

Fill the K-Space and Refine the Image: Prompting for Dynamic and Multi-Contrast MRI Reconstruction

Bingyu Xin¹, Meng Ye¹, Leon Axel², and Dimitris N. Metaxas¹

¹ Department of Computer Science, Rutgers University, Piscataway, NJ 08854, USA

² Department of Radiology, New York University, New York, NY 10016, USA

{bx64, my389, dnm}@cs.rutgers.edu
<https://github.com/hellopipu/PromptMR>

Abstract. The key to dynamic or multi-contrast magnetic resonance imaging (MRI) reconstruction lies in exploring inter-frame or inter-contrast information. Currently, the unrolled model, an approach combining iterative MRI reconstruction steps with learnable neural network layers, stands as the best-performing method for MRI reconstruction. However, there are two main limitations to overcome: firstly, the unrolled model structure and GPU memory constraints restrict the capacity of each denoising block in the network, impeding the effective extraction of detailed features for reconstruction; secondly, the existing model lacks the flexibility to adapt to variations in the input, such as different contrasts, resolutions or views, necessitating the training of separate models for each input type, which is inefficient and may lead to insufficient reconstruction. In this paper, we propose a two-stage MRI reconstruction pipeline to address these limitations. The first stage involves *filling the missing k-space data*, which we approach as a physics-based reconstruction problem. We first propose a simple yet efficient baseline model, which utilizes adjacent frames/contrasts and channel attention to capture the inherent inter-frame/-contrast correlation. Then, we extend the baseline model to a prompt-based learning approach, **PromptMR**, for all-in-one MRI reconstruction from different views, contrasts, adjacent types, and acceleration factors. The second stage is to *refine the reconstruction* from the first stage, which we treat as a general video restoration problem to further fuse features from neighboring frames/contrasts in the image domain. Extensive experiments show that our proposed method significantly outperforms previous state-of-the-art accelerated MRI reconstruction methods.

Keywords: MRI reconstruction · Prompt-based learning · Dynamic · Multi-contrast · Two-stage approach

1 Introduction

Cardiovascular disease, including conditions such as coronary artery disease, heart failure, and arrhythmias, remains the leading cause of death globally. Cardiac magnetic resonance (CMR) imaging is the most accurate and reliable non-invasive

technique for accessing cardiac anatomy, function, and pathology [13]. In the field of accelerated MR imaging (MRI) reconstruction, unrolled networks have achieved state-of-the-art performance. This is attributed to their ability to incorporate the known imaging degradation processes, the undersampling operation in k-space, into the network and to learn image priors from large-scale data [16,2]. As transformers have become predominant in general image restoration tasks [18,9], there is a noticeable trend towards incorporating transformer-based denoising blocks into the unrolled network [2], which enhances reconstruction quality. However, the adoption of transformer blocks concurrently increases the network parameters and computational complexity. The stacking of denoising blocks, in an unrolled manner, further exacerbates this complexity, making the network training challenging. Therefore, one challenging question is how to design efficient denoising blocks within an unrolled model while fully leveraging the k-space information. Another challenge arises from the versatility of MRI, which enables the acquisition of multi-view, multi-contrast, multi-slice, and dynamic image sequences, given specific clinical demands. While there is a prevailing trend towards designing all-in-one models for natural image restoration [7,12], existing MRI reconstruction models cannot offer a unified solution for diverse input types. We thus endeavor to address these challenges with the following contributions:

- Firstly, we propose a simple yet efficient convolution-only baseline model for MRI reconstruction, which outperforms previous state-of-the-art methods on two public multi-coil MRI reconstruction tasks, the CMRxRecon and fastMRI knee image reconstruction.
- Then, by extending our baseline model with prompt-based learning, we are the first to propose an all-in-one approach, **PromptMR**, for multi-view/-contrast and dynamic MRI reconstruction.
- Lastly, we extend our approach to address the capacity limitations of unrolled models, by proposing a two-stage MRI reconstruction pipeline. In the first stage we solve a physics-based inverse problem in k-space domain to fill the missing k-space data, and in the second stage we solve a video restoration problem in the image domain to further refine the MRI reconstruction.

2 Preliminaries

Consider reconstructing a complex-valued MR image x from the multi-coil undersampled measurements y in k-space, such that,

$$y = Ax + \epsilon, \quad (1)$$

where A is the linear forward complex operator which is constructed based on multiplications with the sensitivity maps S , application of the 2D Fourier transform F , while it under-samples the k-space data with a binary mask M ; ϵ is the acquisition noise. According to compressed sensing theory [1], we can estimate x by formulating an optimization problem:

$$\min_x \frac{1}{2} \|y - Ax\|_2^2 + \lambda R(x), \quad (2)$$

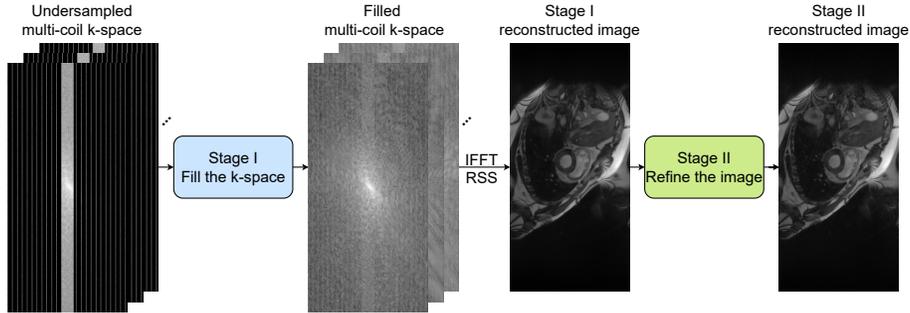


Fig. 1: The proposed two-stage MRI reconstruction pipeline. The first stage solves a physics-based inverse problem to fill the missing k-space data, which are then transformed to the image domain by the inverse Fast Fourier Transformation (IFFT) and root-sum-of-squares (RSS) is applied to get the first-stage reconstructed image. The second stage solves a general denoising problem to further refine the image reconstruction result.

where $\|y - Ax\|_2^2$ is the data consistency term, $R(x)$ is a sparsity regularization term on x (e.g., total variation) and λ is a hyper-parameter which controls the contribution weights of the two terms. E2E-VarNet [16] solves the problem in Eq. 2 by applying an iterative gradient descent method in the k-space domain. In the t -th step, the k-space is updated from k^t to k^{t+1} using:

$$k^{t+1} = k^t - \eta^t M(k^t - y) + G(k^t), \quad (3)$$

where η^t is a learned step size and G is a learned function representing the gradient of the regularization term R . We can unroll the iterative updating algorithm to a sequence of sub-networks, where each cascade represents an unrolled iteration in Eq. 3. The regularization term is applied in the image domain:

$$G(k) = F(\mathcal{E}(\mathbf{D}(\mathcal{R}(F^{-1}(k))))), \quad (4)$$

where $\mathcal{R}(x_1, \dots, x_N) = \sum_{i=1}^N \hat{S}_i^* x_i$ is the reduce operator that combines N coil images $\{x_i\}_{i=1}^N$ via estimated sensitivity maps $\{\hat{S}_i\}_{i=1}^N$, \hat{S}_i^* is the complex conjugate of \hat{S}_i , and $\mathcal{E}(x) = (\hat{S}_1 x, \dots, \hat{S}_N x)$ is the expand operator that computes coil images from image x . Therefore, the linear forward operator A is computed as $A = MF\mathcal{E}$. \mathbf{D} is a denoising neural network used to refine the complex image. $\hat{S} = \text{SME}(y_{\text{ACS}})$ is computed by a sensitivity map estimation (SME) network from the low-frequency region of k-space y_{ACS} , called the Auto-Calibration Signal (ACS), which is typically fully sampled. The final updated multi-coil k-space is converted to the image domain by applying an inverse Fourier transform followed by a root-sum-of-squares (RSS) method reduction [14] for each pixel.

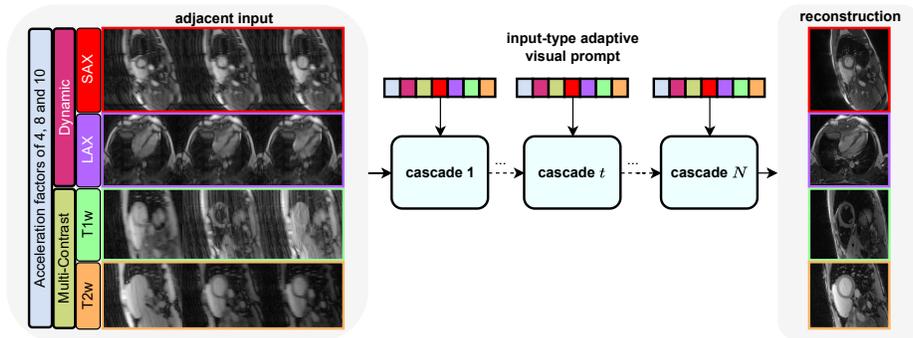


Fig. 2: Overview of PromptMR in Stage I: an all-in-one unrolled model for MRI reconstruction. Adjacent inputs, depicted in image domain for visual clarity, provide neighboring k-space information for reconstruction. To accommodate different input varieties, the input-type adaptive visual prompt is integrated into each cascade of the unrolled architecture to guide the reconstruction process.

3 Method

We propose a two-stage pipeline for dynamic and multi-contrast MRI reconstruction, as shown in Fig. 1. Below, we give more details of each stage.

3.1 Stage I: Filling the K-Space

The center of the k-space preserves image contrast, and the periphery of the k-space contains edge information. In the first stage, we fill the missing k-space data constrained by the existing k-space acquisition and learned image priors.

Baseline Model We follow the implementation of E2E-VarNet [16] to construct an unrolled model in Stage I. Inspired by the adjacent slice reconstruction (ASR) method [2], which learns inter-slice information by jointly reconstructing a set of adjacent slices instead of relying on a single k-space to be reconstructed, we devise the following new method. We generalize ASR to adjacent k-space reconstruction along any dimension, e.g., temporal/slice/view/contrast dimension, and the updating formula of Eq. 3 is improved as follows:

$$k_{adj}^{t+1} = k_{adj}^t - \eta^t A(k_{adj}^t - y_{adj}) + G(k_{adj}^t), \quad (5)$$

where $k_{adj}^t = [k_{c-a}^t, \dots, k_{c-1}^t, k_c^t, k_{c+1}^t, \dots, k_{c+a}^t]$ is the concatenation of the central k-space k_c^t with its $2a$ adjacent k-spaces along a specific dimension. To efficiently extract features from adjacent inputs, we design a Unet-style network [15] with channel attention [3,4], namely CAUnet, for both the denoising network D and the sensitivity map estimation network, as shown in Appendix A.1. The CAUnet has a 3-level encoder-decoder structure. Each level consists of a DownBlock,

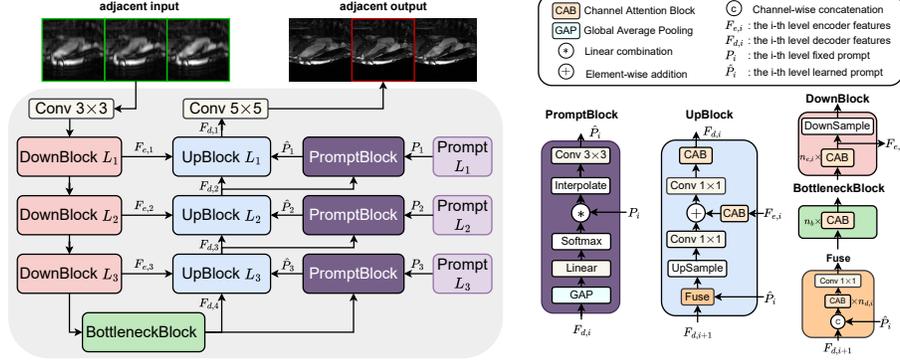


Fig. 3: Overview of the PromptUnet architecture in PromptMR, featuring a 3-level encoder-decoder design. Each level comprises a DownBlock, UpBlock and PromptBlock. The PromptBlock in the i -th level encodes input-specific context into fixed prompt P_i , producing adaptively learned prompt \hat{P}_i . These prompts, across multiple levels, integrate with decoder features $F_{d,i}$ in the UpBlocks to allow rich hierarchical context learning.

UpBlock, and corresponding skip connection. The architecture integrates a BottleneckBlock for high-level semantic feature capturing and employs Channel Attention Blocks (CABs) within each block. The overall unrolled architecture is shown in Appendix A.2.

PromptMR Considering various image types (e.g., different views, different contrasts) with different adjacent types (e.g., dynamic, multi-contrast) under different undersampling rates (e.g., $\times 4, \times 8, \times 10$), instead of training separate models for each specific input, we propose to learn an all-in-one unified model for all possible adjacent inputs. The image structure remains consistent for multi-contrast adjacent input, while only the contrast varies. Conversely, the contrast remains constant for dynamic adjacent input, but the image structure shifts. To achieve effective performance on diverse input types, the unified model should be able to encode the contextual information conditioned on the input type. Inspired by the recent development of visual prompt learning [5,6] and prompt learning-based image restoration method [12], we introduce PromptMR, an all-in-one approach for MRI reconstruction, as illustrated in Fig. 2. While PromptMR retains the same unrolled architecture of the baseline model, it extends CAUnet to PromptUnet by integrating PromptBlocks to learn input-type adaptive prompts and then interact with decoder features in the UpBlocks at multiple levels, to enrich the input-specific context, as shown in Fig. 3. The PromptBlock at i -th level takes features $F_{d,i} \in \mathbb{R}^{H_f \times W_f \times C_f}$ from the decoder and N_p -components fixed prompt $P_i \in \mathbb{R}^{N_p \times H_p \times W_p \times C_p}$ as input. Then, $F_{d,i}$ are processed by a global average pooling (GAP) layer, followed by a linear layer and a softmax layer to generate the normalized prompt weights $\{\omega_{ij}\}_{j=1}^{N_p}$. These weights linearly combine

the N_p prompt components as $\sum_{j=1}^{N_p} \omega_{ij} P_{ij}$, which is then interpolated to match the spatial dimension of $F_{d,i}$, before going through a 3×3 convolution layer to generate the input-type adaptive prompt \hat{P}_i . The process in the PromptBlock can be summarized as:

$$\hat{P}_i = \text{Conv}_{3 \times 3}(\text{Interp}(\sum_{j=1}^{N_p} \omega_{ij} P_{ij})), \quad \omega_i = \text{Softmax}(\text{Linear}(\text{GAP}(F_{d,i}))) \quad (6)$$

The generated prompts by the PromptBlocks at multiple levels can learn hierarchical input-type contextual representations, which are integrated with the decoder features to guide the all-in-one MRI reconstruction.

3.2 Stage II: Refining the Image

After the first stage, the missing k-space data has been filled, and image aliasing artifacts have been largely removed. However, due to the unrolled nature and memory limitations, the capability of the denoising block we can use is constrained, which may prevent the full exploration of dynamic and multi-contrast information. In stage II, we further explore the inter-frame/-contrast coherence in the image domain for multi-frame/-contrast feature aggregation by using a powerful restoration model, ShiftNet [8], as the refinement network. This network employs stacked Unets and grouped spatio-temporal shift operations to expand the effective receptive fields. Details of the ShiftNet are not covered here, since it is not the core part of this paper, and ShiftNet can be replaced by any state-of-the-art video restoration model.

4 Experiments

In this section, we first provide experimental details and results of our proposed method on the CMRxRecon dataset. We use SSIM, PSNR, and NMSE to compare the performance of different reconstruction methods under various acceleration factors ($\times 4$, $\times 8$, $\times 10$). Then, we conduct extensive ablation studies of our proposed method and also benchmark on another large-scale MRI dataset, the fastMRI multi-coil knee dataset. For experiments on fastMRI dataset, we refer readers to the Appendix B.

4.1 CMRxRecon Dataset

The CMRxRecon Dataset [17] includes 120 cardiac MRI cases of fully sampled dynamic cine and multi-contrast raw k-space data obtained on 3 Tesla magnets. The dynamic cine images in each case include short-axis (SAX), two-chamber (2-CH), three-chamber (3-CH), and four-chamber (4-CH) long-axis (LAX) views. Typically 5 ~ 10 slices were acquired for SAX cine, while a single slice was acquired for each LAX view. The cardiac cycle was segmented into 12 ~ 25 phases with a temporal resolution of 50 ms. The multi-contrast cardiac MRI in each case is in the SAX view, which contains 9 T1-weighted (T1w) images

conducted using a modified look-locker inversion recovery (MOLLI) sequence and 3 T2-weighted (T2w) images performed using T2-prepared FLASH sequence.

The shape of each k-space data is [time phases/contrasts, slices, coils, readouts, phase encodings]. All data were compressed into 10 virtual coils. We split the cases in an 8 : 2 ratio, resulting in 14, 964 dynamic images and 6, 516 multi-contrast images for training, and 2, 940 dynamic images and 1, 272 multi-contrast images for testing.

4.2 Results

We assessed the performance of our proposed baseline model, PromptMR, and two-stage reconstruction pipeline using the CMRxRecon dataset. In the first stage, we compared the E2E-VarNet [16] and HUMUS-Net-L [2] with our baseline in a one-by-one setup, in which we trained four separate models from scratch for SAX/LAX/T1w/T2w reconstruction task, respectively. Then we compared our PromptMR and PromptIR [12] in an all-in-one configuration. In the second stage, we deployed ShiftNet to refine the images reconstructed by PromptMR. In our experiment, we minimize the SSIM loss between the target image and the reconstructed image; all unrolled models consist of 12 cascades, except for HUMUS-Net-L, which only has 8 cascades due to its large parameter size; we trained networks using AdamW [10] optimizer with a weight decay of 0.01 for 12 epochs; the learning rate was set as 2×10^{-4} for the first 11 epochs and 2×10^{-5} for the last epoch.

The results are shown in Table 1. Notably, our baseline model outperforms E2E-VarNet and HUMUS-Net-L across all tasks. Moreover, our PromptMR demonstrates significant enhancement in the all-in-one setup when compared to the baseline model trained for individual tasks. PromptIR performs poorly due to the fact that it is not tailored to account for the MRI forward model. The refinement in the second stage offers a marginal boost to the SSIM, but provides considerable improvements for NMSE and PSNR. The qualitative results are shown in Fig. 4. More qualitative comparisons can be found in Appendix C. These qualitative comparisons show that our method can recover more finer details for small anatomical structures on the reconstructed images.

4.3 Ablation Study

Single MRI Reconstruction Task We started with an ablation study on two single MRI reconstruction tasks, dynamic cine SAX image reconstruction and multi-contrast T1-weighted image reconstruction, both under $\times 10$ acceleration, to investigate the impact of adjacent reconstruction and prompt module in the proposed PromptMR. We changed the number of adjacent images to 1, 3, and 5, where ‘1’ indicates the absence of adjacent input. The results, shown in Table 2, underscore the utility of incorporating adjacent input to enhance the reconstruction quality. Moreover, the inclusion of PromptBlocks proves beneficial for individual MRI reconstruction tasks.

Table 1: Comparison of NMSE($\times 10^{-2}$)/PSNR/SSIM of different MRI reconstruction methods on CMRxRecon dataset under $\times 10$ acceleration. The best and second best results are highlighted in red and blue colors, respectively.

Stage	Task	Method	Cine		Mapping	
			SAX	LAX	T1w	T2w
I	One-by-One	E2E-VarNet [16]	1.6/42.05/0.9744	2.1/39.93/0.9673	1.5/43.12/0.9800	1.4/41.20/0.9777
		HUMUS-Net-L [2]	1.3/42.96/0.9791	2.0/40.07/0.9689	1.3/43.85/0.9832	1.1/42.39/0.9824
		Baseline (Ours)	1.1/43.68/0.9814	1.9/40.38/0.9705	1.2/44.14/0.9839	0.9/43.12/0.9845
	All-in-One	PromptIR [12]	2.5/40.16/0.9659	2.7/38.62/0.9581	2.3/41.10/0.9726	1.4/41.10/0.9784
		PromptMR (Ours)	1.1/45.58/0.9865	1.2/43.72/0.9836	1.0/46.84/0.9899	0.7/46.24/0.9903
II	PromptMR+ShiftNet [8]	0.7/45.63/0.9866	0.9/43.76/0.9837	0.7/47.04/0.9903	0.5/46.33/0.9905	

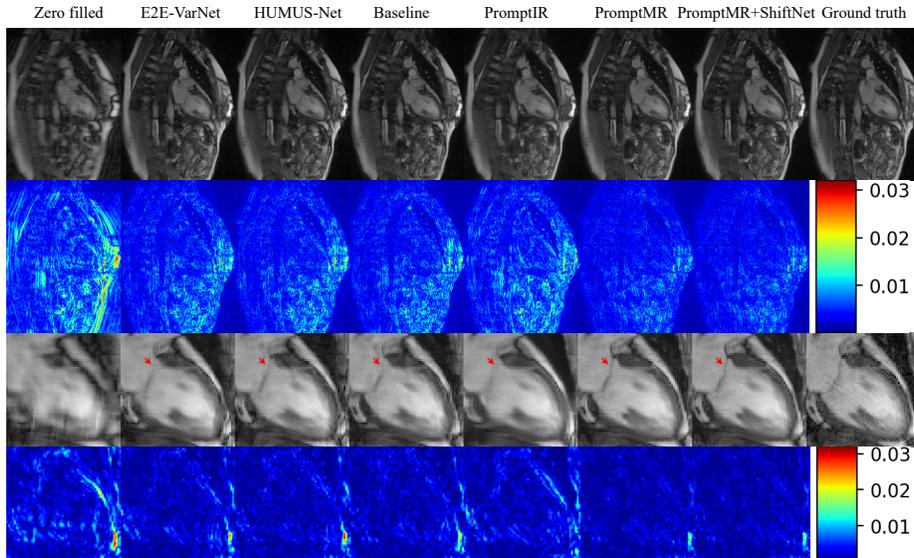


Fig. 4: The reconstruction results and absolute error maps of different methods for LAX 2-CH cine image of case P101 under $\times 10$ acceleration factors. The bottom two rows show the zoomed area. Red arrows show the difference in recovery of the mitral valve structure for different reconstruction methods.

All-In-One MRI Reconstruction Task To investigate the impact of the PromptBlock in the all-in-one MRI reconstruction task, we trained both our baseline and PromptMR model using all possible input data in the CMRxRecon dataset. As depicted in Table 3, the integration of the PromptBlock into our baseline model enables PromptMR to achieve significant improvements across all individual reconstruction tasks. We also used t-SNE [11] to visualize the learned prompts in the 12-th cascade at multiple decoder levels from different types of data in the test set. Fig. 5 shows that the prompts can learn to encode discriminative information for different input types at lower levels.

Table 2: Impact of the adjacent input number and the PromptBlock in PromptMR for two single MRI reconstruction tasks: dynamic cine SAX and multi-contrast T1-weighted (T1w) reconstruction under $\times 10$ acceleration.

# of adj	PromptBlock	SAX	T1w
		PSNR/SSIM	PSNR/SSIM
1	✓	43.19/0.9798	44.36/0.9845
3	✓	43.96/0.9822	44.78/0.9856
5	✓	43.87/0.9820	44.75/0.9856
5	✗	43.68/0.9814	44.14/0.9839

Table 3: Impact of PromptBlock in all-in-one task. Results are reported on the CMRxRecon dataset under $\times 10$ acceleration.

Method	SAX	LAX	T1w	T2w
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Baseline (Ours)	43.97/0.9825	42.11/0.9786	44.90/0.9862	44.45/0.9874
PromptMR (Ours)	45.58/0.9865	43.72/0.9836	46.84/0.9899	46.24/0.9903

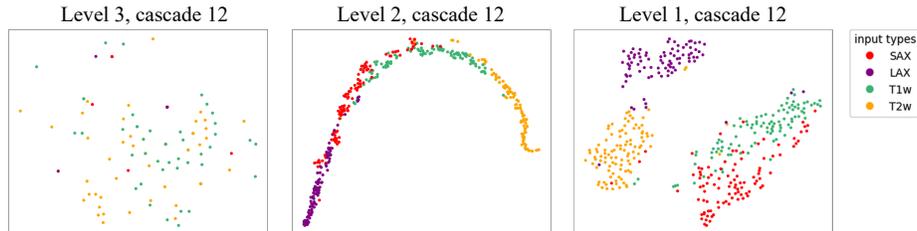


Fig. 5: Visualization of the learned prompts at each decoder level in the 12-th cascade in PromptMR using t-SNE.

5 Conclusion

In this work, we introduce a robust baseline model for MRI reconstruction that utilizes neighboring information of adjacent k-space. To accommodate various input types, adjacent configurations, and undersampling rates within a unified model, we enhance our baseline with prompt-based learning blocks, creating an all-in-one MRI reconstruction model, **PromptMR**. Finally, to overcome the model capacity constraints of unrolled architectures, we propose a second stage of image refinement to delve deeper into the adjacent information, which is particularly useful when immediate reconstruction latency is not a priority.

References

1. Donoho, D.L.: Compressed sensing. *IEEE Transactions on information theory* **52**(4), 1289–1306 (2006)
2. 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)
3. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 7132–7141 (2018)
4. Huang, Q., Yang, D., Wu, P., Qu, H., Yi, J., Metaxas, D.: Mri reconstruction via cascaded channel-wise attention network. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. pp. 1622–1626. IEEE (2019)
5. Jia, M., Tang, L., Chen, B.C., Cardie, C., Belongie, S., Hariharan, B., Lim, S.N.: Visual prompt tuning. In: *European Conference on Computer Vision*. pp. 709–727. Springer (2022)
6. Khattak, M.U., Rasheed, H., Maaz, M., Khan, S., Khan, F.S.: Maple: Multi-modal prompt learning. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 19113–19122 (2023)
7. Li, B., Liu, X., Hu, P., Wu, Z., Lv, J., Peng, X.: All-in-one image restoration for unknown corruption. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 17452–17462 (2022)
8. Li, D., Shi, X., Zhang, Y., Cheung, K.C., See, S., Wang, X., Qin, H., Li, H.: A simple baseline for video restoration with grouped spatial-temporal shift. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 9822–9832 (2023)
9. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 1833–1844 (2021)
10. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017)
11. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. *Journal of machine learning research* **9**(11) (2008)
12. Potlapalli, V., Zamir, S.W., Khan, S., Khan, F.S.: Promptir: Prompting for all-in-one blind image restoration. *arXiv preprint arXiv:2306.13090* (2023)
13. Rajiah, P.S., François, C.J., Leiner, T.: Cardiac mri: state of the art. *Radiology* p. 223008 (2023)
14. Roemer, P.B., Edelstein, W.A., Hayes, C.E., Souza, S.P., Mueller, O.M.: The nmr phased array. *Magnetic resonance in medicine* **16**(2), 192–225 (1990)
15. 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)
16. 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)
17. Wang, C., Lyu, J., Wang, S., Qin, C., Guo, K., Zhang, X., Yu, X., Li, Y., Wang, F., Jin, J., et al.: Cmrrecon: An open cardiac mri dataset for the competition of accelerated image reconstruction. *arXiv preprint arXiv:2309.10836* (2023)

18. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5728–5739 (2022)
19. 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)

Appendix

A Details of the Baseline Model

A.1 CAUnet

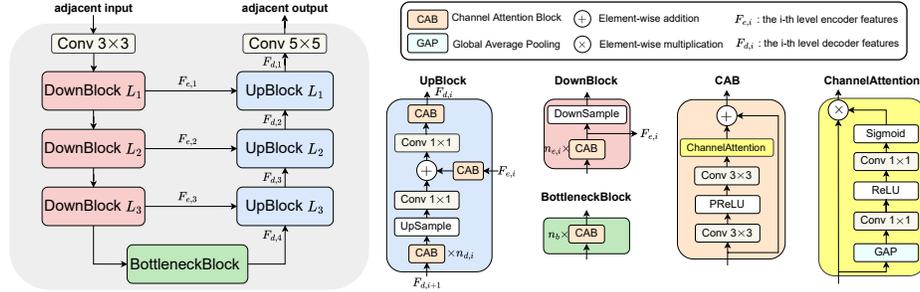


Fig. A.1: Overview of the CAUnet architecture in the proposed baseline model.

A.2 Unrolled model architecture

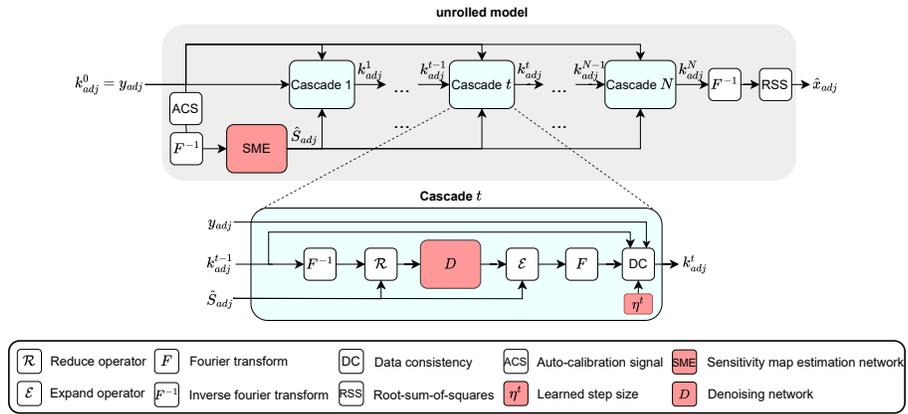


Fig. A.2: Overview of the unrolled model architecture for both our baseline model and PromptMR. The primary distinction is in the denoiser D and sensitivity map estimation (SME) networks: the baseline employs CAUnet, whereas PromptMR utilizes PromptUnet. Each cascade represents an updating step in Eq. 5 in the main text. The red module indicates the learnable part in the unrolled model.

B Experiments on FastMRI Multi-Coil Knee Dataset

Benchmark on FastMRI Multi-Coil Knee Dataset To assess the performance of our proposed method across different anatomies, we benchmarked it on another large-scale MRI reconstruction dataset, the fastMRI multi-coil knee dataset [19]. Since the online evaluation platform for the fastMRI test set is unavailable¹, we divided the original 199 validation cases into 99 for validation and 100 for testing. The results of other methods are reported using their officially pretrained models. As presented in Table B.1, our models outperform all previous state-of-the-art methods, without significantly increasing the number of network parameters compared to E2E-Varnet.

Table B.1: Performance of state-of-the-art accelerated MRI reconstruction techniques on the fastMRI knee multi-coil $\times 8$ test dataset. The best and second best results are highlighted in red and blue colors, respectively.

Method	# of params	NMSE($\times 10^{-2}$)(\downarrow)	PSNR(\uparrow)	SSIM(\uparrow)
E2E-Varnet [16]	30M	0.8690 \pm 0.9279	37.30 \pm 4.925	0.8936 \pm 0.1157
HUMUS-Net [2]	109M	0.8974 \pm 0.9743	37.20 \pm 5.009	0.8946 \pm 0.1162
HUMUS-Net-L [2]	228M	0.8587 \pm 0.9930	37.45 \pm 5.067	0.8955 \pm 0.1161
Baseline (ours)	47.5M	0.8321 \pm 0.9258	37.57 \pm 5.143	0.8964 \pm 0.1162
PromptMR (ours)	79.6M	0.8344 \pm 0.9648	37.63 \pm 5.319	0.8970 \pm 0.1168

Effectiveness of Two-Stage Pipeline We employed ShiftNet to refine the images reconstructed by the pretrained E2E-Varnet on the fastMRI multi-coil knee test dataset with $\times 8$ undersampling. Table B.2 shows that the second-stage refinement substantially improves the reconstruction quality, which implies that the multi-slice information in the fastMRI dataset might not be comprehensively utilized by the single-stage unrolled model.

Table B.2: Effectiveness of the second-stage image refinement on the fastMRI knee multi-coil $\times 8$ test dataset.

Stage	Method	# of params	NMSE($\times 10^{-2}$)(\downarrow)	PSNR(\uparrow)	SSIM(\uparrow)
I	E2E-Varnet [16]	30M	0.8690 \pm 0.9279	37.30 \pm 4.925	0.8936 \pm 0.1157
II	ShiftNet [8]	2M	0.8415 \pm 0.9131	37.46 \pm 4.973	0.8953 \pm 0.1157

¹ <https://github.com/facebookresearch/fastMRI/discussions/293>

C Additional Qualitative Results on CMRxRecon Dataset

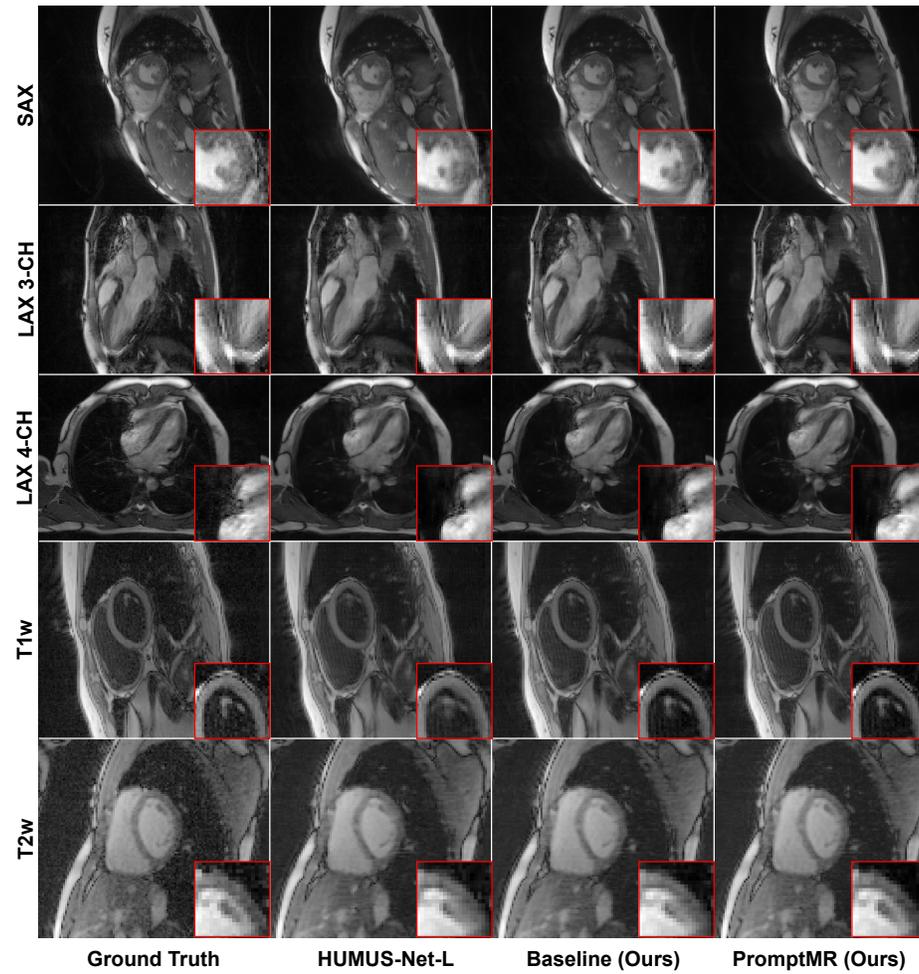


Fig. C.1: Visual comparison of reconstructions from the CMRxRecon dataset with $\times 10$ acceleration. PromptMR can recover fine details (highlighted in red box) on reconstructed images that other state-of-the-art methods may miss.