Preprints.org

Review

A Review of Optimization-Based Deep Learning Models for MRI Reconstruction

Wanyu Bian [*] and Yokhesh Krishnasamy Tamilselvam

Article

A Review of Optimization-Based Deep Learning Models for MRI Reconstruction

Wanyu Bian [1,2,*] and **Yokhesh Krishnasamy Tamilselvam** [3]

[1] Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, Massachusetts, USA; wbian4@mgh.harvard.edu

[2] Harvard Medical School, Boston, Massachusetts, USA

[3] University of Western Ontario, London, ON, N6A 5B9, Canada

[*] Correspondence: wbian4@mgh.harvard.edu

Abstract: Magnetic resonance imaging (MRI) is crucial for its superior soft tissue contrast and high spatial resolution. Integrating deep learning algorithms into MRI reconstruction enhances image quality and efficiency, but a comprehensive review of optimization-based deep learning models for MRI reconstruction has been missing. This study fills that gap by examining the latest optimization-based algorithms in deep learning for MRI reconstruction, including gradient descent algorithms, proximal gradient descent algorithms, ADMM, PDHG, and diffusion models combined with gradient descent. Learnable optimization algorithms (LOA) are highlighted for their ability to map optimization processes to structured neural networks, improving model interpretability and performance. The study demonstrates significant advancements in MRI reconstruction through deep learning, with successful clinical applications. These findings provide valuable insights and resources for researchers aiming to advance medical imaging using innovative deep learning techniques.

Keywords: optimization algorithms; MRI reconstruction; deep learning

1. Introduction

Magnetic Resonance Imaging (MRI) is a crucial medical imaging technology that is non-invasive and non-ionizing, providing precise in-vivo images of tissues vital for disease diagnosis and medical research. As an indispensable instrument in both diagnostic medicine and clinical studies, MRI plays an essential role[1,2].

Although MRI offers superior diagnostic capabilities, its lengthy imaging times, compared to other modalities, restrict patient throughput. This challenge has spurred innovations aimed at speeding up the MRI process, with the shared objective of significantly reducing scan duration while maintaining image quality[3,4]. Accelerating data acquisition during MRI scans is a major focus within the MRI and clinical application community. Typically, scanning one sequence of MR images can take at least 30 minutes, depending on the body part being scanned, which is considerably longer than most other imaging techniques. However, certain groups such as infants, elderly individuals, and patients with serious diseases who cannot control their body movements, may find it difficult to remain still for the duration of the scan. Prolonged scanning can lead to patient discomfort and may introduce motion artifacts that compromise the quality of the MR images, reducing diagnostic accuracy. Consequently, reducing MRI scan times is crucial for enhancing image quality and patient experience.

MRI scan time is largely dependent on the number of phase encoding steps in the frequency domain (k-space), with common methods to accelerate the process involving the reduction of these steps by skipping phase encoding lines and sampling only partial k-space data. However, this approach can lead to aliasing artifacts due to under-sampling, violating the Nyquist criterion[5]. MRI reconstruction involves creating clear MR images from undersampled k-space data, which is then used for diagnostic and clinical purposes. Compressed Sensing (CS)[6] MRI reconstruction and parallel imaging[3,7,8] are effective techniques that address this inverse problem, speeding up MRI scans and reducing artifacts. By allowing for under-sampling and heaving the ability to reconstruct high-quality MRI images from under-sampled data, CS significantly reduces scan time while offering images that are often comparable to those obtained from fully sampled data.

Deep learning has seen extensive application in image processing tasks [9–13] because of its ability to efficiently manage multi-scale data and learn hierarchical structures effectively, both of which are essential for precise image reconstruction and enhancement. Convolution neural network (CNN) also extensively utilized in MRI reconstruction due to its proficiency in handling complex patterns and noise inherent in MRI data[14–23]. By learning from large datasets, deep learning algorithms can improve the accuracy and speed of reconstructing high-quality images, thus significantly enhancing the diagnostic capabilities of MRI technology.

Optimization-based network unrolling algorithms was introduced in recent years [23–33], each inspired by classical optimization techniques tailored to address the challenges inherent in MRI data and end-to-end deep models. We introduced a class of learnable optimization algorithms (LOA) that can increase the interpretability of the deep models that benefit from the MR physical information, and also enhance the training efficiency. Specifically, this paper discusses the Gradient Descent and Proximal Gradient Descent Algorithm Inspired Networks, Variational Networks and Iterative Shrinkage-Thresholding Algorithm (ISTA) Network, and Alternating Direction Method of Multipliers (ADMM) inspired network, which leverage iterative methods to refine MRI reconstructions, reducing artifacts and improving image clarity. Diffusion models integration, specifically the score-based diffusion Model and Domain-conditioned Diffusion Modeling represent a novel approach by combining the strengths of deep learning with diffusion processes to address the under-sampling issue more robustly. These methods collectively represent a robust framework for improving the speed and accuracy of MRI scans, advancing both the theory and application of machine learning in medical imaging. The integration of these sophisticated deep learning techniques with traditional optimization algorithms provides a dual advantage of enhancing diagnostic capabilities while significantly reducing scan times. By reviewing how these optimization methods are used in conjunction with novel deep learning techniques, we aim to shed light on the capabilities of state-of-the-art MRI reconstruction techniques and the scope for future work in this direction.

This paper is organized in the following structure: Section 1 introduces the importance of MRI reconstruction and LOA methods. Section 2 presents the compressed sensing (CS)-based MRI reconstruction model. Section 3 provides a detailed overview of various optimization algorithms utilizing deep learning techniques. Section 4 discusses the current issues and limitations of learnable optimization models. Section 5 concludes the paper by summarizing the key findings and implications of the study.

2. MRI Reconstruction Model

Parallel imaging is a k-space method that utilize coil-by-coil auto-calibration, such as GRAPPA [34] and SPIRiT [35]. CS-based methods were applied on image domain such as SENSE [3], which depend on the accurate knowledge of coil sensitivity maps for optimization. The formulation for the MRI reconstruction problem in CS-based parallel imaging is described by a regularized variational model as follows:

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{f}\|_2^2 + \mu \mathcal{R}(\mathbf{x}), \tag{1}$$

where $\mathbf{x} \in \mathbb{C}^n$ is the MR image to be reconstructed, consisting of n pixels, and $\mathbf{f} \in \mathbb{C}^m$ denotes the corresponding undersampled measurement data in k-space. The data fidelity term $\frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{f}\|_2^2$ enforces physical consistency between the reconstructed image $\mathbf{x}$ and the partial data $\mathbf{f}$ measured in the k-space. The regularization operator $\mathcal{R} : \mathbb{C}^n \to \mathbb{R}$ emphasizes the sparsity of the MRI data or low rankness constraints. The regularizer provides prior information to avoid overfitting of the data fidelity term. The weight parameter $\mu > 0$ balances the data fidelity term and regularization term. The measurement data is typically expressed as $\mathbf{f} = \mathbf{A}\mathbf{x} + \varepsilon$ with $\varepsilon \in \mathbb{C}^m$ representing the noise encountered during acquisition. The forward measurement encoding matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$ utilized in parallel imaging is defined by:

$$\mathbf{A} := \mathcal{P}_\Omega \mathcal{F} \mathcal{S}, \tag{2}$$

where $\mathcal{S} := [\mathcal{S}_1, ..., \mathcal{S}_c]$ refers to the sensitivity maps of c different coils, $\mathcal{F} \in \mathbb{C}^{n \times n}$ represents the 2D discrete Fourier transform, and $\mathcal{P}_\Omega \in \mathbb{N}^{m \times n}(m \ll n)$ is the binary undersampling mask that captures m sampled data points according to the undersampling pattern Ω. The Figure 1 shows the image reconstruction diagram.
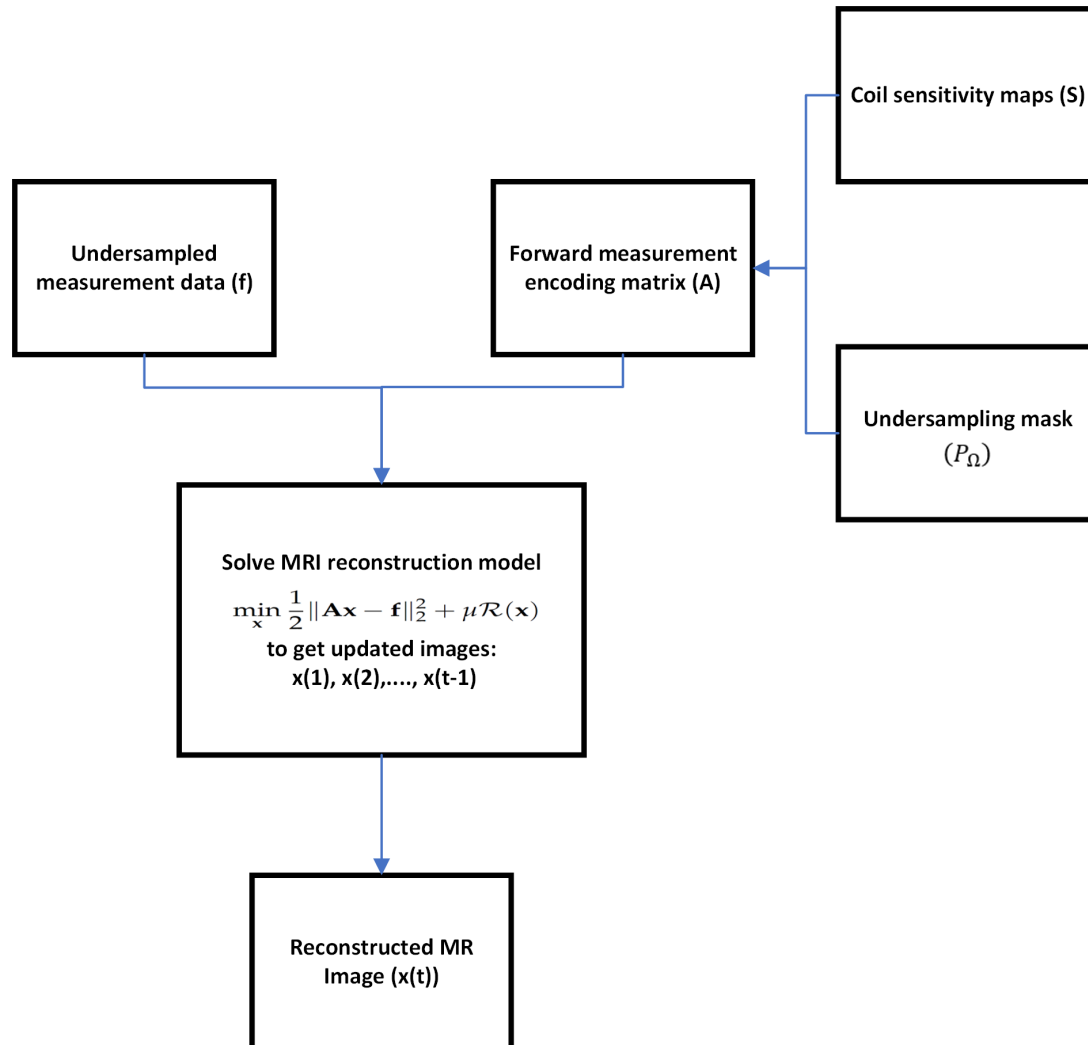

Figure 1. Demonstration of MRI reconstruction process.

3. Optimization-Based Network Unrolling Algorithms for MRI Reconstruction

Deep learning-based model leverages large dataset and further explores the potential improvement of reconstruction performance compared to traditional methods and has successful applications in clinic field[14–20,36,37]. Most existing deep learning-based methods rendering end-to-end neural networks mapping from the partial k-space data to the reconstructed images [38–42]. To improve the interpretability of the relation between the topology of the deep model and reconstruction results, a new emerging class of deep learning-based methods known as *learnable optimization algorithms* (LOA) have attracted much attention e.g. [23–33,43,44]. LOA was proposed to map existing optimization algorithms to structured networks where each phase of the networks corresponds to one iteration of an optimization algorithm.

The architectures of these networks are modeled after iterative optimization algorithms. They retain the data fidelity term, which describes image formation based on well-established physical principles that are already known and do not need to be relearned. Instead of using manually designed and overly simplified regularization as in classical reconstruction methods, these networks employ

deep neural networks for regularization. Typically, these reconstruction networks consist of a few phases, each mimicking one iteration of a traditional optimization-based reconstruction algorithm. The manually designed regularization terms in classical methods are replaced by layers of CNNs, whose parameters are learned during offline training.

For instance, ADMM-Net [45], ISTA-Net$^+$ [46], and cascade network [47] are applied on single-coil MRI reconstruction, where the encoding matrix is reduced to $\mathbf{A} = \mathcal{P}_\Omega \mathcal{F}$ as the sensitivity map $\mathcal{S}$ is identity. Variational network (VN)[14] introduced gradient descent method with a given (pre-calculated) sensitivities $\mathcal{S}$. MoDL [48] proposed a recursive network by unrolling the conjugate gradient algorithm using a weight sharing strategy. Blind-PMRI-Net [49] designed three network blocks to alternatively update multi-channel images, sensitivity maps, and the reconstructed MR image using an iterative algorithm based on half-quadratic splitting. VS-Net [50] derived a variable splitting optimization method. However, existing methods still face the lack of accurate coil sensitivity maps and proper regularization in the parallel imaging problem. Alder et al.[51] proposed a reconstruction network that unrolled primal-dual algorithm where the proximal operator is learnable.

3.1. Gradient Descent Algorithm Inspired Network

3.1.1. Variational Network

Variational Network (VN) solves the model (1) by using gradient descent:

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \alpha^{(t)}(\lambda \mathbf{A}^\top(\mathbf{A}\mathbf{x}^{(t)} - \mathbf{f}) + \nabla \mathcal{R}(\mathbf{x}^{(t)})). \tag{3}$$

This model was applied to multi-coil MRI reconstruction. The regularization term was defined by the Field of Expert model: $\mathcal{R}(\mathbf{x}) = \sum_{i=1}^{N} < \mathcal{H}_i(\mathcal{G}_i \mathbf{x}), \mathbf{1} >$. A convolution neural network $\mathcal{G}_i$ is applied to the MRI data. The function $\mathcal{H}_i$ is defined as nonlinear potential functions which are composed of scalar activation functions. Then take the summation of the inner product of the non-linear term $\mathcal{H}_i(\mathcal{G}_i \mathbf{x})$ and the vector of ones $\mathbf{1}$. The sensitivity maps are pre-calculated and being used in $\mathbf{A}$. The algorithm of VN unrolls the step (3) where the regularizer $\mathcal{R}$ is parameterized by the learnable network $\mathcal{G}_i$ together with nonlinear activation function $\mathcal{H}_i$:

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \alpha^{(t)}(\lambda \mathbf{A}^\top(\mathbf{A}\mathbf{x}^{(t)} - \mathbf{f}) + \sum_{i=1}^{N} (\mathcal{G}_i^{(t)})^\top \mathcal{H}_i(\mathcal{G}_i^{(t)} \mathbf{x}^{(t)})). \tag{4}$$

A simple gradient descent step can be utilized for different designs of the variational model, even for different tasks. The following subsection introduces the customized variational model with joint tasks: MRI reconstruction and multi-contrast synthesis.

3.1.2. Variational Model for Joint Reconstruction and Synthesis

This subsection introduces a provable learnable optimization algorithm[52] for joint MRI reconstruction and synthesis. Consider the partial k-space data $\{\mathbf{f}_1, \mathbf{f}_2\}$ of the source modalities (e.g. T1 and T2) obtained from the measurement domain. The goal is to **reconstruct** the corresponding images $\{\mathbf{x}_1, \mathbf{x}_2\}$ *and* **synthesize** the image $\mathbf{x}_3$ of the missing modality (e.g. FLAIR) without having its k-space data. The following optimization model is designed:

$$\min_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3} \Phi_{\Theta,\gamma}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3) := \frac{1}{2} \sum_{i=1}^{2} \|\mathbf{A}_i \mathbf{x}_i - \mathbf{f}_i\|_2^2 + \frac{1}{3} \sum_{i=1}^{3} \|f_{w_i}(\mathbf{x}_i)\|_{2,1}$$
$$+ \frac{\gamma}{2} \|\zeta_\theta([f_{w_1}(\mathbf{x}_1), f_{w_2}(\mathbf{x}_2)]) - \mathbf{x}_3\|_2^2. \tag{5}$$

The first term $\frac{1}{2}\sum_{i=1}^{2}\|\mathbf{A}_i\mathbf{x}_i - \mathbf{f}_i\|_2^2$ ensures the fidelity of the reconstructed images $\{\mathbf{x}_1, \mathbf{x}_2\}$ to their partial k-space data $\{\mathbf{f}_1, \mathbf{f}_2\}$. The second term $\frac{1}{3}\sum_{i=1}^{3}\|f_{w_i}(\mathbf{x}_i)\|_{2,1}$ regularizes the images using modality-specific feature extraction operators $f_{w_i}, i = 1, 2, 3$. The third term $\frac{\gamma}{2}\|\zeta_\theta([f_{w_1}(\mathbf{x}_1), f_{w_2}(\mathbf{x}_2)]) - \mathbf{x}_3\|_2^2$ enforces consistency between the synthesized image $\mathbf{x}_3$ and the learned correlation relationship from the reconstructed images $\mathbf{x}_1, \mathbf{x}_2$. To synthesize the image $\mathbf{x}_3$ using $\mathbf{x}_1$ and $\mathbf{x}_2$, a feature-fusion operator ζ_θ was employed which learns the mapping from the features $f_{w_1}(\mathbf{x}_1)$ and $f_{w_2}(\mathbf{x}_2)$ to the image $\mathbf{x}_3$.

Denote $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$, the forward Learnable Optimization Algorithm is presented in Algorithm 1. In step 3, the algorithm performs a gradient descent update with a step size found via line search, while keeping the smoothing parameter $\varepsilon > 0$ fixed. In step 4, the reduction of ε ensures the subsequence which met the ε reduction criterion must have an accumulation point that is a Clarke stationary point of the problem.

Algorithm 1 Learnable Descent Algorithm for joint MRI reconstruction and synthesis

---

1: **Input:** Initial estimate $\mathbf{X}_0$, step size range $0 < \eta < 1$, initial smoothing parameter $\varepsilon_0, a, \sigma > 0, t = 0$. Maximum iterations T. Set tolerance $\epsilon_{\text{tol}} > 0$.
2: **for** $t = 0, 1, 2, \ldots, T - 1$ **do**
3:     $\mathbf{X}_{t+1} = \mathbf{X}_t - \alpha_t \nabla \Phi_{\Theta,\gamma}^\varepsilon(\mathbf{X}_t)$, where the step size α_t is determined by a line search such that $\Phi_{\Theta,\gamma}^\varepsilon(\mathbf{X}_{t+1}) - \Phi_{\Theta,\gamma}^\varepsilon(\mathbf{X}_t) \leq -\frac{1}{a}\|\mathbf{X}_{t+1} - \mathbf{X}_t\|^2$ holds.
4:     **if** $\|\nabla\Phi_{\Theta,\gamma}^\varepsilon(\mathbf{X}_{t+1})\| < \sigma\eta\varepsilon$, set $\varepsilon_{t+1} = \eta\varepsilon$; **otherwise**, set $\varepsilon_{t+1} = \varepsilon$.
5:     **if** $\sigma\varepsilon < \epsilon_{\text{tol}}$, **terminate** and go to step 6,
6: **end for** and **output** $\mathbf{X}^{(t)}$.

The Algorithm 1 is a forward MRI reconstruction algorithm. The backward network training algorithm is designed to solve a bilevel optimization problem:

$$\min_\gamma \quad \sum_{i=1}^{\mathcal{M}_{val}} \ell(\Theta(\gamma), \gamma; \mathcal{D}_i^{val}) \quad \text{s.t.} \quad \Theta(\gamma) = \arg\min_\Theta \sum_{i=1}^{\mathcal{M}_{tr}} \ell(\Theta, \gamma; \mathcal{D}_i^{tr}), \tag{6}$$

$$\text{where} \quad \ell(\Theta, \gamma; \mathcal{D}_i) := \frac{\mu}{2}\|\zeta_\theta([f_{w_1}(\mathbf{x}_1^*), f_{w_2}(\mathbf{x}_2^*)]) - \mathbf{x}_3^*\|_2^2$$
$$+ \sum_{j=1}^{3}\left(\frac{1}{2}\|\mathbf{x}_{j,\hat{T}}(\Theta, \gamma; \mathcal{D}_i) - \mathbf{x}_j^*\|_2^2 + (1 - SSIM(\mathbf{x}_{j,\hat{T}}(\Theta, \gamma; \mathcal{D}_i), \mathbf{x}_j^*)))\right). \tag{7}$$

The following Algorithm 2 was proposed for training the model for joint reconstruction and synthesis.

Algorithm 2 Mini-batch alternating direction penalty algorithm

---

1: **Input** Training data $\mathcal{D}^{tr}$, validation data $\mathcal{D}^{val}$, tolerance $\delta_{tol} > 0$. **Initialize** $\Theta, \gamma, \delta, \lambda > 0$ and $\nu_\delta \in (0, 1), \nu_\lambda > 1$.
2: **while** $\delta > \delta_{tol}$ **do**
3:     Sample training batch $\mathcal{B}^{tr} \subset \mathcal{D}^{tr}$ and validation batch $\mathcal{B}^{val} \subset \mathcal{D}^{val}$.
4:     **while** $\|\nabla_\Theta\widetilde{\mathcal{L}}(\Theta, \gamma; \mathcal{B}^{tr}, \mathcal{B}^{val})\|^2 + \|\nabla_\gamma\widetilde{\mathcal{L}}(\Theta, \gamma; \mathcal{B}^{tr}, \mathcal{B}^{val})\|^2 > \delta$ **do**
5:         **for** $k = 1, 2, \ldots, K$ (inner loop) **do**
6:             Update $\Theta \leftarrow \Theta - \rho_\Theta^{(k)}\nabla_\Theta\widetilde{\mathcal{L}}(\Theta, \gamma; \mathcal{B}^{tr}, \mathcal{B}^{val})$
7:         **end for**
8:         Update $\gamma \leftarrow \gamma - \rho_\gamma\nabla_\gamma\widetilde{\mathcal{L}}(\Theta, \gamma; \mathcal{B}^{tr}, \mathcal{B}^{val})$
9:     **end while** and **update** $\delta \leftarrow \nu_\delta\delta, \lambda \leftarrow \nu_\lambda\lambda$.
10: **end while** and **output:** Θ, γ.

3.2. Proximal Gradient Descent Algorithm Inspired Networks

Solving inverse problems using proximal gradient descent has been largely explored and successfully applied in medical imaging reconstruction[23,52–63].

Applying a proximal gradient descent algorithm to approximate a (local) minimizer of (1) is an iterative process. The first step is gradient descent to force data consistency, and the second step applies a proximal operator to obtain the updated image. The following steps iterates the proximal gradient descent algorithm:

$$\mathbf{b}_t = \mathbf{x}_t - \rho_t \mathbf{A}^\top (\mathbf{A}\mathbf{x}_t - \mathbf{f}), \tag{8a}$$

$$\mathbf{x}_{t+1} = \text{prox}_{\rho_t \mathcal{R}(\cdot)}(\mathbf{b}_t), \tag{8b}$$

where $\rho_t > 0$ is the step size and $\text{prox}_{\alpha \mathcal{R}}$ is the proximal operator of $\mathcal{R}$ defined by

$$\text{prox}_{\rho \mathcal{R}}(\mathbf{b}) = \arg\min_{\mathbf{x}} \frac{1}{2\rho} \|\mathbf{x} - \mathbf{b}\|_2^2 + \mathcal{R}(\mathbf{x}). \tag{9}$$

The gradient update step (8a) is straightforward to compute and fully utilizes the relationship between the partial k-space data $\mathbf{f}$ and the image $\mathbf{x}$ to be reconstructed as derived from MRI physics. This step involves implementing the proximal operation for regularization $\mathcal{R}$, which is equivalent to finding the maximum-a-posteriori solution for the Gaussian denoising problem at a noise level $\sqrt{\rho}$ [64,65]. Thus, the proximal operator can be interpreted as a Gaussian denoiser. However, because the proximal operator $\text{prox}_{\rho \mathcal{R}}$ in the objective function (1) does not admit closed form solution, a CNN can be used to substitute $\text{prox}_{\rho \mathcal{R}}$. Constructing the network with residual learning[53,58,58,66] is suitable to avoid the gradient vanishing problem. This approach allows the CNN to effectively approximate the proximal operator and facilitate the optimization process.

Mardani et al.[53] introduced a recurrent neural network (RNN) architecture enhanced by residual learning to learn the proximal operator more effectively. This learnable proximal mapping effectively functions as a denoiser, progressively eliminating aliasing artifacts from the input image.

3.2.1. Iterative Shrinkage-Thresholding Algorithm (ISTA) Network

ISTA-Net$^+$[46] formulate the regularizer as a ℓ_1 norm of non-linear transform $\mathcal{R}(\mathbf{x}) = \|\varphi(\mathbf{x})\|_1$. The proximal gradient descent updates (8) become:

$$\mathbf{b}_t = \mathbf{x}_t - \rho_t \mathbf{A}^\top (\mathbf{A}\mathbf{x}_t - \mathbf{f}), \tag{10a}$$

$$\mathbf{x}_{t+1} = \arg\min_{\mathbf{x}} \frac{1}{2\rho_t} \|\mathbf{x} - \mathbf{b}_t\|_2^2 + \|\varphi(\mathbf{x})\|_1. \tag{10b}$$

The proximal step (10b) can be parameterized as an implicit residual update step due to the lack of closed form solution:

$$\mathbf{x}_{t+1} = \mathbf{b}_t + \mathcal{H}(\mathbf{b}_t), \tag{11}$$

where $\mathcal{H}$ is a deep neural network with residual learning that approximates the proximal point.

Using the mean value theorem, ISTA-Net$^+$ derives an approximation theorem: $\|\varphi(\mathbf{x}) - \varphi(\mathbf{b}_t)\|_2^2 \approx \delta \|\mathbf{x} - \mathbf{b}_t\|_2^2$ with $\delta > 0$. Thus the proximal update step (10b) was written as

$$\mathbf{x}_{t+1} \approx \arg\min_{\mathbf{x}} \frac{1}{2} \|\varphi(\mathbf{x}) - \varphi(\mathbf{b}_t)\|_2^2 + \delta \rho_t \|\varphi(\mathbf{x})\|_1. \tag{12}$$

Assuming φ is orthogonal and invertible, ISTA-Net$^+$ provides the following closed-form solution:

$$\mathbf{x}_{t+1} = \tilde{\varphi}(\mathcal{S}_{\beta_t}(\varphi(\mathbf{b}_t))), \tag{13}$$

where $\beta_t = \delta\rho_t$ and $\mathcal{S}_{\beta_t}(\mathbf{x}) = \text{prox}_{\beta_t\|\cdot\|_1}(\mathbf{x}) = [\text{sign}(x_i)\max(|x_i| - \beta_t, 0)] \in \mathbb{R}^n$ is the soft shrinkage operator with vector $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Thus, equation (11) is reduced to an explicit form as given in (13), which we summarize together with (10a) in the following scheme:

$$\mathbf{b}_t = \mathbf{x}_t - \rho_t \mathbf{A}^\top (\mathbf{A}\mathbf{x}_t - \mathbf{f}), \tag{14a}$$

$$\mathbf{x}_t = \mathbf{b}_t + \tilde{\varphi}_t(\mathcal{S}_{\beta_t}(\varphi_t(\mathbf{b}_t))). \tag{14b}$$

The deep network $\tilde{\varphi}$ is applied with the symmetric structure to φ and is trained separately to enhance the capacity of the network. The initial input $\mathbf{x}_0$ was set to be the zero-filled reconstruction $\mathbf{A}^\top \mathbf{f}$.

The loss function was designed in two parts: The first part is the discrepancy loss:

$$\mathcal{L}_{dis}(\Theta) = \frac{1}{2}\|\mathbf{x}_T(\mathbf{f}, \Theta) - \mathbf{x}^*\|_2^2 \tag{15}$$

This loss measures the squared discrepancy between the reference image $\mathbf{x}^*$ and the reconstructed image from the last iteration $\mathbf{x}_T(\Theta)$. The second part of the loss function is to enforce the consistency of $\tilde{\varphi}_t$ and φ_t:

$$\mathcal{L}_{id}(\Theta) = \frac{1}{2}\sum_{t=1}^{T}\|\tilde{\varphi}_t(\varphi_t(\mathbf{x}^*)) - \mathbf{x}^*\|_2^2 \tag{16}$$

This loss aims to ensure that $\tilde{\varphi}_t\varphi_t = I$, an identity mapping. The training process is minimizes the following loss function:

$$\mathcal{L}_{loss} = \mathcal{L}_{dis}(\Theta) + \mu\mathcal{L}_{id}(\Theta). \tag{17}$$

where μ is a balancing parameter.

3.2.2. Parallel MRI Network

Parallel MRI network[58] leveraged residual learning to learn the proximal mapping and tackle model (8), thus bypassing the requirement for pre-calculated coil sensitivity maps in the encoding matrix (2). Similar to the model for joint reconstruction and synthesis[52], parallel MRI network considers the MRI reconstruction as a bi-level optimization problem:

$$\min_{\Theta} \quad \ell(\mathbf{x}_\Theta, \mathbf{x}^*), \tag{18a}$$

$$\text{s.t.} \quad \mathbf{x}_\Theta = \arg\min_{\mathbf{x}} \Phi_\Theta(\mathbf{x}). \tag{18b}$$

The variable $\mathbf{x} = (x_1, \ldots, x_c) \in \mathbb{C}^{m \times n \times c}$ denotes the multi-coil MRI data scanned from c coil elements, with each x_i corresponding to i-th coil for $i = 1 \cdots c$. The study constructs a model Φ_Θ to incorporate dual regularization terms applied to both image space and k-space, described by:

$$\Phi(\mathbf{x}) := \frac{1}{2}\sum_{i=1}^{c}\|\mathcal{P}_\Omega\mathcal{F}\mathbf{x}_i - \mathbf{f}_i\|_2^2 + \mathcal{R}(\mathcal{J}(\mathbf{x})) + \mathcal{R}_f(\mathcal{F}\mathbf{x}_i). \tag{19}$$

The channel-combination operator $\mathcal{J}$ aims to learn a combination of multi-coil MRI data which integrates the prior information among multiple channels. Then the image domain regularizer $\mathcal{R}$ extracts the information from the channel-integration image $\mathcal{J}(\mathbf{x})$. The regularizer $\mathcal{R}_f$ is designed to obtain prior information from k-space data.

The upper-level optimization (18a) is the network training process where the loss function $\ell(\mathbf{x}_\Theta, \mathbf{x}^*)$ is defined as the discrepancy between learned $\mathbf{x}_\Theta$ and the ground truth $\mathbf{x}^*$. The lower level optimization (18b) is solved by the following redefined algorithm:

$$\mathbf{b}_i^{(t)} = \mathbf{x}_i^{(t)} - \rho_t \mathcal{F}^{-1} \mathcal{P}_\Omega^\top (\mathcal{P}_\Omega \mathcal{F} \mathbf{x}_i^{(t)} - \mathbf{f}_i), \quad i = 1, \cdots, c, \tag{20a}$$

$$\bar{\mathbf{x}}_i^{(t)} = [\text{prox}_{\rho_t \mathcal{R}(\mathcal{J}(\cdot))}(\mathbf{b}^{(t)})]_i, \quad i = 1, \cdots, c, \tag{20b}$$

$$\mathbf{x}_i^{(t+1)} = \text{prox}_{\rho_t \mathcal{R}_f(\mathcal{F}(\cdot))}(\bar{\mathbf{x}}_i^{(t)}), \quad i = 1, \cdots, c. \tag{20c}$$

The proximal operator can be understood as a Gaussian denoiser. Nevertheless, the proximal operator $\text{prox}_{\rho_t \mathcal{R}}$ in the objective function (20b) lacks a closed-form solution, necessitating the use of a CNN as a substitute for $\text{prox}_{\rho_t \mathcal{R}}$. This network is designed as a residual learning network denoted by ϕ in the image domain and φ in the k-space domain, and the algorithm (20) is implemented in the following scheme:

$$\mathbf{b}_i^{(t)} = \mathbf{x}_i^{(t)} - \rho_t \mathcal{F}^{-1} \mathcal{P}_\Omega^\top (\mathcal{P}_\Omega \mathcal{F} \mathbf{x}_i^{(t)} - \mathbf{f}_i), \quad i = 1, \cdots, c, \tag{21a}$$

$$\bar{\mathbf{x}}_i^{(t)} = \mathbf{b}_i(t) + \phi(\mathbf{b}_i^{(t)}), \quad i = 1, \cdots, c, \tag{21b}$$

$$\mathbf{x}_i^{(t+1)} = \bar{\mathbf{x}}_i(t) + \mathcal{F}^{-1} \varphi\big(\mathcal{F}(\bar{\mathbf{x}}_i^{(t)})\big), \quad i = 1, \cdots, c. \tag{21c}$$

The CNN ϕ utilizes channel-integration operator $\mathcal{J}$ and operates with shared weights across iterations, effectively learning spatial features. However, it may erroneously enhance oscillatory artifacts as real features. In the k-space denoising step (21c), the k-space network φ focuses on low-frequency data, helping to remove high-frequency artifacts and restore image structure. Alternating between (21b) and (21c) in their respective domains balances their strengths and weaknesses, improving overall performance.

This network architecture has also been generalized to quantitative MRI (qMRI) reconstruction problems under a self-supervised learning framework. The next subsection introduces a similar learnable optimization algorithm for the qMRI reconstruction network.

3.2.3. RELAX-MORE

RELAX-MORE [66] introduced an optimization algorithm to unroll the proximal gradient for qMRI reconstruction. RELAX-MORE is a self-supervised learning where the loss function minimizes the discrepancy between undersampled reconstructed MRI k-space data and the "true" undersampled k-space data retrospectively. The well-trained model can be applied to other testing data using transfer learning. As new techniques develop [67–69], transfer learning may serve as an effective method to enhance the reconstruction timing efficiency of RELAX-MORE.

The qMRI reconstruction model aims to reconstruct the quantitative parameters $\mathbf{P}$ and this problem can be formulated as a bi-level optimization model:

$$\min_\Theta \ell(\mathcal{P}_\Omega \mathcal{F} \mathcal{S} \mathcal{M}(\mathbf{P}(\mathbf{f}|\Theta)), \mathbf{f}) \quad \text{s.t.} \tag{22a}$$

$$\mathbf{P}(\mathbf{f}|\Theta) = \arg\min_{\mathbf{P}} \mathcal{K}_\Theta(\mathbf{P}), \tag{22b}$$

$$\text{where} \quad \mathcal{K}_\Theta(\mathbf{P}) := \frac{1}{2} \| \mathcal{P}_\Omega \mathcal{F} \mathcal{S} \mathcal{M}(\mathbf{P}) - \mathbf{f} \|_2^2 + \beta \mathcal{R}_\Theta(\mathbf{P}). \tag{22c}$$

The model $\mathcal{M}$ represents the MR signal function that maps the set of quantitative parameters $\mathbf{P} := \{\rho_1, \cdots, \rho_N\}$ to the MRI data. The loss function in (22a) is addressed through a self-supervised learning network, and $\mathbf{P}(\mathbf{f}|\Theta)$ is derived from the network parameterized by Θ. The upper-level problem (22a) focuses on optimizing the learnable parameters for network training, while the lower level problem (22b) concentrates on optimizing the quantitative MR parameters. RELAX-MORE uses

$\mathbf{T}_1$ mapping obtained through the variable flip angle (vFA) method [70] as an example. The MR signal model $\mathcal{M}$ is described by the following equation:

$$\mathcal{M}_k(\mathbf{T}_1, \mathbf{I}_0) = \mathbf{I}_0 \cdot \frac{(1 - e^{-TR/\mathbf{T}_1}) \sin \eta_k}{1 - e^{-TR/\mathbf{T}_1} \cos \eta_k}, \tag{23}$$

where η_k represents flip angle for $k = 1, \ldots, N_k$, where N_k is the total number of the flip angles acquired. $\mathbf{T}_1$ and $\mathbf{I}_0$ are the spin-lattice relaxation time maps and proton density maps, respectively. Therefore, the parameter set need to reconstruct is $\mathbf{P} = \{\mathbf{T}_1, \mathbf{I}_0\}$.

Similar to the Parallel MRI Network[58], RELAX-MORE employs a proximal gradient descent algorithm to address the lower level problem (22b), with a residual network structure designed to learn the proximal mapping. Below is the unrolled learnable algorithm for resolving (22b):

Algorithm 3 Learnable Proximal Gradient Descent Algorithm

Input: $\rho_i^{(0)}, \mu_i^{(1)}, \nu_i^{(1)}, i = 1, \cdots, N$.
1: **for** $t = 1$ to T **do**

2:   **for** $i = 1$ to N **do**

3:     $\bar{\rho}_i^{(t)} = \rho_i^{(t-1)} - \mu_i^{(t)} \nabla_{\rho_i} \frac{1}{2} \|\mathcal{P}_\Omega \mathcal{F} \mathcal{S} \mathcal{M}(\{\rho_i^{(t-1)}\}_{i=1}^N) - \mathbf{f}\|_2^2,$
4:     $\rho_i^{(t)} = \widetilde{\mathcal{W}}_{\Theta_i}^{(t)} \circ \mathcal{T}_{\nu_i}^{(t)} \circ \mathcal{W}_{\Theta_i}^{(t)}(\bar{\rho}_i^{(t)}) + \bar{\rho}_i^{(t)}$
5:   **end for**
6:   $\mathbf{x}^{(t)} = \mathcal{M}(\{\rho_i^{(t)}\}_{i=1}^N),$
7: **end for**
Output: $\{\rho_i^{(T)}\}_{i=1}^N$ and $\mathbf{x}^{(t)}, \forall t \in \{1, \cdots, T\}$.

Step (4) implements the residual network structure to learn the proximal operator with regularization $\beta \mathcal{R}_\Theta$. The learnable operators $\widetilde{\mathcal{W}}_\Theta$ and $\mathcal{W}_\Theta$ has symmetric network structure, and $\mathcal{T}_\nu^{(t)}$ is the soft thresholding operator threshold parameter ν.

3.3. Alternating Direction Method of Multipliers (ADMM) Algorithm Inspired Networks

ADMM introduced an auxiliary variable $\mathbf{v}$ to solve the following bi-level problem:

$$\min_{\mathbf{x}, \mathbf{v}} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{f}\|_2^2 + \mu \mathcal{R}(\mathbf{v}) \tag{24a}$$

$$\text{s.t. } \mathbf{v} = \mathbf{D}\mathbf{x}, \tag{24b}$$

we can consider $\mathbf{D}$ as a gradient operator to reinforce the sparsity of MRI data such as total variation norm.

The first step is to form the augmented Lagrangian for the given problem. The augmented Lagrangian combines the objective function with a penalty term for the constraint violation and a Lagrange multiplier:

$$L_\rho(\mathbf{x}, \mathbf{v}, \lambda) = \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{f}\|_2^2 + \mu \mathcal{R}(\mathbf{v}) + <\lambda, \mathbf{D}\mathbf{x} - \mathbf{v}> + \frac{\rho}{2} \|\mathbf{D}\mathbf{x} - \mathbf{v}\|_2^2, \tag{25}$$

where λ is he Lagrange multiplier and $\rho > 0$ is a penalty parameter.

The ADMM algorithm solves the above problem by alternating the following three subproblems:

$$\mathbf{x}_{t+1} = \arg\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{f}\|_2^2 + \frac{\rho}{2} \|\mathbf{D}\mathbf{x} - (\mathbf{v}_t - \mathbf{u}_t)\|_2^2, \tag{26a}$$

$$\mathbf{v}_{t+1} = \arg\min_{\mathbf{v}} \frac{\rho}{2} \|\mathbf{D}\mathbf{x}_{t+1} - (\mathbf{v} - \mathbf{u}_t)\|_2^2 + \mu \mathcal{R}(\mathbf{v}), \tag{26b}$$

$$\mathbf{u}_{t+1} = \mathbf{u}_t + (\mathbf{D}\mathbf{x}_{t+1} - \mathbf{v}_t). \tag{26c}$$

We can obtain the closed-form solutions for each subproblem as follows:

$$\mathbf{x}_{t+1} = (\mathbf{A}^\top \mathbf{A} + \rho \mathbf{D}^\top \mathbf{D})^{-1}(\mathbf{A}^\top \mathbf{f} + \rho \mathbf{D}^\top(\mathbf{v}_t - \mathbf{u}_t)), \tag{27a}$$

$$\mathbf{v}_{t+1} = \text{prox}_{\mu/\rho}(\mathbf{D}\mathbf{x}_{t+1} + \mathbf{u}_t), \tag{27b}$$

$$\mathbf{u}_{t+1} = \mathbf{u}_t + (\mathbf{D}\mathbf{x}_{t+1} - \mathbf{v}_{t+1}). \tag{27c}$$

If regularizer is l_1 norm $\mathcal{R}(\mathbf{v}) = \|\mathbf{v}\|_1$, then (27b) reduces to $\mathbf{v}_{t+1} = \mathcal{S}_\beta(\mathbf{D}\mathbf{x}_{t+1} + \mathbf{u}_t)$ with the soft shrinkage threshold $\beta = \mu/\rho$.

3.3.1. ADMM-Net

ADMM-Net[45] reformulates these three steps through an augmented Lagrangian method. This approach leverages a cell-based architecture to optimize neural network operations for MRI image reconstruction. The network is structured into several layers, each corresponding to a specific operation in the ADMM optimization process. The gradient operator $\mathbf{D}$ is parameterized as deep neural network $\mathcal{D}$. All the scalars μ and ρ are learnable parameters to be trained and updated through ADMM iterations:

$$\mathbf{x}_{t+1} = (\mathbf{A}^\top \mathbf{A} + \rho_t \mathcal{D}^\top \mathcal{D})^{-1}(\mathbf{A}^\top \mathbf{f} + \rho_t \mathcal{D}^\top(\mathbf{v}_t - \mathbf{u}_t)), \tag{28a}$$

$$\mathbf{v}_{t+1} = \mathcal{S}_{\beta_t}(\mathcal{D}\mathbf{x}_{t+1} + \mathbf{u}_t), \tag{28b}$$

$$\mathbf{u}_{t+1} = \mathbf{u}_t + (\mathcal{D}\mathbf{x}_{t+1} - \mathbf{v}_{t+1}). \tag{28c}$$

The Reconstruction layer (28a) uses a combination of Fourier and penalized transformations to reconstruct images from undersampled k-space data, incorporating learnable penalty parameters and filter matrices. The Convolution layer $c_t = \mathcal{D}\mathbf{x}_{t+1}$ applies a convolution operation, transforming the reconstructed image to enhance feature representation, using distinct, learnable filter matrices to increase the network's capacity. The Non-linear Transform layer (28b) replaces traditional regularization functions with a learnable piecewise linear function, allowing for more flexible and data-driven transformations that go beyond simple thresholding. Finally, the Multiplier Update layer (28c) updates the Lagrangian multipliers, essential for integrating constraints into the learning process, with learnable parameters to adaptively refine the model's accuracy. Each layer's output is methodically fed into the next, ensuring a coherent flow that mimics the iterative ADMM process, thus systematically refining the image reconstruction quality with each pass through the network.

3.4. Primal-Dual Hybrid Gradient (PDHG) Algorithm Inspired Networks

There are several networks[51,71] are developed inspired by the PDHG algorithm. PDHG can be used to solve the model (1) by iterating the following steps:

$$\mathbf{d}_{t+1} = \text{prox}_{\zeta_t H^\top(\cdot)}(\mathbf{d}_t + \zeta_t \mathbf{A}\bar{\mathbf{x}}_t), \tag{29a}$$

$$\mathbf{x}_{t+1} = \text{prox}_{\rho_t \mathcal{R}(\cdot)}(\mathbf{x}_t + \eta_t \mathbf{A}^\top \mathbf{d}_{t+1}), \tag{29b}$$

$$\bar{\mathbf{x}}_{t+1} = \mathbf{x}_{t+1} + \theta(\mathbf{x}_{t+1} - \mathbf{x}_t), \tag{29c}$$

where H is the data fidelity function defined as $H(\mathbf{A}\mathbf{x}, \mathbf{f}) := \|\mathbf{A}\mathbf{x} - \mathbf{f}\|_2^2$ in the model (1). In the Learned PDHG[53], the traditional proximal operators are replaced with learned parametric operators. These operators are not necessarily proximal but are instead learned from training data, aiming to act similarly to denoising operators, such as Block Matching 3D (BM3D). The proximal operators can be parameterized as deep networks $\mathcal{G}_{\theta'}$ and $\mathcal{K}_{\theta''}$. PD-Net[71] iterates the following two steps:

$$\mathbf{d}_{t+1} = \mathcal{G}_{\theta'}(\mathbf{d}_t, \zeta_t \mathbf{A}\mathbf{x}_t, \mathbf{f}), \tag{30a}$$

$$\mathbf{x}_{t+1} = \mathcal{K}_{\theta''}(\mathbf{x}_t, \mathbf{A}^\top \mathbf{d}_{t+1}), \tag{30b}$$

The key innovation here is that these operators—both for the primal and dual variables—are parameterized and optimized during training, allowing the model to learn optimal operation strategies directly from the data. The learned PDHG operates under a fixed number of iterations, which serves as a stopping criterion. This approach ensures that the computation time remains predictable and manageable, which is beneficial for time-sensitive applications. The algorithm maintains its structure but becomes more adaptive to specific data characteristics through the learning process, potentially enhancing reconstruction quality over traditional methods.

3.5. Diffusion Models Meet Gradient Descent for MRI Reconstruction

A notable development for MRI reconstruction using diffusion model is the emergence of De-noising Diffusion Probabilistic Models (DDPMs)[72–75]. In Denoising Diffusion Probabilistic Models (DDPMs), the forward diffusion process systematically introduces noise into the input data, incrementally increasing the noise level until the data becomes pure Gaussian noise. This alteration progressively distorts the original data distribution. Conversely, the reverse diffusion process, or the denoising process, aims to reconstruct the original data structure from this noise-altered distribution. DDPMs effectively employ a Markov chain mechanism to transition from a noise-modified distribution back to the original data distribution via learned Gaussian transitions. The learnable Gaussian noise can be parametrized in a U-net architecture that consists of transformers/attention layers[76] in each diffusion step. The Transformer model has demonstrated promising performance in generating global information and can be effectively utilized for image denoising tasks.

DDPMs represent an innovative class of generative models renowned for their ability to master complex data distributions and achieve high-quality sample generation without relying on adversarial training methods. Their adoption in MRI reconstruction has been met with growing enthusiasm due to their robustness, particularly in handling distribution shifts. Recent studies exploring DDPM-based MRI reconstructions [72–75] demonstrate how these models can generate noisy MR images which are progressively denoised through iterative learning at each diffusion step, either unconditionally or conditionally. This approach has shown promise in enhancing MRI workflows by speeding up the imaging process, improving patient comfort, and boosting clinical throughput. Moreover, the model [73] has proven exceptionally robust, producing high-quality images even when faced with data that deviates from the training set (distribution shifts) [67], accommodating various patient anatomies and conditions, and thus enhancing the accuracy and reliability of diagnostic imaging.

3.5.1. Score-Based Diffusion Model

Chung et al. [72] presented an innovative framework that applies score-based diffusion models to solve inverse imaging problems. The core technique involves training a continuous time-dependent score function using denoising score matching. The score function of the data distribution $\log p_t(\mathbf{x}(t))$ is defined as the gradient of log density w.r.t the input data. This is estimated by a time-conditional deep neural network $S_\Theta(\mathbf{x}(t), t)$. The score model is trained by minimizing the following loss function on the magnitude image:

$$\mathcal{L}(\Theta) = \mathbb{E}_{\mathbf{x}(t) \sim p(\mathbf{x}(t)|\mathbf{x}(0)), \mathbf{x}(0) \sim p_{data}} \left[\| S_\Theta(\mathbf{x}(t), t) - \nabla_{\mathbf{x}(t)} \log p_t(\mathbf{x}(t)|\mathbf{x}(0)) \|_2^2 \right]. \tag{31}$$

During inference, the model alternates between a numerical Stochastic Differential Equation (SDE) solver and a data consistency step to reconstruct images. The method is agnostic to subsampling patterns, enabling its application across various sampling schemes and body parts not included in the training data. Chung et al. [72] proposed the following algorithm with predictor-corrector (PC) sampling algorithm[77]. For $i = N - 1, \cdots, 0$, the predictor is defined as $\mathbf{x}_i = \mathbf{x}_{i+1} + (\sigma_{i+1}^2 - \sigma_i^2) S_\Theta(\mathbf{x}_{i+1}, \sigma_{i+1}) + \sqrt{\sigma_{i+1}^2 - \sigma_i^2} \epsilon$ with $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The corrector is defined as $\mathbf{x}_i = \mathbf{x}_i + \lambda_i S_\Theta(\mathbf{x}_i, \sigma_i) + \sqrt{2\lambda_i} \epsilon$ with step size $\lambda_i > 0$.

Incorporating a gradient descent step to emphasize the data consistency after the predictor and corrector, we can obtain the following algorithm:

Algorithm 4 Score-based sampling for MRI reconstruction [72]

Input: $\mathbf{x}_N \sim \mathcal{N}(\mathbf{0}, \sigma_T^2 \mathbf{I})$, Learned score function Θ, step size $\{\epsilon_i\}$, noise schedule $\{\sigma_i\}$ and MRI encoding matrix $\mathbf{A}$.

1: **for** $i = N - 1, ..., 0$ **do**

2:     $\text{Re}(\mathbf{x}_i) \leftarrow \text{Predictor}(\text{Re}(\mathbf{x}_{i+1}), \sigma_i, \sigma_{i+1})$
3:     $\text{Im}(\mathbf{x}_i) \leftarrow \text{Predictor}(\text{Im}(\mathbf{x}_{i+1}), \sigma_i, \sigma_{i+1})$
4:     $\mathbf{x}_i = \text{Re}(\mathbf{x}_i) + j\text{Im}(\mathbf{x}_i)$
5:     $\mathbf{x}_i \leftarrow \mathbf{x}_i - \mathbf{A}^\top(\mathbf{A}\mathbf{x}_i - \mathbf{f})$
6:     **for** $j = 1, ..., M$ **do**

7:         $\text{Re}(\mathbf{x}_i) \leftarrow \text{Corrector}(\text{Re}(\mathbf{x}_{i+1}), \sigma_i, \epsilon_i)$
8:         $\text{Im}(\mathbf{x}_i) \leftarrow \text{Corrector}(\text{Im}(\mathbf{x}_{i+1}), \sigma_i, \epsilon_i)$
9:         $\mathbf{x}_i = \text{Re}(\mathbf{x}_i) + j\text{Im}(\mathbf{x}_i)$
10:        $\mathbf{x}_i \leftarrow \mathbf{x}_i - \mathbf{A}^\top(\mathbf{A}\mathbf{x}_i - \mathbf{f})$
11:    **end for**
12: **end for**

The above algorithm can be variate to other two different algorithms. One is the parallel implementation for each coil image for parallel MRI reconstruction. The other one considers the correlation among the multiple coil images and eliminates the calculation of sensitivity maps, and the final magnitude image is obtained by using the sum-of-root-sum-of-squares of each coil. The results outperforms conventional deep learning methods: UNet[78], DuDoRNet[79] and E2E-Varnet[80] which requires complex k-space data.

3.5.2. Domain-Conditioned Diffusion Modeling

Domain-conditioned Diffusion Modeling (DiMo)[60] and quantitative DiMo were developed to apply on both accelerated multi-coil MRI and quantitative MRI (qMRI) using diffusion models conditioned on the native data domain rather than the image domain. The method incorporates a gradient descent optimization within the diffusion steps to improve feature learning and denoising effectiveness. The training and sampling algorithm for MRI reconstruction is illustrated in Algorithm 5 and 6.

Algorithm 5 Training Process of Static DiMo

Input: $t \sim \text{Uniform}(\{1, \cdots, T\}), \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, fully scanned k-space $\hat{\mathbf{f}}_0 \sim q(\mathbf{f}_0)$, undersampling mask $\mathcal{P}_\Omega$, partial scanned k-space $\mathbf{f}$, and coil sensitivities $\mathcal{S}$.

  Initialisation : η_0
1: $\hat{\mathbf{f}}_t \leftarrow \sqrt{\bar{\alpha}_t}\hat{\mathbf{f}}_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon.$
2: $\hat{\mathbf{f}}_t \leftarrow \mathcal{P}_\Omega(\lambda_t \mathbf{f} + (1 - \lambda_t)\hat{\mathbf{f}}_t) + (\mathbb{1} - \mathcal{P}_\Omega)\hat{\mathbf{f}}_t$ ▷ DC
3: **for** $k = 0$ to $K - 1$ **do**

4:     $\hat{\mathbf{f}}_t \leftarrow \hat{\mathbf{f}}_t - \eta_k \nabla_{\hat{\mathbf{f}}_t} \frac{1}{2}\|\mathbf{A}\mathcal{F}^{-1}\hat{\mathbf{f}}_t - \mathbf{f}\|_2^2$ ▷ GD
5: **end for**
6: Take gradient descent update step

    $\nabla_\theta \|\epsilon - \epsilon_\theta(\hat{\mathbf{f}}_t, t)\|_2^2$

  Until converge
Output: $\hat{\mathbf{f}}_t, t \in \{1, \cdots, T\}$.

Algorithm 6 Sampling Process of Static DiMo

Input: $\hat{\mathbf{f}}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, undersampling mask $\mathcal{P}_\Omega$, partial scanned k-space $\mathbf{f}$, and coil sensitivities $\mathcal{S}$.
1: **for** $t = T - 1, ..., 0$ **do**

2: $\quad z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 0$, else $z = 0$
3: $\quad \hat{\mathbf{f}}_t = \mu_\theta(\hat{\mathbf{f}}_{t+1}, t+1) + \sigma_{t+1} z$
4: $\quad \hat{\mathbf{f}}_t \leftarrow \mathcal{P}_\Omega(\lambda_t \mathbf{f} + (1 - \lambda_t)\hat{\mathbf{f}}_t) + (\mathbb{1} - \mathcal{P}_\Omega)\hat{\mathbf{f}}_t$ $\qquad\qquad\qquad\qquad\qquad\qquad$ ▷ DC
5: $\quad$ **for** $k = 0$ to $K - 1$ **do**

6: $\qquad \hat{\mathbf{f}}_t \leftarrow \hat{\mathbf{f}}_t - \eta_k \nabla_{\hat{\mathbf{f}}_t} \frac{1}{2}\|\mathbf{A}\mathcal{F}^{-1}\hat{\mathbf{f}}_t - \mathbf{f}\|_2^2$ $\qquad\qquad\qquad\qquad\qquad\qquad$ ▷ GD
7: $\quad$ **end for**
8: **end for**
Output: $\hat{\mathbf{f}}_0$

In the training and sampling algorithm, the data consistency (DC) term was used to emphasize the physical consistency between the partial k-space and reconstructed images. Then gradient descent (GD) algorithm is applied iteratively into the diffusion step to refine k-space data further. The matrix $\mathbb{1}$ only contains value of one. GD in here solves the optimization problem (1) without the regularization term.

Static DiMo performed a qualitative comparison with both image domain diffusion model: Chung et al. [72] and k-space domain diffusion model: MC-DDPM[81] and demonstrated robust performance in reconstruction quality and noise reduction. Quantitative DiMo reconstructs the quantitative parameter maps from the partial k-space data. The MR signal model $\mathcal{M}$ defined in (23) maps MR parameter maps to the static MRI, therefore it is one more function inside the reconstruction model:

$$\min_\Delta \frac{1}{2}\|\mathbf{A}\mathcal{M}(\Delta) - \mathbf{f}\|_2^2. \tag{32}$$

The MR parameter maps are denoted as $\Delta = \{\delta_i\}_{i=1}^N$ where δ_i indicates each MR parameter and N is the total number of MR parameters to be estimated.

The training and sampling diffusion model for quantitative MRI (qMRI) DiMo follows the same steps as static DiMo, where the signal model should take the inverse when calculating the quantitative maps Δ from the updated k-space. Quantitative DiMo showed the least error compared to other methods[82–84]. This is likely achieved through integrating the unrolling gradient descent algorithm and diffusion denoising network, prioritizing noise suppression without compromising the fidelity and clarity of the underlying tissue structure.

4. Discussion

4.1. Selection of Acquisition Parameters

The selection of acquisition parameters is a critical aspect of MRI reconstruction, influencing the efficiency and accuracy of deep learning models. The architecture of deep reconstruction networks and efficient numerical methods play a pivotal role in this selection. In LOAs, one must determine the appropriate number of iterations T and the initial step size for gradient descent to ensure the reconstruction network converges to the local optimum of the problem (1). The convergence to the local optimum is essential for producing high-quality reconstructed images. The required number of iterations and the step size depend on the specific application tasks and whether the step size is learnable or fixed in the gradient descent-based algorithm used for reconstruction. Proper tuning of these parameters is crucial for optimizing the performance of the reconstruction network.

4.2. Computing Memory Consumption and High-Performance Computing

Deep learning-based MRI reconstruction methods, particularly those involving unrolled optimization algorithms, demand significant computational resources. The training process involves substantial GPU memory consumption to store intermediate results and their corresponding gradients.

This high memory requirement, coupled with potentially long training times, arises from the need to repeatedly apply the forward and adjoint operators during training. To address these challenges, leveraging high-performance computing resources such as quantum computing could be explored. Quantum computing offers the potential to dramatically accelerate computations and reduce memory constraints[85], paving the way for more efficient and scalable deep learning models in MRI reconstruction. Additionally, advancements in hardware acceleration, such as the use of specialized AI chips and tensor processing units (TPUs), could further enhance the performance of deep learning-based MRI reconstruction.

4.3. Theoretical Convergence and Practical Considerations

While unrolling-based deep-learning methods are derived from numerical algorithms with convergence guarantees, these guarantees do not always extend to the unrolled methods due to their dynamic nature and the direct replacement of functions by neural networks. Theoretical convergence is compromised, and only a few works have analyzed the convergence behavior of unrolling-based methods in theory. Notable studies include [52,56,86], which provide insights into the theoretical convergence of these methods. For example, reference [52] proved that if $\mathbf{x}^{(t)}$ satisfies the stopping criterion, then there exists a subsequence $\mathbf{x}^{(t_{l+1})}$ at least one accumulation point, and every accumulation point of $\mathbf{x}^{(t_{l+1})}$ is a Clarke stationary point of the (1). Understanding the convergence properties of unrolled networks is crucial for ensuring the reliability and robustness of MRI reconstruction algorithms. Future research should focus on establishing stronger theoretical foundations and convergence guarantees for unrolling-based deep learning methods. This includes developing new theoretical frameworks that can account for the dynamic and adaptive nature of these models, as well as creating more rigorous validation protocols.

4.4. Limitations of the Existing Deep Learning Approaches

Despite the advancements in reconstructing MRI images through deep learning methods, there are several practical challenges that need to be addressed. Most deep learning approaches focus on designing end-to-end networks that are independent of intrinsic MRI physical characteristics, leading to sub-optimal performance. Deep learning methods are also often criticized for their lack of mathematical interpretation, being seen as "black boxes." Acquiring and processing large high-quality datasets that are needed for training deep-learning models may be difficult, especially when dealing with diverse patient populations and varying imaging conditions. Training deep neural networks may require a large quantity of data and may be prone to over-fitting when data is scarce. Additionally, both the training and inference process of deep learning models may require substantial computational resources, which may act as a barrier for certain medical institutions. Hence, it may be a trade-off between cost and time. Ensuring that the model generalizes well across different MRI scans and clinical settings is also essential for the widespread adoption of deep-learning techniques. Finally, the technologies utilized in clinical settings might need to be validated, transparent, and fully interpretable in order to ensure that clinicians trust the decision-making capabilities of the algorithms. Future studies may focus on addressing the above-mentioned challenges, which would accelerate the adoption of deep learning methods and advance the field of medical imaging.

4.5. Future Directions and Research Opportunities

Future research should aim to address the limitations of current models and explore new avenues for enhancement. Emerging AI techniques, such as reinforcement learning, offer promising directions for improving MRI reconstruction. Reinforcement learning can optimize acquisition parameters dynamically during the scan, leading to more efficient data collection and potentially reducing scan times further. Personalized medicine approaches, where models are tailored to individual patient characteristics, could provide significant benefits. These approaches can leverage patient-specific data to enhance the accuracy and reliability of MRI reconstructions, leading to more precise diagnoses and

personalized treatment plans. Additionally, the integration of self-supervised learning techniques, which can operate with limited ground truth data, represents a promising avenue. Self-supervised learning can leverage the inherent structure in MRI data to improve model training, reducing the dependency on extensive labeled datasets.

Moreover, exploring hybrid models that combine multiple algorithms may offer a more comprehensive solution to the challenges of MRI reconstruction. These hybrid models can integrate the strengths of different techniques, such as combining the robustness of classical optimization with the adaptability of deep learning. Collaborative efforts between researchers, clinicians, and industry partners will be essential for advancing the field and translating research innovations into clinical practice. Ensuring that these models are user-friendly and seamlessly integrated into existing clinical workflows will be critical for their successful implementation.

5. Conclusion

In conclusion, this paper provides a comprehensive overview of several optimization algorithms and network unrolling methods for MRI reconstruction. The discussed techniques include gradient descent algorithms, proximal gradient descent algorithms, ADMM, PDHG, and diffusion models combined with gradient descent. By summarizing these advanced methodologies, we aim to offer a valuable resource for researchers seeking to enhance MRI reconstruction through optimization-based deep learning approaches. The insights presented in this review are expected to facilitate further development and application of these algorithms in the field of medical imaging.

References

1. Kim, R.J.; Wu, E.; Rafael, A.; Chen, E.L.; Parker, M.A.; Simonetti, O.; Klocke, F.J.; Bonow, R.O.; Judd, R.M. The use of contrast-enhanced magnetic resonance imaging to identify reversible myocardial dysfunction. *New England Journal of Medicine* **2000**, *343*, 1445–1453.
2. Kasivisvanathan, V.; Rannikko, A.S.; Borghi, M.; Panebianco, V.; Mynderse, L.A.; Vaarala, M.H.; Briganti, A.; Budäus, L.; Hellawell, G.; Hindley, R.G.; others. MRI-targeted or standard biopsy for prostate-cancer diagnosis. *New England Journal of Medicine* **2018**, *378*, 1767–1777.
3. Pruessmann, K.P.; Weiger, M.; Scheidegger, M.B.; Boesiger, P. SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine* **1999**, *42*, 952–962.
4. Griswold, M.A.; Jakob, P.M.; Heidemann, R.M.; Nittka, M.; Jellus, V.; Wang, J.; Kiefer, B.; Haase, A. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine* **2002**, *47*, 1202–1210.
5. Nyquist, H. Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers* **1928**, *47*, 617–644.
6. Donoho, D.L. Compressed sensing. *IEEE Transactions on information theory* **2006**, *52*, 1289–1306.
7. Sodickson, D.K.; Manning, W.J. Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. *Magnetic resonance in medicine* **1997**, *38*, 591–603.
8. Larkman, D.J.; Nunes, R.G. Parallel magnetic resonance imaging. *Physics in Medicine & Biology* **2007**, *52*, R15.
9. Siddique, N.; Paheding, S.; Elkin, C.P.; Devabhaktuni, V. U-net and its variants for medical image segmentation: A review of theory and applications. *Ieee Access* **2021**, *9*, 82031–82057.
10. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer, 2015, pp. 234–241.
11. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS

2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. Springer, 2018, pp. 3–11.

12. Zhan, C.; Ghaderibaneh, M.; Sahu, P.; Gupta, H. Deepmtl pro: Deep learning based multiple transmitter localization and power estimation. *Pervasive and Mobile Computing* **2022**, *82*, 101582.

13. Zhan, C.; Ghaderibaneh, M.; Sahu, P.; Gupta, H. Deepmtl: Deep learning based multiple transmitter localization. 2021 IEEE 22nd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM). IEEE, 2021, pp. 41–50.

14. Hammernik, K.; Klatzer, T.; Kobler, E.; Recht, M.P.; Sodickson, D.K.; Pock, T.; Knoll, F. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic resonance in medicine* **2018**, *79*, 3055–3071.

15. Schlemper, J.; Caballero, J.; Hajnal, J.V.; Price, A.; Rueckert, D. A deep cascade of convolutional neural networks for MR image reconstruction. Information Processing in Medical Imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings 25. Springer, 2017, pp. 647–658.

16. Han, Y.; Yoo, J.; Kim, H.H.; Shin, H.J.; Sung, K.; Ye, J.C. Deep learning with domain adaptation for accelerated projection-reconstruction MR. *Magnetic resonance in medicine* **2018**, *80*, 1189–1205.

17. Zhu, B.; Liu, J.Z.; Cauley, S.F.; Rosen, B.R.; Rosen, M.S. Image reconstruction by domain-transform manifold learning. *Nature* **2018**, *555*, 487–492.

18. Yang, G.; Yu, S.; Dong, H.; Slabaugh, G.; Dragotti, P.L.; Ye, X.; Liu, F.; Arridge, S.; Keegan, J.; Guo, Y.; others. DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE transactions on medical imaging* **2017**, *37*, 1310–1321.

19. Lee, D.; Yoo, J.; Tak, S.; Ye, J.C. Deep residual learning for accelerated MRI using magnitude and phase networks. *IEEE Transactions on Biomedical Engineering* **2018**, *65*, 1985–1995.

20. Knoll, F.; Hammernik, K.; Zhang, C.; Moeller, S.; Pock, T.; Sodickson, D.K.; Akcakaya, M. Deep-learning methods for parallel magnetic resonance imaging reconstruction: A survey of the current approaches, trends, and issues. *IEEE signal processing magazine* **2020**, *37*, 128–140.

21. Yaman, B.; Hosseini, S.A.H.; Moeller, S.; Ellermann, J.; Uğurbil, K.; Akçakaya, M. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magnetic resonance in medicine* **2020**, *84*, 3172–3191.

22. Blumenthal, M.; Luo, G.; Schilling, M.; Haltmeier, M.; Uecker, M. NLINV-Net: Self-Supervised End-2-End Learning for Reconstructing Undersampled Radial Cardiac Real-Time Data. ISMRM annual meeting, 2022.

23. Bian, W. A Brief Overview of Optimization-Based Algorithms for MRI Reconstruction Using Deep Learning. *arXiv preprint arXiv:2406.02626* **2024**.

24. Jin, K.H.; McCann, M.T.; Froustey, E.; Unser, M. Deep convolutional neural network for inverse problems in imaging. *IEEE transactions on image processing* **2017**, *26*, 4509–4522.

25. Lundervold, A.S.; Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik* **2019**, *29*, 102–127.

26. Liang, D.; Cheng, J.; Ke, Z.; Ying, L. Deep magnetic resonance image reconstruction: Inverse problems meet neural networks. *IEEE Signal Processing Magazine* **2020**, *37*, 141–151.

27. Sandino, C.M.; others. Compressed Sensing: From Research to Clinical Practice With Deep Neural Networks: Shortening Scan Times for Magnetic Resonance Imaging. *IEEE Signal Processing Magazine* **2020**, *37*, 117–127.

28. McCann, M.T.; Jin, K.H.; Unser, M. Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Processing Magazine* **2017**, *34*, 85–95.

29. Zhou, S.K.; Greenspan, H.; Davatzikos, C.; Duncan, J.S.; Van Ginneken, B.; Madabhushi, A.; Prince, J.L.; Rueckert, D.; Summers, R.M. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the IEEE* **2021**, *109*, 820–838.

30. Singha, A.; Thakur, R.S.; Patel, T. Deep Learning Applications in Medical Image Analysis. *Biomedical Data Mining for Information Retrieval: Methodologies, Techniques and Applications* **2021**, pp. 293–350.

31. Chandra, S.S.; Bran Lorenzana, M.; Liu, X.; Liu, S.; Bollmann, S.; Crozier, S. Deep learning in magnetic resonance image reconstruction. *Journal of Medical Imaging and Radiation Oncology* **2021**.

32. Ahishakiye, E.; Van Gijzen, M.B.; Tumwiine, J.; Wario, R.; Obungoloch, J. A survey on deep learning in medical image reconstruction. *Intelligent Medicine* **2021**.

33. Liu, R.; Zhang, Y.; Cheng, S.; Luo, Z.; Fan, X. A Deep Framework Assembling Principled Modules for CS-MRI: Unrolling Perspective, Convergence Behaviors, and Practical Modeling. *IEEE Transactions on Medical Imaging* **2020**, *39*, 4150–4163.

34. Griswold, M.A.; others. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **2002**, *47*, 1202–1210.

35. Lustig, M.; Pauly, J.M. SPIRiT: iterative self-consistent parallel imaging reconstruction from arbitrary k-space. *Magnetic resonance in medicine* **2010**, *64*, 457–471.

36. Singh, D.; Monga, A.; de Moura, H.L.; Zhang, X.; Zibetti, M.V.; Regatte, R.R. Emerging trends in fast MRI using deep-learning reconstruction on undersampled k-space data: a systematic review. *Bioengineering* **2023**, *10*, 1012.

37. Sun, H.; Liu, X.; Feng, X.; Liu, C.; Zhu, N.; Gjerswold-Selleck, S.J.; Wei, H.J.; Upadhyayula, P.S.; Mela, A.; Wu, C.C.; others. Substituting gadolinium in brain MRI using DeepContrast. 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, 2020, pp. 908–912.

38. Wang, S.; others. DeepcomplexMRI: Exploiting deep residual network for fast parallel MR imaging with complex convolution. *Magnetic Resonance Imaging* **2020**, *68*, 136 – 147.

39. Wang, S.; Su, Z.; Ying, L.; Peng, X.; Zhu, S.; Liang, F.; Feng, D.; Liang, D. Accelerating magnetic resonance imaging via deep learning. 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), 2016, pp. 514–517.

40. Kwon, K.; Kim, D.; Park, H. A parallel MR imaging method using multilayer perceptron. *Medical physics* **2017**, *44*, 6209–6224.

41. Quan, T.M.; Nguyen-Duc, T.; Jeong, W.K. Compressed Sensing MRI Reconstruction Using a Generative Adversarial Network With a Cyclic Loss. *IEEE Transactions on Medical Imaging* **2018**, *37*, 1488–1497.

42. Mardani, M.; others. Deep Generative Adversarial Neural Networks for Compressive Sensing MRI. *IEEE Transactions on Medical Imaging* **2019**, *38*, 167–179.

43. Yang, Y.; Sun, J.; Li, H.; Xu, Z. Deep ADMM-Net for Compressive Sensing MRI. Advances in Neural Information Processing Systems. Curran Associates, Inc., 2016, Vol. 29.

44. Kofler, A.; Altekrüger, F.; Antarou Ba, F.; Kolbitsch, C.; Papoutsellis, E.; Schote, D.; Sirotenko, C.; Zimmermann, F.F.; Papafitsoros, K. Learning regularization parameter-maps for variational image reconstruction using deep neural networks and algorithm unrolling. *SIAM Journal on Imaging Sciences* **2023**, *16*, 2202–2246.

45. Yang, Y.; Sun, J.; Li, H.; Xu, Z. Deep ADMM-Net for compressive sensing MRI. Proceedings of the 30th international conference on neural information processing systems, 2016, pp. 10–18.

46. Zhang, J.; Ghanem, B. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 1828–1837.

47. Schlemper, J.; others. A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction. *IEEE Transactions on Medical Imaging* **2018**, *37*, 491–503.

48. Aggarwal, H.K.; Mani, M.P.; Jacob, M. MoDL: Model-Based Deep Learning Architecture for Inverse Problems. *IEEE Transactions on Medical Imaging* **2019**, *38*, 394–405.

49. Meng, N.; Yang, Y.; Xu, Z.; Sun, J. A prior learning network for joint image and sensitivity estimation in parallel MR imaging. International conference on medical image computing and computer-assisted intervention. Springer, 2019, pp. 732–740.

50. Duan, J.; Schlemper, J.; Qin, C.; Ouyang, C.; Bai, W.; Biffi, C.; Bello, G.; Statton, B.; O'regan, D.P.; Rueckert, D. VS-Net: Variable splitting network for accelerated parallel MRI reconstruction. Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part IV 22. Springer, 2019, pp. 713–722.

51. Adler, J.; Öktem, O. Learned primal-dual reconstruction. *IEEE transactions on medical imaging* **2018**, *37*, 1322–1332.

52. Bian, W.; Zhang, Q.; Ye, X.; Chen, Y. A learnable variational model for joint multimodal MRI reconstruction and synthesis. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2022, pp. 354–364.

53. Mardani, M.; Sun, Q.; Donoho, D.; Papyan, V.; Monajemi, H.; Vasanawala, S.; Pauly, J. Neural proximal gradient descent for compressive imaging. *Advances in Neural Information Processing Systems* **2018**, *31*.

54. Zeng, G.; Guo, Y.; Zhan, J.; Wang, Z.; Lai, Z.; Du, X.; Qu, X.; Guo, D. A review on deep learning MRI reconstruction without fully sampled k-space. *BMC Medical Imaging* **2021**, *21*, 195.

55. Bian, W.; Chen, Y.; Ye, X. Deep parallel MRI reconstruction network without coil sensitivities. Machine Learning for Medical Image Reconstruction: Third International Workshop, MLMIR 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 3. Springer, 2020, pp. 17–26.

56. Bian, W.; Chen, Y.; Ye, X.; Zhang, Q. An optimization-based meta-learning model for mri reconstruction with diverse dataset. *Journal of Imaging* **2021**, *7*, 231.

57. Bian, W. Optimization-Based Deep learning methods for Magnetic Resonance Imaging Reconstruction and Synthesis. PhD thesis, University of Florida, 2022.

58. Bian, W.; Chen, Y.; Ye, X. An optimal control framework for joint-channel parallel MRI reconstruction without coil sensitivities. *Magnetic Resonance Imaging* **2022**.

59. Bian, W.; Jang, A.; Liu, F. Magnetic Resonance Parameter Mapping using Self-supervised Deep Learning with Model Reinforcement. *ArXiv* **2023**.

60. Bian, W.; Jang, A.; Liu, F. Diffusion Modeling with Domain-conditioned Prior Guidance for Accelerated MRI and qMRI Reconstruction. *arXiv preprint arXiv:2309.00783* **2023**.

61. Bian, W.; Jang, A.; Liu, F. Multi-task Magnetic Resonance Imaging Reconstruction using Meta-learning. *arXiv preprint arXiv:2403.19966* **2024**.

62. Bian, W. A Review of Electromagnetic Elimination Methods for low-field portable MRI scanner. *arXiv preprint arXiv:2406.17804* **2024**.

63. Bian, W.; Jang, A.; Liu, F. Improving quantitative MRI using self-supervised deep learning with model reinforcement: Demonstration for rapid T1 mapping. *Magnetic Resonance in Medicine* **2024**, *92*, 98–111.

64. Heide, F.; Steinberger, M.; Tsai, Y.T.; Rouf, M.; Pająk, D.; Reddy, D.; Gallo, O.; Liu, J.; Heidrich, W.; Egiazarian, K.; others. Flexisp: A flexible camera image processing framework. *ACM Transactions on Graphics (ToG)* **2014**, *33*, 1–13.

65. Venkatakrishnan, S.V.; Bouman, C.A.; Wohlberg, B. Plug-and-play priors for model based reconstruction. 2013 IEEE Global Conference on Signal and Information Processing. IEEE, 2013, pp. 945–948.

66. Bian, W.; Jang, A.; Liu, F. Improving quantitative MRI using self-supervised deep learning with model reinforcement: Demonstration for rapid T1 mapping. *Magnetic Resonance in Medicine* **2024**.

67. Knoll, F.; Hammernik, K.; Kobler, E.; Pock, T.; Recht, M.P.; Sodickson, D.K. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. *Magnetic resonance in medicine* **2019**, *81*, 116–128.

68. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A comprehensive survey on transfer learning. *Proceedings of the IEEE* **2020**, *109*, 43–76.

69. Dar, S.U.H.; Özbey, M.; Çatlı, A.B.; Çukur, T. A transfer-learning approach for accelerated MRI using deep neural networks. *Magnetic resonance in medicine* **2020**, *84*, 663–685.

70. Wang, H.Z.; Riederer, S.J.; Lee, J.N. Optimizing the precision in T1 relaxation estimation using limited flip angles. *Magnetic resonance in medicine* **1987**, *5*, 399–416.

71. Cheng, J.; Wang, H.; Ying, L.; Liang, D. Model learning: Primal dual networks for fast MR imaging. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2019, pp. 21–29.

72. Chung, H.; Ye, J.C. Score-based diffusion models for accelerated MRI. *Medical image analysis* **2022**, *80*, 102479.

73. Güngör, A.; Dar, S.U.; Öztürk, Ş.; Korkmaz, Y.; Bedel, H.A.; Elmas, G.; Ozbey, M.; Çukur, T. Adaptive diffusion priors for accelerated MRI reconstruction. *Medical Image Analysis* **2023**, p. 102872.

74. Kazerouni, A.; Aghdam, E.K.; Heidari, M.; Azad, R.; Fayyaz, M.; Hacihaliloglu, I.; Merhof, D. Diffusion models in medical imaging: A comprehensive survey. *Medical Image Analysis* **2023**, p. 102846.

75. Yang, L.; Zhang, Z.; Song, Y.; Hong, S.; Xu, R.; Zhao, Y.; Zhang, W.; Cui, B.; Yang, M.H. Diffusion models: A comprehensive survey of methods and applications. *ACM Computing Surveys* **2023**, *56*, 1–39.

76. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Advances in neural information processing systems* **2017**, *30*.

77. Song, Y.; Sohl-Dickstein, J.; Kingma, D.P.; Kumar, A.; Ermon, S.; Poole, B. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456* **2020**.

78. Zbontar, J.; Knoll, F.; Sriram, A.; Murrell, T.; Huang, Z.; Muckley, M.J.; Defazio, A.; Stern, R.; Johnson, P.; Bruno, M.; others. fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv preprint arXiv:1811.08839* **2018**.

79. Zhou, B.; Zhou, S.K. DuDoRNet: learning a dual-domain recurrent network for fast MRI reconstruction with deep T1 prior. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 4273–4282.

80. Sriram, A.; Zbontar, J.; Murrell, T.; Defazio, A.; Zitnick, C.L.; Yakubova, N.; Knoll, F.; Johnson, P. End-to-end variational networks for accelerated MRI reconstruction. Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23. Springer, 2020, pp. 64–73.

81. Xie, Y.; Li, Q. Measurement-conditioned denoising diffusion probabilistic model for under-sampled medical image reconstruction. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2022, pp. 655–664.

82. Liu, F.; Kijowski, R.; El Fakhri, G.; Feng, L. Magnetic resonance parameter mapping using model-guided self-supervised deep learning. *Magnetic resonance in medicine* **2021**, *85*, 3211–3226.

83. Maier, O.; Schoormans, J.; Schloegl, M.; Strijkers, G.J.; Lesch, A.; Benkert, T.; Block, T.; Coolen, B.F.; Bredies, K.; Stollberger, R. Rapid T1 quantification from high resolution 3D data with model-based reconstruction. *Magnetic resonance in medicine* **2019**, *81*, 2072–2089.

84. Zhang, T.; Pauly, J.M.; Levesque, I.R. Accelerating parameter mapping with a locally low rank constraint. *Magnetic resonance in medicine* **2015**, *73*, 655–661.

85. Alsharabi, N.; Shahwar, T.; Rehman, A.U.; Alharbi, Y. Implementing magnetic resonance imaging brain disorder classification via AlexNet–quantum learning. *Mathematics* **2023**, *11*, 376.

86. Chen, Y.; Liu, H.; Ye, X.; Zhang, Q. Learnable descent algorithm for nonsmooth nonconvex image reconstruction. *SIAM Journal on Imaging Sciences* **2021**, *14*, 1532–1564.