Rethinking Deep Unrolled Model for Accelerated MRI Reconstruction

Bingyu Xin[1], Meng Ye[1], Leon Axel[2], and Dimitris N. Metaxas[1]

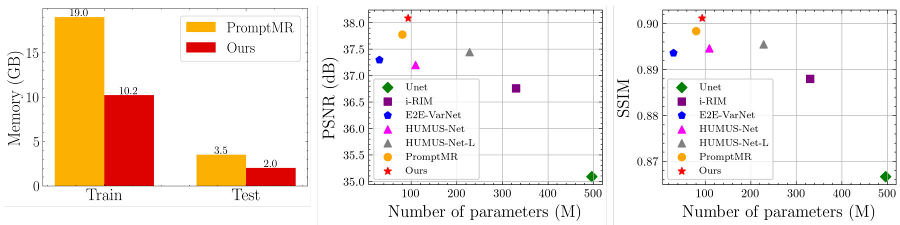
[1] Rutgers University, New Brunswick, NJ 08901, USA
[2] New York University Langone Medical Center, New York, NY 10016, USA
bx64@rutgers.edu, dnm@cs.rutgers.edu

Abstract. Magnetic Resonance Imaging (MRI) is a widely used imaging modality for clinical diagnostics and the planning of surgical interventions. Accelerated MRI seeks to mitigate the inherent limitation of long scanning time by reducing the amount of raw k-space data required for image reconstruction. Recently, the deep unrolled model (DUM) has demonstrated significant effectiveness and improved interpretability for MRI reconstruction, by truncating and unrolling the conventional iterative reconstruction algorithms with deep neural networks. However, the potential of DUM for MRI reconstruction has not been fully exploited. In this paper, we first enhance the gradient and information flow within and across iteration stages of DUM, then we highlight the importance of using various adjacent information for accurate and memory-efficient sensitivity map estimation and improved multi-coil MRI reconstruction. Extensive experiments on several public MRI reconstruction datasets show that our method outperforms existing MRI reconstruction methods by a large margin. The code is available at `https://github.com/hellopipu/PromptMR-plus`.

Keywords: MRI Reconstruction · Deep Unrolled Model · Gradient-based Learning · Sensitivity Map Estimation

1 Introduction

Magnetic Resonance Imaging (MRI) provides a radiation-free and highly versatile method for imaging the organs, tissues, and skeletal systems within the human body. Over the past 50 years since Paul Lauterbur produced the first MR image, MRI has evolved to become a cornerstone in clinical diagnostics [19]. However, the process of raw k-space data acquisition in MRI is typically time-consuming. Accelerated MRI techniques tackle this issue by minimizing the amount of raw data that needs to be collected for image reconstruction, thereby shortening the duration of the scan. Modern advances in MRI technology, including Parallel Imaging (PI) [8] and Compressed Sensing (CS) [7, 22], have significantly enhanced the efficiency and quality of MRI scans, making it possible to acquire high-resolution images within a considerably reduced time frame. However, with increased patient throughput and the requirements of emerging



(a) (b)

Fig. 1: Our proposed method is memory-efficient and shows better reconstruction performance. (a) GPU memory consumption on Calgary-Campinas brain dataset, (b) PSNR/SSIM results vs. the number of parameters on the fastMRI knee $\times 8$ testset.

technologies, such as real-time MRI [25] and low-field MRI [3], the development of accelerated MRI reconstruction methods remains a hot research topic [24].

Accelerated MRI reconstruction is a regularized inverse problem, which aims to reconstruct an unknown MR image from highly undersampled measurements in k-space. Conventional iterative MRI reconstruction methods minimize a cost function comprising two main components: a data-consistency term, which assesses the alignment between k-space predicted from the reconstructed image and the observed measurements, and a regularization term, which incorporates prior knowledge to encourage the emergence of desirable image attributes, e.g., sparsity. Over the past decade, deep learning [20] has emerged as a transformative approach to MRI reconstruction [26, 37]. Recently, the release of several large-scale public MRI reconstruction benchmarks, including the fastMRI dataset [49], Calgary-Campinas dataset [39], and CMRxRecon dataset [44], has significantly propelled the advancement of MRI reconstruction methods. Among deep learning-based MRI reconstruction approaches, the deep unrolled model (DUM) has garnered significant attention for its exceptional performance and ability to establish a concrete and systematic link between widely used iterative MRI reconstruction methods and deep neural networks [23]. However, as pointed out in [51], the current architecture design of DUM is inefficient, characterized by limited information transmission capacity and low flexibility and robustness. These shortcomings significantly restrict DUM's performance and its applicability to unseen data.

In this work, we rethink and improve key components of the DUM-based MRI reconstruction approach. Our main contributions can be summarized as follows:

- We improve the adaptive gradient algorithm and apply it to DUM, which can achieve self-adaptive dynamic learning rates adjusting for different spatial areas in an MR image. This approach provides a more flexible updating strategy in each iteration stage of DUM.
- We incorporate the momentum technique used in gradient descent acceleration and propose a multi-stage and multi-level feature aggregation scheme to accelerate the iteration convergence of DUM.

- We highlight the importance of adjacent information which can be used to improve multi-coil MRI reconstruction, especially for accurate and memory-efficient sensitivity map estimation. As shown in Fig. 1(a), our method reduces GPU consumption to $\sim 55\%$ than the previous state-of-the-art.
- We demonstrate the effectiveness of our proposed method on three public MRI reconstruction benchmarks of different anatomies: the fastMRI knee, the Calgary-Campinas brain and the CMRxRecon cardiac dataset. As shown in Fig. 1(b), our method achieves better PSNR and SSIM with fewer parameters than existing MRI Reconstruction methods.

2 Related Work

2.1 Deep Unrolled Model (DUM)

DUM maps a truncated optimization algorithm into a deep neural network, iteratively alternating between gradient descent steps and proximal mapping steps. Its superior performance and good interpretability make it a preferred method for a wide range of inverse problems, including compressive sensing, image restoration and image reconstruction [23,51]. Learned ISTA (LISTA) was the first unrolling method proposed for fast approximation of sparse coding [13]. LISTA unrolls the traditional ISTA algorithm [5] and truncates it to a fixed number of iterations. By learning from a training dataset, this trainable version of ISTA shows significant computational benefits. The success of deep learning further promotes the idea of relating conventional iterative optimization methods to deep neural networks, including ADMM-Net [41], ISTA-Net [52] and AMP-Net [54]. By introducing a condition module to transmit input information to each stage of the unrolled model, ISTA-Net++ [48] provides more flexibility to handle multi-ratio and multi-scene images for compressive sensing. In MADUN [38], a memory-augmented DUM is proposed to enhance the short-term and long-term information transmission in traditional DUMs.

2.2 DUM for MRI Reconstruction

ADMM-Net [41] and DC-CNN [36] are the pioneering DUMs for MRI reconstruction, establishing themselves as widely used baseline DUM-based MRI reconstruction methods. These approaches, based on Convolutional Neural Networks (CNNs), unroll the alternating direction method of multipliers (ADMM) [6] and the alternating minimization (AM) algorithm, respectively. KIKI-Net [11] employs a novel approach by alternating between k-space completion blocks and image restoration blocks, demonstrating the advantages of cross-domain learning [10,21,56]. Adaptive-CS-Net finds it beneficial to inject information from neighboring slices into the model. This 2.5D reconstruction strategy has been adopted in subsequent works [12,45].

The development of DUM has led to various innovative approaches beyond simple unrolling, categorized broadly into three main variants for enhanced

MRI reconstruction. (1) Firstly, inspired by Recurrent Neural Network (RNN), RIM [31,32] uses a latent variable as temporal memory to track the progression of iterations. Recurrent-VarNet [46] extends this idea by designing a recurrent state initializer and recurrent units for the hidden state. In parallel, LPD-Net [1] and NCPD-Net [34] simplify this concept by utilizing a **buffer** design as additional learnable memory transmitting across iterations. (2) Secondly, inspired by efficient variants of proximal gradient descent (PGD), HC-PGD [16] proposes the history-cognizant unrolling method by concatenating all previous proximal operator output as input to each iteration. DIRCN [27] further proposes to add feature map interconnections between neighboring stages. These works suggest the importance of gradient flow and shared information between cascades for MRI reconstruction. (3) Moreover, Adaptive-CS-Net [28] incorporates soft MRI priors, including soft data consistency, image phase information and background priors, as additional input channels to each iteration. Similarly, HiTDUN [53] proposes the design of multi-channel input for every iteration. These methods underscore the advantages of introducing additional priors into the input at each iteration stage, aiming to dismantle the information bottleneck encountered within each iteration.

2.3 Sensitivity Map Estimation in DUM

Numerous works on DUM have achieved promising results on MRI reconstruction [11, 21, 36, 41, 56]. However, most of these methods focus exclusively on synthesized single-coil k-space data, overlooking the more prevalent multi-coil imaging in real clinical MRI scans. This oversight neglects a more realistic scenario, given the widespread application of parallel imaging. Parallel imaging (PI) can speed up MRI, with the idea of using multiple receiving coils to reduce the number of phase encoding lines needed in acquiring raw k-space data. The sensitivity map of a given coil is its local field profile, describing through reciprocity where that coil can efficiently pick up MR signals from the imaged region. The reconstruction of multiple coil k-space data needs accurate estimation of the sensitivity maps, which are either directly measured, as in SENSE [30], or indirectly accounted using a center region of k-space data, as in GRAPPA [14].

Robust and accurate sensitivity map estimation is crucial for multi-coil MRI reconstruction. The i-RIM directly reconstructs coil images without explicitly estimating the sensitivity maps. Adaptive-CS-Net [28] uses the normalized low-pass filtered coil images as sensitivity maps. VarNet [15] and MoDL [2] handle multi-coil data with pre-estimated sensitivity maps [43]. E2E-VarNet [40] advances the VarNet by learning sensitivity maps from the Auto-Calibration Signal (ACS) data of the k-space in an end-to-end manner. Due to its superior performance and end-to-end approach, many subsequent studies are based on E2E-VarNet [12,45,46]. When the ACS data is insufficient for accurate sensitivity map estimation, [4,17] suggest iterative sensitivity map estimation and image reconstruction, which can be considered as a deep version of J-SENSE [47].

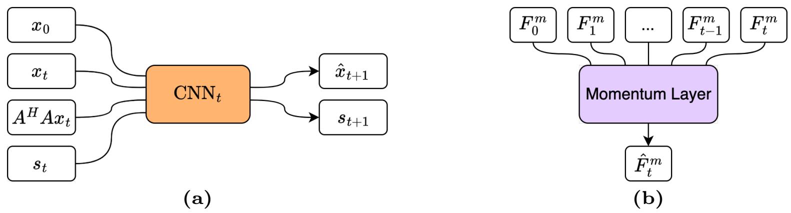


Fig. 2: Overview of our proposed (a) self-adaptive and (b) momentum-accelerated gradient algorithm. $\hat{x}_{t+1}$ and $\hat{F}_t^m$ are the output of CNN and Momentum layer in Eq. (7) and Eq. (10), respectively.

3 Proposed Method

Drawing inspiration from adaptive gradient and momentum-accelerated algorithms in gradient descent optimization, we dynamically incorporate prior knowledge in the data flow, to enable more informative and faster convergence for gradient-based MRI reconstruction learning, as shown in Fig. 2. Subsequently, we enhance the concept of adjacent reconstruction by focusing on more robust and memory-efficient multi-coil sensitivity map estimation, as shown in Fig. 4. Please refer to Supplementary Section B for architecture details.

3.1 Problem Formulation

We can estimate the complex MR image $\boldsymbol{x}$ from its undersampled measurements $\boldsymbol{y}$ in k-space by solving an optimization problem [22],

$$\min_{\boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x}) = \min_{\boldsymbol{x}} \frac{1}{2}||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}||_2^2 + \boldsymbol{R}(\boldsymbol{x}), \tag{1}$$

where $\boldsymbol{A} = \boldsymbol{M}\boldsymbol{\mathcal{F}}\boldsymbol{S}$ is the forward process of acquiring the k-space measurements, which is a combination of coil sensitivity maps $\boldsymbol{S}$, the Fourier transform $\boldsymbol{\mathcal{F}}$ and an undersampling mask $\boldsymbol{M}$. The first term $\frac{1}{2}||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}||_2^2$ represents the data consistency term, which ensures data fidelity, while the second term $\boldsymbol{R}(\boldsymbol{x})$ serves as the regularization term, aimed at promoting image domain priors, such as sparsity [7]. We can solve Eq. (1) iteratively via the gradient descent method,

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t - \eta_t \boldsymbol{g}_t, \tag{2}$$

$$\boldsymbol{g}_t = \boldsymbol{A}^H(\boldsymbol{A}\boldsymbol{x}_t - \boldsymbol{y}) + \nabla \boldsymbol{R}(\boldsymbol{x}_t), \tag{3}$$

where $\nabla \boldsymbol{R}(\boldsymbol{x}_t)$ is the gradient or the proximal mapping of $\boldsymbol{R}$, η^t is the scalar learning rate at iteration t, and $t \in \{0, 1, ..., T-1\}$, with T being the total number of truncated iterations. In the context of the deep unrolled models, it is customary to substitute $\nabla \boldsymbol{R}(\boldsymbol{x}_t)$ with a neural network whose parameters are learned from large-scale training datasets. This allows for the learning of more sophisticated image domain priors, compared to those that are hand-crafted.

3.2 Self-Adaptive Gradient Algorithm

During the iterative MRI reconstruction process, some regions in an image are straightforward to reconstruct, such as the background, which contains minimal information. However, other areas are more prone to aliasing artifacts resulting from downsampling the k-space, which are challenging to eliminate. This difficulty often arises because the aliasing artifacts may closely mimic the appearance of anatomical structures within these regions, necessitating a more meticulous gradient-based correction at each iteration. Inspired by the adaptive gradient algorithm [9,18,42,50], instead of applying a uniform learning rate across the entire image space at each iteration, it is reasonable to utilize self-adaptive learning rates for individual pixels to achieve a more refined reconstruction result. Our generalization of the standard gradient descent in Eq. (2) employs the following update for each iteration:

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t - \eta_t \boldsymbol{g}_t - \boldsymbol{E}_t \odot \boldsymbol{g}_t, \tag{4}$$

$$\boldsymbol{E}_t = \frac{\eta_t}{\sqrt{\sum_{\tau=0}^{t} \boldsymbol{g}_\tau^2} + \epsilon} - \eta_t, \tag{5}$$

where $\boldsymbol{E}_t$ is the residual pixel-wise learning rate at the t-th iteration, and the denominator in Eq. (5) computes the root of the sum of squares of all previous gradients of individual pixels. ϵ is a small number to improve the numerical stability. The operation $\odot$ denotes element-wise multiplication. We note that, the accumulation of historical pixel-wise gradients in Eq. (5) can help improve iteration convergence by

(1) scaling down the updating steps for regions with historically large gradients, which may be prone to overshooting;
(2) scaling up the updating steps for regions with historically small gradients, which may benefit from a more aggressive update.

To unroll our self-adaptive gradient algorithm with a DUM, we deviate from the common practice of substituting $\nabla \boldsymbol{R}(\boldsymbol{x}_t)$ with a neural network in Eq. (4). There are compelling reasons for this. First, learning the $\nabla \boldsymbol{R}(\boldsymbol{x}_t)$ term in $\boldsymbol{g}_t$ (see Eq. (3)) can lead to numerical instability in $\boldsymbol{E}_t$ during the early stage of training. Second, the adaptive gradient strategy in Eq. (5), which is the basic variant utilized in AdaGrad [9], may not be the most suitable choice for a DUM in the context of MRI reconstruction. To leverage the robust nonlinear representation capabilities of neural networks, we improve upon the adaptive gradient algorithm by reformulating Eq. (4) as follows:

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t - \eta_t \boldsymbol{A}^H(\boldsymbol{A}\boldsymbol{x}_t - \boldsymbol{y}) - (\eta_t \nabla \boldsymbol{R}(\boldsymbol{x}_t) + \boldsymbol{E}_t \odot \boldsymbol{g}_t) \tag{6}$$

$$= \boldsymbol{x}_t - \eta_t \boldsymbol{A}^H(\boldsymbol{A}\boldsymbol{x}_t - \boldsymbol{y}) + \mathbf{CNN}_t(\boldsymbol{x}_t, \boldsymbol{A}^H\boldsymbol{A}\boldsymbol{x}_t, \boldsymbol{A}^H\boldsymbol{y}, \boldsymbol{s}_t). \tag{7}$$

For the last term in Eq. (6), we reparametrize it with a CNN, as indicated in Eq. (7). This neural network takes as input the fundamental components from which the term is constructed, namely, $\boldsymbol{x}_t$, $\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{x}_t$, $\boldsymbol{A}^H\boldsymbol{y}$, and a learnable auxiliary memory $\boldsymbol{s}_t$, which serves as a hidden state to track the historical gradient information implicitly. Note that $\boldsymbol{A}^H\boldsymbol{y}$ yields $\boldsymbol{x}_0$, which represents the image transformed from zero-filled k-space. For further discussion, please refer to Supplementary Section E.

3.3 Momentum-Accelerated Gradient Algorithm

The momentum method [29] accelerates gradient descent in Eq. (2) by accumulating a velocity vector in directions of persistent reduction in the optimization objective across iterations:

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t - \eta_t \boldsymbol{m}_t, \tag{8}$$

$$\boldsymbol{m}_t = \mu \boldsymbol{m}_{t-1} + (1-\mu)\boldsymbol{g}_t, \tag{9}$$

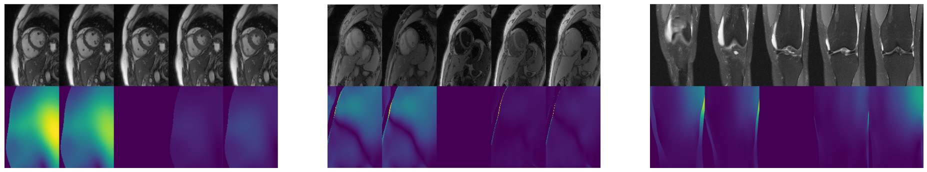
where $\mu \in [0,1]$ is the momentum coefficient. The momentum $\boldsymbol{m}_t$ is known as the exponential moving average of historical gradients. It can benefit the optimization of the current iteration, thereby considerably accelerating convergence. In Sec. 3.2, we introduced the hidden state $\boldsymbol{s}_t$, which is used to implicitly track the pixel-wise historical gradient information. Here, we propose explicit multi-stage and multi-level feature aggregation, to mimic the idea of momentum and increase the information flow across the iterations. For a typical multi-level encoder-decoder denoiser, such as `Unet`, in the t-th iteration, we insert a Momentum layer before the m-th level decoder layer. This layer concatenates features at the current stage and features from all previous iteration stages at the same level, followed by a 1×1 convolution layer to reduce the dimensionality, and a Channel Attention Block (CAB) to adaptively fuse the multi-stage information, as described below:

$$\text{Momentum}(\boldsymbol{F}_0^m, ..., \boldsymbol{F}_t^m) = \text{CAB}(\text{Conv}_{1\times 1}(\text{Concat}(\boldsymbol{F}_0^m, ..., \boldsymbol{F}_t^m))), \tag{10}$$

where $\boldsymbol{F}_t^m$ is the input feature of the m-th level decoder layer at the t-th iteration.

3.4 Adjacent Reconstruction

Incorporating information from adjacent slices in multi-coil MRI reconstruction has been substantiated in prior work [12, 28, 45]. However, a significant limitation of adjacent slice reconstruction lies in its substantial memory consumption during the estimation of multi-coil sensitivity maps. This issue stems from the design of the sensitivity map estimation (SME) network in earlier work [40]. The network adopts a coil-by-coil estimation approach to accommodate a variable number of coils, increasing flexibility but leading to higher memory demands in the adjacent reconstruction setting (see Fig. 4(a)). To be more specific, for k-space data with N coils and $2a+1$ adjacent slices, the total number of forward passes required for the SME process is calculated as $(2a+1)N$. Typically, the



(a) time **(b)** contrast **(c)** slice

Fig. 3: The sensitivity maps exhibit variations across (a) time, (b) contrast and (c) slice dimension. First row: original images, second row: magnitude of differences between the sensitivity map of the central image and each of those of its adjacent images.

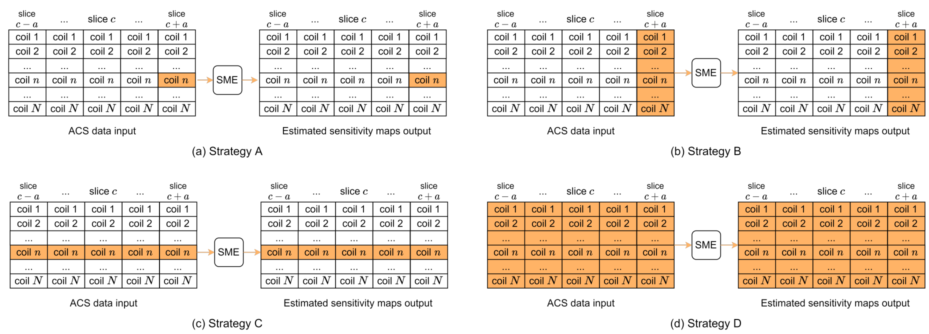


Fig. 4: Concept illustration of different strategies for multi-coil sensitivity map estimation (SME). The main difference between the four strategies lies in the Auto-Calibration-Signal (ACS) information sharing between coils and adjacent MR image slices. Each single coil sensitivity map gets ACS information from (a) a single coil and a single image slice; (b) multiple coils but a single image slice; (c) a single coil but multiple adjacent image slices; (d) multiple coils and multiple adjacent image slices.

value of N ranges from 8 to 34 and a is set to 2. We propose to reduce the number of forward passes, by utilizing the correlations between adjacent sensitivity maps.

A sensitivity map is subject to variation, based on the position of the imaging subject and changes over time due to the subject motion. Additionally, in multi-contrast MRI acquisition, variations in acquisition parameters, such as inversion time (TI) and repetition time (TR), can indirectly influence the sensitivity map estimation through the acquired k-space data, as shown in Fig. 3. Consequently, the estimation of adjacent sensitivity maps is a spatially, temporally, and contrastively correlated process. We therefore propose to estimate sensitivity maps by utilizing information from spatially, temporally, and contrastively adjacent slices. As shown in Fig. 4(c), we simultaneously estimate a single coil sensitivity map at all adjacent $2a+1$ slices using information across these slices. In this way, we can reduce the number of forward passes from $(2a+1)N$ to N. Our estimation of the n-th sensitivity map set S^n_{adj} can be described as:

$$\boldsymbol{S}^n_{adj} = \mathbf{SME}(\mathbf{ACS}(\boldsymbol{x}^n_{0,adj})), \tag{11}$$

where $\boldsymbol{S}^n_{adj} = [\boldsymbol{S}^n_{a-c}, ..., \boldsymbol{S}^n_c, ..., \boldsymbol{S}^n_{a+c}]$ is an adjacent sensitivity map set which is the concatenation of central slice sensitivity map $\boldsymbol{S}^n_c$ with its $2a$-adjacent slice sensitivity maps of the n-th coil. $\mathbf{ACS}(\boldsymbol{x}^n_{0,adj})$ is the n-th coil Auto-Calibration Signal (ACS)[1] within all adjacent $2a+1$ slices.

4 Experiments and Results

In this section, we present the experimental results of our proposed approach. We showcase the reconstruction performance of our model across multiple publicly available datasets, including Calgary-Campinas brain [39], CMRxRecon cardiac [44], and fastMRI multi-coil knee [49] datasets. These datasets encompass a diverse range of anatomies, contrasts, views, slices and motion states. For details about the datasets, please refer to Supplementary Section A. For information on the experimental setup for each dataset, see Supplementary Section C.

4.1 Calgary-Campinas Brain Results

Tab. 1 presents the quantitative results for $\times 5$ and $\times 10$ acceleration factors on the Calgary-Campinas brain test set. Our method is compared with two state-of-the-art (SOTA) methods: Recurrent-VarNet [46] and PromptMR [45]. Our proposed method significantly outperforms these benchmarks, demonstrating its outstanding performance. Qualitative results under $\times 10$ acceleration are shown in Fig. 5. Interestingly, while PromptMR achieves significantly higher PSNR and SSIM values compared to Recurrent-VarNet, it exhibits less accurate reconstruction in the area highlighted by the red box.

The highly undersampled k-space data makes the ill-posed reconstruction challenging at aliasing artifacts removal. While leveraging large-scale training data to learn the image priors can improve the reconstruction accuracy, we note that Recurrent-VarNet learns a weaker prior, resulting in reconstructions that appear blurry. Conversely, PromptMR adopts a strong but overly biased prior, leading to sharper reconstructions but introducing incorrect structures. Our proposed method, in contrast, successfully learns the appropriate prior, achieving a reconstruction that closely mirrors the ground truth in terms of structural accuracy.

4.2 CMRxRecon Cardiac Results

We compare our method with PromptMR, the winning method of the MICCAI2023 CMRxRecon challenge, for $\times 4$, $\times 8$ and $\times 10$ acceleration factors on the CMRxRecon cardiac test set. In Tab. 2, our method exhibits superior performance to PromptMR across different views and contrasts. Qualitative results under $\times 10$ acceleration are shown in Fig. 6. Our method reconstructs fine details of the interventricular septum, highlighted in the red box, with significantly higher clarity than the PromptMR method, which is severely compromised by aliasing artifacts.

[1] It is usually a part of the acquired central k-space data.

Table 1: Quantitative comparison of PSNR/SSIM (mean ± std) of different MRI reconstruction methods on the Calgary-Campinas brain dataset under ×5 and ×10 acceleration. The best and second best performance are in red and blue colors, respectively.

Method	Acc = 5×		Acc = 10×	
	PSNR(dB)	SSIM(%)	PSNR(dB)	SSIM(%)
Zero-filled	25.27±1.31	73.75±3.93	24.35±1.21	69.04±3.67
Recurrent-Varnet [46]	36.27±1.72	94.37±1.33	33.27±2.12	91.47±2.23
PromptMR [45]	36.83±1.68	94.85±1.26	34.10±2.11	92.69±1.92
Ours	37.24±1.57	95.15±1.21	34.74±1.97	93.38±1.65

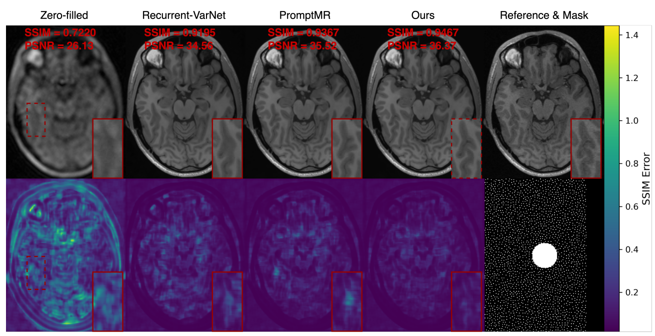


Fig. 5: The reconstruction results and SSIM error maps of various methods for the Calgary-Campinas brain imaging under ×10 acceleration factor. The red boxes highlight the differences in the recovery of white and gray matter structures.

4.3 FastMRI Knee Results

We evaluated the PSNR and SSIM at acceleration factors of ×4 and ×8 on the fastMRI multi-coil knee test set. Our method is compared with several methods, including SOTA methods HUMUS-Net-L [12] and PromptMR. In the table, 'PD' and 'PDFS' denote proton density-weighted images without and with fat suppression, respectively. Tab. 3 demonstrates that our approach achieves the best performance across all other competitive methods on this large-scale dataset. Qualitative results under ×8 acceleration are shown in Fig. 7. Our approach achieves a more detailed reconstruction of the meniscus region as shown in the red box, closely approximating the ground truth. In this region, meniscus tears were identified by radiologists from the fastMRI+ dataset [55]. The qualitative comparison suggests that our approach is more robust in reconstructing abnormal knee regions for highly accelerated acquisition.

5 Abalation study

In this section, we justify our design choices through ablation studies on the brain and cardiac datasets.

Table 2: Quantitative comparison of PSNR/SSIM (mean ± std) of different MRI reconstruction methods on the CMRxRecon cardiac dataset under ×4, ×8 and ×10 acceleration. The best and second best performance are in red and blue colors, respectively.

Acc	Method	Cine				Mapping			
		SAX		LAX		T1w		T2w	
		PSNR(dB)	SSIM(%)	PSNR(dB)	SSIM(%)	PSNR(dB)	SSIM(%)	PSNR(dB)	SSIM(%)
4×	Zero-filled	26.20±1.54	72.99±5.14	25.11±1.62	69.87±4.73	24.23±1.16	67.27±3.84	24.97±1.10	77.16±2.91
	PromptMR [45]	45.78±1.89	98.78±0.43	45.58±1.72	98.72±0.37	46.49±2.23	98.93±0.57	42.02±1.93	98.05±0.72
	Ours	46.26±1.76	98.87±0.34	46.13±1.70	98.83±0.34	47.45±2.13	99.09±0.42	42.68±1.84	98.25±0.59
8×	Zero-filled	25.18±1.67	70.21±5.87	24.34±1.66	67.67±5.80	23.41±1.25	64.12±4.43	24.21±1.12	74.84±3.07
	PromptMR [45]	40.65±1.52	96.97±0.68	39.64±1.87	96.36±1.04	40.94±1.69	97.28±0.85	38.24±1.82	96.53±1.09
	Ours	41.44±1.48	97.32±0.58	40.49±1.83	96.76±0.94	42.08±1.63	97.70±0.67	39.23±1.71	96.98±0.89
10×	Zero-filled	24.83±1.67	69.35±5.92	24.13±1.67	66.89±5.58	23.16±1.27	64.14±4.47	24.25±1.14	76.08±2.91
	PromptMR [45]	39.18±1.50	96.15±0.80	38.28±1.62	95.60±1.07	38.99±1.58	96.61±0.99	37.21±1.76	96.22±1.16
	Ours	39.99±1.49	96.58±0.72	39.13±1.66	96.05±0.99	40.37±1.58	97.19±0.79	38.22±1.70	96.70±0.96

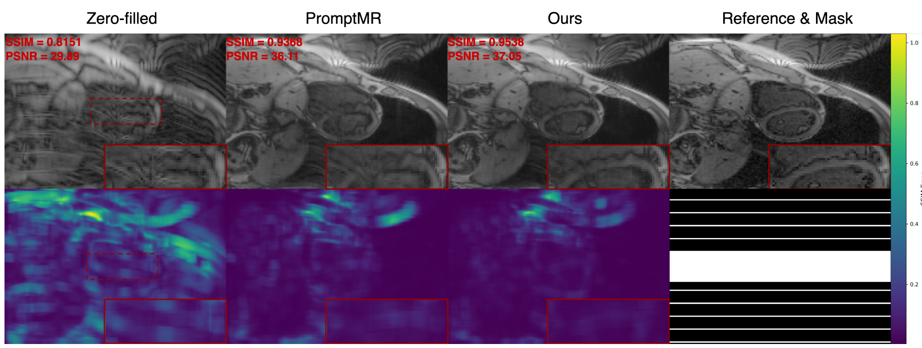


Fig. 6: The reconstruction results and SSIM error maps of various methods for the CMRxRecon cardiac T1-weighted imaging under ×10 acceleration factor. The red boxes highlight the differences in the recovery of the interventricular septum.

Table 3: Quantitative comparison of PSNR/SSIM (mean ± std) of different MRI reconstruction methods on the fastMRI knee dataset under ×4 and ×8 acceleration. The best and second best performance are in red and blue colors, respectively.

Method	Acc = 4×				Acc = 8×			
	PD		PDFS		PD		PDFS	
	PSNR(dB)	SSIM(%)	PSNR(dB)	SSIM(%)	PSNR(dB)	SSIM(%)	PSNR(dB)	SSIM(%)
Zero-filled	30.86±1.73	80.65±3.76	31.00±3.33	78.48±6.16	27.70±1.84	74.12±4.94	27.97±2.02	68.78±5.58
Unet [35]	37.27±1.76	92.11±2.72	37.07±2.47	88.03±5.04	35.41±2.14	90.36±3.13	34.84±1.59	83.71±5.04
i-RIM [33]	39.61±2.33	94.16±2.87	38.23±3.13	89.27±4.96	37.91±2.47	93.22±2.85	35.84±1.71	85.27±4.93
E2E-VarNet [40]	40.09±2.32	94.56±2.64	38.43±3.24	89.50±4.95	38.72±2.50	93.93±2.68	36.16±1.76	85.70±4.94
HUMUS-Net [12]	-	-	-	-	38.72±2.50	94.07±2.68	35.98±1.75	85.77±4.93
HUMUS-Net-L [12]	-	-	-	-	39.03±2.47	94.18±2.61	36.19±1.79	85.84±4.92
PromptMR [45]	40.58±2.59	94.87±2.75	38.59±3.33	89.75±4.94	39.48±2.54	94.51±2.54	36.42±1.79	86.10±4.93
Ours	40.85±2.74	95.00±2.84	38.71±3.37	89.85±4.92	40.09±2.58	94.90±2.45	36.49±1.88	86.29±4.95

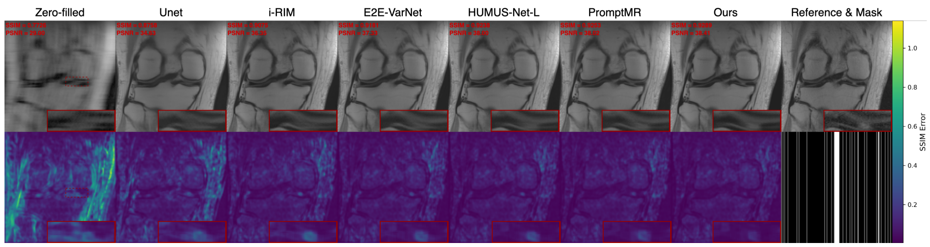


Fig. 7: The reconstruction results and SSIM error maps of various methods for the fastMRI multi-coil knee proton density (PD) imaging under ×8 acceleration. The red boxes highlight the differences in the recovery of the meniscus region. The red box in the reference image was annotated by radiologists from the fastMRI+ dataset [55] to indicate the knee abnormality of meniscus tears.

Table 4: Abalation of different input combinations of CNN on the Calgary-Campinas brain dataset under ×10 acceleration.

Input	Baseline	-	-	-	-	-	-	Ours
$\boldsymbol{x}_t$	✓	✓	✓	✓	✓	✓	✓	✓
$\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{x}_t$		✓				✓	✓	✓
$\boldsymbol{A}^H\boldsymbol{y}$			✓		✓	✓		✓
$\boldsymbol{s}_t$				✓	✓		✓	✓
PSNR(dB)	35.76	+0.00	+0.01	+0.09	+0.08	+0.31	+0.36	+0.36
SSIM(%)	94.33	+0.00	−0.01	+0.06	+0.05	+0.25	+0.27	+0.28

5.1 Input Variations to CNN

We conducted an ablation study for different input combinations to CNN, to validate the effectiveness of our proposed method. Results are shown in Tab. 4. The baseline network in each iteration only accepts the output from the last iteration $\boldsymbol{x}_t$ as input, which is the setting for most unrolled models, such as E2E-VartNet and PromptMR. Our proposed method takes additional input, $\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{x}_t$, $\boldsymbol{A}^H\boldsymbol{y}$ and $\boldsymbol{s}_t$, which outperforms the baseline by a large margin. More specifically, adding $\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{x}_t$ or $\boldsymbol{A}^H\boldsymbol{y}$ solely to the input has no impact on the performance. Adding $\boldsymbol{s}_t$ can improve the performance, as it serves as additional memory to track the progression of the reconstruction process. Adding $\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{x}_t$ together with $\boldsymbol{A}^H\boldsymbol{y}$ greatly improves the gains. We assume the reason is that the residual connection of two terms into the network can provide soft data consistency information to the model. Adding $\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{x}_t$ together with $\boldsymbol{s}_t$ performs the closest to our final design. It seems the missing of $\boldsymbol{A}^H\boldsymbol{y}$ has little impact, as it is a constant value and its information may be compactly represented in the $\boldsymbol{s}_t$. This assumption is implied by the evidence that the result of adding $\boldsymbol{A}^H\boldsymbol{y}$ together with $\boldsymbol{s}_t$ into the network does not perform better than adding $\boldsymbol{s}_t$ alone.

Table 5: Quantitative comparison of PSNR/SSIM using different numbers of previous stages' outputs for multi-stage feature fusion on the Calgary-Campinas brain dataset under $\times 10$ acceleration.

Input	# of previous iterations				
	0	1	5	9	11
PSNR(dB)	35.88	+0.02	+0.04	+0.04	+0.05
SSIM(%)	94.43	+0.02	+0.05	+0.06	+0.07

5.2 Multi-Stage Information Flow

In Tab. 5, we investigate the effect of incorporating different numbers of previous iteration results into the current iteration for multi-stage information fusion. As the number increases, the reconstruction performance increases gradually. This implies the benefit of incorporating the information from previous iterations in DUM-based MRI reconstruction.

5.3 Adjacent Sensitivity Map Estimation

In Tab. 6, we examine the impact of various strategies shown in Fig. 4 for sensitivity map estimation on both reconstruction performance and memory consumption. The ACS data input to the SME network is represented as a complex tensor with shape $(B, 2a+1, N, h, w)$, where $2a+1$ represents the adjacent number, N denotes the number of coils of each slice and B indicates the batch size. Here, the batch size is linearly related to the memory consumption. Strategy A reshapes the tensor to a batch size of $(2a+1)BN$ and a channel size of 1 for coil-by-coil estimation. Strategy B concatenates the coils within each slice, reducing the batch size to $(2a+1)B$. Strategy C concatenates the same coils from adjacent slices, decreasing the batch size to BN. Strategy D combines all coils within a single batch. Note that only strategies A and C can be used for data with variable numbers of coils. Our experiments demonstrate that Strategy C, which utilizes adjacent information, not only shows the best performance but also reduces memory consumption by a factor of $2a+1$ compared to Strategy A. Intriguingly, strategy B is found to be the least effective. Ideally, since the coils within the same slice should be uncorrelated, this strategy should not negatively impact performance. However, the noise that is coupled across coils during acquisition may be amplified under this strategy, degrading the SME performance.

In Tab. 7, we further explore the impact of different adjacent types on Strategy C using the CMRxRecon cardiac datasets. These types include temporal information in the cine LAX dataset, contrast information in the T1 mapping dataset, and slice information in the cine SAX dataset. Strategy C shows consistent improvement over Strategy A for adjacent types of time and contrast. However, the result of the slice adjacency experiment deviates from our expectations, as previously observed in the brain dataset in Tab. 6. This discrepancy

Table 6: Comparison of different strategies for sensitivity map estimation on the Calgary-Campinas brain dataset under $\times 10$ acceleration. The "Batch" number is linearly related to memory consumption. "Flexibility" indicates the ability to accept data with varying numbers of coils.

Strategy	Batch	Channel	PSNR(dB)/SSIM(%)	Flexibility
A	$(2a+1)BN$	1	35.60/94.22	✓
B	$(2a+1)B$	N	35.30/93.98	✗
C	BN	$2a+1$	35.88/94.43	✓
D	B	$(2a+1)N$	35.76/94.32	✗

Table 7: Quantitative comparison of PSNR(dB)/SSIM(%) of different sensitivity map estimation strategies using different adjacent information types on the CMRxRecon cardiac dataset under $\times 10$ acceleration.

Strategy	Adjacent type		
	Time	Contrast	Slice
A	41.67/97.35	42.62/98.15	42.75/97.58
C	41.94/97.54	42.77/98.22	42.53/97.54

is attributed to the considerable slice gap present in the cardiac dataset, which is 4.0 mm compared to no slice gap in the brain dataset. This suggests that adjacent reconstruction can be beneficial only when adjacent slices show high correlations; otherwise, it might even detract from performance by introducing confounding factors for the network.

6 Limitations

Like other deep unrolled models, our approach may exhibit limited generalization to unseen sampling masks, modalities, and anatomies, and it may be susceptible to adversarial attacks. Hallucinations and artifacts may be observed when the model is applied to highly undersampled k-space data (e.g., at acceleration factors of 10), when trained on a limited amount of data, or when tested on out-of-distribution datasets.

7 Conclusion

In this paper, we improve the deep unrolled model for multi-coil MRI reconstruction through informative gradient-based learning and memory-efficient sensitivity map estimation. Our proposed method achieves state-of-the-art performance on several public MRI reconstruction benchmarks.

Acknowledgments. This research has been partially funded by research grants to D. Metaxas through NSF: 2310966, 2235405, 2212301, 2003874, and FA9550-23-1-0417 and NIH 2R01HL127661.

References

1. Adler, J., Öktem, O.: Learned primal-dual reconstruction. IEEE transactions on medical imaging **37**(6), 1322–1332 (2018)
2. Aggarwal, H.K., Mani, M.P., Jacob, M.: Modl: Model-based deep learning architecture for inverse problems. IEEE transactions on medical imaging **38**(2), 394–405 (2018)
3. Arnold, T.C., Freeman, C.W., Litt, B., Stein, J.M.: Low-field mri: clinical promise and challenges. Journal of Magnetic Resonance Imaging **57**(1), 25–44 (2023)
4. Arvinte, M., Vishwanath, S., Tewfik, A.H., Tamir, J.I.: Deep j-sense: Accelerated mri reconstruction via unrolled alternating optimization. In: International conference on medical image computing and computer-assisted intervention. pp. 350–360. Springer (2021)
5. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM journal on imaging sciences **2**(1), 183–202 (2009)
6. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine learning **3**(1), 1–122 (2011)
7. Candès, E.J., et al.: Compressive sampling. In: Proceedings of the international congress of mathematicians. vol. 3, pp. 1433–1452. Madrid, Spain (2006)
8. Deshmane, A., Gulani, V., Griswold, M.A., Seiberlich, N.: Parallel mr imaging. Journal of Magnetic Resonance Imaging **36**(1), 55–72 (2012)
9. Duchi, J., Hazan, E., Singer, Y.: Adaptive subgradient methods for online learning and stochastic optimization. Journal of machine learning research **12**(7) (2011)
10. El-Rewaidy, H., Fahmy, A.S., Pashakhanloo, F., Cai, X., Kucukseymen, S., Csecs, I., Neisius, U., Haji-Valizadeh, H., Menze, B., Nezafat, R.: Multi-domain convolutional neural network (md-cnn) for radial reconstruction of dynamic cardiac mri. Magnetic Resonance in Medicine **85**(3), 1195–1208 (2021)
11. Eo, T., Jun, Y., Kim, T., Jang, J., Lee, H.J., Hwang, D.: Kiki-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. Magnetic resonance in medicine **80**(5), 2188–2201 (2018)
12. Fabian, Z., Tinaz, B., Soltanolkotabi, M.: Humus-net: Hybrid unrolled multi-scale network architecture for accelerated mri reconstruction. Advances in Neural Information Processing Systems **35**, 25306–25319 (2022)
13. Gregor, K., LeCun, Y.: Learning fast approximations of sparse coding. In: Proceedings of the 27th international conference on international conference on machine learning. pp. 399–406 (2010)
14. Griswold, M.A., Jakob, P.M., Heidemann, R.M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., Haase, A.: Generalized autocalibrating partially parallel acquisitions (grappa). Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine **47**(6), 1202–1210 (2002)
15. Hammernik, K., Klatzer, T., Kobler, E., Recht, M.P., Sodickson, D.K., Pock, T., Knoll, F.: Learning a variational network for reconstruction of accelerated mri data. Magnetic resonance in medicine **79**(6), 3055–3071 (2018)

16. Hosseini, S.A.H., Yaman, B., Moeller, S., Hong, M., Akçakaya, M.: Dense recurrent neural networks for accelerated mri: History-cognizant unrolling of optimization algorithms. IEEE Journal of Selected Topics in Signal Processing **14**(6), 1280–1291 (2020)
17. 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. pp. 5270–5279 (2021)
18. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
19. Lauterbur, P.C.: Image formation by induced local interactions: examples employing nuclear magnetic resonance. nature **242**(5394), 190–191 (1973)
20. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. nature **521**(7553), 436–444 (2015)
21. Liu, X., Pang, Y., Jin, R., Liu, Y., Wang, Z.: Dual-domain reconstruction network with v-net and k-net for fast mri. Magnetic Resonance in Medicine **88**(6), 2694–2708 (2022)
22. Lustig, M., Donoho, D., Pauly, J.M.: Sparse mri: The application of compressed sensing for rapid mr imaging. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine **58**(6), 1182–1195 (2007)
23. Monga, V., Li, Y., Eldar, Y.C.: Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. IEEE Signal Processing Magazine **38**(2), 18–44 (2021)
24. Munoz, C., Fotaki, A., Botnar, R.M., Prieto, C.: Latest advances in image acceleration: all dimensions are fair game. Journal of Magnetic Resonance Imaging **57**(2), 387–402 (2023)
25. Nayak, K.S., Lim, Y., Campbell-Washburn, A.E., Steeden, J.: Real-time magnetic resonance imaging. Journal of Magnetic Resonance Imaging **55**(1), 81–99 (2022)
26. Ongie, G., Jalal, A., Metzler, C.A., Baraniuk, R.G., Dimakis, A.G., Willett, R.: Deep learning techniques for inverse problems in imaging. IEEE Journal on Selected Areas in Information Theory **1**(1), 39–56 (2020)
27. Ottesen, J.A., Caan, M.W., Groote, I.R., Bjørnerud, A.: A densely interconnected network for deep learning accelerated mri. Magnetic Resonance Materials in Physics, Biology and Medicine **36**(1), 65–77 (2023)
28. Pezzotti, N., Yousefi, S., Elmahdy, M.S., Van Gemert, J.H.F., Schuelke, C., Doneva, M., Nielsen, T., Kastryulin, S., Lelieveldt, B.P., Van Osch, M.J., et al.: An adaptive intelligence algorithm for undersampled knee mri reconstruction. IEEE Access **8**, 204825–204838 (2020)
29. Polyak, B.T.: Some methods of speeding up the convergence of iteration methods. Ussr computational mathematics and mathematical physics **4**(5), 1–17 (1964)
30. Pruessmann, K.P., Weiger, M., Scheidegger, M.B., Boesiger, P.: Sense: sensitivity encoding for fast mri. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine **42**(5), 952–962 (1999)
31. Putzky, P., Karkalousos, D., Teuwen, J., Miriakov, N., Bakker, B., Caan, M., Welling, M.: i-rim applied to the fastmri challenge. arXiv preprint arXiv:1910.08952 (2019)
32. Putzky, P., Welling, M.: Recurrent inference machines for solving inverse problems. arXiv preprint arXiv:1706.04008 (2017)
33. Putzky, P., Welling, M.: Invert to learn to invert. Advances in neural information processing systems **32** (2019)

34. Ramzi, Z., Chaithya, G., Starck, J.L., Ciuciu, P.: Nc-pdnet: A density-compensated unrolled network for 2d and 3d non-cartesian mri reconstruction. IEEE Transactions on Medical Imaging **41**(7), 1625–1638 (2022)
35. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)
36. Schlemper, J., Caballero, J., Hajnal, J.V., Price, A.N., Rueckert, D.: A deep cascade of convolutional neural networks for dynamic mr image reconstruction. IEEE transactions on Medical Imaging **37**(2), 491–503 (2017)
37. 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 **10**(9), 1012 (2023)
38. Song, J., Chen, B., Zhang, J.: Memory-augmented deep unfolding network for compressive sensing. In: Proceedings of the 29th ACM international conference on multimedia. pp. 4249–4258 (2021)
39. Souza, R., Lucena, O., Garrafa, J., Gobbi, D., Saluzzi, M., Appenzeller, S., Rittner, L., Frayne, R., Lotufo, R.: An open, multi-vendor, multi-field-strength brain mr dataset and analysis of publicly available skull stripping methods agreement. NeuroImage **170**, 482–494 (2018)
40. 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: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23. pp. 64–73. Springer (2020)
41. Sun, J., Li, H., Xu, Z., et al.: Deep admm-net for compressive sensing mri. Advances in neural information processing systems **29** (2016)
42. Tieleman, T.: Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural networks for machine learning **4**(2), 26 (2012)
43. 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. Magnetic resonance in medicine **71**(3), 990–1001 (2014)
44. Wang, C., Lyu, J., Wang, S., Qin, C., Guo, K., Zhang, X., Yu, X., Li, Y., Wang, F., Jin, J., et al.: Cmrxrecon: an open cardiac mri dataset for the competition of accelerated image reconstruction. arXiv preprint arXiv:2309.10836 (2023)
45. Xin, B., Ye, M., Axel, L., Metaxas, D.N.: Fill the k-space and refine the image: Prompting for dynamic and multi-contrast mri reconstruction. In: International Workshop on Statistical Atlases and Computational Models of the Heart. pp. 261–273. Springer (2023)
46. 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. pp. 732–741 (2022)
47. Ying, L., Sheng, J.: Joint image reconstruction and sensitivity estimation in sense (jsense). Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine **57**(6), 1196–1202 (2007)
48. You, D., Xie, J., Zhang, J.: Ista-net++: Flexible deep unfolding network for compressive sensing. In: 2021 IEEE International Conference on Multimedia and Expo (ICME). pp. 1–6. IEEE (2021)

49. 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 preprint arXiv:1811.08839 (2018)
50. Zeiler, M.D.: Adadelta: an adaptive learning rate method. arXiv preprint arXiv:1212.5701 (2012)
51. Zhang, J., Chen, B., Xiong, R., Zhang, Y.: Physics-inspired compressive sensing: Beyond deep unrolling. IEEE Signal Processing Magazine **40**(1), 58–72 (2023)
52. 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. pp. 1828–1837 (2018)
53. Zhang, J., Zhang, Z., Xie, J., Zhang, Y.: High-throughput deep unfolding network for compressive sensing mri. IEEE Journal of Selected Topics in Signal Processing **16**(4), 750–761 (2022)
54. Zhang, Z., Liu, Y., Liu, J., Wen, F., Zhu, C.: Amp-net: Denoising-based deep unfolding for compressive image sensing. IEEE Transactions on Image Processing **30**, 1487–1500 (2020)
55. Zhao, R., Yaman, B., Zhang, Y., Stewart, R., Dixon, A., Knoll, F., Huang, Z., Lui, Y.W., Hansen, M.S., Lungren, M.P.: fastmri+, clinical pathology annotations for knee and brain fully sampled magnetic resonance imaging data. Scientific Data **9**(1), 152 (2022)
56. 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. pp. 4273–4282 (2020)