GF-BrainSR: Gated Frequency-Aware Selective State Space Model-Based Brain MRI Image Super-resolution

Anonymized Authors

Anonymized Affiliations
email@anonymized.com

Abstract. Deep learning-based super-resolution (SR) mitigates resolution bottlenecks in brain MRI, yet existing CNNs and Transformers struggle to balance global context, computational efficiency, and high-frequency detail fidelity. We propose GF-BrainSR, an efficient linear-complexity State Space Model (SSM) framework. GF-BrainSR introduces a DWT-driven Frequency Adaptive Gate that explicitly decouples and enhances structural boundaries, while a depthwise spatial gate robustly suppresses noise in homogeneous soft tissues. Additionally, an Implicit Coordinate Upsampling module fuses continuous spatial coordinates with local features, eliminating discrete interpolation artifacts to reconstruct morphologically continuous tissue interfaces. Experiments demonstrate that GF-BrainSR achieves competitive performance, delivering superior anatomical clarity and spatial resolution with linear computational complexity. The source code will be released upon acceptance.

Keywords: Magnetic Resonance Imaging · Brain Imaging · Super-Resolution · State Space Models · Frequency-Aware Network · Implicit Coordinate Upsampling

1 Introduction

High-quality brain MRI is critical for neurological diagnosis and automated neuroimage analysis. However, prohibitive hardware costs and clinical constraints, such as accelerated scanning to minimize patient motion, often yield scans with insufficient spatial resolution, Rician noise, and truncation artifacts. Deep learning-based super-resolution (SR) offers a cost-effective software alternative to enhance degraded images without costly hardware upgrades, democratizing high-quality imaging across under-resourced clinical settings [1,2,3].

While general SR methods have advanced significantly, they face distinct bottlenecks when applied to medical imaging. CNNs [4,5] struggle to capture global anatomical consistency, Transformers [6,7] incur prohibitive quadratic complexity at high resolutions, and Diffusion models [8,9] impose unacceptable inference delays for clinical workflows. Recently, State Space Models (SSMs) [10,11,12] have emerged to achieve linear-complexity global modeling. However, most existing

SSM architectures ignore the unique anatomical priors of brain MRI, which features large homogeneous soft tissues (e.g., white matter and cerebrospinal fluid) interspersed with highly complex, high-frequency boundaries (e.g., cortical folds). Without explicit domain constraints, generic models tend to blindly amplify noise in uniform regions or over-smooth critical pathological details. To address these

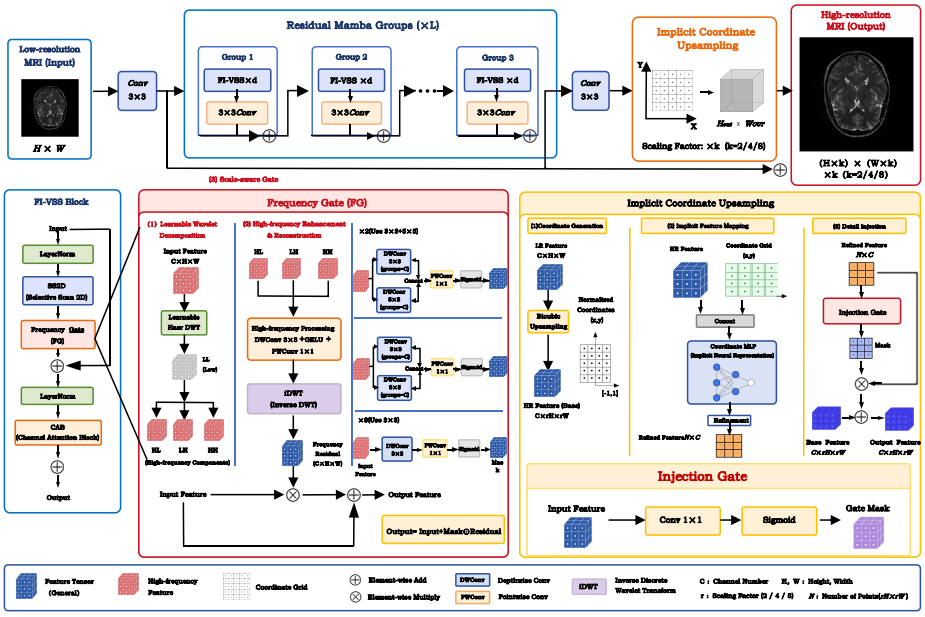

Fig. 1: Architecture of the proposed GF-BrainSR. Each FI-VSS block integrates fixed orthogonal Haar wavelet decomposition, multi-scale frequency gating, and selective state-space scanning for efficient global-local feature refinement.

challenges, we propose an efficient linear-complexity SSM framework tailored to cerebral MRI. Our main contributions are:

- A selective-scanning SR architecture that achieves a strict balance between global anatomical context modeling and linear computational efficiency.
- A DWT-driven frequency adaptive gate that couples orthogonal frequency decoupling with depthwise spatial masking, selectively enhancing cortical boundaries while suppressing noise in homogeneous tissues.
- An implicit coordinate upsampling module that fuses features with continuous 2D coordinates, eliminating checkerboard artifacts to reconstruct morphologically continuous tissue interfaces.
- Comparison experiments demonstrating that our method achieves state-of-the-art quantitative accuracy and superior visual recovery of clinical anatomical structures.

2 Methodology: GF-BrainSR

Fig. 1 illustrates the proposed GF-BrainSR. A 3×3 convolution first extracts shallow features F_0 from the low-resolution MRI I_{in}. These are hierarchically refined by N cascaded Residual Mamba Groups, comprising Frequency-Integrated Visual State Space (FI-VSS) and Channel Attention Blocks, to jointly model long-range anatomical dependencies and localized high-frequency structures. To preserve low-frequency structural consistency, the output deep feature F_{deep} is residually fused with F_0. Finally, an Implicit Coordinate Upsampling module (H_{UP}) reconstructs the high-resolution image via $I_{\text{out}} = H_{\text{UP}}(F_{\text{deep}} + F_0)$. Unlike conventional discrete operators, this continuous coordinate-guided reconstruction effectively alleviates interpolation discontinuities and sharpens anatomical boundaries.

2.1 Frequency-Integrated Visual State Space Block

MRI acquisition artifacts (e.g., Rician noise) are highly coupled with high-frequency anatomical details. Consequently, standard SR models often blindly amplify noise in homogeneous tissues like white matter and cerebrospinal fluid (CSF). To mitigate this, we propose a frequency-aware gate that decouples structural boundaries from noise, selectively enhancing valid details prior to state-space propagation.

Frequency Gate via DWT Given an input feature map $X \in \mathbb{R}^{C \times H \times W}$, we first apply a 2D discrete wavelet transform (DWT) to decompose it into one low-frequency approximation subband X_{LL} and three high-frequency detail subbands $\{X_{HL}, X_{LH}, X_{HH}\}$:

$$X_{LL}, \{X_{HL}, X_{LH}, X_{HH}\} = \text{DWT}(X). \tag{1}$$

The high-frequency components, which encapsulate critical directional edge information, are concatenated along the channel dimension to form:

$$X_{\text{high}} = \text{Concat}(X_{HL}, X_{LH}, X_{HH}). \tag{2}$$

The processed high-frequency representation is formulated as:

$$\hat{X}_{\text{high}} = W_{1\times1} * \delta\left(\text{DWConv}_{3\times3}(X_{\text{high}})\right), \tag{3}$$

where $\delta(\cdot)$ denotes the GELU activation function, $\text{DWConv}_{3\times3}$ represents the 3×3 depthwise convolution operator for spatial grouping, and $W_{1\times1}$ denotes point-wise frequency fusion. The enhanced residual component is then reconstructed through inverse discrete wavelet transform (IDWT):

$$X_{\text{res}} = \text{IDWT}(\hat{X}_{\text{high}}). \tag{4}$$

To suppress noise in homogeneous tissues and adapt to varying degradation levels, we introduce a **scale-aware depthwise gating module**. Under severe

degradation ($\times 8$), spatial dimensions are highly compressed; thus, compact 3×3 kernels provide an adequate relative receptive field while preventing the over-smoothing of fragile high-frequency signals. Conversely, milder degradations ($\times 2$ and $\times 4$) preserve higher resolution and richer structural information. For these scales, we fuse both 3×3 and 5×5 convolutions to simultaneously capture fine local details and broader spatial context, effectively balancing detail retention and structural continuity. The operation $\mathcal{G}(X)$ is formalized as:

$$\mathcal{G}(X) = \begin{cases} \text{Concat}\left(\text{DWConv}_{3\times3}(X), \text{DWConv}_{5\times5}(X)\right), & \text{for } \times 2, \times 4 \\ \text{DWConv}_{3\times3}(X), & \text{for } \times 8 \end{cases} \tag{5}$$

The gating mask M is then estimated through a lightweight bottleneck network:

$$M = \sigma\left(W_{1\times1}^{up} * \text{LReLU}\left(W_{1\times1}^{down} * \mathcal{G}(X)\right)\right), \tag{6}$$

where $\sigma(\cdot)$ denotes the Sigmoid activation function and LReLU represents Leaky ReLU.

The final gated frequency enhancement is:

$$X_{\text{out}} = X + M \odot X_{\text{res}}, \tag{7}$$

where $\odot$ denotes element-wise multiplication.

Frequency-Integrated 2D Selective Scan Standard state-space models exhibit strong long-range dependency modeling capability but tend to suppress subtle high-frequency structural variations. To alleviate this limitation, we inject frequency-enhanced priors before selective state propagation:

$$Z_{\text{freq}} = \text{Gate}_{\text{freq}}\left(\text{SiLU}\left(\text{Conv}_{3\times3}(Z)\right)\right). \tag{8}$$

Unlike post-hoc frequency refinement strategies, the proposed frequency gate is inserted before selective scanning, enabling the state-space model to encode enhanced anatomical boundaries during long-range dependency propagation. To capture global non-causal anatomical dependencies, the enhanced feature Z_{freq} is processed through a 4-direction (horizontal, vertical, and their reverses) selective state-space model (SSM). The multi-directional outputs are then reshaped and aggregated to form the global representation:

$$Y = \sum_{k=1}^{4} \text{Unflatten}^{(k)}\left(\text{SSM}^{(k)}\left(\text{Flatten}^{(k)}(Z_{\text{freq}})\right)\right), \tag{9}$$

where $k \in \{1, 2, 3, 4\}$ denotes the four scanning directions. Outputs from all directions are reshaped back to the spatial domain and aggregated to form the global contextual representation:

$$Y = \sum_{k=1}^{4} \text{Unflatten}^{(k)}\left(y^{(k)}\right). \tag{10}$$

2.2 Implicit Coordinate Upsampling

Conventional discrete upsampling operators (e.g., transposed convolution or pixel shuffle) are constrained to integer scale factors and frequently introduce checkerboard artifacts, which compromise the morphological continuity of fine anatomical structures. In clinical practice, acquiring arbitrary-scale continuous representations is crucial for standardizing MR images from multi-center cohorts with varying slice thicknesses. Motivated by this, we introduce an Implicit Coordinate Upsampling module. Given a low-resolution feature map F_{LR} and scale factor s, a bicubic-interpolated base feature is first generated:

$$F_{\mathrm{up}} = \mathrm{Bicubic}(F_{\mathrm{LR}}, s) \in \mathbb{R}^{C \times sH \times sW}. \tag{11}$$

A normalized coordinate grid $V_{\mathrm{coord}} \in \mathbb{R}^{2 \times sH \times sW}$ is then constructed with coordinates linearly distributed in $[-1, 1]$. The coordinate grid is concatenated with F_{up} and projected into coordinate-conditioned feature representations through a lightweight coordinate fusion network ϕ_{coord}:

$$F_{\mathrm{imp}} = \phi_{\mathrm{coord}} \left(\mathrm{Concat}(F_{\mathrm{up}}, V_{\mathrm{coord}}) \right). \tag{12}$$

A refinement network ψ_{refine} predicts high-frequency residual details from F_{imp}, while a sigmoid gate dynamically modulates the enhancement intensity to alleviate artifacts in homogeneous regions. The final reconstructed high-resolution feature is formulated as:

$$F_{\mathrm{HR}} = F_{\mathrm{up}} + \sigma \left(W_{\mathrm{gate}} * F_{\mathrm{imp}} \right) \odot \psi_{\mathrm{refine}}(F_{\mathrm{imp}}). \tag{13}$$

The proposed upsampling strategy improves boundary sharpness while preserving smooth anatomical tissue regions.

3 Experiments

3.1 Datasets

IXI Dataset. The IXI dataset is a publicly available brain imaging dataset comprising multi-modal MRI scans (T1, T2, and PD sequences) of healthy adults from diverse scanning sources. Following preprocessing and spatial sampling, we partition the dataset into disjoint subsets, yielding 4,524 slices for training, 1,508 for validation, and 1,536 for testing to robustly evaluate our model.

3.2 Implementation Details

Experiments are implemented in PyTorch and conducted on NVIDIA RTX 4090 GPUs. High-resolution training patches are cropped to 192×192 and augmented with random flips and rotations. Our model is trained for 250K iterations using L_1 loss (batch size 4) and the Adam optimizer ($\beta_1 = 0.9, \beta_2 = 0.99$). The initial learning rate is 1×10^{-4}, halved at 125K, 200K, 225K, and 237.5K iterations. Notably, for fair comparison, baselines including Restore-RWKV [12] and SwinIR [6] are trained for 750K iterations with 2×10^{-4} initial learning rate to ensure full convergence. Performance is evaluated every 5K iterations using PSNR and SSIM metrics.

3.3 Results and Analysis

Table 1 compares the quantitative performance on the IXI dataset. The proposed GF-BrainSR consistently outperforms representative CNN, Transformer, and SSM baselines across all scales. Notably, under severe degradation, it surpasses the runner-up MambaIR by 0.30 dB and 0.22 dB in PSNR at ×4 and ×8, respectively. These improvements demonstrate that our scale-adaptive frequency gating and implicit upsampling effectively capture unique anatomical boundaries while maintaining linear-complexity global context modeling. Qualitative comparisons are visualized in Fig. 2.

Table 1: Quantitative results with different methods and upsampling factors on the IXI dataset. Best and second-best results are highlighted in **bold** and underlined, respectively.

Method	2×		4×		8×	
	PSNR↑	**SSIM↑**	**PSNR↑**	**SSIM↑**	**PSNR↑**	**SSIM↑**
SRResNet [13]	31.78	0.9467	27.73	0.8598	24.23	0.7389
EDSR [14]	34.38	0.9605	29.22	0.8841	24.51	0.7499
RCAN [4]	36.98	0.9702	28.98	0.8802	24.30	0.7463
ECBSR [5]	31.24	0.9374	25.32	0.7953	22.29	0.6796
SwinIR [6]	38.69	0.9782	30.79	0.9075	25.92	0.7891
MambaIR [11]	<u>39.36</u>	<u>0.9805</u>	<u>31.25</u>	<u>0.9139</u>	<u>26.42</u>	<u>0.8068</u>
Restore-RWKV [12]	38.13	0.9782	30.08	0.9117	25.23	0.7973
MambaIRV2 [15]	38.97	0.9790	30.78	0.9085	25.80	0.7866
GF-BrainSR (ours)	**39.46**	**0.9809**	**31.55**	**0.9201**	**26.64**	**0.8163**

3.4 Ablation Study

Frequency-Aware Encoder Mechanisms. Table 2 ablates encoder variants for ×8 SR using our implicit decoder. The baseline VSS block (26.44 dB) struggles

Table 2: Quantitative ablation study on the evolution of the frequency-aware encoder for the ×8 super-resolution task. All variants are evaluated using the proposed Implicit Coordinate Upsampling decoder to ensure a fair comparison. Best results are highlighted in **bold**.

Encoder Architecture	**PSNR↑**	**SSIM↑**
VSS Block (Baseline)	26.44	0.8094
Parallel DWT VSS Block	26.43	0.8081
Pseudo-IDWT VSS Block	26.58	0.8150
FI-VSS Block w/o gate	26.57	**0.8164**
FI-VSS Block (Ours)	**26.64**	0.8163

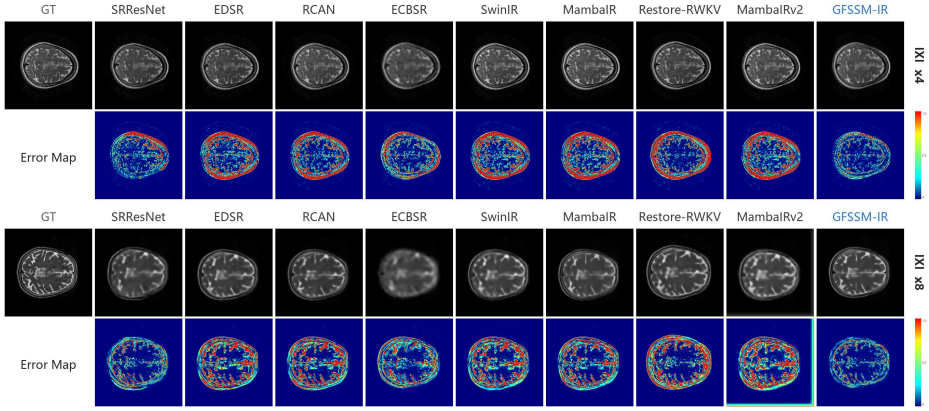

Fig. 2: Qualitative comparison of ×4 and ×8 MRI super-resolution on the IXI dataset. Error maps visualize per-pixel absolute residuals between SR predictions and ground truth.

with fine anatomical boundaries due to absent high-frequency priors. Naively integrating DWT (Parallel DWT) degrades performance (26.43 dB) by blindly amplifying noise in homogeneous tissues and breaking wavelet orthogonality. While refining residual reconstruction (Pseudo-IDWT) improves PSNR to 26.58 dB, it still suffers from aliasing due to non-orthogonal bases. Omitting the spatial gate (FI-VSS w/o gate) drops PSNR to 26.57 dB, proving that unmodulated high-frequency enhancement inevitably corrupts smooth backgrounds. Our complete FI-VSS Block achieves the highest PSNR (26.64 dB) by coupling strictly orthogonal IDWT with localized spatial gating, optimally balancing boundary enhancement and soft-tissue noise suppression.

Scale-aware Depthwise Gating.

Scale-aware Depthwise Gating. Table 3 investigates the impact of spatial kernel sizes in the frequency gating module across different degradation scales. Interestingly, the optimal kernel size exhibits a strictly scale-dependent behavior. Under mild degradation (×2), the joint $3 \times 3 + 5 \times 5$ kernel achieves the highest PSNR (39.46 dB), as the complementary receptive fields effectively capture both local details and broader spatial context required for high-resolution frequency modulation. Under the intermediate ×4 scale, fusing both 3×3 and 5×5 convolutions also optimally balances local detail retention and medium-range structural continuity, peaking at 31.55 dB. Conversely, under severe degradation (×8), the compact 3×3 kernel alone performs best (26.64 dB). This is because the ×8 latent representations are highly compressed; a 3×3 kernel already yields a massive effective receptive field relative to the original image, whereas larger or fused kernels tend to over-smooth the fragile, severely degraded high-frequency anatomical signals. This justifies our scale-aware gating design that adaptively tailors kernel configurations to different scales.

Table 3: Ablation study of different kernel sizes on the IXI dataset for depthwise gating in the frequency gate of the FI-VSS block. Best results are highlighted in **bold**.

Method	2×		4×		8×	
	PSNR↑	**SSIM↑**	**PSNR↑**	**SSIM↑**	**PSNR↑**	**SSIM↑**
3×3	39.33	0.9803	31.37	0.9167	**26.64**	**0.8163**
5×5	39.38	0.9805	31.47	0.9170	26.48	0.8131
3×3 + 5×5	**39.46**	**0.9809**	**31.55**	**0.9201**	26.56	0.8150

Table 4: Ablation study on different upsampling decoder architectures for the ×8 super-resolution task on the IXI dataset. "Implicit" denotes continuous coordinate upsampling, and "Gate" refers to the spatial gating mechanism for detail refinement. Best results are highlighted in **bold**.

Decoder Architecture	**PSNR↑**	**SSIM↑**
Standard PixelShuffle	26.38	0.8095
Implicit Upsample w/o Gate	26.59	0.8156
Implicit Upsample w/ Gate (Ours)	**26.64**	**0.8163**

Decoder Architecture. Table 4 validates our continuous spatial decoder. The PixelShuffle baseline yields the lowest performance (26.38 dB), as its discrete nature introduces severe aliasing artifacts and struggles to reconstruct smooth anatomical boundaries at large scales. Replacing it with coordinate-based implicit upsampling improves PSNR to 26.59 dB, demonstrating that continuous spatial modeling effectively mitigates discrete interpolation bottlenecks. Finally, incorporating our spatial gating mechanism achieves the best result (26.64 dB). This gate dynamically regulates high-frequency detail injection, successfully preventing spurious textures in homogeneous soft-tissue regions.

4 Conclusion

We present GF-BrainSR, an efficient linear-complexity super-resolution framework tailored for brain MRI. By coupling a DWT-driven frequency gate with selective state-space models, it successfully decouples high-frequency anatomical boundaries from acquisition noise. Furthermore, our gated implicit upsampling eliminates discrete interpolation artifacts, reconstructing morphologically continuous tissue interfaces. Experiments confirm that GF-BrainSR consistently outperforms state-of-the-art baselines, delivering superior anatomical clarity for clinical diagnosis with strict computational efficiency. Future work will extend the framework to a 3D state-space architecture to exploit full spatial continuity. We will also validate its generalization on multi-center pathological datasets and explore its utility in enhancing downstream clinical tasks, such as automated brain MRI image classification and segmentation.

References

1. B Sahu and G Madani. Imaging inequality: exploring the differences in radiology between high-and low-income countries, 2024.
2. TA Awolowo, GM OLUMIDE, OO OLAPOSI, OO ADENIYI, AO AYOBAMI, DO ALFRED, AC IYABO, and A BENJAMIN. A critical appraisal of the unmet needs for cross-sectional imaging techniques in nigeria. *International Journal of Radiology and Diagnostic Imaging*, 7:19–24, 2024.
3. Chun Wai Chin, Haniza Yazid, and Hoi Leong Lee. Challenges, advances, and evaluation metrics in medical image enhancement: A systematic literature review. *arXiv preprint arXiv:2510.13638*, 2025.
4. Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *ECCV*, 2018.
5. Xindong Zhang, Hui Zeng, and Lei Zhang. Edge-oriented convolution block for real-time super resolution on mobile devices. In *Proceedings of the 29th ACM international conference on multimedia*, pages 4034–4043, 2021.
6. Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021.
7. Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022.
8. Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE transactions on pattern analysis and machine intelligence*, 45(4):4713–4726, 2022.
9. Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.
10. Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.
11. Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. Mambair: A simple baseline for image restoration with state-space model. In *European conference on computer vision*, pages 222–241. Springer, 2024.
12. Zhiwen Yang, Jiayin Li, Hui Zhang, Dan Zhao, Bingzheng Wei, and Yan Xu. Restore-rwkv: Efficient and effective medical image restoration with rwkv. *IEEE Journal of Biomedical and Health Informatics*, 2025.
13. Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.
14. Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.
15. Hang Guo, Yong Guo, Yaohua Zha, Yulun Zhang, Wenbo Li, Tao Dai, Shu-Tao Xia, and Yawei Li. Mambairv2: Attentive state space restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 28124–28133, 2025.