Image Content Matters: An Image Content Aware State Space Model for Accelerated MRI Reconstruction

Yucong Meng[1 2], **Zhiwei Yang**[1 2 3], **Kexue Fu**[4], **Zhijian Song**[1 2*], **Yonghong Shi**[1 2*]

[1]Digital Medical Research Center, School of Basic Medical Science, Fudan University, Shanghai 200032, China
[2]Shanghai Key Laboratory of Medical Image Computing and Computer Assisted Intervention, Shanghai 200032, China
[3]Academy for Engineering and Technology, Fudan University, Shanghai 200433, China
[4] Shandong Computer Science Center, Qilu University of Technology (Shandong Academy of Sciences), 250101 Jinan, China
{ ycmeng21, zwyang21 } @m.fudan.edu.cn, {fukx@sdas.org} { yonghong.shi, zjsong } @fudan.edu.cn

Abstract

The challenge of accelerated MRI reconstruction lies in recovering high-quality images from undersampled k-space. Recently, the selective state space model (Mamba) has shown promising results in various tasks with balanced global receptive field and computational efficiency, shedding new light on MRI reconstruction. However, existing approaches directly flatten 2D images based on spatial positions and apply Mamba to vision tasks, failing to preserve and explore the content properties. In this paper, we posit that the key to unlocking Mamba's full potential for MRI reconstruction lies in content-aware sequence modeling. We investigate two fundamental challenges: (1) how to reasonably preserve semantic information when converting 2D images into 1D sequences, and (2) how to effectively identify and recover the crucial high-frequency textures. To this end, we introduce CAM, a novel framework that shifts Mamba-based MRI reconstruction from position-based to content-aware sequence modeling. Specifically, we introduce three modules: (1) the Semantic Preservation Scanning Module (SPSM) introduces learnable clustering centers to group similar pixels, establishing the semantic preserved sequence. (2) The Texture Extraction Scanning Module (TESM) acts as a differentiable local texture descriptor to estimate crucial high-frequency information, forming the texture emphasized sequence. (3) The Texture Enhancement Mamba Module (TEMM) further modulates the semantic sequence with texture-informed system matrices derived from the texture sequence, yielding both context- and texture-aware sequential representations. With these enhancements, CAM significantly outperforms existing methods across various datasets and under-sampling masks.

Introduction

MRI is a widely used diagnostic tool that provides detailed visualization of anatomical structures, but its long acquisition time can cause patient discomfort and motion artifacts. To accelerate MRI, (1) Parallel Imaging (PI) (Deshmane et al. 2012) uses multiple coils to collect signals, and (2) Compressed Sensing (CS) (Donoho 2006) undersamples the k-space and reconstructs high-quality MRI images. CS, as a more economical approach that is less dependent on scanner hardware, has attracted increasing interest (Guo et al. 2024; Meng et al. 2025a). In this paper, we focus on CS paradigm.

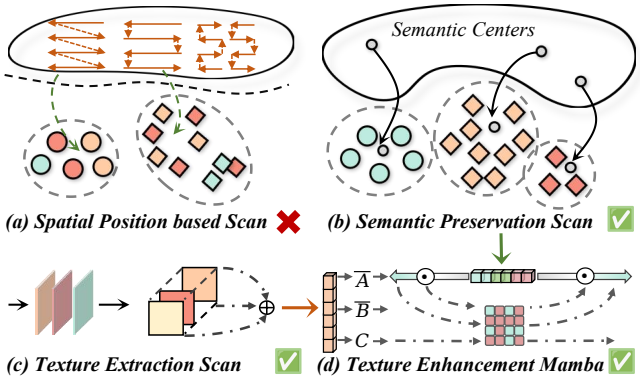

Figure 1: (a) Existing methods flatten images based on pixel positions, neglecting the preservation of content properties. We propose CAM to explicitly construct semantic preserved and texture highlighted sequences through (b) semantic preservation scanning and (c) texture extraction scanning, and combine them via our (d) texture Enhancement Mamba.

In recent years, various deep learning architectures like Convolutional Neural Networks (CNNs) (Zeng et al. 2020) and Vision Transformers (ViTs) (Huang et al. 2022; Guo et al. 2024) have been widely used for MRI reconstruction. However, CNNs' performance is limited by constrained receptive field while ViTs' quadratic complexity of self-attention creates a dilemma between global modeling and computational efficiency (Yang et al. 2024b, 2025a).

Recently, Mamba (Gu and Dao 2023) has emerged as a powerful alternative, offering new insights into these challenges. With near-linear scaling complexity, Mamba excels at long-sequence tasks and has been extended to vision tasks (Liu, Zhang, and Zhang 2024). However, Mamba is inherently a causal modeling approach, where each token can only capture information from its preceding tokens. This is at odds with the non-causal nature of the 2D spatial pattern in images. Traditional flattening methods disrupt the natural spatial dependencies within the image, weakening the model's ability to accurately interpret spatial relationships.

*Corresponding author.

To address this, recent research concentrates on effective image unfolding as shown in Figure 1 (a). VMamba (Liu et al. 2024) introduces SS2D to scan images in both horizontal and vertical directions. LocalMamba (Huang et al. 2025) divides images into distinct windows to capture local dependencies. MaIR (Li et al. 2025a) further proposes to adopt S-shaped scanning within and across the windows, thus preserving both locality and continuity. However, existing methods primarily focus on spatial relationships while overlooking the underlying content information, i.e., semantic dependencies and high-frequency textures among pixels. As a result, they fail to preserve the both related tokens in scanned sequences and texture details of MRI images, bottle-necking the performance of MRI reconstruction.

In this work, we propose Content Aware Mamba (CAM) to efficiently utilize content properties for MRI reconstruction. Specifically, we identify two key content-dependent characteristics within MRI images: (1) Similar pixels are often distantly located across entire MRI image. However, these related tokens is particularly useful to enhance mutual information complementarity. (2) MRI images contain crucial high-frequency textures that represent key anatomical details. Despite being challenging to capture and recover, these textures are essential for clinical diagnosis.

To leverage property (1), thus facilitating information complementarity among helpful pixels, we propose Semantic Preservation Scanning Module (SPSM) as Figure 1 (b). Instead of predefining scanning routes based on pixel positions, SPSM dynamically establishes a content-aware scanning path guided by a set of learnable semantic centers. These centers cluster related pixels and arrange the corresponding image tokens adjacently in the 1D sequence. Considering the low contrast and inherent ambiguity of MRI images, we design a contrastive loss to further enhance the discrimination among these centers. It effectively facilitates the ambiguous pixel grouping and guarantees semantic correlations in the scanned sequence, thereby supporting more informative global modeling for MRI images.

To consider property (2), thereby mining fine-grained textures from MRI images, we propose the Texture Extraction Scanning Module (TESM) as Figure 1 (c). Inspired by the traditional Local Binary Pattern (LBP) operators (Zhao et al. 2015), TESM applies differential convolutional kernels to compare each pixel with its neighbors, generating feature maps that capture local texture distributions. These features are then combined into a texture map by introducing encoding weights. This effectively captures rich textures within MRI images, forming sequence that highlights high-frequency details for accurate MRI reconstruction.

With the semantic-aware and texture-aware sequences generated by SPSM and TESM, we further integrate them through a Texture Enhancement Mamba Module (TEMM). As shown in Figure 1(d), the texture sequence is used to generate the Mamba system matrices $\overline{\mathbf{A}}$, $\overline{\mathbf{B}}$, and $\mathbf{C}$. These matrices serve as distribution shifts to modulate the semantic sequence with finer textures. In this way, high-frequency details are effectively injected, yielding both context- and texture-aware sequential representations.

The main contributions of our work are listed as follows:

- We introduce a novel approach, CAM, which shifts Mamba-based MRI reconstruction from position-based to content-aware sequence modeling.
- We explore two key content properties of MRI images, i.e., semantic relevance and texture structures, and introduce Semantic Preservation Scanning Module (SPSM) and Texture Extraction Scanning Module (TESM) to generate semantic- and texture-aware 1D sequences.
- We propose the Texture Enhancement Mamba Module (TEMM) to enhance semantic sequence modeling by injecting high-frequency components from texture sequences, thereby improving texture details and yielding more accurate reconstruction results.
- Extensive experiments on both single-coil and multi-coil datasets under various undersampling patterns show the superiority of our CAM over other competitors.

Related Work

MRI Reconstruction

Deep learning has gained attention for CS-based MRI reconstruction (Meng et al. 2024; Ye 2019). Methods like UNet (Zbontar et al. 2018), D5C5 (Schlemper et al. 2018), and DCRCN (Aghabiglou 2021) exploit CNNs' nonlinearity to restore images from low-quality inputs. However, the limited receptive field of convolutional layers restricts the potential of these studies (Khan et al. 2020; Wang et al. 2023).

In contrast to CNNs, ViT utilizes the self-attention to model global relationships (Yang et al. 2024a, 2025b). This capability has garnered considerable attention for ViT-based MRI reconstruction. ReconFormer (Guo et al. 2024) introduces a local pyramid but global columnar structure, incorporating multi-scale information into ViT for improved MRI reconstruction. FPS-Former (Meng et al. 2025a) enhances ViT-MRI from the perspectives of high-frequency capture, spatial purification, and multi-scale modeling. However, the quadratic computational complexity of ViT limits its ability to efficiently model images at high resolutions, thus hindering the performance of MRI reconstruction (Ali et al. 2023).

Vision Mamba

Mamba is emerging as a linear-complexity alternative to ViTs (Xu et al. 2024; Peng et al. 2025; Meng et al. 2025b; Wang et al. 2025). However, its 1D sequential processing requires flattening images, which disrupts spatial relationships. Existing adaptations for 2D data (e.g., VMamba (Liu et al. 2024), LocalMamba(Huang et al. 2025), and MaIR (Li et al. 2025a)) mitigate this by developing various scanning strategies. MambaRecon (Korkmaz 2025) employs state space model as its core framework. MMR-Mamba (Zou et al. 2025) uses Mamba for multi-modal MRI reconstruction while exploiting k-space information.

However, all these approaches rely on a fixed, spatially-defined scanning path, rendering them agnostic to the image's underlying content. This overlooks a key opportunity: guiding Mamba's powerful long-range modeling with essential image properties. In this paper, we introduce content-aware Mamba, which leverages image content to unlock the full potential of SSM for high-fidelity MRI reconstruction.

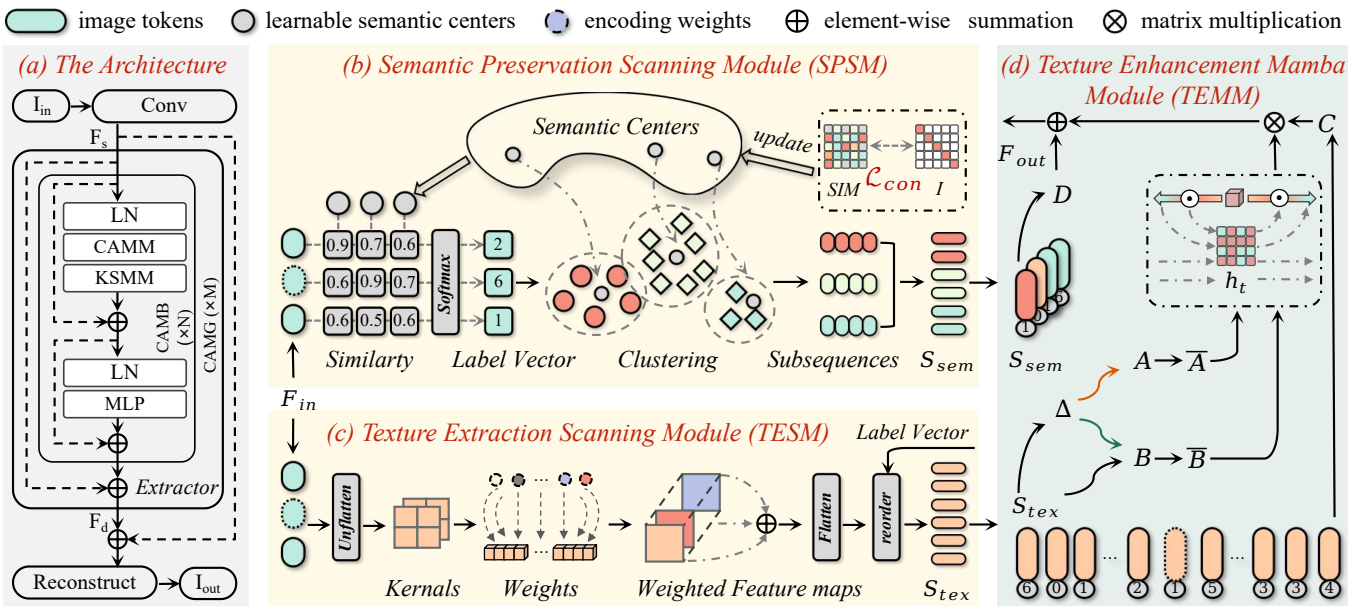

Figure 2: (a) CAM Architecture. Given input I_{in}, we first apply convolution to obtain F_s. In the backbone $Extractor(\cdot)$, we stack multiple CAMGs, each containing recurrent CAMBs. The core of CAMB is the Content-Aware Mamba Module (CAMM), which comprises SPSM (b), TESM (c), and TEMM (d). Besides, we introduce a K-space Mamba Module (KSMM) to enable iterative recovery across image and k-space domains. Finally, F_d from $Extractor(\cdot)$ are fused with F_s to obtain I_{out}.

Methodology

Overall Pipeline

Building upon previous work (Li et al. 2025a), our CAM is structured into three main stages: shallow feature extraction , deep feature extraction, and high-quality reconstruction, as illustrated in Figure 2 (a). Specifically, given input $I_{in} \in \mathbb{R}^{H\times W\times D}$ with resolution $H\times W$ and channel dimension D, we first apply a 3×3 convolution to extract shallow features:

$$F_s = conv(I_{in}). \tag{1}$$

Next, F_s is flattened into $F_f \in \mathbb{R}^{L\times D}$, where $L = H \times W$. Then, we feed F_f into the proposed deep feature extractor $Extractor(\cdot)$, which includes M stacked Content Aware Mamba Groups (CAMGs). Each CAMG consists of N Content Aware Mamba Blocks (CAMBs). Within each CAMB, a Content Aware Mamba Module (CAMM) is introduced to integrate content information with Mamba. Following CAMM, we further design a K-space Mamba Module (KSMM) to process the k-space data, enabling iterative reconstruction across both image and k-space domains.

Mathematically, given the input features f_n for n-th CAMB, its processing is formulated as:

$$\begin{aligned} f_n^{'} &= \alpha \cdot f_n + KSMM(CAMM(LN(f_n))), \\ f_{n+1} &= \beta \cdot f_n^{'} + MLP(LN(f_n^{'})), \end{aligned} \tag{2}$$

where $LN(\cdot)$ is the layer normalization, MLP is the multilayer perceptron, α and $\beta \in \mathbb{R}^C$ are learnable factors to control the skip connection. Finally, we generate the output feature F_d from the $Extractor(\cdot)$ and fuse it with F_s, which then go through a reconstruction head implemented by convolution layers, to generate the high-quality image I_{out}.

Content Aware Mamba Module (CAMM)

As shown in Figure 2, to explore content information within the image and thereby strengthen semantic guidance, CAMM enhances Mamba-based methods through two key strategies: preservation of semantics of relevant pixels and enhancement of critical texture features. Specifically, the training pipeline of CAMM is generalized as follows.

Semantic Preservation Scanning Module (SPSM) Due to Mamba's causal nature, each token can only access information from its preceding ones, and the interactions of distant tokens are diminished due to the long-range forgetting (Shi, Dong, and Xu 2024). However, in MRI reconstruction, distant yet semantically related tokens are crucial. Prior methods flatten images based purely on spatial positions, ignoring semantic correlations and increasing the distance between similar tokens, which leads to suboptimal reconstruction performance. To address this, we propose the Semantic Preservation Scanning Module (SPSM) to strengthen the interaction among semantically related tokens for information complementarity as Figure 2 (b).

Specifically, given the flattened feature $F_{in} \in \mathbb{R}^{L\times D}$, we first propose K learnable semantic centers $C \in \mathbb{R}^{K\times D}$ to group tokens with similar semantics. Then, we compute the similarity matrix between F_{in} and C. The softmax is applied to normalize it into a probability distribution $P \in \mathbb{R}^{L\times K}$ as:

$$P = softmax(F_{in} \times C^{\mathsf{T}}), \tag{3}$$

Based on $P \in \mathbb{R}^{L\times K}$, we assign each token of F_{in} a semantic label, resulting in a label vector $\mathcal{L} \in \mathbb{R}^{L}$ as follows:

$$\mathcal{L} = \{\mathcal{L}_i\}_{i=1}^{L}, \mathcal{L}_i = argmax_k P_i[k], \tag{4}$$

where $\mathcal{L}_i$ is the cluster index of the i-th token F_{in}^i in F_{in}. Then, we define a subsequence $S_k \in \mathbb{R}^{n_k \times D}$ to group tokens with the same cluster index k, where n_k is the token number of each group. The subsequence is defined as:

$$S_k = \{F_{in}^i \mid L_i = k\}, k \in \{1, 2, \ldots, K\}. \tag{5}$$

Finally, we combine all subsequences according to the cluster index k and concatenate ($[\cdot]$) it with semantic centers C, generating semantic preserved sequence $\mathcal{S}_{sem}$ as:

$$\begin{aligned} S &= [S_{k \in \{1,2,\ldots,K\}}], \\ \mathcal{S}_{sem} &= [S, C] \in \mathbb{R}^{(L+K) \times D}. \end{aligned} \tag{6}$$

Texture Extraction Scanning Module (TESM) Recovering high-frequency textures has always been a critical yet challenging objective in MRI reconstruction. Existing networks constrained by pixel-wise losses mainly focus on stable edge details, failing to recover fine-grained textures. Motivated by traditional Local Binary Pattern (LBP) operators (Zhao et al. 2015), we design a Texture Extraction Scanning Module (TESM) to explicitly capture texture features.

Specifically, the traditional LBP operator encodes the local texture of an image by thresholding a pixel's neighborhood against its center value. For a given pixel I_c and its neighboring pixels $\{I_p\}_{p=0}^{7}$ within a 3×3 local region, the LBP value v is computed using a sign function $s(x)$ as:

$$v = \sum_{p=0}^{7} s(I_p - I_c) \cdot 2^p, s(x) = \begin{cases} 1, & \text{if } x \geq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Motivated by LBP, we design differentiable TESM to explicitly capture fine-grained texture sequence $\mathcal{S}_{tex}$ as shown in Figure 2 (c). Specifically, given input F_{in}, we first unflatten it to obtain $M_{in} \in \mathbb{R}^{H \times W \times D}$. Next, instead of using hard thresholding, our TESM first computes the differences between each pixel and its neighbors using 3×3 convolution kernels w_i. Then, we multiply the obtained features maps with designed encoding weights $2^{i \in \{0,1,2,\ldots,7\}}$. The weighted maps are summed to obtain the texture map $\mathcal{T}_{\mathcal{M}}$:

$$\mathcal{T}_{\mathcal{M}} = IN\left[\sum_{i=0}^{7} \frac{2^i}{255}\left(\frac{w_i * M_{in} + 1}{2}\right)\right], \tag{8}$$

here, kernel w_i contains a central value of 1 and one neighbor value of -1, and the remaining values are all 0. $IN(\cdot)$ denotes the instance normalization. Finally, we flatten $\mathcal{T}_{\mathcal{M}}$ to obtain feature $F_{tex} \in \mathbb{R}^{L \times D}$ and use the label vector $\mathcal{L}$ from SPSM to reorder it. Thus, we obtain texture enhanced sequence $\mathcal{S}_{tex}$ that aligns with semantic preserved $\mathcal{S}_{sem}$:

$$\begin{aligned} S_k^{'} &= \{F_{tex}^i \mid L_i = k\}, k \in \{1, 2, \ldots, K\}, \\ S &= [S_{k \in \{1,2,\ldots,K\}}^{'}], \\ \mathcal{S}_{tex} &= [S, C] \in \mathbb{R}^{(L+K) \times D}. \end{aligned} \tag{9}$$

Texture Enhancement Mamba Module (TEMM) With extracted semantic aware sequence $\mathcal{S}_{sem}$ and texture aware feature $\mathcal{S}_{tex}$, we propose Texture Enhancement Mamba Module (TEMM) to transfer high-frequency details from $\mathcal{S}_{tex}$ to semantically rich $\mathcal{S}_{sem}$, as shown in Figure 2 (d).

Algorithm 1: Texture Enhancement Mamba Module

Inputs: $\mathcal{S}_{\mathbf{sem}}, \mathcal{S}_{\mathbf{tex}} : (L, D)$
Output: $F_{out} : (L, D)$
$\mathbf{A} : (D, N) \leftarrow \mathbf{Parameter}_{\mathrm{A}}$ /* N is the state size */
$\mathbf{B} : (L, N) \leftarrow \mathbf{Linear}_{\mathrm{B}}(\mathcal{S}_{\mathbf{tex}})$
$\mathbf{C} : (L, N) \leftarrow \mathbf{Linear}_{\mathrm{C}}(\mathcal{S}_{\mathbf{tex}})$
$\mathbf{\Delta} : (L, D) \leftarrow \log(1 + \exp(\mathbf{Linear}_{\Delta}(\mathcal{S}_{\mathbf{tex}}) + \mathbf{Param}._{\Delta}))$
$\overline{\mathbf{A}}, \overline{\mathbf{B}} : (L, D, N) \leftarrow \mathrm{discretize}(\mathbf{\Delta}, \mathbf{A}, \mathbf{B})$
$F_{out} \leftarrow \mathrm{SSM}(\overline{\mathbf{A}}, \overline{\mathbf{B}}, \mathbf{C})(\mathcal{S}_{\mathbf{sem}})$
/* SSM represents Eq. 10 implemented using selective scan */
return F_{out}

First, let's briefly review the Selective State Space Model (S6) of Mamba. The vanilla S6 capture long-range dependencies via an intermediate latent state $h(t) \in \mathbb{R}^N$ as:

$$\begin{aligned} h(t) &= \overline{\mathbf{A}} h(t-1) + \overline{\mathbf{B}} x(t), \\ y(t) &= \mathbf{C} h(t) + \mathbf{D} x(t), \end{aligned} \tag{10}$$

where $x(t)$, $y(t) \in \mathbb{R}^{L \times D}$ are input and output sequences. The system matrices $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$ discretized by a timescale parameter $\mathbf{\Delta}$, drive the state evolution. $\mathbf{C}$ governs the output mapping, and $\mathbf{D}$ represents the skip connection. In vanilla S6, to make the matrices input-dependent, $\overline{\mathbf{B}}$, $\mathbf{C}$, and $\mathbf{\Delta}$ are all calculated from the input $x(t)$ via linear projection.

Building on this, our Texture Enhancement Mamba Module (TEMM) proposes to inject the high-frequency details from $\mathcal{S}_{tex}$ to $\mathcal{S}_{seq}$ via modulating the system matrices $\overline{\mathbf{A}}$, $\overline{\mathbf{B}}$, and $\mathbf{C}$, as detailed in Algorithm 1. Specifically, because the content-aware sequence $\mathcal{S}_{seq}$ contains rich semantic information, we let it serve as the primary driver of the SSM global modeling. Meanwhile, instead of generating matrices from the main input $\mathcal{S}_{seq}$, we apply linear projections to the texture sequence $\mathcal{S}_{tex}$, obtaining $\overline{\mathbf{A}}$, $\overline{\mathbf{B}}$, and $\mathbf{C}$. This design allows texture feature to influence the state evolution and output mapping of $\mathcal{S}_{seq}$, thereby enriching high-frequency details in the restored images. In this way, our TEMM not only learns the 2D dependencies with rich semantic information, but also enhances high-frequency details to facilitate accurate and detail-preserving MRI reconstruction.

K-space Mamba Module (KSMM)

We propose the K-space Mamba Module (KSMM) that models the k-space for precise recovery. Specifically, given the output sequence $F_{out} \in \mathbb{R}^{L \times D}$ processed by CAMM, we first apply the inverse transformations of Eq. 5 and Eq. 6 to unflatten it back into feature map $F_f \in \mathbb{R}^{H \times W \times D}$. Then, we apply the Fast Fourier Transform (FFT) to F_f, obtaining the k-space spectrum $F_s \in \mathbb{R}^{H \times W \times D}$ and modeling the global dependency as follows:

$$F_s = FFT(F_f), F_s^{'} = SSM(flatten(F_s)), \tag{11}$$

where $flatten(\cdot)$ denotes the traditional image flattening, $SSM(\cdot)$ is the vanilla SSM implemented via Eq. 10. Finally, we obtain the output feature K_{out} via unflattening $unflatten(\cdot)$ and Inverse Fast Fourier Transform (IFFT) as:

$$K_{out} = IFFT(unflatten(F_s^{'})) \tag{12}$$

Method	Type	CC359 (Brain)						fastMRI (Knee)					
		NMSE↓		SSIM↑		PSNR↑		NMSE↓		SSIM↑		PSNR↑	
		AF=4	AF=8	AF=4	AF=8	AF=4	AF=8	AF=4	AF=8	AF=4	AF=8	AF=4	AF=8
KIKI-Net (Eo et al. 2018)	$\mathcal{C}$	0.0221	0.0417	0.8415	0.7773	28.97	26.24	0.0353	0.0546	0.7172	0.6355	31.87	29.27
UNet-32 (Zbontar et al. 2018)		0.0197	0.0385	0.8898	0.8348	31.54	28.66	0.0337	0.0477	0.7248	0.6570	31.99	30.02
D5C5 (Schlemper et al. 2018)		0.0177	0.0428	0.8977	0.8267	31.59	28.20	0.0332	0.0512	0.7256	0.6457	32.25	29.65
DCRCN (Aghabiglou 2021)		0.0119	0.0291	0.9100	0.8649	32.01	29.49	0.0351	0.0443	0.7332	0.6635	32.18	30.76
SwinMR (Huang et al. 2022)	$\mathcal{T}$	0.0109	0.0260	0.9298	0.8695	34.14	30.36	0.0342	0.0476	0.7213	0.6537	32.14	30.21
ReconFormer (Guo et al. 2024)		0.0108	0.0276	0.9297	0.8650	34.16	30.11	0.0320	0.0431	0.7327	0.6672	32.53	30.76
FPS-Former (Meng et al. 2025a)		0.0103	0.0217	0.9321	0.8828	34.38	31.15	0.0316	0.0408	0.7337	0.6692	32.51	31.03
HQS-Net (Xin et al. 2022)	$\mathcal{U}$	0.0117	0.0276	0.9370	0.8765	33.80	30.11	0.0451	0.0485	0.7319	0.6582	32.38	30.28
PDAC (Wang et al. 2024)		0.0103	0.0243	0.9431	0.8894	34.36	30.68	0.0315	0.0420	0.7356	0.6698	32.58	31.06
MaIR (Li et al. 2025a)	$\mathcal{M}$	0.0110	0.0269	0.9405	0.8807	34.09	30.22	0.0319	0.0443	0.7323	0.6587	32.47	30.55
MambaRecon (Korkmaz 2025)		0.0065	0.0230	0.9524	0.9001	35.28	31.14	0.0316	0.0421	0.7352	0.6679	32.55	30.89
LMO (Li et al. 2025b)		0.0104	0.0237	0.9430	0.8936	34.35	30.79	0.0428	0.0471	0.7328	0.6542	32.40	30.33
Ours		**0.0034**	**0.0134**	**0.9731**	**0.9272**	**39.37**	**33.47**	**0.0310**	**0.0406**	**0.7428**	**0.6719**	**32.68**	**31.15**

Table 1: Performance comparison under $4\times$ and $8\times$ Acceleration Factors (AF) on the single-coil datasets, including CC359 and fastMRI. $\mathcal{C}$: CNN-based methods. $\mathcal{T}$: transformer-based methods. $\mathcal{U}$: deep unfolding methods. $\mathcal{M}$: Mamba-based methods.

Method	SKM-TEA (Brain)					
	NMSE ↓		SSIM ↑		PSNR ↑	
	AF=4	AF=8	AF=4	AF=8	AF=4	AF=8
KIKI-Net	0.0196	0.0271	0.8577	0.7941	34.26	31.42
D5C5	0.0188	0.0257	0.8648	0.8030	34.63	31.89
ReconFormer	0.0179	0.0239	0.8730	0.8158	35.06	32.51
FPS-Former	0.0158	0.0200	0.8975	0.8527	35.64	32.85
PDAC	0.0157	0.0202	0.8952	0.8503	35.20	32.65
MambaRecon	0.0182	0.0249	0.8691	0.8116	34.78	32.25
MoDL	0.0195	0.0220	0.8711	0.8296	34.85	32.58
CAMP-Net	0.0156	0.0200	0.8995	0.8528	35.36	32.92
Ours	**0.0154**	**0.0199**	**0.9001**	**0.8532**	**35.66**	**32.95**

Table 2: Performance comparison on SKM-TEA dataset.

Method	Mask	NMSE ↓		SSIM ↑		PSNR ↑	
		AF=5	AF=10	AF=5	AF=10	AF=5	AF=10
ReconFormer	$\mathcal{I}$	0.0057	0.0136	0.9589	0.9247	36.91	33.16
FPS-Former		0.0056	0.0135	0.9593	0.9249	36.97	33.18
PDAC		0.0055	0.0130	0.9601	0.9255	37.34	33.51
MambaRecon		0.0049	0.0096	0.9544	0.9203	37.68	33.76
Ours		**0.0032**	**0.0065**	**0.9703**	**0.9539**	**39.51**	**36.34**
ReconFormer	$\mathcal{R}$	0.0110	0.0160	0.9331	0.9125	34.06	32.43
FPS-Former		0.0110	0.0160	0.9334	0.9137	34.08	32.47
PDAC		0.0106	0.0159	0.9351	0.9140	34.22	32.51
MambaRecon		0.0108	0.0156	0.9348	0.9144	34.17	32.56
Ours		**0.0081**	**0.0150**	**0.9444**	**0.9152**	**35.41**	**32.71**

Table 3: Performance of Radial ($\mathcal{I}$) and Random ($\mathcal{R}$) masks.

Training Objectives

Considering inherent ambiguity of MRI images, we design a contrastive loss $\mathcal{L}_{con}$ to further enhance the diversity among semantic centers C. Specifically, given the normalized C_j from the j-th CAMB, we first compute a self similarity matrix $SIM_j \in \mathbb{R}^{K\times K}$ by: $SIM_j = C_j \times (C_j)^{\mathsf{T}}$.

Next, we define an identity matrix $I \in \mathbb{R}^{K\times K}$ with 1 on the diagonal and 0s elsewhere, and compute a cross-entropy loss between SIM_j and I. This encourages each center to be self-similar and distinct from others, promoting the discrepancy of semantic centers and facilitating effective pixel grouping. The center-contrastive loss $\mathcal{L}_{con}$ is defined as:

$$\mathcal{L}_{con} = \frac{1}{J}\sum_{j=1}^{J} CrossEntropy(SIM_j, I), \tag{13}$$

where J is the number of CAMBs. In addition, we use L1-norm to ensure accurate MRI reconstruction: $\mathcal{L}_{\text{rec}} = \|I_{out} - I_{GT}\|_1$. Denoting the weight factor as γ, the overall loss is:

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \gamma\mathcal{L}_{\text{con}}. \tag{14}$$

Experiments

Experimental Settings

Datasets CAM is evaluated on CC359 (Warfield, Zou, and Wells 2004), fastMRI (Zbontar et al. 2018), and SKM-TEA (Desai et al. 2022). CC359 contains 35 raw brain MRI volumes. Following prior methods, $4,524$ slices from 25 subjects and $1,700$ slices from 10 subjects are used for training and testing; The fastMRI dataset includes $1,172$ single-coil knee MRI scans, in which 973 scans are used for training and 199 scans are used for testing; The SKM-TEA is a raw multi-coil T2 knee MRI dataset. Following (Guo et al. 2024; Meng et al. 2025a), $124, 10$, and 21 volumes are used for training, validation, and testing, respectively.

Implementation Details Please refer to *Supplementary Materials* for our implementation details.

Comparisons with State-of-the-art Methods

Single-coil datasets Table 1 reports the comprehensive results conducted on two single-coil datasets. Key observations include: (1) CAM significantly outperforms CNN-based methods, demonstrating a 0.43 dB improvement over D5C5 under $4\times$ AF on fastMRI. (2) Recently reported transformer-based and deep unfolding methods show superior performance over CNNs. However, they suffer significant degradation in challenging conditions, such as $AF = 8$. (3) Building on Mamba's global receptive field, our CAM further effectively captures semantic dependencies

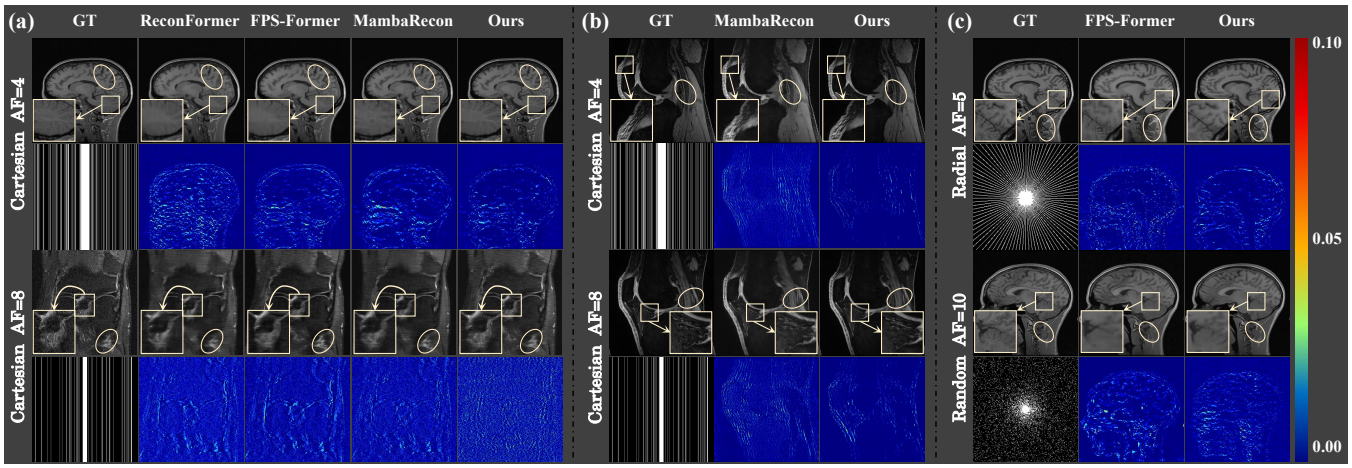

Figure 3: Comparison on (a) the single-coil dataset including CC359 and fastMRI, (b) the multi-coil dataset SKM-TEA, and (c) the CC359 dataset using different masks. The second row of each subplot shows the corresponding error maps.

Model	SPSM	TESM	TEMM	KSMM	SSIM↑	PSNR↑
(a)		✓	✓	✓	0.9677$_{\text{-0.0054}}$	38.20$_{\text{-1.17}}$
(b)	✓		✓	✓	0.9682$_{\text{-0.0049}}$	38.30$_{\text{-1.07}}$
(c)	✓	✓		✓	0.9687$_{\text{-0.0044}}$	38.43$_{\text{-0.94}}$
(d)	✓	✓	✓		0.9723$_{\text{-0.0008}}$	39.15$_{\text{-0.22}}$
Ours	✓	✓	✓	✓	**0.9731**	**39.37**

Table 4: Ablation results of CAM on CC359 dataset (AF=4).

Model	Module	SSIM↑	PSNR↑
Kmeans (Lloyd 1982)	SPSM	0.9701$_{\text{-0.0030}}$	38.79$_{\text{-0.58}}$
Without $\mathcal{L}_{con}$		0.9706$_{\text{-0.0025}}$	39.02$_{\text{-0.35}}$
Random noise	TESM	0.9511$_{\text{-0.0220}}$	35.71$_{\text{-3.66}}$
LBP (Zhao et al. 2015)		0.9696$_{\text{-0.0035}}$	38.59$_{\text{-0.78}}$
Ours		**0.9731**	**39.37**

Table 5: Ablations of SPSM and TESM on CC359 (AF=4).

K	None	2	4	**6**	8	10
SSIM ↑	0.9677	0.9682	0.9700	**0.9731**	0.9724	0.9712
PSNR ↑	38.20	38.36	38.74	**39.37**	39.20	38.96

Table 6: Ablation study of the semantic centers number K.

and high-frequency textures in MRI data, consistently surpassing competitors across various settings. For instance, CAM achieves PSNR of 39.37 and 33.47 on CC359 under $AF = 4$ and $AF = 8$, respectively, outperforming other methods by at least 4.09 dB and 2.32 dB.

Multi-coil datasets We further conduct comparisons on multi-coil SKM-TEA against several high-performing methods, additionally including two dedicated multi-coil approaches, MoDL (Aggarwal and Mani 2018) and CAMP-Net (Zhang and Li 2025). As shown in Table 2, CAM maintains superior performance in different settings. It achieves 35.66 and 32.95 PSNR at $4\times$ and $8\times$ AF respectively, outperforming all the other approaches.

Experiments on different masks Table 3 reports the comparison results using radial and random undersampling patterns under $5\times$ and $10\times$ acceleration factors. Evidently, our CAM again scores the highest, leading other competitors at least 1.83 and 1.19 dB under $5\times$ radial and random masks, respectively. This further validates the robustness of our CAM, highlighting its ability to effectively reconstruct MRI images from various undersampling masks.

Visualization Results Figure 3 presents the qualitative results. Our CAM excels at reconstructing high-quality images while exhibiting fewer visual artifacts on both single-coil and multi-coil datasets, outperforming all the other competitors. Thanks to our designs of semantic relation guided and high-frequency enhanced Mamba, CAM successfully recovers clear and rich anatomical details as highlighted by the zoomed-in boxes and ellipses. Furthermore, as shown in Figure 3 (c), CAM demonstrates strong robustness and accurate performance across various undersampling patterns and acceleration rates, further confirming its effectiveness.

Ablation Studies and Further Analysis

Efficacy of Key Components We conduct a breakdown ablation study, the results are shown in Table 4. (a) Replacing SPSM with SS2D (Liu et al. 2024) leads to a 1.17 dB drop in PSNR, indicating that SPSM effectively aggregates semantically related pixels in 1D sequences to enhance feature representation. (b) Excluding TESM and only using semantic sequence for vanilla SSM modeling results in a significant performance drop of 1.07 dB in PSNR and 0.0049 in SSIM, demonstrating its importance in extracting and injecting high-frequency details. (c) Removing TEMM and using the summation of semantic and texture sequences for vanilla SSM modeling reduces PSNR from 39.37 to 38.43 dB, confirming TEMM's effectiveness in injecting high-frequency textures. (d) Disabling KSMM causes a 0.22 dB drop, validating its crucial role in enhancing k-space recovery.

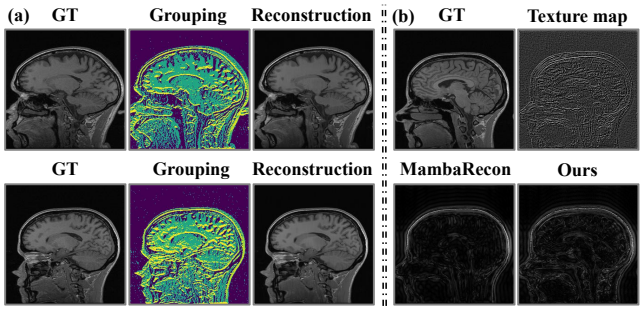

Figure 4: (a) Visualization of token grouping via SPSM. (b) Visualization of local texture extraction via TESM.

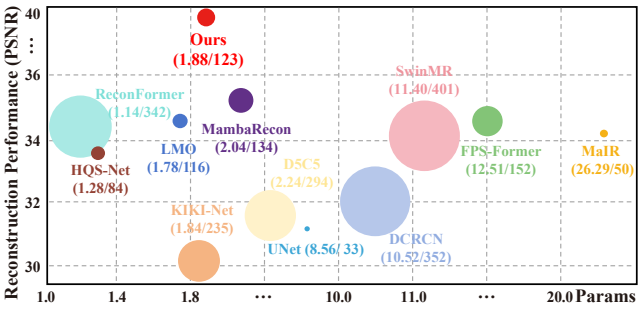

Figure 5: Comparison on PSNR and Params (M) / FLOPs (G), which is conducted on CC359 dataset under $4\times$ AF.

Efficacy of Semantic Preservation Scanning Module We further analyze the effectiveness of the proposed SPSM. (1) Comparison with Kmeans: As shown in Table 5, SPSM shows advantage over Kmeans. By using learnable centers, SPSM dynamically groups pixels, resulting in more accurate semantic-preserving representations. (2) Impact of $\mathcal{L}_{con}$: Without supervision from $\mathcal{L}_{con}$, PSNR drops significantly by 0.35, highlighting that $\mathcal{L}_{con}$ plays a crucial role in enhancing class discrimination. (3) Effect of varying K: As shown in Table 6, both overly small and excessively large K values degrade performance, with the best reconstruction achieved at $K = 6$. A too-small K fails to capture the diverse semantics, while a too-large K causes semantic fragmentation. (4) Visualization of token grouping: As Figure 4 (a), similar pixels are successfully grouped. Based on this, related tokens become close after reordering in the scanned sequence. This enhances the information complementarity among relevant tokens, improving the global representation.

Efficacy of Texture Extraction Scanning Module To further validate the effectiveness of TESM, we design two experimental settings, as shown in Table 5. (1) We replace the texture features extracted by TESM with random Gaussian noise and inject it into the Texture Enhancement Mamba Module (TEMM). This results in a significant performance drop of 3.66 dB, indicating that the benefits of TESM arise from its accurate modeling of local textures rather than the mere injection of high-frequency signals. (2) We compare TESM with traditional Local Binary Patterns (LBP). Benefiting from its learnable optimization, TESM outperforms the unsupervised LBP. (3) Figure 4 (b) visualizes the texture map extracted by TESM. Besides, we compare the high-frequency details of CAM with MambaRecon using a high-pass filter. TESM provides vivid textures, enabling CAM to better reconstruct rich details.

Efficacy of Texture Enhancement Mamba Module The core contribution of our TEMM lies in its information interaction mechanism. Considering that the system matrices $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$ govern state evolution, while $\mathbf{C}$ determines the output mapping, our TEMM injects texture information into all of them for more comprehensive feature interaction. Here, we evaluate alternative strategies in Table 7. We can find that: (1) our proposed setup achieves the best performance, confirming the importance of fully incorporating texture features during the entire process; (2) the greatest improvement is observed when injecting texture into $\mathbf{C}$, likely because $\mathbf{C}$ directly influences the output in the SSM as shown in Eq. 10, thereby having a stronger impact on the final results.

Interaction	$\overline{\mathbf{A}}$	$\overline{\mathbf{B}}$	$\mathbf{C}$	SSIM ↑	PSNR ↑
			✓	0.9632$_{-0.0099}$	37.36$_{-2.01}$
		✓	✓	0.9708$_{-0.0023}$	38.88$_{-0.49}$
	✓	✓	✓	0.9720$_{-0.0011}$	39.05$_{-0.32}$
	✓	✓	✓	**0.9731**	**39.37**

Table 7: Ablation study on interaction strategies of TEMM.

Analysis of Training Efficiency The training efficiency comparison is given in Figure 5, in which the radius of the circle denotes the metric of FLOPs. We can find that: (1) HQS-Net and LMO enjoy both low computational complexity and minimal number of parameters, but they obtain suboptimal results. (2) The lightweight ReconFormer maintains a small number of trainable parameters while suffering high complexity (FLOPs=342G). (3) Overall, our CAM achieves significant performance improvements with a favorable balance between parameter count and computational cost.

Analysis of Hyper-parameters The analysis of important hyper-parameters, such as the the number of CAMGs M, the number of CAMBs N, and the loss weight γ, etc., are specifically discussed in the *Supplementary Materials*. The results show that our CAM demonstrates consistent performance across different hyper-parameter variations.

Conclusion

In this work, we propose CAM, a novel Mamba based method to explore image content related properties for accurate and efficient MRI reconstruction. To this end, Semantic Preservation Scanning Module (SPSM) and Texture Extraction Scanning Module (TESM) are designed to construct semantic structure preserved sequence and texture information enhanced sequence, respectively. A Texture Enhancement Mamba Module (TEMM) is further proposed to effectively inject the high-frequency textures into the reconstructed MRI images during the state space modeling process. Extensive experiments and analysis are conducted on both single-coil and multi-coil datasets, demonstrating the superiority of our CAM over existing methods.

Acknowledgments

This work was funded by the 2025 Shanghai Explorer Program for Basic Research (25TS1411800), the National Natural Science Foundation of China (82072021, 62501340), the Taishan Scholars Program (TSQN202408245), the Shandong Provincial Natural Science Foundation (ZR2024QF110), the Linyi People's Hospital Horizontal Project (2024LYKC002), and the Key R&D Program of Shandong Province (2025CXPT110).

References

Aggarwal, H. K.; and Mani, M. P. 2018. MoDL: Model-based deep learning architecture for inverse problems. *IEEE transactions on medical imaging*, 38(2): 394–405.

Aghabiglou, A. 2021. MR image reconstruction using densely connected residual convolutional networks. *Computers in Biology and Medicine*, 139: 105010.

Ali, A. M.; Benjdira, B.; Koubaa, A.; El-Shafai, W.; Khan, Z.; and Boulila, W. 2023. Vision transformers in image restoration: A survey. *Sensors*, 23(5): 2385.

Desai, A. D.; Schmidt, A. M.; Rubin, E. B.; Sandino, C. M.; Black, M. S.; Mazzoli, V.; Stevens, K. J.; Boutin, R.; Ré, C.; Gold, G. E.; et al. 2022. Skm-tea: A dataset for accelerated mri reconstruction with dense image labels for quantitative clinical evaluation. *arXiv preprint arXiv:2203.06823*.

Deshmane, A.; Gulani, V.; Griswold, M. A.; and Seiberlich, N. 2012. Parallel MR imaging. *Journal of Magnetic Resonance Imaging*, 36(1): 55–72.

Donoho, D. L. 2006. Compressed sensing. *IEEE Transactions on information theory*, 52(4): 1289–1306.

Eo, T.; Jun, Y.; Kim, T.; Jang, J.; Lee, H.-J.; and Hwang, D. 2018. KIKI-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. *Magnetic resonance in medicine*, 80(5): 2188–2201.

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

Guo, P.; Mei, Y.; Zhou, J.; Jiang, S.; and Patel, V. M. 2024. ReconFormer: Accelerated MRI Reconstruction Using Recurrent Transformer. *IEEE Transactions on Medical Imaging*, 43(1): 582–593.

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, T.; Pei, X.; You, S.; Wang, F.; Qian, C.; and Xu, C. 2025. Localmamba: Visual state space model with windowed selective scan. In *European Conference on Computer Vision*, 12–22. Springer.

Khan, A.; Sohail, A.; Zahoora, U.; and Qureshi, A. S. 2020. A survey of the recent architectures of deep convolutional neural networks. *Artificial intelligence review*, 53: 5455–5516.

Korkmaz, Y. 2025. MambaRecon: MRI reconstruction with structured state space models. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 4142–4152. IEEE.

Li, B.; Zhao, H.; Wang, W.; Hu, P.; Gou, Y.; and Peng, X. 2025a. Mair: A locality-and continuity-preserving mamba for image restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 7491–7501.

Li, W.; Jiang, J.; Wu, J.; Yu, K.; and Zheng, J. 2025b. LMO: Linear Mamba Operator for MRI Reconstruction. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, 5112–5122.

Liu, X.; Zhang, C.; and Zhang, L. 2024. Vision mamba: A comprehensive survey and taxonomy. *arXiv preprint arXiv:2405.04404*.

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. arXiv:2401.10166.

Lloyd, S. 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2): 129–137.

Meng, Y.; Yang, Z.; Duan, M.; Shi, Y.; and Song, Z. 2024. Continuous k-space recovery network with image guidance for fast mri reconstruction. *arXiv preprint arXiv:2411.11282*.

Meng, Y.; Yang, Z.; Shi, Y.; and Song, Z. 2025a. 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*, volume 39, 6135–6143.

Meng, Y.; Yang, Z.; Song, Z.; and Shi, Y. 2025b. DM-Mamba: Dual-domain Multi-scale Mamba for MRI reconstruction. *arXiv e-prints*, arXiv–2501.

Peng, Z.; Xu, Z.; Liu, Q.; Yang, X.; and Shen, W. 2025. HyperET: Efficient Training in Hyperbolic Space for Multimodal Large Language Models. *Advances in Neural Information Processing Systems*.

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.

Shi, Y.; Dong, M.; and Xu, C. 2024. Multi-scale vmamba: Hierarchy in hierarchy visual state space model. *Advances in Neural Information Processing Systems*, 37: 25687–25708.

Wang, C.; Guo, L.; Wang, Y.; Cheng, H.; Yu, Y.; and Wen, B. 2024. Progressive Divide-and-Conquer via Subsampling Decomposition for Accelerated MRI. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 25128–25137.

Wang, S.; Zhu, Y.; Luo, X.; Yang, Z.; Zhang, Y.; Fu, P.; Wang, M.; Song, Z.; Li, Q.; Zhou, P.; et al. 2023. Knowledge extraction and distillation from large-scale image-text colonoscopy records leveraging large language and vision models. *arXiv preprint arXiv:2310.11173*.

Wang, S.; Zhu, Y.; Yang, Z.; Luo, X.; Zhang, Y.; Fu, P.; Wang, H.; Wang, M.; Song, Z.; Li, Q.; et al. 2025. Leveraging large language and vision models for knowledge extraction from large-scale image–text colonoscopy records. *Nature Biomedical Engineering*, 1–12.

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.

Xin, B.; Phan, T.; Axel, L.; and Metaxas, D. 2022. Learned half-quadratic splitting network for MR image reconstruction. In *International Conference on Medical Imaging with Deep Learning*, 1403–1412. PMLR.

Xu, R.; Yang, S.; Wang, Y.; Cai, Y.; Du, B.; and Chen, H. 2024. Visual mamba: A survey and new outlooks. *arXiv preprint arXiv:2404.18861*.

Yang, Z.; Fu, K.; Duan, M.; Qu, L.; Wang, S.; and Song, Z. 2024a. Separate and conquer: Decoupling co-occurrence via decomposition and representation for weakly supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3606–3615.

Yang, Z.; Meng, Y.; Fu, K.; Tang, F.; Wang, S.; and Song, Z. 2025a. Exploring CLIP's Dense Knowledge for Weakly Supervised Semantic Segmentation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 20223–20232.

Yang, Z.; Meng, Y.; Fu, K.; Wang, S.; and Song, Z. 2024b. Tackling Ambiguity from Perspective of Uncertainty Inference and Affinity Diversification for Weakly Supervised Semantic Segmentation. *ArXiv*, abs/2404.08195.

Yang, Z.; Meng, Y.; Fu, K.; Wang, S.; and Song, Z. 2025b. More: Class patch attention needs regularization for weakly supervised semantic segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 9400–9408.

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.; Murrell, T.; Huang, Z.; Muckley, M. J.; Defazio, A.; Stern, R.; Johnson, P.; Bruno, M.; et al. 2018. fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv preprint arXiv:1811.08839*.

Zeng, W.; Peng, J.; Wang, S.; and Liu, Q. 2020. A comparative study of CNN-based super-resolution methods in MRI reconstruction and its beyond. *Signal Processing: Image Communication*, 81: 115701.

Zhang, L.; and Li, X. 2025. CAMP-Net: Consistency-Aware Multi-Prior Network for Accelerated MRI Reconstruction. *IEEE journal of biomedical and health informatics*, 29(3): 2006–2019.

Zhao, Y.; Wang, R.; Wang, W.; and Gao, W. 2015. High resolution local structure-constrained image upsampling. *IEEE Transactions on Image Processing*, 24(11): 4394–4407.

Zou, J.; Liu, L.; Chen, Q.; Wang, S.; Hu, Z.; Xing, X.; and Qin, J. 2025. MMR-Mamba: Multi-modal MRI reconstruction with Mamba and spatial-frequency information fusion. *Medical Image Analysis*, 102: 103549.