DDHMamba: Dual Domains Hilbert Mamba for Multi-Contrast MRI Super-Resolution

Yuguang Yan
Guangdong University of Technology
Guangzhou, China
ygyan@gdut.edu.cn

Yonglin Xu
Guangdong University of Technology
Guangzhou, China
1076577207@qq.com

Zipeng Zhu
Guangdong University of Technology
Guangzhou, China
zhuzipeng@mails.gdut.edu.cn

Boyan Xu
Guangdong University of Technology
Guangzhou, China
hpakyim@gmail.com

Ruichu Cai†
Guangdong University of Technology
Guangzhou, China
cairuichu@gmail.com

***Abstract*—Driven by advancements in deep learning and increasing clinical demands in Magnetic Resonance Imaging (MRI), Multi-Contrast MRI Super-Resolution (MCSR) has been developed. It utilizes an easily obtainable reference contrast MRI as an auxiliary to improve the quality of low-resolution target contrast MRI. While the existing approach shows promise, convolution-based methods cannot capture long-range dependencies, and Transformers suffer from quadratic complexity. Recently, image super-resolution technology based on Mamba has received significant attention due to its effectiveness. However, Mamba's sequential scanning hinders the preservation of local details and the learning of dependencies among multi-modal images. Moreover, the frequency features across different contrast MRIs exhibit inherent relationships, modeling their complex correlations within the Fourier domain remains a significant challenge. To mitigate these challenges, we propose Dual Domains Hilbert Mamba (DDHMamba), a new Mamba-based framework that captures long-range dependencies with linear complexity through dual domain modeling, leveraging multi-modal features for efficient MRI super-resolution. Specifically, we propose a Hilbert Scan Fusion Mamba (HSFM) module with two novel cross-modal Hilbert scanning mechanisms: one operating in the spatial domain and the other in the Fourier domain to better capture features of distinct modalities across their respective domains. In addition, we introduce a Frequency Assisted Local Fusion (FALF) module to enhance the fusion capability of local image information from two complementary modalities. Experiments on BraTS and fastMRI datasets demonstrate that DDHMamba achieves superior MCSR performance across multiple scales.**

***Index Terms*—Multi-contrast MRI; Mamba; Dual Domains Learning; Hilbert curves; Local Fusion**

I. Introduction

Magnetic Resonance Imaging (MRI) has become one of the most widely utilized and powerful modalities in clinical diagnosis and research. Its principal advantages lie in providing exceptional soft-tissue contrast [1]. However, a significant limitation restricts its broader practicality: the long-duration data acquisition required to obtain high-resolution (HR) images. In clinical practice, it not only reduces diagnostic efficiency but also increases patient discomfort, which frequently leads to motion artifacts that can degrade image quality and compromise diagnostic accuracy [2].

Multi-Contrast Super-Resolution for MRI has surfaced as a promising computational strategy to mitigate this trade-off [3]–[6]. MCSR techniques aim to synthesize a high-resolution image for a target MRI contrast (e.g., T2-weighted) by synergistically integrating information from a complementary, more readily acquired reference contrast (e.g., T1-weighted) that is often already part of standard clinical protocols. [7], [8]. Compared with Single-Contrast MRI Super-Resolution, MCSR offers the following advantages: 1) Due to the complexity of human anatomy, acquiring multi-contrast MRIs for disease diagnosis is more common in clinical practice than relying on single contrast MRI; 2) It leverages complementary information from different modalities to enhance the restoration of anatomical details in low-resolution images of target modalities

The primary challenge of MCSR lies in modeling the intrinsic dependencies within each modality and fully leveraging their complementary information, while filtering out irrelevant noise to enhance the restoration of anatomical details [9]. Convolutional neural networks (CNNs), as one of the core architectures of deep learning, have been widely applied to MRI super-resolution tasks due to their powerful feature extraction capabilities [10]–[12]. By stacking multiple convolutional layers, CNNs can capture local structural information, but they have limitations in handling long-distance dependencies. To address this issue, Transformer-based models effectively capture global relationships across modalities through attention mechanisms [13], [14]. While achieving notable success, these models impose substantial computational and memory burdens due to the quadratic complexity of attention with respect to sequence length.

Very recently, the Mamba model has received a lot of attention. Its core innovation—a selection scanning mecha-

†Corresponding author.

This research was supported in part by National Natural Science Foundation of China (62206061,62406078,U24A20233), Guangdong Basic and Applied Basic Research Foundation (2024A1515011901), Guangzhou Basic and Applied Basic Research Foundation (2023A04J1700).

nism and a hardware-aware parallel algorithm—allows it to efficiently capture long-range dependencies while maintaining linear computational complexity with sequence length. [15]. This makes Mamba particularly intriguing for vision tasks [16], [17]. However, due to Mamba's sequential scanning mechanism, it becomes challenging to preserve the local details and effectively learn the correspondence between MRIs of different modalities. Moreover, while the Mamba model demonstrates promising capability in capturing intra-modal frequency interactions [18], [19], the inherent spectral dependencies across multi-modal MRI data pose persistent challenges in effectively modeling cross-domain correlations within Fourier-learned representations.

To address the above challenges, we propose Dual Domains Hilbert Mamba (DDHMamba), a new deep learning framework built upon Mamba that effectively captures long-range dependencies in MRI data, extracts informative features across different modalities, and jointly models them in both spatial and Fourier domains to achieve efficient MCSR. Specifically, we propose a Hilbert Scan Fusion Mamba (HSFM) module featuring dual-domain scanning mechanisms: a spatial domain 2D Hilbert alternating scanner and a Fourier domain counterpart (3D Hilbert curve), which synergistically capture complementary visual patterns from different modalities within their respective domains. In addition, we introduce a Frequency Assisted Local Fusion (FALF) module that enhances the fusion of local image information from these two complementary modalities. The main contributions are summarized as follows:

- We propose DDHMamba, a new framework for efficient Multi-Contrast MRI Super-resolution. The proposed HSFM block uses two cross-modal Hilbert scanning mechanisms, one operating in the spatial domain and the other in the Fourier domain, which enable more effective capture of features from different modalities in two different domains.
- We introduce a FALF module to significantly improves the super-resolution performance of local image details from two different modalities of MRI.
- We demonstrate the effectiveness of our approach by comparing it with previous MCSR methods on the BraTS and fastMRI datasets.

II. Related Workds

Early research on Multi-Contrast MRI Super-Resolution relied primarily on convolutional neural networks (CNNs), leveraging their strong local feature extraction capabilities to learn mappings from low-resolution (LR) to high-resolution (HR) images. Pioneering works such as MCSR [3] introduced CNNs into MCSR through a simple multi-layer stacked convolutional architecture. However, CNNs are inherently limited by their local receptive fields, which makes it difficult to model long-range dependencies in MRI data.

To address the locality constraint of CNNs, Vision Transformer (ViT) is introduced to MCSR. Transformers were also adopted to fuse complementary information across modalities [13]. Methods such as McMRSR [14] integrated features from T1 and T2 modalities via cross-attention mechanisms. FSMNet [20] designed a frequency-spatial mutual learning framework that leverages Fourier transforms for global feature extraction and cross-modal selective fusion to combine complementary information across modalities. DiffMSR [6] extracts prior knowledge through a prior feature extractor and performs MCSR tasks by combining a diffusion model based on a Transformer architecture. Despite their effectiveness, Transformers suffer from quadratic computational complexity, imposing a heavy computational burden when processing high-resolution medical images.

Recently, Mamba, have gained attention due to their linear computational complexity and strong ability to model long sequences. In medical imaging, Mamba has been explored for MRI Super-Resolution task. For example, GLMamba [21] introduced a dual-branch architecture combining a global Mamba branch for capturing long-range dependencies in LR images and a local Mamba branch for extracting fine details from HR reference images, enhanced by deformable convolution and a multimodal fusion module. Similarly, MMR-Mamba [18] incorporated target-guided cross Mamba and selective frequency fusion to efficiently integrate multimodal features in both spatial and frequency domains.

Unlike the MCSR method based on Mamba mentioned above, we employed a carefully designed Hilbert curve to expand MRI images in both spatial and Fourier domains, preserving the locality of the images in their one-dimensional sequence, ensuring that images within the same region are arranged in close proximity in the sequence.

III. Method

A. Preliminaries: State Space Models

State Space Models (SSMs) are typically defined as linear time-invariant systems that map a sequence $x(t) \in \mathbb{R}^L$ to an output sequence $y(t) \in \mathbb{R}^L$ through an intermediate hidden state $h(t) \in \mathbb{R}^N$. Mathematically, SSMs employ the following ordinary differential equation (ODE) to express the systems:

$$\begin{aligned} h'(t) &= Ah(t) + Bx(t), \\ y(t) &= Ch(t) + Dx(t), \end{aligned} \tag{1}$$

where $A \in \mathbb{R}^{N\times N}$ denotes the state transition matrix, $B \in \mathbb{R}^{N\times 1}$ and $C \in \mathbb{R}^{1\times N}$ are the projection parameters, and $D \in \mathbb{R}^1$ is a skip connection.

B. Proposed Architecture

Fig. 1 illustrates the overall framework of the proposed DDHMamba. Given a data pair that includes $X_{lr} \in \mathbb{R}^{H\times W}$ from the low-resolution MRI of the target modality and $X_{ref} \in \mathbb{R}^{H\times W}$ from high-resolution MRI of the reference modality, DDHMamba aims to generate a high-quality MRI X_{hr} of the target modality. In the beginning, we feed X_{lr} and X_{ref} into the model, and employ convolution layers to map them into the feature space. After that, we apply Mamba

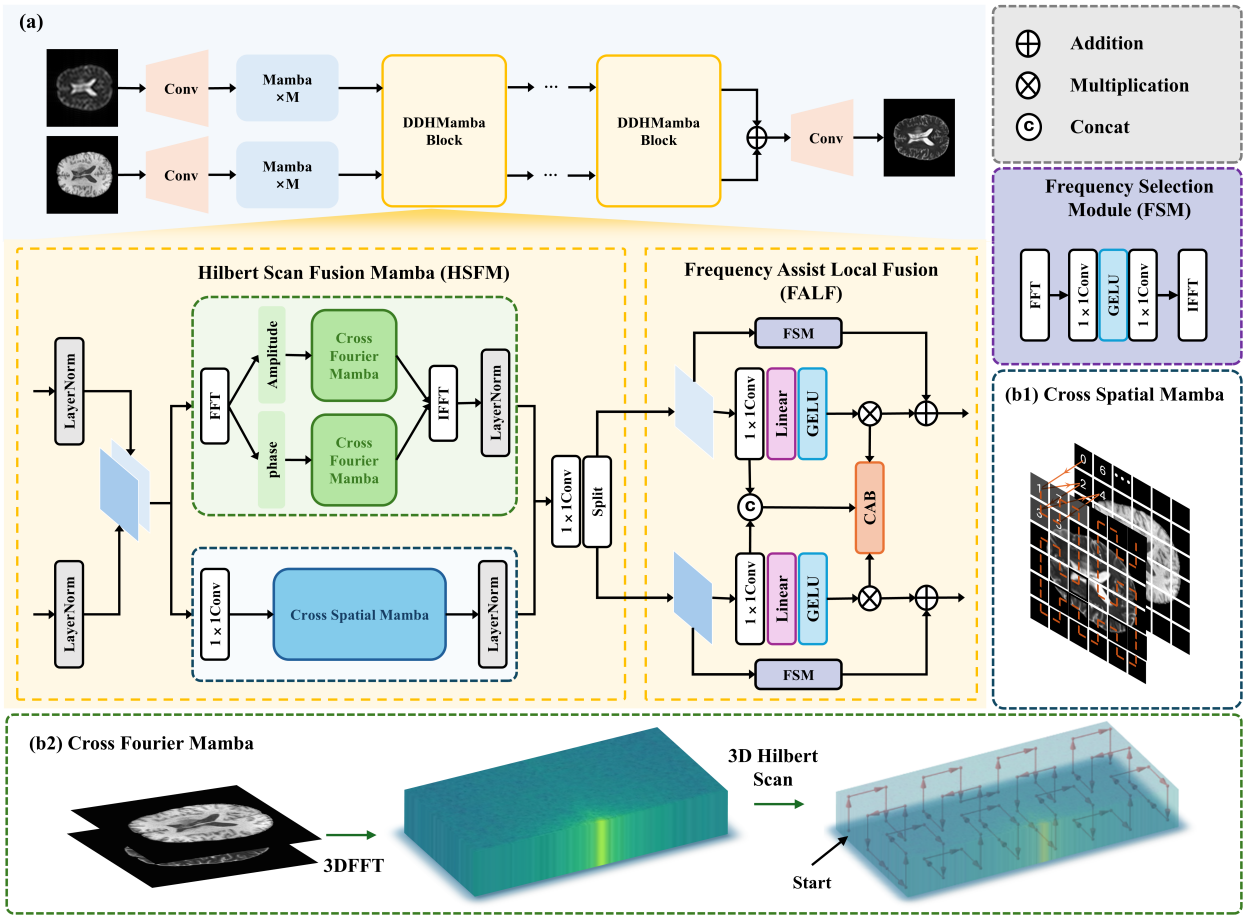

Fig. 1. (a) The overview of DDHMamba. It contains M Mamba blocks for feature extraction and N DDHMamba modules to learn multi-modal feature representations. Each DDHMamba integrates a HSFM block and a FALF block. (b1) The scanning path of HSFM's Cross Spatial Mamba branch. The solid line indicates the actual scanning path that traverses between two images. The dashed line represents the Hilbert curve trajectory, guiding the solid line through the entire spatial domain. (b2) The illustration of the Cross Fourier Mamba scanning method in phase and amplitude spectra.

blocks to capture the long-range dependencies and obtain feature maps $\mathcal{F}_{lr}$ and $\mathcal{F}_{ref}$ as follows:

$$\begin{aligned} \mathcal{F}_{lr} &= \psi_{lr}(Conv_{lr}(X_{lr})), \\ \mathcal{F}_{ref} &= \psi_{ref}(Conv_{ref}(X_{ref})), \end{aligned} \tag{2}$$

where $\mathcal{F}_{lr} \in \mathbb{R}^{H\times W\times C}$, $\mathcal{F}_{ref} \in \mathbb{R}^{H\times W\times C}$, C is the channel dimension of the feature map, and $\psi(.)$ represents Mamba operations for extracting image features. Thereafter, we design Hilbert Scan Fusion Mamba (HSFM) and Frequency Assisted Local Fusion (FALF) to deeply integrate two different feature maps $\mathcal{F}_{lr}$ and $\mathcal{F}_{ref}$ to obtain $\mathcal{F}'_{lr}$ and $\mathcal{F}'_{ref}$. Finally, we combine them to obtain the super-resolution MR images of the target modality through a convolution layer:

$$\tilde{X}_{hr} = \mathrm{Conv}(\mathcal{F}'_{lr} + \mathcal{F}'_{ref}), \tag{3}$$

and employ the $L1$ loss to optimize our model:

$$\mathcal{L} = \left\| \tilde{X}_{hr} - X_{hr} \right\|_1. \tag{4}$$

C. Hilbert Scan Fusion Mamba

For multi-modal images, previous works [18], [22], [23] based on Mamba tend to aggregate the multi-modal features using multiple Mamba models. However, such approaches introduce redundant parameters and ignore the relations between images of different modalities at the same local position. Although previous work has explored the modeling of single-modal data in the Fourier domain [19], [24], how to capture cross-modal image interaction in spatial and Fourier domains remains a challenge. Inspired by [19], [24], we propose a cross-modal scanning method based on Hilbert curves to replace the simple scanning path used in classical Mamba. As shown in Fig. 1, the Hilbert Scan Fusion Mamba(HSFM) stacks the input $\mathcal{F}_{lr} \in \mathbb{R}^{H\times W\times C}$ and $\mathcal{F}_{ref} \in \mathbb{R}^{H\times W\times C}$ in the third dimension of space to obtain $\mathcal{F}_{stack} \in \mathbb{R}^{2\times H\times W\times C}$. After that, we design two scanning strategies in spatial and Fourier domains.

Cross Spatial Mamba. Hilbert curves can map high-dimensional spatial data into 1D sequences, allowing adjacent spatial position pixels to maintain their proximity in the 1D sequence. Although the Mamba model based on Hilbert curves enjoys a strong capability to preserve intra-modality locality, simply connecting Hilbert curves of different modalities may unexpectedly ignore inter-modality local pixel-level relation-

TABLE I
QUANTITATIVE COMPARISON ON BRATS AND FASTMRI DATASETS, USING PSNR AND SSIM AS PERFORMANCE METRICS. PARAM REFERS TO THE MODEL PARAMETERS.

Method	Param	FLOPs	2×				4×			
			BraTS		fastMRI		BraTS		fastMRI	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
MCSR	2.5M	85.82G	40.53	0.9818	30.75	0.7517	34.72	0.9587	28.68	0.6034
MCMRSR	3.5M	130.94G	40.35	0.9796	30.71	0.7513	34.64	0.9597	28.60	0.6012
MINet	3.9M	103.72G	41.29	0.9861	31.00	0.7630	35.75	0.9624	28.91	0.6137
DCAMSR	8.0M	83.07G	41.43	0.9870	31.01	0.7634	35.80	0.9629	28.95	0.6143
FSMNet	14.1M	137.82G	41.63	0.9872	31.04	0.7630	35.82	0.9626	28.99	0.6128
Panmamba	0.48M	9.13G	41.42	0.9873	31.04	0.7642	35.76	0.9617	28.98	0.6141
Ours	0.64M	15.21G	**42.21**	**0.9884**	**31.13**	**0.7660**	**36.35**	**0.9651**	**29.09**	**0.6181**

ships. To tackle this, in the spatial domain, we adopt a 2D symmetric Hilbert alternating scan in Mamba to extract the cross-correlation of multi-modal data. As shown in Fig. 1(b1), for each modality, we construct identical Hilbert curve that traverses every point. Subsequently, a path is devised to alternately scan the symmetric points on both of two curves along the direction of the Hilbert curve. We then convert each coordinate (d, h, w) of $\mathcal{F}_{stack}$ on the Hilbert curve into a unique one-dimensional index:

$$\mathcal{I}(d, h, w) = w + hW + dHW. \tag{5}$$

Through this path, two local pixels from different modalities at the same position are rearranged into adjacent points, enhancing Mamba's capability to model local dependencies among intra and inter-modalities.

Cross Fourier Mamba. We apply the 3D Fourier transform to convert $\mathcal{F}_{stack}$ into the frequency domain, which allows every pixel in the Fourier space include the global information of each modality:

$$f(x)(o, p, q) = \frac{1}{\sqrt{DHW}} \sum_{d=0}^{D-1} \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} x(d, h, w) e^{-j2\pi\left(\frac{d}{D}o + \frac{h}{H}p + \frac{w}{W}q\right)} \tag{6}$$

Due to the conjugate symmetry of the Fourier transform, we retain only half of the spectrum. The amplitude spectrum $\mathcal{A}(x)(o, p, q)$ and phase spectrum $\mathcal{P}(x)(o, p, q)$ are derived by calculating the real $R(x)(o, p, q)$ and icomparemaginary $I(x)(o, p, q)$ parts of the frequency features as follows:

$$\begin{aligned} \mathcal{A}(x)(o, p, q) &= \sqrt{R^2(x)(o, p, q) + I^2(x)(o, p, q)}, \\ \mathcal{P}(x)(o, p, q) &= \arctan\left[\frac{I(x)(o, p, q)}{R(x)(o, p, q)}\right]. \end{aligned} \tag{7}$$

Since the 3D Hilbert curve has the same symmetry as the frequency, it exploits the spectral symmetry more effectively than the traditional Mamba scanning method does. As illustrated in Fig. 1(b2), in contrast to the spatial domain, 3D Hilbert curves are used to scan the amplitude spectrum and phase spectrum separately, which can capture the diverse dependence of the frequency features in each modality. Similarly, we use Eq. (5) to transform the coordinates in the Fourier domain into a one-dimensional index.

D. Frequency Assisted Local Fusion

Although Mamba has excellent long-distance dependency modeling capability, its inherent linear scanning mode makes it difficult to extract local features. Inspired by [25], we propose a Frequency Assisted Local Fusion module to overcome this challenge. As illustrated in Fig. 1, taking the input features $\mathcal{F}_{lr}$ and $\mathcal{F}_{ref}$, FALF applies two layers of 1×1 convolution operation and GELU activation functions to get feature maps $\mathcal{F}_{lr}^{fre}$ and $\mathcal{F}_{ref}^{fre}$, after applying the 2D Fast Fourier Transform for frequency selection. Considering the complementary

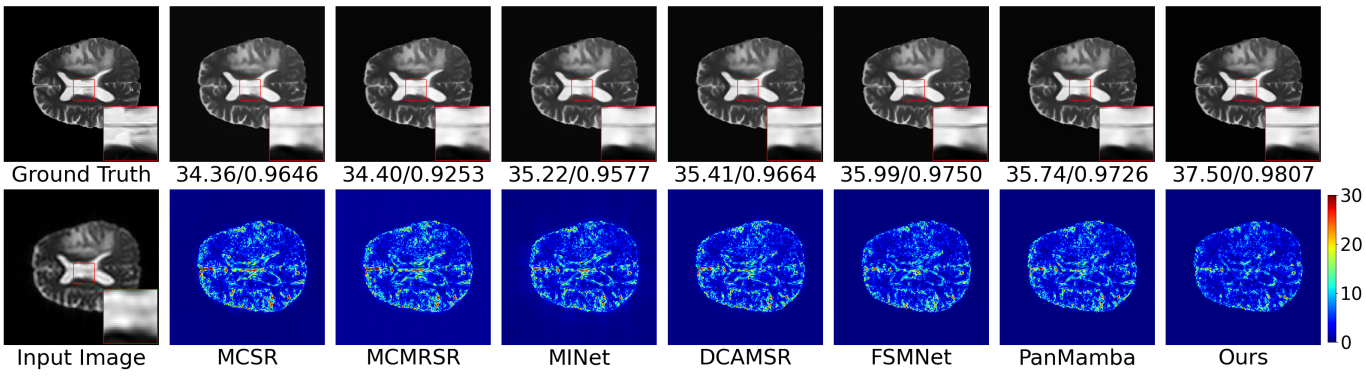

Fig. 2. Visual comparison of different MCSR methods under the scaling factor of ×4 on the BraTS2021 dataset. 1st row: SR images. 2nd row: error maps.

information between different modalities, FALF adaptively integrates features from each modality to generate enhanced local features. Specifically, we get $\tilde{\mathcal{F}}_{lr}$ and $\tilde{\mathcal{F}}_{ref}$ by adopting Layer Normalization and 1×1 convolution to each modality. Subsequently, we concat them and apply a channel attention (CA) block [26] to learn the scores of complementary local information of modalities:

$$\mathcal{S}_{lr}, \mathcal{S}_{ref} = \mathrm{CA}(\mathrm{Concat}(\tilde{\mathcal{F}}_{lr}, \tilde{\mathcal{F}}_{ref})), \tag{8}$$

After that, two enhanced local features $\mathcal{F}_{lr}^{spa}$ and $\mathcal{F}_{ref}^{spa}$ that fuse complementary information are obtained by:

$$\begin{aligned} \mathcal{F}_{lr}^{spa} &= \mathcal{S}_{lr} \cdot \mathrm{GELU}(\mathrm{Linear}(\tilde{\mathcal{F}}_{lr})), \\ \mathcal{F}_{ref}^{spa} &= \mathcal{S}_{ref} \cdot \mathrm{GELU}(\mathrm{Linear}(\tilde{\mathcal{F}}_{ref})). \end{aligned} \tag{9}$$

Finally, we add frequency feature maps and local features for outputs:

$$\mathcal{F}'_{lr} = \mathcal{F}_{lr}^{spa} + \mathcal{F}_{lr}^{fre}, \tag{10}$$

$$\mathcal{F}'_{ref} = \mathcal{F}_{ref}^{spa} + \mathcal{F}_{ref}^{fre}. \tag{11}$$

IV. Experiments

A. Datasets

The BraTS2021 [27] (Brain Tumor Segmentation) dataset is widely used for brain tumor segmentation research. It contains multi-modal brain MRI scans and serves as a common benchmark for evaluating cross-modal image generation and reconstruction algorithms. The fastMRI [28] dataset is dedicated to advancing accelerated Magnetic Resonance Imaging (MRI) reconstruction. It provides a large collection of raw k-space data and multi-contrast knee scans, establishing an important benchmark for image reconstruction and synthesis tasks.

We evaluate our method on the BraTS2021 and fastMRI datasets. On both datasets, we apply the k-space truncation [29] and then adopt zero-padding (ZP) to obtain low-resolution MRIs of the target modality. The scaling factor is set to 2 and 4. For the BraTS2021 dataset, we adopt T1-Weighted (T1WI) MRI as the reference modality and T2-Weighted (T2WI) MRI as the target modality. 300 and 100 pairs of T1WI and T2WI volumes are selected for training and testing, respectively. The middle 20 slices cropped to 224×224 resolution in each volume are chosen for our experiments. The fastMRI dataset contains both single-coil proton density-weighted (PDWI) and fat-suppressed proton density-weighted (FS-PDWI) scans of the knee MRI volumes. Following [5], we use 227 pairs of PDWI (reference modality) and FS-PDWI (target modality) as the training set and another 25 pairs as the test set. Each slice is resized to 256×256 in our experiments.

B. Compared Methods and Implementation Details

We compare our proposed method with several state-of-the-art methods, including MCSR [3], McMRSR [14], MINet [5], DCAMSR [30], FSMNet [20], Panmamba [23].

We using PSNR and SSIM as the evaluation metrics. PSNR quantifies the pixel-level fidelity by measuring the ratio between the maximum possible signal power and the power of the reconstruction error, with higher values indicating better quality. SSIM evaluates the perceptual similarity by considering luminance, contrast, and structure between the reconstructed and reference images, with a value closer to 1 representing higher structural preservation.

The entire framework of DDHMamba is implemented in PyTorch, with both the training and testing processes executed on an NVIDIA RTX4090 GPU. During the training phase, we employ the Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ to optimize the model parameters. The model is trained for 150,000 iterations with a batch size of 4, and the learning rate initialized to 2×10^{-4}.

C. Experimental Results

The quantitative results in Table I demonstrate that our method achieves higher PSNR values on average by 0.58 dB and 0.53dB under $2\times$ and $4\times$ scaling factor on the BraTS dataset, respectively. On the fastMRI dataset, our method also outperforms the other methods, showing improvements of 0.09dB and 0.10dB in PSNR values under $2\times$ and $4\times$ scaling factors compared to the second-ranked method. These significant improvements indicate that our feature fusion mechanism effectively preserves anatomical details during MRI super-resolution and highlights the robustness of our method in handling diverse MRI data. Remarkably, our method achieves these improvements while maintaining a low number of parameters and FLOPs among all compared methods.

Fig. 2 shows a pair of images from the BraTS test set to visualize the super-resolution results and error maps of different methods. We observe that MCSR and McMRSR both suffer from blurring and over-smoothing in fine tissue boundaries, resulting in loss of contrast and visible high-error patches in their error maps. MINet and DCAMSR produce artifacts and unresolved gradient details near edges, leading to concentrated high-error areas in those critical anatomical regions. PanMamba introduces slight noise-like distortions in the homogeneous tissue regions, which increases the overall error level and leads to a more fragmented error distribution. FSMNet preserves most large-scale structures but still exhibits mild blurring in the fine details of the white matter, with its error map displaying persistent mid-to-high error values in the zoomed area. In contrast, our method reconstructs the sharpest tissue boundaries and most accurate gray-level transitions, with its error map showing predominantly blue and minimal high-intensity regions, demonstrating superior preservation of anatomical details.

We observe that the proposed method reconstructs brain texture details more precisely, with lower error distribution.

D. Ablation Study

To further analyze the effectiveness of different components in our model, we conduct an ablation study on BraTS dataset. As shown in Table 2, removing the Cross Spatial Mamba (CS-Mamba) branch, Cross Fourier Mamba (CFMamba) branch, or FALF individually leads to performance degradation. Introducing CSMamba alone brings improvements (+0.73 dB

TABLE II
ABLATION STUDY ON THE BRATS DATASET WITH DIFFERENT SR FACTORS

CSMamba	CFMamba	FALF	2×		4×	
			PSNR	SSIM	PSNR	SSIM
✗	✗	✗	40.52	0.9838	34.89	0.9565
✓	✗	✗	41.25	0.9850	35.45	0.9611
✓	✓	✗	41.92	0.9878	35.97	0.9628
✓	✓	✓	42.21	0.9884	36.35	0.9651

PSNR) in $2\times$ scaling factor, while combining both CSMamba and CFMamba modules further boosts performance (+0.67 dB PSNR). Incorporating FALF ultimately enhances results to 42.21dB. Similar trends are observed under $4\times$ scaling. The results show that removing any one of them led to a significant decrease in both PSNR and SSIM values. This indicates that these modules play a crucial role in improving the super-resolution performance of our model, and their contribution to the network's overall effectiveness are indispensable.

V. CONCLUSION

In this paper, we present DDHMamba, a novel framework to effectively model long-range dependencies within MRI data via dual-domain learning. Specifically, the HSFM module introduces two innovative cross-modal scanning methods based on the Hilbert curve to synergistically fuse complementary features from both domains. Furthermore, we design the FALF module to enhance the cross-modal integration of local texture details through adaptive feature aggregation. Extensive experiments on the BraTS and fastMRI datasets demonstrate our framework's state-of-the-art performance across multiple super-resolution scaling factors.

REFERENCES

[1] R. Reda, A. Zanza, A. Mazzoni, A. Cicconetti *et al.*, "An update of the possible applications of magnetic resonance imaging (mri) in dentistry: a literature review," *Journal of imaging*, vol. 7, no. 5, p. 75, 2021.

[2] F. Shi, J. Cheng, L. Wang, P.-T. Yap *et al.*, "Lrtv: Mr image super-resolution with low-rank and total variation regularizations," *IEEE transactions on medical imaging*, vol. 34, no. 12, pp. 2459–2466, 2015.

[3] Q. Lyu, H. Shan, C. Steber, C. Helis *et al.*, "Multi-contrast super-resolution mri through a progressive network," *IEEE transactions on medical imaging*, vol. 39, no. 9, pp. 2738–2749, 2020.

[4] E. Tsiligianni, M. Zerva, I. Marivani, N. Deligiannis *et al.*, "Interpretable deep learning for multimodal super-resolution of medical images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2021, pp. 421–429.

[5] C.-M. Feng, H. Fu, S. Yuan, and Y. Xu, "Multi-contrast mri super-resolution via a multi-stage integration network," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24*. Springer, 2021, pp. 140–149.

[6] G. Li, C. Rao, J. Mo, Z. Zhang *et al.*, "Rethinking diffusion model for multi-contrast mri super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 11 365–11 374.

[7] W. Wang, H. Shen, J. Chen, and F. Xing, "Mhan: Multi-stage hybrid attention network for mri reconstruction and super-resolution," *Computers in Biology and Medicine*, vol. 163, p. 107181, 2023.

[8] Y. Mao, L. Jiang, X. Chen, and C. Li, "Disc-diff: Disentangled conditional diffusion model for multi-contrast mri super-resolution," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 387–397.

[9] G. Yang, L. Zhang, A. Liu, X. Fu *et al.*, "Mgdun: An interpretable network for multi-contrast mri image super-resolution reconstruction," *Computers in Biology and Medicine*, vol. 167, p. 107605, 2023.

[10] J. Shi, Q. Liu, C. Wang, Q. Zhang *et al.*, "Super-resolution reconstruction of mr image with a novel residual learning network algorithm," *Physics in Medicine & Biology*, vol. 63, no. 8, p. 085011, 2018.

[11] C.-Y. Shin, Y.-P. Chao, L.-W. Kuo, Y.-P. E. Chang *et al.*, "Improving the brain image resolution of generalized q-sampling mri revealed by a three-dimensional cnn-based method," *Frontiers in Neuroinformatics*, vol. 17, p. 956600, 2023.

[12] L. Xiang, Y. Chen, W. Chang, Y. Zhan *et al.*, "Deep-learning-based multi-modal fusion for fast mr reconstruction," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 7, pp. 2105–2114, 2018.

[13] C.-M. Feng, Y. Yan, G. Chen, Y. Xu *et al.*, "Multimodal transformer for accelerated mr imaging," *IEEE Transactions on Medical Imaging*, vol. 42, no. 10, pp. 2804–2816, 2022.

[14] G. Li, J. Lv, Y. Tian, Q. Dou *et al.*, "Transformer-empowered multi-scale contextual matching and aggregation for multi-contrast mri super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 20 636–20 645.

[15] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," *arXiv preprint arXiv:2312.00752*, 2023.

[16] L. Zhu, B. Liao, Q. Zhang, X. Wang *et al.*, "Vision mamba: Efficient visual representation learning with bidirectional state space model," *arXiv preprint arXiv:2401.09417*, 2024.

[17] Y. Shi, B. Xia, X. Jin, X. Wang *et al.*, "Vmambair: Visual state space model for image restoration," *IEEE Transactions on Circuits and Systems for Video Technology*, 2025.

[18] J. Zou, L. Liu, Q. Chen, S. Wang *et al.*, "Mmr-mamba: Multi-modal mri reconstruction with mamba and spatial-frequency information fusion," *arXiv preprint arXiv:2406.18950*, 2024.

[19] D. Li, Y. Liu, X. Fu, S. Xu *et al.*, "Fouriermamba: Fourier learning integration with state space models for image deraining," *arXiv preprint arXiv:2405.19450*, 2024.

[20] Q. Chen, X. Xing, Z. Chen, and Z. Xiong, "Accelerated multi-contrast mri reconstruction via frequency and spatial mutual learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2024, pp. 56–66.

[21] Z. Ji, B. Zou, X. Kui, S. Thureau *et al.*, "Global and local mamba network for multi-modality medical image super-resolution," *Pattern Recognition*, p. 112888, 2025.

[22] F. Gao, X. Jin, X. Zhou, J. Dong *et al.*, "Msfmamba: Multi-scale feature fusion state space model for multi-source remote sensing image classification," *arXiv preprint arXiv:2408.14255*, 2024.

[23] X. He, K. Cao, J. Zhang, K. Yan *et al.*, "Pan-mamba: Effective pan-sharpening with state space model," *Information Fusion*, vol. 115, p. 102779, 2025.

[24] H. Wu, Y. Yang, H. Xu, W. Wang *et al.*, "Rainmamba: Enhanced locality learning with state space models for video deraining," in *Proceedings of the 32nd ACM International Conference on Multimedia*, 2024, pp. 7881–7890.

[25] Y. Xiao, Q. Yuan, K. Jiang, Y. Chen *et al.*, "Frequency-assisted mamba for remote sensing image super-resolution," *arXiv preprint arXiv:2405.04964*, 2024.

[26] Y. Zhang, K. Li, K. Li, L. Wang *et al.*, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 286–301.

[27] U. Baid, S. Ghodasara, S. Mohan, M. Bilello *et al.*, "The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification," *arXiv preprint arXiv:2107.02314*, 2021.

[28] J. Zbontar, F. Knoll, A. Sriram, T. Murrell *et al.*, "fastmri: An open dataset and benchmarks for accelerated mri," *arXiv preprint arXiv:1811.08839*, 2018.

[29] X. Zhao, Y. Zhang, T. Zhang, and X. Zou, "Channel splitting network for single mr image super-resolution," *IEEE transactions on image processing*, vol. 28, no. 11, pp. 5649–5662, 2019.

[30] S. Huang, J. Li, L. Mei, T. Zhang *et al.*, "Accurate multi-contrast mri super-resolution via a dual cross-attention transformer network," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 313–322.