Review

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, MA 02129, USA
2   Department of Radiology, Harvard Medical School, Boston, MA 02115, USA
3   Department of Electrical and Computer Engineering, University of Western Ontario,
    London, ON N6A 5B9, Canada; ykrishn4@uwo.ca
*   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 has significantly enhanced image quality and efficiency. This paper provides a comprehensive review of optimization-based deep learning models for MRI reconstruction, focusing on recent advancements in gradient descent algorithms, proximal gradient descent algorithms, ADMM, PDHG, and diffusion models combined with gradient descent. We highlight the development and effectiveness of learnable optimization algorithms (LOAs) in improving model interpretability and performance. Our findings demonstrate substantial improvements in MRI reconstruction in handling undersampled data, which directly contribute to reducing scan times and enhancing diagnostic accuracy. The review offers valuable insights and resources for researchers and practitioners aiming to advance medical imaging using state-of-the-art 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 highly detailed and accurate images of tissues in their natural, living state, which is 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 min, 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 undersampling, violating

the Nyquist criterion [5]. MRI reconstruction involves generating high-quality, artifact-free MR images from undersampled k-space data, which are then used for diagnostic and clinical purposes. In MRI reconstruction, the challenge lies in solving an inverse problem, where the goal is to recover an image from partially sampled and noisy k-space data. 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 undersampling and heaving the ability to reconstruct high-quality MRI images from undersampled data, CS significantly reduces scan time while offering images that are often comparable to those obtained from fully sampled data.

Traditional MRI reconstruction techniques suffer from several limitations and challenges. While faster MRI scans are desirable to reduce patient discomfort and improve throughput, this could lead to reduced sampling which in turn could compromise the integrity of the reconstructed images. In addition to undersampling, certain techniques may be susceptible to noise and other artifacts which could contaminate the images [9,10]. Challenges pertaining to reconstructing multi-contrast images have also been discussed by these papers [11,12]. These multi-contrast images are often considered to provide rich and more useful information as they involve combining different types of image modalities and MRI sequences. However, reconstructing these images may be computationally expensive with traditional methods and may also involve managing large and complex datasets. Recent studies have proposed numerous deep-learning models to address these challenges.

Deep learning has seen extensive applications in image processing tasks [13–17] 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 networks (CNNs) are also extensively utilized in MRI reconstruction due to their proficiency in handling complex patterns and noise inherent in MRI data [18–27]. 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. Additionally, the deep learning models may be particularly suited for reconstructing the MRI data. This is because these optimization-based models are designed to iteratively refine their outputs, thereby allowing them to efficiently handle the complex inverse problems associated with the MRI reconstruction. Furthermore, deep learning models can effectively incorporate prior knowledge such as sparsity or smoothness constraints, which are essential for reconstructing high-quality images from undersampled data. Owing to the flexibility available during the model design, it is possible to design models that can enforce the physics of MRI acquisition, such as Fourier encoding of spatial information. This would ensure that the reconstructed images are consistent with the underlying data.

In recent years, optimization-based algorithm unrolling networks have gained significant attention in the field of MRI reconstruction [27–37]. These algorithms are inspired by classical optimization techniques and are designed to address the unique challenges posed by MRI data. One notable development in this area is the introduction of learnable optimization algorithms (LOAs). LOAs enhance the interpretability of deep learning models by incorporating MR physics, thereby improving both model performance and training efficiency. The convergence properties of these LOAs can support the fast convergence of the reconstruction process and speed up the model training [38,39]. This paper explores several key approaches within this framework, including gradient descent and proximal gradient descent algorithm-inspired networks, variational networks, iterative shrinkage-thresholding algorithm (ISTA) networks, and alternating direction method of multipliers (ADMM)-inspired networks. These methods leverage iterative optimization techniques to refine MRI reconstructions, effectively reducing artifacts and enhancing image clarity. Additionally, the integration of diffusion models, such as the score-based diffusion model and Domain-conditioned diffusion modeling, represents a novel approach that combines deep learning with diffusion processes to more robustly tackle undersampling issues. LOAs, as a subset of physics-driven machine learning methods [40], explicitly

incorporate known physics-based forward imaging models into deep learning architectures. This integration ensures consistency with k-space measurements during the reconstruction process, offering a comprehensive framework for improving the speed and accuracy of MRI scans and advancing the application of machine learning in medical imaging. 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 methods, such as the generalized auto-calibrating partially parallel acquisition (GRAPPA) [4] and ESPIRiT [41], are k-space techniques that focus on manipulating or reconstructing the k-space data before converting it into the image domains using an inverse Fourier transform [42]. These methods utilize coil-by-coil auto-calibration to achieve accurate reconstruction.

On the other hand, compressed sensing (CS) exploits the sparsity of MR images in a specific transform domain (e.g., wavelet or total variation) to reconstruct images. For CS to be effective, it requires incoherent sampling, which helps to spread the aliasing artifacts in the image domain in a way that makes them easier to remove. CS is primarily applied in the image domain, removing aliasing artifacts by solving a system of equations that relate the image to be reconstructed and the partial k-space data through coil sensitivities. An example of this approach is sensitivity encoding (SENSE) [3].

This paper focuses on CS-based methods and different algorithms for solving the system equations derived from them. The formulation for the MRI reconstruction problem in CS-based methods 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 choice of regularization operator plays a critical role in enforcing sparsity or low-rank constraints to stabilize the reconstruction from undersampled k-space data. A common regularization used is L_1 Norm which promotes sparse solutions by penalizing the sum of absolute values of coefficients in transformed domains. Alternatively, a low-rank or non-convex regularization may also be used depending on the application. While a low-rank regularization leverages the fact that MR images exhibit low-rank structures in certain matrix forms, the non-convex penalties provide stronger sparsity-promoting properties similar to L_p norm. However, the downside is that the non-convex regularization problems are harder to solve due to non-convexity. A practical example illustrating the effectiveness or advantage of enforcing sparsity in a clinical setting is that by enforcing sparsity in the wavelet domain, CS-based methods have been found to accelerate MRI acquisitions by 4 to 6 times while maintaining the image quality [3].

The regularization operator $\mathcal{R}:\mathbb{C}^n \to \mathbb{R}$ enforces sparsity or low-rank constraints on the MRI data, incorporating prior knowledge to guide the reconstruction and prevent overfitting of the data fidelity term. It is important to note that MR images are generally not sparse in their original spatial domain. Therefore, to effectively apply CS-based methods, the image must be transformed into a domain where it exhibits sparsity, such as the wavelet or Fourier domain. The regularization term $\mathcal{R}(\mathbf{x})$ typically enforces sparsity in this transformed domain, allowing for accurate image reconstruction from undersampled data.

The weight parameter $\mu > 0$ balances the data fidelity term and regularization term. The measurement data is typically expressed as $\mathbf{f} = \mathbf{A}\mathbf{x} + \boldsymbol{\varepsilon}$ with $\boldsymbol{\varepsilon} \in \mathbb{C}^m$ 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 $\mathcal{S} := [\mathcal{S}_1, \ldots, \mathcal{S}_c]$ 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. Figure 1 shows the image reconstruction diagram.
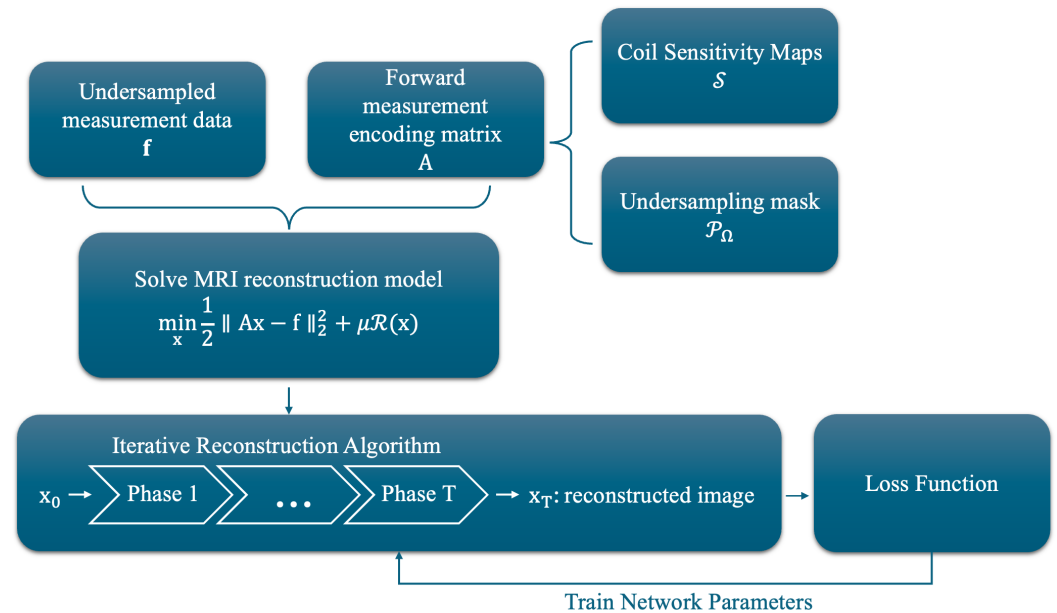

Figure 1. Demonstration of MRI reconstruction process.

Optimization-based reconstruction methods encompass a variety of techniques for solving complex problems such as (1). These include gradient descent methods like steepest descent [43], proximal methods such as ADMM for non-smooth optimization, interior point methods for constrained problems, and Newton-type methods for faster convergence. Other approaches include iterative shrinkage-thresholding algorithms for sparse reconstruction, coordinate descent for large-scale problems, stochastic methods like SGD for machine learning applications, and primal-dual methods that optimize in both spaces simultaneously. Each method offers unique advantages, making them suitable for different types of reconstruction problems based on factors such as problem structure, size, and computational requirements.

3. Optimization-Based Network Unrolling Algorithms for MRI Reconstruction

The deep learning-based model has the capability to leverage large datasets and further improve reconstruction performance compared to traditional methods and therefore has had successful applications in clinical fields [18–24,44,45]. Most existing deep learning-

based methods employ end-to-end neural networks that either map partial k-space data directly to reconstructed images [46–50], or map partial k-space data to an estimated fully sampled k-space such as RAKI [51] and Grappa-Net [52]. By incorporating optimization algorithms in the end-to-end training, both the acquisition scanning times and image reconstruction time can be drastically reduced to reconstruct high-quality images from undersampled k-space data. This could allow for faster imaging without compromising the MRI's diagnostic accuracy [51–53]. 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., [27–37,54,55]. 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 re-learned. 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 [54], ISTA-Net$^+$ [56], and cascade network [19] 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 its identity. Variational networks (VNs) [18] introduced the gradient descent method with given (pre-calculated) sensitivities $\mathcal{S}$. MoDL [57] proposed a recursive network by unrolling the conjugate gradient algorithm using a weight-sharing strategy. Blind-PMRI-Net [58] 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 [59] 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. [60] proposed a reconstruction network that unrolled a primal-dual algorithm where the proximal operator is learnable.

3.1. Gradient Descent Algorithm-Inspired Network

3.1.1. Variational Network

The use of a variational network (VN) [18] solves model (1) by using gradient descent:

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

This model was applied to multi-coil MRI reconstruction. The regularization term was defined by the field-of-experts 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 non-linear 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 non-linear activation function $\mathcal{H}_i$:

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

Figure 2 shows the iterative process of the reconstruction algorithm of VN.
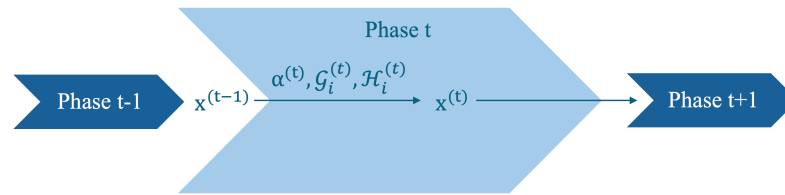
Figure 2. Network architecture of VN at *t*-th phase, illustrating the updates of $\mathbf{x}^{(t)}$.

Owing to the numerous advantages of CNNs, they can function as implicit regularizers, replacing traditional techniques such as L_2 regularization or dropout layers. CNNs excel at feature extraction and enable weight sharing, which reduces the number of parameters compared to fully connected layers, making the model less susceptible to overfitting. Additionally, the hierarchical learning process of CNNs imposes a structured learning approach, naturally limiting the network's complexity and providing an implicit regularization effect. Learning regularization terms from training data is becoming a popular trend for solving inverse problems. Some methods were developed via a hybrid domain-specific learning model. For instance, E2EVarNet [61] performs iterative optimization steps in the k-space domain and uses CNNs to learn the gradient of the regularization term in the image domain within each cascade iteratively.

$$\mathbf{f}^{(t+1)} = \mathbf{f}^{(t)} - \alpha^{(t)}\mathcal{P}_\Omega(\mathbf{f}^{(t)} - \mathbf{f}) + \mathcal{W}(\mathbf{f}^{(t)}), \tag{5a}$$

$$\text{where} \quad \mathcal{W}(\mathbf{f}^{(t)}) = \mathcal{F} \circ \mathcal{J} \circ \mathcal{H}_\Theta \circ \tilde{\mathcal{J}} \circ \mathcal{F}^\top(\mathbf{f}^{(t)}). \tag{5b}$$

where $\mathcal{H}_\Theta$ is a CNN applied on the complex image, $\mathcal{J}$ and $\tilde{\mathcal{J}}$ are expand and reduce operators that take care of coil sensitivity maps, respectively. Recurrent VarNet [62] utilizes recurrent neural networks (RNN) for learning the image-refinement model $\mathcal{H}_\Theta$. The coil sensitivity maps are estimated and refined during the training phase. It has strong performance in refining image quality and sensitivity maps, particularly in multi-coil MRI. It is also effective in long-term sequence data, but is computationally intensive. Additionally, other variations of VN [61,63] have also been developed. E2E-VarNet [61] employs hybrid domain learning, combining both k-space and image domain approaches through iterative optimization steps using gradient descent. While this method delivers excellent performance in reconstructing undersampled MRI data in both domains, it comes with a complex training process that demands substantial computational resources. Despite its effectiveness, the high computational requirements for training and deployment present a significant challenge. These works demonstrate the potential of combining variational methods with deep learning for solving complex inverse problems in medical imaging.

3.1.2. Denoising Model-Based Regularizations

A group of optimization models employed denoising model-based regularizations [57,64,65]. They help in the reconstruction of high-quality images by iteratively refining the image while reducing noise and preserving important image details. These types of models balance the reconstruction between fitting the observed data and adhering to the learned priors about noise and artifact patterns. The model-based deep learning (MoDL) [57] framework incorporates a CNN-based regularization prior within a model-based reconstruction scheme. This framework unrolls an alternating recursive algorithm to solve a variational model, where the regularization term is learnable and designed to estimate the noise and alias artifacts:

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{f}\|_2^2 + \mu\|\mathbf{x} - \mathcal{D}_\Theta(\mathbf{x})\|_2^2. \tag{6}$$

In Equation (6), the non-linear denoising operator $\mathcal{D}_\Theta(\mathbf{x})$ parameterized by learnable variables Θ is trained to eliminate noise and artifacts from the image. The regularization term ensures that the reconstructed image $\mathbf{x}$ closely approximates the denoised version

provided by $\mathcal{D}_\Theta(\mathbf{x})$. This regularized optimization model is solved by using the normal equations and iterates the following steps:

$$\mathbf{x}^{(t)} = \frac{\mathbf{A}^\top \mathbf{f} + \lambda \mathcal{Z}^{(t-1)}}{\mathbf{A}^\top \mathbf{A} + \lambda I}. \tag{7a}$$

$$\mathcal{Z}^{(t)} = \mathcal{D}_\Theta(\mathbf{x}^{(t)}). \tag{7b}$$

When dealing with single-channel MRI data, the step (7a) has an analytical solution which can be expressed as a data consistency (DC) layer, where $\mathbf{x}^{(t)} = \mathcal{F}^\top \mathbf{f}_{DC}^{(t)}$, and

$$\mathbf{f}_{DC}^{(t)}[i] = \begin{cases} \frac{(\mathbf{f}[i] + \lambda \mathcal{Z}^{(t-1)}[i])}{1+\lambda} & \text{if } i\text{th sample is acquired,} \\ \mathcal{Z}^{(t-1)}[i] & \text{else.} \end{cases} \tag{8}$$

For the multi-coil MRI data, (7a) is solved using a conjugate gradient optimization algorithm. Figure 3 shows the iterative process of the reconstruction algorithm of VN.
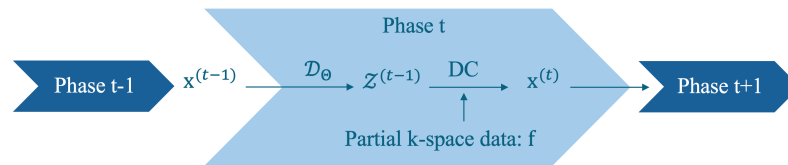
Figure 3. Network architecture of MoDL at t-th phase, illustrating the updates of $\mathbf{x}^{(t)}$.

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 [27,39,66–76].

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:

$$\mathbf{b}_t = \mathbf{x}_t - \rho_t \mathbf{A}^\top (\mathbf{A}\mathbf{x}_t - \mathbf{f}), \tag{9a}$$

$$\mathbf{x}_{t+1} = \text{prox}_{\rho_t \mathcal{R}(\cdot)}(\mathbf{b}_t), \tag{9b}$$

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

$$\text{prox}_{\rho \mathcal{R}}(\mathbf{b}) = \arg\min_{\mathbf{x}} \frac{1}{2\rho} \|\mathbf{x} - \mathbf{b}\|_2^2 + \mathcal{R}(\mathbf{x}). \tag{10}$$

The gradient update step (9a) 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 $\sqrt{\rho}$ [77,78]. 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 a closed-form solution, a CNN can be used to substitute $\text{prox}_{\rho \mathcal{R}}$. Constructing the network with residual learning [66,71,76] 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. [66] 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.

The step size ρ_t plays a crucial role in determining the convergence and performance of the proximal gradient descent algorithm. LOAs usually apply a learnable step size which adds a layer of adaptability to the optimization process, allowing the algorithm to adjust dynamically to the specific characteristics of the data and the model at each iteration. Traditional fixed step sizes may be either too small, leading to slow convergence, or too large, potentially causing the algorithm to overshoot and oscillate. In contrast, a learnable step size can adjust itself based on the gradient's magnitude and direction, promoting a more stable and faster convergence. By fine-tuning the step size during the training process, the algorithm can navigate the loss landscape more efficiently, avoiding regions where a fixed step size might struggle. The performance of the proximal gradient descent algorithm in MRI reconstruction is closely tied to how well it can minimize the objective function. A learnable step size has the flexibility that helps in striking a balance between convergence speed and reconstruction accuracy, leading to better overall performance. Additionally, it allows the model to adapt to different noise levels and data inconsistencies, which are common in MRI tasks, further improving the robustness of the reconstruction.

3.2.1. Iterative Shrinkage-Thresholding Algorithm (ISTA) Network

ISTA-Net$^+$ [56] 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 (9) become:

$$\mathbf{b}_t = \mathbf{x}_t - \rho_t \mathbf{A}^\top (\mathbf{A}\mathbf{x}_t - \mathbf{f}), \tag{11a}$$

$$\mathbf{x}_{t+1} = \arg\min_{\mathbf{x}} \frac{1}{2\rho_t} \|\mathbf{x} - \mathbf{b}_t\|_2^2 + \|\varphi(\mathbf{x})\|_1. \tag{11b}$$

The proximal step (11b) can be parameterized as an implicit residual update step due to the lack of a closed-form solution:

$$\mathbf{x}_{t+1} = \mathbf{b}_t + \mathcal{H}(\mathbf{b}_t), \tag{12}$$

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 (11b) was written as

$$\mathbf{x}_{t+1} \approx \arg\min_{\mathbf{x}} \frac{1}{2} \|\varphi(\mathbf{x}) - \varphi(\mathbf{b}_t)\|_2^2 + \delta \rho_t \|\varphi(\mathbf{x})\|_1. \tag{13}$$

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

$$\mathbf{x}_{t+1} = \tilde{\varphi}(\mathcal{S}_{\beta_t}(\varphi(\mathbf{b}_t))), \tag{14}$$

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 (12) is reduced to an explicit form as given in (14), which we summarize together with (11a) in the following scheme:

$$\mathbf{b}_t = \mathbf{x}_t - \rho_t \mathbf{A}^\top (\mathbf{A}\mathbf{x}_t - \mathbf{f}), \tag{15a}$$

$$\mathbf{x}_t = \mathbf{b}_t + \tilde{\varphi}_t(\mathcal{S}_{\beta_t}(\varphi_t(\mathbf{b}_t))). \tag{15b}$$

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:

$$\mathcal{L}_{dis}(\Theta) = \frac{1}{2} \|\mathbf{x}_T(\mathbf{f}, \Theta) - \mathbf{x}^*\|_2^2 \tag{16}$$

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:

$$\mathcal{L}_{id}(\Theta) = \frac{1}{2}\sum_{t=1}^{T}\|\tilde{\varphi}_t(\varphi_t(\mathbf{x}^*)) - \mathbf{x}^*\|_2^2 \tag{17}$$

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

$$\mathcal{L}_{loss} = \mathcal{L}_{dis}(\Theta) + \mu\mathcal{L}_{id}(\Theta). \tag{18}$$

where μ is a balancing parameter. Figure 4 shows the iterative process of the reconstruction algorithm of VN.
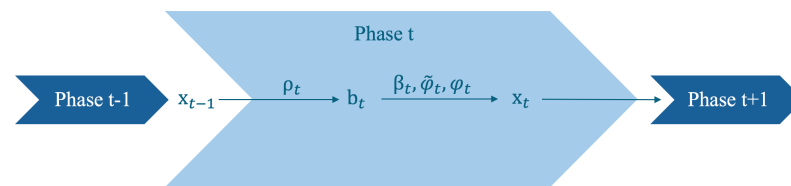
Figure 4. Network architecture of ISTA-Net at the *t*-th phase, illustrating the updates of $\mathbf{b}^{(t)}$ and $\mathbf{x}^{(t)}$.

The derivation of the approximation theorem in ISTA-Net+ relies on several critical assumptions including orthogonality and invertibility of non-linear transform (φ). One of the primary assumptions is that φ is the orthogonal means that the transform φ satisfies:

$$\varphi^T\varphi = I, \tag{19}$$

where I is the identity matrix. Another assumption is that φ is invertible and it ensures that there exists a unique inverse transform. This assumption ensures that no information is lost and allows for a complete reconstruction of the original signal after transformation. The final assumption is the mean value theorem application. The approximation theorem assumes that φ satisfies the conditions necessary for applying the mean value theorem. Discussing the practical implications of these assumptions, the orthogonality assumption allows for more straightforward updates leading to a closed-form solution. Similarly, the invertibility is an important requirement to ensure that the transformation φ does not lose any critical information. Finally, the mean value theorem provides a convenient way to approximate the relationship between the transformed variables.

3.2.2. Parallel MRI Network

Parallel MRI networks [71] leverage residual learning to learn the proximal mapping and tackle model (9), thus bypassing the requirement for pre-calculated coil sensitivity maps in the encoding matrix (2). Similar to the model for joint reconstruction and synthesis [70], parallel MRI networks consider the MRI reconstruction as a bi-level optimization problem:

$$\min_{\Theta}\ \ell(\mathbf{x}_{\Theta}, \mathbf{x}^*), \tag{20a}$$

$$\text{s.t.}\ \mathbf{x}_{\Theta} = \arg\min_{\mathbf{x}}\Phi_{\Theta}(\mathbf{x}). \tag{20b}$$

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:

$$\Phi(\mathbf{x}) := \frac{1}{2}\sum_{i=1}^{c}\|\mathcal{P}_\Omega\mathcal{F}\mathbf{x}_i - \mathbf{f}_i\|_2^2 + \mathcal{R}(\mathcal{J}(\mathbf{x})) + \mathcal{R}_f(\mathcal{F}\mathbf{x}_i). \tag{21}$$

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 (20a) 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 (20b) is solved by the following redefined algorithm:

$$\mathbf{b}_i^{(t)} = \mathbf{x}_i^{(t)} - \rho_t \mathcal{F}^{-1}\mathcal{P}_\Omega^\top(\mathcal{P}_\Omega\mathcal{F}\mathbf{x}_i^{(t)} - \mathbf{f}_i), \quad i = 1, \cdots, c, \tag{22a}$$

$$\bar{\mathbf{x}}_i^{(t)} = [\text{prox}_{\rho_t\mathcal{R}(\mathcal{J}(\cdot))}(\mathbf{b}^{(t)})]_i, \quad i = 1, \cdots, c, \tag{22b}$$

$$\mathbf{x}_i^{(t+1)} = \text{prox}_{\rho_t\mathcal{R}_f(\mathcal{F}(\cdot))}(\bar{\mathbf{x}}_i^{(t)}), \quad i = 1, \cdots, c. \tag{22c}$$

The proximal operator can be understood as a Gaussian denoiser. Nevertheless, the proximal operator $\text{prox}_{\rho_t\mathcal{R}}$ in the objective function (22b) 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 (22a), (22b) and (22c) are implemented in the following scheme:

$$\mathbf{b}_i^{(t)} = \mathbf{x}_i^{(t)} - \rho_t \mathcal{F}^{-1}\mathcal{P}_\Omega^\top(\mathcal{P}_\Omega\mathcal{F}\mathbf{x}_i^{(t)} - \mathbf{f}_i), \quad i = 1, \cdots, c, \tag{23a}$$

$$\bar{\mathbf{x}}_i^{(t)} = \mathbf{b}_i(t) + \phi_t(\mathbf{b}_i^{(t)}), \quad i = 1, \cdots, c, \tag{23b}$$

$$\mathbf{x}_i^{(t+1)} = \bar{\mathbf{x}}_i(t) + \mathcal{F}^{-1}\varphi_t(\mathcal{F}(\bar{\mathbf{x}}_i^{(t)})), \quad i = 1, \cdots, c. \tag{23c}$$

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 (23c), the k-space network φ focuses on low-frequency data, helping to remove high-frequency artifacts and restore image structure. Alternating between (23b) and (23c) in their respective domains balances their strengths and weaknesses, improving overall performance. Figure 5 shows the iterative process of reconstruction algorithm of the parallel MRI network.
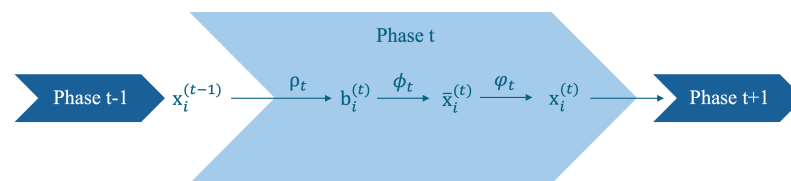
Figure 5. Architecture of parallel MRI network at the t-th phase, illustrating the updates of $\mathbf{b}^{(t)}, \bar{\mathbf{x}}_i^{(t)}$ and $\mathbf{x}^{(t)}$.

The construction of iterative algorithm (23) is inspired by cross-domain reconstruction methods [79–82], which aimed at improving the quality and speed of MRI image reconstruction by leveraging information from multiple domains—typically the image domain and the k-space (frequency) domain. This approach integrates data and knowledge across different domains to enhance the accuracy and efficiency of reconstructing high-quality images from undersampled MRI data.

Learning sensitivity maps also represent a related group of reconstruction techniques [61,83–86]. Deep J-Sense builds on the joint SENSE model, which solves for both the image and coil sensitivity maps simultaneously during MRI reconstruction. This work

incorporates unrolled alternating optimization to refine both the magnetization (image) kernel and the sensitivity maps iteratively. The optimization problem is defined as

$$\min_{\mathbf{S},\mathbf{x}} \frac{1}{2}\|\mathbf{f} - \mathcal{P}_\Omega\mathcal{F}(\mathbf{S}*\mathbf{x})\|_2^2 + \lambda_x R_x(\mathbf{x}) + \lambda_s R_s(\mathbf{S}), \tag{24}$$

$R_x(\mathbf{x}), R_s(\mathbf{S})$ are regularization terms for the image and sensitivity maps, respectively. λ_x and λ_s are regularization parameters. This optimization problem is solved iteratively by alternating between updating the image $\mathbf{x}$ and the sensitivity maps $\mathbf{S}$.

For the sensitivity maps update:

$$\mathbf{S}_{t+1} = \arg\min_{\mathbf{S}} \frac{1}{2}\|\mathbf{f} - \mathcal{P}_\Omega\mathcal{F}(\mathbf{S}*\mathbf{x})\|_2^2 + \lambda_s R_s(\mathbf{S}). \tag{25a}$$

$$\mathbf{S}^+ = \mathcal{D}_s(\mathbf{S}) \tag{25b}$$

For the image update:

$$\mathbf{x}_{t+1} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{f} - \mathcal{P}_\Omega\mathcal{F}(\mathbf{S}*\mathbf{x})\|_2^2 + \lambda_x R_x(\mathbf{x}). \tag{26a}$$

$$\mathbf{x}^+ = \mathcal{D}_x(\mathbf{x}) \tag{26b}$$

The end-to-end deep neural networks $\mathcal{D}_s$ and $\mathcal{D}_x$ are applied to refine the sensitivity maps and refine the image, respectively.

Parallel MRI 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. Self-Supervised Approaches for Quantitative MRI Reconstruction

Recent advancements in quantitative MRI (qMRI) reconstruction have seen the incorporation of self-supervised learning techniques, which have proven effective in reconstructing quantitative mapping from the undersampled k-space MRI data. Among these, the RELAX-MORE [76] algorithm has introduced a self-supervised learning framework that optimizes reconstruction by leveraging the underlying physics of MRI signal acquisition.

RELAX-MORE introduced an optimization algorithm to unroll the proximal gradient for qMRI reconstruction. The loss function minimizes the discrepancy between undersampled reconstructed MRI k-space data and the "true" undersampled k-space data retrospectively. The model, once thoroughly trained, can be adapted to other testing data through the use of transfer learning.

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

$$\min_{\Theta} \ell(\mathcal{P}_\Omega\mathcal{F}\mathcal{S}\mathcal{M}(\mathbf{P}(\mathbf{f}|\Theta)), \mathbf{f}) \quad \text{s.t.} \tag{27a}$$

$$\mathbf{P}(\mathbf{f}|\Theta) = \arg\min_{\mathbf{P}} \mathcal{K}_\Theta(\mathbf{P}), \tag{27b}$$

$$\text{where} \quad \mathcal{K}_\Theta(\mathbf{P}) := \frac{1}{2}\|\mathcal{P}_\Omega\mathcal{F}\mathcal{S}\mathcal{M}(\mathbf{P}) - \mathbf{f}\|_2^2 + \beta\mathcal{R}_\Theta(\mathbf{P}). \tag{27c}$$

The model $\mathcal{M}$ represents the MR signal function that maps the set of quantitative parameters $\mathbf{P} := \{\boldsymbol{\rho}_1, \cdots, \boldsymbol{\rho}_N\}$ to the MRI data. The loss function in (27a) 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 (27a) focuses on optimizing the learnable parameters for network training, while the lower-level problem (27b) concentrates on optimizing the quantitative MR parameters. RELAX-MORE uses $\mathbf{T}_1$ mapping obtained through the variable flip angle (vFA) method [87] as an example. The MR signal model $\mathcal{M}$ is described by the following equation:

$$\mathcal{M}_k(\mathbf{T}_1, \mathbf{I}_0) = \mathbf{I}_0 \cdot \frac{(1 - e^{-TR/\mathbf{T}_1}) \sin \eta_k}{1 - e^{-TR/\mathbf{T}_1} \cos \eta_k}, \tag{28}$$

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 needed for reconstruction is $\mathbf{P} = \{\mathbf{T}_1, \mathbf{I}_0\}$.

Similar to the parallel MRI network [71], RELAX-MORE employs a proximal gradient descent algorithm to address the lower level problem (27b), with a residual network structure designed to learn the proximal mapping. Below is the unrolled learnable Algorithm 1 for resolving (27b):

Algorithm 1 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:          $\bar{\rho}_i^{(t)} = \rho_i^{(t-1)} - \mu_i^{(t)} \nabla_{\rho_i} \frac{1}{2} \| \mathcal{P}_\Omega \mathcal{FSM}(\{\rho_i^{(t-1)}\}_{i=1}^N) - \mathbf{f} \|_2^2$,
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 $\mathcal{W}_\Theta$ have a 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:

$$\min_{\mathbf{x}, \mathbf{v}} \frac{1}{2} \| \mathbf{Ax} - \mathbf{f} \|_2^2 + \mu \mathcal{R}(\mathbf{v}) \tag{29a}$$

$$\text{s.t. } \mathbf{v} = \mathbf{Dx}, \tag{29b}$$

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:

$$L_\rho(\mathbf{x}, \mathbf{v}, \lambda) = \frac{1}{2} \| \mathbf{Ax} - \mathbf{f} \|_2^2 + \mu \mathcal{R}(\mathbf{v}) + <\lambda, \mathbf{Dx} - \mathbf{v}> + \frac{\rho}{2} \| \mathbf{Dx} - \mathbf{v} \|_2^2, \tag{30}$$

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

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

$$\mathbf{x}_{t+1} = \arg\min_{\mathbf{x}} \frac{1}{2} \| \mathbf{Ax} - \mathbf{f} \|_2^2 + \frac{\rho}{2} \| \mathbf{Dx} - (\mathbf{v}_t - \mathbf{u}_t) \|_2^2, \tag{31a}$$

$$\mathbf{v}_{t+1} = \arg\min_{\mathbf{v}} \frac{\rho}{2} \| \mathbf{Dx}_{t+1} - (\mathbf{v} - \mathbf{u}_t) \|_2^2 + \mu \mathcal{R}(\mathbf{v}), \tag{31b}$$

$$\mathbf{u}_{t+1} = \mathbf{u}_t + (\mathbf{Dx}_{t+1} - \mathbf{v}_t). \tag{31c}$$

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

$$\mathbf{x}_{t+1} = (\mathbf{A}^\top \mathbf{A} + \rho \mathbf{D}^\top \mathbf{D})^{-1}(\mathbf{A}^\top \mathbf{f} + \rho \mathbf{D}^\top(\mathbf{v}_t - \mathbf{u}_t)), \tag{32a}$$

$$\mathbf{v}_{t+1} = \mathrm{prox}_{\mu/\rho}(\mathbf{D}\mathbf{x}_{t+1} + \mathbf{u}_t), \tag{32b}$$

$$\mathbf{u}_{t+1} = \mathbf{u}_t + (\mathbf{D}\mathbf{x}_{t+1} - \mathbf{v}_{t+1}). \tag{32c}$$

If the regularizer is l_1 norm $\mathcal{R}(\mathbf{v}) = \|\mathbf{v}\|_1$, then (32b) reduces to $\mathbf{v}_{t+1} = \mathcal{S}_\beta(\mathbf{D}\mathbf{x}_{t+1} + \mathbf{u}_t)$ with the soft shrinkage threshold $\beta = \mu/\rho$.

In the ADMM, tuning penalty and regularization parameters play a crucial role in the algorithm's performance—particularly the penalty parameter ρ and the augmented Lagrangian parameter μ. These parameters significantly influence the convergence and performance of the algorithm. A well-chosen μ can accelerate convergence by appropriately weighting the constraints in the problem. However, if μ is too large, the algorithm may become unstable or oscillate between iterations. Conversely, if μ is too small, convergence can be excessively slow. The parameter ρ controls the weight of the augmented Lagrangian term in ADMM. It influences the convergence speed and stability of the algorithm. A large ρ can enforce constraints more strictly but may lead to numerical instability, while a small ρ might result in slower convergence. Tuning parameters typically rely on heuristic methods or cross-validation to find a suitable value. One approach to mitigating the challenges of parameter selection is to adaptively update ρ and μ during the iterations based on the observed residuals. This strategy can help balance the trade-off between convergence speed and stability, but it adds complexity to the algorithm.

Gradient descent is simple and widely used but it suffers from slow convergence, particularly in ill-conditioned problems. ADMM can converge faster for certain problem classes, especially when the objective function can be split into simpler subproblems. Proximal gradient descent extends gradient descent by incorporating proximal operators, making it suitable for optimization problems with non-smooth terms. ADMM can be seen as a more general approach that also leverages proximal operators but in a way that allows for better decomposition of the problem.
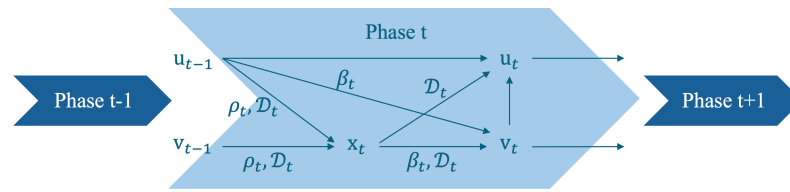
ADMM-Net

ADMM-Net [54] 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 a deep neural network $\mathcal{D}$. All the scalars μ and ρ are learnable parameters to be trained and updated through ADMM iterations:

$$\mathbf{x}_t = (\mathbf{A}^\top \mathbf{A} + \rho_t \mathcal{D}_t^\top \mathcal{D}_t)^{-1}(\mathbf{A}^\top \mathbf{f} + \rho_t \mathcal{D}_t^\top(\mathbf{v}_{t-1} - \mathbf{u}_{t-1})), \tag{33a}$$

$$\mathbf{v}_t = \mathcal{S}_{\beta_t}(\mathcal{D}_t \mathbf{x}_t + \mathbf{u}_{t-1}), \tag{33b}$$

$$\mathbf{u}_t = \mathbf{u}_{t-1} + (\mathcal{D}_t \mathbf{x}_t - \mathbf{v}_t). \tag{33c}$$

The Reconstruction layer (33a) 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 (33b) 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 (33c) 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. Figure 6 shows the iterative process of the reconstruction algorithm of the parallel MRI network.

Figure 6. Network architecture of ADMM-Net at the *t*-th phase, illustrating the updates of $\mathbf{x}_t, \mathbf{v}_t$ and $\mathbf{u}_t$.

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

There are several networks [60,88] are developed inspired by the PDHG algorithm. PDHG can be used to solve the model (1) by iterating the following steps:

$$\mathbf{d}_{t+1} = \text{prox}_{\zeta_t H^\top(\cdot)}(\mathbf{d}_t + \zeta_t \mathbf{A}\bar{\mathbf{x}}_t), \tag{34a}$$

$$\mathbf{x}_{t+1} = \text{prox}_{\rho_t \mathcal{R}(\cdot)}(\mathbf{x}_t + \eta_t \mathbf{A}^\top \mathbf{d}_{t+1}), \tag{34b}$$

$$\bar{\mathbf{x}}_{t+1} = \mathbf{x}_{t+1} + \theta(\mathbf{x}_{t+1} - \mathbf{x}_t), \tag{34c}$$

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 primal-dual model [60], 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 [88] iterates the following two steps:

$$\mathbf{d}_{t+1} = \mathcal{G}_{\theta'}(\mathbf{d}_t, \zeta_t \mathbf{A}\mathbf{x}_t, \mathbf{f}), \tag{35a}$$

$$\mathbf{x}_{t+1} = \mathcal{K}_{\theta''}(\mathbf{x}_t, \mathbf{A}^\top \mathbf{d}_{t+1}), \tag{35b}$$

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 primal-dual model 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. Figure 7 shows the iterative process of the reconstruction algorithm of a parallel MRI network.
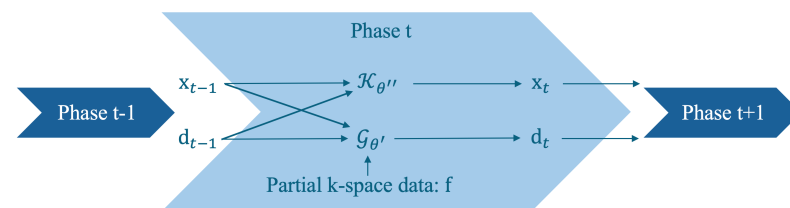
Figure 7. Network architecture of PD-Net at the *t*-th phase, illustrating the updates of $\mathbf{d}_t, \mathbf{x}_t$.

3.5. Diffusion Models Meet Gradient Descent for MRI Reconstruction

A notable development for MRI reconstruction using diffusion models is the emergence of denoising diffusion probabilistic models (DDPMs) [89–92]. 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 [93] 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 [89–92] 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 [90] has proven exceptionally robust, producing high-quality images even when faced with data that deviates from the training set (distribution shifts) [94], 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. [89] 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:

$$\mathcal{L}(\Theta) = \mathbb{E}_{\mathbf{x}(t) \sim p(\mathbf{x}(t)|\mathbf{x}(0)), \mathbf{x}(0) \sim p_{data}} \left[\| S_\Theta(\mathbf{x}(t), t) - \nabla_{\mathbf{x}(t)} \log p_t(\mathbf{x}(t)|\mathbf{x}(0)) \|_2^2 \right]. \quad (36)$$

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. [89] proposed the following Algorithm (2) with the predictor-corrector (PC) sampling algorithm [95]. 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 2:

Algorithm 2 Score-based sampling for MRI reconstruction [89]

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** $i = N - 1, \ldots, 0$ **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** $j = 1, \ldots, M$ **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 2 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, including UNet [96], DuDoRNet [97], and E2E-Varnet [61], which requires complex k-space data.

3.5.2. Domain-Conditioned Diffusion Modeling

Domain-conditioned diffusion modeling (DiMo) [73] and quantitative DiMo were developed for application 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 Algorithms 3 and 4.

Algorithm 3 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 $\mathcal{S}$.
    Initialization : η_0
1: $\hat{\mathbf{f}}_t \leftarrow \sqrt{\bar{\alpha}_t}\hat{\mathbf{f}}_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon$.
2: $\hat{\mathbf{f}}_t \leftarrow \mathcal{P}_\Omega(\lambda_t\mathbf{f} + (1 - \lambda_t)\hat{\mathbf{f}}_t) + (\mathbb{1} - \mathcal{P}_\Omega)\hat{\mathbf{f}}_t$ $\triangleright$ DC
3: **for** $k = 0$ to $K - 1$ **do**
4:    $\hat{\mathbf{f}}_t \leftarrow \hat{\mathbf{f}}_t - \eta_k\nabla_{\hat{\mathbf{f}}_t}\frac{1}{2}\|\mathbf{A}\mathcal{F}^{-1}\hat{\mathbf{f}}_t - \mathbf{f}\|_2^2$ $\triangleright$ 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 4 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** $t = T - 1, \ldots, 0$ **do**
2:    $z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 0$, else $z = 0$
3:    $\hat{\mathbf{f}}_t = \mu_\theta(\hat{\mathbf{f}}_{t+1}, t + 1) + \sigma_{t+1}z$
4:    $\hat{\mathbf{f}}_t \leftarrow \mathcal{P}_\Omega(\lambda_t\mathbf{f} + (1 - \lambda_t)\hat{\mathbf{f}}_t) + (\mathbb{1} - \mathcal{P}_\Omega)\hat{\mathbf{f}}_t$ $\triangleright$ DC
5:    **for** $k = 0$ to $K - 1$ **do**
6:       $\hat{\mathbf{f}}_t \leftarrow \hat{\mathbf{f}}_t - \eta_k\nabla_{\hat{\mathbf{f}}_t}\frac{1}{2}\|\mathbf{A}\mathcal{F}^{-1}\hat{\mathbf{f}}_t - \mathbf{f}\|_2^2$ $\triangleright$ GD
7:    **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, the gradient descent (GD) algorithm is applied iteratively into the diffusion step to refine k-space data further. The matrix $\mathbb{1}$ only contains a value of one. GD in here solves the optimization problem (1) without the regularization term.

Static DiMo performed a qualitative comparison with both the image domain diffusion model presented by Chung et al. [89] and the k-space domain diffusion model MC-DDPM [98] 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 (28) maps MR parameter maps to the static MRI; therefore, it is one more function inside the reconstruction model:

$$\min_{\Delta} \frac{1}{2} \|\mathbf{A}\mathcal{M}(\Delta) - \mathbf{f}\|_2^2. \tag{37}$$

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 [99–101]. 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.

3.6. Bi-Level Optimization Model for Multi-Task Learning

In recent years, a large amount of work introduced the customized variational model for multi-task learning using bi-level optimization models. For example, joint tasks of reconstruction and multi-contrast synthesis [70] and meta-learning model for MRI reconstruction [39].

Consider a clinical situation where a patient with suspected multiple sclerosis (MS) undergoes an MRI scan. The clinician requires both T1-weighted and T2-weighted images to assess different tissue characteristics—T1 for detailed anatomical structures and T2 for detecting lesions with high water content, commonly associated with MS. Typically, reconstructing these sequences separately is time-consuming due to the high-resolution requirements. Concurrently, there is a need to synthesize a FLAIR (Fluid-Attenuated Inversion Recovery) image, which is crucial for suppressing cerebrospinal fluid signals and enhancing the visibility of lesions. Given the urgency of diagnosis and the need for comprehensive imaging, a joint reconstruction and synthesis model can be employed. This model not only reconstructs the T1 and T2 sequences simultaneously, thereby reducing overall scan and processing time, but also synthesizes the FLAIR image directly from the acquired data. By leveraging the complementary information between the T1 and T2 sequences, the model ensures that the synthesized FLAIR image is consistent and reliable, providing the clinician with a complete set of images for accurate diagnosis without the need for additional scans.

A provable learnable optimization algorithm [70] was introduced 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:

$$\min_{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3} \Phi_{\Theta, \gamma}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) := \frac{1}{2} \sum_{i=1}^{2} \|\mathbf{A}_i \mathbf{x}_i - \mathbf{f}_i\|_2^2 + \frac{1}{3} \sum_{i=1}^{3} \|f_{w_i}(\mathbf{x}_i)\|_{2,1}$$
$$+ \frac{\gamma}{2} \|\zeta_\theta([f_{w_1}(\mathbf{x}_1), f_{w_2}(\mathbf{x}_2)]) - \mathbf{x}_3\|_2^2. \tag{38}$$

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 5. 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 5 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 ε_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:     $\mathbf{X}_{t+1} = \mathbf{X}_t - \alpha_t \nabla \Phi^{\varepsilon}_{\Theta, \gamma}(\mathbf{X}_t)$, where the step size α_t is determined by a line search such that $\Phi^{\varepsilon}_{\Theta, \gamma}(\mathbf{X}_{t+1}) - \Phi^{\varepsilon}_{\Theta, \gamma}(\mathbf{X}_t) \leq -\frac{1}{a}\|\mathbf{X}_{t+1} - \mathbf{X}_t\|^2$ holds.

4:     **if** $\|\nabla \Phi^{\varepsilon}_{\Theta, \gamma}(\mathbf{X}_{t+1})\| < \sigma \eta \varepsilon$, 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** $\mathbf{X}^{(t)}$.

Algorithm 5 is a forward MRI reconstruction algorithm. Let $\Theta = \{w_1, w_2, w_3, \theta\}$ denote the collection of all the learnable parameters and γ denote a parameter to balance the reconstruction part and image synthesis part. The backward network training algorithm is designed to solve a bilevel optimization problem:

$$\min_{\gamma} \quad \sum_{i=1}^{\mathcal{M}_{val}} \ell(\Theta(\gamma), \gamma; \mathcal{D}_i^{val}) \quad \text{s.t.} \quad \Theta(\gamma) = \arg\min_{\Theta} \sum_{i=1}^{\mathcal{M}_{tr}} \ell(\Theta, \gamma; \mathcal{D}_i^{tr}), \tag{39}$$

$$\text{where} \quad \ell(\Theta, \gamma; \mathcal{D}_i) := \frac{\mu}{2}\|\zeta_\theta([f_{w_1}(\mathbf{x}_1^*), f_{w_2}(\mathbf{x}_2^*)]) - \mathbf{x}_3^*\|_2^2$$
$$+ \sum_{j=1}^{3}\left(\frac{1}{2}\|\mathbf{x}_{j,\hat{T}}(\Theta, \gamma; \mathcal{D}_i) - \mathbf{x}_j^*\|_2^2 + (1 - SSIM(\mathbf{x}_{j,\hat{T}}(\Theta, \gamma; \mathcal{D}_i), \mathbf{x}_j^*))\right). \tag{40}$$

The following Algorithm 6 was proposed for training the model for joint reconstruction and synthesis.

Algorithm 6 Mini-batch alternating direction penalty algorithm

---

1: **Input** Training data $\mathcal{D}^{tr}$, validation data $\mathcal{D}^{val}$, tolerance $\delta_{tol} > 0$. **Initialize** Θ, γ, δ, $\lambda > 0$ 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.

The training Algorithm 6 considers updating the network parameters Θ and the balancing parameter γ by minimizing the loss function (40) on both validation and training data sets for each task $j = 1, 2, 3$. Figure 8 shows the overall network architecture of iterating forward optimization Algorithm 5 with backward training Algorithm 6.
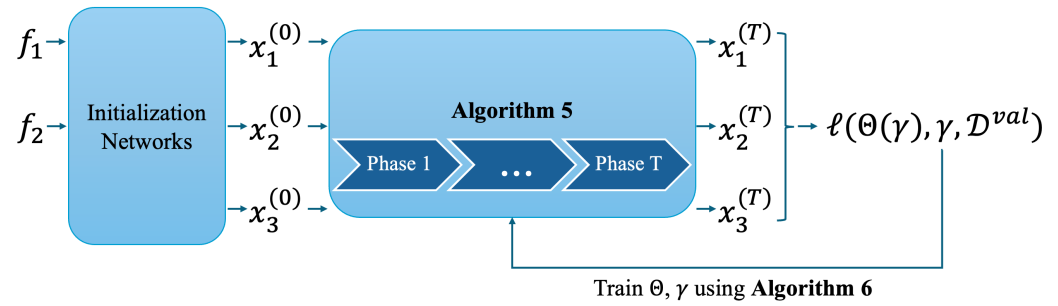
Figure 8. The model framework of joint MRI reconstruction and synthesis.

4. Discussion

4.1. Evaluation Metrics and Loss Functions

To quantitatively compare the performance of several MRI reconstruction algorithms, evaluation metrics are essential. These metrics need to be standardized in order to provide a consistent comparison between algorithms proposed in different studies and provide insights into effectiveness of the algorithms. In this paper, we introduce a few evaluation metrics used in previous studies. A common evaluation metric used across multiple studies is the root mean squared error (RMSE), which provides the square root of the average squared difference between predicted and actual images. The RMSE between the reconstruction $\mathbf{v}$ and the ground truth $\mathbf{v}^*$ is defined as

$$RMSE = \|\mathbf{v}^* - \mathbf{v}\| / \|\mathbf{v}^*\|. \tag{41}$$

$$RMSE = \sqrt{\sum_{i=1}^{c} \|\mathbf{v}_i^* - \mathbf{v}_i\|^2 / \sum_{i=1}^{c} \|\mathbf{v}_i^*\|^2}. \tag{42}$$

Studies have also discussed the peak signal-to-noise ratio (PSNR), which expresses the ratio between the maximum possible pixel value and the power of the noise. A higher PSNR values indicate better predictive quality. Typically, a PSNR of above 30 dB is considered acceptable for reconstructions. The PSNR is defined as follows.

$$PSNR = 20 \log_{10} \left(\max(|\mathbf{v}^*|) / \frac{1}{N} \|\mathbf{v}^* - \mathbf{v}\|^2 \right), \tag{43}$$

where N is the total number of pixels in the magnitude of ground truth.

Finally, the structural similarity index (SSIM) evaluates the quality of the model predictions by comparing luminance, contrast, and structure between the reconstructed and original images. A higher SSIM indicates closer structural similarity. The SSIM ranges between -1 and 1, with 1 being considered as being perfect structural similarity. The following equation calculates SSIM between reconstruction $\mathbf{v}$ and reference $\mathbf{v}^*$:

$$SSIM = \frac{(2\mu_{\mathbf{v}}\mu_{\mathbf{v}^*} + C_1)(2\sigma_{\mathbf{v}\mathbf{v}^*} + C_2)}{(\mu_{\mathbf{v}}^2 + \mu_{\mathbf{v}^*}^2 + C_1)(\sigma_{\mathbf{v}}^2 + \sigma_{\mathbf{v}^*}^2 + C_2)}, \tag{44}$$

where $\mu_{\mathbf{v}}, \mu_{\mathbf{v}^*}$ are local means of pixel intensity, $\sigma_{\mathbf{v}}, \sigma_{\mathbf{v}^*}$ denote the standard deviation and $\sigma_{\mathbf{v}\mathbf{v}^*}$ is covariance between $\mathbf{v}$ and $\mathbf{v}^*$. $C_1 = (k_1 L)^2, C_2 = (k_2 L)^2$ are constants to avoid the denominator being zero, where $k_1 = 0.01, k_2 = 0.03$. L is the largest pixel value of the magnitude of images.

Loss functions play crucial roles in training the network and optimizing model performance. Studies proposing various networks have also discussed novel loss functions that may enhance image reconstruction. We listed the loss functions that are being used in several key representative methods. Most of the LOAs are supervised learning which requires ground truth in the loss function. RELAX-MORE [76] is subject-dependent self-supervised learning method that does not require fully sampled k-space data. The Table 1 shows the comparisons of different loss functions.

Table 1. Loss functions for variational networks in MRI reconstruction.

Model	Loss Function
Variational Network [18]	$\frac{1}{2}\|\mathbf{x}_T - \mathbf{x}^*\|_2^2$
MoDL [57]	$\frac{1}{2}\|\mathbf{x}_T - \mathbf{x}^*\|_2^2$
ISTA-Net [56]	$\mathcal{L}_{dis} + \mu \mathcal{L}_{id}$ $\mathcal{L}_{dis} = \frac{1}{2}\|\mathbf{x}_T - \mathbf{x}^*\|_2^2$ $\mathcal{L}_{id} = \frac{1}{2}\sum_{t=1}^{T}\|\tilde{\varphi}_t(\varphi_t(\mathbf{x}^*)) - \mathbf{x}^*\|_2^2$
pMRI-Net [71]	$\sum_{i=1}^{c}\gamma\|\mathbf{x}_i^{(T)} - \hat{\mathbf{x}}_i\| + \||\mathcal{J}(\bar{\mathbf{x}}^{(T)})| - \mathbf{s}(\hat{\mathbf{x}})\|$ $\quad + \eta\|\mathbf{s}(\bar{\mathbf{x}}^{(T)}) - \mathbf{s}(\hat{\mathbf{x}})\|$
ADMM-Net [102]	$\frac{1}{2}\|\mathbf{x}_T - \mathbf{x}^*\|_2^2$
Learned Primal-Dual [60]	$\frac{1}{2}\|\mathbf{x}_T - \mathbf{x}^*\|_2^2$
RELAX-MORE [76]	$\sum_{t=0}^{T}\gamma_t\|\mathbf{U}\mathcal{F}\mathbf{C}\mathbf{x}_t - \mathbf{f}\|_2^2$
DiMo [73]	$\mathbb{E}_{\hat{f}_t, t, \epsilon}\|\epsilon - \epsilon_\theta(\hat{f}_t, t)\|_2^2$

It is worth noting that the goal for diffusion models is to learn the Gaussian noise added in each diffusion step, the network ϵ_θ learns to remove the noise that is added in the forward process and matches the target distribution. Therefore, the loss function measures the discrepancy between the estimated learned noise in each step and the actual added noise in the training process.

4.2. Comparing Learnable Optimization Algorithms (LOA) with Traditional Optimization Methods

LOA has emerged as an effective alternative to traditional optimization methods. Despite its advantages, the LOA is computationally expensive, especially during the training phase. LOAs need a substantial amount of training data to generalize well across optimization tasks. Furthermore, owing to the requirement of large-scale training, the training process for the LOA could also be quite slow, especially when suffering the curse of dimensionality. Discussing the predictive performance and accuracy of the LOA and comparing it with the traditional optimization methods, LOA often tends to outperform traditional methods on more complex non-convex problems with irregular landscapes. Focusing on MRI reconstruction, LOAs have shown impressive results in recovering high-quality MRI images from undersampled k-space data. Additionally, studies have demonstrated that LOAs can handle noise and artifacts better, which are common in clinical MRI scans. Another advantage is that once the LOA is trained, they can reconstruct MRI images much quicker than traditional methods. This efficiency makes LOAs more attractive for real-time applications. However, as indicated earlier, this efficiency comes at the cost of increasing training time.

4.3. Strengths, Weakness, and Performance of the Reviewed Algorithms in Different Scenarios

In this paper, several algorithms focusing on MRI reconstruction have been discussed. To fully understand the limitations of existing algorithms and the need for future research, it is vital to perform a comparative analysis between the algorithms in this paper. The Table 2 summarized a detailed comparisons among several well-known LOA inspired methods.

Table 2. Strengths, weaknesses, performance of various optimization-based network unrolling algorithms for MRI reconstruction and their code availability.

Algorithm	Strengths	Weaknesses	Performance in Various Scenarios	Code Availability
Variational Network (VN) [18]	Uses gradient descent for reconstruction. Effective in multi-coil MRI. Incorporates CNNs for regularization.	Needs pre-calculated sensitivity maps. Limited flexibility in handling unknown or inaccurate sensitivity maps.	Strong performance in multi-coil MRI reconstruction, but dependent on the accuracy of sensitivity maps.	https://github.com/VLOGroup/mri-variationalnetwork (accessed on 28 August 2024)
E2E-VarNet [61]	Hybrid domain learning (k-space and image domain). Learn iterative optimization steps using gradient descent.	Complex training process. Requires high computational resources.	Excellent performance in k-space and image domain reconstructions. Effective in handling undersampled MRI data, but requires significant computational resources for training and deployment.	https://github.com/facebookresearch/fastMRI/tree/main/fastmri_examples/varnet (accessed on 28 August 2024)
MoDL [57]	Recursive architecture. Weight sharing across iterations. Effective denoising-based regularization.	May enhance artifacts due to over-smoothing. Computationally intensive due to its recursive nature.	Performs well in denoising and removing aliasing artifacts, particularly in single-coil and multi-coil MRI reconstruction. May suffer from artifact enhancement when over-smoothed.	https://github.com/hkaggarwal/modl (accessed on 28 August 2024)
ISTA-Net$^+$ [56]	Closed-form solution using soft-thresholding. Efficient for solving L1 norm problems. Incorporates residual learning.	May be less effective with complex, multi-coil data. Requires careful tuning of hyperparameters.	Excellent performance in sparse image reconstruction and L1-based regularization scenarios. Struggles in complex multi-coil scenarios.	https://github.com/jianzhangcs/ISTA-Net-PyTorch (accessed on 28 August 2024)
ADMM-Net [54]	Effective for single-coil MRI reconstruction. Retains iterative optimization principles. Utilizes CNN-based regularization.	Limited application in multi-coil settings. Requires accurate coil sensitivity maps.	Performs well in single-coil MRI reconstruction but struggles in multi-coil settings due to lack of sensitivity map adaptation.	https://github.com/yangyan92/Deep-ADMM-Net (accessed on 28 August 2024)
Primal-Dual Network [60]	Utilizes primal-dual optimization principles. Proximal operator is learnable. Good flexibility.	Requires significant training data. Computationally expensive due to dual operations.	Performs well across different MRI settings, including multi-coil and accelerated imaging. Effective in handling complex data with well-learned priors.	https://github.com/adler-j/learned_primal_dual (accessed on 28 August 2024)

Table 2. *Cont.*

Algorithm	Strengths	Weaknesses	Performance in Various Scenarios	Code Availability
Parallel MRI Network [71]	Bi-level optimization for MRI reconstruction. Regularization in both image and k-space domains. Cross-domain learning.	Complex design with dual regularization terms. Sensitive to training data.	Excellent performance in parallel MRI, including multi-coil settings. Balances between k-space and image domain reconstruction, effective at removing artifacts.	https://github.com/Wanyu624/pMRI_optimal_control (accessed on 28 August 2024)
RELAX-MORE [76]	Self-supervised learning. Transfer learning for quick adaptation. Optimizes quantitative MRI parameters.	Dependent on transfer learning for efficiency. Requires well-prepared training datasets.	Strong performance in quantitative MRI reconstruction, especially with transfer learning. Effective in undersampled scenarios and cross-domain applications.	N/A

4.4. 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.

4.5. 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 [38,39,70], which provide insights into the theoretical convergence of these methods. For example, reference [70] 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.

A significant application of deep learning in clinical MRI is its use in accelerating image acquisition, making it possible to acquire images up to 10 times faster than conventional methods without compromising diagnostic quality [24,94,103]. Artifact reduction is a critical challenge in MRI, where motion artifacts or metal implants can significantly degrade image quality. Deep learning models have shown scalability across different MRI modalities

and anatomical regions, which is crucial for their widespread adoption in clinical practice. The ability to handle large datasets and perform real-time processing is also vital for integrating these models into routine workflows. For deep learning models to be truly effective in clinical settings, they must generalize well across diverse patient populations, imaging protocols, and MRI scanners. This requires robust training on large, heterogeneous datasets and careful validation across multiple sites and clinical conditions.

4.6. 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.

4.7. Computational Burden, Memory Consumption, and Inference Time

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.

However, diffusion models, for example, require long inference times due to the pre-scheduled denoising steps involved in the sampling process. Additionally, training diffusion models is time-consuming and memory-intensive, particularly because of the self-attention modules incorporated in the denoising network at each diffusion step.

The LOAs such as ISTA-Net, and PD-Net unroll iterative optimization procedures, with each phase of the network being trained independently without parameter sharing. As the number of iterations increases, so does the computational and memory burden, making these methods more resource-intensive. It is crucial to acknowledge these limitations in the design of the network architecture, such as long inference times, model complexity, memory consumption, and computational burden, alongside the advantages of these methods. Addressing these challenges will be key to improving the feasibility and efficiency of deep learning approaches in MRI reconstruction.

To address these challenges, techniques such as pruning, quantization, and knowledge distillation can reduce the model size and memory footprint without significantly sacrificing performance. In addition, implement gradient checkpointing, where only a subset of activations is stored during the forward pass, and others are recomputed during the backward pass. This can significantly reduce memory usage during training. Splitting the training process across multiple GPUs or even across different nodes in a cluster can help manage memory constraints by distributing the workload. 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.

4.8. Other Related Reconstruction Methods

Federated learning (FL) for MRI reconstruction [104–106] is an innovative approach that enables multiple institutions to collaboratively train a deep learning model without sharing patient data, thus preserving privacy. This method is particularly valuable in the medical imaging field, where data privacy and security are paramount, and where institutions may have diverse datasets collected under different conditions, such as varying sensors, disease types, and acquisition protocols. For example, the FL-MR framework [105] trains local models at multiple institutions. Each local model computes reconstruction losses and update parameters through gradient descent. Then the updated model parameters are sent to the central server, where they are averaged to update the global model. The updated global model is then redistributed to local institutions for further training, iterating until the model converges.

Plug-and-play (PnP) methods [78,107–109] integrate state-of-the-art denoising algorithms as priors into the image reconstruction process. The PnP approach allows for the decoupling of image modeling (denoising) from forward modeling (data acquisition), which is particularly advantageous in MRI, where the forward model can vary significantly between different scans. For example, proximal-based PnP methods [78] leverage the ADMM framework to decouple the regularization term (which encodes prior knowledge of the image) from the data fidelity term. In these methods, a denoising algorithm replaces the proximal operator that would typically be used in the optimization process. Gradient-based PnP methods [108] involve using gradient descent-based algorithms like the fast iterative shrinkage-thresholding algorithm (FISTA) in combination with denoisers. Instead of solving the proximal update exactly, a gradient descent step is performed to handle the data fidelity term, followed by a denoising step that uses the denoiser as a regularizer. This category emphasizes computational efficiency, as gradient steps are generally less expensive than solving proximal updates. The consensus equilibrium (CE) framework [110–113] provides a theoretical understanding of PnP methods. It interprets the denoiser used in the PnP algorithm as a solution to an equilibrium equation rather than an exact minimization of a cost function. This framework helps to address questions related to the convergence of PnP methods, particularly when the denoiser does not correspond to any known regularizer.

Blind-PMRI-Net [58] is a method that alternates between updating images and sensitivity maps to solve multi-channel MRI problems. It employs a half-quadratic splitting approach, resulting in a complex network design that requires careful balancing of updates to maintain stability. This method is particularly effective in multi-coil MRI scenarios where sensitivity maps are not known a priori, making it well-suited for handling complex imaging tasks. However, optimizing this approach can be challenging due to its intricate structure.

VS-Net [59] utilizes variable splitting optimization to manage complex multi-coil data. While it performs well in parallel imaging tasks, its robustness is limited by its handling of sensitivity maps, making it less adaptable to certain coil configurations. VS-Net delivers good performance, but its sensitivity to inaccuracies in the sensitivity maps can impact its effectiveness in some scenarios.

4.9. 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. To further enhance MRI reconstruction, future work may focus on leveraging reinforcement learning (RL). RL can optimize acquisition parameters dynamically during the scan. This could lead to more efficient data collection and potentially reduce scan times further, thereby making MRI

procedures more responsive to specific patient needs and scanning conditions. In addition to RL, self-supervised learning also holds immense promise in improving the efficacy, quality, and accuracy of MRI reconstruction. Self-supervised learning can exploit the structure in MRI data, such as the physical constraints of the imaging process, to generate useful training signals without the need for a large quantity of ground truth data. This can serve as a catalyst for the development of novel models and reduce any dependency the researchers may have on expensive and time-consuming data labeling. Future work may see the development of more sophisticated self-supervised to leverage domain-specific knowledge and enhance model performance.

In addition to leveraging various techniques, 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. Conclusions

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. One of the most promising directions is the integration of these optimization techniques with emerging AI-driven methods such as generative models and RL, which could further improve reconstruction accuracy and reduce computational costs. Additionally, research on reinforcement and self-supervised learning could greatly improve the efficacy and accuracy of the MRI reconstruction process. The insights presented in this review are expected to facilitate further development and application of these LOAs in the fields of inverse problems and medical imaging reconstruction.

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. *N. Engl. J. Med.* **2000**, *343*, 1445–1453. [CrossRef]
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.; et al. MRI-targeted or standard biopsy for prostate-cancer diagnosis. *N. Engl. J. Med.* **2018**, *378*, 1767–1777. [CrossRef] [PubMed]
3. Pruessmann, K.P.; Weiger, M.; Scheidegger, M.B.; Boesiger, P. SENSE: Sensitivity encoding for fast MRI. *Magn. Reson. Med.* **1999**, *42*, 952–962. [CrossRef]
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). *Magn. Reson. Med.* **2002**, *47*, 1202–1210. [CrossRef] [PubMed]
5. Nyquist, H. Certain topics in telegraph transmission theory. *Trans. Am. Inst. Electr. Eng.* **1928**, *47*, 617–644. [CrossRef]
6. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306. [CrossRef]
7. Sodickson, D.K.; Manning, W.J. Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays. *Magn. Reson. Med.* **1997**, *38*, 591–603. [CrossRef]
8. Larkman, D.J.; Nunes, R.G. Parallel magnetic resonance imaging. *Phys. Med. Biol.* **2007**, *52*, R15. [CrossRef]
9. Fessler, J.A. Model-based image reconstruction for MRI. *IEEE Signal Process. Mag.* **2010**, *27*, 81–89. [CrossRef]
10. Yang, A.C.; Kretzler, M.; Sudarski, S.; Gulani, V.; Seiberlich, N. Sparse reconstruction techniques in magnetic resonance imaging: Methods, applications, and challenges to clinical adoption. *Investig. Radiol.* **2016**, *51*, 349–364. [CrossRef]
11. Bilgic, B.; Goyal, V.K.; Adalsteinsson, E. Multi-contrast reconstruction with Bayesian compressed sensing. *Magn. Reson. Med.* **2011**, *66*, 1601–1615. [CrossRef]
12. Huang, J.; Chen, C.; Axel, L. Fast multi-contrast MRI reconstruction. *Magn. Reson. Imaging* **2014**, *32*, 1344–1352. [CrossRef]
13. 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. [CrossRef]
14. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015, Proceedings of the 18th International Conference, Munich, Germany, 5–9 October 2015*; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
15. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In Proceedings of the 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, 20 September 2018; Proceedings 4; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–11.
16. Zhan, C.; Ghaderibaneh, M.; Sahu, P.; Gupta, H. Deepmtl pro: Deep learning based multiple transmitter localization and power estimation. *Pervasive Mob. Comput.* **2022**, *82*, 101582. [CrossRef]
17. Zhan, C.; Ghaderibaneh, M.; Sahu, P.; Gupta, H. Deepmtl: Deep learning based multiple transmitter localization. In Proceedings of the 2021 IEEE 22nd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), Pisa, Italy, 7–11 June 2021; pp. 41–50.
18. 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. *Magn. Reson. Med.* **2018**, *79*, 3055–3071. [CrossRef] [PubMed]
19. Schlemper, J.; Caballero, J.; Hajnal, J.V.; Price, A.; Rueckert, D. A deep cascade of convolutional neural networks for MR image reconstruction. In *Information Processing in Medical Imaging, Proceedings of the 25th International Conference, IPMI 2017, Boone, NC, USA, 25–30 June 2017*; Proceedings 25; Springer: Berlin/Heidelberg, Germany, 2017; pp. 647–658.
20. 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. *Magn. Reson. Med.* **2018**, *80*, 1189–1205. [CrossRef] [PubMed]
21. 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. [CrossRef]
22. Yang, G.; Yu, S.; Dong, H.; Slabaugh, G.; Dragotti, P.L.; Ye, X.; Liu, F.; Arridge, S.; Keegan, J.; Guo, Y.; et al. DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Trans. Med. Imaging* **2017**, *37*, 1310–1321. [CrossRef]
23. Lee, D.; Yoo, J.; Tak, S.; Ye, J.C. Deep residual learning for accelerated MRI using magnitude and phase networks. *IEEE Trans. Biomed. Eng.* **2018**, *65*, 1985–1995. [CrossRef]
24. 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 Process. Mag.* **2020**, *37*, 128–140. [CrossRef]
25. 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. *Magn. Reson. Med.* **2020**, *84*, 3172–3191. [CrossRef] [PubMed]
26. 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. In Proceedings of the ISMRM Annual Meeting, London, UK, 7–12 May 2022.

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

28. Jin, K.H.; McCann, M.T.; Froustey, E.; Unser, M. Deep convolutional neural network for inverse problems in imaging. *IEEE Trans. Image Process.* **2017**, *26*, 4509–4522. [CrossRef] [PubMed]

29. Lundervold, A.S.; Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. *Z. Med. Phys.* **2019**, *29*, 102–127. [CrossRef]

30. Liang, D.; Cheng, J.; Ke, Z.; Ying, L. Deep magnetic resonance image reconstruction: Inverse problems meet neural networks. *IEEE Signal Process. Mag.* **2020**, *37*, 141–151. [CrossRef]

31. Sandino, C.M.; Cheng, J.Y.; Chen, F.; Mardani, M.; Pauly, J.M.; Vasanawala, S.S. Compressed Sensing: From Research to Clinical Practice with Deep Neural Networks: Shortening Scan Times for Magnetic Resonance Imaging. *IEEE Signal Process. Mag.* **2020**, *37*, 117–127. [CrossRef]

32. McCann, M.T.; Jin, K.H.; Unser, M. Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Process. Mag.* **2017**, *34*, 85–95. [CrossRef]

33. 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. *Proc. IEEE* **2021**, *109*, 820–838. [CrossRef]

34. Singha, A.; Thakur, R.S.; Patel, T. Deep Learning Applications in Medical Image Analysis. In *Biomedical Data Mining for Information Retrieval: Methodologies, Techniques and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2021; pp. 293–350

35. Chandra, S.S.; Bran Lorenzana, M.; Liu, X.; Liu, S.; Bollmann, S.; Crozier, S. Deep learning in magnetic resonance image reconstruction. *J. Med. Imaging Radiat. Oncol.* **2021**, *65*, 564–577. [CrossRef]

36. Ahishakiye, E.; Van Gijzen, M.B.; Tumwiine, J.; Wario, R.; Obungoloch, J. A survey on deep learning in medical image reconstruction. *Intell. Med.* **2021**, *1*, 118–127. [CrossRef]

37. 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 Trans. Med. Imaging* **2020**, *39*, 4150–4163. [CrossRef] [PubMed]

38. Chen, Y.; Liu, H.; Ye, X.; Zhang, Q. Learnable descent algorithm for nonsmooth nonconvex image reconstruction. *SIAM J. Imaging Sci.* **2021**, *14*, 1532–1564. [CrossRef]

39. Bian, W.; Chen, Y.; Ye, X.; Zhang, Q. An optimization-based meta-learning model for mri reconstruction with diverse dataset. *J. Imaging* **2021**, *7*, 231. [CrossRef]

40. Hammernik, K.; Küstner, T.; Yaman, B.; Huang, Z.; Rueckert, D.; Knoll, F.; Akçakaya, M. Physics-Driven Deep Learning for Computational Magnetic Resonance Imaging: Combining physics and machine learning for improved medical imaging. *IEEE Signal Process. Mag.* **2023**, *40*, 98–114. [CrossRef]

41. Uecker, M.; Lai, P.; Murphy, M.J.; Virtue, P.; Elad, M.; Pauly, J.M.; Vasanawala, S.S.; Lustig, M. ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA. *Magn. Reson. Med.* **2014**, *71*, 990–1001. [CrossRef]

42. Deshmane, A.; Gulani, V.; Griswold, M.A.; Seiberlich, N. Parallel MR imaging. *J. Magn. Reson. Imaging* **2012**, *36*, 55–72. [CrossRef] [PubMed]

43. Fliege, J.; Svaiter, B.F. Steepest descent methods for multicriteria optimization. *Math. Methods Oper. Res.* **2000**, *51*, 479–494. [CrossRef]

44. 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. [CrossRef]

45. 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.; et al. Substituting gadolinium in brain MRI using DeepContrast. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; pp. 908–912.

46. Others, S.W. DeepcomplexMRI: Exploiting deep residual network for fast parallel MR imaging with complex convolution. *Magn. Reson. Imaging* **2020**, *68*, 136–147.

47. Wang, S.; Su, Z.; Ying, L.; Peng, X.; Zhu, S.; Liang, F.; Feng, D.; Liang, D. Accelerating magnetic resonance imaging via deep learning. In Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, Czech Republic, 13–16 April 2016; pp. 514–517.

48. Kwon, K.; Kim, D.; Park, H. A parallel MR imaging method using multilayer perceptron. *Med. Phys.* **2017**, *44*, 6209–6224. [CrossRef]

49. Quan, T.M.; Nguyen-Duc, T.; Jeong, W.K. Compressed Sensing MRI Reconstruction Using a Generative Adversarial Network with a Cyclic Loss. *IEEE Trans. Med. Imaging* **2018**, *37*, 1488–1497. [CrossRef]

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

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

52. Sriram, A.; Zbontar, J.; Murrell, T.; Zitnick, C.L.; Defazio, A.; Sodickson, D.K. GrappaNet: Combining parallel imaging with deep learning for multi-coil MRI reconstruction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 14315–14322.

53. Han, Y.; Sunwoo, L.; Ye, J.C. k-space deep learning for accelerated MRI. *IEEE Trans. Med. Imaging* **2019**, *39*, 377–386. [CrossRef]

54. Yang, Y.; Sun, J.; Li, H.; Xu, Z. Deep ADMM-Net for Compressive Sensing MRI. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: New York, NY, USA, 2016; Volume 29.

55. 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 J. Imaging Sci.* **2023**, *16*, 2202–2246. [CrossRef]

56. Zhang, J.; Ghanem, B. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1828–1837.

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

58. Meng, N.; Yang, Y.; Xu, Z.; Sun, J. A prior learning network for joint image and sensitivity estimation in parallel MR imaging. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October; Springer: Berlin/Heidelberg, Germany, 2019; pp. 732–740.

59. 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. In *Medical Image Computing and Computer Assisted Intervention– MICCAI 2019, Proceedings of the 22nd International Conference, Shenzhen, China, 13–17 October 2019*; Proceedings, Part IV 22; Springer: Berlin/Heidelberg, Germany, 2019; pp. 713–722.

60. Adler, J.; Öktem, O. Learned primal-dual reconstruction. *IEEE Trans. Med. Imaging* **2018**, *37*, 1322–1332. [CrossRef] [PubMed]

61. 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 In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020, Proceedings of the 23rd International Conference, Lima, Peru, 4–8 October 2020*; Proceedings, Part II 23; Springer: Berlin/Heidelberg, Germany, 2020; pp. 64–73.

62. Yiasemis, G.; Sonke, J.J.; Sánchez, C.; Teuwen, J. Recurrent variational network: A deep learning inverse problem solver applied to the task of accelerated MRI reconstruction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 732–741.

63. Polak, D.; Cauley, S.; Bilgic, B.; Gong, E.; Bachert, P.; Adalsteinsson, E.; Setsompop, K. Joint multi-contrast variational network reconstruction (jVN) with application to rapid 2D and 3D imaging. *Magn. Reson. Med.* **2020**, *84*, 1456–1469. [CrossRef]

64. Schlemper, J.; Caballero, J.; Hajnal, J.V.; Price, A.N.; Rueckert, D. A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction. *IEEE Trans. Med. Imaging* **2018**, *37*, 491–503. [CrossRef]

65. Jun, Y.; Shin, H.; Eo, T.; Kim, T.; Hwang, D. Deep model-based magnetic resonance parameter mapping network (DOPAMINE) for fast T1 mapping using variable flip angle method. *Med. Image Anal.* **2021**, *70*, 102017. [CrossRef]

66. Mardani, M.; Sun, Q.; Donoho, D.; Papyan, V.; Monajemi, H.; Vasanawala, S.; Pauly, J. Neural proximal gradient descent for compressive imaging. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 9596–9606.

67. 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 Med. Imaging* **2021**, *21*, 195. [CrossRef] [PubMed]

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

69. Bian, W. Optimization-Based Deep Learning Methods for Magnetic Resonance Imaging Reconstruction and Synthesis. Ph.D. Thesis, University of Florida, Gainesville, FL, USA, 2022.

70. Bian, W.; Zhang, Q.; Ye, X.; Chen, Y. A learnable variational model for joint multimodal MRI reconstruction and synthesis. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Singapore, 18–22 September 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 354–364.

71. Bian, W.; Chen, Y.; Ye, X. An optimal control framework for joint-channel parallel MRI reconstruction without coil sensitivities. *Magn. Reson. Imaging* **2022**, *89*, 1–11. [CrossRef] [PubMed]

72. Bian, W.; Jang, A.; Liu, F. Magnetic Resonance Parameter Mapping using Self-supervised Deep Learning with Model Reinforcement. *arXiv* **2023**, arXiv:2307.13211v1.

73. Bian, W.; Jang, A.; Zhang, L.; Yang, X.; Stewart, Z.; Liu, F. Diffusion modeling with domain-conditioned prior guidance for accelerated mri and qmri reconstruction. *IEEE Trans. Med. Imaging* **2024**. [CrossRef]

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

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

76. Bian, W.; Jang, A.; Liu, F. Improving quantitative MRI using self-supervised deep learning with model reinforcement: Demonstration for rapid T1 mapping. *Magn. Reson. Med.* **2024**, *92*, 98–111. [CrossRef]

77. Heide, F.; Steinberger, M.; Tsai, Y.T.; Rouf, M.; Pająk, D.; Reddy, D.; Gallo, O.; Liu, J.; Heidrich, W.; Egiazarian, K.; et al. Flexisp: A flexible camera image processing framework. *ACM Trans. Graph.* **2014**, *33*, 1–13. [CrossRef]

78. Venkatakrishnan, S.V.; Bouman, C.A.; Wohlberg, B. Plug-and-play priors for model based reconstruction. In Proceedings of the 2013 IEEE Global Conference on Signal and Information Processing, Austin, TX, USA, 3–5 December 2013; pp. 945–948.

79. Sun, L.; Fan, Z.; Fu, X.; Huang, Y.; Ding, X.; Paisley, J. A deep information sharing network for multi-contrast compressed sensing MRI reconstruction. *IEEE Trans. Image Process.* **2019**, *28*, 6141–6153. [CrossRef]

80. Dar, S.U.; Yurt, M.; Shahdloo, M.; Ildız, M.E.; Tınaz, B.; Çukur, T. Prior-guided image reconstruction for accelerated multi-contrast MRI via generative adversarial networks. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 1072–1087. [CrossRef]

81. Liu, X.; Wang, J.; Jin, J.; Li, M.; Tang, F.; Crozier, S.; Liu, F. Deep unregistered multi-contrast MRI reconstruction. *Magn. Reson. Imaging* **2021**, *81*, 33–41. [CrossRef] [PubMed]

82. Zhou, B.; Dey, N.; Schlemper, J.; Salehi, S.S.M.; Liu, C.; Duncan, J.S.; Sofka, M. DSFormer: A dual-domain self-supervised transformer for accelerated multi-contrast MRI reconstruction. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–7 January 2023; pp. 4966–4975.

83. Jun, Y.; Shin, H.; Eo, T.; Hwang, D. Joint deep model-based MR image and coil sensitivity reconstruction network (joint-ICNet) for fast MRI. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 5270–5279.

84. Arvinte, M.; Vishwanath, S.; Tewfik, A.H.; Tamir, J.I. Deep J-Sense: Accelerated MRI reconstruction via unrolled alternating optimization. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Strasbourg, France, 27 September–1 October 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 350–360.

85. Peng, X.; Sutton, B.P.; Lam, F.; Liang, Z.P. DeepSENSE: Learning coil sensitivity functions for SENSE reconstruction using deep learning. *Magn. Reson. Med.* **2022**, *87*, 1894–1902. [CrossRef] [PubMed]

86. Tang, L.; Zhao, Y.; Li, Y.; Guo, R.; Cai, B.; Wang, J.; Li, Y.; Liang, Z.P.; Peng, X.; Luo, J. Jsense-pro: Joint sensitivity estimation and image reconstruction in parallel imaging using p re-learned subspaces of coil sensitivity functions. *Magn. Reson. Med.* **2023**, *89*, 1531–1542. [CrossRef]

87. Wang, H.Z.; Riederer, S.J.; Lee, J.N. Optimizing the precision in T1 relaxation estimation using limited flip angles. *Magn. Reson. Med.* **1987**, *5*, 399–416. [CrossRef]

88. Cheng, J.; Wang, H.; Ying, L.; Liang, D. Model learning: Primal dual networks for fast MR imaging. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 21–29.

89. Chung, H.; Ye, J.C. Score-based diffusion models for accelerated MRI. *Med. Image Anal.* **2022**, *80*, 102479. [CrossRef]

90. 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. *Med. Image Anal.* **2023**, *88*, 102872. [CrossRef] [PubMed]

91. Kazerouni, A.; Aghdam, E.K.; Heidari, M.; Azad, R.; Fayyaz, M.; Hacihaliloglu, I.; Merhof, D. Diffusion models in medical imaging: A comprehensive survey. *Med. Image Anal.* **2023**, *88*, 102846. [CrossRef]

92. 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 Comput. Surv.* **2023**, *56*, 1–39. [CrossRef]

93. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 6000–6010.

94. 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. *Magn. Reson. Med.* **2019**, *81*, 116–128. [CrossRef]

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

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

97. Zhou, B.; Zhou, S.K. DuDoRNet: Learning a dual-domain recurrent network for fast MRI reconstruction with deep T1 prior. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 4273–4282.

98. Xie, Y.; Li, Q. Measurement-conditioned denoising diffusion probabilistic model for under-sampled medical image reconstruction. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Singapore, 18–22 September 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 655–664.

99. Liu, F.; Kijowski, R.; El Fakhri, G.; Feng, L. Magnetic resonance parameter mapping using model-guided self-supervised deep learning. *Magn. Reson. Med.* **2021**, *85*, 3211–3226. [CrossRef] [PubMed]

100. 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. *Magn. Reson. Med.* **2019**, *81*, 2072–2089. [CrossRef] [PubMed]

101. Zhang, T.; Pauly, J.M.; Levesque, I.R. Accelerating parameter mapping with a locally low rank constraint. *Magn. Reson. Med.* **2015**, *73*, 655–661. [CrossRef] [PubMed]

102. Yang, Y.; Sun, J.; Li, H.; Xu, Z. ADMM-CSNet: A deep learning approach for image compressive sensing. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *42*, 521–538. [CrossRef]

103. Radmanesh, A.; Muckley, M.; Murrell, T.; Lindsey, E.; Sriram, A.; Knoll, F.; Sodickson, D.; Lui, Y. Exploring the Acceleration Limits of Deep Learning VarNet-based Two-dimensional Brain MRI. *Radiol. Artif. Intell.* **2022**, *4*, e210313. [CrossRef]

104. Guo, P.; Wang, P.; Zhou, J.; Jiang, S.; Patel, V.M. Multi-institutional collaborations for improving deep learning-based magnetic resonance image reconstruction using federated learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 2423–2432.

105. Elmas, G.; Dar, S.U.; Korkmaz, Y.; Ceyani, E.; Susam, B.; Ozbey, M.; Avestimehr, S.; Çukur, T. Federated learning of generative image priors for MRI reconstruction. *IEEE Trans. Med. Imaging* **2022**, *42*, 1996–2009. [CrossRef]
106. Feng, C.M.; Yan, Y.; Wang, S.; Xu, Y.; Shao, L.; Fu, H. Specificity-preserving federated learning for MR image reconstruction. *IEEE Trans. Med. Imaging* **2022**, *42*, 2010–2021. [CrossRef]
107. Yazdanpanah, A.P.; Afacan, O.; Warfield, S.K. Deep Plug-and-Play Prior for Parallel MRI Reconstruction. In Proceedings of the International Conference on Computer Vision (ICCV 2019) Workshop on Learning for Computational Imaging, Seoul, Republic of Korea, 27 October–2 November 2019.
108. Kamilov, U.S.; Mansour, H.; Wohlberg, B. A plug-and-play priors approach for solving nonlinear imaging inverse problems. *IEEE Signal Process. Lett.* **2017**, *24*, 1872–1876. [CrossRef]
109. Ahmad, R.; Bouman, C.A.; Buzzard, G.T.; Chan, S.; Liu, S.; Reehorst, E.T.; Schniter, P. Plug-and-play methods for magnetic resonance imaging: Using denoisers for image recovery. *IEEE Signal Process. Mag.* **2020**, *37*, 105–116. [CrossRef] [PubMed]
110. Sun, Y.; Wohlberg, B.; Kamilov, U.S. An online plug-and-play algorithm for regularized image reconstruction. *IEEE Trans. Comput. Imaging* **2019**, *5*, 395–408. [CrossRef]
111. Reehorst, E.T.; Schniter, P. Regularization by denoising: Clarifications and new interpretations. *IEEE Trans. Comput. Imaging* **2018**, *5*, 52–67. [CrossRef] [PubMed]
112. Chan, S.H. Performance analysis of plug-and-play ADMM: A graph signal processing perspective. *IEEE Trans. Comput. Imaging* **2019**, *5*, 274–286. [CrossRef]
113. Buzzard, G.T.; Chan, S.H.; Sreehari, S.; Bouman, C.A. Plug-and-play unplugged: Optimization-free reconstruction using consensus equilibrium. *SIAM J. Imaging Sci.* **2018**, *11*, 2001–2020. [CrossRef]