

# LMO: Linear Mamba Operator for MRI Reconstruction

Wei Li<sup>1\*</sup>, Jiawei Jiang<sup>1\*</sup>, Jie Wu<sup>1</sup>, Kaihao Yu<sup>1</sup>, Jianwei Zheng<sup>1†</sup>  
<sup>1</sup>Zhejiang University of Technology, Hangzhou, Zhejiang

## Abstract

*Interpretability and consistency have long been crucial factors in MRI reconstruction. While interpretability has been significantly innovated with the emerging deep unfolding networks, current solutions still suffer from inconsistency issues and produce inferior anatomical structures. Especially in out-of-distribution cases, e.g., when the acceleration rate (AR) varies, the generalization performance is often catastrophic. To counteract the dilemma, we propose an innovative Linear Mamba Operator (LMO) to ensure consistency and generalization, while still enjoying desirable interpretability. Theoretically, we argue that mapping between function spaces, rather than between signal instances, provides a solid foundation of high generalization. Technically, LMO achieves a good balance between global integration facilitated by a state space model that scans the whole function domain, and local integration engaged with an appealing property of continuous-discrete equivalence. On that basis, learning holistic features can be guaranteed, tapping the potential of maximizing data consistency. Quantitative and qualitative results demonstrate that LMO significantly outperforms other state-of-the-arts. More importantly, LMO is the unique model that, with AR changed, achieves retraining performance without retraining steps. Codes are available at <https://github.com/ZhengJianwei2/LMO>.*

## 1. Introduction

As a dual-free technology, i.e., radiation-free and invasive-free, Magnetic Resonance Imaging (MRI) has become the most effective diagnostic tool for diseases of the brain [41], spinal cord [30], and tumors [24] in modern biomedical science. However, due to hardware limitations, high-quality imaging often requires long scanning times, which may cause discomfort to the patient. A common approach to MRI acceleration is recovering high-quality images from undersampled k-space data, with Compressed Sensing (CS) [31] being the most widely used technique. Unfortunately,

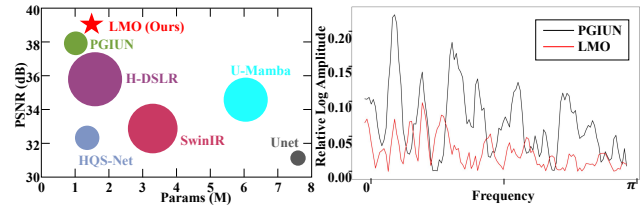


Figure 1. (Left) Comparisons on PSNR and Params/GFLOPs. (Right) Relative log amplitudes of spectral feature maps between the reconstructed image and the Ground Truth.

although CS has significantly sped up MRI acquisition, undersampled k-space data may suffer from the issue of aliasing error, as it violates the Nyquist theorem. The presence of aliasing artifacts greatly impacts the clinical diagnosis, which is unacceptable in the life and health sciences. To that end, developing new algorithms with excellent data consistency remains an urgent topic.

In common practice, manually added priors such as sparsity [35], total variation (TV) [23], Laplacian [43] and Curvature [59] are used to assist in the reconstruction of the originated signal. Despite excellent performance demonstrated in certain applications, these manually designed priors are often empirically established, which may not entirely capture the complexity of actual MRI data. The approach of jointing multiple priors is considered as palliative, yet introduces additional challenges of tricky model selection and heavy computational burden.

In contrast, deep learning methods capture the implicit knowledge in a data-driven fashion, rather than assuming some explicit priors on MRIs. To date, massive network architectures have been shifted into the accomplishment of MRI reconstruction task, such as convolutional neural networks (CNNs) [47, 54], UNet [14], transformer [10, 11], and Mamba [15, 34, 63], all of which are capable of establishing complex mappings between paired inputs and outputs. However, despite the appealing performance, there are still some notable drawbacks. First, these methods are not specifically tailored for MRI acceleration. Adapting general frameworks that have succeeded in other fields to specific medical applications may not be entirely appropriate. Second, the lack of interpretability is also a significant issue, as

\*Equal contribution.

†Corresponding author. {zjw@zjut.edu.cn}

a purely deep network is generally known as a black box.

Deep Unrolling Networks (DUNs) [5, 46] and Neural Operators (NOs) [25, 26, 39] are two of the most promising directions to enforce the network architecture with better interpretability. DUNs frame the MRI acceleration problem as an optimization task, embedding domain-specific knowledge directly into network architecture. The representative approaches [16, 52] not only promote interpretability by making the process more transparent but also achieve acceptable performance by leveraging the inherent structure of the MRI data. However, DUNs also have their own suffering. To the best of our knowledge, current unrolling methods typically operate in the original signal space, limiting their ability to capture the intrinsic features of complex signals. Besides, approaches that focus on learning frequency information often fall short in capturing local details, since the k-space, where the original measurements are taken, corresponds to the global spatial domain. Last but not least, the generalizability of such algorithms often leaves much to be desired. NOs, on the other hand, are well-suited to address these issues. By integral computation in a high-dimensional space, neural operators naturally learn richer features of input functions. More importantly, NOs learn mappings between continuous spaces [22], rather than instances, with the potential for high generalization being maintained. Representatively, NOs based on Fourier [25] and transformer [9, 26] enable the capture of long-range dependencies, yet with relatively heavier complexities of  $O(n \log n)$  and  $O(n^2)$ , respectively. In contrast, based on the convolutions, CNO [39] with  $O(n)$  time complexity is effective in learning localized and fine-grained features. Unfortunately, to date, NOs have never been used for MRI acceleration. In this study, we attempt to fill this gap by innovating a Mamba-based NO variant, which enjoys merits of both holistic feature learning and linear complexity. Some initial results are given in Fig. 1, in which the radius of the circle denotes the metric of GFLOPs. Evidently, other competitors lag behind LMO by a large margin.

Note that despite the advantages of Mamba in  $O(n)$  computational complexity and long-range dependency [62], the absence of a suitable integral form has plagued its naive use in the context of operator learning. We argue that this limited use of the state space model for operator learning stands in complete contrast to the fact that Mambas and their variants are extensively used architectures in visual community [55]. Given their innate globality and computational efficiency, we believe it could be very advantageous to bring Mamba into the reckoning for operator learning. The practical contributions are threefold.

- We individually propose two integral forms, i.e., scanning integration (SI) and convolution integration (CI), to capture both the global and local function features. More importantly, holistic information is achieved with a com-

putational complexity of only  $O(n)$ , tapping the potential for attractive consistency under high efficiency.

- By feeding the two newly elaborated integrations into traditional NO architecture, we then propose an innovative Linear Mamba Operator (LMO), which performs network mapping entirely in bandlimited function spaces, hence enjoying superior generalization, as k-space is a special case of bandlimited function spaces.
- Theoretically, all the computations involved are rooted in solid mathematical deduction, enabling the final model with high interpretability. Empirically, extensive experiments on single- and multi-coil reconstructions demonstrate that LMO outperforms current state-of-the-art both numerically and visually. Codes are available.

Hence, we offer a new NO-based MRI reconstruction model incorporating more holistic features, with favorable properties in theory and excellent performance in practice.

## 2. Related Work

### 2.1. MRI Acceleration

MRI is a non-invasive medical imaging technique that generates detailed images of the internal tissues using magnetic fields and radio waves. Due to the relatively slow imaging speed of MRI, extensive research has been devoted to accelerating the imaging process to enhance its clinical efficiency, among which compressed sensing, deep learning, and deep unrolling networks have drawn the most intensive attention. Compressed sensing [61] seeks to promote reconstruction speed by minimizing data inconsistency during MRI scans. A variety of compressed sensing methods have been introduced, such as sparse regularization [58], low rankness [21, 36], non-local similarity [49] and edge sharpness [18]. Although these handcrafted priors have delivered exceptional performance in specific scenarios, they are usually developed through experience that may not fully capture the complexities of actual MRI data, leading to poor generalization [4]. Recently, deep learning methods have also demonstrated promising results in MRI acceleration. Jethi et al. [14] designed a dual-encoder UNet to capture multi-scale features, yet is still trapped in issues of restricted receptive fields and sensitivity to input variations. The advent of transformers [6, 29, 51, 56] has significantly alleviated these limitations, holding the ability for global modeling with a certain sacrifice of more computational resources. Recently, Ma et al. [34] introduced the U-Mamba architecture, marrying the advantage of multi-scale learning with linear complexity, yet the generalization is still limited. Following the DUN branch, HQS-Net [53] is a pioneering work that utilizes the half-quadratic splitting (HQS) algorithm to process highly undersampled k-space MRI data. Similarly, Jiang et al. [17] innovatively employed a parallel dual-domain update mechanism within a first-order optimizer,

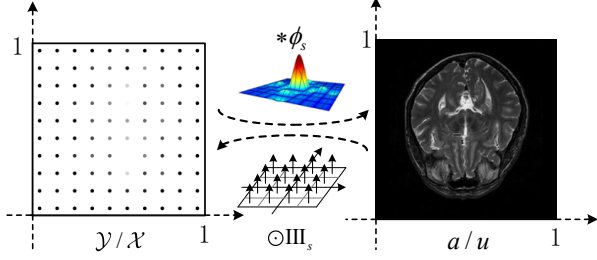


Figure 2. Discrete representation  $\mathcal{Y}/\mathcal{X}$  and continuous function  $a/u$  are related to each other via convolution with ideal interpolation filter  $\phi_s$  and pointwise multiplication with Dirac comb  $\text{III}_s$ .

setting state-of-the-art scores in in-distribution cases. To further enhance generalization performance, H-DSLR [38] employs a single pre-trained CNN, which adapts to different datasets by functioning as various linear annihilation filter-banks. Overall, current techniques have shown promising results in accelerating MRI reconstruction. However, most of them enjoy part of the practical demands in facilitating clinical diagnosis, leaving much room for improvement.

## 2.2. Neural Operator

By combining linear integral operators with non-linear activation functions, neural operators [25, 60] were initially proposed for the solving of partial differential equations. The combination allows neural operators to fit mappings between arbitrary function spaces [13], thus achieving promising generalization capabilities. With the booming technology, researchers have moved beyond the vanilla integral forms, engineering new kernels and the concerned operators, such as Oformer [26] using attention-based integral and CNO [39] based on the convolutional integral. Moreover, these days have witnessed the shift of the NO paradigm into the computer vision field. For example, SRNO [48], FNOseg3D [50], and FNO have been successfully mitigated into image super-resolution, 3D medical segmentation, and pattern classification [19], respectively, knocking out most of the traditional solutions. These successful cases have also inspired us to employ NOs for MRI acceleration.

## 3. Methodology

### 3.1. Problem Setting

Fast MRI reconstruction is considered to recover the target image  $\mathcal{X} \in \mathbb{R}^{H \times W \times 1}$  from the undersampled measurement  $\mathcal{Y} \in \mathbb{R}^{H \times W \times 1}$ . Additionally, we can define  $\mathcal{Y}$  and  $\mathcal{X}$  as discrete representations of vector-valued functions  $a : \Omega \rightarrow \mathbb{R}^1, u : \Omega \rightarrow \mathbb{R}^1$  in Sobolev spaces  $\mathcal{A}$  and  $\mathcal{U}$  respectively, where  $\Omega$  represents a 2-dimensional bounded domain. As illustrated in Fig. 2, given a discrete

grid  $\mathcal{Y}[y]$  and a sampling rate  $s$ , the Whittaker–Shannon interpolation formula [44] states that the continuous representation  $a(y)$  is obtained by convolving  $\mathcal{Y}[y]$  with an ideal interpolation filter  $\phi_s$ . This is expressed as  $a(y) = (\phi_s * \mathcal{Y})(y)$ , where  $\phi_s(y) = \text{sinc}(sy_0) \cdot \text{sinc}(sy_1)$  and  $\text{sinc}(sy_i) = \frac{\sin(\pi sy_i)}{\pi sy_i}, i = 1, 2$ . The filter  $\phi_s$  has a bandlimit of  $s/2$  along the horizontal and vertical dimensions, ensuring that the continuous signal  $a(y)$  captures all representable frequencies at the sampling rate  $s$ . To convert from the continuous to the discrete domain,  $a(y)$  is sampled at the grid points of  $\mathcal{Y}[y]$ , offset by half the sample spacing to lie at the “pixel centers”. This is represented by pointwise multiplication with a two-dimensional Dirac comb  $\text{III}_s(y) = \sum_{Y \in \mathbb{Z}^2} \delta(y - (Y + \frac{1}{2})/s)$ , where  $Y$  represents the position indices of sampling points on the integer grid. Similarly, we can achieve the conversion between the continuous function  $u$  and its discrete representation  $\mathcal{X}$ .

Now, the core purpose here involves learning a mapping between two infinite-dimensional continuous spaces via discrete representations of input-output functions, i.e.,  $\mathcal{G}^\dagger : \mathcal{A} \rightarrow \mathcal{U}$ . Generally, the problem concerned can be formally delineated as follows. By constructing a parametric map  $\mathcal{G}_\theta : \mathcal{A} \rightarrow \mathcal{U}$ , the practice would be to build an approximation of  $\mathcal{G}^\dagger$  from data pairs  $(a_i, u_i)_{i=1}^N \in \mathcal{A} \times \mathcal{U}$ , with  $\theta$  in the finite-dimensional space  $\mathbb{R}^p$ . The practical learning of  $\mathcal{G}_\theta$  can be naturally addressed through the empirical-risk minimization problem, such as

$$\begin{aligned} & \min_{\theta \in \mathbb{R}^p} \mathbb{E} \|\mathcal{G}^\dagger(a) - \mathcal{G}_\theta(a)\|_{\mathcal{U}}^2 \\ & \approx \min_{\theta \in \mathbb{R}^p} \frac{1}{N} \sum_{i=1}^N \|\mathcal{G}^\dagger(a_i) - \mathcal{G}_\theta(a_i)\|_{\mathcal{U}}^2, \quad a \in \mathcal{A}. \end{aligned} \quad (1)$$

Considering that the bandlimited function space  $\mathcal{B}$ , as a subspace of the Sobolev space  $\mathcal{A}$  and  $\mathcal{U}$ , facilitates to achieve equivalence between the underlying operator and its discrete representations more easily [2]. Therefore, we select subspace  $\mathcal{B}$  instead of the original  $\mathcal{A}$  and  $\mathcal{U}$ . In this respect, we elaborate the space of bandlimited functions defined by

$$\mathcal{B}_w(\Omega) = \{f \in L^2(\Omega) : \text{supp} \hat{f} \subseteq [-w, w]^2\}, \quad (2)$$

in which  $w > 0$  denotes the frequency bound and  $\hat{f}$  represents the Fourier transform of  $f$ . Note that if a bandlimited function can approximate the original function with arbitrary precision (depending on  $w$ ), then a bandlimited operator mapping between bandlimited functions can also approximate the original operator with arbitrary precision [3, 39]. In other words, for any  $\varepsilon > 0$ , there exist a  $w$  and a continuous operator  $\mathcal{G}_w : \mathcal{B}_w(\Omega) \rightarrow \mathcal{B}_w(\Omega)$  such that  $\|\mathcal{G}^\dagger - \mathcal{G}_w\| < \varepsilon$ , with  $\|\cdot\|$  pertaining to the corresponding operator norm. In addition, let  $P_w$  denote a certain frequency projection (FP), as shown in Fig. 3,  $P_w(g)$  is capable of

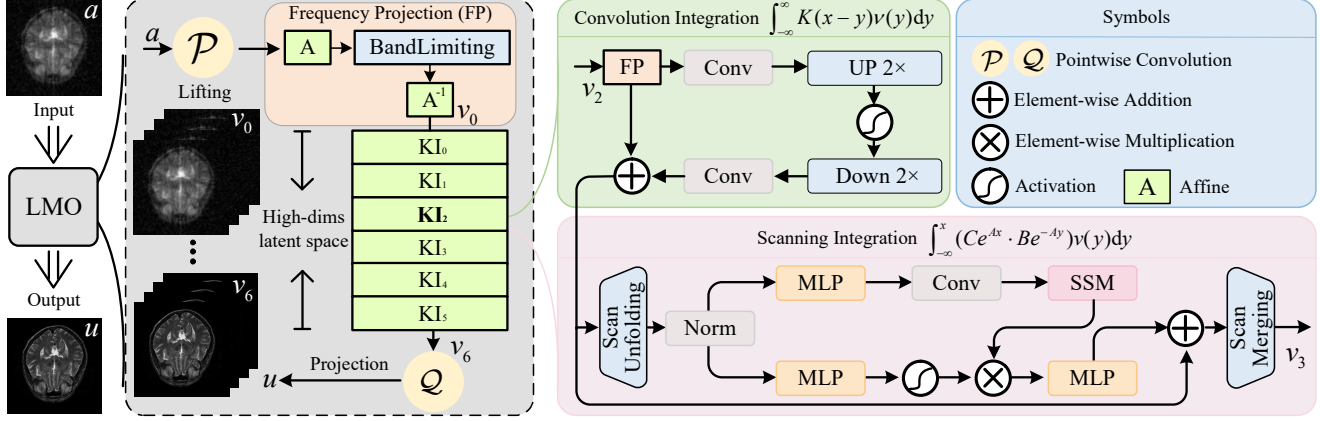


Figure 3. The overall architecture of LMO. The core ingredient lies in the joint scanning and convolution integration.

discarding the high-frequency components higher than frequency  $w$ , where  $g \in \mathcal{A}(\Omega)$  is any function in that space. As the underlying operator maps between spaces of bandlimited functions, the operator approximation architecture should be designed in a manner that preserves this structure. That is to say, it implies that our network architecture should be equivalent to the operator handling bandlimited functions when processing discrete versions of bandlimited functions, which we refer to as *continuous-discrete equivalence* (CDE) in MRI reconstruction. To that end, a Linear Mamba Operator for MRI Reconstruction is engineered to approximate the operator  $\mathcal{G}_\theta$ . Following the standard paradigm adopted by existing NOs [12, 60], our practical approach is similarly based on a compositional mapping:

$$\mathcal{G}_\theta := \mathcal{Q} \circ \eta_{t-1}(W_{t-1} + \mathcal{K}_{t-1}) \circ \dots \circ \eta_0(W_0 + \mathcal{K}_0) \circ \mathcal{P}, \quad (3)$$

where

$$\mathcal{P} : \{a \in \mathcal{B}_w(\Omega, \mathbb{R}^1)\} \rightarrow \{v_0 \in \mathcal{B}_w(\Omega, \mathbb{R}^{d_0})\}, \quad (4a)$$

$$\mathcal{Q} : \{v_t \in \mathcal{B}_w(\Omega, \mathbb{R}^{d_t})\} \rightarrow \{u \in \mathcal{B}_w(\Omega, \mathbb{R}^1)\}, \quad (4b)$$

are the lifting and projection mappings respectively,  $W_t \in \mathbb{R}^{d_{t+1} \times d_t}$  are linear operators in a matrix form,  $\eta_t$  are activation functions. It is worth noting that all these operations are performed locally or pointwisely, adhering to the principle of discretization invariance [19] to some extent. The general formulation of the kernel integral operator  $\mathcal{K}_t : \{v_t : \Omega \rightarrow \mathbb{R}^{d_t}\} \rightarrow \{v_{t+1} : \Omega \rightarrow \mathbb{R}^{d_{t+1}}\}$  is parameterized by  $\alpha$  as follows:

$$(\mathcal{K}_t(v_t); \alpha)(x) = \int_{\Omega} K_t(x, y, v_t(x), v_t(y))v_t(y)dy, \quad (5)$$

in which the parameter of kernel  $K_t$  is learnt from given data and  $x, y \in \Omega$ . For example, transformer-based neural operators [26, 48] use attention mechanisms as its primary operation, while FNO [25, 32] relies on global convolution.

In this work, a new integral form sequentially jointing scanning integration that captures long-range dependencies and convolution integration that learns local bias is crafted, ensuring a holistic feature learning.

### 3.2. Scanning Integration

Building on the scheme in FNO [25, 27], we initially assume that  $\mathcal{K}_t : \mathbb{R}^\Omega \times \mathbb{R}^\Omega \rightarrow \mathbb{R}^{d_{t+1} \times d_t}$  is primarily influenced by the input pair  $(x, y)$ , with minimal dependence on the spatial variables  $(v_t(x), v_t(y))$ . Then, we let

$$K_t(x, y) = Ce^{Ax} \cdot Be^{-Ay}, \quad (6)$$

where  $A, B$ , and  $C$  are temporarily constant parameters, as for an easier deduction purpose. To further ensure a possible employment of the scanning pattern used in Mamba [62], we set the integration interval to  $y \in (-\infty, x)$  rather than the entire definition domain  $\Omega$  and omit the subscript  $t$  from Eq. (5) hereafter, which is originally used to denote the number of iterations during integration flow. Hence, Eq. (5) can be rewritten as follows.

$$(\mathcal{K}(v); \alpha)(x) = \int_{-\infty}^x (Ce^{Ax} \cdot Be^{-Ay})v(y)dy. \quad (7)$$

It is evident that  $Ce^{Ax}$  does not depend on the integration variable  $y$ , allowing Eq. (7) to be reformulated as:

$$(\mathcal{K}(v); \alpha)(x) = Ch(x), \quad (8)$$

with  $h(x) = e^{Ax} \int_{-\infty}^x B(e^{-Ay})v(y)dy$ . Additionally, by applying a straightforward differentiation with respect to  $x$ , we can obtain the following equation.

$$h'(x) = Ah(x) + Bv(x), \quad (9)$$

which together with Eq. (8) leads to

$$(\mathcal{K}(v))(x) = Ch(x), \quad (10)$$

By now, we have seamlessly married the computation of kernel integral in Eq. (5) with a State Space Model (SSM) [8, 55]. Inspired by the theory of continuous systems, the objective of Eq. (10) is to transform a 2-dimensional function  $v(x)$  to  $(\mathcal{K}(v))(x)$  through the hidden space  $h(x)$ . Within this context,  $A$  serves as the evolution parameter, while  $B$  and  $C$  act as the projection parameters. To incorporate Eq. (10) into deep learning paradigm, a discretization process is first required. Note this transformation is crucial to align the model with the sampling rate of the underlying signal embodied in the input data, enabling computationally efficient operations. In contrast to the Monte Carlo approximation commonly applied in traditional integrals, our method utilizes the zero-order hold (ZOH) technique, which can be formally defined as follows.

$$\begin{aligned}\bar{A} &= \exp(\Delta A), \\ \bar{B} &= (\Delta A)^{-1} (\exp(\Delta A) - \mathbf{I}) \cdot \Delta B,\end{aligned}\quad (11)$$

where  $\Delta$  is a timescale parameter converting continuous parameters  $A$  and  $B$  into their discrete counterparts  $\bar{A}$  and  $\bar{B}$ . Therefore, the discrete representation of Eq. (10) can be formulated as follows.

$$\begin{aligned}h(x_k) &= \bar{A}h(x_{k-1}) + \bar{B}v(x_k), \\ (\mathcal{K}(v))(x_k) &= C\bar{h}(x_k).\end{aligned}\quad (12)$$

The detailed derivations from Eq. (10) to Eq. (12) can be found in the supplementary material. Our objective is to perform the integration of the 2-dimensional function using a scanning approach, efficiently capturing global information with  $O(n)$  complexity. In Eq. (12),  $h(x_k)$ , representing the hidden space, encapsulates relevant information about the integrated points before  $x$ . Therefore, the output can be obtained through global convolution [20, 42]:

$$\begin{aligned}\bar{\mathbf{K}} &= (C\bar{B}, C\bar{A}\bar{B}, \dots, C\bar{A}^{(s-1)}\bar{B}), \\ (\mathcal{K}(v))(x)(x_s) &= v(x_s) * \bar{\mathbf{K}},\end{aligned}\quad (13)$$

where  $\bar{\mathbf{K}} \in \mathbb{R}^s$  represents a structured convolution kernel and  $s$  is the sampling rate of input  $v$ . The pseudo-code for the Scanning Integration is presented in Algorithm 1.

### 3.3. Convolution Integration

While scanning integration benefits from a global receptive field, we believe that incorporating local convolution integration could further enhance the capture of more holistic features. To commence with, the convolution integration for  $\mathcal{K} : \mathcal{B}_w(\Omega) \rightarrow \mathcal{B}_w(\Omega)$  is defined as:

$$\begin{aligned}(\mathcal{K}(v); \alpha)(x) &= \int_{\Omega} K(x-y)v(y)dy \\ &\approx \sum_{i,j=1}^k \kappa_{ij}v(x-y_{ij})\end{aligned}\quad (14)$$

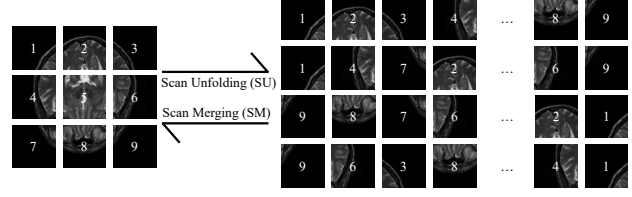


Figure 4. The scan unfolding and scan merging operation integrates pixels from four directions with  $O(n)$  complexity.

where  $v \in \mathcal{B}_w$ ,  $\kappa$  is a discrete kernel with size  $k \in \mathbb{N}$  and  $v(x-y_{ij})$  represents the value of function  $v$  at position  $x$  shifted by  $y_{ij}$ . Thus, the convolution operator can be intuitively parameterized in physical space, deviating far from the treatments of Fourier transformation and then followed by matrix multiplication, as in FNO [25, 28]. Additionally, since our convolution operates within a bandlimited function space [39], it cleverly avoids the challenge faced by conventional convolutions as discussed in [33]: that as resolution increases, a standard convolutional layer converges to a pointwise operator.

### 3.4. Linear Mamba Operator

With continuous functions as inputs/outputs, the concerns can be organized into an operator architecture, as given in Fig. 3. The input function, represented as  $a \in \mathcal{B}_w(\Omega, \mathbb{R}^1)$ , is initially lifted and then processed through a Frequency Projection module and a sequence of kernel integration (KI) layers. Recall that the kernel integration consists of two main layers, i.e., scanning integration (Eq. 10) and convolution integration (Eq. 14). Overall, integration is employed functionally to learn mappings between continuous spaces.

The form of convolution integration is as Eq. (14), which is defined using a finite sum, with each term  $v(x-y_{ij})$  being regarded as a translation of  $v$ . Since  $v \in \mathcal{B}_w(\Omega, \mathbb{R}^d)$ , the translation is still in  $\mathcal{B}_w(\Omega, \mathbb{R}^d)$  [39]. Actions such as multiplying by a finite number of constant coefficients  $\kappa_{ij}$  and taking the sum will not cause the function to escape from this space. From another perspective, if the functions in  $\mathcal{B}_w(\Omega, \mathbb{R}^d)$  have certain properties, such as smoothness and rapid decay, then the weighted sum and movement in convolution will not change the basic properties of the function. Therefore, we present the commutative diagram of the convolution integration:

$$\begin{array}{ccc} \mathcal{B}_w(\Omega, \mathbb{R}^d) & \xrightarrow{CI} & \mathcal{B}_w(\Omega, \mathbb{R}^d) \\ \phi_s \uparrow & & \downarrow \text{III}_s \\ \ell^2(\mathbb{Z}^2) & \xrightarrow{\sum kv} & \ell^2(\mathbb{Z}^2) \end{array}$$

where  $\sum \kappa v$  is an abbreviated expression form of  $\sum_{i,j=1}^k \kappa_{ij}v(x-y_{ij})$ . Accordingly, the output of scanning

integration (also known as SSM) can be defined as a global convolution integral like Eq. (13). Similar to the defined convolution integral, such a global integral is a well-defined operator from  $\mathcal{B}_w(\Omega, \mathbb{R}^d)$  to itself,

$$\begin{array}{ccc} \mathcal{B}_w(\Omega, \mathbb{R}^d) & \xrightarrow{SSM/SI} & \mathcal{B}_w(\Omega, \mathbb{R}^d) \\ \phi_s \uparrow & & \downarrow \Pi_s \\ \ell^2(\mathbb{Z}^2) & \xrightarrow{\bar{\mathbf{K}}v} & \ell^2(\mathbb{Z}^2) \end{array}$$

where  $\bar{\mathbf{K}}$  is global convolution kernel in Eq. (13). Practically, the constant parameters  $A$ ,  $B$ , and  $C$  in Eq. (8) are better learned from the data, allowing attractive model adaptability. This together with the shifted integral interval, as shown in Eq. (7), enables scan operation within the state space. Finally, as illustrated in Fig. 4, we choose to unfold sampling points into sequences along with rows and columns and then proceed with scanning along four different directions. These sequences are then processed by the SSM block for kernel integration, ensuring that information from multiple directions is thoroughly scanned, thereby capturing a diverse range of function features. Subsequently, the sequences from the four directions are merged to reconstruct the output function. In summary, the above content provides the implementation of the discretized structure of the underlying continuous operator, demonstrating that our algorithm conforms to the *continuous-discrete equivalence*.

### 3.5. Complexity Analysis of Kernel Integral

**Scanning Integration:** given a discrete counterpart  $v \in R^{H \times W \times D}$  of a 2-dimensional continuous function, one can reshape it to get  $v \in R^{n \times D}$  with  $n = H \times W$ . Now, we assume the sequence dimensions of  $\Delta, A, B, C$  are all  $D$ . As outlined in step 1 of Algorithm 1, the complexity of generating three learnable projections is  $O(3nD^2)$ . The discretization steps 2 and 3 involve four matrix multiplications, which cost  $O(4nD^2)$ . Step 4, which implements the state space model, has a complexity of  $O(n)$  complexity [7]. In summary, the computations of Algorithm 1 are all linear with the sequence length, i.e.,  $O(n)$ . **Convolution Integration:** The local convolution operator is defined by Eq. (14). Evaluating each point involves  $k^2$  multiplications. So the time complexity is  $O(nk^2) \Rightarrow O(n)$ .

Due to space constraints, the running efficiency comparisons can be found in the supplementary materials (SM).

## 4. Experiments

In this section, we conduct extensive experiments on open-sourced datasets to evaluate the performance of our proposal. The experimental settings are as follows.

**Datasets:** For evaluating single-channel MRI, two datasets were employed to assess the clinical efficacy of the

---

### Algorithm 1 Scanning Integration

---

**Input:**  $v(x)$ , a function with shape [sampling points ( $n$ ), dimension ( $D$ )]

**Parameters:**  $A$ , an evolution parameter;  $\Delta$ , a timescale parameter;  $B$  and  $C$ , projection parameters

**Linear Projection Layer:** Linear( $\cdot$ )

**Output:**  $v'(x) = (\mathcal{K}(v))(x)$ , a function with shape [sampling points ( $n$ ), dimension ( $D$ )]

1:  $\Delta, B, C = \text{Linear}(v), \text{Linear}(v), \text{Linear}(v)$ .  $\underline{\underline{//O(3nD^2)}}$

2:  $\bar{\mathbf{A}} = \exp(\Delta A)$ .  $\underline{\underline{//O(nD^2)}}$

3:  $\bar{\mathbf{B}} = (\Delta A)^{-1} (\exp(\Delta A) - \mathbf{I}) \cdot \Delta B$ .  $\underline{\underline{//O(3nD^2)}}$

4:  $v'(x) = SSM(\bar{\mathbf{A}}, \bar{\mathbf{B}}, C)(v(X))$ ,  $X$  is a discrete sequence that contains  $x_1, x_2, \dots, x_N$ .  $\underline{\underline{//O(n)}}$

4.1:  $h(x_k) = \bar{\mathbf{A}}h(x_{k-1}) + \bar{\mathbf{B}}v(x_k)$

4.2:  $v'(x_k) = Ch(x_k)$

4.3:  $v'(x) = [v'(x_1), v'(x_2), \dots, v'(x_N)]$

**return**  $(\mathcal{K}(v))(x) = v'(x)$

---

proposed method: the IXI dataset\*, comprising 578 registered T2 images, and the knee fastMRI dataset[57].

For multi-channel MRI evaluation, a brain dataset [1] is employed, which includes fully sampled multi-coil images from five volunteers. Coil sensitivity maps are estimated using ESPIRiT [45] from the central k-space region of each slice. More details about the dataset can be found in the SM.

**Competing methods:** Six previously reported state-of-the-art methods are selected, comprising of three deep unfolding methods including HQS-Net [53], H-DSSLR [37], and PGIUN [17], as well as three purely deep learning-based methods including Unet [57], SwinIR [29] and U-Mamba [40]. For all competitors, the parameter configurations suggested by the original authors are employed for a fair comparison.

For further details on the experimental implementation, please refer to the SM.

### 4.1. Results and Analysis

**Signal Recovery for Single-Coil and Multi-Coil MRI.** Table 1 presents the average results of 10 runs for IXI and fastMRI datasets, in which the best and second-best performers are highlighted in different colors. Our first observation is that LMO consistently ranks first. Yet with some marginal lead in partial cells, we believe that the rolling victory is sufficient to demonstrate our advantages over others. Specifically, under a  $\times 4$  radial acceleration, LMO achieves a PSNR result of 48.16dB, outperforming others by a large margin. In contrast, Unet, SwinIR, and U-Mamba exhibit

\*<https://brain-development.org/ixi-dataset/>

AR	Methods	IXI (Brain)						fastMRI (Knee)					
		Random		Radial		Equispaced		Random		Radial		Equispaced	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
×4	Unet	31.28	0.954	34.03	0.935	30.22	0.946	27.96	0.811	28.69	0.830	27.26	0.780
	SwinIR	32.51	0.962	35.57	0.940	31.30	0.951	28.45	0.822	29.50	0.840	28.12	0.794
	U-Mamba	32.10	0.958	34.03	0.930	30.92	0.945	28.17	0.813	28.93	0.833	27.57	0.782
	HQS-Net	32.49	0.948	35.14	0.969	30.34	0.942	28.57	0.819	29.32	0.839	27.82	0.787
	H-DSLRL	36.01	0.982	45.31	0.994	33.65	0.968	29.04	0.834	30.23	0.866	28.25	0.799
	PGIUN	37.98	0.985	47.09	0.994	35.51	0.978	30.02	0.850	30.98	0.876	28.55	0.809
	LMO (Ours)	39.14	0.986	48.16	0.996	35.70	0.980	30.17	0.853	31.11	0.883	28.65	0.832
×8	Unet	29.06	0.932	29.86	0.890	27.91	0.922	26.38	0.754	26.42	0.723	26.21	0.745
	SwinIR	30.13	0.947	29.79	0.891	28.54	0.925	27.41	0.768	27.88	0.742	27.26	0.752
	U-Mamba	29.88	0.935	29.31	0.879	28.06	0.923	26.89	0.758	27.75	0.765	26.89	0.748
	HQS-Net	28.70	0.923	29.02	0.921	27.40	0.909	26.64	0.755	27.82	0.764	26.42	0.747
	H-DSLRL	31.96	0.955	34.50	0.971	29.80	0.938	27.04	0.762	27.28	0.741	26.84	0.752
	PGIUN	34.15	0.970	36.21	0.980	31.74	0.957	27.83	0.783	28.01	0.791	27.41	0.768
	LMO (Ours)	34.26	0.971	36.53	0.983	31.80	0.960	27.88	0.808	28.29	0.837	27.51	0.787

Table 1. The numerical results under three masks as well as ×4 and ×8 ARs.

Methods	PSNR	SSIM	FLOPs (G)	Params (M)
HQS-Net	38.72▼19.6%	0.983▼1.3%	116.68▼0.0%	1.78▼0.0%
H-DSLRL	46.19▼4.1%	0.994▼0.2%	295.52▲153.8%	1.79▲0.6%
PGIUN	47.33▼1.7%	0.995▼0.1%	115.38▼0.9%	1.78▼0.0%
LMO (Ours)	48.16	0.996	116.44	1.78

Table 2. Comparisons of HQS-Net, H-DSLRL, PGIUN, and LMO under comparable parameters (IXI-Radial-×4).

AR	Methods	Multi-coil Reconstructions	
		PSNR	SSIM
×6	Unet	32.37▼22.0%	0.922▼7.2%
	SwinIR	32.28▼22.2%	0.887▼10.8%
	U-Mamba	32.80▼21.0%	0.931▼6.3%
	HQS-Net	35.51▼14.5%	0.947▼4.7%
	H-DSLRL	38.09▼8.3%	0.963▼3.1%
	PGIUN	40.33▼2.9%	0.989▼0.5%
	LMO (Ours)	41.52	0.994

Table 3. The numerical results on multi-coil reconstructions.

relatively unsatisfactory performance, possibly stemming from their disregard for data consistency in network design. Overall, these results verify the positive roles of our employed mechanisms, such as scanning integration and convolution integration. Taking Fig. 1 into account, while HQS-Net and PGIUN lag behind our model consistently in these cases, they enjoy relatively fewer parameters. Besides, while H-DSLRL outperforms others except PGIUN and LMO, it suffers from the heaviest model burden. For better fairness, experiments with comparable parameters are also conducted. In practice, we change the number of internal operations, e.g., unrolling stages for PGIUN and convolutions for H-DSLRL, until a similar quantity of parameters is reached. The results are shown in Table 2. Evidently, LMO again scores the highest, leading PGIUN, H-DSLRL,

SG	LMO		PGIUN		H-DSLRL	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
×4 to ×4	48.16	0.996	47.09	0.994	45.31	0.994
×4 to ×6	43.76	0.989	41.06	0.982	29.80	0.835
×4 to ×8	30.25	0.860	28.31	0.826	24.10	0.746
×8 to ×4	<b>43.50</b>	0.988	29.80	0.941	29.20	0.754
×8 to ×6	<b>41.98</b>	0.985	32.58	0.963	31.50	0.850
×8 to ×8	<b>36.53</b>	0.983	36.21	0.980	34.50	0.971

Table 4. Generalization comparisons across various scales. Bold and italic denote different performance trends (IXI-Radial).

and HQS-Net by 0.83dB, 1.97dB, and 9.44dB, respectively.

Comparisons of target contrast and the error maps concerned are given in Fig. 5. The texture of the error serves as an indicator of the efficacy of the restoration, with a smoother and bluer surface indicating a better reconstruction. As evidenced, given the input that suffers substantial aliasing artifacts and lacks anatomical details, most of the selected methods are qualified for recovering a clear MR image. Yet comparatively, it is evident that from the error maps, LMO is capable of recovering more details and exhibiting fewer visible artifacts, surpassing all the other competitors. The improvement in visual quality can again be attributed to the employment of global scanning integration and local convolution integration.

To transition from a single-coil to a multi-coil scenario, it is necessary to use coil sensitivity maps that weight the signals from 12 different positions, reducing artifacts and noise [1]. Given this operation, the proposed LMO is readily available for multi-coil reconstruction, with the representative results from ×6 accelerate rate of random masks given in Table 3. As shown, LMO again consistently outperforms other competing models by a large margin. For refer-

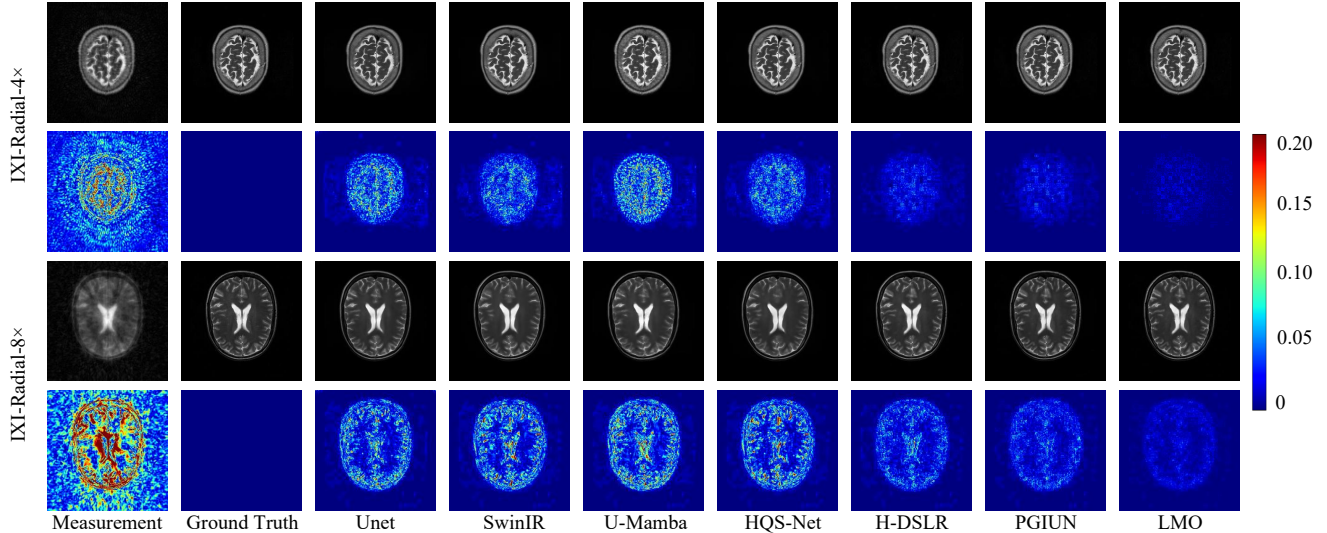


Figure 5. Visual results and error maps on IXI dataset with radial mask  $\times 4$  and radial mask  $\times 8$  ARs.

ence, the SM include a visual comparison of multi-coil reconstructions performed under identical conditions. As evidenced, LMO has delivered strong stability and reliability in both single- and multi-coil MRI acceleration. We attribute this to the neural operator architecture that performs mapping between bandlimited function spaces, as they maintain consistency with the physical sampling constraints of k-space, while the neural operator provides generalization to handle different sampling constraints.

**Generalization Analysis.** Taking PGIUN and H-DSLRL as competing baselines, the experimental results of scale generalization (SG) with the radial mask are given in Table 4. Representatively, the symbol ‘ $\times 4$  to  $\times 8$ ’ denotes the case of training at a  $\times 4$  scale and testing at a  $\times 8$  scale. Notably, LMO is the only algorithm that shows improved performance when trained at a scale of 8 and generalized to lower scales. We attribute this to the advantage of LMO which mappings between function spaces, rather than the typical instance-to-instance mappings. This also empirically demonstrates that our proposal enjoys *continuous-discrete equivalence* with the underlying continuous operator.

**Ablation Experiments.** To affirm the benefit of combining both global and local information, Table 5 discloses results pertaining to the ablation of Scanning and Convolution integration. Based on our final configuration, the scanning integration is replaced with convolution integral, thus obtaining baseline 1 with pure local terms. Similarly, we can also obtain baseline 3, which holds pure scanning integration. To ensure a fair comparison, the parameters of all three competitors are configured at the same level. As expected, due to the lack of partial information, the performance of pure convolution integration (CI) and pure scanning integration (SI) drops by 13.48dB and 7.71dB, respectively. In

Config	PSNR	SSIM	Params (M)	FLOPs (G)
1. Pure CI	34.68 $\nabla$ 28.0%	0.807 $\nabla$ 19.0%	1.74 $\nabla$ 2.2%	89.66 $\nabla$ 23.0%
2. CI+SI *	48.16	0.996	1.78	116.44
3. Pure SI	40.45 $\nabla$ 16.0%	0.896 $\nabla$ 10.0%	1.81 $\blacktriangle$ 1.7%	138.56 $\blacktriangle$ 19.0%

Table 5. The ablation results by using different forms of kernel integrals. \* indicates our default choice.

terms of efficiency, while baseline 1 enjoys fewer Params and FLOPs in this case, it cannot be ensured in practice since more network layers are often needed for a larger receptive field. Due to space limitations, we have placed more ablation experiments in SM.

## 5. Conclusion

In conclusion, this study presents a novel Linear Mamba Operator for MRI reconstruction (LMO), enjoying both theoretical support and empirical evidence, hence effectively addressing the limitations, such as weak consistency or poor interpretability, of traditional or deep learning methods. Specifically, by integrating the scanning and convolution integral layers, our proposal ensures both global and local feature extraction, significantly enhancing the reconstruction quality. More importantly, due to the nature of learning mappings between high-dimensional continuous functions, LMO enjoys a high generalization in cases of different acceleration rates, which is a challenging issue for most current methods. Extensive experiments conducted on single- and multi-coil MRI recovery both demonstrate that our proposed method outperforms existing state-of-the-art methods in both quantitative and qualitative metrics. Future work will focus on exploring the applicability of LMO to other medical imaging modalities, thus broadening its impact on the medical field.

## Acknowledgements

This work was supported in part by the Key Program of Natural Science Foundation of Zhejiang Province under Grant LZ24F030012, the National Natural Science Foundation of China under Grant 62276232, Zhejiang Students' Technology and Innovation Program under Grant 2024R403B071, and the China Postdoctoral Science Foundation under grant ZY24211190002.

## References

- [1] Hemant K Aggarwal, Merry P Mani, and Mathews Jacob. Modl: Model-based deep learning architecture for inverse problems. *IEEE Transactions on Medical Imaging*, 38(2): 394–405, 2018. 6, 7
- [2] Francesca Bartolucci, Emmanuel de Bézenac, Bogdan Raonić, Roberto Molinaro, Siddhartha Mishra, and Rima Alaifari. Are neural operators really neural operators? frame theory meets operator learning. *SAM Research Report*, 2023, 2023. 3
- [3] Francesca Bartolucci, Emmanuel de Bézenac, Bogdan Raonic, Roberto Molinaro, Siddhartha Mishra, and Rima Alaifari. Representation equivalent neural operators: a framework for alias-free operator learning. *Advances in Neural Information Processing Systems*, 36, 2024. 3
- [4] Jiacheng Chen, Jiawei Jiang, Fei Wu, and Jianwei Zheng. Null space matters: Range-null decomposition for consistent multi-contrast mri reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1081–1090, 2024. 2
- [5] Zhile Chen, Yuhui Quan, and Hui Ji. Unsupervised deep unrolling networks for phase unwrapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25182–25192, 2024. 2
- [6] Yuchao Feng, Honghui Xu, Jiawei Jiang, Hao Liu, and Jianwei Zheng. Icif-net: Intra-scale cross-interaction and inter-scale feature fusion network for bitemporal remote sensing images change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022. 2
- [7] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023. 6
- [8] Albert Gu, Karan Goel, and Christopher Re. Efficiently modeling long sequences with structured state spaces. In *International Conference on Learning Representations*, 2021. 5
- [9] Yubin Gu, Honghui Xu, Yueqian Quan, Wanjun Chen, and Jianwei Zheng. Orsi salient object detection via bidimensional attention and full-stage semantic guidance. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–13, 2023. 2
- [10] Yubin Gu, Siting Chen, Xiaoshuai Sun, Jiayi Ji, Yiyi Zhou, and Rongrong Ji. Optical remote sensing image salient object detection via bidirectional cross-attention and attention restoration. *Pattern Recognition*, page 111478, 2025. 1
- [11] Pengfei Guo, Yiqun Mei, Jinyuan Zhou, Shanshan Jiang, and Vishal M Patel. Reconformer: Accelerated mri reconstruction using recurrent transformer. *IEEE Transactions on Medical Imaging*, 2023. 1
- [12] Zhongkai Hao, Zhengyi Wang, Hang Su, Chengyang Ying, Yinpeng Dong, Songming Liu, Ze Cheng, Jian Song, and Jun Zhu. Gnot: A general neural operator transformer for operator learning. In *International Conference on Machine Learning*, pages 12556–12569. PMLR, 2023. 4
- [13] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989. 3
- [14] Amrit Kumar Jethi, Balamurali Murugesan, Keerthi Ram, and Mohanasankar Sivaprakasam. Dual-encoder-unet for fast mri reconstruction. In *2020 IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops)*, pages 1–4. IEEE, 2020. 1, 2
- [15] Zexin Ji, Beiji Zou, Xiaoyan Kui, Pierre Vera, and Su Ruan. Deform-mamba network for mri super-resolution. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 242–252. Springer, 2024. 1
- [16] Jiawei Jiang, Jiacheng Chen, Honghui Xu, Yuchao Feng, and Jianwei Zheng. Ga-hqs: Mri reconstruction via a generically accelerated unfolding approach. In *2023 IEEE International Conference on Multimedia and Expo (ICME)*, pages 186–191. IEEE, 2023. 2
- [17] Jiawei Jiang, Zihan He, Yueqian Quan, Jie Wu, and Jianwei Zheng. Pgiun: Physics-guided implicit unrolling network for accelerated mri. *IEEE Transactions on Computational Imaging*, 2024. 2, 6
- [18] Min Jiang, Fuhao Zhai, and Jun Kong. A novel deep learning model ddu-net using edge features to enhance brain tumor segmentation on mr images. *Artificial Intelligence in Medicine*, 121:102180, 2021. 2
- [19] Samira Kabri, Tim Roith, Daniel Tenbrinck, and Martin Burger. Resolution-invariant image classification based on fourier neural operators. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 236–249. Springer, 2023. 3, 4
- [20] Yoshinobu Kawahara. Dynamic mode decomposition with reproducing kernels for koopman spectral analysis. *Advances in neural information processing systems*, 29, 2016. 5
- [21] Ziwen Ke, Wenqi Huang, Zhuo-Xu Cui, Jing Cheng, Sen Jia, Haifeng Wang, Xin Liu, Hairong Zheng, Leslie Ying, Yanjie Zhu, et al. Learned low-rank priors in dynamic mr imaging. *IEEE Transactions on Medical Imaging*, 40(12):3698–3710, 2021. 2
- [22] Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Aziz-zadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023. 2
- [23] Peng Li, Wengu Chen, and Michael K Ng. Compressive total variation for image reconstruction and restoration. *Computers & Mathematics with Applications*, 80(5):874–893, 2020. 1

- [24] Xiaodong Li, Yunkai Bao, Zhuheng Li, Peihong Teng, Lina Ma, Hua Zhang, Guifeng Liu, and Zhenxin Wang. Employing antagonistic cxc motif chemokine receptor 4 antagonistic peptide functionalized nagdf4 nanodots for magnetic resonance imaging-guided biotherapy of breast cancer. *Scientific Reports*, 14(1):15764, 2024. 1
- [25] Zongyi Li, Nikola Borislavov Kovachki, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, Anima Anandkumar, et al. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2020. 2, 3, 4, 5
- [26] Zijie Li, Kazem Meidani, and Amir Barati Farimani. Transformer for partial differential equations’ operator learning. *Transactions on Machine Learning Research*, 2022. 2, 3, 4
- [27] Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, and Anima Anandkumar. Fourier neural operator with learned deformations for pdes on general geometries. *Journal of Machine Learning Research*, 24(388):1–26, 2023. 4
- [28] Zongyi Li, Nikola Kovachki, Chris Choy, Boyi Li, Jean Kossaifi, Shourya Otta, Mohammad Amin Nabian, Maximilian Stadler, Christian Hundt, Kamyar Azizzadenesheli, et al. Geometry-informed neural operator for large-scale 3d pdes. *Advances in Neural Information Processing Systems*, 36, 2024. 5
- [29] Jingyun Liang, Jiezhong Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844, 2021. 2, 6
- [30] Wei-Jie Liao, Bo-Lin Sun, Jia-Bao Wu, Ning Zhang, Rong-Ping Zhou, Shan-Hu Huang, Zhi-Li Liu, and Jia-Ming Liu. Role of magnetic resonance imaging features in diagnosing and localization of disc rupture related to cervical spinal cord injury without radiographic abnormalities. *Spinal Cord*, 61(6):323–329, 2023. 1
- [31] Jingshuai Liu, Chen Qin, and Mehrdad Yaghoobi. High-fidelity mri reconstruction using adaptive spatial attention selection and deep data consistency prior. *IEEE Transactions on Computational Imaging*, 9:298–313, 2023. 1
- [32] Ning Liu, Siavash Jafarzadeh, and Yue Yu. Domain agnostic fourier neural operators. *Advances in Neural Information Processing Systems*, 36, 2024. 4
- [33] Miguel Liu-Schiaffini, Julius Berner, Boris Bonev, Thorsten Kurth, Kamyar Azizzadenesheli, and Anima Anandkumar. Neural operators with localized integral and differential kernels. In *Forty-first International Conference on Machine Learning*, 2024. 5
- [34] Jun Ma, Feifei Li, and Bo Wang. U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722*, 2024. 1, 2
- [35] Marko Panić, Jan Aelterman, Vladimir Crnojević, and Aleksandra Pižurica. Sparse recovery in magnetic resonance imaging with a markov random field prior. *IEEE Transactions on Medical Imaging*, 36(10):2104–2115, 2017. 1
- [36] Jiangjun Peng, Yao Wang, Hongying Zhang, Jianjun Wang, and Deyu Meng. Exact decomposition of joint low rankness and local smoothness plus sparse matrices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5766–5781, 2022. 2
- [37] Aniket Pramanik, Hemant Kumar Aggarwal, and Mathews Jacob. Deep generalization of structured low-rank algorithms (deep-slr). *IEEE Transactions on Medical Imaging*, 39(12):4186–4197, 2020. 6
- [38] Aniket Pramanik, Hemant Kumar Aggarwal, and Mathews Jacob. Deep generalization of structured low-rank algorithms (deep-slr). *IEEE Transactions on Medical Imaging*, 39(12):4186–4197, 2020. 3
- [39] Bogdan Raonic, Roberto Molinaro, Tim De Ryck, Tobias Rohner, Francesca Bartolucci, Rima Alaifari, Siddhartha Mishra, and Emmanuel de Bézenac. Convolutional neural operators for robust and accurate learning of pdes. *Advances in Neural Information Processing Systems*, 36, 2024. 2, 3, 5
- [40] Jiacheng Ruan and Suncheng Xiang. Vm-unet: Vision mamba unet for medical image segmentation. *arXiv preprint arXiv:2402.02491*, 2024. 6
- [41] Mariam Saad, Isaac V Manzanera Esteve, Adam G Evans, Huseyin Karagoz, Tigran Kesayan, Krista Brooks-Horrar, Saikat Sengupta, Ryan Robison, Brian Johnson, Richard Dortch, et al. Preoperative visualization of the greater occipital nerve with magnetic resonance imaging in candidates for occipital nerve decompression for headaches. *Scientific Reports*, 14(1):15248, 2024. 1
- [42] Yousef Saad. Analysis of some krylov subspace approximations to the matrix exponential operator. *SIAM Journal on Numerical Analysis*, 29(1):209–228, 1992. 5
- [43] Mrinmoy Sandilya and SR Nirmala. Compressed sensing mri reconstruction using convolutional dictionary learning and laplacian prior. In *IOT with Smart Systems: Proceedings of ICTIS 2021, Volume 2*, pages 661–669. Springer, 2022. 1
- [44] Claude Elwood Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949. 3
- [45] Martin Uecker, Peng Lai, Mark J Murphy, Patrick Virtue, Michael Elad, John M Pauly, Shreyas S Vasanawala, and Michael Lustig. Espirit—an eigenvalue approach to autocalibrating parallel mri: where sense meets grappa. *Magnetic Resonance in Medicine*, 71(3):990–1001, 2014. 6
- [46] Kaidong Wang, Xiuwu Liao, Jun Li, Deyu Meng, and Yao Wang. Hyperspectral image super-resolution via knowledge-driven deep unrolling and transformer embedded convolutional recurrent neural network. *IEEE Transactions on Image Processing*, 2023. 2
- [47] Yueze Wang, Yanwei Pang, and Chuan Tong. Dsmenet: Detail and structure mutually enhancing network for under-sampled mri reconstruction. *Computers in Biology and Medicine*, 154:106204, 2023. 1
- [48] Min Wei and Xuesong Zhang. Super-resolution neural operator. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18247–18256, 2023. 3, 4
- [49] Yu-Hui Wen, Lin Gao, Hongbo Fu, Fang-Lue Zhang, Shihong Xia, and Yong-Jin Liu. 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, 2022. 2

- [50] Ken CL Wong, Hongzhi Wang, and Tanveer Syeda-Mahmood. Fnoseg3d: Resolution-robust 3d image segmentation with fourier neural operator. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2023. 3
- [51] Jie Wu, Yuchao Feng, Honghui Xu, Chuanmeng Zhu, and Jianwei Zheng. Syformer: structure-guided synergism transformer for large-portion image inpainting. In *Proceedings of the AAAI conference on artificial intelligence*, pages 6021–6029, 2024. 2
- [52] Jingfen Xie, Jian Zhang, Yongbing Zhang, and Xiangyang Ji. Puert: Probabilistic under-sampling and explicable reconstruction network for cs-mri. *IEEE Journal of Selected Topics in Signal Processing*, 16(4):737–749, 2022. 2
- [53] Bingyu Xin, Timothy Phan, Leon Axel, and Dimitris Metaxas. Learned half-quadratic splitting network for mr image reconstruction. In *International Conference on Medical Imaging with Deep Learning*, pages 1403–1412. PMLR, 2022. 2, 6
- [54] Honghui Xu, Chuangjie Fang, Yilin Ge, Yubin Gu, and Jianwei Zheng. Cascade-transform-based tensor nuclear norm for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 2024. 1
- [55] Rui Xu, Shu Yang, Yihui Wang, Bo Du, and Hao Chen. A survey on vision mamba: Models, applications and challenges. *arXiv preprint arXiv:2404.18861*, 2024. 2, 5
- [56] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5728–5739, 2022. 2
- [57] Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, et al. fastmri: An open dataset and benchmarks for accelerated mri. *arXiv preprint arXiv:1811.08839*, 2018. 6
- [58] Mingli Zhang, Mingyan Zhang, Fan Zhang, Ahmad Chad-dad, and Alan Evans. Robust brain mr image compressive sensing via re-weighted total variation and sparse regression. *Magnetic Resonance Imaging*, 85:271–286, 2022. 2
- [59] Xuemin Zhang, Jianwei Ma, and Hao Zhang. Curvature-regularized manifold for seismic data interpolation. *Geophysics*, 88(1):WA37–WA53, 2023. 1
- [60] Jianwei Zheng, Wei Li, Ni Xu, Junwei Zhu, and Xiaoqin Zhang. Alias-free mamba neural operator. *Advances in Neural Information Processing Systems*, 37:52962–52995, 2024. 3, 4
- [61] Shenglong Zhou and Geoffrey Ye Li. Federated learning via inexact admm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8):9699–9708, 2023. 2
- [62] Lianghui Zhu, Bencheng Liao, Qian Zhang, Xinlong Wang, Wenyu Liu, and Xinggang Wang. Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint arXiv:2401.09417*, 2024. 2, 4
- [63] Jing Zou, Lanqing Liu, Qi Chen, Shujun Wang, Xiaohan Xing, and Jing Qin. Mmr-mamba: Multi-contrast mri re-construction with mamba and spatial-frequency information fusion. *arXiv preprint arXiv:2406.18950*, 2024. 1