electronics

MDPI

Article

ARMANI: Dictionary-Learning-Inspired Data-Free Deep Generative Modeling with Meta-Attention and Implicit Preconditioning for Compressively Sampled Magnetic Resonance Imaging

Ming Wu [1], Jing Cheng [2], Qingyong Zhu [3] and Dong Liang [1,2,3,*]

[1] School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China
[2] Paul C. Lauterbur Research Centre for Biomedical Imaging, Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Sciences, Shenzhen 518055, China
[3] Research Center for Medical AI, Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Sciences, Shenzhen 518055, China; qy.zhu@siat.ac.cn
* Correspondence: dong.liang@siat.ac.cn

Abstract

Magnetic resonance imaging (MRI) reconstruction from undersampled k-space data enables accelerated acquisition but leads to a severely ill-posed inverse problem. Although supervised deep learning methods have achieved strong performance, they typically rely on large paired datasets that are difficult to obtain in clinical practice. To address these limitations, we propose a dictionary-learning-inspired dAta-fRee deep generative modeling with Meta-Attention and implicit precoNditIoning for compressively sampled MRI (CS-MRI), termed ARMANI. Specifically, a meta-attention-augmented deep image prior (MA-DIP) generator performs a joint optimization over the latent input η and the network parameter θ, where η is regularized via gradient-domain sparsity and θ is constrained by a ridge penalty, mirroring the adaptive estimation of sparse coefficients and an empirical sparsifying dictionary. Furthermore, we integrate a single-step pseudo-orthogonal projection to achieve implicit preconditioning, which modulates the loss landscape and mitigates ill-conditioning of the forward operator. Experimental results demonstrate that ARMANI consistently outperforms existing SOTA data-free and self-supervised methods, and, with limited training data, achieves performance comparable to or slightly better than the supervised benchmark MoDL, with effective artifact suppression and faithful recovery of fine structural details. Overall, ARMANI shows strong scalability and potential for practical deployment in fully data-free CS-MRI reconstruction scenarios.

Keywords: magnetic resonance imaging; deep generative model; deep learning; meta-attention; implicit preconditioning

Check for updates

Academic Editor: Panagiota Spyridonos

Received: 14 February 2026
Revised: 21 March 2026
Accepted: 25 March 2026
Published: 27 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Magnetic resonance imaging (MRI) is an essential clinical tool owing to the superior soft-tissue contrast and multiparametric imaging capabilities. Nevertheless, prolonged data acquisition remains a critical limitation, adversely affecting both workflow efficiency and patient comfort. In recent years, k-space undersampling-based accelerated imaging techniques [1–5], collectively referred to as compressively sampled MRI (CS-MRI), have been devised to overcome hardware-imposed speed bottlenecks [6,7] by casting an ill-posed reconstruction problem with advanced learnable models [8,9].

Deep learning-involved CS-MRI methods have gained considerable traction [10–14]. Among them, most existing approaches employ deep networks to learn an end-to-end mapping from undersampled data to fully sampled images. However, whether purely data-driven [15,16] or model-guided [17,18], their effectiveness still depends heavily on large paired training datasets, which are costly and often impractical to obtain in clinical settings. Motivated by these limitations, a class of unsupervised CS-MRI strategies has emerged that enables acceptable image reconstruction without relying on external paired training data. One category is label-free deep generative models (DGMs) [19–21], which learn the target-image distribution $p(x)$ from large collections of unpaired data by maximizing likelihood, minimizing Kullback–Leibler (KL) divergence [22,23], or employing adversarial losses, and then use the learned distribution as a powerful prior for CS-MRI reconstruction. However, such methods still require substantial data collection and curation. In contrast, another category consists of data-free DGMs [24–34], such as deep image prior (DIP) [25], which completely dispense with training datasets. Instead, they exploit architectural priors and the intrinsic inductive bias of convolutional or non-convolutional networks, and naturally capture dominant low-frequency structures during the early stages of optimization, thereby offering notable advantages in data-scarce scenarios. Nevertheless, these methods still face two major challenges: limited representational capacity, which constrains reconstruction fidelity, and unstable or inefficient optimization, which may lead to suboptimal reconstruction quality [35].

In this work, we present ARMANI, a data-free DGM conceptually inspired by dictionary learning (DL) frameworks [36,37]. Specifically, a meta-attention-augmented DIP (MA-DIP) generator is introduced to jointly optimize the latent input η and the network parameters θ, where η is regularized by gradient-domain sparsity and θ is constrained by a ridge penalty. From a DL-inspired perspective, this joint optimization can be interpreted as being analogous to the adaptive estimation of representation coefficients and an empirical sparsifying dictionary. To further improve reconstruction accuracy and convergence behavior, we incorporate a single-step pseudo-orthogonal projection to realize implicit preconditioning (IP), which economically reshapes the loss landscape and alleviates the ill-conditioning of the forward operator. Experimental results demonstrate that ARMANI consistently outperforms existing state-of-the-art data-free CS-MRI methods and, in some cases, even slightly surpasses supervised benchmarks such as MoDL [38], while simultaneously improving structural fidelity and suppressing reconstruction artifacts with favorable computational efficiency.

2. Materials and Methods

In the CS-MRI task, the data acquisition process can be mathematically expressed as follows:

$$y = \mathcal{A}x + n, \tag{1}$$

where y denotes the acquired k-space measurements, x is the underlying image to be reconstructed, and n represents additive white Gaussian noise. The forward operator $\mathcal{A}$ is defined as $\mathcal{A} = \mathrm{MFS}$, where M is the downsampling pattern mask, F denotes the Fourier encoding, and S represents the coil sensitivity.

2.1. Data-Free DGM for CS-MRI

Definition 1 (Deep Network Prior [28,39])**.** *A particular image x is deemed to follow an untrained deep network prior if it falls within a set Ω defined as $\Omega := \{x \mid x = \phi_\theta(\eta)\}$, where η is a (randomly chosen, fixed) latent code vector.*

As a classical data-free DGM, DIP [25] reveals that even an untrained convolutional network embodies a strong inductive bias: during the fitting process, it inherently prioritizes capturing dominant image components before eventually overfitting image corruptions. Consequently, optimizing the network parameters θ to map a fixed random input η onto a target image x that naturally exhibits favorable prior characteristics can be formalized as the following optimization problem (Definition 1):

$$\hat{\theta} = \arg\min_{\theta} \|\mathcal{A}\phi_\theta(\eta) - y\|_2^2, \quad \hat{x} = \phi_{\hat{\theta}}(\eta), \tag{2}$$

where $\hat{x}$ denotes the reconstructed image. However, existing DIP-based methods frequently encounter challenges, particularly suboptimal reconstruction accuracy due to overfitting and inefficient convergence.

2.2. ARMANI

We propose a DL-inspired data-free DGM, termed ARMANI (see Figure 1). Firstly, one of the central insights is that the input latent code η is jointly optimized with θ under a dual-orientation sparsity constraint on the gradient domain, within an objective function akin to a classical DL framework. Specifically, a convex ℓ_1-norm regularization imposed on the horizontal and vertical gradients of η promotes piecewise smoothness in homogeneous regions while preserving salient structural edges. Meanwhile, a ridge penalty on θ can be viewed, from a DL-inspired perspective, as constraining the adaptive update of the empirical sparsifying dictionary represented by the network parameters, thereby enforcing scale consistency across the naturally encoded dictionary atoms.

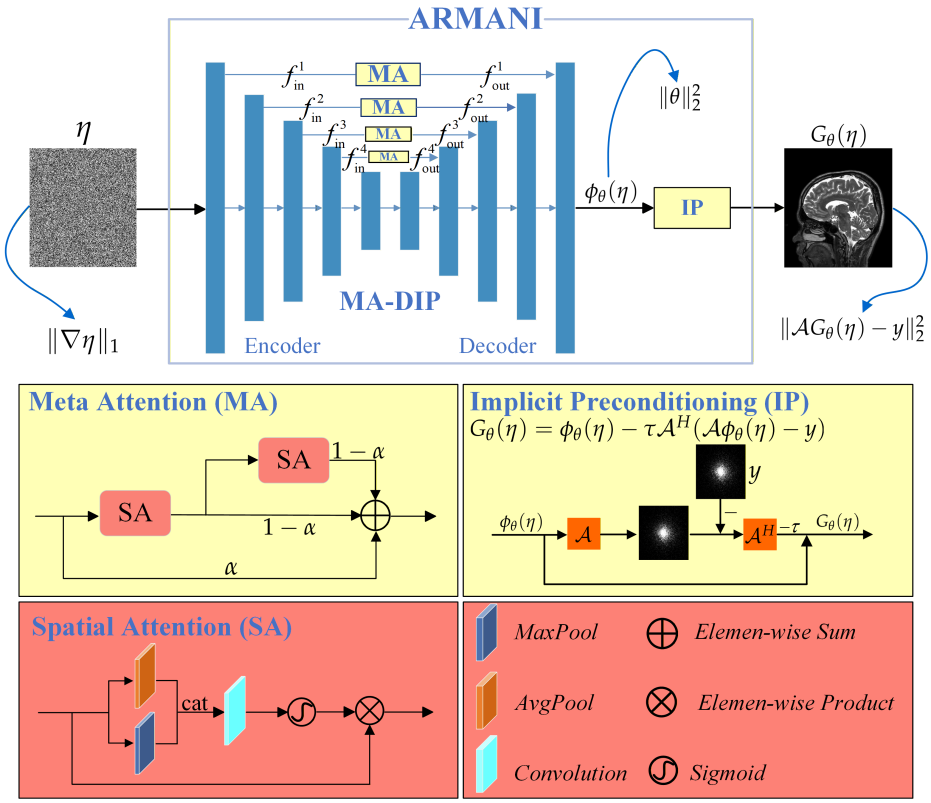


Figure 1. The ARMANI framework. The design consists of two key components: (i) The MA module is inserted into the skip connections and adopts an attention-in-attention design built upon the standard spatial attention (SA) mechanism. (ii) A pseudo-orthogonal projection of the MA-DIP output serves as an implicit preconditioning (IP) stage. The encoder consists of downsampling blocks with 128 convolutional filters of size 3×3, while the decoder symmetrically mirrors this structure through upsampling blocks.

Then, building upon the DIP architecture, we further integrate an MA module into the encoder–decoder generator $\phi_\theta(\eta)$. The MA module adopts a hierarchical refinement strategy, i.e., an "attention-in-attention" design, which acts as a meta-level transformation to adaptively reallocate attention weights with improved numerical stability and convergence behavior. Specifically, an initial spatial attention (SA) [40,41] provides a first weighting, while a subsequent refinement $\mathrm{SA}(\mathrm{SA}(\cdot))$ sharpens this distribution to emphasize salient anatomical regions. Formally, the features in the i-th skip connection, $f_{\text{in}}^{(i)}$, are updated via a multi-step, numerical-scheme-inspired rule:

$$\begin{aligned} f_{\text{out}}^{(i)} &= \alpha f_{\text{in}}^{(i)} + \beta \mathrm{MA}\left(f_{\text{in}}^{(i)}\right) \\ &= \alpha f_{\text{in}}^{(i)} + \beta \mathrm{SA}\left(f_{\text{in}}^{(i)}\right) + \beta \mathrm{SA}\left(\mathrm{SA}\left(f_{\text{in}}^{(i)}\right)\right), \end{aligned} \tag{3}$$

where the learnable parameter α (with $\beta = 1 - \alpha$) mediates the trade-off between preserving intrinsic structural information and incorporating guidance from the refined attention. This multi-stage design can be interpreted as a discrete instantiation of an underlying continuous dynamical system [42], where classical stability theory offers theoretical guarantees for robust and reliable multi-scale feature interactions.

Furthermore, to improve numerical stability and convergence, a preconditioning strategy is incorporated. Rather than explicitly constructing a preconditioner, which typically involves computationally intensive matrix inversions, we utilize $\phi_\theta(\eta)$ as an initial guess and perform a pseudo-Euclidean projection onto the set C, where $C = \{x \in \mathbb{C}^N | y = \mathcal{A}x + n\}$:

$$G_\theta(\eta) = \phi_\theta(\eta) - \tau \mathcal{A}^H(\mathcal{A}\phi_\theta(\eta) - y), \tag{4}$$

where τ is a step size. Next, we obtain a new data-fidelity term,

$$\begin{aligned} &\|\mathcal{A}G_\theta(\eta) - y\|_2^2 \\ &= \|\mathcal{A}(\phi_\theta(\eta) - \tau\mathcal{A}^H(\mathcal{A}\phi_\theta(\eta) - y)) - y\|_2^2 \\ &= \|(I - \tau\mathcal{A}\mathcal{A}^H)(\mathcal{A}\phi_\theta(\eta) - y)\|_2^2 \\ &= \|P^{-1}(\mathcal{A}\phi_\theta(\eta) - y)\|_2^2, \end{aligned} \tag{5}$$

Therefore, we get the total hybrid loss function for θ, η :

$$\begin{aligned} (\hat{\theta}, \hat{\eta}) = \arg\min_{\theta,\eta} &\|P^{-1}(\mathcal{A}\phi_\theta(\eta) - y)\|_2^2 \\ &+ \lambda_1\|\theta\|_2^2 + \lambda_2(\|\nabla_h\eta\|_1 + \|\nabla_v\eta\|_1), \end{aligned} \tag{6}$$

where $\nabla_{\{h,v\}}$ denote discrete finite-difference operators. $\lambda_1 > 0$ and $\lambda_2 > 0$ serve as balancing factors. This formulation arises naturally from interpreting $P = (I - \tau\mathcal{A}\mathcal{A}^H)^{-1}$ as a spectrally adjusted operator in the residual domain [43,44], which suppresses components along ill-conditioned directions that amplify noise, while preserving modes relevant to the signal. The preconditioning effect on the system $P^{-1}\mathcal{A}$ is theoretically analyzed in Appendix A under an idealized linear operator model. The complete iterative reconstruction procedure of ARMANI is summarized in Algorithm 1. Consequently, the ARMANI framework is essentially equivalent to the unrolling of a single-step proximal gradient descent algorithm [45], which underscores the validity of model-driven deep learning flow [46].

Algorithm 1 ARMANI

Require: Measurement y, forward encoding operator $\mathcal{A}$, maximum number of iterations T, step size τ, regularization weights λ_1, λ_2

1: Randomly initialize the latent input η and network parameters θ
2: **for** $t = 1$ to T **do**
3: Generate the image estimate $\phi_\theta(\eta)$
4: Perform implicit preconditioning:

$$G_\theta(\eta) = \phi_\theta(\eta) - \tau \mathcal{A}^H\left(\mathcal{A}\phi_\theta(\eta) - y\right)$$

5: Compute the hybrid loss:

$$\mathcal{L} = \|\mathcal{A}G_\theta(\eta) - y\|_2^2 + \lambda_1\|\theta\|_2^2 + \lambda_2(\|\nabla_h\eta\|_1 + \|\nabla_v\eta\|_1)$$

6: Update θ and η using Adam
7: **end for**
8: **return** the reconstructed image $G_\theta(\eta)$

3. Experiment and Results

3.1. Experimental Setting

To assess the performance of our ARMANI approach, two retrospective and one prospective experiments are conducted on the following three in vivo human datasets: Dataset #1 [38]: 3D brain scans of five healthy subjects were acquired with a 12-channel coil using a 3D T2 CUBE sequence (the readout (RO) $\times$ phase-encoding (PE) $\times$ slice: 256 $\times$ 232 $\times$ 208). A single slice was retrospectively undersampled using a 2D variable-density (VD) pattern with an acceleration factor (AF) of 8. Dataset #2 [47]: 2D knee scans were collected on a 3T Siemens scanner using a TSE sequence with a 15-channel coil (RO $\times$ PE $\times$ slice: 384 $\times$ 384 $\times$ 227). A single slice was retrospectively undersampled along the PE direction using a 1D VD sampling pattern with an AF of 5. Dataset #3: Prospectively undersampled data were obtained from healthy volunteers on a 3T Siemens scanner following a 1D VD pattern with an AF of 3, consisting of 106 phase-encoding lines and 64 fully sampled central lines. Each scan has a matrix size of 320 $\times$ 320, TE/TR = 5.1 ms/200 ms, FOV = 220 $\times$ 220 mm^2, and slice thickness = 5.0 mm.

We compared the proposed ARMANI approach against six classical SOTA methods, including BM3D-FISTA [48], Self2Self [49], SSDU [50], ConvDecoder-TV [26], Self-Guided DIP [27], and MoDL [38]. BM3D-FISTA is a conventional iterative reconstruction method incorporating plug-and-play priors. Self2Self and SSDU are self-supervised learning methods, whereas ConvDecoder-TV and Self-Guided DIP are unsupervised deep learning approaches that exploit either generative architectures or generative training. MoDL serves as a representative supervised baseline trained on paired datasets. It is worth noting that recent diffusion-model-based MRI reconstruction methods were not included in the present comparison, mainly because such methods generally rely on pretrained generative priors learned from external datasets, and their performance is often sensitive to the amount, quality, and distribution matching of the training data.

For implementation, BM3D-FISTA was realized in MATLAB (R2018a), while all deep learning models were implemented in PyTorch 1.13.1 on an NVIDIA V100 GPU. All intermediate convolutional layers were configured with 128 filters with a kernel size of 3 $\times$ 3. The network weights θ were optimized using Adam with a learning rate of 1×10^{-4} for 30,000 iterations, while the latent code η was initialized with random noise and jointly optimized using Adam with a learning rate of 1×10^{-3}. Coil sensitivity maps were estimated from undersampled k-space using the ESPIRiT algorithm [51]. For hyperparameter tuning, the step size τ was selected according to the condition in Appendix A, i.e.,

$0 < \tau \leq 1/(3\sigma_{\max}^2)$. The regularization weights λ_1 and λ_2 were initialized in proportion to the relative magnitudes of $\|\theta\|_2^2$ and $\|\nabla_h \eta\|_1 + \|\nabla_v \eta\|_1$ with respect to the preconditioned data-fidelity term $\|P^{-1}(A\phi_\theta(\eta) - y)\|_2^2$, and were then mildly adjusted empirically to ensure stable convergence.

Reconstruction quality was assessed using the relative ℓ_2-norm error (RLNE), the peak signal-to-noise ratio (PSNR), and the structural similarity index (SSIM), all computed within the tissue-containing region of interest (ROI). These metrics are defined as follows:

$$\begin{aligned} \mathrm{RLNE}_{\mathrm{ROI}} &= \|\hat{x}_{\mathrm{ROI}} - x^*_{\mathrm{ROI}}\|_2 \;/\; \|x^*_{\mathrm{ROI}}\|_2, \\ \mathrm{PSNR}_{\mathrm{ROI}} &= 20\log_{10}\Big(\max(x^*_{\mathrm{ROI}})\sqrt{N} \;/\; \|\hat{x}_{\mathrm{ROI}} - x^*_{\mathrm{ROI}}\|_2\Big), \\ \mathrm{SSIM}_{\mathrm{ROI}} &= \mathrm{SSIM}(\hat{x}_{\mathrm{ROI}}, x^*_{\mathrm{ROI}}), \end{aligned} \tag{7}$$

where x^* is the ground truth (GT), $\hat{x}$ denotes the reconstructed image, and N denotes the number of pixels in the ROI.

3.2. Results

Figure 2 presents the retrospective reconstruction results for retrospectively undersampled VD acquisitions from Dataset #1 (A) and Dataset #2 (B). As shown in Figure 2A, BM3D-FISTA exhibits the lowest reconstruction quality, characterized by evident blurring and noise-like artifacts, which are reflected by widespread residuals in the error maps. Self2Self does not present pronounced granular artifacts; however, edge definition remains suboptimal and certain fine anatomical structures are not fully recovered. SSDU and ConvDecoder-TV demonstrate comparable performance, providing clearer delineation of major anatomical structures and reduced residual errors. Self-Guided DIP achieves higher reconstruction fidelity, with sharper structural boundaries and lower error magnitudes in the corresponding difference maps. In comparison, the proposed ARMANI method delivers the most accurate reconstruction, with minimal residual artifacts and superior preservation of fine-scale anatomical details. A similar trend can be observed in Figure 2B, where ARMANI produces reconstructions that are closest to the fully sampled reference. These qualitative observations are consistent with the quantitative results summarized in Tables 1 and 2. Across both datasets, ARMANI achieves the lowest $\mathrm{RLNE}_{\mathrm{ROI}}$, the highest $\mathrm{PSNR}_{\mathrm{ROI}}$, and the highest $\mathrm{SSIM}_{\mathrm{ROI}}$ among all considered unsupervised and self-supervised methods. In addition, the reconstruction time comparison in Table 3 shows that ARMANI achieves better reconstruction quality within an acceptable computational cost.

Table 1. Quantitative reconstruction metrics on Dataset #1 (Brain, 2D, AF = 8). The mean value and standard deviation are reported for each metric, and the best performance per metric is highlighted in bold.

Methods	$\mathrm{RLNE}_{\mathrm{ROI}}$	$\mathrm{PSNR}_{\mathrm{ROI}}$	$\mathrm{SSIM}_{\mathrm{ROI}}$
BM3D-FISTA	0.0983 ± 0.0005	38.33 ± 0.03	0.9750 ± 0.0006
Self2Self	0.0945 ± 0.0014	38.59 ± 0.11	0.9751 ± 0.0011
SSDU	0.0894 ± 0.0006	39.17 ± 0.05	0.9774 ± 0.0006
ConvDecoder-TV	0.0871 ± 0.0012	39.23 ± 0.11	0.9787 ± 0.0009
Self-guided DIP	0.0823 ± 0.0011	39.71 ± 0.09	0.9789 ± 0.0007
ARMANI	**0.0752 ± 0.0009**	**40.54 ± 0.07**	**0.9810 ± 0.0007**

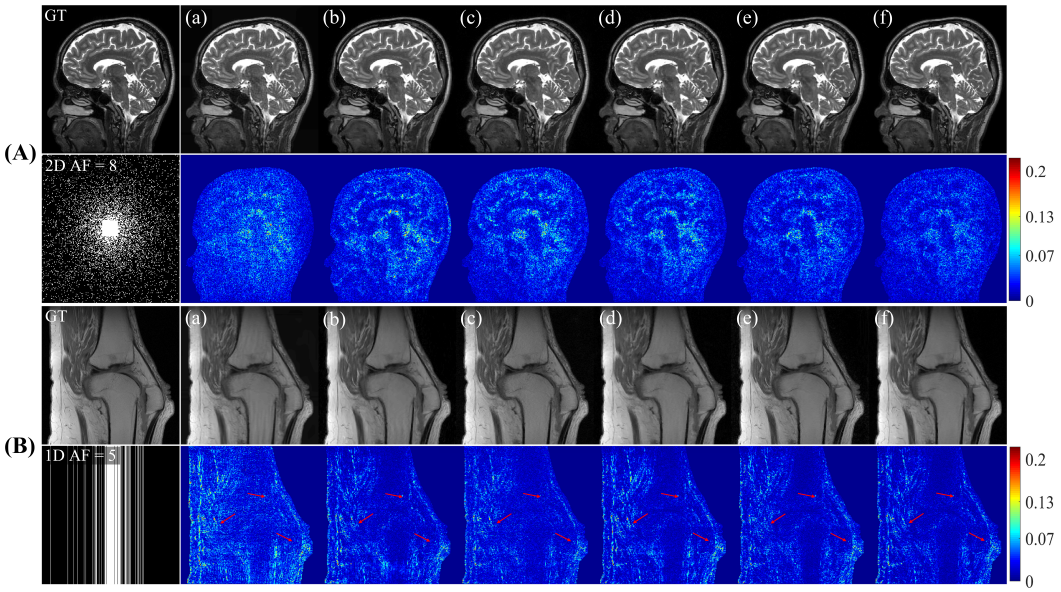


Figure 2. Retrospective reconstruction results on Dataset #1 (**A**) (Brain, 2D, AF = 8) and Dataset #2 (**B**) (Knee, 1D, AF = 5). In both (**A**) and (**B**), reconstruction methods are ordered identically as follows: (**a**) BM3D-FISTA, (**b**) Self2Self, (**c**) SSDU, (**d**) ConvDecoder-TV, (**e**) Self-Guided DIP, and (**f**) ARMANI. The corresponding error maps show the pixel-wise absolute differences between the reconstructed images and the ground truth (GT). Red arrows mark regions of visible differences.

Table 2. Quantitative reconstruction metrics on Dataset #2 (Knee, 1D, AF = 5). The mean value and standard deviation are reported for each metric, and the best performance per metric is highlighted in bold.

Methods	$RLNE_{ROI}$	$PSNR_{ROI}$	$SSIM_{ROI}$
BM3D-FISTA	0.1082 ± 0.0006	32.84 ± 0.05	0.8686 ± 0.0007
Self2Self	0.0810 ± 0.0018	35.38 ± 0.16	0.8794 ± 0.0015
SSDU	0.0792 ± 0.0009	35.84 ± 0.06	0.8821 ± 0.0006
ConvDecoder-TV	0.0777 ± 0.0016	35.91 ± 0.13	0.8911 ± 0.0011
Self-guided DIP	0.0753 ± 0.0014	36.16 ± 0.12	0.8944 ± 0.0009
ARMANI	**0.0711 ± 0.0012**	**36.89 ± 0.09**	**0.9054 ± 0.0008**

Table 3. Reconstruction time on the selected slice from Dataset #1 (brain, 2D, AF = 8) and Dataset #2 (knee, 1D, AF = 5). The reported values are given in minutes.

Methods	Brain	Knee
BM3D-FISTA	2	3.5
Self2Self	20	30
SSDU	1	1.2
ConvDecoder-TV	4	5.9
Self-guided DIP	9	13.5
ARMANI	8	12

To further evaluate the effectiveness of ARMANI under practical conditions, we additionally tested it on the prospectively acquired brain dataset with 1D undersampling at an AF of 3. Figure 3 demonstrates that our proposed method achieves superior preservation of structural details (indicated by red arrows), thereby confirming the superiority under realistic clinical acquisition settings.

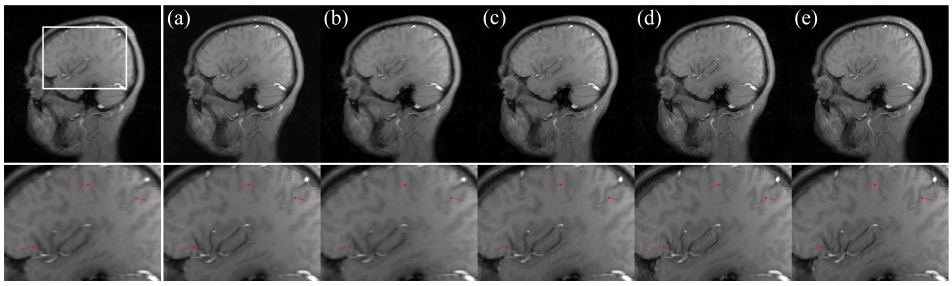


Figure 3. Prospective reconstruction results on Dataset #3 (Brain, 1D, AF = 3). Reconstruction methods: (**a**) BM3D-FISTA, (**b**) Self2Self, (**c**) ConvDecoder-TV, (**d**) Self-Guided DIP, and (**e**) ARMANI. Red arrows in the zoomed-in regions highlight areas where our ARMANI method achieves superior reconstruction quality for fine structural details. ZF denotes zero-filling reconstruction.

4. Discussion

4.1. Pointed Assessments with MoDL

Here, we specifically conducted a comparative experiment with the supervised deep learning method MoDL on Dataset #1 under the same experimental setup. MoDL served as the benchmark and was trained for 100 epochs using 360 brain images drawn from the same distribution as the test data. The qualitative results are presented in Figure 4, while the corresponding quantitative metrics are summarized in Table 4. Based on both visual inspection of the zoomed-in regions and quantitative metrics, it is evident that our approach achieves visually comparable, and in some cases slightly superior, reconstruction accuracy relative to MoDL. Notably, under these limited data conditions, our approach demonstrates a clear advantage, as it attains high-quality reconstruction without relying on extensive training samples, whereas MoDL is constrained by the small dataset size. These results underscore the significant potential of the proposed model and its training paradigm.

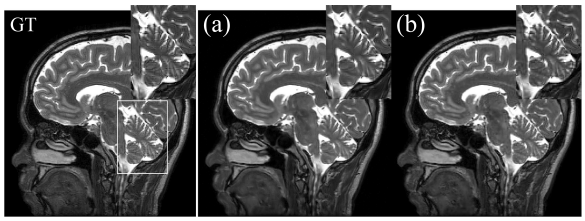


Figure 4. Pointed comparison with MoDL on Dataset #1 (Brain, 2D, AF = 8). (**a**) MoDL, (**b**) ARMANI. The upper-right corner of each image displays a zoomed-in view of a representative region to highlight structural differences.

Table 4. Quantitative comparison between ARMANI and MoDL on Dataset #1 (Brain, 2D, AF = 8). The mean value and standard deviation are reported for each metric, and the best performance per metric is highlighted in bold.

Methods	$RLNE_{ROI}$	$PSNR_{ROI}$	$SSIM_{ROI}$
MoDL	0.0795 ± 0.0006	40.21 ± 0.02	0.9797 ± 0.0004
ARMANI	**0.0752 ± 0.0009**	**40.54 ± 0.07**	**0.9810 ± 0.0007**

4.2. Ablation Study

An ablation study was performed on Dataset #1 to assess the contributions of different architectural modules and loss designs. The qualitative results are illustrated in Figure 5, while the corresponding quantitative metrics are summarized in Table 5. First, replacing the MA module with a simpler SA counterpart led to the presence of noticeable local noise in the reconstructed images (Figure 5b), although many structural details were still

preserved. From a quantitative standpoint, this substitution degraded the reconstruction quality, yielding higher $RLNE_{ROI}$, lower $PSNR_{ROI}$, and lower $SSIM_{ROI}$. These results highlight the advantage of the proposed MA module over a conventional attention design in suppressing noise while maintaining fine structural fidelity. Second, we further evaluated a variant without the MA module. As shown in Figure 5c, removing the attention mechanism entirely led to more severe structural degradation and increased residual artifacts. The corresponding quantitative results also became worse, indicating that the attention mechanism itself makes a practical contribution to the final reconstruction performance. Third, IP was found to play a pivotal role in stabilizing the optimization process by alleviating the ill-conditioning inherent to the inverse reconstruction problem. The removal of this component led to reconstructed images with pronounced artifacts and amplified noise (Figure 5d), alongside a significant deterioration in quantitative metrics. Overall, these ablation results confirm that both the MA module and the IP module are important for achieving the superior reconstruction quality of ARMANI.

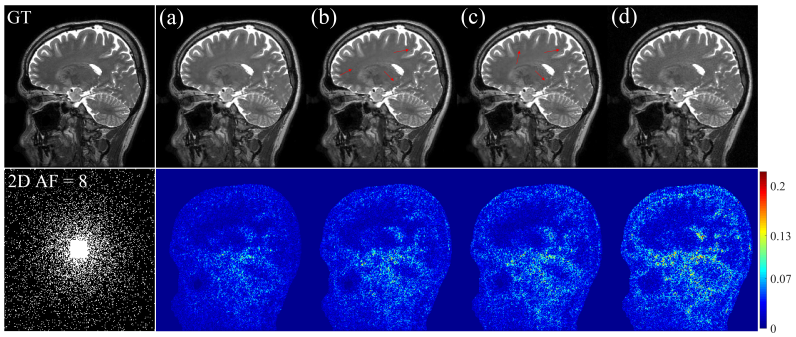


Figure 5. An ablation study on Dataset #1 (Brain, 2D, AF = 8). (**a**) ARMANI, (**b**) ARMANI with SA replacing MA, (**c**) ARMANI without MA, and (**d**) ARMANI without IP. The corresponding error maps show the pixel-wise absolute differences between the reconstructed images and the ground-truth (GT). Red arrows mark regions of visible differences. The impact of the core components on the overall performance is readily apparent.

Table 5. Quantitative ablation study results on Dataset #1 (Brain, 2D, AF = 8). The mean value and standard deviation are reported for each metric, and the best performance per metric is highlighted in bold.

Methods	$RLNE_{ROI}$	$PSNR_{ROI}$	$SSIM_{ROI}$
ARMANI	**0.0722 ± 0.0011**	**42.17 ± 0.08**	**0.9830 ± 0.0010**
ARMANI with SA replacing MA	0.0853 ± 0.0010	40.58 ± 0.06	0.9792 ± 0.0007
ARMANI without MA	0.0941 ± 0.0014	39.80 ± 0.11	0.9752 ± 0.0013
ARMANI without IP	0.1274 ± 0.0017	38.31 ± 0.14	0.9634 ± 0.0025

4.3. Future Work

Looking ahead, we plan to extend the proposed ARMANI framework to real-time cardiac imaging, which is a pressing scenario for unsupervised reconstruction. We also aim to refine the model through meta-learning-based attention strategies to enhance its performance in extremely data-scarce environments. Furthermore, we will investigate adaptive, learnable preconditioning mechanisms to accelerate convergence without compromising numerical stability. Beyond algorithmic refinement, future work will also include more systematic validation of algorithm–system integration in clinically realistic settings, in order to further strengthen the translational relevance of the proposed framework [52].

5. Conclusions

We present a novel unsupervised CS-MRI framework ARMANI that integrates DL formulation with a data-free DGM. Experimental results demonstrate that ARMANI exhibits strong scalability and shows great potential for deployment in fully data-free scenarios.

Author Contributions: Conceptualization, M.W. and Q.Z.; methodology, M.W. and Q.Z.; software, M.W.; validation, M.W. and J.C.; formal analysis, J.C.; resources, D.L.; writing—original draft preparation, M.W. and J.C.; writing—review and editing, Q.Z. and D.L.; visualization, M.W. and J.C.; supervision, Q.Z. and D.L.; project administration, Q.Z. and D.L.; funding acquisition, D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (62331028, 62125111), Guangdong Basic and Applied Basic Research Foundation (2023A1515110476) and Guangdong Provincial Key Laboratory of Multimodality Non-Invasive Brain-Computer Interfaces (2024B1212010010).

Data Availability Statement: The 3D brain dataset is publicly available at https://zenodo.org/records/6481291 (accessed on 15 March 2025). The 2D knee dataset is publicly available in the fastMRI repository at https://fastmri.org/ (accessed on 15 March 2025). The prospectively acquired 1D undersampled dataset (AF = 3) is available upon reasonable request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

We now proceed to analyze the condition number of the preconditioned system, $P^{-1}\mathcal{A}$. Consider the singular value decomposition $\mathcal{A} = U\Sigma V^H$, where Σ contains the singular values $\{\sigma_i\}$, and the preconditioned system is rewritten as

$$P^{-1}\mathcal{A} = (I - \tau\mathcal{A}\mathcal{A}^H)\mathcal{A} = U(I - \tau\Sigma^2)\Sigma V^H,$$

where the singular values of this system are $\sigma_i(P^{-1}\mathcal{A}) = \sigma_i|1 - \tau\sigma_i^2|$. For the range $0 < \tau \leq 1/(3\sigma_{\max}^2)$, we have

$$1 - \tau\sigma_i^2 \geq 1 - \tau\sigma_{\max}^2 \geq \frac{2}{3} > 0,$$

so the absolute value can be removed. Consider the function $f(\sigma) = \sigma(1 - \tau\sigma^2)$ with derivative $f'(\sigma) = 1 - 3\tau\sigma^2$. Since $\tau \leq 1/(3\sigma_{\max}^2)$, we have $f'(\sigma) \geq 0$ for all $\sigma \in [\sigma_{\min}, \sigma_{\max}]$; hence, $f(\sigma)$ is monotonically increasing. Therefore,

$$\sigma_{\max}(P^{-1}\mathcal{A}) = \sigma_{\max}(1 - \tau\sigma_{\max}^2),$$
$$\sigma_{\min}(P^{-1}\mathcal{A}) = \sigma_{\min}(1 - \tau\sigma_{\min}^2).$$

Thus, the condition number of $P^{-1}\mathcal{A}$ follows

$$\begin{aligned}\kappa(P^{-1}\mathcal{A}) &= \frac{\sigma_{\max}(1 - \tau\sigma_{\max}^2)}{\sigma_{\min}(1 - \tau\sigma_{\min}^2)} \\ &= \kappa(\mathcal{A}) \cdot \frac{1 - \tau\sigma_{\max}^2}{1 - \tau\sigma_{\min}^2} \\ &< \kappa(\mathcal{A}),\end{aligned}$$

where the strict inequality holds because $\sigma_{\max} > \sigma_{\min}$ and $\tau > 0$ imply $1 - \tau\sigma_{\max}^2 < 1 - \tau\sigma_{\min}^2$.

References

1. Zhu, Q.; Cui, Z.; Liu, Y.; Cheng, J.; Zhao, K.; Wang, H.; Zhu, Y.; Liang, D. Characteristic-constrained accelerating MR T1rho mapping with blockwise infimal convolution of matrix elastic-net regularization. *Med. Phys.* **2022**, *50*, 2224–2238. [CrossRef]
2. Liu, Y.; Liang, D.; Cui, Z.X.; Yang, Y.; Cao, C.; Zhu, Q.; Cheng, J.; Shi, C.; Wang, H.; Zhu, Y. Accelerating Magnetic Resonance T1ρ Mapping Using Simultaneously Spatial Patch-Based and Parametric Group-Based Low-Rank Tensors (SMART). *IEEE Trans. Med. Imaging* **2023**, *42*, 2247–2261. [CrossRef] [PubMed]
3. Hong, T.; Xu, X.; Hu, J.; Fessler, J.A. Provable Preconditioned Plug-and-Play Approach for Compressed Sensing MRI Reconstruction. *IEEE Trans. Comput. Imaging* **2024**, *10*, 1476–1488. [CrossRef] [PubMed]
4. Cao, C.; Cui, Z.; Zhu, Q.; Liu, C.; Liang, D.; Zhu, Y. Annihilation-Net: Learned annihilation relation for dynamic MR imaging. *Med. Phys.* **2024**, *51*, 1883–1898. [CrossRef] [PubMed]
5. Gutierrez, E.B.; Delplancke, C.; Ehrhardt, M.J. Randomized Primal-Dual Algorithm with Arbitrary Sampling: Convergence Theory and Applications to Parallel MRI. *J. Math. Imaging Vis.* **2025**, *67*, 1–17. [CrossRef]
6. Keil, B.; Wald, L.L. Massively parallel MRI detector arrays. *J. Magn. Reson.* **2013**, *229*, 75–89. [CrossRef]
7. Pfrommer, A.; Henning, A. The ultimate intrinsic signal-to-noise ratio of loop- and dipole-like current patterns in a realistic human head model. *Magn. Reson. Med.* **2018**, *80*, 2122–2138. [CrossRef]
8. Doneva, M. Mathematical Models for Magnetic Resonance Imaging Reconstruction: An Overview of the Approaches, Problems, and Future Research Areas. *IEEE Signal Process. Mag.* **2020**, *37*, 24–32. [CrossRef]
9. Fessler, J.A. Optimization Methods for Magnetic Resonance Image Reconstruction: Key Models and Optimization Algorithms. *IEEE Signal Process. Mag.* **2020**, *37*, 33–40. [CrossRef]
10. Du, T.; Zhang, H.; Li, Y.; Pickup, S.; Rosen, M.; Zhou, R.; Song, H.K.; Fan, Y. Adaptive convolutional neural networks for accelerating magnetic resonance imaging via k-space data interpolation. *Med. Image Anal.* **2021**, *72*, 102098. [CrossRef]
11. Chung, H.; Ye, J.C. Score-based diffusion models for accelerated MRI. *Med. Image Anal.* **2022**, *80*, 102479. [CrossRef]
12. Akcakaya, M.; Oh, G.; Ye, J.C. k-Space Deep Learning for MR Reconstruction and Artifact Removal. In *Deep Learning for Biomedical Image Reconstruction*; Cambridge University Press: Cambridge, UK, 2023; pp. 200–222.
13. Guo, P.; Mei, Y.; Zhou, J.; Jiang, S.; Patel, V.M. ReconFormer: Accelerated MRI Reconstruction Using Recurrent Transformer. *IEEE Trans. Med. Imaging* **2024**, *43*, 582–593. [CrossRef]
14. Xin, B.; Ye, M.; Axel, L.; Metaxas, D.N. Rethinking Deep Unrolled Model forAccelerated MRI Reconstruction. In Proceedings of the 18th European Conference on Computer Vision—ECCV 2024, Milan, Italy, 29 September–4 October 2024; pp. 164–181.
15. 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 Proceedings of the 23rd International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2020), Lima, Peru, 4–8 October 2020; pp. 64–73.
16. Huang, Q.; Xian, Y.; Yang, D.; Qu, H.; Yi, J.; Wu, P.; Metaxas, D.N. Dynamic MRI reconstruction with end-to-end motion-guided network. *Med. Image Anal.* **2021**, *68*, 101901. [CrossRef]
17. Zhang, J.; Ghanem, B. ISTA-Net: Interpretable Optimization-Inspired Deep Network for Image Compressive Sensing. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 1828–1837.
18. Song, J.; Chen, B.; Zhang, J. Memory-Augmented Deep Unfolding Network for Compressive Sensing. In Proceedings of the 29th ACM International Conference on Multimedia (ACM MM 2021), Online, 20–24 October 2021; pp. 4249–4258.
19. Oh, G.; Sim, B.; Chung, H.; Sunwoo, L.; Ye, J.C. Unpaired Deep Learning for Accelerated MRI Using Optimal Transport Driven CycleGAN. *IEEE Trans. Comput. Imaging* **2020**, *6*, 1285–1296. [CrossRef]
20. Chong, Z.Y.; Zhao, Y.; Wang, Z.; Lam, E.Y. Solving Inverse Problems in Compressive Imaging with Score-Based Generative Models. In Proceedings of the 2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA), Thessaloniki, Greece, 9–13 October 2023; pp. 1–10.
21. Cui, Z.; Cao, C.; Liu, S.; Zhu, Q.; Cheng, J.; Wang, H.; Zhu, Y.; Liang, D. Self-Score: Self-Supervised Learning on Score-Based Models for MRI Reconstruction. *arXiv* **2022**, arXiv:2209.00835.
22. Bouhlel, N.; Dziri, A. Kullback–Leibler Divergence Between Multivariate Generalized Gaussian Distributions. *IEEE Signal Process. Lett.* **2019**, *26*, 1021–1025. [CrossRef]
23. Zhang, Y.; Pan, J.; Li, K.; Liu, W.; Chen, Z.; Liu, X.; Wang, J. On the properties of kullback-leibler divergence between multivariate Gaussian distributions. In Proceedings of the 37th Conference on Neural Information Processing Systems (NeurIPS 2023), New Orleans, LA, USA, 10–16 December 2023; pp. 58152–58165.
24. Heckel, R.; Hand, P. Deep Decoder: Concise Image Representations from Untrained Non-convolutional Networks. *arXiv* **2018**, arXiv:1810.03982.
25. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Deep Image Prior. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.

26. Darestani, M.; Heckel, R. Accelerated MRI with Un-Trained Neural Networks. *IEEE Trans. Comput. Imaging* **2021**, *7*, 724–733. [CrossRef]
27. Liang, S.; Bell, E.; Qu, Q.; Wang, R.; Ravishankar, S. Analysis of Deep Image Prior and Exploiting Self-Guidance for Image Reconstruction. *IEEE Trans. Comput. Imaging* **2025**, *11*, 435–451. [CrossRef]
28. Zhu, Q.; Liu, B.; Cui, Z.X.; Cao, C.; Yan, X.; Liu, Y.; Cheng, J.; Zhou, Y.; Zhu, Y.; Wang, H.; et al. PEARL: Cascaded Self-Supervised Cross-Fusion Learning for Parallel MRI Acceleration. *IEEE J. Biomed. Health Inform.* **2025**, *29*, 3086–3097. [CrossRef]
29. Shen, L.; Pauly, J.; Xing, L. NeRP: Implicit Neural Representation Learning with Prior Embedding for Sparsely Sampled Image Reconstruction. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *35*, 770–782. [CrossRef] [PubMed]
30. Sultan, M.A.; Chen, C.; Liu, Y.; Gil, K.; Zareba, K.; Ahmad, R. An unsupervised method for MRI recovery: Deep image prior with structured sparsity. *Magn. Reson. Mater. Phys. Biol. Med.* **2025**, *38*, 859–871. [CrossRef] [PubMed]
31. Feng, J.; Feng, R.; Wu, Q.; Shen, X.; Chen, L.; Li, X.; Feng, L.; Chen, J.; Zhang, Z.; Liu, C.; et al. Spatiotemporal Implicit Neural Representation for Unsupervised Dynamic MRI Reconstruction. *IEEE Trans. Med. Imaging* **2025**, *44*, 2143–2156. [CrossRef] [PubMed]
32. Shen, X.; Feng, J.; Li, Z.; Zou, Q.; Zhang, Y.; Wei, H. Free-breathing dynamic MRI reconstruction via joint time-dependent coil sensitivity estimation using implicit neural representation. *Med. Image Anal.* **2026**, *108*, 103847. [CrossRef]
33. Joo, J.; Kim, H.; Won, H.; Lee, D.; Eo, T.; Hwang, D. AeSPa: Attention-guided Self-supervised Parallel imaging for MRI Reconstruction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2025), Nashville, TN, USA, 11–15 June 2025; pp. 5217–5226.
34. Alkhouri, I.; Liang, S.; Bell, E.; Qu, Q.; Wang, R.; Ravishankar, S. Image Reconstruction Via Autoencoding Sequential Deep Image Prior. In Proceedings of the Thirty-eighth Annual Conference on Neural Information Processing Systems (NeurIPS 2024), Vancouver, BC, Canada, 10–15 December 2024; pp. 18988–19012.
35. Qayyum, A.; Ilahi, I.; Shamshad, F.; Boussaid, F.; Bennamoun, M.; Qadir, J. Untrained Neural Network Priors for Inverse Imaging Problems: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 6511–6536. [CrossRef]
36. Liu, T.; Chaman, A.; Belius, D.; Dokmanić, I. Learning Multiscale Convolutional Dictionaries for Image Reconstruction. *IEEE Trans. Comput. Imaging* **2022**, *8*, 425–437. [CrossRef]
37. Kofler, A.; Pali, M.; Schaeffter, T.; Kolbitsch, C. Deep supervised dictionary learning by algorithm unrolling—Application to fast 2D dynamic MR image reconstruction. *Med. Phys.* **2023**, *50*, 2939–2960. [CrossRef]
38. 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]
39. Jagatap, G.; Hegde, C. Phase Retrieval using Untrained Neural Network Priors. In Proceedings of the NeurIPS 2019 Workshop on Solving Inverse Problems with Deep Networks, Vancouver, BC, Canada, 13 December 2019.
40. Guo, M.H.; Xu, T.X.; Liu, J.J.; Liu, Z.N.; Jiang, P.T.; Mu, T.J.; Zhang, S.H.; Martin, R.R.; Cheng, M.M.; Hu, S.M. Attention Mechanisms in Computer Vision: A Survey. *Comp. Vis. Media* **2021**, *8*, 331–368. [CrossRef]
41. Zhang, Y.; Li, K.; Li, K.; Fu, Y.R. MR Image Super-Resolution with Squeeze and Excitation Reasoning Attention Network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021), Nashville, TN, USA, 20–25 June 2021; pp. 13420–13429.
42. Lu, Y.; Zhong, A.; Li, Q.; Dong, B. Beyond Finite Layer Neural Networks: Bridging Deep Architectures and Numerical Differential Equations. In Proceedings of the 35th International Conference on Machine Learning (ICML 2018), Stockholm, Sweden, 10–15 July 2018; Volume 80, pp. 3276–3285.
43. Gilton, D.; Ongie, G.; Willett, R. Learning to Regularize Using Neumann Networks. In *Proceedings of the 2019 IEEE Data Science Workshop (DSW)*; IEEE: New York, NY, USA, 2019; pp. 201–207.
44. Gilton, D.; Ongie, G.; Willett, R. Neumann networks for linear inverse problems in imaging. *IEEE Trans. Comput. Imaging* **2019**, *6*, 328–343. [CrossRef]
45. Malitsky, Y.; Mishchenko, K. Adaptive proximal gradient method for convex optimization. In Proceedings of the 38th Conference on Neural Information Processing Systems (NeurIPS 2024), Vancouver, BC, Canada, 10–15 December 2024; pp. 100670–100697.
46. Shlezinger, N.; Whang, J.; Eldar, Y.C.; Dimakis, A.G. Model-Based Deep Learning. *Proc. IEEE* **2023**, *111*, 465–499. [CrossRef]
47. Zbontar, J.; Knoll, F.; Sriram, A.; Muckley, M.J.; Murrell, T. fastMRI: An Open Dataset and Benchmarks for Accelerated MRI. *arXiv* **2018**, arXiv:1811.08839.
48. Kamilov, U.; 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]
49. Quan, Y.; Chen, M.; Pang, T.; Ji, H. Self2Self with Dropout: Learning Self-Supervised Denoising from Single Image. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020), Seattle, WA, USA, 13–19 June 2020; pp. 1887–1895.
50. Yaman, B.; Hosseini, S.A.H.; Moeller, S.; Ellermann, J.; Uurbil, K.; Akakaya, M. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magn. Reson. Med.* **2020**, *84*, 3172–3191. [CrossRef]

51. Uecker, M.; Lai, P.; Murphy, M.; Virtue, P.; Elad, M. ESPIRiT-an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA. *Magn. Reson. Med.* **2014**, *71*, 990–1001. [CrossRef]
52. Laganà, F.; Prattico, D.; De Carlo, D.; Oliva, G.; Pullano, S.A.; Calcagno, S. Engineering Biomedical Problems to Detect Carcinomas: A Tomographic Impedance Approach. *Eng* **2024**, *5*, 1594–1614. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.