Breaking Information Isolation: Accelerating MRI via Inter-sequence Mapping and Progressive Masking

Jianwei Zheng, Xiaomin Yao, Guojiang Shen, Wei Li, and Jiawei Jiang*

Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China
jjw@zjut.edu.cn, zjw@zjut.edu.cn

Abstract

Deep unfolding network (DUN) has shed new light on multi-sequence MRI reconstruction, providing both high interpretability and acceptable performance. However, current approaches still suffer from the plight of information isolation, *i.e.,* learning features of multi-suquences individually and leaving the mask departed from model updating. In this work, we propose a new unfolding solution, namely Information-coupled MRI Acceleration (IMA), to address the isolation issue. Concretely, two specific mechanisms are presented. On the one hand, the latent connections across different sequences are explicitly molded via two auxiliary matrices. While the first matrix is meticulously engineered to assemble the spatial details, the second one hammers at capturing the depth information conditioned on the enriched channels. On the other hand, following a deep analysis on the non-uniform distribution in low- and high-frequency components of the given mask, we elaborate a new unfolding flow using a progressive masking scheme, featuring a dilation-contraction mechanism during forward propagation of successive stages. Massive experiments are conducted under various sampling patterns and acceleration rates, whose results demonstrate that, without any sophisticated architectures, our IMA outperforms the current cutting-edge methods both visually and numerically. Codes are available as an attachment and will be publicly released.

Introduction

Magnetic Resonance Imaging (MRI) is a key non-invasive and non-ionizing tool in medical diagnostics, offering superior soft tissue contrast and detailed anatomical visualization without radiation risks (Lei et al. 2023a; Wei et al. 2023). However, the issue of prolonged acquisition period, inherent in fully sampling the k-space, is pronounced in multi-sequence scans with long Echo Time (TE) and Repetition Time (TR). Practically, this issue often causes patient discomfort, motion-induced image distortion, and care delays, necessitating research into faster acquisition while maintaining high image quality.

In recent years, the acceleration of MRI reconstruction from single-sequence image has been well-established (Jiang et al. 2023b; Zhang et al. 2022; Yang et al. 2018), yet it overlooks the rich information inherent in diverse sequences. As

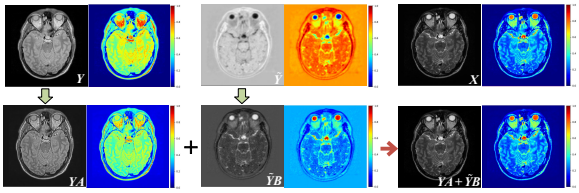

Figure 1: Explicit mapping between reference channel $\boldsymbol{Y}$ and target channel $\boldsymbol{X}$, *viz.*, $\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{A} + \tilde{\boldsymbol{Y}}\boldsymbol{B}$. $\boldsymbol{A}$ and $\boldsymbol{B}$ denote two transformation operators, respectively, while $\tilde{\boldsymbol{Y}}$ is a self-learned substitution used to bridge the sequence differences.

is known, during multi-sequence imaging, each entity characterizes distinct physical properties of the equal anatomy structure, making different sequences highly correlated. Clinically, radiologists also lean towards acquiring multiple in vivo images (*e.g.,* T1, PD, T2) for a comprehensive patient evaluation (Feng et al. 2023). Moreover, the large period gaps in acquiring different sequences presents an opportunity to speed up the reconstruction. That is, employing fully-sampled yet shorter sequences to aid in reconstructing longer sequences offers considerable potential to improve diagnostic efficiency and accuracy (Zhou and Zhou 2020).

In the realm of multi-sequence MRI reconstruction, most hand-crafted priors have been conventionally employed to inject visually authentic properties to the recovered images. The priors concerned, including sparse representation (Wang et al. 2023a), low rankness (Peng et al. 2022), edge sharpness (Huang et al. 2022), and non-local similarity (Wen et al. 2022), have shown commendable efficacy with solid theretical support. However, the applicability to practices of these methods is still hindered by the lengthy running time, a tendency towards over-smoothness, and the tricky selection of hyperparameters. Alternatively, the surge of deep neural networks (DNNs) has ushered in a new era of MRI techniques. Early DNNs often adapted architectures from general image processing, such as Convolution Neural Network (CNN) (Yang et al. 2023; Xu et al. 2021) and Transformer (Zhou et al. 2023; Sun, Wang, and Shen 2024; Fang et al. 2022; Sun et al. 2024; Sun 2024), to this field. Yet, most treatments are dedicated to designing a standalone network that is independent of intrinsic MRI characteristics, which often leads to sub-optimal performance. Additionally, deep learning methods

* Corresponding author.

are known with ill interpretability, being perceived as "black boxes". A pivotal advancement in alleviating these issues is the advent of Deep Unfolding Networks (DUNs) (Jiang et al. 2024; Xie et al. 2022; Song, Chen, and Zhang 2023; Jiang et al. 2023a), whose core idea lies in the joint consideration of both traditional optimizers and neural networks. On that basis, the ultimates not only achieve high performance, but also enjoy a better interpretability. Emerging practices (Zhou et al. 2024; Song et al. 2019; Yang et al. 2023) involve using quickly acquired sequences like T1WI and PDWI to facilitate the reconstruction of longer sequences such as T2WI and FSPDWI. For example, Ref. (Yang et al. 2020) capitalizes on the cross-sequence prior and combines channel-spatial attention to refine the proximal operators. MC-VarNet (Lei et al. 2023a) further introduces convolutional sparse coding technique to this field, striving for better accuracy with the sacrifice of running efficiency. More recently, Ref. (Chen et al. 2024) proposes a range-null decomposition-based unfolding method (RNU) to individually assimilate shared and unique information across sequences, thereby promoting the consistency of reconstructions.

Yet with high interpretability, current DUNs fall short in dealing with the issue of information isolation. Most standard treatments individually perform feature extraction before capturing interactions between sequences (Chen et al. 2024; Wang et al. 2023b). Some others simply concatenate the inputs along the channel direction and then feed them into certain brute-force networks (Pooja et al. 2022; Guo et al. 2023). Several efforts have recently emerged to investigate the establishment of inter-sequence relationships. For example, Ref. (Lei et al. 2023a) implicitly connects multiple inputs by constraining some shared components. Ref. (Yang et al. 2023) employs an indirect transformation matrix for the collection of interactions. However, either the implicit or indirect solution is suboptimal for the final reconstruction. Moreover, leaving the interaction away from the loss setting would also cause ambiguity during model updating (Chen et al. 2024; Yang et al. 2023). In addition, existing research typically focuses on fixed masks to ensure data fidelity in known areas. In other words, a standalone model update is usually performed, which suffers a poor convergence due to the non-uniform distribution of different in-mask regions.

To mitigate the aforementioned issues, in this study, we propose an information-coupled MRI acceleration model. The principal motivation lies in breaking the information isolation that exists in current DUNs. On the one hand, two treatments are endeavored to engineer the explicit inter-sequnece connections. First, a direct mapping from embedding reference $\boldsymbol{Y}$ to target $\boldsymbol{X}$ is established. As shown in Fig. 1, we decompose $\boldsymbol{X}$ into a combination of spatial-spectral complementary approximation $\boldsymbol{YA}$ and $\tilde{\boldsymbol{Y}}\boldsymbol{B}$. While the latter bridges protocol gaps with the aid of a feature-enriched $\tilde{\boldsymbol{Y}}$, these two jointly enrich the texture details of target $\boldsymbol{X}$. Second, a cross-sequence converging attention (CSCA) is crafted to jointly consider the cross- and self- attentions. On the other hand, we propose a dilation-contraction masking (DCM) mechanism to facilitate a smooth unfolding flow. Specifically, a self-updated mask is embedded into the forward process, enabling the network to learn progressively in regions from known to unknown ones. The overall framework of the proposed IMA is given in Fig. 2.

Methodology

Problem Analysis

Let $\bar{\boldsymbol{X}} \in \mathbb{R}^{HW\times 1}$ and $\bar{\boldsymbol{Y}} \in \mathbb{R}^{HW\times 1}$ denote two fully-sampled MR sequences belonging to different contrasts, *e.g.,* T2 and PD. Accordingly, the undersampled k-space observation of $\bar{\boldsymbol{X}}$ is represented by $\bar{\boldsymbol{Z}} \in \mathbb{C}^{HW\times 1}$, with the masked elements being 0. In the context of multi-sequences MRI reconstruction, the rooted goal hammers at recovering $\bar{\boldsymbol{X}}$ from $\bar{\boldsymbol{Z}}$ with the assistance of reference $\bar{\boldsymbol{Y}}$. To achieve this, we consider $\boldsymbol{X}$ and $\boldsymbol{Z}$ as the enriched embeddings of $\bar{\boldsymbol{X}}$ and $\bar{\boldsymbol{Z}}$ in a C-dimensional space, i.e., $\boldsymbol{X} \in \mathbb{R}^{HW\times C}$ and $\boldsymbol{Z} \in \mathbb{C}^{HW\times C}$. Additionally, we define $\boldsymbol{Y}$ as the embedding of $\bar{\boldsymbol{Y}}$ in a c-dimensional subspace with $c < C$, i.e., $\boldsymbol{Y} \in \mathbb{R}^{HW\times c}$. The above transformations can be achieved through learnable networks. Typically, we assume that $\boldsymbol{X}$ and $\boldsymbol{Z}$ adhere to the original MRI acceleration model physically, which can be represented by the following formulation:

$$\boldsymbol{Z} = \boldsymbol{M} \odot \mathcal{F}(\boldsymbol{X}) + \boldsymbol{N}_z, \tag{1}$$

where $\mathcal{F}$ denotes 2D Fourier transform, $\boldsymbol{M}$ is a binary undersampling mask of size $H \times W$, $\odot$ denotes Hadamard product, $\boldsymbol{N}_z$ represents the noise in $\boldsymbol{Z}$. The explicit relationship between heterogeneous sequences $\boldsymbol{Y}$ and $\boldsymbol{X}$ has been rarely discussed. However, we consider this relationship is of significant importance for understanding the reconstruction process, since $\boldsymbol{Y}$ possesses the entire sampling information. Note that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are acquired from the same targets with different protocols. Therefore, we further explicitly represent the relationship between $\boldsymbol{Y}$ and $\boldsymbol{X}$ as follows:

$$\boldsymbol{Y} = \boldsymbol{XR} + \boldsymbol{N}_y, \tag{2}$$

where $\boldsymbol{R} \in \mathbb{R}^{C\times c}$ represents a transformation, and $\boldsymbol{N}_y$ denotes the noise present in the auxiliary sequence $\boldsymbol{Y}$. Subsequently, based on Theorem 1, we present an equivalent representation of Eq. (2).

Theorem 1. *For any $\boldsymbol{X} \in \mathbb{R}^{HW\times C}$ and $\hat{\boldsymbol{Y}} \in \mathbb{R}^{HW\times c}$, if rank$(\boldsymbol{X}) = r > c$ and rank$(\hat{\boldsymbol{Y}}) = c$, then the following statements are equivalent:*

(a) There exists $\boldsymbol{R} \in \mathbb{R}^{C\times c}$, such that

$$\hat{\boldsymbol{Y}} = \boldsymbol{XR}. \tag{3}$$

(b) There exist $\boldsymbol{A} \in \mathbb{R}^{c\times C}$, $\boldsymbol{B} \in \mathbb{R}^{(r-c)\times C}$, and $\tilde{\boldsymbol{Y}} \in \mathbb{R}^{HW\times(r-c)}$, such that

$$\boldsymbol{X} = \hat{\boldsymbol{Y}}\boldsymbol{A} + \tilde{\boldsymbol{Y}}\boldsymbol{B}. \tag{4}$$

The proof is presented in Appendix. Note that $\hat{\boldsymbol{Y}}$ is considered to be the noisyless version of $\boldsymbol{Y}$, which is the transformed embedding discussed in *Problems Analysis*. In practice, the rank-1 input can be enriched via certain operations such as spectral abundant convolution kerels and the concerned nonlinear activations, achieving a rank$(\hat{\boldsymbol{Y}}) = c$ counterpart. On that basis, it could be observed that $\hat{\boldsymbol{Y}}$ and $\boldsymbol{X}$ satisfy the conditions in Theorem 1. By ensuring $\hat{\boldsymbol{Y}} = \boldsymbol{Y} - \boldsymbol{N}_y$,

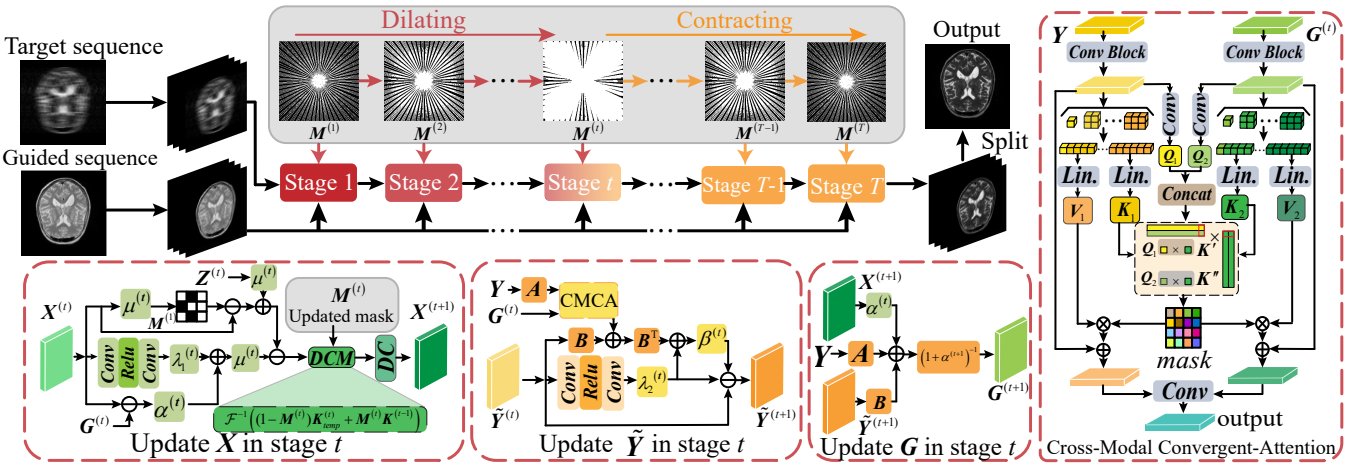

Figure 2: The illustration of our proposed IMA. The mask is progressively updated during different unfolding stages, which favors feature learning in high-frequency areas.

model (2) can be reformulated as follows:

$$\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{A} + \tilde{\boldsymbol{Y}}\boldsymbol{B} + \boldsymbol{N}_x, \tag{5}$$

where $\boldsymbol{N}_x = -\boldsymbol{N}_y\boldsymbol{A}$ denotes the noise present in $\boldsymbol{X}$. Note that in (5), $\tilde{\boldsymbol{Y}}$ is also unknown and requires to be estimated. By combining Eq. (1) and Eq. (5), we can intuitively design the following optimization problem for multi-sequence MRI reconstruction:

$$\min_{\boldsymbol{X},\tilde{\boldsymbol{Y}}} \frac{1}{2}\|\boldsymbol{M}\odot\mathcal{F}(\boldsymbol{X}) - \boldsymbol{Z}\|_F^2 + \frac{1}{2}\|(\boldsymbol{Y}\boldsymbol{A} + \tilde{\boldsymbol{Y}}\boldsymbol{B}) - \boldsymbol{X}\|_F^2 + \lambda_1\mathcal{R}_1(\boldsymbol{X}) + \lambda_2\mathcal{R}_2(\tilde{\boldsymbol{Y}}), \tag{6}$$

where λ_1 and λ_2 are trade-off parameters, $\mathcal{R}_1(\cdot)$ and $\mathcal{R}_2(\cdot)$ are prior constraints. Contrary to traditional hand-crafted ways, (Yang et al. 2022) shows data-driven priors are more effective. So, our method doesn't specify $\mathcal{R}_1$ and $\mathcal{R}_2$ explicitly, but designs an unfolding network to implicitly learn the intrinsic features of MRI.

Model Optimization

Most first-order optimization algorithms are readily available to solve model (6). We simply employ the typical half quadratic splitting scheme (Jiang et al. 2023a), yet put more efforts on network architecture during unfolding stages. With an auxiliary variable $\boldsymbol{G}$ introduced, the optimization can be reformulated as follows.

$$\min_{\boldsymbol{X},\tilde{\boldsymbol{Y}},\boldsymbol{G}} \frac{1}{2}\|\boldsymbol{M}\odot\mathcal{F}(\boldsymbol{X}) - \boldsymbol{Z}\|_F^2 + \frac{1}{2}\|(\boldsymbol{Y}\boldsymbol{A} + \tilde{\boldsymbol{Y}}\boldsymbol{B}) - \boldsymbol{G}\|_F^2 + \frac{\alpha}{2}\|\boldsymbol{X} - \boldsymbol{G}\|_F^2 + \lambda_1\mathcal{R}_1(\boldsymbol{X}) + \lambda_2\mathcal{R}_2(\tilde{\boldsymbol{Y}}), \tag{7}$$

where α is also a trade-off parameter. By sequentially solving one targe variable with the others fixed, Eq. (7) can be practiced via three subproblems.

$$\boldsymbol{X}^{(t+1)} = \arg\min_{\boldsymbol{X}} \frac{1}{2}\|\boldsymbol{M}\odot\mathcal{F}(\boldsymbol{X}) - \boldsymbol{Z}\|_F^2 + \frac{\alpha^{(t)}}{2}\|\boldsymbol{X} - \boldsymbol{G}^{(t)}\|_F^2 + \lambda_1^{(t)}\mathcal{R}_1(\boldsymbol{X}), \tag{8}$$

$$\tilde{\boldsymbol{Y}}^{(t+1)} = \arg\min_{\tilde{\boldsymbol{Y}}} \frac{1}{2}\|(\boldsymbol{Y}\boldsymbol{A} + \tilde{\boldsymbol{Y}}\boldsymbol{B}) - \boldsymbol{G}^{(t)}\|_F^2 + \lambda_2^{(t)}\mathcal{R}_2(\tilde{\boldsymbol{Y}}), \tag{9}$$

$$\boldsymbol{G}^{(t+1)} = \arg\min_{\boldsymbol{G}} \frac{1}{2}\|(\boldsymbol{Y}\boldsymbol{A} + \tilde{\boldsymbol{Y}}^{(t+1)}\boldsymbol{B}) - \boldsymbol{G}\|_F^2 + \frac{\alpha^{(t)}}{2}\|\boldsymbol{X}^{(t+1)} - \boldsymbol{G}\|_F^2, \tag{10}$$

where t denotes the iteration step.

1) ***Update of*** $\boldsymbol{X}$. By introducing gradient descent algorithm (GDA) (Haji and Abdulazeez 2021), image $\boldsymbol{X}^{(t+1)}$ can be learned as:

$$\boldsymbol{X}^{(t+1)} = \boldsymbol{X}^{(t)} - \mu^{(t)}\boldsymbol{U}^*\left(\boldsymbol{U}\boldsymbol{X}^{(t)} - \boldsymbol{Z}\right) - \mu^{(t)}\left(\alpha^{(t)}(\boldsymbol{X}^{(t)} - \boldsymbol{G}^{(t)}) + \lambda_1^{(t)}\nabla\mathcal{R}_1(\boldsymbol{X}^{(t)})\right), \tag{11}$$

where $\mu^{(t)}$ is the learning rate controlling the step size, $\boldsymbol{U} = \boldsymbol{M}\odot\mathcal{F}(\cdot)$ is the linear forward operator that applies 2D Fourier Transform and then undersamples data using mask $\boldsymbol{M}$, $\boldsymbol{U}^*$ is the adjoint (backward) operator of $\boldsymbol{U}$. Note that $\mathcal{R}_1(\boldsymbol{X})$ is traditionally indifferentiable given an explicit prior. This issue is naturally avoided within the scope of adaptive prior learning using networks, which essentially empower a backpropagation process along with a gradient flow.

2) ***Update of*** $\tilde{\boldsymbol{Y}}$. Similarly, with the aid of GDA, the updated formulation for $\tilde{\boldsymbol{Y}}$ can be derived as:

$$\tilde{\boldsymbol{Y}}^{(t+1)} = \tilde{\boldsymbol{Y}}^{(t)} - \beta^{(t)}\left(\boldsymbol{Y}\boldsymbol{A} - \boldsymbol{G}^{(t)} + \tilde{\boldsymbol{Y}}^{(t)}\boldsymbol{B}\right)\boldsymbol{B}^{\mathrm{T}} - \beta^{(t)}\lambda_2^{(t)}\nabla\mathcal{R}_2(\tilde{\boldsymbol{Y}}^{(t)}), \tag{12}$$

where $\beta^{(t)}$ is the learning rate, $\nabla\mathcal{R}_2(\tilde{\boldsymbol{Y}})$ is the gradient of $\mathcal{R}_2$ with respect to $\tilde{\boldsymbol{Y}}$.

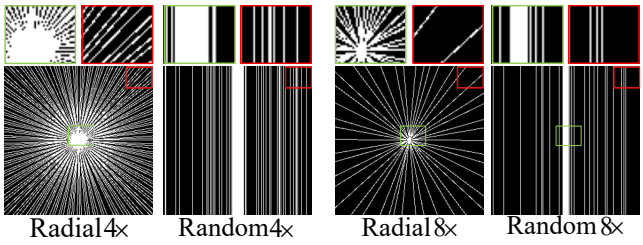

Figure 3: The distribution of low- and high-frequency information in masks with various acceleration ratios, where white and black represent known and unknown areas, respectively.

3) ***Update of*** $\boldsymbol{G}$. Due to the differentiability of Frobenius norm, the solution of subproblem (13) can be analytically derived as:

$$\boldsymbol{G}^{(t+1)} = \left(1+\alpha^{(t)}\right)^{-1}\left(\boldsymbol{Y}\boldsymbol{A} + \tilde{\boldsymbol{Y}}^{(t+1)}\boldsymbol{B} + \alpha^{(t)}\boldsymbol{X}^{(t+1)}\right). \tag{14}$$

With Eqs. (12)-(14) evolved into a data-driven refinement, an unfolding network would naturally be established. In the sequel, we will present our elaboration to fit MRI acceleration.

Beyond Unfolding: Network Design

Along with the explicit mapping given in Eq. (5), we further propose two mechanisms to address the issue of information isolation. Across distinct stages, a dilation-contraction masking (DCM) is crafted to favor a smooth learning flow, which employs a gradual diffusion from known to unknown areas. During each unfolding stage, a cross-sequence converging attention (CSCA) is engineered to facilitate both feature interaction and information aggregation from two inputs. In the following, we shall expound in detail the practical unfolding steps.

1) **The unfolding of** $\boldsymbol{X}$: The first two terms of Eq. (12) is regarded as the basic data fidelity in k-space, as described in (Lei et al. 2023b). The remaining term can then be seen as an operation to refine details in the image domain. Note that rather than approximating $\mathcal{R}_1(\boldsymbol{X}^{(t)})$ first and then computing the concerned gradients, we directly learn $\nabla\mathcal{R}_1(\boldsymbol{X}^{(t)})$ with a convolution block $\mathcal{C}^{(t)}_{\theta_{\mathcal{R}1}}$. By introducing the operator $\mathcal{F}^{-1}\mathcal{F}$ to the first two terms, Eq. (12) can be temporarily reformulated as:

$$\boldsymbol{X}^{(t)}_{temp} = \mathcal{F}^{-1}\underbrace{\left(\boldsymbol{X}^{(t)}_k - \mu^{(t)}(\boldsymbol{M}\boldsymbol{X}^{(t)}_k - \boldsymbol{Z}_k)\right)}_{\text{data consistency}}$$
$$-\mu^{(t)}\left(\alpha^{(t)}(\boldsymbol{X}^{(t)} - \boldsymbol{G}^{(t)}) + \lambda^{(t)}_1\mathcal{C}^{(t)}_{\theta_{\mathcal{R}1}}(\boldsymbol{X}^{(t)})\right), \tag{15}$$

where $\boldsymbol{X}^{(t)}_k$ and $\boldsymbol{Z}_k$ are short for $\mathcal{F}\boldsymbol{X}^{(t)}$ and $\mathcal{F}\boldsymbol{Z}^{(t)}$, respectively, $\mathcal{F}^{-1}$ denotes the inverse Fourier transform, $\mathcal{C}^{(t)}_{\theta_{\mathcal{R}1}}$ follows a simple architecture of [3×3 *Conv, ReLU,* 3×3 *Conv*]. Within Eq. (15), the used mask $\boldsymbol{M}$ is originally fixed and given in advance. Fig. 3 illustrates two prevailing masks with acceleration rates $4\times$ and $8\times$. As is evidenced, while most

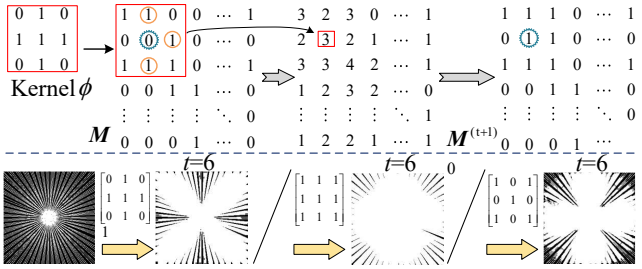

Figure 4: Mask updating against different kernels.

of the low-frequency regions enjoy the known pixels, merely sparse areas in the high-frequency portion hold known information. The divisive practice aligns closely with the physical reality since most of the energy lies in the low-frequency components, yet it poses challenges in learning features for high-frequency areas. To alleviate this issue and further refine the intermediate variable $\boldsymbol{X}^{(t)}_{temp}$, we schedule a progressive learning mechanism termed dilation-contraction masking.

DCM mechanism. The schematic of DCM is given in the upper part of Fig. 2. The detailed masking flow involves two phases holding an inversely symmetric structure, *i.e.,* dilation and contraction. During the dilation phase, the initially known pixels gradually spread to the surroundings, favoring easier learning of higher-frequency features. The contraction process inversely restores the mask to the initial state, providing an opportunity for a twice-learning paradigm. Mathematically, the dilation can be expressed as follows.

$$\boldsymbol{M}^{(t+1)} = \begin{cases} 1, & \phi(\boldsymbol{M}^{(t)}) \geq \delta \text{ or } \boldsymbol{M}^{(0)} \geq 1 \\ 0, & \text{else} \end{cases}, \tag{16}$$

where ϕ is a predefined diffusion kernel, δ serves as a confidence threshold. The upper part of Fig. 4 provides a numerical illustration of the dilation process, while the lower part shows partial results from different kernel selections. As shown, the practical ϕ values would clearly affect the diffusion shapes and masking speed. Due to the symmetry of dilation and contraction, the discussion concerning the contraction process is omitted for simplicity. The condition to terminate the diffusion process of $\boldsymbol{M}^{(t+1)}$ is defined as

$$\text{Termination if } \left(\frac{n_1}{N} > ratio\right) \text{ or } \left(t > \frac{T}{2}\right), \tag{17}$$

where N represents the total pixels, T is the maximum iteration, n_1 denotes the count of value-1 elements, $ratio \in [0,1]$ is a hyperparameter denoting the proportion occupied by known regions.

Given the updated mask, $\boldsymbol{X}^{(t)}_{temp}$ then undergoes a renewed update.

$$\hat{\boldsymbol{X}}^{(t)} = \mathcal{F}^{-1}\left((1-\boldsymbol{M}^{(t)})\boldsymbol{K}^{(t)}_{temp} + \boldsymbol{M}^{(t)}\boldsymbol{K}^{(t-1)}\right), \tag{18}$$

$$\boldsymbol{X}^{(t+1)} = \mathcal{F}^{-1}\left((1-\boldsymbol{M}^{(0)})\hat{\boldsymbol{K}}^{(t)} + \boldsymbol{M}^{(0)}\boldsymbol{K}^{(0)}\right), \tag{19}$$

where $\boldsymbol{K}^{(t)}_{temp} = \mathcal{F}\boldsymbol{X}^{(t)}_{temp}$, $\boldsymbol{K}^{(t-1)} = \mathcal{F}\boldsymbol{X}^{(t-1)}$, $\hat{\boldsymbol{K}}^{(t)} = \mathcal{F}\hat{\boldsymbol{X}}^{(t)}$, and $\boldsymbol{K}^{(0)} = \mathcal{F}\boldsymbol{Z}$. As previously mentioned, Eq.

Methods	IXI (4×)						IXI (8×)					
	Random			Radial			Random			Radial		
	PSNR ↑	SSIM ↑	*P* values	PSNR ↑	SSIM ↑	*P* values	PSNR ↑	SSIM ↑	*P* values	PSNR ↑	SSIM ↑	*P* values
Unet_{15}	30.636▼30%	0.947▼5%	$< 0.001/ < 0.001$	33.634▼34%	0.963▼4%	$< 0.001/ < 0.001$	28.662▼31%	0.927▼7%	$< 0.001/ < 0.001$	29.379▼30%	0.929▼6%	$< 0.001/ < 0.001$
MICCAN_{19}	35.187▼20%	0.972▼2%	$< 0.001/ < 0.001$	42.776▼16%	0.993▼1%	$< 0.001/ < 0.001$	31.554▼24%	0.949▼4%	$< 0.001/ < 0.001$	33.919▼19%	0.967▼3%	$< 0.001/ < 0.001$
DuDoRNet_{20}	38.967▼11%	0.984▼1%	$< 0.001/ < 0.001$	44.855▼12%	0.994▼1%	$< 0.001/ < 0.001$	36.193▼12%	0.974▼2%	$< 0.001/ < 0.001$	38.272▼9%	0.982▼1%	$< 0.001/ < 0.001$
DuDoCAF_{22}	39.037▼11%	0.985▼1%	$< 0.001/ < 0.001$	45.001▼11%	0.994▼1%	$< 0.001/ < 0.001$	36.319▼12%	0.976▼2%	$< 0.001/ < 0.001$	37.574▼10%	0.977▼2%	$< 0.001/ < 0.001$
MC-VarNet_{23}	41.532▼6%	0.992▼0.3%	$< 0.001/ < 0.001$	47.471▼7%	0.998▼0.1%	$< 0.001/ < 0.001$	38.704▼6%	0.988▼0.4%	$= 0.001/ < 0.001$	40.354▼4%	0.990▼0.2%	$< 0.001/ < 0.001$
RNU_{24}	42.323▼4%	0.992▼0.3%	$< 0.001/ < 0.001$	48.374▼5%	0.998▼0.1%	$< 0.001/ < 0.001$	39.894▼4%	0.990▼0.2%	$< 0.001/ < 0.001$	40.901▼2%	0.991▼0.1%	$< 0.001/ < 0.001$
Ours	43.984	0.995	–	50.810	0.999	–	41.291	0.992	–	41.948	0.992	–
Methods	fastMRI (4×)						fastMRI (8×)					
	Random			Radial			Random			Radial		
	PSNR ↑	SSIM ↑	*P* values	PSNR ↑	SSIM ↑	*P* values	PSNR ↑	SSIM ↑	*P* values	PSNR ↑	SSIM ↑	*P* values
Unet_{15}	27.792▼8%	0.808▼7%	$< 0.001/ < 0.001$	28.619▼9%	0.829▼7%	$< 0.001/ < 0.001$	26.330▼8%	0.752▼8%	$< 0.001/ < 0.001$	26.256▼8%	0.720▼8%	$< 0.001/ < 0.001$
MICCAN_{19}	29.112▼4%	0.835▼4%	$< 0.001/ < 0.001$	30.186▼4%	0.865▼3%	$< 0.001/ < 0.001$	27.377▼5%	0.769▼5%	$< 0.001/ < 0.001$	27.297▼4%	0.737▼6%	$< 0.001/ < 0.001$
DuDoRNet_{20}	29.034▼4%	0.834▼4%	$< 0.001/ < 0.001$	30.004▼5%	0.861▼4%	$< 0.001/ < 0.001$	27.367▼5%	0.774▼5%	$< 0.001/ < 0.001$	27.137▼5%	0.742▼5%	$< 0.001/ < 0.001$
DuDoCAF_{22}	29.163▼3%	0.837▼4%	$< 0.001/ < 0.001$	30.244▼4%	0.866▼3%	$< 0.001/ < 0.001$	27.557▼4%	0.776▼5%	$< 0.001/ < 0.001$	27.515▼3%	0.751▼4%	$= 0.001/ < 0.001$
MC-VarNet_{23}	29.498▼2%	0.845▼3%	$< 0.001/ < 0.001$	30.404▼4%	0.871▼3%	$< 0.001/ < 0.001$	28.089▼2%	0.790▼3%	$< 0.001/ < 0.001$	27.819▼2%	0.762▼3%	$< 0.001/ < 0.001$
RNU_{24}	29.949▼1%	0.851▼2%	$= 0.001/ < 0.001$	31.228▼1%	0.878▼2%	$< 0.001/ < 0.001$	28.379▼1%	0.795▼2%	$= 0.001/ < 0.001$	28.047▼2%	0.766▼2%	$< 0.001/ < 0.001$
Ours	30.125	0.871	–	31.518	0.894	–	28.764	0.813	–	28.493	0.783	–

Table 1: The numerical results from two masks and with acceleration rates of 4× and 8×. ▼ represents the dropping percentage compared to ours. $P < 0.001$ denotes a statistically significant level.

(18) ensures that the known regions remain unchanged while updating the unknown ones according to the updated mask. Then, Eq. (19) guarantees that the initial fidelity of the data is maintained once again.

2) **The unfolding of $\tilde{\boldsymbol{Y}}$ and $\boldsymbol{G}$**: Note again $\boldsymbol{G}^{(t)}$ is an auxiliary variable that approximates the target sequence $\boldsymbol{X}$. On this basis, the term $\boldsymbol{Y}\boldsymbol{A} - \boldsymbol{G}^{(t)}$ in Eq. (13) plays the core role in inter-sequence interaction and aggregation. However, the computation in current form simply employs pixel-level operations, which are inadequate for a holistic feature learning. Instead, we propose a lightweight CSCA as a substitute. Besides, similar to the treatment for Eq. (15), the unknown terms $\boldsymbol{B}^{\mathrm{T}}$ and $\nabla\mathcal{R}_2(\cdot)$ are also replaced with convolution blocks. The reformulated Eq. (13) is then written as:

$$\tilde{\boldsymbol{Y}}^{(t+1)} = \tilde{\boldsymbol{Y}}^{(t)} - \beta^{(t)}\mathcal{B}^{(t)}_{\theta_{B^{\mathrm{T}}}}\left(CSCA(\boldsymbol{Y}, \boldsymbol{G}^{(t)}) + \tilde{\boldsymbol{Y}}^{(t)}\right) - \beta^{(t)}\lambda_2^{(t)}\mathcal{C}^{(t)}_{\theta_{\mathcal{R}2}}(\tilde{\boldsymbol{Y}}^{(t)}), \tag{20}$$

where $\mathcal{B}^{(t)}_{\theta_{B^{\mathrm{T}}}}$ and $\mathcal{C}^{(t)}_{\theta_{\mathcal{R}2}}$ represent the network of $\boldsymbol{B}^{\mathrm{T}}$ and $\nabla\mathcal{R}_2(\cdot)$, respectively. As discussed, the principal role of CSCA lies in deep-level feature interaction and aggregation, whose basic routine follows the typical self- or cross-attention (Sun, Wang, and Shen 2024). As illustrated in Fig. 2 (right), CSCA commences with multiscale feature extraction from $\boldsymbol{Y}$ and $\boldsymbol{G}^{(t)}$, generating ($\boldsymbol{Q}_1$, $\boldsymbol{K}_1$, $\boldsymbol{V}_1$) and ($\boldsymbol{Q}_2$, $\boldsymbol{K}_2$, $\boldsymbol{V}_2$), respectively. Subsequently, with the aim of achieving both lightweight module and long-range information, we concatenate $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ and then individually perform interaction with $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$, allowing self- and cross-attention in a single computation, as shown in Eq. (21).

$$\underbrace{\boldsymbol{Q}^{\mathrm{T}}_{\mathrm{cat}} \cdot \boldsymbol{K}_i}_{\text{Assume } i=1} = \underbrace{\boldsymbol{Q}^{\mathrm{T}}_1 \cdot \boldsymbol{K}'_i}_{\text{Self-attention}} + \underbrace{\boldsymbol{Q}^{\mathrm{T}}_2 \cdot \boldsymbol{K}''_i}_{\text{Cross-attention}}, \tag{21}$$

Here $i = \{1, 2\}$, $\boldsymbol{Q}_{\mathrm{cat}}$ is the concatenated result, $\boldsymbol{K}'_i$ and $\boldsymbol{K}''_i$ denote two individual halves of either $\boldsymbol{K}_1$ or $\boldsymbol{K}_2$. The remaining computations follow the traditions and are omitted here to save space.

The solution of $\boldsymbol{G}$ is relatively simple due to the closed-form of Eq. (14). The only concern during unfolding is to replace unknown operators with learnable modules.

$$\boldsymbol{G}^{(t+1)} = \left(1+\alpha^{(t)}\right)^{-1}(\mathcal{A}^{(t)}_{\theta_A}(\boldsymbol{Y}) + \mathcal{B}^{(t)}_{\theta_B}(\tilde{\boldsymbol{Y}}^{(t+1)}) \tag{22}$$

$$+ \alpha^{(t)}\boldsymbol{X}^{(t+1)}), \tag{23}$$

where $\mathcal{A}^{(t)}_{\theta_A}$ and $\mathcal{B}^{(t)}_{\theta_B}$ represent the replacements of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. In our implementation, $\mathcal{B}^{(t)}_{\theta_{B^{\mathrm{T}}}}$, $\mathcal{C}^{(t)}_{\theta_{\mathcal{R}2}}$, $\mathcal{A}^{(t)}_{\theta_A}$ and $\mathcal{B}^{(t)}_{\theta_B}$ all simply employ the same convolution block as $\mathcal{C}^{(t)}_{\theta_{\mathcal{R}1}}$, *i.e.,* [$3 \times 3$ *Conv, ReLU,* 3×3 *Conv*]. This choice is motivated by two considerations: 1) Rather than meticulously designing complex architectures for deeper features, we focus on alleviating the issue of information isolation with three newly presented mechanisms, including explicit inter-sequence mapping, cross-sequence converging attention, and dilation-contraction masking. 2) The experimental results support that convolutions are sufficient to approximate the concerned operators. Moreover, the simple treatment avoid to burden the entire unfolding network with redundant parameter quantities and training/testing overhead.

3) **Network parameters and loss function**: Let Θ represent the set of learnable parameters in our method. Building on the earlier analysis, Θ holds five transformation operator $\mathcal{C}^{(t)}_{\theta_{\mathcal{R}1}}$, $\mathcal{C}^{(t)}_{\theta_{\mathcal{R}2}}$, $\mathcal{B}^{(t)}_{\theta_{B^{\mathrm{T}}}}$, $\mathcal{A}^{(t)}_{\theta_A}$ and $\mathcal{B}^{(t)}_{\theta_B}$, one attention module $CSCA(\cdot)$, and five trade-off parameters $\beta^{(t)}$, $\alpha^{(t)}$, $\mu^{(t)}$, $\lambda_1^{(t)}$, $\lambda_2^{(t)}$, which can be written as

$$\Theta = \{\mathcal{C}^{(t)}_{\theta_{\mathcal{R}1}}, \mathcal{C}^{(t)}_{\theta_{\mathcal{R}2}}, \mathcal{B}^{(t)}_{\theta_{B^{\mathrm{T}}}}, \mathcal{A}^{(t)}_{\theta_A}, \mathcal{B}^{(t)}_{\theta_B}, CSCA(\cdot), \beta^{(t)}, \alpha^{(t)}, \mu^{(t)}, \lambda_1^{(t)}, \lambda_2^{(t)}\}_{t=1}^{T}. \tag{24}$$

With all elements organically assembled, the overall loss is derived by measuring the differences as

$$\mathbb{L}_S(\Theta) = \mathbb{L}\left(\boldsymbol{X}_{out} - \boldsymbol{X}_{GT}\right) + \gamma\mathbb{L}\left(\boldsymbol{X}_{out} - (\boldsymbol{Y}\boldsymbol{A} + \tilde{\boldsymbol{Y}}\boldsymbol{B})\right), \tag{25}$$

where $\mathbb{L}$ denotes an l_1 loss, $\boldsymbol{X}_{out}$ is the final output of the forward network, and $\boldsymbol{X}_{GT}$ represents the ground truth. Note the

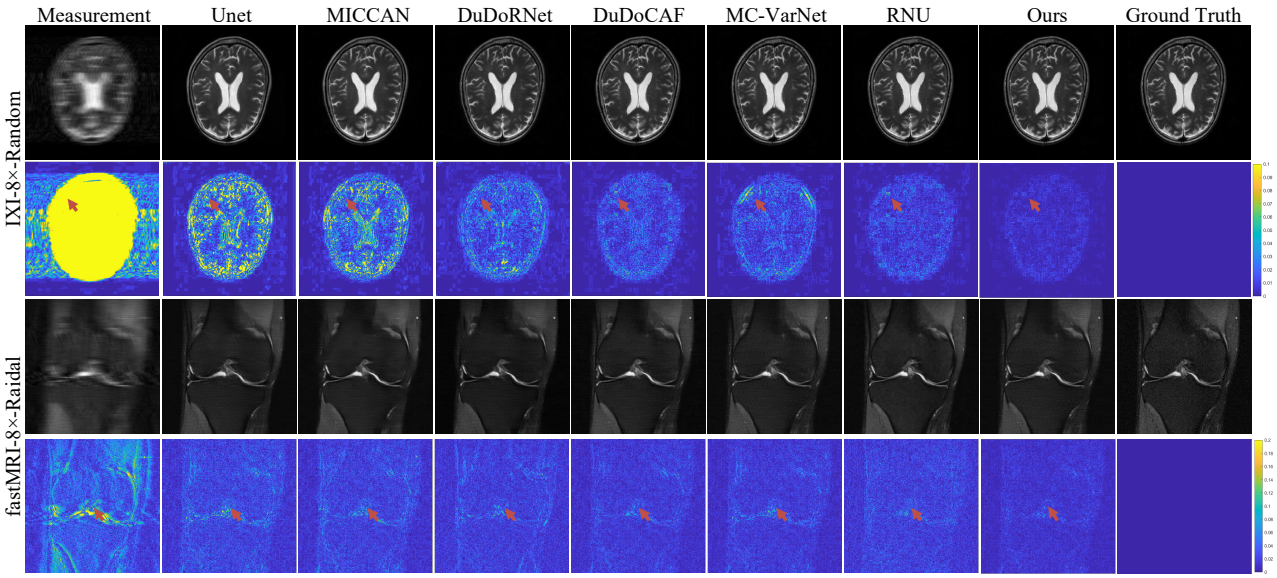

Figure 5: Visual results with different masks. The corresponding error maps are also given for immediate inspection.

second term of (25) imposes a proximity constraint to clearly enforce the explicit mapping, *i.e.,* Eq. (5), guaranteeing the model update to the expectations.

Experiments and Analysis

Datasets.We evaluate our proposal on two publicly available datasets: IXI (578 registered T2 and PD image pairs, 256×256) and knee fastMRI (240 PD image pairs, 320×320). The data is split into training, validation, and testing sets (7:1:2). Two undersampling masks (1D Cartesian and 2D radial) with acceleration rates of $4\times$ and $8\times$ are used, following (Chen et al. 2024).

Implementation Details: All experiments are conducted using PyTorch on an NVIDIA GeForce RTX 3090 GPU. For our IMA, we use l_1-norm loss, the Adam optimizer (0.9, 0.999), 100 epochs, an initial learning rate of 10^{-4}, batch size of 1, *ratio*=0.9, $\gamma = 0.5$, $\delta = 3$, and T=12. Other competitors' settings are taken from the original authors.

Metrics: We use Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) for quantitative evaluation, along with efficiency metrics like parameters (Params) and GMACs in the ablation studies.

Comparison to Cutting-edge Methods

Six state-of-the-art methods from previous reports are chosen as competitors, consisting of two DUNs including MC-VarNet (Lei et al. 2023b) and RNU (Chen et al. 2024), as well as four DNNs including Unet (Ronneberger, Fischer, and Brox 2015), MICCAN (Huang et al. 2019), DuDoRNet (Zhou and Zhou 2020), and DuDoCAF (Lyu et al. 2022).

Table 1 presents the oncerned results, together with the percentage improvements over the competing methods. In terms of IXI dataset, the results achieved from IMA demonstrate superior performance. Specifically, under an $8\times$ radial acceleration, IMA reaches a PSNR value of 41.948, outperforming other competitors by a large margin. Despite the challenging nature of fastMRI dataset, our proposal still ranks first across all cells. Among the compared methods, it is evident that RNU and MC-VarNet achieve relative better results, verifying the advantages of integrating traditional optimizers with deep network frameworks. In contrast, Unet consistently generates sub-par performance, possibly due to the ignorance of data consistency during network design. Owing to the implementation of deeper feature fusion, DuDoCAF slightly outperforms DuDoRNet. Furthermore, almost all p-values are less than 0.001, with some minority equal to 0.001 yet still significantly below the threshold of 0.05, demonstrating statistically the improvements of IMA over the others. Overall, these results validate the positive impact of key mechanisms adopted in our approach, especially the dilation-contraction masking and the explicit inter-sequence mapping.

Fig. 5 visualizes the reconstructions of two representative scenes on T2 and PD sequences of the IXI and fastMRI datasets, respectively. At first glance, all competing methods demonstrate satisfactory preservation of tissue structures, recovering the main features across various regions. Zooming in for a deeper view, we can observe that our results conform closer to the ground truth, which is clearly supported by the error maps. Much more residuals scatter in the error maps from Unet, MICCAN, DuDoRNet, and MC-VarNet. Comparatively, DuDoCAF and RNU generate much bluer contrasts. Our IMA yields smaller and more evenly distributed errors in the reconstruction results. Fig. 6 further visualizes the overall performance in terms of PSNR, GMACs, and Params. Evidently, DuDoCAF and DuDoRNet enjoy the most lighweigt architecture, yet they suffer from many more GMACs. On the contrary, Unet and MICCAN share the least computational cost, yet the model parameters are much heavier. MC-VarNet, RNU, and our proposal hold a well balance between the running efficiency and model parameters. However, the two

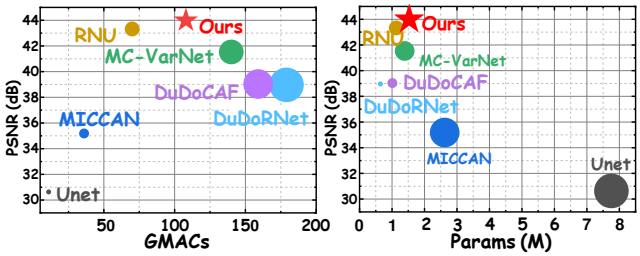

Figure 6: Comparisons on PSNR, GMACs, and Params.

Figure 7: Error and SSIM curves of stage-wise reconstructions with and without DCM.

competitors, together with all others, clearly lag behind our IMA in terms of PSNR values.

Ablation Study

Impact of Proposed Modules. To evaluate the roles of CSCA and DCM, two ablation studies using an $8\times$ radial mask are carried out on the IXI dataset (Table 2(a)). The addition of CSCA, with few extra parameters, increases the PSNR by 1.181dB, showing its effectiveness in gathering complementary information. Adopting DCM improves the PSNR by 2.134dB without additional parameters. To further verify DCM's advantages, Fig. 7 shows the stage-wise error and SSIM curves under a $4\times$ radial mask on IXI. The DCM-enabled model optimizes smoothly. In contrast, without DCM, the curves fluctuate severely, highlighting the benefits of progressive masking for data fidelity.

Impact of Different Unfolding Stages. To evaluate the performance variations across different iterations, the results traversed from {6, 8, 10, 12, 14} are shown in Fig. 8. As expected, more iterations consistently boost performance, yet with the cost of more computational demands. Since the performance improvement tends to be trivial when the

(a) Ablation of CSCA and DCM (mask: $8\times$ radial)

	PSNR	SSIM	Params(M)	GMACs
with CSCA	41.948▲2%	0.992▲0.2%	1.53	107.92
w/o CSCA	40.767	0.990	1.33	103.28
with DCM	41.948▲5%	0.992▲0.3%	1.53	107.92
w/o DCM	39.814	0.989	1.53	107.92

(b) Ablation of kernel type (mask: $8\times$ radial)

	IXI			fastMRI		
Kernels	1	2	3	1	2	3
PSNR	41.948	40.286 ▼4%	40.872▼3%	28.493	28.167▼1%	28.319▼1%
SSIM	0.992	0.991▼0.1%	0.991▼0.1%	0.783	0.776▼0.9%	0.779▼0.5%

Table 2: Ablation study on several key components, ▲ and ▼ represent the lifting and dropping percentage, respectively.

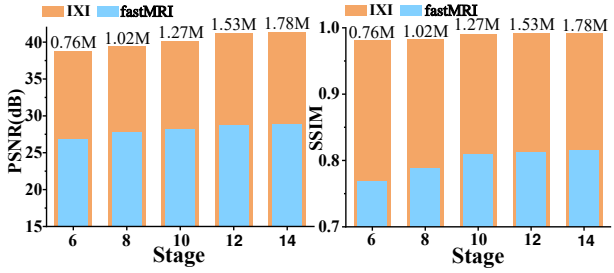

Figure 8: Comparisons of different stages ($8\times$ random mask).

stage reaches 12, which is then set as the default value. Note that even with only 6 iterations, the performance has already approximated closely to the peak value, which again verifies the positive contributions of breaking information isolation.

Impact of Various Kernel Selections. As in Fig. 4, choosing DCM kernels under the same mask is empirical, depending on data and sampling. Table 2(b) ablates their impact on reconstruction. Results vary with kernels, showing masking-strategy study's importance. The first kernel performs best on both datasets, so we use it. Recall that the masking mechanism's efficacy was ablated in Table 2(a).

Conclusion

In this study, a new MRI acceleration scheme, namely IMA, is meticulously engineered. Specifically, the formulation of target sequence is explicitly molded via the summation of two transformations from references and a self-learned variable. With specialized operators that capture exclusive features, spatially complementary information can be assembled. In addition, a lightweight cross-modal converging attention is elaborated to perform direct interaction of multiple sequences, which further alleviates the issue of information isolation. More importantly, IMA is stage-wise equipped with a newly designed DCM mechanism, which enables the network to progressively spread information from known to unknown areas, thus promoting the capability in high-frequency learning. Extensive experiments are conducted with guarantees of new state-of-the-art scores. Moreover, all results are achieved with well-balanced GMACs and Params. We hope that our research would offer valuable insights to the community in designing effective MRI architectures.

Acknowledgments

This work was supported in part by the Key Program of Natural Science Foundation of Zhejiang Province under Grant LZ24F030012, the China Postdoctoral Science Foundation under grant ZY24211190002, and the National Natural Science Foundation of China under Grant 62276232.

References

Chen, J.; Jiang, J.; Wu, F.; and Zheng, J. 2024. Null Space Matters: Range-Null Decomposition for Consistent Multi-ContrastMRI Reconstructio. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Fang, C.; Zhang, D.; Wang, L.; Zhang, Y.; Cheng, L.; and Han, J. 2022. Cross-modality high-frequency transformer for mr image super-resolution. In *Proceedings of the 30th ACM International Conference on Multimedia*, 1584–1592.

Feng, C.-M.; Yan, Y.; Chen, G.; Xu, Y.; Hu, Y.; Shao, L.; and Fu, H. 2023. Multimodal Transformer for Accelerated MR Imaging. *IEEE Transactions on Medical Imaging*, 42(10): 2804–2816.

Guo, D.; Zeng, G.; Fu, H.; Wang, Z.; Yang, Y.; and Qu, X. 2023. A Joint Group Sparsity-based deep learning for multi-contrast MRI reconstruction. *Journal of Magnetic Resonance*, 346: 107354.

Haji, S. H.; and Abdulazeez, A. M. 2021. Comparison of optimization techniques based on gradient descent algorithm: A review. *PalArch's Journal of Archaeology of Egypt/Egyptology*, 18(4): 2715–2743.

Huang, J.; Ding, W.; Lv, J.; Yang, J.; Dong, H.; Del Ser, J.; Xia, J.; Ren, T.; Wong, S. T.; and Yang, G. 2022. Edge-enhanced dual discriminator generative adversarial network for fast MRI with parallel imaging using multi-view information. *Applied Intelligence*, 52(13): 14693–14710.

Huang, Q.; Yang, D.; Wu, P.; Qu, H.; Yi, J.; and Metaxas, D. 2019. MRI reconstruction via cascaded channel-wise attention network. In *IEEE 16th International Symposium on Biomedical Imaging (ISBI)*, 1622–1626. IEEE.

Jiang, J.; Chen, J.; Xu, H.; Feng, Y.; and Zheng, J. 2023a. GA-HQS: MRI reconstruction via a generically accelerated unfolding approach. In *IEEE International Conference on Multimedia and Expo (ICME)*, 186–191. IEEE.

Jiang, J.; Feng, Y.; Chen, J.; Guo, D.; and Zheng, J. 2023b. Latent-space Unfolding for MRI Reconstruction. In *Proceedings of the 31st ACM International Conference on Multimedia*, 1294–1302.

Jiang, J.; He, Z.; Quan, Y.; Wu, J.; and Zheng, J. 2024. PGIUN: Physics-Guided Implicit Unrolling Network for Accelerated MRI. *IEEE Transactions on Computational Imaging*, 10: 1055–1068.

Lei, P.; Fang, F.; Zhang, G.; and Xu, M. 2023a. Deep unfolding convolutional dictionary model for multi-contrast MRI super-resolution and reconstruction. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, 1008–1016.

Lei, P.; Fang, F.; Zhang, G.; and Zeng, T. 2023b. Decomposition-Based Variational Network for Multi-Contrast MRI Super-Resolution and Reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 21296–21306.

Lyu, J.; Sui, B.; Wang, C.; Tian, Y.; Dou, Q.; and Qin, J. 2022. DuDoCAF: Dual-domain cross-attention fusion with recurrent transformer for fast multi-contrast MR imaging. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 474–484. Springer.

Peng, J.; Wang, Y.; Zhang, H.; Wang, J.; and Meng, D. 2022. Exact decomposition of joint low rankness and local smoothness plus sparse matrices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5): 5766–5781.

Pooja, K.; Ramzi, Z.; Chaithya, G.; and Ciuciu, P. 2022. MC-PDNET: Deep unrolled neural network for multi-contrast mr image reconstruction from undersampled k-space data. In *IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, 1–5. IEEE.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI*, 234–241. Springer.

Song, J.; Chen, B.; and Zhang, J. 2023. Deep Memory-Augmented Proximal Unrolling Network for Compressive Sensing. *International Journal of Computer Vision*, 131(6): 1477–1496.

Song, P.; Weizman, L.; Mota, J. F.; Eldar, Y. C.; and Rodrigues, M. R. 2019. Coupled dictionary learning for multi-contrast MRI reconstruction. *IEEE Transactions on Medical Imaging*, 39(3): 621–633.

Sun, H. 2024. Ultra-High Resolution Segmentation via Boundary-Enhanced Patch-Merging Transformer. arXiv:2412.10181.

Sun, H.; Xu, L.; Jin, S.; Luo, P.; Qian, C.; and Liu, W. 2024. PROGRAM: PROtotype GRAph Model based Pseudo-Label Learning for Test-Time Adaptation. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Sun, K.; Wang, Q.; and Shen, D. 2024. Joint Cross-Attention Network With Deep Modality Prior for Fast MRI Reconstruction. *IEEE Transactions on Medical Imaging*, 43(1): 558–569.

Wang, J.; Ding, D.; Li, Z.; Feng, X.; Cao, C.; and Ma, Z. 2023a. Sparse Tensor-Based Multiscale Representation for Point Cloud Geometry Compression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7): 9055–9071.

Wang, J.; Yang, Y.; Yang, H.; Lian, C.; Xu, Z.; and Sun, J. 2023b. MD-GraphFormer: A Model-Driven Graph Transformer for Fast Multi-Contrast MR Imaging. *IEEE Transactions on Computational Imaging*, 9: 1018–1030.

Wei, J.; Wang, Z.; Wang, K.; Guo, L.; Fu, X.; Liu, J.; and Chen, X. 2023. Accurate MRI reconstruction via multi-domain recurrent networks. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, 1524–1532.

Wen, Y.-H.; Gao, L.; Fu, H.; Zhang, F.-L.; Xia, S.; and Liu, Y.-J. 2022. Motif-GCNs with local and non-local temporal blocks for skeleton-based action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2): 2009–2023.

Xie, J.; Zhang, J.; Zhang, Y.; and Ji, X. 2022. PUERT: Probabilistic under-sampling and explicable reconstruction network for CS-MRI. *IEEE Journal of Selected Topics in Signal Processing*, 16(4): 737–749.

Xu, H.; Zheng, J.; Yao, X.; Feng, Y.; and Chen, S. 2021. Fast tensor nuclear norm for structured low-rank visual inpainting. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(2): 538–552.

Yang, G.; Zhang, L.; Liu, A.; Fu, X.; Chen, X.; and Wang, R. 2023. MGDUN: An interpretable network for multi-contrast MRI image super-resolution reconstruction. *Computers in Biology and Medicine*, 167: 107605.

Yang, G.; Zhang, L.; Zhou, M.; Liu, A.; Chen, X.; Xiong, Z.; and Wu, F. 2022. Model-Guided Multi-Contrast Deep Unfolding Network for MRI Super-resolution Reconstruction. In *Proceedings of the 30th ACM International Conference on Multimedia*, 3974–3982.

Yang, Y.; Sun, J.; Li, H.; and Xu, Z. 2018. ADMM-CSNet: A deep learning approach for image compressive sensing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(3): 521–538.

Yang, Y.; Wang, N.; Yang, H.; Sun, J.; and Xu, Z. 2020. Model-driven deep attention network for ultra-fast compressive sensing MRI guided by cross-contrast MR image. In *Medical Image Computing and Computer Assisted Intervention–MICCAI*, 188–198. Springer.

Zhang, J.; Zhang, Z.; Xie, J.; and Zhang, Y. 2022. High-throughput deep unfolding network for compressive sensing MRI. *IEEE Journal of Selected Topics in Signal Processing*, 16(4): 750–761.

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

Zhou, B.; and Zhou, S. K. 2020. DuDoRNet: learning a dual-domain recurrent network for fast MRI reconstruction with deep T1 prior. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4273–4282.

Zhou, X.; Zhang, Z.; Du, H.; and Qiu, B. 2024. MLMFNet: A multi-level modality fusion network for multi-modal accelerated MRI reconstruction. *Magnetic Resonance Imaging*.