

# HiFi-Mamba: Dual-Stream $\omega$ -Laplacian Enhanced Mamba for High-Fidelity MRI Reconstruction

Hongli Chen<sup>1,\*</sup>, Pengcheng Fang<sup>2,\*</sup>, Yuxia Chen<sup>4</sup>, Yingxuan Ren<sup>5</sup>, Jing Hao<sup>6</sup>, Fangfang Tang<sup>1</sup>, Xiaohao Cai<sup>2</sup>, Shanshan Shan<sup>3,†</sup>, Feng Liu<sup>1</sup>

<sup>1</sup>The University of Queensland

<sup>2</sup>University of Southampton

<sup>3</sup>Soochow University

<sup>4</sup>Chengdu Jiaozi Financial Holding Group Co., Ltd

<sup>5</sup>National University of Singapore

<sup>6</sup>The University of Hong Kong

hongli.chen@uq.edu.au, p.fang@soton.ac.uk, craveuchen@gmail.com, e1664832@u.nus.edu, jinghao@connect.hku.hk, f.tang1@uq.edu.au, x.cai@soton.ac.uk, ssshanshan@suda.edu.cn, feng@eecs.uq.edu.au

## Abstract

Reconstructing high-fidelity MR images from undersampled k-space data remains a challenging problem in MRI. While Mamba variants for vision tasks offer promising long-range modeling capabilities with linear-time complexity, their direct application to MRI reconstruction inherits two key limitations: (1) insensitivity to high-frequency anatomical details; and (2) reliance on redundant multi-directional scanning. To address these limitations, we introduce High-Fidelity Mamba (HiFi-Mamba), a novel dual-stream Mamba-based architecture comprising stacked  $\omega$ -Laplacian (WL) and HiFi-Mamba blocks. Specifically, the WL block performs fidelity-preserving spectral decoupling, producing complementary low- and high-frequency streams. This separation enables the HiFi-Mamba block to focus on low-frequency structures, enhancing global feature modeling. Concurrently, the HiFi-Mamba block selectively integrates high-frequency features through adaptive state-space modulation, preserving comprehensive spectral details. To eliminate the scanning redundancy, the HiFi-Mamba block adopts a streamlined unidirectional traversal strategy that preserves long-range modeling capability with improved computational efficiency. Extensive experiments on standard MRI reconstruction benchmarks demonstrate that HiFi-Mamba consistently outperforms state-of-the-art CNN-based, Transformer-based, and other Mamba-based models in reconstruction accuracy while maintaining a compact and efficient model design.

## Code —

<https://github.com/HongliChen-Ali/HiFi-Mamba.git>

## Introduction

Magnetic Resonance Imaging (MRI) is a clinically indispensable modality due to its non-invasive nature and excellent soft-tissue contrast (Jerban, Chang, and Du 2020;

\*These authors contributed equally.

†Corresponding author.

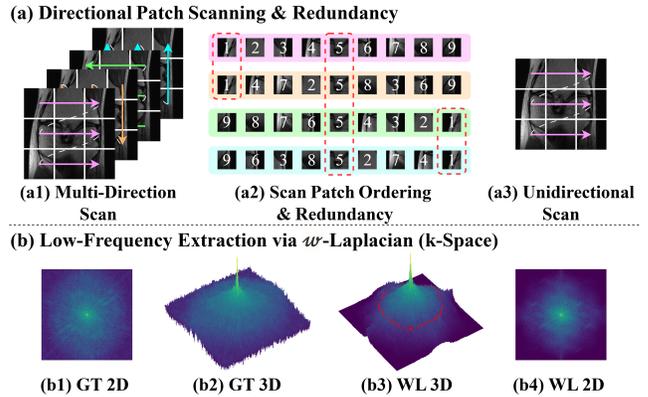


Figure 1: Illustration of scanning and decoupling. (a) Scanning strategies. Multi-directional scanning introduces redundancy, while the unidirectional approach avoids repeated access. Colors denote scan orders; red dashed boxes highlight redundant regions. (b) Visualization of k-space before and after the  $\omega$ -Laplacian decomposition. Subfigures (b3) and (b4) show only the output branch retained for Mamba. The red circle marks the theoretical boundary between low- and high-frequency regions in k-space. This retained branch exhibits a cleaner, concentrated low-frequency spectrum and is better aligned with Mamba’s global modeling needs.

Varela-Mattatall and D’Arcy 2017). However, a primary limitation lies in its long acquisition time, which can cause patient discomfort and increase the risk of motion artifacts (Hammernik et al. 2018; Knoll et al. 2020). To accelerate scans, modern protocols commonly adopt undersampling in the frequency domain. While this reduces scan time, it violates the Nyquist-Shannon sampling criterion (Ye 2019; Zbontar et al. 2018), leading to aliasing artifacts and degraded image quality. Addressing this challenge requires advanced reconstruction methods capable of recovering high-fidelity images from incomplete k-space data. Tra-

ditional compressed sensing (CS) methods (Lustig, Donoho, and Pauly 2007) exploit sparsity priors in transform domains (e.g., wavelets), but they require extensive hyperparameter tuning and often lack robustness to variations in sampling patterns or noise conditions.

Recent deep learning frameworks have substantially advanced MRI reconstruction by leveraging data-driven priors from large-scale datasets (Zhou et al. 2018). Convolutional neural networks (CNNs) have demonstrated strong performance by modeling hierarchical and localized anatomical features (Schlemper et al. 2018; Qin et al. 2019; Chen et al. 2024). Model-based CNNs further improve reconstruction quality by integrating the MRI forward model and enforcing data consistency (Hammernik et al. 2018; Aggarwal, Mani, and Jacob 2019; Shan et al. 2024). However, the inherent locality of CNNs limits their capacity to capture long-range dependencies, which are crucial for reconstructing global anatomical structures, especially under highly undersampled conditions.

Transformer-based architectures model global dependencies through self-attention by computing pairwise correlations across all spatial tokens (Vaswani et al. 2017; Fu et al. 2025). This capacity has shown promise in MRI reconstruction (Guo et al. 2023; Wu et al. 2023) by enabling global context modeling to restore anatomical structures. Nonetheless, standard Transformers incur quadratic complexity, posing challenges for high-resolution MRI applications. To improve efficiency, variants such as the Swin Transformer (Liu et al. 2021) employ hierarchical designs and shifted window-based attention to restrict computations to local neighborhoods. While computationally efficient, these approaches inherently constrain the receptive field, reigniting the fundamental trade-off between computational efficiency and global modeling capability, leaving a critical gap for a more holistic solution.

Structured State Space Models (SSMs) (Gu et al. 2022) have recently emerged as promising alternatives to self-attention for global context modeling, offering linear-time scalability with strong sequence modeling capacity. Among them, Mamba (Gu, Dao et al. 2023) employs input-independent state transitions and efficient gating mechanisms, enabling expressive long-range interactions with reduced computational complexity. While initially proposed for language tasks, Mamba is rapidly emerging as a powerful alternative in visual domains such as image classification and restoration (Guo et al. 2024; Liu et al. 2024).

However, a direct application of Mamba to MRI reconstruction reveals three fundamental limitations rooted in its original design. First, existing vision-specific Mamba architectures utilize multi-directional spatial scanning (Rahman et al. 2024; Zhang et al. 2024) to enhance coverage, but this introduces significant computational redundancy (see Figure 1(a)).

Second, its state-space parameters are derived independently for each spatial token through local transformations, limiting spatial awareness. This design neglects neighboring context, which is essential for modeling coherent anatomical structures in MRI. Third, existing Mamba-based variants lack sensitivity to high-frequency components (Ma, Ni, and

Chen 2024), which are critical for preserving fine anatomical details in MRI reconstruction.

To address these limitations, we propose HiFi-Mamba (short for **H**igh-**F**idelity Mamba), a novel reconstruction framework built upon an efficient, dual-stream Mamba-based architecture including:

- A *novel HiFi-Mamba block* that embodies an efficient, dual-stream architecture. It employs a unidirectional scan for efficiency and a cross-stream guidance mechanism to resolve the locality and high-frequency insensitivity inherent to Mamba.
- A *lightweight  $w$ -Laplacian block* that decomposes features into high- and low-frequency streams, enabling frequency-aware dual-stream processing in our HiFi-Mamba.
- *State-of-the-art results.* On public MRI benchmarks, HiFi-Mamba consistently outperforms leading CNN-, Transformer-, and Mamba-based baselines, while establishing a superior trade-off between reconstruction fidelity and computational efficiency.

## Related Work

**CNN-based MRI Reconstruction.** CNNs have been widely employed in MRI reconstruction for their ability to extract hierarchical features efficiently. Early models such as DeepCascade (Schlemper et al. 2018) and ISTA-Net (Zhang and Ghanem 2018) framed reconstruction as an unrolled optimization process, integrating data fidelity with learnable, network-based priors. Subsequent approaches, including KIKI-Net (Eo et al. 2018) and DuDoRNet (Zhou et al. 2020), incorporated architectural advances such as residual connections, dilated convolutions, and hybrid modeling across image and  $k$ -space domains to enhance reconstruction quality and optimization stability. However, the inherently localized receptive fields of CNNs constrain their ability to capture global anatomical context. This limitation has motivated the development of architectures with enhanced global modeling capacity.

**Transformer-based MRI Reconstruction.** Transformers have been increasingly adopted in MRI reconstruction for their ability to model long-range dependencies through global self-attention, which facilitates the preservation of anatomical structures. Early models such as SLATER (Korkmaz et al. 2022) and TTM (Gu et al. 2021) applied Transformer blocks in the image domain. DuDReTLU-Net (Wang et al. 2023) jointly models image and  $k$ -space domains through a transformer-based architecture to enhance reconstruction quality. Later variants such as SwinMR (Huang et al. 2022) and ReconFormer (Guo et al. 2023) adopted hierarchical and windowed self-attention mechanisms to improve multi-scale feature modeling and computational efficiency. Although these designs reduce computational overhead, they often rely on localized attention and staged aggregation, which may still limit the capacity for global context modeling in high-resolution MRI reconstruction.

**Structured State Space Models and Mamba.** State space models (SSMs) have recently gained attention for their

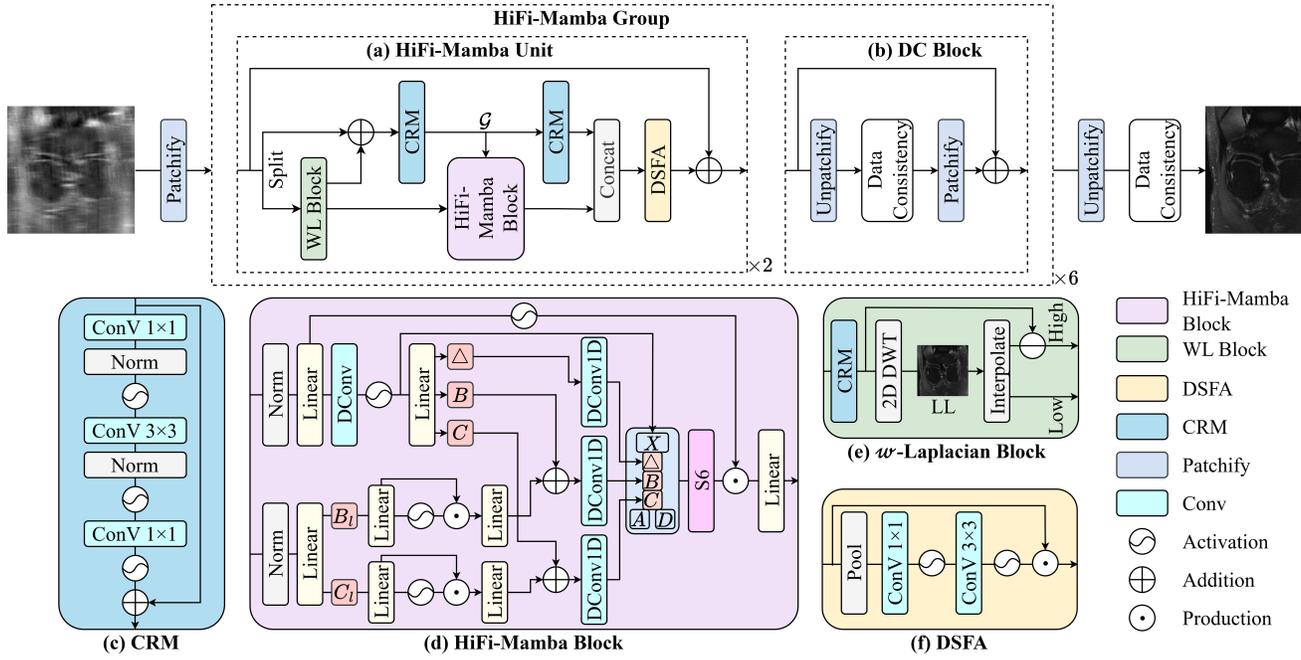


Figure 2: Overview of the proposed HiFi-Mamba architecture. (a) The HiFi-Mamba Unit splits the input into high- and low-frequency components via the  $w$ -Laplacian Block, processes them using the HiFi-Mamba Block and CRM (Condition Refinement Module), and fuses them with DSFA (Dual-Stream Fusion Attention). (b) The data consistency block. (c) CRM performs cross-resolution feature transformation. (d) The HiFi-Mamba block models frequency-aware sequences using Mamba-based token mixing. (e) The  $w$ -Laplacian block performs idelity-preserving spectral decoupling. (f) DSFA fuses dual-frequency streams with adaptive weighting.

ability to model long-range dependencies with linear complexity (Li et al. 2024). The vision-centric Mamba variant, vMamba (Liu et al. 2024), introduced a four-directional scanning strategy to enhance 2D spatial context modeling. Recent works, such as MambaMIR (Huang et al. 2025) and LMO (Li et al. 2025), have extended this paradigm for MRI reconstruction. However, these adaptations remain suboptimal in both efficiency and fidelity. Their reliance on multi-directional scanning induces considerable computational overhead due to repeated processing of spatial features—an issue exacerbated in high-resolution scenarios. Moreover, they inherit two core limitations of the original Mamba design: purely local state-space parameterization and insensitivity to high-frequency anatomical details. These constraints highlight a gap between current vision-based Mamba variants and the spectral characteristics of MRI.

### Overall Pipeline

Our proposed HiFi-Mamba network follows an unrolled optimization framework, a powerful paradigm for solving inverse problems like MRI reconstruction. The network backbone consists of  $K = 8$  cascaded HiFi-Mamba Groups. As shown in Figure 2, each Group consists of two sequential Mamba Units for feature refinement, followed by a Data Consistency (DC) block. Specifically, the input undersampled image  $I_{in} \in \mathbb{R}^{H \times W \times 2}$  is a two-channel tensor representing the real and imaginary parts of the complex-valued

MR data. The input tensor is first transformed into patch embeddings as:

$$F_{in} = \mathcal{P}(I_{in}), \quad F_{in} \in \mathbb{R}^{H/P \times W/P \times C}. \quad (1)$$

where  $\mathcal{P}(\cdot)$  denotes the patch embedding operation. The patch embeddings  $F_{in}$  are then processed sequentially through the  $K$  Groups. Within each Group, the two Mamba Units progressively refine the features by modeling both global and local dependencies via a novel asymmetric dual-stream architecture. The subsequent DC block then enforces data fidelity by incorporating the originally acquired k-space measurements.

After the final Group and its subsequent DC block, an unpatchifying layer restores the features to the full image resolution  $\mathbb{R}^{H \times W \times 2}$ . Finally, a DC block is applied in the image domain to ensure global data fidelity before producing the output reconstruction  $I_{out} \in \mathbb{R}^{H \times W \times 2}$ .

### The Mamba Unit

The Mamba Unit is the core module in each HiFi-Mamba Group, following a dual-stream, frequency-aware architecture. As shown in Figure 2(a), it adopts an asymmetric structure to adaptively process MRI-specific features.

**Stream Preparation via Frequency Decoupling.** The unit’s workflow begins by splitting the input feature map  $F_{in} \in \mathbb{R}^{H \times W \times C}$  into two feature maps,  $F_1, F_2 \in \mathbb{R}^{H \times W \times \frac{C}{2}}$ . The feature map  $F_1$  is processed by the

lightweight  $\omega$ -Laplacian (WL) block to yield a low-frequency component  $F_{\text{low}}$  and a residual high-frequency component  $F'_{\text{high}}$ , i.e.,

$$F_{\text{low}}, F'_{\text{high}} = \text{WL}(F_1). \quad (2)$$

The second feature map,  $F_2$ , is then fused with  $F'_{\text{high}}$  via element-wise addition to form the final high-frequency stream for parallel processing, i.e.,

$$F_{\text{high}} = F_2 + F'_{\text{high}}. \quad (3)$$

**Asymmetric Parallel Processing.** The high-frequency feature map  $F_{\text{high}}$  is first processed by a dedicated Condition Refinement Module (CRM; see Figure 2(c)) to extract an anatomical guidance feature:

$$\mathcal{G} = \text{CRM}(F_{\text{high}}). \quad (4)$$

Here,  $\mathcal{G}$  captures spatial high-frequency structures and serves as a guidance prior for modulating the low-frequency stream. To further enhance the high-frequency representation,  $\mathcal{G}$  is subsequently refined by a second CRM:

$$\tilde{F}_{\text{high}} = \text{CRM}(\mathcal{G}). \quad (5)$$

Concurrently, the low-frequency feature map  $F_{\text{low}}$  is processed by our novel HiFi-Mamba Block. We denote this operation as  $\mathcal{H}(\cdot)$ , which performs long-range dependency modeling under the explicit guidance of the anatomical map  $\mathcal{G}$ , i.e.,

$$\tilde{F}_{\text{low}} = \mathcal{H}(F_{\text{low}} | \mathcal{G}). \quad (6)$$

This cross-stream guidance, detailed in the HiFi-Mamba Block section, infuses essential high-frequency cues into the global context modeling.

**Dual-Stream Fusion.** The two enhanced feature maps,  $\tilde{F}_{\text{low}}$  and  $\tilde{F}_{\text{high}}$ , are concatenated and fused by a Dual-Stream Fusion Attention (DSFA) module (Figure 2(f)) to a fused feature map,  $F_{\text{fused}}$ :

$$F_{\text{fused}} = \text{DSFA}(\text{concat}([\tilde{F}_{\text{low}}, \tilde{F}_{\text{high}}])). \quad (7)$$

The final output of the unit,  $F_{\text{out}}$ , is then obtained by applying a residual connection with the unit's original input,  $F_{\text{in}}$ :

$$F_{\text{out}} = F_{\text{fused}} + F_{\text{in}}. \quad (8)$$

### $\omega$ -Laplacian Block

To enable frequency-aware dual-stream processing, we implement the  $\omega$ -Laplacian block (Figure 2(e)) to perform fidelity-preserving spectral decoupling. This operation serves two purposes: (1) providing a dedicated low-frequency input for the Mamba stream (see Figure 1(b)), and (2) extracting high-frequency features into a parallel stream for fine-detail enhancement.

While the Laplacian pyramid (Burt and Adelson 1987) offers residual-based multiscale representations, its decomposition is resolution-oriented rather than frequency-structured. The resulting low-frequency component is a blurred, downsampled approximation without explicit spectral semantics, limiting its utility for preserving anatomical

context in MRI. To overcome this limitation, we replace the resolution-based hierarchy with a wavelet-based formulation that enables structured and reversible frequency separation.

The  $\omega$ -Laplacian block begins by refining the input feature map  $F_1 \in \mathbb{R}^{H \times W \times \frac{C}{2}}$  using a CRM to enhance local feature representation, yielding:

$$F'_1 = \text{CRM}(F_1). \quad (9)$$

A channel-wise 2D discrete wavelet transform (DWT) is then applied to  $F'_1$  to obtain the four standard subbands:

$$\text{DWT}(F'_1) = \{LL, LH, HL, HH\}. \quad (10)$$

To avoid potential information loss and basis dependency when using the high-frequency subbands directly, a smoothed low-frequency base is first formed by upsampling the  $LL$  subband, i.e.,

$$F_{\text{low}} = \text{Upsample}(LL). \quad (11)$$

The complementary high-frequency component is then defined as the residual between the refined map and this base:

$$F_{\text{high}} = F'_1 - F_{\text{low}}. \quad (12)$$

This decoupling enables our specialized dual-stream processing: the low-frequency stream routes  $F_{\text{low}}$  to the HiFi-Mamba block for long-range anatomical modeling, while the high-frequency stream processes  $F_{\text{high}}$  for targeted enhancement.

### HiFi-Mamba Block

As a central component of our dual-stream architecture, the HiFi-Mamba block processes the low-frequency feature map,  $F_{\text{low}}$ , conditioned by the anatomical guidance map,  $\mathcal{G}$ . This design directly addresses the core limitations of standard Mamba for MRI reconstruction by introducing two key modifications: a cross-frequency guidance mechanism and a spatially-aware parameter refinement process. The block's structure is illustrated in Figure 2(d).

**Initial Parameter Generation.** The block first produces initial state-space parameters from the low-frequency feature map  $F_{\text{low}}$ . To embed local spatial context prior to the main selective scan, this map is projected and split into a main feature map,  $F_c$ , and a residual map,  $Z$ :

$$F_c, Z = \text{Split}(\text{Linear}(\text{Norm}(F_{\text{low}}))). \quad (13)$$

The main feature map  $F_c$  is then passed through a 2D convolution and a SiLU activation to produce a context-aware feature map,  $F_{\text{conv}}$ , i.e.,

$$F_{\text{conv}} = \text{SiLU}(\text{Conv2D}(F_c)). \quad (14)$$

After reshaping  $F_{\text{conv}}$  into a sequence  $F_s$ , a subsequent linear projection is applied, followed by a split operation, to generate the initial state projection matrices  $B, C \in \mathbb{R}^{B \times d_s \times L}$  and the timestep parameter  $\Delta \in \mathbb{R}^{B \times d_t \times L}$ , i.e.,

$$[\Delta, B, C] = \text{Split}(\text{Linear}(F_s)). \quad (15)$$

At this stage, while containing some local context from the convolution, these parameters have not yet been informed by the high-frequency stream.

**Cross-Frequency Guidance.** To address the first core limitation—Mamba’s insensitivity to high frequencies—conditioning terms are derived from the anatomical guidance map  $\mathcal{G}$ . An initial projection is applied to  $\mathcal{G}$  to produce pre-conditioned tensors  $B'_h$  and  $C'_h$ . These tensors are then processed by two independent gating mechanisms to generate the final conditioning terms  $B_h$  and  $C_h$ . The process for generating  $B_h$  from  $B'_h$  is as follows:

$$\begin{aligned} B'_{h,1}, B'_{h,2} &= \text{Split}(\text{Linear}(B'_h)), \\ B_h &= \text{GELU}(B'_{h,1}) \odot B'_{h,2}. \end{aligned} \quad (16)$$

The term  $C_h$  is generated from  $C'_h$  through an identical process but with a separate set of learned weights. These conditioning terms are subsequently integrated through element-wise addition:

$$B = B + B_h, \quad C = C + C_h. \quad (17)$$

The efficacy of this mechanism lies in its targeted modulation of the two fundamental processes of the SSM, governed by the state and output equations:

$$h'(t) = Ah(t) + Bx(t), \quad y(t) = Ch(t). \quad (18)$$

Here, the gating mechanisms act as dynamic filters, selectively distilling salient high-frequency information (e.g., anatomical edges, aliasing artifacts) from  $\mathcal{G}$  into the conditioning terms. The conditioning term  $B_h$  modulates the input projection matrix  $B$ , thereby conditioning the influence of the input signal  $x(t)$  on the state vector  $h(t)$  based on critical local details. Simultaneously, the term  $C_h$  modulates the output projection matrix  $C$ , which determines the projection from the state to the output, ensuring that these distilled high-frequency details are accurately rendered in the final reconstruction.

**Spatially-Aware Refinement.** To address the second limitation—the strictly local parameter generation in standard Mamba—the conditioned state projection matrices ( $B, C$ ) and the timestep parameter ( $\Delta$ ) are refined. Each is processed by a dedicated 1D depth-wise convolution with a kernel size of  $k = 7$ . This step allows the parameters for each token to be influenced by its spatial neighbors, injecting essential local context. Applying convolution to  $\Delta$  additionally ensures that the state transition dynamics evolve smoothly across the spatial sequence.

**Output Generation.** The refined parameters ( $\Delta, B, C$ ) are then supplied to the selective scan operation to produce an output feature map,  $F_{ssm}$ . The final output of the block,  $\tilde{F}_{low}$ , is obtained by modulating  $F_{ssm}$  with the residual feature map  $Z$  and applying a final linear projection, i.e.,

$$\tilde{F}_{low} = \text{Linear}_{out}(F_{ssm} \odot \text{SiLU}(Z)). \quad (19)$$

This final operation combines the long-range context from  $F_{ssm}$  with local features from  $Z$  through a gated modulation, followed by a final linear projection.

## Experiments

### Experimental Settings

**Datasets and Evaluation Metrics.** We evaluate HiFi-Mamba on two public MRI datasets: fastMRI (knee) (Zbontar et al. 2018) and CC359 (brain) (Warfield, Zou, and

Wells 2004). The fastMRI dataset comprises 1,172 complex-valued single-coil coronal knee scans acquired with Proton Density (PD) weighting. Each scan contains approximately 35 coronal slices with a matrix size of  $320 \times 320$ . We exclusively use the Proton Density Fat-Suppressed (PDFS) subset for both training and evaluation, following the official data split. The CC359 dataset contains raw brain MR scans acquired using clinical MR scanners (Discovery MR750; GE Healthcare, USA). Following the official split, 25 subjects are used for training and 10 for testing. Each slice has a matrix size of  $256 \times 256$ . To eliminate slices with limited anatomical content, we discard the first and last five slices for fastMRI and the first and last 15 slices for CC359.

In our experiments, the inputs are generated by applying equispaced 1D Cartesian undersampling masks, as provided by the fastMRI challenge (Zbontar et al. 2018). Specifically, for an acceleration factor (AF) of 4, the central 8% of k-space lines are fully sampled; for AF = 8, this portion is reduced to 4%.

We evaluate reconstruction performance using three widely used metrics: PSNR (Peak Signal-to-Noise Ratio) (Huynh-Thu and Ghanbari 2008), SSIM (Structural Similarity Index) (Wang et al. 2004), and NMSE (Normalized Mean Squared Error) (Zhao et al. 2016).

**Data Preprocessing Strategy.** Previous MRI reconstruction studies employ diverse preprocessing strategies (e.g., normalization schemes), leading to inconsistencies in intensity distributions that hinder fair comparisons. To mitigate this, we adopt a unified and transparent preprocessing strategy, applied consistently across experiments. This ensures reproducibility and minimizes confounding variables during inter-method comparison. Detailed pipeline specifications (e.g., normalization, k-space undersampling) are provided in the appendix to support reproducibility and future benchmarking.

**Training Details.** Our model consists of a stacked  $6 \times 2$  configuration of HiFi-Mamba units. We use the AdamW optimizer with an initial learning rate of  $1 \times 10^{-3}$ . A cosine annealing schedule with a 5-epoch warm-up is used, and training is performed for 100 epochs with a batch size of 4. An  $\ell_1$  loss is adopted for MRI reconstruction.

All experiments are conducted on two NVIDIA H100 GPUs. FLOPs are measured on a single NVIDIA A100 GPU. The implementation is based on PyTorch.

### Comparison with State-of-the-Art Methods

We compare HiFi-Mamba against representative state-of-the-art methods spanning three major model paradigms: CNN-based (UNet, ISTA-Net), Transformer-based (ReconFormer, FPSFormer), and Mamba-based (LMO). Evaluations are performed using equispaced 1D Cartesian undersampling at acceleration factors (AF) of  $4 \times$  and  $8 \times$  on the fastMRI and CC359 datasets. The quantitative results are summarized in Table 1.

HiFi-Mamba (P2), where P2 denotes a patch size of 2, consistently outperforms existing methods across most evaluation scenarios. Notably, it achieves significant gains at AF = 8 on both datasets, e.g., reaching 31.38 dB PSNR

Method	fastMRI						CC359					
	PSNR $\uparrow$		SSIM $\uparrow$		NMSE $\downarrow$		PSNR $\uparrow$		SSIM $\uparrow$		NMSE $\downarrow$	
	4 $\times$	8 $\times$	4 $\times$	8 $\times$	4 $\times$	8 $\times$	4 $\times$	8 $\times$	4 $\times$	8 $\times$	4 $\times$	8 $\times$
Zero-Filling	29.25	25.95	0.723	0.620	0.035	0.064	24.79	21.27	0.725	0.576	0.053	0.120
UNet (Zhou et al. 2018)	31.66	28.60	0.798	0.697	0.021	0.035	28.27	24.28	0.847	0.720	0.025	0.059
ISTA (Zhang and Ghanem 2018)	33.27	29.44	0.832	0.714	0.012	0.030	32.03	25.44	0.902	0.744	0.010	0.046
ReconFormer (Guo et al. 2023)	33.75	30.42	0.837	0.728	0.011	0.026	32.46	26.47	0.906	0.766	0.010	0.039
FpsFormer (Meng et al. 2025)	33.74	30.63	0.841	0.732	0.011	0.026	32.35	26.65	0.897	0.768	0.010	0.038
LMO (Li et al. 2025)	34.49	31.10	0.846	0.744	0.011	0.023	35.35	27.99	0.921	0.787	0.006	0.028
HiFi-Mamba (P2)	34.47	31.38	0.853	0.758	0.010	0.021	35.74	28.08	0.935	0.802	0.005	0.027
<b>HiFi-Mamba (P1)</b>	<b>34.85</b>	<b>31.81</b>	<b>0.855</b>	<b>0.762</b>	<b>0.010</b>	<b>0.020</b>	<b>36.93</b>	<b>28.49</b>	<b>0.942</b>	<b>0.810</b>	<b>0.004</b>	<b>0.026</b>

Table 1: Quantitative results on fastMRI-Equispaced and CC359-Equispaced under 4 $\times$  and 8 $\times$  accelerations.

Method	Scanning	FLOPs	PSNR $\uparrow$	SSIM $\uparrow$
ReconFormer	Attention	270.60G	30.42	0.728
FpsFormer	Attention	200.45G	30.63	0.732
LMO	4-Directional	484.98G	31.10	0.744
HiFi-Mamba (P1)	1-Directional	270.37G	<b>31.81</b>	<b>0.762</b>
<b>HiFi-Mamba (P2)</b>	1-Directional	<b>67.87G</b>	31.38	0.758

Table 2: Efficiency comparison on fastMRI (AF = 8, 320  $\times$  320) using NVIDIA A100.

and 0.758 SSIM on fastMRI, and 28.08 dB PSNR and 0.802 SSIM on CC359. Although its PSNR (34.47 dB) on fastMRI at AF = 4 is slightly lower than LMO (34.49 dB), HiFi-Mamba (P2) still surpasses LMO in SSIM and NMSE and consistently outperforms all other methods across the remaining settings, demonstrating strong generalization capability.

The finer-grained variant HiFi-Mamba (P1), with a patch size of 1, further improves reconstruction performance, establishing a new state-of-the-art. Specifically, it achieves 34.85 / 31.81 dB PSNR and 0.855 / 0.762 SSIM on fastMRI (AF = 4 / 8), with corresponding NMSE of 0.010 / 0.020. On CC359, it achieves 36.93 dB PSNR and 0.942 SSIM at AF = 4, and maintains robust performance at AF = 8, reaching 28.49 dB PSNR, 0.810 SSIM, and 0.026 NMSE, consistently surpassing prior state-of-the-art methods across all metrics.

Collectively, these results highlight the effectiveness of our frequency-aware architecture and fine-grained feature modeling in achieving robust, high-fidelity MRI reconstruction under aggressive acceleration.

**Visualization Results.** Figure 3 presents qualitative comparisons under 4 $\times$  and 8 $\times$  acceleration on representative slices from the fastMRI (knee) and CC359 (brain) datasets. The top row shows the reconstructed images, while the bottom row displays the corresponding error maps (difference from ground truth), color-coded from 0 (blue) to 0.2 (red). Yellow circles and red arrows indicate discrepancies in structural detail, while blue boxes mark zoomed-in regions for closer inspection.

ISTA-Net and UNet exhibit edge blurring and loss of

$\omega$ -Lap	HiFi-Mamba	DSFA	CRM	PSNR $\uparrow$	SSIM $\uparrow$	NMSE $\downarrow$
$\checkmark$				27.07	0.790	0.032
$\checkmark$	$\checkmark$			27.46	0.794	0.030
$\checkmark$	$\checkmark$	$\checkmark$		27.99	0.799	0.028
$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	<b>28.07</b>	<b>0.802</b>	<b>0.027</b>

Table 3: Ablations on CC359 (AF = 8, patch = 2).

structural detail. ReconFormer and FPS-Former partially alleviate these issues but still fail to reconstruct fine anatomical features with high acceleration factors. Notably, FPS-Former produces visually sharp contours in AF = 8 brain reconstructions, but introduces hallucinated structures inconsistent with the ground truth. The Mamba-based LMO also exhibits boundary degradation and missing details.

In contrast, HiFi-Mamba demonstrates robustness to various anatomical structures and acceleration factors. These advantages are also reflected in the corresponding error maps.

**Efficiency** We evaluate computational efficiency on the fastMRI dataset (AF = 8, 320  $\times$  320) using a single NVIDIA A100 GPU. As shown in Table 2, HiFi-Mamba achieves a favorable balance between performance and efficiency. HiFi-Mamba (P2) achieves 31.38 dB PSNR and 0.758 SSIM with only 67.87G FLOPs—substantially lower than all competing models, while matching or surpassing heavier attention-based architectures such as ReconFormer and FpsFormer. HiFi-Mamba (P1), though requiring 270.37G FLOPs, delivers the best overall performance (31.81 dB PSNR, 0.762 SSIM), outperforming the state-of-the-art LMO (484.98G FLOPs) in both accuracy and efficiency. These results demonstrate the resource-awareness of our design for high-fidelity MRI reconstruction.

### Ablation Studies and Analysis

**Ablation on Mamba Unit.** We perform an ablation study on the CC359 dataset with AF = 8 to assess the contribution of each component. Starting with the proposed  $\omega$ -Laplacian transform, we obtain 27.07 dB PSNR and 0.790 SSIM. Incorporating the HiFi-Mamba block, which enables frequency-aware interaction, increases the PSNR to 27.46 dB and the SSIM to 0.794. Adding the DSFA module refines the fused features, improving performance to 27.99 dB PSNR and 0.799 SSIM. Finally, appending the

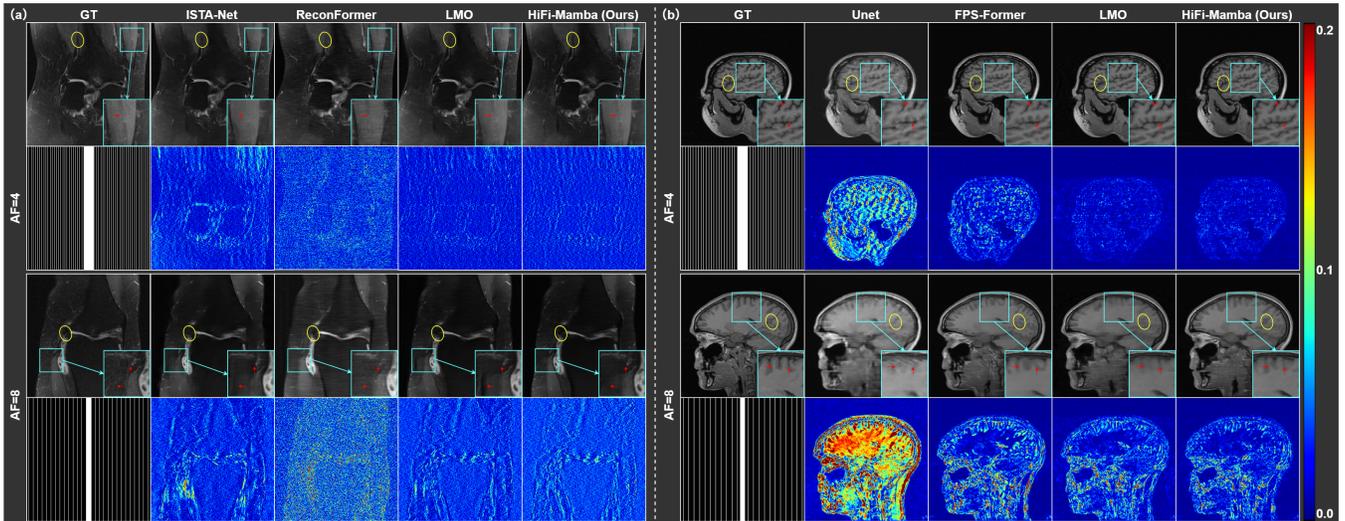


Figure 3: Qualitative comparison on the fastMRI and CC359 datasets under single-coil settings. (a) Reconstruction results on the fastMRI knee dataset with acceleration factors  $AF = 4$  and  $AF = 8$ . (b) Reconstruction results on the CC359 brain dataset under the same acceleration factors. The second row of each subplot shows the corresponding error maps. The blue boxes, yellow ellipses and red arrow highlight the details in the reconstruction results.

Patch-Size	Depth	PSNR	SSIM	NMSE
2	$3 \times 2$	27.92	0.800	0.028
2	$4 \times 2$	27.95	0.800	0.028
2	$6 \times 2$	28.07	0.802	0.027
2	$8 \times 2$	28.15	0.805	0.027
4	$6 \times 2$	27.71	0.793	0.029
1	$6 \times 2$	<b>28.49</b>	<b>0.810</b>	<b>0.026</b>

Table 4: Model architecture experiment is conducted on the CC359 dataset with  $AF = 8$ .

CRM module enhances high-frequency stream representations and achieves the best performance: 28.07 dB PSNR, 0.802 SSIM, and 0.027 NMSE. These results demonstrate the complementary benefits of each module and their collective contribution to reconstruction quality.

**Ablation on Model Depth and Patch Size.** We perform an ablation study on model depth and patch size under an  $8 \times$  acceleration setting using the CC359 dataset. As shown in Table 4, with patch size fixed at 2, increasing depth from  $3 \times 2$  to  $6 \times 2$  consistently improves PSNR and SSIM, indicating better anatomical modeling. Further increasing to  $8 \times 2$  offers only marginal gains, suggesting saturation beyond moderate depth.

Conversely, with depth fixed at  $6 \times 2$ , decreasing patch size from 4 to 1 steadily improves all metrics. The best performance is achieved with patch size 1 and depth  $6 \times 2$ , reaching 28.49 dB PSNR, 0.810 SSIM, and 0.026 NMSE.

These results confirm the scalability of our design—deeper models and finer spatial granularity enhance reconstruction without overfitting, highlighting the robustness and flexibility of the proposed architecture.

Mechanism	PSNR	SSIM	NMSE
DConv1D ( $k = 3$ )	27.81	0.796	0.030
DConv1D ( $k = 5$ )	28.05	0.805	0.028
DConv1D ( $k = 7$ )	28.49	0.810	0.026

Table 5: Spatially-Aware experiment is conducted on the CC359 dataset with  $AF = 8$ , and patch size = 1. Here,  $k$  denotes the kernel size.

**Ablation on Kernel Size.** We evaluate the impact of kernel size in the spatially-aware refinement module by varying the receptive field of the depthwise 1D convolution applied to the conditioned parameters ( $B$ ,  $C$ ,  $\Delta$ ). As shown in Table 5, increasing the kernel size from 3 to 7 yields consistent improvements across all metrics. The kernel size 7 configuration achieves the best performance, indicating that larger spatial context enhances the anatomical coherence of the learned dynamics.

## Conclusion

In this paper, we present HiFi-Mamba, a frequency-aware dual-stream architecture for MRI reconstruction. By coupling a  $w$ -Laplacian block for spectral decoupling with a guided Mamba block that models global anatomy while integrating high-frequency detail, HiFi-Mamba addresses key limitations of prior state-space models, including redundant scanning, local-only parameterization, and frequency insensitivity. Extensive experiments demonstrate that HiFi-Mamba achieves state-of-the-art reconstruction accuracy while substantially reducing computational cost. We believe it offers a promising direction for frequency-structured and efficiency-aware modeling in MRI reconstruction.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 62301352.

## References

- Aggarwal, H. K.; Mani, M. P.; and Jacob, M. 2019. MoDL: Model-based deep learning architecture for inverse problems. *IEEE Transactions on Medical Imaging*, 38(2): 394–405.
- Burt, P. J.; and Adelson, E. H. 1987. The Laplacian pyramid as a compact image code. In *Readings in computer vision*, 671–679. Elsevier.
- Chen, Y.; Fang, P.; Zhong, X.; Yu, J.; Zhang, X.; and Li, T. 2024. Hi-ResNet: Edge detail enhancement for high-resolution remote sensing segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
- Eo, T.; Jun, Y.; Kim, T.; Jang, J.; Lee, H.; Hwang, D.; and Ye, J. C. 2018. KIKI-net: Cross-domain Convolutional Neural Networks for Reconstructing Undersampled Magnetic Resonance Images. *Magnetic Resonance in Medicine*, 80(5): 2188–2201.
- Fu, D.; Sun, T.; Fang, P.; Cai, X.; and Kim, H. 2025. MOGO: Residual Quantized Hierarchical Causal Transformer for High-Quality and Real-Time 3D Human Motion Generation. *arXiv preprint arXiv:2506.05952*.
- Gu, A.; Dao, T.; et al. 2023. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. *arXiv preprint arXiv:2312.00752*.
- Gu, A.; Goel, K.; Dao, T.; et al. 2022. Efficiently Modeling Long Sequences with Structured State Spaces. In *Advances in Neural Information Processing Systems*, volume 35, 21915–21929.
- Gu, Y.; Lu, Y.; You, H.; Zhan, Y.; Zhou, S.; and Shen, D. 2021. Reference-based magnetic resonance image reconstruction using texture transformer. *arXiv preprint arXiv:2111.09492*.
- Guo, H.; Li, J.; Dai, T.; Ouyang, Z.; Ren, X.; and Xia, S.-T. 2024. Mambair: A simple baseline for image restoration with state-space model. In *European conference on computer vision*, 222–241. Springer.
- Guo, P.; Mei, Y.; Zhou, J.; Jiang, S.; and Patel, V. M. 2023. Reformer: Accelerated mri reconstruction using recurrent transformer. *IEEE transactions on medical imaging*, 43(1): 582–593.
- Hammernik, K.; Klatzer, T.; Kobler, E.; Recht, M. P.; Sodickson, D. K.; Pock, T.; and Knoll, F. 2018. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, 79(6): 3055–3071.
- Huang, J.; Fang, Y.; Wu, Y.; Wu, H.; Gao, Z.; Li, Y.; Del Ser, J.; Xia, J.; and Yang, G. 2022. Swin transformer for fast MRI. *Neurocomputing*, 493: 281–304.
- Huang, J.; Yang, L.; Wang, F.; Wu, Y.; Nan, Y.; Wu, W.; Wang, C.; Shi, K.; Aviles-Rivero, A. I.; Schoenlieb, C.-B.; et al. 2025. Enhancing global sensitivity and uncertainty quantification in medical image reconstruction with Monte Carlo arbitrary-masked mamba. *Medical Image Analysis*, 99: 103334.
- Huynh-Thu, Q.; and Ghanbari, M. 2008. The scope of PSNR in image and video quality assessment. *Electronics letters*, 44(13): 800–801.
- Jerban, S.; Chang, E. Y.; and Du, J. 2020. Magnetic resonance imaging (MRI) studies of knee joint under mechanical loading. *Magnetic Resonance Imaging*, 65: 27–36.
- Knoll, F.; Hammernik, K.; Zhang, C.; Moeller, S.; Pock, T.; and Sodickson, D. K. 2020. Deep-learning methods for parallel magnetic resonance imaging reconstruction: A survey of the current approaches, trends, and issues. *IEEE Signal Processing Magazine*, 37(1): 128–140.
- Korkmaz, Y.; Dar, S. U.; Yurt, M.; Özbey, M.; and Cukur, T. 2022. Unsupervised MRI reconstruction via zero-shot learned adversarial transformers. *IEEE Transactions on Medical Imaging*, 41(7): 1747–1763.
- Li, H.; Wang, Y.; Xu, Y.; Ding, Z.; Xu, C.; Lu, Y.; Ye, X.; and Bai, S. 2024. A Survey on Visual Mamba. *arXiv preprint arXiv:2404.15956*.
- Li, J.; Wang, C.; Xu, Y.; Qian, Y.; Yang, Y.; and Shen, D. 2025. LMO: Linear Mamba Operator for MRI Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5112–5122.
- Liu, Y.; Tian, Y.; Zhao, Y.; Yu, H.; Xie, L.; Wang, Y.; Ye, Q.; Jiao, J.; and Liu, Y. 2024. VMamba: Visual State Space Model. In *Advances in Neural Information Processing Systems*, volume 37, 103031–103063.
- Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. 2021. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 10012–10022.
- Lustig, M.; Donoho, D.; and Pauly, J. M. 2007. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6): 1182–1195.
- Ma, X.; Ni, Z.; and Chen, X. 2024. Tinyvim: Frequency decoupling for tiny hybrid vision mamba. *arXiv preprint arXiv:2411.17473*.
- Meng, Y.; Yang, Z.; Shi, Y.; and Song, Z. 2025. Boosting vit-based mri reconstruction from the perspectives of frequency modulation, spatial purification, and scale diversification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 6135–6143.
- Qin, C.; Schlemper, J.; Caballero, J.; Price, A. N.; Hajnal, J. V.; and Rueckert, D. 2019. Convolutional recurrent neural networks for dynamic MR image reconstruction. *IEEE Transactions on Medical Imaging*, 38(1): 280–290.
- Rahman, M. M.; Tutul, A. A.; Nath, A.; Laishram, L.; Jung, S. K.; and Hammond, T. 2024. Mamba in Vision: A Comprehensive Survey of Techniques and Applications. *arXiv preprint arXiv:2410.03105*.
- Schlemper, J.; Caballero, J.; Hajnal, J. V.; Price, A. N.; and Rueckert, D. 2018. A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction. *IEEE Transactions on Medical Imaging*, 37(2): 491–503.

Shan, S.; Gao, Y.; Waddington, D.; Chen, H.; Whelan, B.; Liu, P.; Wang, Y.; Liu, C.; Gan, H.; Gao, M.; et al. 2024. Image Reconstruction with B 0 Inhomogeneity using a Deep Unrolled Network on an Open-bore MRI-Linac. *IEEE Transactions on Instrumentation and Measurement*.

Varela-Mattatall, L.; and D'Arcy, R. C. N. 2017. Motion artifact reduction techniques in MRI: A review. *Journal of Magnetic Resonance Imaging*, 45(6): 1779–1790.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is All You Need. In *Advances in Neural Information Processing Systems*, volume 30.

Wang, J.; Wu, S.; Xu, Z.; Shi, R.; Qian, Y.; Cai, J.; Huang, Y.; et al. 2023. Dual-domain accelerated MRI reconstruction using transformers with learning-based undersampling. *Computerized Medical Imaging and Graphics*, 106: 102179.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612.

Warfield, S. K.; Zou, K. H.; and Wells, W. M. 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23(7): 903–921.

Wu, Z.; Liao, W.; Yan, C.; Zhao, M.; Liu, G.; and Ma, N. 2023. Deep learning based MRI reconstruction with transformer. *Computer Methods and Programs in Biomedicine*, 234: 107602.

Ye, J. C. 2019. Compressed sensing MRI: a review from signal processing perspective. *BMC Biomedical Engineering*, 1(1): 8.

Zbontar, J.; Knoll, F.; Sriram, A.; Muckley, M. J.; Bruno, M.; Defazio, A.; Parente, M.; Zitnick, C. L.; Sodickson, D. K.; Yakubova, N.; et al. 2018. fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv preprint arXiv:1811.08839*.

Zhang, J.; and Ghanem, B. 2018. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1828–1837.

Zhang, X.; He, R.; Wang, F.; and Liu, Q. 2024. Computation-Efficient Era: A Comprehensive Survey of State Space Models in Medical Image Analysis. *arXiv preprint arXiv:2405.07639*.

Zhao, H.; Gallo, O.; Frosio, I.; and Kautz, J. 2016. Loss functions for neural networks for image processing. *arXiv preprint arXiv:1511.08861*.

Zhou, B.; Zhou, S.; Wang, L.; Xing, Y.; Wang, Q.; Zhang, S.; Liu, C.; and Lu, H. 2020. DuDoRNet: Learning a Dual-Domain Recurrent Network for Fast MRI Reconstruction With Deep T1 Prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4273–4282.

Zhou, Z.; Rahman Siddiquee, M. M.; Tajbakhsh, N.; and Liang, J. 2018. Unet++: A nested u-net architecture for medical image segmentation. In *International workshop on deep learning in medical image analysis*, 3–11. Springer.