervision of the network. The third partition Γ is a set of acquired k-space indices set aside for defining a k-space validation loss. Thus, ZS-SSL training is both self-supervised and self-validated.

In general, since zero-shot learning approaches perform training using a single dataset, generation of multiple data pairs from this single dataset is necessary to self-supervise the neural network [126]. Hence, we generate multiple (Θ, Λ) pairs from the acquired locations Ω of the single scan. In ZS-SSL, this is achieved by fixing the k-space validation partition $\Gamma \subset \Omega$, and performing the retrospective masking on $\Omega \backslash \Gamma$ multiple times. Formally, $\Omega \backslash \Gamma$ is partitioned K times such that

$$\Omega \backslash \Gamma = \Theta_k \sqcup \Lambda_k, \quad k \in \{1, \ldots, K\}, \tag{5.9}$$

where Λ_k, Θ_k and Γ are pairwise disjoint, i.e. $\Omega = \Gamma \sqcup \Theta_k \sqcup \Lambda_k, \forall\ k$. ZS-SSL training minimizes

$$\min_{\boldsymbol{\theta}} \frac{1}{K} \sum_{k=1}^{K} \mathcal{L}\Big(\mathbf{y}_{\Lambda_k},\ \mathbf{E}_{\Lambda_k}\big(f(\mathbf{y}_{\Theta_k}, \mathbf{E}_{\Theta_k}; \boldsymbol{\theta})\big)\Big)$$

In the proposed ZS-SSL, this is supplemented by a k-space self-validation loss, which tests the generalization performance of the trained network on the k-space validation partition Γ. For the l^{th} epoch, where the learned network weights are specified by $\boldsymbol{\theta}^{(l)}$, this validation loss is given by:

$$\mathcal{L}\Big(\mathbf{y}_{\Gamma},\ \mathbf{E}_{\Gamma}\big(f(\mathbf{y}_{\Omega\backslash\Gamma}, \mathbf{E}_{\Omega\backslash\Gamma}; \boldsymbol{\theta}^{(l)})\big)\Big). \tag{5.10}$$

Note that in (5.10), the network output is calculated by applying the DC units on $\Omega \backslash \Gamma = \Theta \sqcup \Lambda$, i.e. all acquired points outside of Γ, to better assess its generalizability performance. Our key motivation is that while the training loss will decrease over epochs, the k-space validation loss will start increasing once overfitting is observed. Thus, we monitor the loss in (5.10) during training to define an early stopping criterion to avoid overfitting. Let L be the epoch in which training needs to be stopped. Then at inference time, the network output is calculated as $f(\mathbf{y}_\Omega, \mathbf{E}_\Omega; \boldsymbol{\theta}^{(L)})$, i.e. all acquired points are used to calculate the network output.

ZS-SSL with Transfer Learning (TL): While pretrained models are very efficient in reconstructing new unseen measurements from similar MRI scan protocols, their performance degrades significantly when acquisition parameters vary [42]. Moreover, retraining a new model on a large database for each acquisition parameter(e.g. sampling, contrast, anatomy, acceleration), may be very computationally expensive [40]. Hence, TL has been used for re-training DL models pre-trained on large databases to reconstruct MRI data with different characteristics [40]. However, such transfer still requires another, often smaller, database for re-training. In contrast, in the presence of pre-trained models, ZS-SSL can be combined with TL, referred to as ZS-SSL-TL, to reconstruct a *single* slice/instance with different characteristics by using weights of the pre-trained model for initialization. Thus, ZS-SSL-TL ensures that the pretrained model is adapted for each patient/subject, while facilitating faster convergence time and reduced reconstruction time.

5.4 Experiments

5.4.1 Datasets

We performed experiments on publicly available fully-sampled multi-coil knee and brain MRI from fastMRI database [77]. Knee and brain MRI datasets contained data from 15 and 16 receiver coils, respectively. Fully-sampled datasets were retrospectively undersampled by keeping 24 lines of autocalibrated signal (ACS) from center of k-space. FastMRI database contains different contrast weightings. For knee MRI, we used coronal proton density (Cor-PD) and coronal proton density with fat suppression (Cor-PDFS),

and for brain MRI, axial FLAIR (Ax-FLAIR) and axial T2 (Ax-T2). Different types of datasets and undersampling masks used in this study are provided in **Figure A.21** in the Appendix.

5.4.2 Implementation Details

All PG-DLR approaches were trained end-to-end using 10 unrolled iterations. CG method and a ResNet structure [76] were employed in the DC and regularizer units of the unrolled network, respectively [10]. The ResNet is comprised of a layer of input and output convolution layers, and 15 residual blocks (RB) each containing two convolutional layers, where the first layer is followed by ReLU and the second layer is followed by a constant multiplication [76]. All layers had a kernel size of 3×3, 64 channels. The real and imaginary parts of the complex MR images were concatenated prior to being input to the ResNet as 2-channel images. The unrolled network, which shares parameters across the unrolled iterations had a total of 592,129 trainable parameters. Coil sensitivity maps were generated from the central 24×24 ACS using ESPIRiT [106]. End-to-end training was performed with a normalized ℓ_1-ℓ_2 loss (Adam optimizer, LR = $5 \cdot 10^{-4}$, batch size = 1) [10]. Peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) were used for quantitative evaluation.

5.4.3 Reconstruction Method Comparisons

In this work, we focus on comparing training strategies for accelerated MRI reconstruction. Thus, we use the same network architecture from Section 5.4.2 for all training methods in all experiments. We note that the proposed ZS-SSL strategy is agnostic to the specifics of the neural network architecture. In fact, the number of network parameters is higher than the number of undersampled measurements available on a single slice, i.e. dimension of $\mathbf{y}_\Omega$. As such, different neural networks may be used for the regularizer unit in the unrolled network, but this is not the focus of our study.

Supervised PG-DLR: Supervised PG-DLR models for knee and brain MRI were trained on 300 slices from 15 and 30 different subjects, respectively. For each knee and brain contrast weighting, two networks were trained separately using random and uniform masks [7] at an acceleration rate (R) of 4 [41]. Trained networks were used for

comparison and TL purposes. We note that random undersampling results in incoherent artifacts, whereas uniform undersampling leads to coherent artifacts that are harder to remove (**Figure A.21** in Appendix) [40]. Hence, we focus on the more difficult problem of uniform undersampling, while presenting random undersampling results in the Appendix.

Self-Supervision via Data Undersampling (SSDU) PG-DLR: SSDU [10] PG-DLR was trained using the same database approach as supervised PG-DLR, with the exception that SSDU performed training only using the undersampled data (Sec. 5.2.4).

DIP-Recon: We employ a DIP-type subject-specific MRI reconstruction that uses all acquired measurements in both DC and defining loss [80, 127]

$$\mathcal{L}\Big(\mathbf{y}_\Omega,\, \mathbf{E}_\Omega\big(f(\mathbf{y}_\Omega, \mathbf{E}_\Omega; \boldsymbol{\theta})\big)\Big). \tag{5.11}$$

We refer to the reconstruction from this training mechanism as DIP-Recon. DIP-Recon-TL refers to combining (5.11) with TL. As mentioned, DIP-Recon does not have a stopping criterion, hence early stopping was heuristically determined (**Figure A.22** in the Appendix).

Parallel Imaging: We include CG-SENSE, which is a commonly used subject-specific conventional PI method [1, 78], as the clinical baseline quality for comparison purposes.

5.4.4 Automated Stopping and Ablation Study

The stopping criterion for the proposed ZS-SSL was investigated on slices from the knee dataset. The k-space self-validation set Γ was selected from the acquired measurements Ω using a uniformly random selection with $|\Gamma|/|\Omega| = 0.2$. The remaining acquired measurements $\Omega\backslash\Gamma$ were retrospectively partitioned into disjoint 2-tuples multiple times based on uniformly random selection with the ratio $\rho = |\Lambda_k|/|\Omega\backslash\Gamma| = 0.4$ $\forall k \in \{1, \ldots, K\}$ [10].

Figure 5.2a shows representative subject-specific training and validation loss curves at R = 4 for $K \in \{1, 10, 25, 50, 100\}$. As expected, training loss decreases with increasing epochs for all K. The k-space validation loss for $K = 1$ decreases without showing a clear breaking point for stopping. For $K > 1$, the validation loss forms an L-curve, and the breaking point of the L-curve is used as the stopping criterion. $K = 10$ is used for

the rest of the study, while noting $K = 25$, 50 and 100 also show similar performance. **Figure 5.2b** shows loss curves on a Cor-PD slice with and without transfer learning. ZS-SSL-TL, which uses pre-trained supervised PG-DLR parameters as initial starting parameters, converges faster in time compared to ZS-SSL, substantially reducing the total training time. Average computation times for single-instance reconstruction methods are presented in Table A.3 in the Appendix. Similarly, corresponding reconstruction results for the loss curves in **Figure 5.2a** and **b** are provided in **Figure A.23** in the Appendix.

5.4.5 Reconstruction Results

In the first set of experiments, we compare all methods for the case when the testing and training data belong to the same knee/brain MRI contrast weighting with the same acceleration rate and undersampling mask. These experiments aim to show the efficacy of the proposed approach in performing subject-specific MRI reconstruction, while removing residual aliasing artifacts. We also note that this is the most favorable setup for database-trained supervised PG-DLR.

In the subsequent experiments, we focus on the reported generalization and robustness issues with database-trained PG-DLR methods [40–42, 128]. We investigate banding artifacts, as well as in-domain and cross-domain transfer cases. For these experiments, we concentrate on ZS-SSL-TL, since ZS-SSL has no prior domain information, and is inherently not susceptible to such generalizability issues.

Comparison of Reconstruction Methods: In these experiments, supervised and SSDU PG-DLR are trained and tested using uniform undersampling at R = 4, representing a perfect match for training and testing conditions. **Figure 5.3a** and **b** show reconstruction results for Cor-PD knee and Ax-FLAIR brain MRI datasets in this setting. CG-SENSE reconstruction suffers from significant residual artifacts and noise amplification in Cor-PD knee and Ax-FLAIR brain MRIs, respectively. Similarly, both

Table 5.1: Average PSNR and SSIM values on 30 test slices.

	Metrics	CG-SENSE	Supervised PG-DLR	SSDU PG-DLR	DIP-Recon	DIP-Recon-TL	ZS-SSL	ZS-SSL-TL
Cor-PD	SSIM	0.862	**0.952**	0.949	0.793	0.819	0.948	0.951
	PSNR	34.521	39.966	39.545	32.668	33.583	39.550	**40.102**
Ax-FLAIR	SSIM	0.836	0.934	0.929	0.799	0.818	0.935	**0.937**
	PSNR	31.969	**37.375**	36.761	30.637	31.249	36.861	37.250

DIP-Recon and DIP-Recon-TL suffer from residual artifacts and noise amplification. Supervised PG-DLR achieves artifact-free reconstruction. Both ZS-SSL and ZS-SSL-TL also perform artifact-free reconstruction with similar image quality. Table 5.1 shows the average SSIM and PSNR values on 30 test slices. Similar observations apply when random undersampling is employed (**Figure A.24** in the Appendix). For the remaining experiments, we investigate the generalizability of database-pretrained models using supervised PG-DLR as baseline due to its higher performance, while noting SSDU PG-DLR, which is a self-supervised database-trained model, may also be used as a baseline if needed.

Banding Artifacts: Banding artifacts appear in the form of streaking horizontal lines, and occur due to high acceleration rates and anisotropic sampling [128]. These hinder radiological evaluation and are regarded as a barrier for the translation of DL reconstruction methods into clinical practice [128]. This set of experiments explored training and testing on Cor-PDFS data, where database-trained PG-DLR reconstruction has been reported to show such artifacts [42, 128]. **Figure 5.4** shows reconstructions for a Cor-PDFS test slice. While DIP-Recon-TL suffers from clearly visible noise amplification, supervised PG-DLR suffers from banding artifacts shown with yellow arrows. ZS-SSL-TL significantly alleviates these banding artifacts in the reconstruction. While supervised PG-DLR achieves slightly better SSIM and PSNR (Table A.4 in the Appendix), we note that the banding artifacts do not necessarily correlate with such metrics, and are usually picked up in expert readings [41, 128].

In-Domain Transfer: In these experiments, we compared the in-domain generalizability of database-trained PG-DLR and subject-specific PG-DLR. For in-domain transfer, training and test datasets are of the same type of data, but may differ from each other in terms of acceleration and undersampling pattern (**Figure A.25** in the Appendix). In **Figure 5.5a**, supervised PG-DLR was trained with random undersampling and tested on uniform undersampling, both at R = 4. Supervised PG-DLR fails to generalize and suffers from residual aliasing artifacts (red arrows), consistent with previous reports [40, 42]. Similarly, DIP-Recon-TL suffers from artifacts and noise amplification. Proposed ZS-SSL-TL achieves an artifact-free and improved reconstruction quality. In **Figure 5.5b**, supervised PG-DLR was trained with uniform undersampling at R = 4

and tested on uniform undersampling at R = 6. While both supervised PG-DLR and DIP-Recon-TL suffers from aliasing artifacts, ZS-SSL-TL successfully removes these artifacts. Average PSNR and SSIM values align with the observations (Table A.4 in the Appendix).

Cross-Domain Transfer: In the last set of experiments, we investigated the cross-domain generalizability of database-trained PG-DLR compared to subject-specific trained PG-DLR. For cross-domain transfer, training and test datasets are of the different data characteristics and generally differ in terms of contrast, SNR, and anatomy (**Figure A.25** in the Appendix). **Figure 5.6** shows results for the case when the testing contrast/SNR and anatomy differs from training contrast/SNR and anatomy, even though the same R = 4 uniform undersampling is used for both training and testing. In **Figure 5.6a**, supervised PG-DLR was trained on Cor-PDFS (low-SNR), but tested on Cor-PD (high-SNR and different contrast). In **Figure 5.6b**, supervised PG-DLR was trained on Ax-FLAIR (brain MRI) and tested on Cor-PD (knee MRI). In both cases, supervised PG-DLR fails to generalize and has residual artifacts (red arrows). Similarly, DIP-Recon-TL suffers from artifacts and noise. ZS-SSL-TL achieves an artifact-free improved reconstruction. For both cross-domain transfer experiments, similar results were observed for brain MRI (**Figure A.26** in the Appendix). Average PSNR and SSIM values match these observations (Table A.4 in the Appendix).

5.5 Conclusions

We proposed a zero-shot self-supervised deep learning method, ZS-SSL, for subject-specific accelerated DL MRI reconstruction from a single undersampled dataset. The proposed ZS-SSL partitions the acquired measurements from a single scan into three types of disjoint sets, which are used only in the PG-DLR network, in defining the training loss, and in establishing a validation strategy for early stopping to avoid overfitting. In particular, we showed that with our training methodology and automated stopping criterion, subject-specific zero-shot learning of PG-DLR for MRI can be achieved even when the number of tunable network parameters is higher than the number of available measurements.

Finally, we also combined ZS-SSL with transfer learning, in cases where a pre-trained model may be available, for faster convergence time and reduced reconstruction time. Our results showed that ZS-SSL methods perform similarly to database-trained supervised PG-DLR when training and testing data are matched, and they significantly outperform database-trained methods in terms of artifact reduction and generalizability when the training and testing data differ in terms of image characteristics and acquisition parameters. In fact, the subject-specific nature of ZS-SSL ensures that it is agnostic to such changes in acquisition parameters. As such, the proposed work is able to provide good reconstruction quality for each subject, and may have significant implications in the integration of DL reconstruction to clinical studies. We note that hyperparameters, such as learning rate may be adjusted based on the domain for further improvements. It is also noteworthy that the subject-specific ZS-SSL eliminates the requirement for large training sets. This may also facilitate the use and appeal of DL reconstruction for recently developed acquisitions, as well as pilot studies that are often performed to determine the acquisition parameters/acceleration rates of large-scale imaging studies, such as the Human Connectome Project (HCP) [34]. Finally, while we concentrated on physics-guided models in MRI reconstruction, our ideas and results may inspire further work in related image restoration problems, as well as for generative models or data-driven problems without a data consistency term.

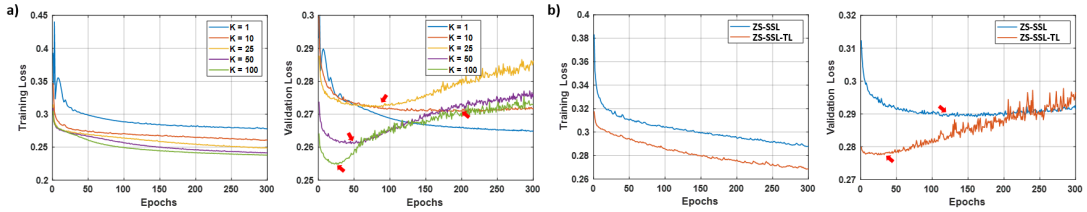

Figure 5.2: a) Representative training and k-space validation loss curves for ZS-SSL with multiple $K \in \{1, 10, 25, 50, 100\}$ masks on Cor-PD knee MRI using uniform undersampling at R = 4. For $K > 1$ the validation loss forms an L-curve, whose breaking point (red arrows) dictates the automated early stopping criterion for training. b) Loss curves for ZS-SSL with/without TL for $K = 10$ on a Cor-PD knee MRI slice. ZS-SSL with TL converges faster compared to ZS-SSL (red arrows).

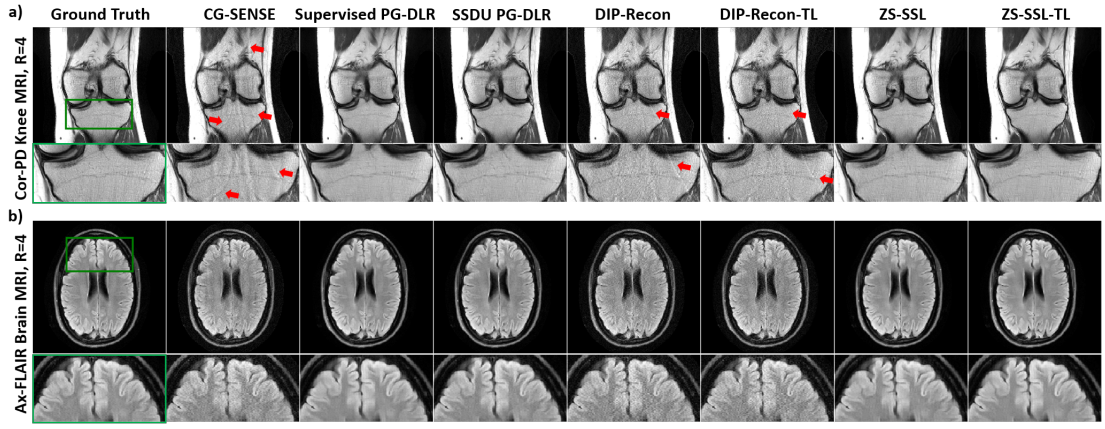

Figure 5.3: Reconstruction results on a representative test slice from a) Cor-PD knee MRI and b) Ax-FLAIR brain MRI at R = 4 with uniform undersampling. CG-SENSE, DIP-Recon, DIP-Recon-TL suffer from noise amplification and residual artifacts shown with red arrows, especially in knee MRI due to the unfavorable coil geometry. Subject-specific ZS-SSL and ZS-SSL-TL achieve artifact-free and improved reconstruction quality, similar to the database-trained SSDU and supervised PG-DLR.

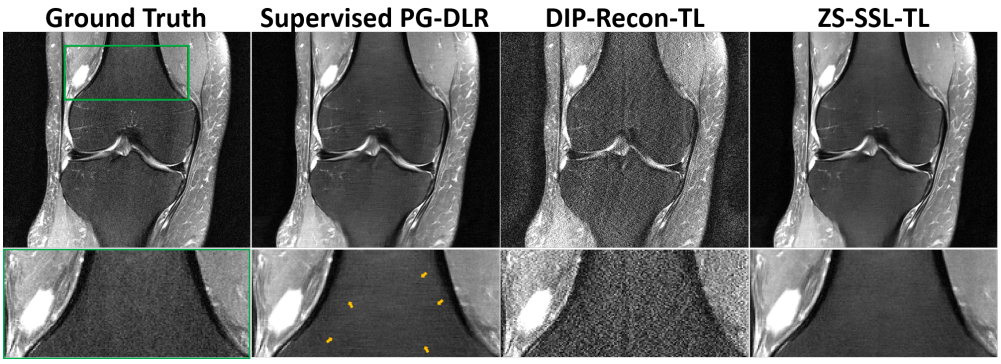

Figure 5.4: Supervised PG-DLR suffers from banding artifacts (yellow arrows), while ZS-SSL-TL significantly alleviates these artifacts. DIP-Recon-TL suffers from clear noise amplification.

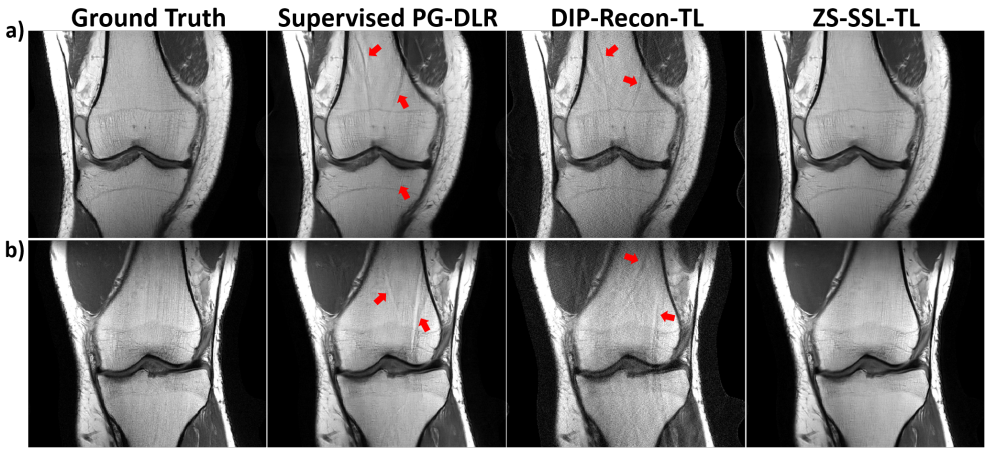

Figure 5.5: Supervised PG-DLR was trained with a) random mask and tested on uniform mask, both R = 4; b) uniform mask at R = 4 and tested on R = 6 uniform mask. Supervised PG-DLR and DIP-Recon-TL suffer from visible artifacts (red arrows). ZS-SSL-TL yields artifact-free reconstruction.

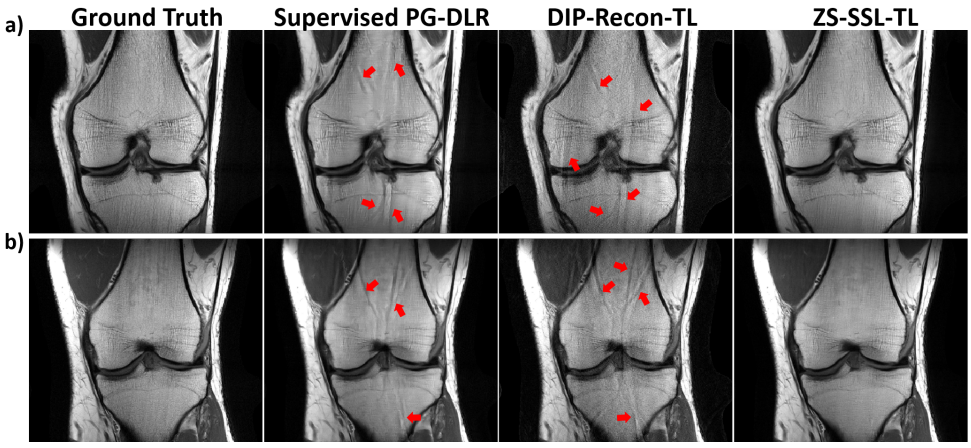

Figure 5.6: Using pre-trained a) Cor-PDFS (low-SNR) and b) Ax-FLAIR (brain MRI) models for Cor-PD. Supervised PG-DLR fails to generalize for both contrast/SNR and anatomy changes, suffering from residual artifacts (red arrows). DIP-Recon-TL also shows artifacts. ZS-SSL-TL successfully removes noise and artifacts for both cases.

Chapter 6

Self-Supervised Denoising with Inpainting Unrolling

6.1 Introduction

Image denoising aims to recover clean images from noisy measurements, since it is not feasible to avoid noise contamination in numerous scenarios due to instrumental imperfection or environmental conditions. The implicit assumption of many denoising approaches is that pixels of the underlying clean images are spatially correlated, while the contaminating noise instances are uncorrelated [129–132]. In recent years, convolutional neural networks (CNNs) have gained immense attention for image denoising [45–51]. In CNN-based denoising, parameters of the convolutional kernels are traditionally tuned to minimize the discrepancy between pairs of noisy and clean target images as measured by a pre-specified loss metric [45, 133]. The network is consequently expected to generalize to denoise future images with similar statistical properties.

The classical supervised training of CNN-based denoiser requires availability of clean target images pertinent to noisy ones, which may not be readily available in some scenarios [48, 50, 53]. A number of recent research studies have attempted to address this issue by training CNNs without requiring ground truth. Noise2Noise [50] was the first method that proposed to perform the training on pairs of noisy images rather

This chapter is based on [54].

than noisy and clean images. The main underlying assumption of Noise2Noise is the availability of two noisy instances of the same image with independent noise, which may be difficult to obtain in practice, such as medical imaging applications.

To tackle this challenge, several self-supervised approaches have been proposed to learn denoising from only noisy images [48, 53, 126, 134]. The main underlying assumptions in these approaches are the statistical independence of noise across pixels, and the existence of spatial correlations across the pixels of the true underlying image. These self-supervised approaches split image pixels into two disjoint sets following a masking operation, in which image pixels in one of these set is used as input to the network while the other is used to define the loss. Among such self-supervised approaches, Noise2Self (N2S) theoretically shows that under a certain masking choice, minimizing the self-supervised loss on only noisy images is equivalent to minimizing the supervised loss function up to a constant under the aforementioned assumptions. While self-supervised approaches learn denoising using only noisy images, all these approaches relies on a purely data-driven regularization neural network without explicitly incorporating the masking model in the network architecture.

In this work, we propose a novel self-supervised image denoising approach, called Noise2Inpaint. Noise2Inpaint utilizes the masking model of N2S, but recasts the denoising problem as an image inpainting inverse problem with a well-defined objective function. Subsequently, an iterative optimization procedure for solving this objective function is unrolled for a fixed number of iterations via a procedure known as algorithm unrolling [22,135], incorporating a CNN-based regularizer and a linear data fidelity (DF) unit in each iteration [7, 28, 136]. This unrolled network has only one more trainable parameter compared to N2S, which arises from the penalty term in the DF. N2I follows N2S and splits noisy pixels into two disjoint sets. One of these sets is used both as network input and in the data fidelity units for ensuring consistency, while the other set is used to define the loss as in N2S.

Our main contributions are summarized as follows:

- We introduce a self-supervised learning approach for referenceless denoising by recasting the denoising problem as an inpainting task with an objective function to be minimized in an optimization framework.

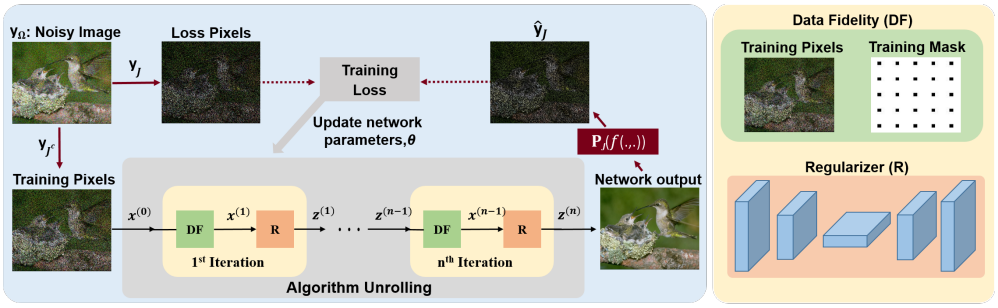

Figure 6.1: Overview of the self-supervised training mechanism of the proposed Noise2Inpaint approach. The noisy pixels of each image are split into training and loss pixels. Training pixels are input to the unrolled network with fixed number of iterations where each iteration consist of regularizer and data fidelity (DF) terms. These training pixels are also used in the DF units to ensure data consistency. The loss is defined between loss pixels that are not used in the training and network output at corresponding loss pixel locations.

- We train an unrolled neural network with data fidelity and CNN-based regularization units to solve the optimization problem pertinent to the inpainting challenge using self-supervision.

- Our unrolled network in N2I shares parameters units across iterations. Thus, when using the same neural network architecture as its purely data-driven counterparts (e.g. N2S, N2V) in the regularization unit, the N2I network *only has one additional trainable parameter*. This single additional parameter arises from the quadratic penalty parameter in the data fidelity unit.

- We apply the proposed Noise2Inpaint approach to real world datasets, and showing its superiority compared to its purely data-driven counterparts.

6.2 Related Work

In this section, we discuss CNN-based denoising algorithms, including approaches that do not require ground truth clean data for training.

Image denoising focuses on recovering a clean target image, $\mathbf{x} \in \mathbb{R}^m$ from a noisy

image, $\mathbf{y} \in \mathbb{R}^m$. Typically, an additive noise model is assumed [130,131] with

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \tag{6.1}$$

where $\mathbf{n}$ denotes the noise which is generally modeled as Gaussian and is independent from $\mathbf{x}$ [129, 130]. More complicated statistical models may also be encountered in practical applications [137–140].

6.2.1 Noise2True Training

A common setup for deep learning methods in image denoising is the supervised setting, which is also referred to as Noise2True (N2T), and requires clean ground-truth images for training. Specifically, N2T training learns a denoising CNN $f : \mathbb{R}^m \to \mathbb{R}^m$ that is parameterized by $\boldsymbol{\theta}$ and which minimizes the supervised loss

$$\mathcal{L}(\boldsymbol{\theta}) \triangleq \mathbb{E}\|f(\mathbf{y};\boldsymbol{\theta}) - \mathbf{x}\|^2. \tag{6.2}$$

Denoising function f in Eq. 6.2 is approximated by minimizing the empirical loss on a database of pairs of noisy input and ground truth clean images $\{\mathbf{y}^i, \mathbf{x}^i\}_{i=1}^N$ as

$$\min_{\boldsymbol{\theta}} \mathcal{L}_N(\boldsymbol{\theta}) \triangleq \sum_{i=1}^{N} \|f(\mathbf{y}^i;\boldsymbol{\theta}) - \mathbf{x}^i\|^2, \tag{6.3}$$

where $f(.;\boldsymbol{\theta})$ is the output of the CNN with parameters $\boldsymbol{\theta}$, and $\mathcal{L}_N(\boldsymbol{\theta})$ is the empirical loss function [45, 50, 52, 53].

6.2.2 Noise2Noise Training

Noise2Noise (N2N) training does not require any clean images [50] as opposed to N2T. N2N performs training by learning a mapping function between pairs of noisy images that have the same underlying clean images but independently drawn noises from the same distribution. The key concept of N2N is that given a set of two such degraded images, $\{\mathbf{y}^i = \mathbf{x}^i + n^i, \hat{\mathbf{y}}^i = \mathbf{x}^i + \hat{n}^i\}$, the expected value of both of these noisy images are equivalent to the clean signal. Hence, N2N modifies loss function in Eq. 6.3 into

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \|f(\mathbf{y}^i;\boldsymbol{\theta}) - \hat{\mathbf{y}}^i\|^2. \tag{6.4}$$

6.2.3 Noise2Self Training

Self-supervised training approaches enable training of CNNs without requiring a ground-truth image or pairs of noisy images. Under the assumption of independent zero-mean noise across pixels, these methods minimize a self-supervised loss on noisy images as

$$\mathcal{L}(\boldsymbol{\theta}) \triangleq \mathbb{E}\|f(\mathbf{y};\boldsymbol{\theta}) - \mathbf{y}\|^2. \tag{6.5}$$

The early pioneering work in this field is Noise2Void (N2V), which takes random patches from images and replaces the central pixel with a random pixel in the patch. Training is performed by minimizing a loss function between only the true central pixel value and estimated central pixel value of the network. While N2V empirically works well, it lacks theoretical guarantees as Eq. 6.5 simplifies to learning an identity mapping. Noise2Self (N2S) provides strong theoretical guarantees by proposing a $\mathcal{J}$-invariant f that avoids learning the identity function [53].

Definition. Let $\mathcal{J} = \{J_1, \ldots, J_N\}$ be a set of partitions of the pixels set of an image, where $\sum_{i=1}^{N} |J_i| = m$. A function $f : \mathbb{R}^m \to \mathbb{R}^m$ is $\mathcal{J}$-invariant if the value of $f(\mathbf{y})_J$ does not depend on the value of $\mathbf{y}_J$ for all $J \in \mathcal{J}$ [53].

In other words, the pixel values of an image is split into two disjoint sets J and J^c with $|J| + |J|^c = m$, and denoising function $f(\mathbf{y})_J$ uses pixels in $\mathbf{y}_{J^c}$ to denoise $\mathbf{y}_J$. Hence, rewriting the self-supevised loss function in Eq. 6.5 over $\mathcal{J}$-invariant functions leads to [53]

$$\mathbb{E}\|f(\mathbf{y};\boldsymbol{\theta}) - \mathbf{y}\|^2 = \mathbb{E}\|f(\mathbf{y};\boldsymbol{\theta}) - \mathbf{x}\|^2 + \|\mathbf{y} - \mathbf{x}\|^2. \tag{6.6}$$

Thus, minimizing the self-supervised loss over $\mathcal{J}$-invariant function is equivalent to minimizing supervised loss up to a constant defined by the variance of noise (last term). Hence, $\mathcal{J}$-invariant denoising function f can be empirically approximated over a database of noisy images as

$$\arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \sum_{J \in \mathcal{J}} \|\mathbf{P}_J f(\mathbf{y}^i_{J^c};\boldsymbol{\theta}) - \mathbf{y}^i_J\|^2, \tag{6.7}$$

where $\mathbf{P}_J$ is defined as the masking operator specified by the index set J in order to perform the loss.

6.3 Methods

Our proposed Noise2Inpaint (N2I) method builds on the $\mathcal{J}$-invariant masking idea from N2S [53], as well as a recent self-supervision method from image reconstruction that uses an optimization-focused algorithm unrolling for neural networks [10]. In conventional image denoising, a regularized objective function is typically used [46]:

$$\arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{x}\|_2^2 + \mathcal{R}(\mathbf{x}), \tag{6.8}$$

where the first term denotes a data fidelity (DF) term between the desired output and the noisy input, while the second term $\mathcal{R}(\cdot)$ is a regularizer. These regularizers can have explicit closed forms [141, 142], or the whole objective function can be solved implicitly either with traditional methods [129, 130] or using CNNs [45, 47, 52, 143, 144].

Here, we formulate the masking approach that estimates a pixel in J from its complement J^c in N2S as an image inpainting problem [145]. In image inpainting, missing pixels are estimated from available pixels using a regularized objective function. Similar to N2S, let J be the masked pixels, and J^c be the complement pixels that are available at the input of the neural network. Then, the available non-masked data in Eq. (6.1) is given as

$$\mathbf{y}_{J^C} = \mathbf{P}_{J^C}\mathbf{x} + \mathbf{n}_{J^C} \tag{6.9}$$

where $\mathbf{P}_{J^C}$ is the masking operator as defined in Section 6.2.3. While inpainting seems like a more difficult problem than denoising, this recasting allows us to write an objective function that can be solved using algorithms that enforce data fidelity and regularization.

6.3.1 Algorithm Unrolling for Inpainting

The objective function corresponding to the measurement model in (6.9) for the inpainting problem is given as

$$\arg\min_{\mathbf{x}} \|\mathbf{y}_{J^C} - \mathbf{P}_{J^C}\mathbf{x}\|_2^2 + \mathcal{R}(\mathbf{x}). \tag{6.10}$$

The regularized inpainting problem has been extensively studied using only CNNs [146–148]. An alternative approach to solve the regularized inpainting problem is to use algorithm unrolling [135]. In these methods, an iterative optimization algorithm, such as proximal gradient descent or variable splitting [149, 150] for solving the objective

function in Equation (6.10) is unrolled for a fixed number of iterations. Each iteration consists of a DF and regularizer term, as shown in Figure 6.1. The unrolled network is trained end-to-end by minimizing a loss function that characterizes the discrepancy between a reference and network output. Algorithm unrolling has gained significant popularity in many fields such as image reconstruction tasks in MRI or computational tomography due to its improved precision and accuracy [7, 10, 89, 151–153].

One approach to solve the objective function in Equation (6.10) is to use variable splitting [150], which decouples the DF and regularizer term by introducing an auxilliary variable $\mathbf{z}$ that is constrained to be equal to $\mathbf{x}$. Following variable spliting approach, the objective function in Equation (6.10) can be rewritten using quadratic relaxation:

$$\arg\min_{\mathbf{x},\mathbf{z}} \|\mathbf{y}_{J^C} - \mathbf{P}_{J^C}\mathbf{x}\|_2^2 + \mu\|\mathbf{x} - \mathbf{z}\|_2^2 + \mathcal{R}(\mathbf{z}), \tag{6.11}$$

where μ denotes the penalty parameter. This is solved by alternating minimization over $\mathbf{x}$ and $\mathbf{z}$ as

$$\mathbf{x}^{(k)} = \arg\min_{\mathbf{x}} \|\mathbf{y}_{J^C} - \mathbf{P}_{J^C}\mathbf{x}\|_2^2 + \mu\|\mathbf{x} - \mathbf{z}^{(k-1)}\|_2^2 \tag{6.12a}$$

$$\mathbf{z}^{(k)} = \arg\min_{\mathbf{z}} \mu\|\mathbf{x}^{(k)} - \mathbf{z}\|_2^2 + \mathcal{R}(\mathbf{z}) \tag{6.12b}$$

In algorithm unrolling, this problem is unrolled for a fixed number of iterations, with each iteration including a DF and a regularization block. The regularization subproblem in Equation (6.12b) does not have a closed form solution and is solved implicitly by CNNs. Equation (6.12a) corresponds to the DF term, with a closed form solution

$$\mathbf{x}_j^{(k)} = \begin{cases} \mathbf{z}_j^{(k-1)} & \text{if } j \in J \\ \frac{1}{1+\mu}\mathbf{y}_j + \frac{\mu}{1+\mu}\mathbf{z}_j^{(k-1)} & \text{if } j \in J^c \end{cases} \tag{6.13}$$

where j indicates the pixel location in the image. In other words, at iteration k in the unrolled network, the denoised image is comprised of the CNN output at the masked locations and a weighted average for the non-masked locations.

6.3.2 Noise2Inpaint Self-Supervised Training

The proposed Noise2Inpaint method performs end-to-end training by minimizing

$$\arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \sum_{J\in\mathcal{J}} \|\mathbf{P}_J\big(f_{\text{unroll}}(\mathbf{y}^i_{J^c}, \mathbf{P}_{J^c}; \boldsymbol{\theta})\big) - \mathbf{y}^i_J\|^2, \tag{6.14}$$

where $f_{\text{unroll}}(\mathbf{y}^i_{J^c}, \mathbf{P}^i_{J^c}; \boldsymbol{\theta})$ denotes the output of the unrolled network for inpainting described by Equations (6.12a)-(6.12b), with $\mathbf{y}^i_{J^c}$ and $\mathbf{P}^i_{J^c}$ denoting the inputs used at the DF units of the unrolled network. $\boldsymbol{\theta}$ includes the parameters of the CNN that implements the regularization unit of Equation (6.12b), as well as the learnable quadratic penalty parameter μ used in Equation (6.12b).

We note a few important points about the objective function in Equation (6.14):

- First, we use the same masking strategy in N2S and split noisy pixels into two set of pixels: 1) training pixels are used as network input and also in the DF units to ensure data consistency, 2) loss pixels are used to define the loss between noisy pixels excluded from the training and the network output at corresponding unseen locations.

- Second, in contrast to N2S, the N2I network has a well-defined separation between linear data consistency and the CNN-based regularization units.

- Third, as the weights of the regularization CNNs are shared across the iterations of the unrolled network, *N2I has only a single additional parameter (the penalty term μ) compared to N2S when using the same CNN architecture.*

- Finally, when using the masking scheme described in [53] for selecting J, the N2I enjoys the same theoretical guarantees as N2S.

6.4 Experiments

The proposed Noise2Inpaint method is evaluated on various denoising tasks. We compare our results with Noise2True, Noise2Noise, Noise2Self and a conventional denoising algorithm BM3D [130].

6.4.1 Datasets

BSD. A grey-scale natural images dataset is generated from the Berkeley Segmentation Dataset (BSD) [154], following [155]. This dataset contains 400 cropped images of size 180×180 with a pixel intensity range of [0-255]. BSD68 dataset (68 grey-scale images) is used for testing.

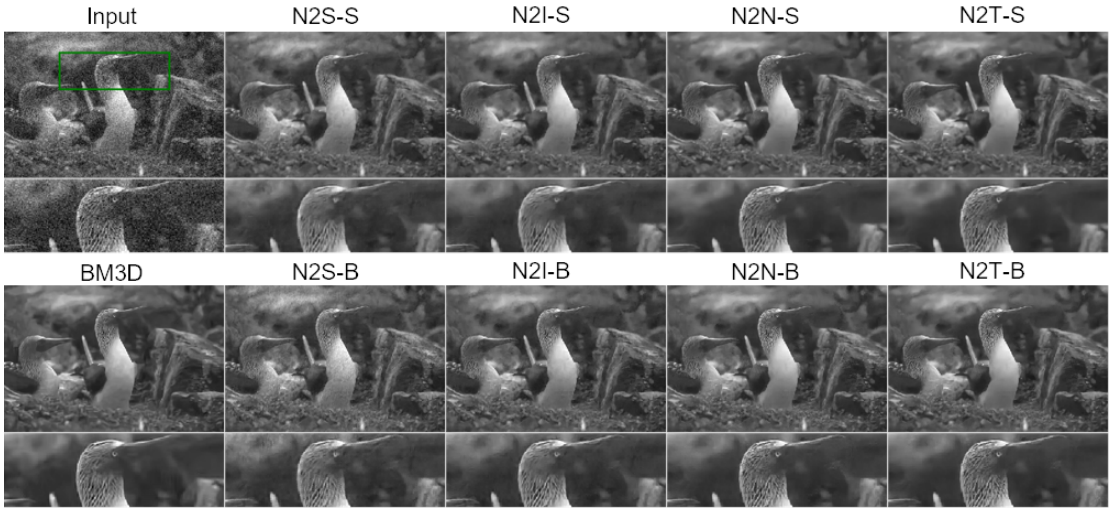

Figure 6.2: Denoising results of one representative image from BSD68 with noise level 25 for training with known specific (-S) and blind (-B) Gaussian noise. (N2S: Noise2Self, N2I: Noise2Inpaint (Ours), N2N: Noise2Noise, N2T: Noise2Truth)

Hànzì. We construct a dataset of 13029 Chinese characters (Hànzì). As in [53], the whole dataset comprises of 78174 images (each character is repeated 6 times) of size 64 $\times$ 64 and a pixel intensity range of [0-1]. The dataset is split into two as 90% and 10 % for training and testing, respectively.

ImageNet. To generate a dataset of RGB natural images, ImageNet LSVRC 2012 Validation dataset consisting of 50,000 images is used [53, 134]. The training dataset of 60,000 cropped images of size 128 $\times$ 128 wih a pixel intensity range of [0-255] is constructed from the first 20,000 images. Another 1,000 different images are used for

Methods	BM3D	N2S-S	N2S-B	N2I-S	N2I-B	N2N-S	N2N-B	N2T-S	N2T-B
σ=15	31.09	25.54	27.31	30.14	30.1	31.21	30.74	31.23	30.85
σ=25	28.59	26.51	25.64	28.02	28.03	28.72	28.35	28.75	28.49
σ=50	25.68	23.55	23.58	25.17	24.92	25.75	25.31	25.75	25.56

Table 6.1: Average PSNR results on the BSD68 dataset for known specific and blind Gaussian denoising using BM3D, Noise2Self-Specific/Blind(N2S-S,N2S-B), Noise2Inpaint-Specific/Blind (N2I-S,N2I-B), Noise2Noise-Specific/Blind (N2N-S,N2N-B) and Noise2True-Specific/Blind (N2T-S,N2T-B).

testing.

Fluorescence Microscopy. To show the utility of self-supervised denoising methods in real world applications, two fluorescence microscopy datasets from the Cell Tracking challenge [156] (Fluo-C2DL-MSC and Fluo-N2DH-GOWT1) are used. These datasets only contain single noisy images [48]. Training dataset for each of these microscopy datasets contain 100 cropped images of size 512×512 with a pixel intensity range of [0-255].

6.4.2 Implementation Details

Experiments are performed on two U-Net architectures. We use a shallow U-Net architecture of depth 2, with a linear function in the last layer [48] for all experiments on BSD and fluorescence microscopy datasets. For the experiments on Hanzi and ImageNet datasets, we use a deeper U-Net (depth 4) architecture with batch normalization and a batch size of 64 as in [53]. For both networks: The number of channel in initial level is set to 32 channels and it doubles as it goes deeper; kernel size 3; Adam optimizer with learning rate of 10^{-5}. Since our study focuses on enabling self-supervised learning from an algorithm unrolling perspective, we use CNNs that have been previously utilized in self-supervised denoising literature. However, further improvements may be possible with other well-designed neural networks. PSNR is used as evaluation criterion, when a reference image is available. Training datasets are augmented by rotating each image 90° three times and and mirroring them.

Noise2True, Noise2Noise and Noise2Self are trained as described in Section 6.2. These methods use a purely data-driven CNN as previously described. Our Noise2Inpaint method is trained end-to-end by unrolling the iterative algorithm in Eq. 6.12a, and 6.12b for 10 iterations. Each iteration contains a data fidelity and regularization term, where the trainable parameters are shared across iterations. Hence, Noise2Inpaint has only 1 more trainable parameter, which is the μ penalty parameter for quadratic relaxation, compared to N2T, N2N and N2S. During training, we follow [53] and randomly choose one single mask J for each image with density $1/25$ to speed up the training process. During testing, we input the full noisy image on the trained network as this has been reported [53] to outperform the strategy of applying a partition $\mathcal{J}$ containing $|\mathcal{J}|$ sets and averaging them.

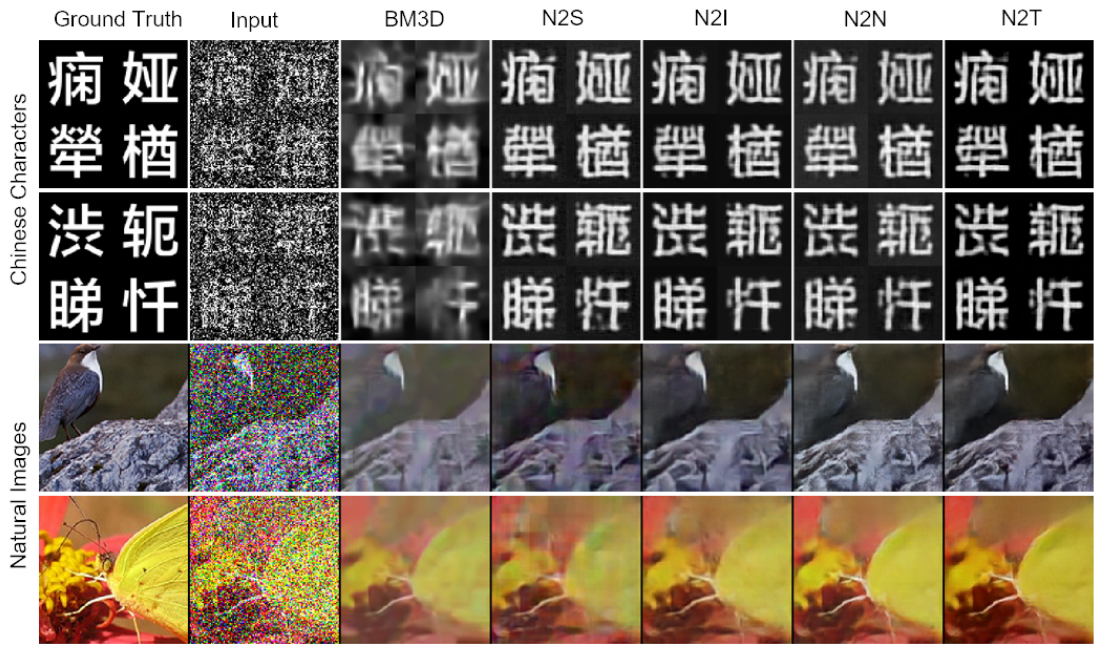

Figure 6.3: Denoising performance on Chinese characters (Hànzì) and RGB natural images (ImageNet) test datasets using traditional denoising method BM3D, supervised method Noise2True, supervised with second noisy image Noise2Noise, self-supervised approaches Noise2Self and Noise2Inpaint.

6.4.3 Known and Blind Gaussian Noise Removal

We perform two set of experiments using BSD400 dataset to analyze denoising performance in the presence of known and blind Gaussian noise. For known noise level, we use three noise levels σ =15, 25 and 50, and train a network for each noise level [155]. For blind Gaussian denoising, we train a single model using noisy images selected from noise levels $\sigma \in [0, 50]$.

Average PSNR results on test BSD68 dataset for known and blind Gaussian denoising are shown in Table 6.1. We refer to each method trained with known specific and blind noise level as Method-S and Method-B, respectively following the convention of [155]. As anticipated, Noise2True and Noise2Noise with known noise levels achieves the best PSNR results as they use extra information. Among the self-supervised approaches, our method Noise2Inpaint outperforms Noise2Self for both known and blind

noise cases. Figure 6.2 displays denoising results at noise level $\sigma = 25$. The proposed Noise2Inpaint approach shows superior reconstruction quality compared to Noise2Self by preserving more details for known noise removal, and reducing non-uniform background artifacts with subjectively appealing reconstruction in blind denoising.

6.4.4 Mixture Noise Removal

Hànzì and ImageNet datasets are evaluated with a mixture of different noise models. For the Hànzì dataset, a mixture of Gaussian ($\sigma = 0.7$) and Bernoulli noise (half the pixels blacked out) are applied to each clean image, as in [53]. For the ImageNet dataset, multiplicative Poisson noise ($\lambda = 30$), additive Gaussian noise ($\sigma = 80$) and Bernoulli noise ($p = 0.2$) is applied to each clean image, following [53].

Average PSNR values over the test datasets for both Hànzì and ImageNet are listed in Table 6.2. While N2T achieves the highest metrics among all methods, Noise2Inpaint achieves higher PSNR compared to the other ground-truth free approaches, N2S and BM3D. Figure 6.3 illustrates the visual results on representative test images. N2I achieves a better denoising quality compared to N2S and BM3D, which aligns with the quantitative metrics.

6.4.5 Denoising of Fluorescence Microscopy Data

We evaluate performance of blind denoising methods on Fluorence microscopy datasets, which contain only single sets of noisy images. Hence, Noise2True and Noise2Noise are not applicable. Figure 6.4 shows the performance of BM3D, and self-supervised N2S and N2I approaches. We assess the results qualitatively with visual inspection as quantitative metrics such as PSNR cannot be reported due to lack of ground-truth

Methods	BM3D	N2S	N2I	N2N	N2T
Hanzi	10.69	12.70	13.55	12.79	13.99
ImageNet	18.18	19.12	20.26	20.72	20.97

Table 6.2: Average PSNR results on the Hànzì and ImageNet dataset for mixture noise levels.

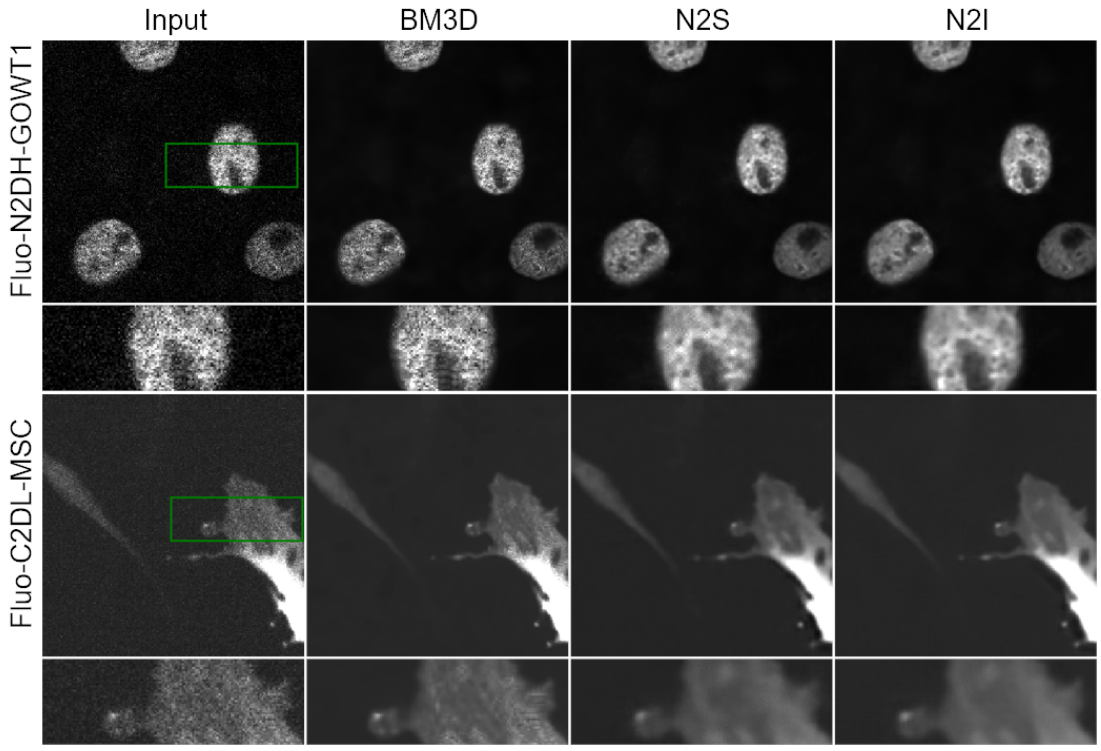

Figure 6.4: Representative results from fluorescence microscopy datasets Fluo-N2DH-GOWT1 and Fluo-C2DL-MSC for traditional denoising method BM3D and self-supervision methods Noise2Self and Noise2Inpaint. Note that Noise2True and Noise2Noise are not applicable as microscopy datasets contain only single noisy images.

reference data. The zoomed-in regions show that Noise2Inpaint achieves a superior denoising quality compared to BM3D and N2S by suppressing the noise further and achieving a more spatially uniform and visually appealing result.

6.5 Conclusions

We proposed the Noise2Inpaint approach, a self-supervised deep learning algorithm for image denoising from only noisy images. In particular, we first recast the denoising problem with holdout self-supervision as an iterative regularized inpainting problem

consisting of data fidelity and regularizer terms, and then unroll the the iterative algorithm for fixed number of iterations. The training of this network was performed end-to-end by partitioning the noisy image pixels into two disjoint sets, similar to the purely data-driven Noise2Self, where one set was utilized in the data fidelity units of the unrolled network, while the other was used to define the loss. The experiments on different datasets showed that the proposed Noise2Inpaint outperforms its purely data-driven counterpart Noise2Self.

Chapter 7

Conclusion

7.1 Thesis Summary

In this thesis, we introduced novel self-supervised learning techniques to solve inverse problems for image reconstruction. While MRI is a commonly used a medical imaging modality as it is non-invasive and radiation-free, its lengthy acquisition times has remained a major limitation. Conventional accelerated MRI techniques has been used in clinical studies, however their acceleration is limited by noise amplification and/or residual artifacts. Recently, deep learning approaches have emerged as an alternative approach for accelerated MRI to reduce the scan time further, while providing enhanced high-quality reconstructions that are clinically relevant. Most of these current deep learning approaches rely on the availability of fully-sampled datasets to perform supervised training. However, training these powerful algorithms with the current approaches is difficult, since in many scenarios, acquisition of fully-sampled dataset is either infeasible or impractical. Impediments include organ motion (e.g. cardiac motion), signal decay during some MR scans, as well as excessively long scan times during which the subject has to remain still. In addition to these challenges, database-trained approaches may face robustness issues due to shifts in acquisition parameters or distribution. Moreover, acquisition of a large databases may not be feasible. This thesis aimed to deliver state-of-the-art deep learning solutions to tackle these challenges.

In Chapter 2, we proposed the SSDU approach, a self-supervised physics-guided deep learning technique. Unlike the existent supervised learning techniques, SSDU enables

training for MRI reconstruction without fully-sampled data. Experimental results on fully-sampled knee datasets showed that conventional approaches suffered from aliasing artifacts and noise amplification, whereas deep learning approaches achieved artifact free improved reconstruction. More importantly, proposed self-supervised learning approach showed on-par performance with supervised deep learning approaches despite using only undersampled measurements in contrast to its supervised learning counterpart. Experiments on prospectively sub-sampled brain datasets, where supervised deep learning becomes inoperative, showed that proposed SSDU approach significantly outperformed the clinical conventional method. A radiologist reader study further validated the observations, which highlighted the potential utility of SSDU for clinical studies. SSDU performs end-to-end training and evaluation through only the acquired measurements without making any other assumptions about image output or characteristics. Thus, SSDU is broadly applicable in practical applications as it is not limited to a specific network architecture or acquisition parameters such as under-sampling pattern.

In Chapter 3, we investigated the extension of self-supervised learning for 3D datasets. Specifically, we analyzed the self-supervised training of 3D LGE CMR datasets, which is a main diagnostic tool for assessment of myocardial fibrosis. In the 3D SSDU extension, we introduced partitioning of the large volumes into sub-volumes which helped to tackle the GPU constraints and data scarcity issues. Experimental results showed that self-supervised 3D processing of LGE CMR datasets outperformed 2D processing, further indicating efficiency of 3D processing in capturing multi-dimensional interactions. Moreover, expert reader study results showed that the self-supervised 3D processing at rate 6 outperformed a current clinically used accelerated MRI approach at rate 3. The promising results on 3D LGE CMR datasets further validated the utility of deploying self-supervised learning in practical settings for further facilitating healthcare.

In Chapter 4, we proposed a multi-mask SSDU framework. The performance of SSDU degrades at high acceleration rates due to scarcity of acquired data. Hence, SSDU performance might suffer from artifacts in highly accelerated MRI reconstruction. The proposed multi-mask SSDU efficiently utilizes the available acquired dataset by augmenting the existent dataset by retrospectively splitting available undersampled measurements into multiple pairs of disjoint sets, in which one of them is used in DC units and the other is used to define loss. A uniform masking strategy used in multi-mask

SSDU, ensuring both low and high frequency signals are well represented in training and loss sets. Experimental results at high acceleration rates showed that multi-mask SSDU achieves artifact-free reconstruction, while SSDU and supervised deep learning approaches suffer from artifacts. Reader studies further consolidate findings by consistently rating multi-mask SSDU higher than other reconstruction methods.

In Chapter 5, we introduced a zero-shot self-supervised learning approach. While self-supervised learning enabled the training of networks without fully-sampled data, requirement on a database of undersampled measurements remains a challenge for numerous medical imaging applications. We showed that ZS-SSL enables training and testing on a single slice unlike database training approaches. Zero-shot deep learning algorithms such as deep image prior suffers from overfitting, thus requires a heuristic early stopping. In ZS-SSL, we presented a rigorous stopping criterion based on self-validation that ensures avoid overfitting. Moreover, we showed that ZS-SSL can be efficiently combined with models pre-trained on a database with different characteristics and distribution via transfer learning for faster convergence time and reduced computational complexity. Experimental results showed that ZS-SSL significant outperforms conventional parallel imaging and its zero-shot counterpart deep image prior based reconstruction. Moreover, ZS-SSL achieved on-par performance with database-trained methods, when training and testing data were matched, and significantly outperformed them when the training and testing data differed in terms of acquisition parameters or domain.

Finally, we introduced a new method for self-supervised denoising in Chapter 6. Although there exists various self-supervised learning techniques, all these approaches are purely data-driven, which obscures re-utilizing the set of pixels separated as input to the network in end-to-end training. We showed that recasting denoising as an inpainting task led to an objective function that can be minimized using algorithm unrolling. Thus, the developed framework provided an efficient solution for leveraging both training and loss pixel sets in a single task for training. The experimental results showed that algorithm unrolling concept was beneficial for denoising as Noise2Inpaint outperformed its purely data-driven self-supervised counterparts.

In summary, the work presented in this thesis facilitates developing advanced self-supervised learning algorithms to tackle the real-world challenges. We hope that the

contributions will inspire new research directions and provide insights for inverse imaging and deep learning practitioners.

7.2 Limitations and Future Directions

Selection of the sets plays an important role for the SSDU study presented in Chapter 2. We currently partition measurements based on either a Gaussian selection or a uniform random selection. In a recent supervised learning work, sub-sampling pattern was simultaneously optimized with the reconstruction model for achieving improved reconstruction quality [157]. Thus, a self-supervised learning technique that performs adaptive partitioning of measurements based on the reconstruction error warrants further study. Devising such a self-supervised sampling strategy would be beneficial in gaining further improvements at moderate acceleration rates.

Another limitation of deep learning based reconstruction techniques is the blurring artifacts. While these approaches attempt to further accelerate MRI and preserve reconstruction quality, recent fastMRI challenge has considered reconstructions from different methods at high acceleration rates as clinically not relevant due to blurring [42].Blurring artifacts seen in the current reconstruction methods can be alleviated by using loss functions that can have finer texture, sharper edges and enhanced reconstruction quality. Perceptual losses are often used for such purposes. In a recent supervised learning study [158], perceptual losses are shown to outperform distance based losses such as ℓ_1 and ℓ_2 in terms of image quality and visual assesment by radiologists. However, such perceptual losses require a ground truth reference image which may not be possible to have in many applications. While development of perceptual loss function without a ground truth warrants further study, such a development can enable clinically relevant reconstructions at high acceleration rates .

The quantitative metrics used for evaluating deep learning based reconstruction methods is another limitation in medical imaging applications. Several studies and fastMRI challenges has reported the quantitative metrics may not always align with the reconstruction quality. For example, banding artifacts can not be captured by any currently available quantitative method and it can be assessed only visually [41,128]. In another striking case reported in the fastMRI challenge [41], none of the deep learning

reconstruction method reconstructed a lesion despite having very high quantitative metrics. Therefore, developing new clinically relevant metrics that can capture and provide pathology is essential for clinical translation of deep learning methods. More recently, a new dataset named fastMRI+ has been released and it contains bounding box and annotation for various pathologies in fastMRI dataset [159]. These annotations would enable reporting new clinically relevant metrics such as false positive rate which can also help in developing methods that are clinically more relevant.

Finally, we showed the zero-shot learning removes dependency on a database. However, zero-shot learning requires long computational times compared to inference time of database-trained deep learning approaches. The computational times can be further decreased with the design of more compact architectures. Despite the long reconstruction times, zero-shot learning might be very important in certain clinical and practical applications, where it is difficult to procure large training databases, such as whole 3D/4D volumes of functional, diffusion or perfusion MRI. This is especially important for translational studies that propose new acquisition schemes at ultra-high resolutions. Thus, future studies exploring the utility ZS-SSL for such datasets could further facilitate the efficientevaluation of new acquisition schemes.

References

[1] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger. SENSE: sensitivity encoding for fast MRI. *Magn Reson Med*, 42:952–962, 1999.

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

[3] M. Lustig and J. M. Pauly. SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space. *Magn Reson Med*, 64(2):457–471, Aug 2010.

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

[5] J. P. Haldar, D. Hernando, and Z. P. Liang. Compressed-sensing MRI with random encoding. *IEEE Trans Med Imaging*, 30(4):893–903, Apr 2011.

[6] M. Akcakaya, T. A. Basha, B. Goddu, L. A. Goepfert, K. Kissinger, V. Tarokh, W. J. Manning, and R. Nezafat. Low-dimensional-structure self-learning and thresholding: regularization beyond compressed sensing for MRI reconstruction. *Magn Reson Med*, 66(3):756–767, Sep 2011.

[7] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson Med*, 79:3055–3071, 2018.

[8] M. Akçakaya, S. Moeller, S. Weingärtner, and K. Uğurbil. Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging. *Magn Reson Med*, 81:439–453, 2019.

[9] H. K. Aggarwal, M. P. Mani, and M. Jacob. MoDL: Model-Based Deep Learning Architecture for Inverse Problems. *IEEE Trans Med Imaging*, 38:394–405, 2019.

[10] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Ugurbil, and M. Akcakaya. Self-Supervised Learning of Physics-Guided Reconstruction Neural Networks without Fully-Sampled Reference Data. *Magn Reson Med*, 84(6):3172–3191, Dec 2020.

[11] K. T. Block, M. Uecker, and J. Frahm. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. *Magn Reson Med*, 57(6):1086–1098, Jun 2007.

[12] F. Knoll, K. Bredies, T. Pock, and R. Stollberger. Second order total generalized variation (TGV) for MRI. *Magn Reson Med*, 65(2):480–491, 2011.

[13] Y. Hu and M. Jacob. Higher degree total variation (hdtv) regularization for image recovery. *IEEE Transactions on Image Processing*, 21(5):2559–2571, 2012.

[14] M. Akcakaya, S. Nam, P. Hu, M. H. Moghari, L. H. Ngo, V. Tarokh, W. J. Manning, and R. Nezafat. Compressed sensing with wavelet domain dependencies for coronary MRI: a retrospective study. *IEEE Trans Med Imaging*, 30(5):1090–1099, May 2011.

[15] M. Doneva, P. Börnert, H. Eggers, C. Stehning, J. Sénégas, and A. Mertins. Compressed sensing reconstruction for magnetic resonance parameter mapping. *Magn Reson Med*, 64(4):1114–1120, Oct 2010.

[16] S. Ravishankar and Y. Bresler. Mr image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE transactions on medical imaging*, 30(5):1028–1041, 2010.

[17] S. A. H. Hosseini, B. Yaman, S. Moeller, M. Hong, and M. Akçakaya. Dense recurrent neural networks for accelerated mri: History-cognizant unrolling of optimization algorithms. *IEEE J Sel Top Sig Proc*, 14(6):1280–1291, 2020.

[18] D. Liang, J. Cheng, Z. Ke, and L. Ying. Deep magnetic resonance image reconstruction: Inverse problems meet neural networks. *IEEE Sig Proc Mag*, 37(1):141–151, 2020.

[19] J. Trzasko and A. Manduca. Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0-minimization. *IEEE Trans Med Imaging*, 28(1):106–121, 2008.

[20] H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye. k-t FOCUSS: a general compressed sensing framework for high resolution dynamic MRI. *Magn Reson Med*, 61:103–116, 2009.

[21] C. M. Sandino, J. Y. Cheng, F. Chen, M. Mardani, J. M. Pauly, and S. S. Vasanawala. Compressed sensing: From research to clinical practice with deep neural networks: Shortening scan times for magnetic resonance imaging. *IEEE Signal Processing Magazine*, 37(1):117–127, 2020.

[22] V. Monga, Y. Li, and Y. C. Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2):18–44, 2021.

[23] S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng, and D. Liang. Accelerating magnetic resonance imaging via deep learning. In *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI)*, pages 514–517. IEEE, 2016.

[24] D. Lee, J. Yoo, S. Tak, and J. C. Ye. Deep residual learning for accelerated MRI using magnitude and phase networks. *IEEE Trans Bio Eng*, pages 1985–1995, 2018.

[25] M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanawala, G. Zaharchuk, L. Xing, and J. M. Pauly. Deep Generative Adversarial Neural Networks for Compressive Sensing MRI. *IEEE Trans Med Imaging*, 38:167–179, 2019.

[26] Y. Han, L. Sunwoo, and J. C. Ye. k-Space Deep Learning for Accelerated MRI. *IEEE Trans Med Imaging*, Jul 2019.

[27] J. Zhang and B. Ghanem. Ista-net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1828–1837, 2018.

[28] Y. Yang, J. Sun, H. Li, and Z. Xu. Deep ADMM-Net for compressive sensing MRI. In *Advances in neural information processing systems*, pages 10–18, 2016.

[29] 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*, 37:491–503, 2018.

[30] C. Qin, J. Schlemper, J. Caballero, A. N. Price, J. V. Hajnal, and D. Rueckert. Convolutional Recurrent Neural Networks for Dynamic MR Image Reconstruction. *IEEE Trans Med Imaging*, 38:280–290, 2019.

[31] M. Mardani, H. Monajemi, V. Papyan, S. Vasanawala, D. Donoho, and J. Pauly. Recurrent generative adversarial networks for proximal learning and automated compressive image recovery. *arXiv preprint arXiv:1711.10046*, 2017.

[32] K. Gregor and Y. LeCun. Learning fast approximations of sparse coding. In *Proc. Int. Conf. Mach. Learn*, pages 399–406. Omnipress, 2010.

[33] H. Haji-Valizadeh, A. A. Rahsepar, J. D. Collins, E. Bassett, T. Isakova, T. Block, G. Adluru, E. V. R. DiBella, D. C. Lee, J. C. Carr, and D. Kim. Validation of highly accelerated real-time cardiac cine MRI with radial k-space sampling and compressed sensing in patients at 1.5T and 3T. *Magn Reson Med*, 79(5):2745–2751, 05 2018.

[34] K. Ugurbil, J. Xu, E. J. Auerbach, S. Moeller, A. T. Vu, J. M. Duarte-Carvajalino, C. Lenglet, X. Wu, S. Schmitter, P. van de Moortele, J. P. Strupp, G. Sapiro, F. D. Martino, D. Wang, N. Harel, M. Garwood, L. Chen, D. A. Feinberg, S. M. Smith, K. L. Miller, S. N. Sotiropoulos, S. Jbabdi, J. L. R. Andersson, T. E. J. Behrens, M. F. Glasser, D. C. V. Essen, and E. Yacoub. Pushing spatial and temporal

resolution for functional and diffusion MRI in the human connectome project. *NeuroImage*, 80:80–104, 2013.

[35] B. Yaman, S. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya. Self-supervised physics-based deep learning mri reconstruction without fully-sampled data. In *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI)*, pages 921–925. IEEE, 2020.

[36] B. Yaman, C. Shenoy, Z. Deng, S. Moeller, H. El-Rewaidy, R. Nezafat, and M. Akçakaya. Self-supervised physics-guided deep learning reconstruction for high-resolution 3d lge cmr. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 100–104. IEEE, 2021.

[37] B. Yaman, S. Weingärtner, N. Kargas, N. D. Sidiropoulos, and M. Akçakaya. Low-Rank Tensor Models for Improved Multidimensional MRI: Application to Dynamic Cardiac T_1 Mapping. *IEEE Trans Comput Imaging*, 6:194–207, 2019.

[38] B. Yaman, S. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya. Multi-mask self-supervised learning for physics-guided neural networks in highly accelerated MRI. *arXiv preprint arXiv:2008.06029*, 2020.

[39] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya. Ground-truth free multi-mask self-supervised physics-guided deep learning in highly accelerated mri. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1850–1854. IEEE, 2021.

[40] F. Knoll, K. Hammernik, E. Kobler, T. Pock, M. P. Recht, and D. K. Sodickson. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. *Magn Reson Med*, 81(1):116–128, 01 2019.

[41] F. Knoll, T. Murrell, A. Sriram, N. Yakubova, J. Zbontar, M. G. Rabbat, A. Defazio, M. J. Muckley, D. K. Sodickson, C. L. Zitnick, and M. P. Recht. Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge. *Magn Reson Med*, Jun 2020.

[42] M. J. Muckley, B. Riemenschneider, A. Radmanesh, S. Kim, G. Jeong, J. Ko, Y. Jun, H. Shin, D. Hwang, M. Mostapha, S. Arberet, D. Nickel, Z. Ramzi,

P. Ciuciu, J. Starck, J. Teuwen, D. Karkalousos, C. Zhang, A. Sriram, Z. Huang, N. Yakubova, Y. W. Lui, and F. Knoll. Results of the 2020 fastmri challenge for machine learning MR image reconstruction. *IEEE Trans. Medical Imaging*, 40(9):2306–2317, 2021.

[43] B. Yaman, S. A. H. Hosseini, and M. Akçakaya. Zero-shot physics-guided deep learning for subject-specific MRI reconstruction. *Advances in neural information processing systems workshops*, 2021.

[44] B. Yaman, S. A. H. Hosseini, and M. Akcakaya. Zero-shot self-supervised learning for MRI reconstruction. In *International Conference on Learning Representations*, 2022.

[45] K. Zhang, W. Zuo, S. Gu, and L. Zhang. Learning deep CNN denoiser prior for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3929–3938, 2017.

[46] Y. Romano, M. Elad, and P. Milanfar. The little engine that could: Regularization by denoising (red). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017.

[47] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. In *Proc. IEEE CVPR*, June 2018.

[48] A. Krull, T.-O. Buchholz, and F. Jug. Noise2void-learning denoising from single noisy images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2129–2137, 2019.

[49] S. Soltanayev and S. Y. Chun. Training deep learning based denoisers without ground truth data. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31, pages 3257–3267. Curran Associates, Inc., 2018.

[50] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila. Noise2Noise: Learning image restoration without clean data. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pages 2965–2974. PMLR, 2018.

[51] G. Vaksman, M. Elad, and P. Milanfar. Lidia: Lightweight learned image denoising with instance adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020.

[52] K. Zhang, W. Zuo, and L. Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018.

[53] J. Batson and L. Royer. Noise2self: Blind denoising by self-supervision. In *Proceedings of the International Conference on Machine Learning*, pages 524–533, 2019.

[54] B. Yaman, S. A. H. Hosseini, and M. Akçakaya. Noise2inpaint: Learning referenceless denoising by inpainting unrolling. *arXiv preprint arXiv:2006.09450*, 2020.

[55] D. Liang, B. Liu, J. Wang, and L. Ying. Accelerating SENSE using compressed sensing. *Magn Reson Med*, 62(6):1574–1584, Dec 2009.

[56] R. Otazo, D. Kim, L. Axel, and D. K. Sodickson. Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI. *Magn Reson Med*, 64(3):767–776, Sep 2010.

[57] P. M. Robson, A. K. Grant, A. J. Madhuranthakam, R. Lattanzi, D. K. Sodickson, and C. A. McKenzie. Comprehensive quantification of signal-to-noise ratio and g-factor for image-based and k-space-based parallel imaging reconstructions. *Magn Reson Med*, 60(4):895–907, Oct 2008.

[58] Y. Chang, D. Liang, and L. Ying. Nonlinear GRAPPA: a kernel approach to parallel MRI reconstruction. *Magn Reson Med*, 68(3):730–740, Sep 2012.

[59] B. Madore. UNFOLD-SENSE: a parallel MRI method with self-calibration and artifact suppression. *Magn Reson Med*, 52(2):310–320, Aug 2004.

[60] K. Sung and B. A. Hargreaves. High-frequency subband compressed sensing MRI using quadruplet sampling. *Magn Reson Med*, 70(5):1306–1318, Nov 2013.

[61] Y. Yang, J. Sun, H. Li, and Z. Xu. Admm-csnet: A deep learning approach for image compressive sensing. *IEEE transactions on pattern analysis and machine intelligence*, 42(3):521–538, 2018.

[62] M. Shahdloo, E. Ilicak, M. Tofighi, E. U. Saritas, A. E. Çetin, and T. Çukur. Projection onto epigraph sets for rapid self-tuning compressed sensing mri. *IEEE transactions on medical imaging*, 38(7):1677–1689, 2018.

[63] S. Ramani, Z. Liu, J. Rosen, J.-F. Nielsen, and J. A. Fessler. Regularization parameter selection for nonlinear iterative image restoration and mri reconstruction using gcv and sure-based methods. *IEEE Transactions on Image Processing*, 21(8):3659–3672, 2012.

[64] J. Cheng, J. Pauly, and S. Vasanawala. Multi-channel image reconstruction with latent coils and adverserial loss. In *Proceedings of the 27th Annual Meeting of ISMRM, Montréal, Canada*, 2019.

[65] P. Wang, E. Z. Chen, T. Chen, V. M. Patel, and S. Sun. Pyramid convolutional rnn for mri reconstruction. *Advances in neural information processing systems workshops*, 2019.

[66] O. R. Coelho-Filho, C. Rickers, R. Y. Kwong, and M. Jerosch-Herold. MR myocardial perfusion imaging. *Radiology*, 266(3):701–715, Mar 2013.

[67] P. Kellman and M. S. Hansen. T1-mapping in the heart: accuracy and precision. *J Cardiovasc Magn Reson*, 16:2, Jan 2014.

[68] K. Setsompop, R. Kimmlingen, E. Eberlein, T. Witzel, J. Cohen-Adad, J. A. McNab, B. Keil, M. D. Tisdall, P. Hoecht, P. Dietz, S. F. Cauley, V. Tountcheva, V. Matschl, V. H. Lenz, K. Heberlein, A. Potthast, H. Thein, J. Van Horn, A. Toga, F. Schmitt, D. Lehne, B. R. Rosen, V. Wedeen, and L. L. Wald. Pushing the limits of in vivo diffusion MRI for the Human Connectome Project. *Neuroimage*, 80:220–233, Oct 2013.

[69] U. Gamper, P. Boesiger, and S. Kozerke. Compressed sensing in dynamic MRI. *Magn Reson Med*, 59(2):365–373, Feb 2008.

[70] J. A. Fessler. Optimization methods for magnetic resonance image reconstruction: Key models and optimization algorithms. *IEEE Sig Proc Mag*, 37(1):33–40, 2020.

[71] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE transactions on image processing*, 19(9):2345–2356, 2010.

[72] K. Hammernik, F. Knoll, D. K. Sodickson, and T. Pock. L2 or not L2: impact of loss function design for deep learning MRI reconstruction. In *ISMRM 25th Annual Meeting*, page 0687, 2017.

[73] F. Knoll, K. Hammernik, C. Zhang, S. Moeller, T. Pock, D. K. Sodickson, and M. Akçakaya. Deep-learning methods for parallel magnetic resonance imaging reconstruction: A survey of the current approaches, trends, and issues. *IEEE Sig Proc Mag*, 37(1):128–140, 2020.

[74] T. M. Quan, T. Nguyen-Duc, and W.-K. Jeong. Compressed sensing mri reconstruction using a generative adversarial network with a cyclic loss. *IEEE transactions on medical imaging*, 37(6):1488–1497, 2018.

[75] S. Arlot and M. Lerasle. Choice of v for v-fold cross-validation in least-squares density estimation. *The Journal of Machine Learning Research*, 17(1):7256–7305, 2016.

[76] R. Timofte, E. Agustsson, L. V. Gool, M. Yang, L. Zhang, B. Lim, S. Son, H. Kim, S. Nah, K. M. Lee, X. Wang, Y. Tian, K. Yu, Y. Zhang, S. Wu, C. Dong, L. Lin, Y. Qiao, C. C. Loy, W. Bae, J. J. Yoo, Y. Han, J. C. Ye, J. Choi, M. Kim, Y. Fan, J. Yu, W. Han, D. Liu, H. Yu, Z. Wang, H. Shi, X. Wang, T. S. Huang, Y. Chen, K. Zhang, W. Zuo, Z. Tang, L. Luo, S. Li, M. Fu, L. Cao, W. Heng, G. Bui, T. Le, Y. Duan, D. Tao, R. Wang, X. Lin, J. Pang, J. Xu, Y. Zhao, X. Xu, J. Pan, D. Sun, Y. Zhang, X. Song, Y. Dai, X. Qin, X. Huynh, T. Guo, H. S. Mousavi, T. H. Vu, V. Monga, C. Cruz, K. O. Egiazarian, V. Katkovnik, R. Mehta, A. K. Jain, A. Agarwalla, C. V. S. Praveen, R. Zhou, H. Wen, C. Zhu, Z. Xia, Z. Wang, and Q. Guo. NTIRE 2017 challenge on single image super-resolution: Methods and results. In *CVPR Workshops*, pages 1110–1121. IEEE Computer Society, 2017.

[77] F. Knoll, J. Zbontar, A. Sriram, M. J. Muckley, M. Bruno, A. Defazio, M. Parente, K. J. Geras, J. Katsnelson, H. Chandarana, Z. Zhang, M. Drozdzalv, A. Romero, M. Rabbat, P. Vincent, J. Pinkerton, D. Wang, N. Yakubova, E. Owens, C. L. Zitnick, M. P. Recht, D. K. Sodickson, and Y. W. Lui. fastMRI: A Publicly Available Raw k-Space and DICOM Dataset of Knee Images for Accelerated MR Image Reconstruction Using Machine Learning. *Radiol Artif Intell*, 2(1):e190007, Jan 2020.

[78] K. P. Pruessmann, M. Weiger, P. Bornert, and P. Boesiger. Advances in sensitivity encoding with arbitrary k-space trajectories. *Magn Reson Med*, 46:638–651, 2001.

[79] F. A. Breuer, M. Blaimer, M. F. Mueller, N. Seiberlich, R. M. Heidemann, M. A. Griswold, and P. M. Jakob. Controlled aliasing in volumetric parallel imaging (2D CAIPIRINHA). *Magn Reson Med*, 55(3):549–556, Mar 2006.

[80] O. Senouf, S. Vedula, T. Weiss, A. Bronstein, O. Michailovich, and M. Zibulevsky. Self-supervised learning of inverse problem solvers in medical imaging. In *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data*, pages 111–119. Springer, 2019.

[81] P. Huang, C. Zhang, H. Li, S. Gaire, R. Liu, X. Zhang, X. Li, and L. Ying. Deep mri reconstruction without ground truth for training. In *Proceedings of the 27th Annual Meeting of the ISMRM, Montréal, Canada*, 2019.

[82] J. Liu, Y. Sun, C. Eldeniz, W. Gan, H. An, and U. S. Kamilov. Rare: Image reconstruction using deep priors learned without groundtruth. *IEEE J Sel Top Sig Proc*, 14:1088–1099, 2020.

[83] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila. Noise2Noise: Learning image restoration without clean data. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pages 2965–2974. PMLR, 2018.

[84] B. Sim, G. Oh, J. Kim, C. Jung, and J. C. Ye. Optimal transport driven cyclegan for unsupervised learning in inverse problems. *SIAM Journal on Imaging Sciences*, 13(4):2281–2306, 2020.

[85] K. Lei, M. Mardani, J. M. Pauly, and S. S. Vasanawala. Wasserstein gans for mr imaging: from paired to unpaired training. *IEEE transactions on medical imaging*, 40(1):105–115, 2020.

[86] J. I. Tamir, S. X. Yu, and M. Lustig. Unsupervised deep basis pursuit: Learning inverse problems without ground-truth data. *Advances in neural information processing systems workshops*, 2019.

[87] K. Hammernik, F. Knoll, D. K. Sodickson, and T. Pock. On the influence of sampling pattern design on deep learning-based mri reconstruction. In *ISMRM 25th Annual Meeting*, page 0644, 2017.

[88] M. W. Browne. Cross-validation methods. *Journal of mathematical psychology*, 44(1):108–132, 2000.

[89] M. Kellman, K. Zhang, E. Markley, J. Tamir, E. Bostan, M. Lustig, and L. Waller. Memory-efficient learning for large-scale computational imaging. *IEEE Trans Comp Imaging*, 6:1403–1414, 2020.

[90] S. U. H. Dar, M. Özbey, A. B. Çatlı, and T. Çukur. A transfer-learning approach for accelerated MRI using deep neural networks. *Magn Reson Med*, 84(2):663–685, 2020.

[91] Y. Han, J. Yoo, H. H. Kim, H. J. Shin, K. Sung, and J. C. Ye. Deep learning with domain adaptation for accelerated projection-reconstruction MR. *Magn Reson Med*, 80(3):1189–1205, 09 2018.

[92] Y. Hu, X. Shi, Q. Tian, H. Guo, M. Deng, M. Yu, C. Moran, G. Yang, J. McNab, B. Daniel, et al. Reconstruction of multi-shot diffusion-weighted mri using unrolled network with u-nets as priors. In *Proc. 27th Annu. Meeting ISMRM*, 2019.

[93] B. Yaman, S. A. H. Hosseini, S. Moeller, and M. Akçakaya. Comparison of neural network architectures for physics-driven deep learning mri reconstruction. In *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pages 0155–0159. IEEE, 2019.

[94] I. Sánchez and V. Vilaplana. Brain mri super-resolution using 3d generative adversarial networks. *arXiv preprint arXiv:1812.11440*, 2018.

[95] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, et al. Dagan: Deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction. *IEEE transactions on medical imaging*, 37(6):1310–1321, 2017.

[96] S. U. Dar, M. Yurt, M. Shahdloo, M. E. Ildız, B. Tınaz, and T. Çukur. Prior-guided image reconstruction for accelerated multi-contrast mri via generative adversarial networks. *IEEE Journal of Selected Topics in Signal Processing*, 14(6):1072–1087, 2020.

[97] R. J. Kim, D. S. Fieno, T. B. Parrish, K. Harris, E. L. Chen, O. Simonetti, J. Bundy, J. P. Finn, F. J. Klocke, and R. M. Judd. Relationship of MRI delayed contrast enhancement to irreversible injury, infarct age, and contractile function. *Circulation*, pages 1992–2002, Nov 1999.

[98] M. Saranathan, C. E. Rochitte, and T. K. Foo. Fast, three-dimensional free-breathing MR imaging of myocardial infarction: a feasibility study. *Magn Reson Med*, 51(5):1055–1060, May 2004.

[99] M. Akcakaya, H. Rayatzadeh, T. A. Basha, S. N. Hong, R. H. Chan, K. V. Kissinger, T. H. Hauser, M. E. Josephson, W. J. Manning, and R. Nezafat. Accelerated late gadolinium enhancement cardiac MR imaging with isotropic spatial resolution using compressed sensing: initial experience. *Radiology*, 264(3):691–699, Sep 2012.

[100] T. A. Basha, M. Akcakaya, C. Liew, C. W. Tsao, F. N. Delling, G. Addae, L. Ngo, W. J. Manning, and R. Nezafat. Clinical performance of high-resolution late gadolinium enhancement imaging with compressed sensing. *J Magn Reson Imaging*, 46(6):1829–1838, 12 2017.

[101] G. Adluru, L. Chen, S.-E. Kim, N. Hu, K. H. Sabey, D. A. Bull, E. Kholmovski, N. Marrouche, and E. DiBella. Late gadolinium enhancement imaging using stack of stars and compressed sensing. *J Cardiovasc Magn Reson*, 13, 2011.

[102] G. Adluru, L. Chen, S. E. Kim, N. Burgon, E. G. Kholmovski, N. F. Marrouche, and E. V. Dibella. Three-dimensional late gadolinium enhancement imaging of the left atrium with a hybrid radial acquisition and compressed sensing. *J Magn Reson Imaging*, 34:1465–1471, 2011.

[103] S. Kamesh Iyer, T. Tasdizen, N. Burgon, E. Kholmovski, N. Marrouche, G. Adluru, and E. DiBella. Compressed sensing for rapid late gadolinium enhanced imaging of the left atrium: A preliminary study. *Magn Reson Imaging*, 34(7):846–854, Sep 2016.

[104] 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 Comp Imaging*, 6:1285–1296, 2020.

[105] Z. Ke, J. Cheng, L. Ying, H. Zheng, Y. Zhu, and D. Liang. An unsupervised deep learning method for multi-coil cine MRI. *Physics in Medicine & Biology*, 2020.

[106] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig. ESPIRiT–an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA. *Magn Reson Med*, 71(3):990–1001, Mar 2014.

[107] T. Küstner, N. Fuin, K. Hammernik, A. Bustin, H. Qi, R. Hajhosseiny, P. G. Masci, R. Neji, D. Rueckert, R. M. Botnar, et al. Cinenet: deep learning-based 3D cardiac cine MRI reconstruction with multi-coil complex-valued 4D spatio-temporal convolutions. *Scientific Reports*, 2020.

[108] S. G. Lingala and M. Jacob. Blind compressive sensing dynamic mri. *IEEE Transactions on Medical Imaging*, 32(6):1132–1145, 2013.

[109] S. D. Sharma, C. L. Fong, B. S. Tzung, M. Law, and K. S. Nayak. Clinical image quality assessment of accelerated magnetic resonance neuroimaging using compressed sensing. *Invest Radiol*, 48(9):638–645, Sep 2013.

[110] F. Ong, S. Amin, S. Vasanawala, and M. Lustig. Mridata. org: An open archive for sharing mri raw data. In *Proc. Intl. Soc. Mag. Reson. Med*, volume 26, page 1, 2018.

[111] B. Yaman, S. A. H. Hosseini, S. Moeller, and M. Akçakaya. Improved supervised training of physics-guided deep learning image reconstruction with multi-masking. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1150–1154. IEEE, 2021.

[112] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick. Momentum contrast for unsupervised visual representation learning. In *Proc IEEE CVPR*, June 2020.

[113] I. Misra and L. v. d. Maaten. Self-supervised learning of pretext-invariant representations. In *Proc IEEE CVPR*, June 2020.

[114] C. Belthangady and L. A. Royer. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nat Methods*, 16(12):1215–1225, 12 2019.

[115] C. Shorten and T. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60, 2019.

[116] P. Henderson, J. Hu, J. Romoff, E. Brunskill, D. Jurafsky, and J. Pineau. Towards the systematic reporting of the energy and carbon footprints of machine learning. *Journal of Machine Learning Research*, 21(248):1–43, 2020.

[117] O. B. Demirel, B. Yaman, L. Dowdle, S. Moeller, L. Vizioli, E. Yacoub, J. Strupp, C. A. Olman, K. Uğurbil, and M. Akçakaya. 20-fold accelerated 7t fmri using referenceless self-supervised deep learning reconstruction. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 3765–3769, 2021.

[118] C. M. Sandino, P. Lai, S. S. Vasanawala, and J. Y. Cheng. Accelerating cardiac cine MRI using a deep learning-based ESPIRiT reconstruction. *Magn Reson Med*, 85(1):152–167, 01 2021.

[119] C. Qin, J. Schlemper, J. Duan, G. Seegoolam, A. Price, J. Hajnal, and D. Rueckert. k-t next: dynamic mr image reconstruction exploiting spatio-temporal correlations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 505–513. Springer, 2019.

[120] Y. Korkmaz, S. U. Dar, M. Yurt, M. Özbey, and T. Çukur. Unsupervised mri reconstruction via zero-shot learned adversarial transformers. *arXiv preprint arXiv:2105.08059*, 2021.

[121] M. Akçakaya, B. Yaman, H. Chung, and J. C. Ye. Unsupervised deep learning methods for biological image reconstruction and enhancement. *IEEE Sig Proc Mag*, 2022.

[122] Y. C. Eldar, A. O. H. III, L. Deng, J. A. Fessler, J. Kovacevic, H. V. Poor, and S. J. Young. Challenges and open problems in signal processing: Panel discussion summary from ICASSP. *IEEE Signal Process. Mag.*, 34(6):8–23, 2017.

[123] S. A. H. Hosseini, B. Yaman, S. Moeller, and M. Akçakaya. High-fidelity accelerated mri reconstruction by scan-specific fine-tuning of physics-based neural networks. In *IEEE Engineering in Medicine Biology Society (EMBC)*, pages 1481–1484, 2020.

[124] A. Shocher, N. Cohen, and M. Irani. "Zero-shot" super-resolution using deep internal learning. In *Proc IEEE CVPR*, June 2018.

[125] M. Z. Darestani, A. S. Chaudhari, and R. Heckel. Measuring robustness in deep learning based compressive sensing. In *ICML*, volume 139 of *Proceedings of Machine Learning Research*, pages 2433–2444. PMLR, 2021.

[126] Y. Quan, M. Chen, T. Pang, and H. Ji. Self2self with dropout: Learning self-supervised denoising from single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2020.

[127] R. Jafari, P. Spincemaille, J. Zhang, T. D. Nguyen, M. R. Prince, X. Luo, J. Cho, D. Margolis, and Y. Wang. Deep neural network for water/fat separation: Supervised training, unsupervised training, and no training. *Magn Reson Med*, 85(4):2263–2277, 04 2021.

[128] A. Defazio, T. Murrell, and M. Recht. MRI banding removal via adversarial training. In *Advances in Neural Information Processing Systems*, volume 33, pages 7660–7670, 2020.

[129] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65. IEEE, 2005.

[130] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007.

[131] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12):3736–3745, 2006.

[132] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. In *2011 International Conference on Computer Vision*, pages 479–486. IEEE, 2011.

[133] S. Lefkimmiatis. Universal denoising networks: a novel cnn architecture for image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3204–3213, 2018.

[134] Y. Xie, Z. Wang, and S. Ji. Noise2same: Optimizing a self-supervised bound for image denoising. *Advances in Neural Information Processing Systems*, 33, 2020.

[135] K. Gregor and Y. LeCun. Learning fast approximations of sparse coding. In *Proc Int Conf Mach Learning*, pages 399–406, 2010.

[136] J. Adler and O. Oktem. Learned Primal-Dual Reconstruction. *IEEE Trans Med Imaging*, 37(6):1322–1332, 06 2018.

[137] H. Liu, C. Yang, N. Pan, E. Song, and R. Green. Denoising 3d mr images by the enhanced non-local means filter for rician noise. *Magnetic resonance imaging*, 28(10):1485–1496, 2010.

[138] J. V. Manjón, P. Coupé, A. Buades, D. L. Collins, and M. Robles. New methods for mri denoising based on sparseness and self-similarity. *Medical image analysis*, 16(1):18–27, 2012.

[139] A. Buades, B. Coll, and J.-M. Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4(2):490–530, 2005.

[140] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin. Deep learning on image denoising: An overview. *Neural Networks*, 2020.

[141] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010.

[142] D. L. Donoho. De-noising by soft-thresholding. *IEEE transactions on information theory*, 41(3):613–627, 1995.

[143] M. Gharbi, G. Chaurasia, S. Paris, and F. Durand. Deep joint demosaicking and denoising. *ACM Transactions on Graphics (TOG)*, 35(6):1–12, 2016.

[144] J. Xu, L. Zhang, W. Zuo, D. Zhang, and X. Feng. Patch group based nonlocal self-similarity prior learning for image denoising. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[145] C. Guillemot and O. Le Meur. Image inpainting: Overview and recent advances. *IEEE signal processing magazine*, 31(1):127–144, 2013.

[146] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. Efros. Context encoders: Feature learning by inpainting. In *Computer Vision and Pattern Recognition (CVPR)*, 2016.

[147] Z. Yan, X. Li, M. Li, W. Zuo, and S. Shan. Shift-net: Image inpainting via deep feature rearrangement. In *The European Conference on Computer Vision (ECCV)*, September 2018.

[148] C. Yang, X. Lu, Z. Lin, E. Shechtman, O. Wang, and H. Li. High-resolution image inpainting using multi-scale neural patch synthesis. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[149] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer, 2011.

[150] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008.

[151] M. Zhussip, S. Soltanayev, and S. Y. Chun. Training deep learning based image denoisers from undersampled measurements without ground truth and without image prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[152] S. Sreehari, S. V. Venkatakrishnan, B. Wohlberg, G. T. Buzzard, L. F. Drummy, J. P. Simmons, and C. A. Bouman. Plug-and-play priors for bright field electron tomography and sparse interpolation. *IEEE Trans Comp Imaging*, 2(4):408–423, 2016.

[153] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang. Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE Trans Med Imaging*, 37(6):1348–1357, 2018.

[154] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 416–423. IEEE, 2001.

[155] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.

[156] V. Ulman, M. Maška, K. E. Magnusson, O. Ronneberger, C. Haubold, N. Harder, P. Matula, P. Matula, D. Svoboda, M. Radojevic, et al. An objective comparison of cell-tracking algorithms. *Nature methods*, 14(12):1141, 2017.

[157] C. D. Bahadir, A. V. Dalca, and M. R. Sabuncu. Learning-based optimization of the under-sampling pattern in mri. In *International Conference on Information Processing in Medical Imaging*, pages 780–792. Springer, 2019.

[158] V. Ghodrati, J. Shao, M. Bydder, Z. Zhou, W. Yin, K. L. Nguyen, Y. Yang, and P. Hu. MR image reconstruction using deep learning: evaluation of network structure and loss functions. *Quant Imaging Med Surg*, 9(9):1516–1527, Sep 2019.

[159] R. Zhao, B. Yaman, Y. Zhang, R. Stewart, A. Dixon, F. Knoll, Z. Huang, Y. W. Lui, M. S. Hansen, and M. P. Lungren. fastmri+: Clinical pathology annotations for knee and brain fully sampled multi-coil mri data. *arXiv preprint arXiv:2109.03812*, 2021.

Appendix A

A.1 Supporting Information for Chapter 2

See **Figures A.1, A.2, A.3, A.4,A.5, A.6, A.7, A.8,A.9, A.10, A.11, A.12, A.13**, and **Tables A.1** & **A.2**.

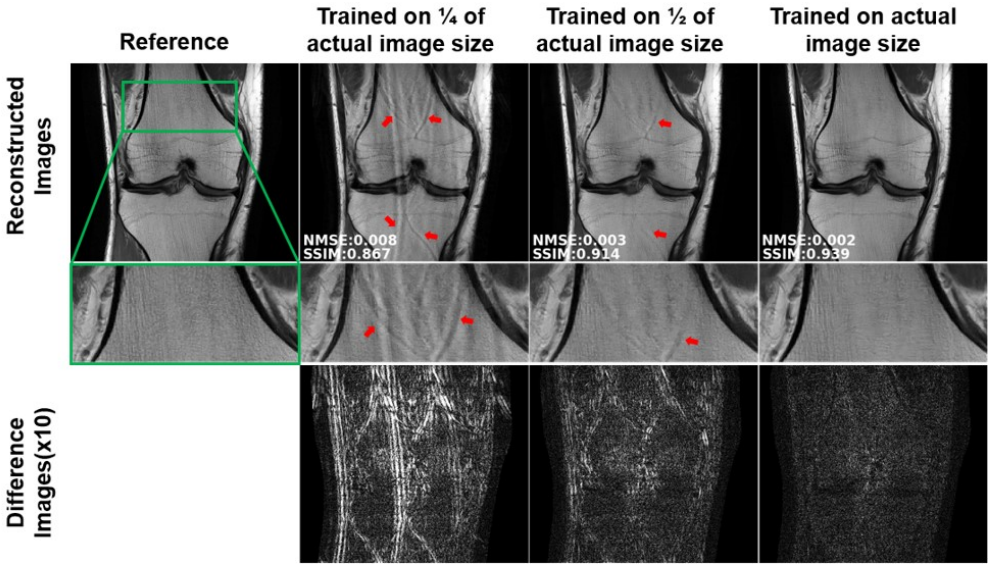

Figure A.1: Reconstruction results for the generalization performance of supervised training across different image matrix sizes. The networks are trained in by taking actual k-space, the central $\frac{1}{2}$ of the k-space (i.e. reducing the resolution by 2-fold), and the central $\frac{1}{4}$ of the k-space (i.e. reducing the resolution by 4-fold). All trained networks are then applied on actual size data to test generalization. The generalization performance of CNNs on actual image size degrades as training image size get smaller, with 1/4 k-space performing the worst.

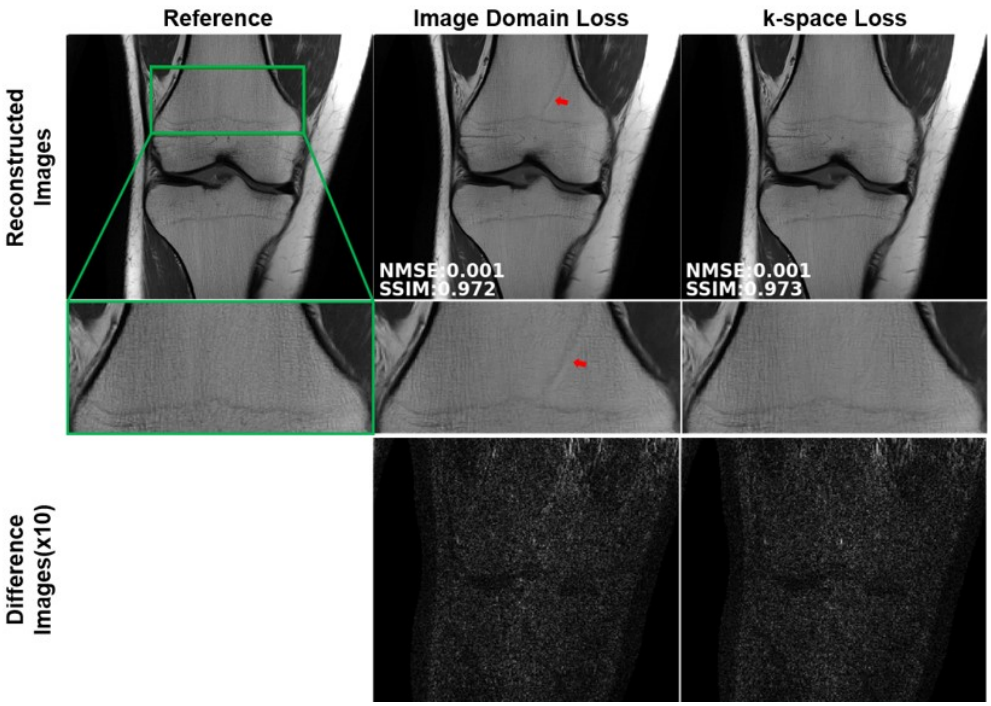

Figure A.2: Reconstruction results for supervised training with image domain and k-space losses. When using image domain loss, the reconstruction suffers from residual artifacts (red arrows), whereas using k-space loss suppresses these artifacts. Difference images also show that the supervised training with k-space loss has fewer residual artifacts. Across the dataset, the two approaches perform quantitatively similar. The median and interquartile range for SSIM values across test dataset were 0.967 [0.955, 0.978], 0.966 [0.956, 0.0977], and for NMSE values were 0.001 [0.001, 0.002], 0.001 [0.001, 0.002] for supervised with image domain and k-space losses, respectively.

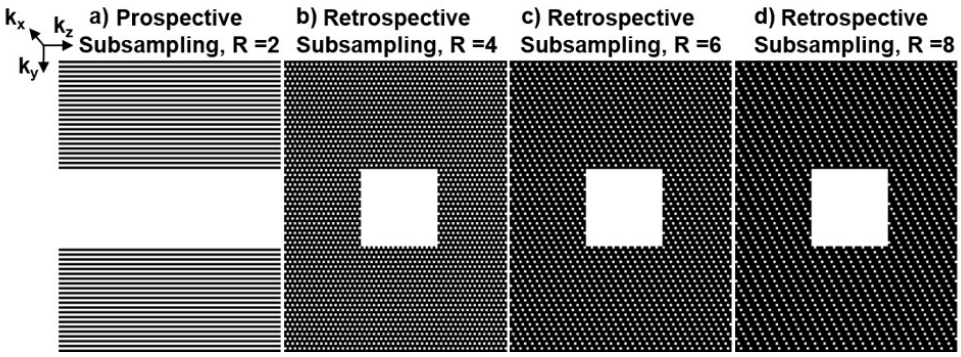

Figure A.3: Sub-sampling masks used in the brain MRI study. Prospective subsampling was equispaced with R = 2 in ky and 32 ACS lines. Subsampling patterns for R = 4, 6, 8 were obtained by sheared sub-sampling, while keeping the center 32×32 ACS region in the $k_y - k_z$ plane.

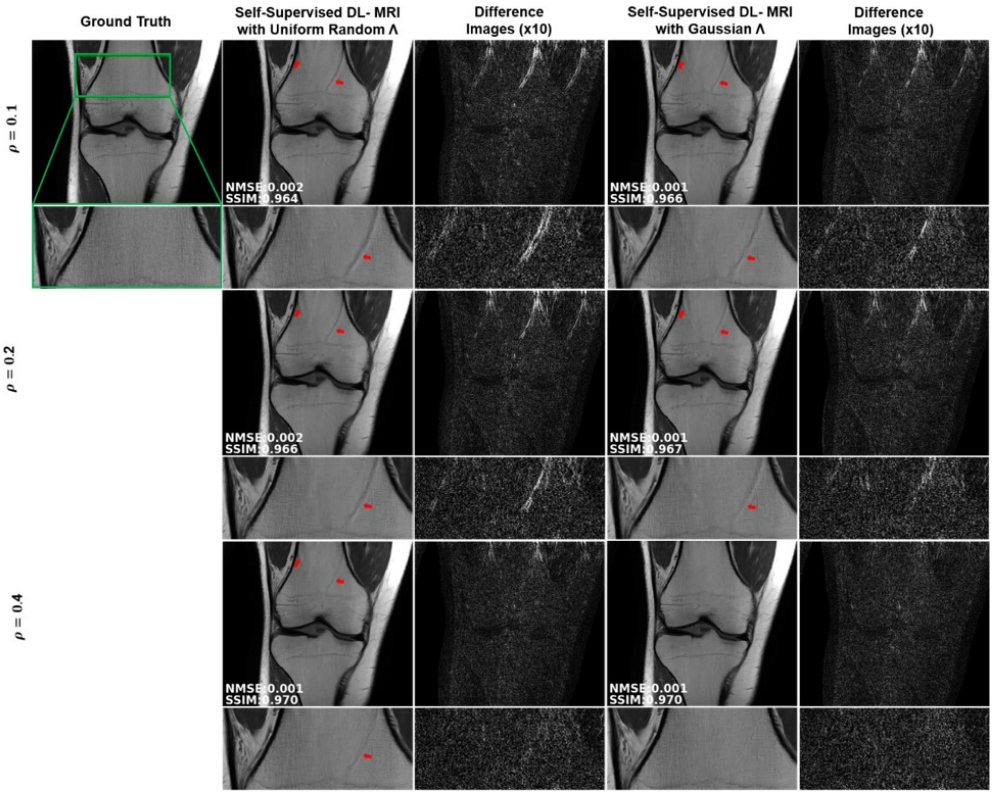

Figure A.4: Reconstruction results from self-supervised training with uniform random selection and variable-density Gaussian selection of Λ for $\rho \in \{0.1, 0.2, 0.4\}$. Gaussian random selection consistently outperforms the uniform random selection at all ρ values in terms of reconstruction quality and suppression of residual artifacts, which is also highlighted in the difference images. For $\rho \in \{0.1, 0.2\}$ both uniform and Gaussian random selection show visible residual artifacts, marked by red arrows, with former showing more residual artifacts. For $\rho = 0.4$, uniform random selection still suffers from visible residual artifacts, whereas Gaussian selection further suppress those artifacts and achieves artifact free reconstruction. Difference images further confirms the observations.

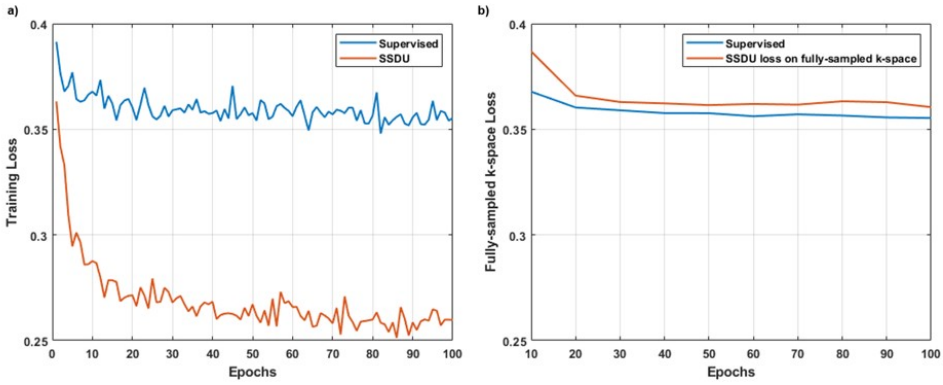

Figure A.5: a) Training loss for supervised and self-supervised training approaches. In both cases, the loss decreases over epochs. Self-supervised approach achieves a lower loss value, as the loss is only measured on Λ, whereas the supervised loss is measured on the fully-sampled k-space. b) For both supervised and self-supervised training, the outputs of the networks is evaluated on the fully-sampled k-space loss, for every 10th epoch. Using a similar metric, the two approaches show similar trends over epochs, with the supervised training achieving a slightly lower loss than the self-supervised approach.

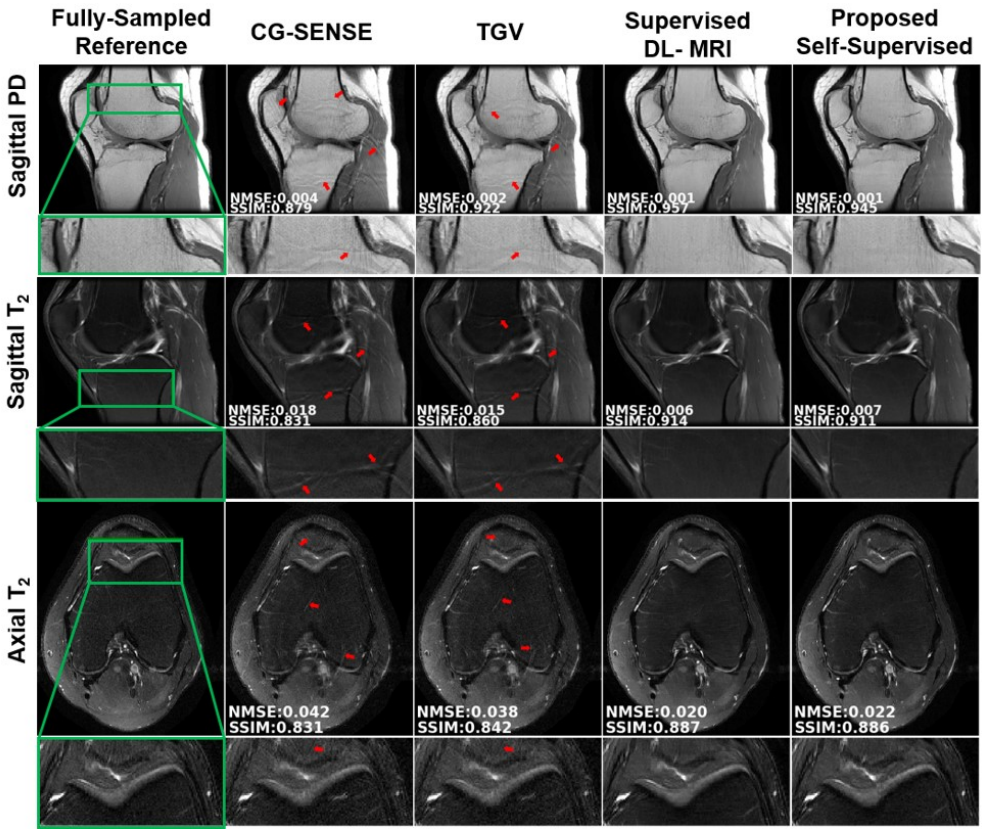

Figure A.6: Representative reconstructed test slices from fastMRI sagittal PD, sagittal T_2 and axial T_2 knee sequences for retrospective equispaced undersampling R = 4. In all three sequences, CG-SENSE and TGV suffer from visible residual artifacts, marked by red arrows. Both proposed self-supervised and fully-supervised DL-MRI approaches successfully remove these residual artifacts, while showing similar quantitative and qualitative performance. Note the former does not require any fully-sampled data for training unlike the latter supervised approach.

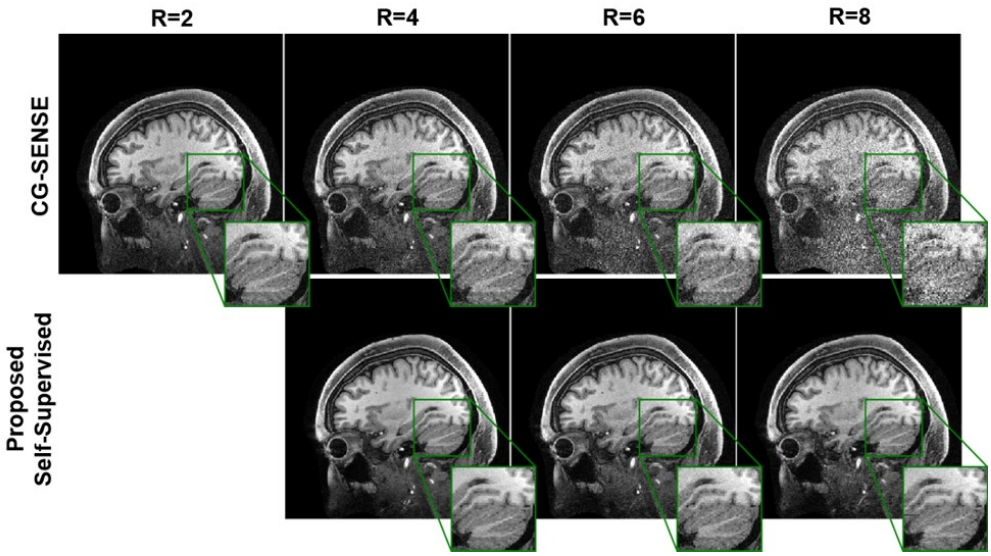

Figure A.7: Reconstruction results for CG-SENSE and proposed self-supervised approach for brain MRI. CG-SENSE suffers from significant noise amplification at high acceleration rates. Proposed self-supervised approach achieves high-quality reconstruction at high acceleration rates, and achieves a lower noise amplification at rate 8 compared to CG-SENSE at acquisition acceleration rate 2.

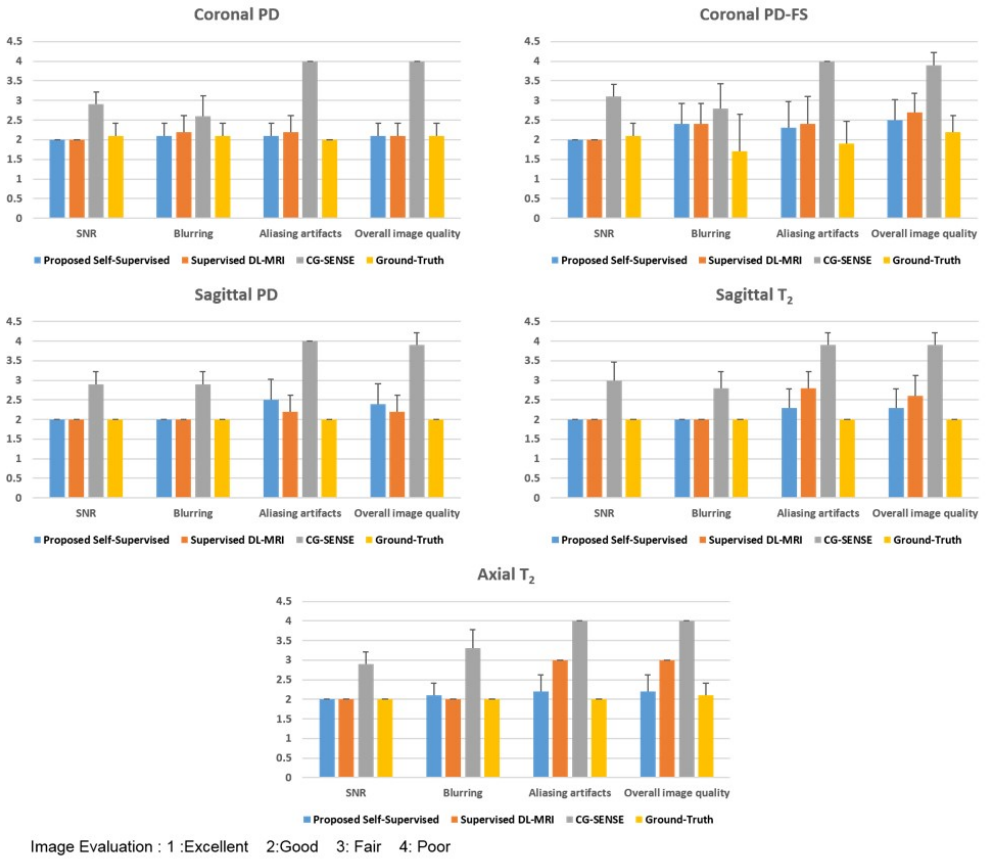

Figure A.8: Average reader scores for all knee sequences for the proposed self-supervised training, supervised training with image domain loss and CG-SENSE. Both supervised and self-supervised DL-MRI approaches get comparable scores to the reference image in terms of SNR, blurring, aliasing artifacts and overall image quality. There was no statistical difference between reference and DL-MRI approaches in terms of SNR and blurring in the knee sequences in general, except for blurring between reference and DL-MRI approaches in coronal PD-FS. In terms of aliasing artifacts and overall image quality, there were no statistical difference between reference and the two DL-MRI approaches for coronal PD, coronal PD-FS and sagittal PD sequences. However, for sagittal T_2 sequence, supervised DL-MRI was ranked statistically worse than the reference, while for axial T_2, it was ranked lower than both the reference and self-supervised DL-MRI. Thus, in general, both DL-MRI approaches performed well, but the self-supervised approach was slightly more favored by the reader, who was blinded to the reconstruction method. CG-SENSE was significantly outperformed by both DL-MRI approaches, while showing statistically significant differences to the reference and both DL-MRI approaches for all knee sequences, except in blurring for coronal PD and PD-FS sequences. Finally, we also note that the supervised training with k-space loss (Figure 10) outperforms supervised training with image domain loss in terms of reader scores for axial T_2, coronal PD-FS and sagittal T_2 sequences.

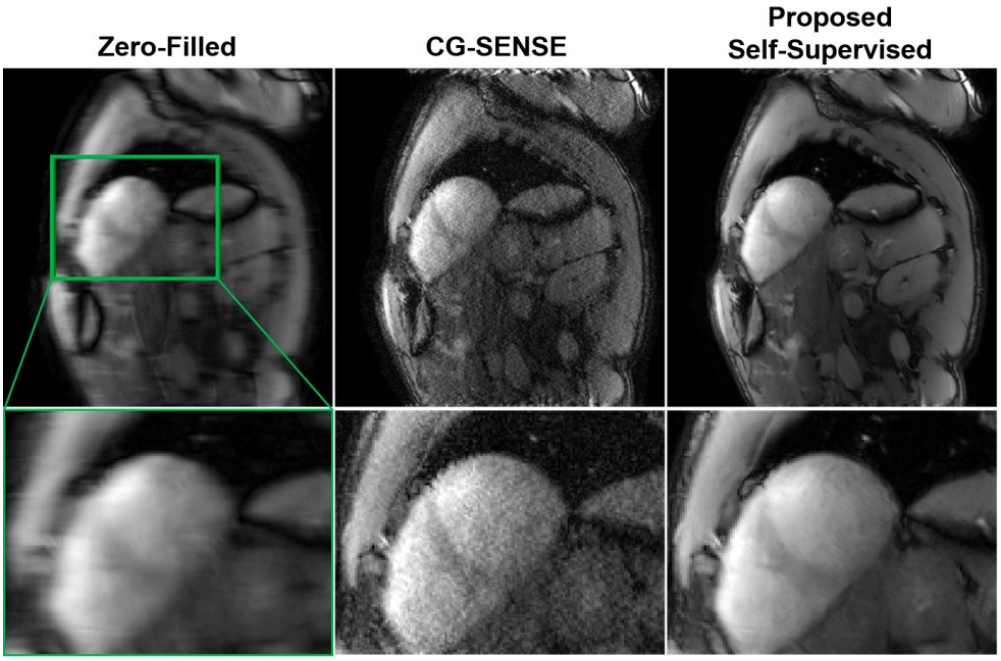

Figure A.9: Reconstructed images from an 8-fold accelerated snapshot cardiac MRI data with 1.3×1.3 mm^2 in-plane resolution, acquired using a transient bSSFP sequence. These type of acquisitions are commonly used in cardiac parametric mapping, where the image data for one contrast weighting need to be acquired within the diastolic quiescence of one heartbeat. A fully-sampled acquisition at this higher resolution would take >700 ms, which is impossible to fit in the diastolic quiescence of a single heart-beat. Training data was acquired on 14 subjects, and testing was performed on a different subject, using the approach described in the manuscript. The proposed self-supervised approach achieves high-quality reconstruction, outperforming CG-SENSE, which suffers from residual artifacts and high noise.

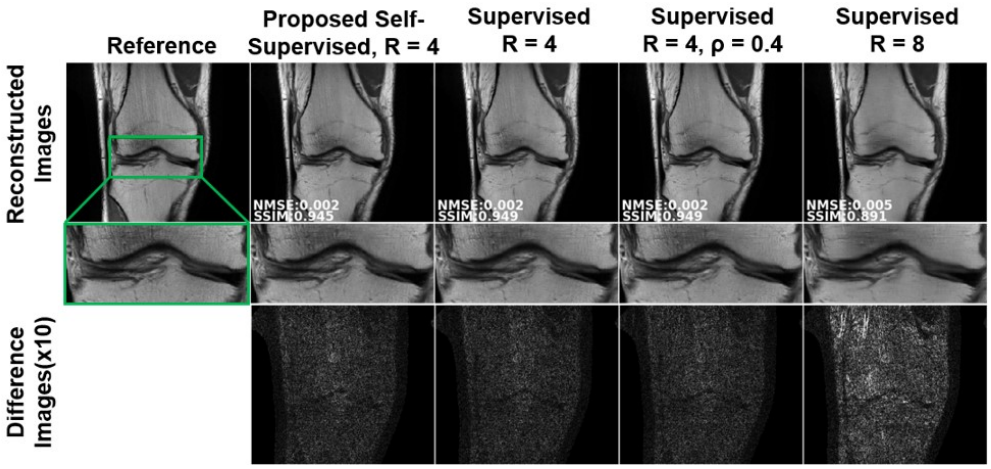

Figure A.10: Reconstruction results for proposed self-supervised training at R = 4, supervised training at R = 4, R = 4 with $\rho = 0.4$, and R = 8. The amount of data used for self-supervised/supervised training at R = 4 (24 ACS lines) with $\rho = 0.4$ is 21120 k-space points, which is approximately equivalent to training the network with an equispaced undersampling pattern of R = 8 (24 ACS lines) with 21440 k-space points. The results show that supervised training at R = 4 with $\rho = 0.4$ is visibly similar with supervised and proposed self-supervised training at R = 4, and outperforms supervised training at R = 8. These results are visibly highlighted in difference images, which show supervised training at R = 8 suffering from residual artifacts, while other approaches show similar performance. Quantitative metrics on test dataset aligns with these qualitative assessments. The median and interquartile range for SSIM across test dataset were 0.961 [0.947, 0.972], 0.966 [0.956, 0.977], 0.966 [0.954, 0.976], 0.929 [0.908, 0.950], and NMSE were 0.002 [0.001, 0.002], 0.001 [0.001, 0.002], 0.002 [0.001, 0.002], 0.004 [0.003, 0.005] for proposed self-supervised at R = 4, supervised at R = 4, supervised at R = 4 with $\rho = 0.4$, and supervised at R = 8, respectively.

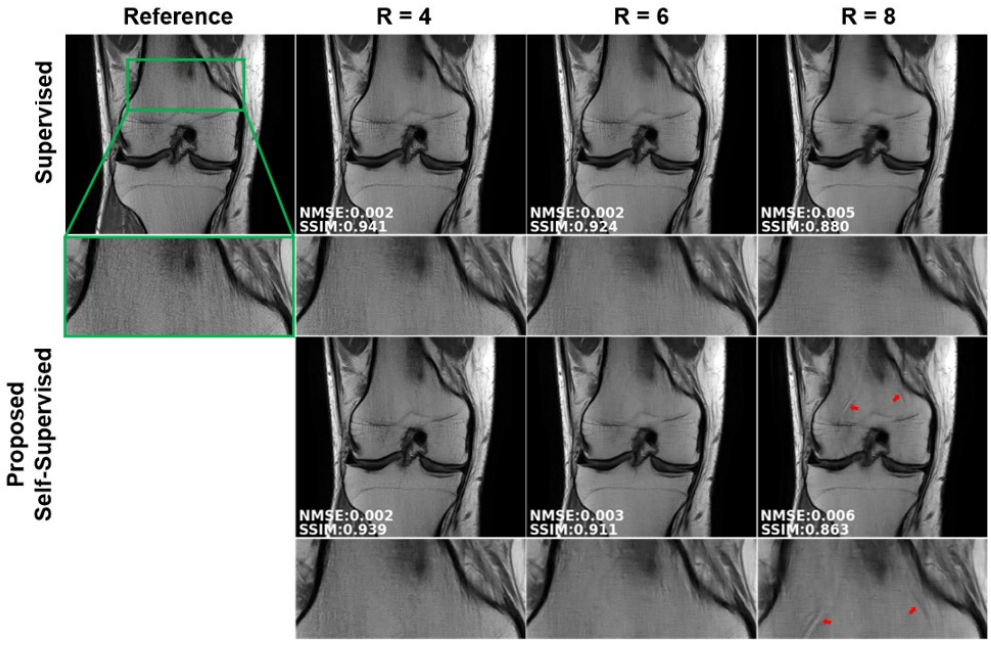

Figure A.11: Reconstruction results for the coronal PD-weighted dataset at acceleration rates of 4, 6 and 8. For R = 4 and 6, the proposed self-supervised approach performs similarly with the supervised approach. However, at R = 8, the image quality degrades for both methods with more pronounced blurring, while the self-supervised approach further suffers from visible residual aliasing artifacts.

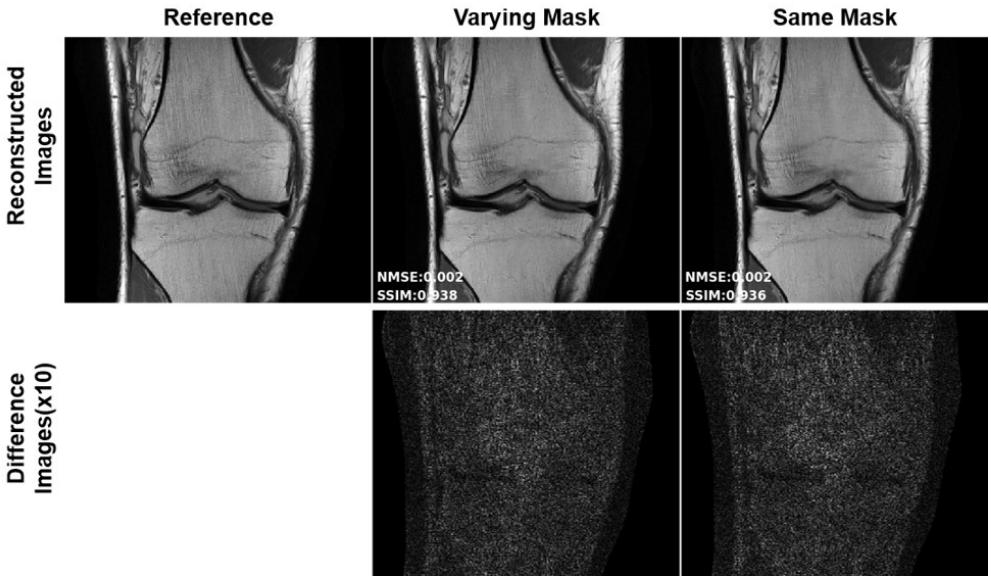

Figure A.12: Reconstruction results for the proposed self-supervised approach when using same or varying sets, Θ and Λ, across different training slices. The two approaches perform similarly with the varying mask approach showing slight improvement. The median and interquartile ranges for SSIM across the test dataset were 0.959 [0.945, 0.970], 0.960 [0.947, 0.0971], and for NMSEs were 0.002 [0.001, 0.002], 0.002 [0.001, 0.002] for varying mask and same mask scenarios, respectively.

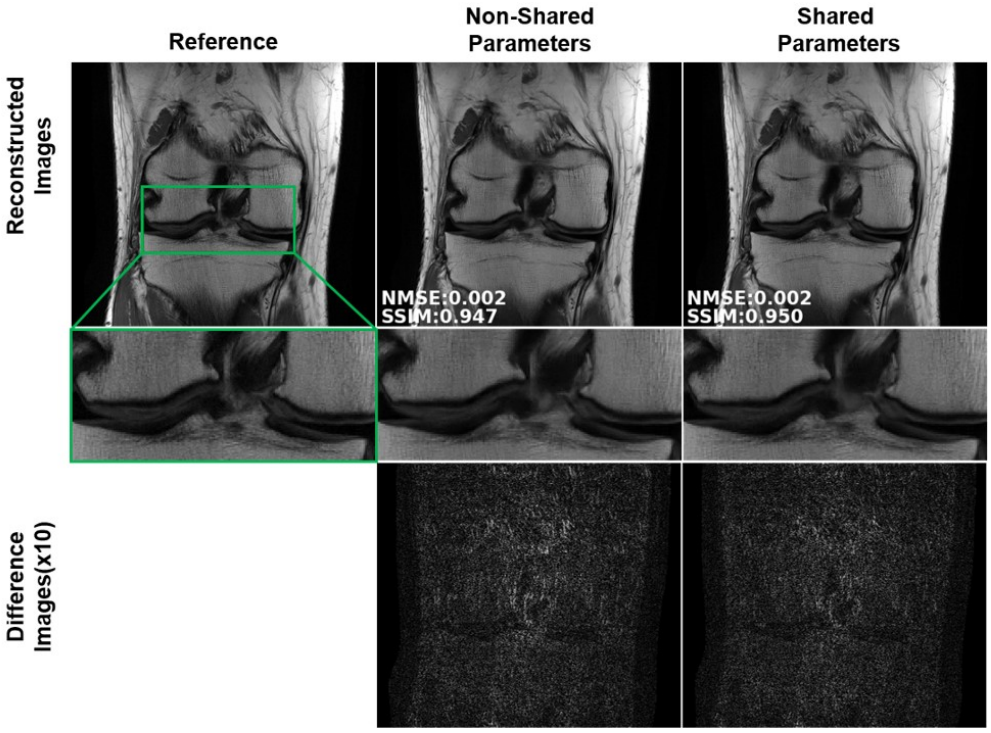

Figure A.13: Reconstruction results for supervised training when using shared and distinct (non-shared) parameters across the unrolled network. The two approaches perform similarly both visually and quantitatively. The interquartile range of SSIM values across the test dataset were 0.966 [0.956, 0.0977], 0.964 [0.952, 0.974], and NMSE values were 0.001 [0.001, 0.002], 0.001 [0.001, 0.002] for shared and non-shared scenarios, respectively. Note that the same training database was used for the two approaches. The non-shared approach has 10 times as many trainable parameters, and its generalization performance may benefit from a larger training database. This was not studied as it is not the focus of our study.

Knee Sequence	TR	TE	TF	Matrix Size	In-Plane Resolution	Slice Thickness	Scan Time
Coronal-PD	2750 ms	27 ms	4	320×368	0.49×0.44 mm^2	3 mm	17 min
Coronal-PDFS	2870 ms	33 ms	4	320×368	0.49×0.44 mm^2	3 mm	18 min
Sagittal PD	2800 ms	27 ms	4	384×304	0.46×0.36 mm^2	3 mm	14 min
Sagittal T_2	4300 ms	50 ms	11	320×256	0.55×0.44 mm^2	3 mm	18 min
Axial T_2	4000 ms	65 ms	9	320×256	0.55×0.44 mm^2	3 mm	17 min

Table A.1: Imaging parameters for the knee datasets.

	Disjoint sets ($\Theta = \Omega \backslash \Lambda$)	50 % Overlap of Θ and Λ	100 % Overlap of Θ and Λ ($\Omega = \Theta$)	Identical sets ($\Omega = \Theta = \Lambda$)
SSIM	0.961 [0.947, 0.972]	0.958 [0.947, 0.970]	0.796 [0.753, 0.862]	0.802 [0.762, 0.867]
NMSE	0.002 [0.001, 0.002]	0.002 [0.001, 0.002]	0.009 [0.006, 0.012]	0.009 [0.006, 0.011]

Table A.2: Median and interquartile range (25^{th} -75^{th} percentile) of the quantitative evaluation of SSIM and NMSE values for different overlap scenarios between Λ and Θ when $\rho = 0.4$. Overlap %, defined as $|\Lambda \cap \Theta|/|\Lambda|$ refers to the amount of data in the loss mask Λ that was also included in the training mask Θ. Performance of the self-supervised training degrades as the amount of overlap increases.

A.2 Supporting Information for Chapter 4

See **Figures A.14, A.15,A.16, A.17,A.18, A.19**,and **A.20**.

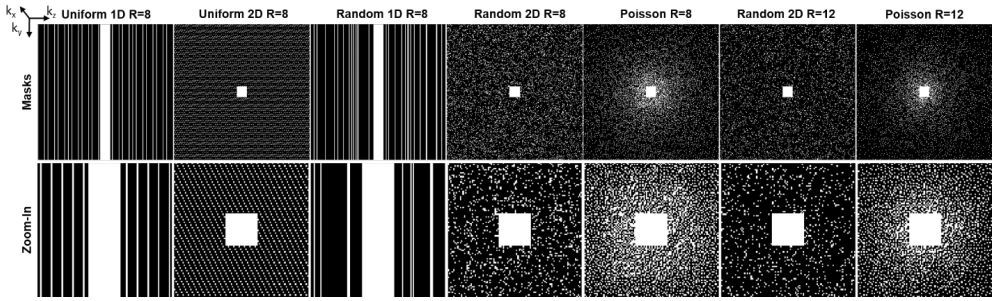

Figure A.14: Undersampling masks used in the study. Note that due to the different size of the ACS data, 1D masks correspond to an effective acceleration rate of 5.2, while the 2D masks yield an effective acceleration rate of 7.7.

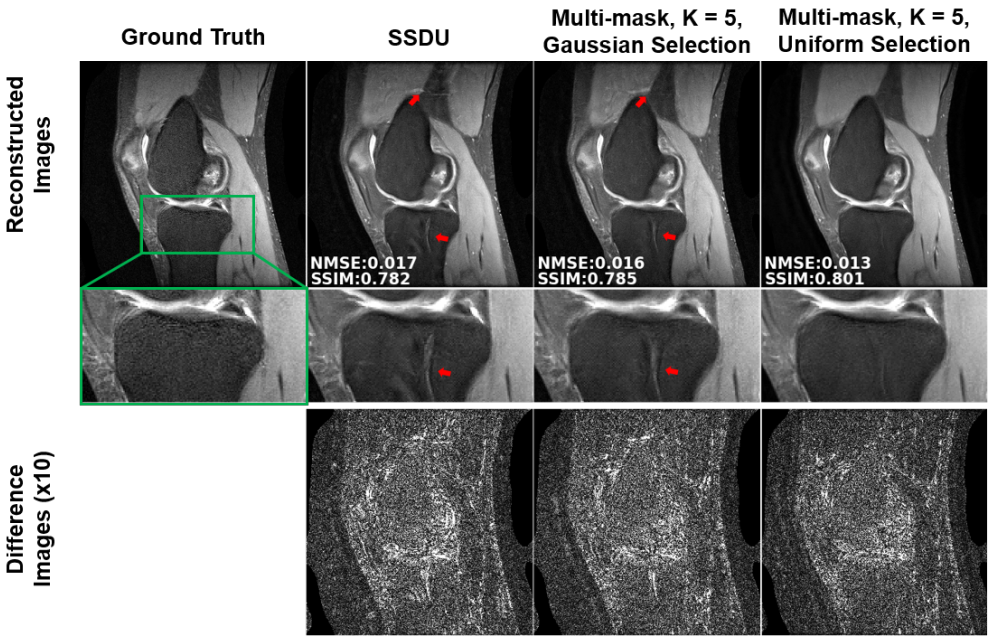

Figure A.15: Reconstruction results from SSDU, and multi-mask SSDU with uniform random selection and variable-density Gaussian selection for K = 5 and $\rho = 0.4$. Multi-mask SSDU with Gaussian random selection fails to remove the artifacts apparent in SSDU, whereas multi-mask SSDU with uniformly random selection significantly suppresses these artifacts. Difference images show that multi-mask SSDU with uniformly random selection shows fewer residual artifacts compared to its multi-mask Gaussian counterpart. The median and interquartile range of SSIM values across the validation dataset were 0.7974 [0.7723, 0.8293], 0.8009 [0.7789, 0.8313], 0.8260 [0.8002, 0.8516], and NMSE values were 0.0166 [0.0142, 0.0202], 0.0159 [0.0139, 0.0191], 0.0135 [0.0119, 0.0157] for SSDU, multi-mask SSDU with Gaussian selection and uniformly random selection, respectively.

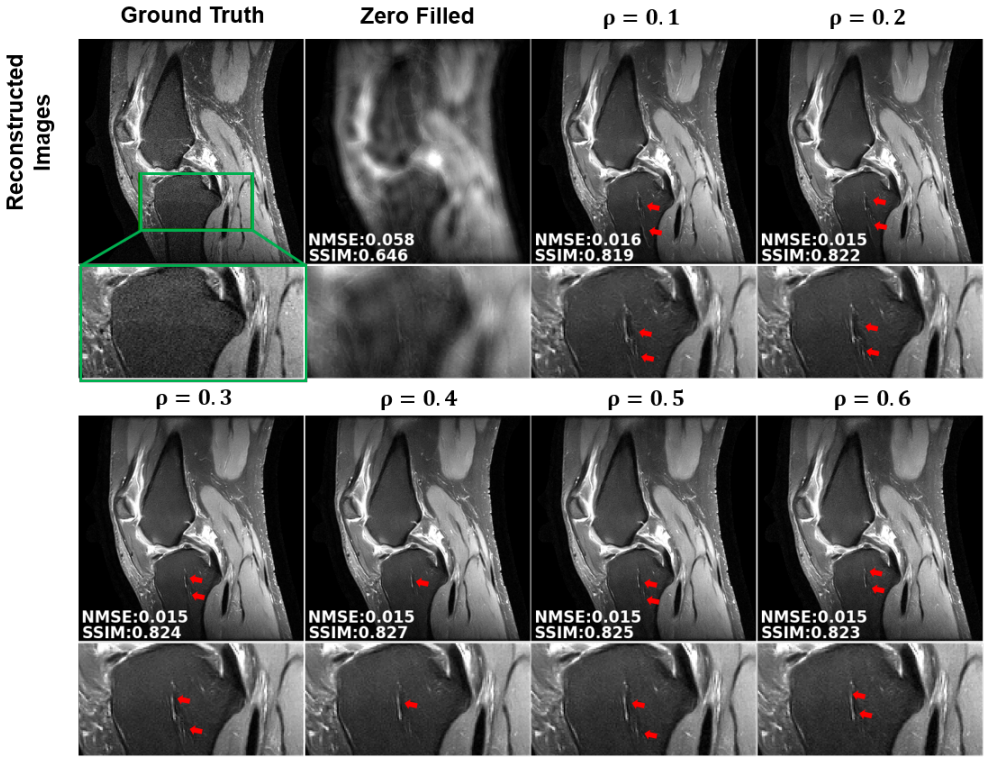

Figure A.16: Reconstruction results from SSDU with uniform random selection of Λ for $\rho \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$. SSDU reconstructions suffers from residual artifacts for low ρ values of 0.1, 0.2 and 0.3. The best reconstruction quality is achieved at ρ = 0.4. Residual artifacts start to reappear after ρ = 0.5, becoming more pronounced as ρ increases. The quantitative assessment from hold-out validation set align with these qualitative assessments. The median and interquartile range of SSIM values were 0.8166 [0.7875, 0.8408], 0.8208 [0.7928, 0.8451], 0.8230 [0.7967, 0.8486], 0.8236 [0.7964, 0.8494], 0.8229 [0.7960, 0.8499], 0.8192 [0.7937, 0.8473], and NMSE values were 0.0149 [0.0136, 0.0175], 0.0143 [0.0128, 0.0167], 0.0141 [0.0123, 0.0163], 0.0140 [0.0122, 0.0161], 0.0145 [0.0125, 0.0168], 0.0145 [0.0127, 0.0169] using uniformly random selection for $\rho \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$, respectively.

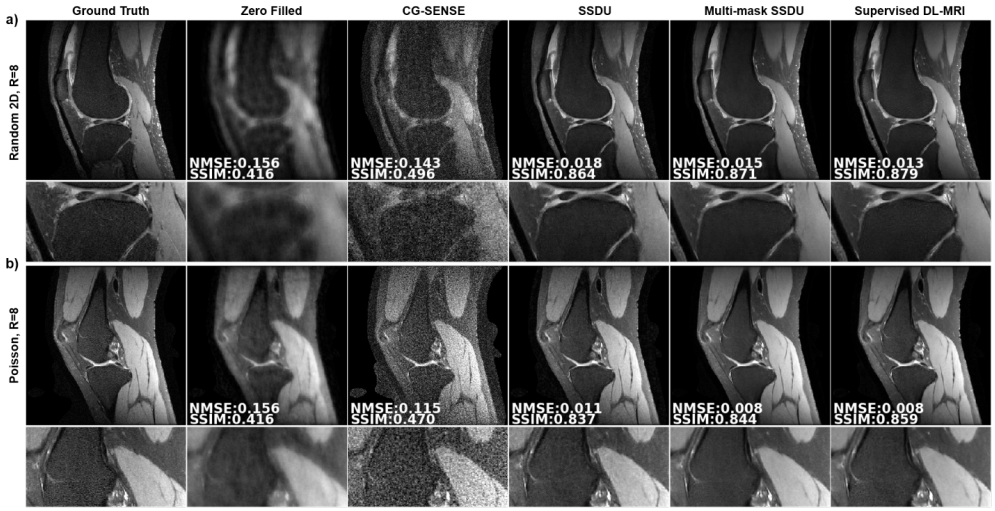

Figure A.17: Reconstruction results using 2D a) random and b) Poisson undersampling masks at R = 8. CG-SENSE suffers from noise and incoherent residual artifacts. All DL approaches achieve artifact-free and improved reconstruction quality.

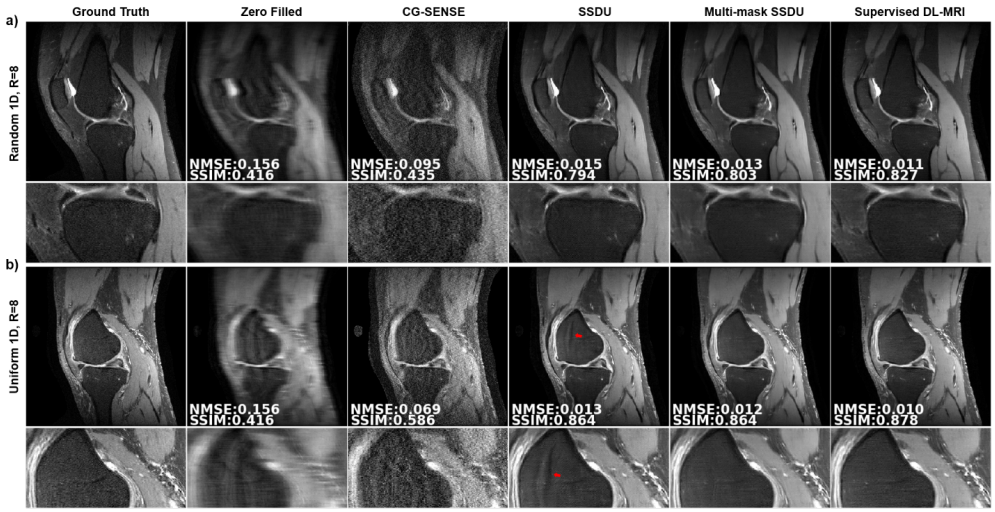

Figure A.18: Reconstruction results at R = 8 using 1D a) random and b) uniform undersampling masks. CG-SENSE suffers from noise and residual artifacts for both of these undersampling masks. All DL reconstructions achieve artifact-free reconstruction with random undersampling. In uniform undersampling, SSDU suffers from residual artifacts shown with red arrows, whereas multi-mask SSDU improves upon SSDU and achieve similar reconstruction quality with supervised DL-MRI.

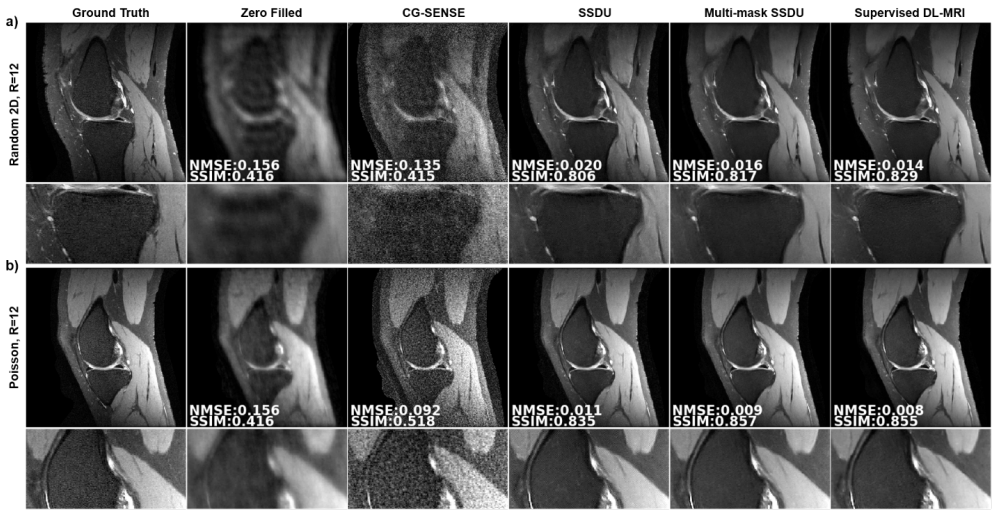

Figure A.19: Reconstruction results at R = 12 using 2D a) random and b) Poisson undersampling masks. CG-SENSE suffers from noise and incoherent residual artifacts. All DL approaches achieve artifact-free and improved reconstruction quality.

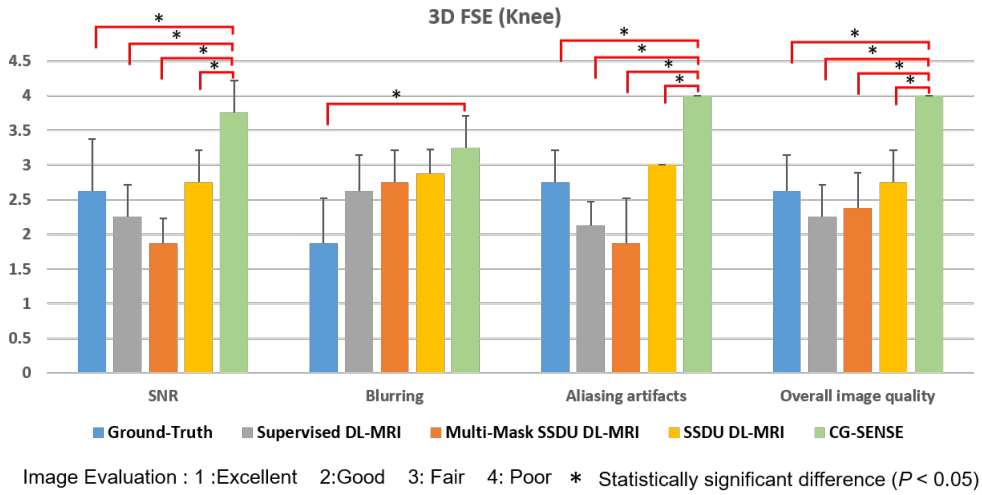

Figure A.20: The image reading results from the clinical reader study for the 3D FSE knee dataset. CG-SENSE was consistently rated lowest in terms of all evaluation criteria. CG-SENSE was significantly worse than all other methods and ground truth in terms of SNR, aliasing artifacts and overall image quality. For blurring, it was only statistically different than the ground truth.

A.3 Supporting Information for Chapter 5

See **Figures A.21, A.22, A.23, A.24, A.25, A.26**, and **Tables A.3 & A.4**.

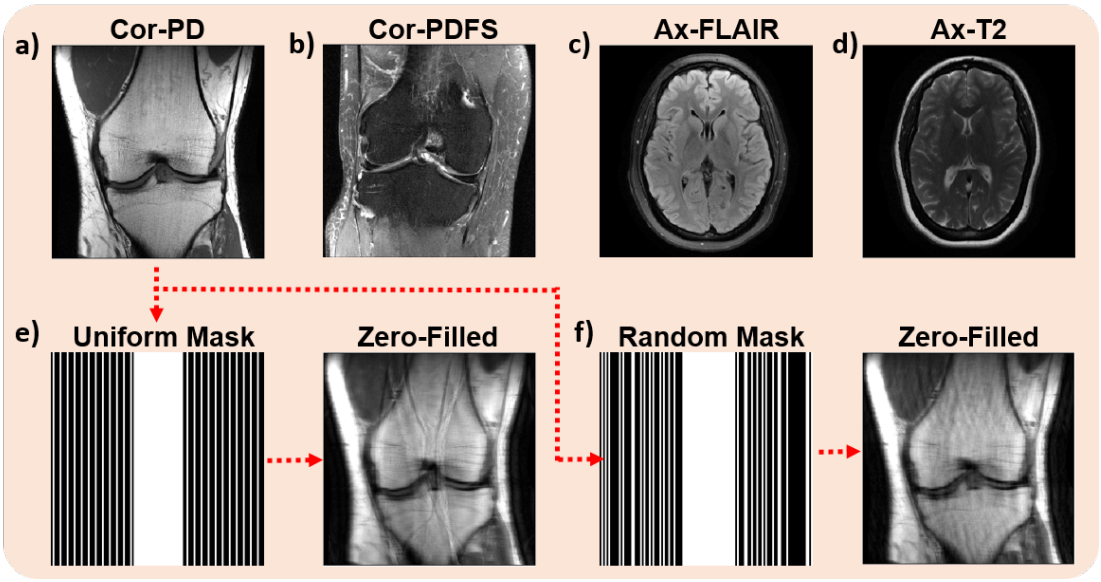

Figure A.21: Different contrast weightings and anatomies used in this study: a) Cor-PD, b) Cor-PDFS, c) Ax-FLAIR, d) Ax-T_2, as well as undersampling patterns: e) Uniform, f) Random mask. Zero-filled images generated by uniform and random undersampling masks have coherent and incoherent aliasing artifacts, respectively. Coherent aliasing artifacts are generally harder to remove than incoherent artifacts.

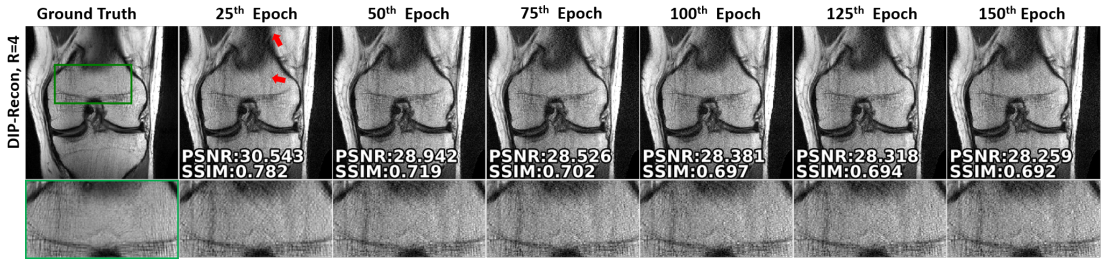

Figure A.22: Cor-PD Knee MRI reconstruction results across different epochs for DIP-Recon using uniform undersampling at R = 4. At the 25th epoch, the reconstruction suffers from artifacts, with the zoom-in area showing texture that does not resemble the ground truth. With more epochs, this aspect of the reconstruction improves, but the reconstruction starts to suffer from noise amplification as the number of epochs increases. Hence, the 50th epoch was used in the experiments.

Figure A.23: a) and b) show reconstruction results corresponding to the loss curves in Figure 5.2a and b, respectively.

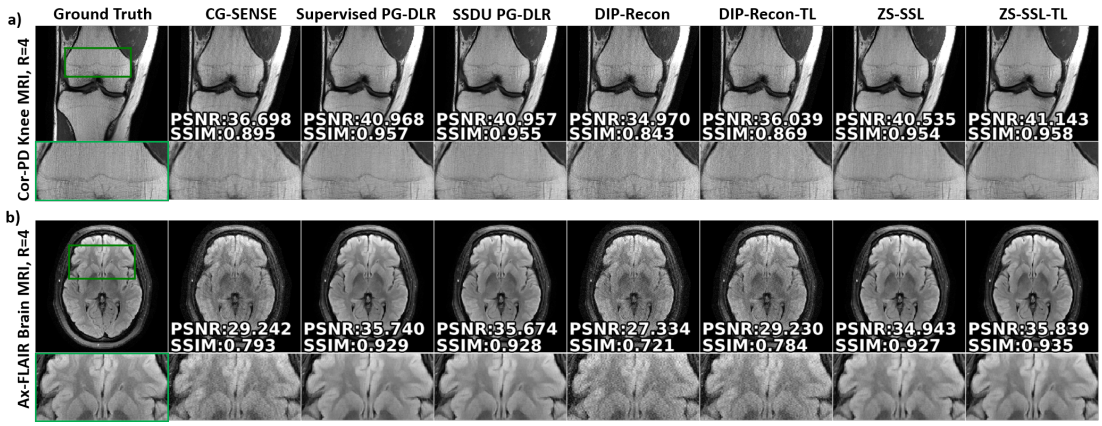

Figure A.24: Reconstruction results from R = 4 with random undersampling on representative test slices from a) Cor-PD knee MRI and b) Ax-FLAIR brain MRI. CG-SENSE, DIP-Recon and DIP-Recon-TL suffer from noise amplification. Supervised PG-DLR, SSDU PG-DLR, ZS-SSL and ZS-SSL-TL all show artifact-free reconstruction quality, with similar quantitative metrics.

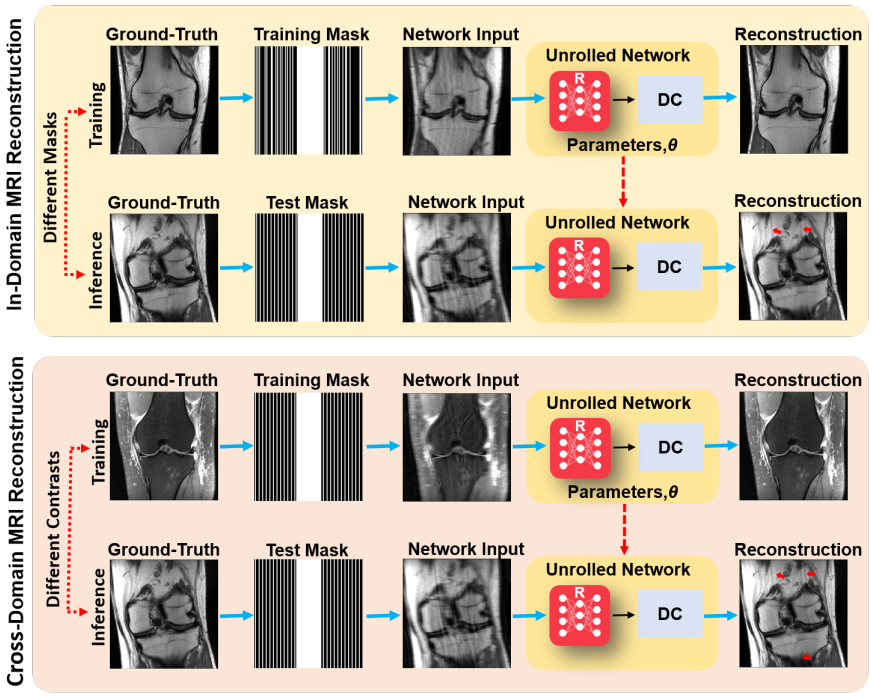

Figure A.25: Test datasets may differ from the training datasets in terms of sampling pattern, SNR, contrast and anatomy. Such differences lead to sub-optimal reconstructions in the test datasets, raising robustness and generalizability concerns for translation of trained MRI reconstruction models to clinical practice.

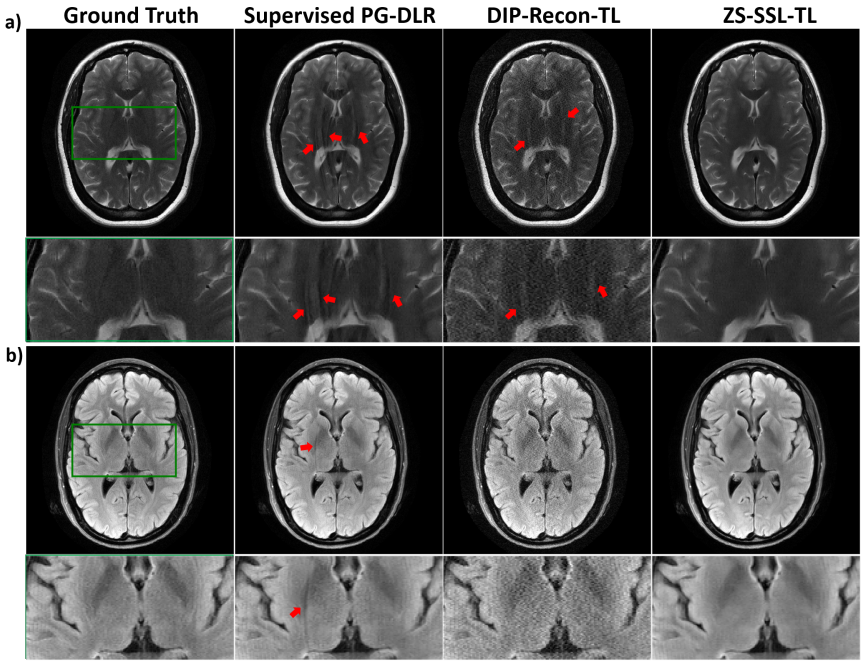

Figure A.26: a) Using pre-trained Ax-Flair for Ax-T_2 reconstruction. b) Using a pre-trained Cor-PD (knee MRI) for Ax-Flair (brain MRI) reconstructions. Supervised PG-DLR fails to generalize when contrast, SNR and anatomy changes, with residual artifacts (red arrows). DIP-Recon-TL also shows artifacts. ZS-SSL-TL successfully removes noise and artifacts.

	CG-SENSE	DIP-Recon	DIP-Recon-TL	ZS-SSL	ZS-SSL-TL
Average Time (sec)	<<1	75	75	640	85

Table A.3: Average reconstruction times for single-instance reconstruction methods. The computation times were measured on the machines equipped with 4 NVIDIA V100 GPUs (each with 32 GB memory). While CG-SENSE and DIP methods have lower computational times, their reconstruction quality is severely degraded hindering clinical usage. ZS-SSL-TL ($K = 10$) provides an 8-fold faster convergence time compared to ZS-SSL ($K = 10$). We note that ZS-SSL methods reconstruction times may further be reduced by means of more compact architectures. Additionally, the increased computational times may be tolerable within the workflow, for instance in clinical settings where image readings are done the next day, or scans such as high-resolution functional or diffusion MRI, where it is challenging to have high-quality high-resolution data, while post-reconstruction analyses readily take hours to days.

	Metrics	Supervised PG-DLR	DIP-Recon-TL	ZS-SSL-TL
Figure 5.4: Banding Artifacts	SSIM	**0.873**	0.530	0.861
	PSNR	**36.365**	26.924	36.121
Figure 5.5a: In-Domain Transfer - Different Mask	SSIM	0.949	0.836	**0.951**
	PSNR	39.167	34.093	**40.088**
Figure 5.5b: In-Domain Transfer - Different Rates	SSIM	0.937	0.792	**0.940**
	PSNR	38.262	32.658	**38.301**
Figure 5.6a: Cross-Domain Transfer - Knee-Different Contrast	SSIM	0.931	0.859	**0.949**
	PSNR	37.566	34.855	**39.855**
Figure 5.6b: Anatomy Change - Trained on Brain & Tested on Knee	SSIM	0.936	0.890	**0.957**
	PSNR	37.494	35.458	**40.407**
Figure A.26a: Cross-Domain Transfer - Brain-Different Contrast	SSIM	0.929	0.834	**0.950**
	PSNR	35.578	32.655	**38.767**
Figure A.26b: Anatomy Change - Trained on Knee & Tested on Brain	SSIM	0.929	0.806	**0.936**
	PSNR	36.242	30.849	**37.134**

Table A.4: Average PSNR and SSIM values on 30 test slices for the experiments associated with Figures 4-6 (in the main text) and 12. We note that ZS-SSL-TL successfully fine-tunes the network for each new dataset/instance regardless of the starting pre-trained model. Interestingly, there are cases where out-of-domain transfer has slightly higher metrics than in-domain transfer. In these cases, since the metrics are already high, the slight quantitative differences do not affect the overall quality. However, the main difference in these cases is in the convergence/stopping time, where in-domain transfer is typically converges/stops in $\sim$2-fold fewer iterations than out-of-domain transfer.