Directing Mamba to Complex Textures: An Efficient Texture-Aware State Space Model for Image Restoration

Long Peng[1,2†], **Xin Di**[1†], **Zhanfeng Feng**[1], **Wenbo Li**[2], **Renjing Pei**[2*], **Yang Wang**[1,3], **Xueyang Fu**[1], **Yang Cao**[1], **Zheng-Jun Zha**[1]
[1] University of Science and Technology of China
[2] Huawei Noah's Ark Lab
[3] Chang'an University
{longp2001@mail.,ywang120@}ustc.edu.cn

Abstract

Image restoration aims to recover details and enhance contrast in degraded images. With the growing demand for high-quality imaging (*e.g.*, 4K and 8K), achieving a balance between restoration quality and computational efficiency has become increasingly critical. Existing methods, primarily based on CNNs, Transformers, or their hybrid approaches, apply uniform deep representation extraction across the image. However, these methods often struggle to effectively model long-range dependencies and largely overlook the spatial characteristics of image degradation (regions with richer textures tend to suffer more severe damage), making it hard to achieve the best trade-off between restoration quality and efficiency. To address these issues, we propose a novel texture-aware image restoration method, TAMambaIR, which simultaneously perceives image textures and achieves a trade-off between performance and efficiency. Specifically, we introduce a novel Texture-Aware State Space Model, which enhances texture awareness and improves efficiency by modulating the transition matrix of the state-space equation and focusing on regions with complex textures. Additionally, we design a Multi-Directional Perception Block to improve multi-directional receptive fields while maintaining low computational overhead. Extensive experiments on benchmarks for image super-resolution, deraining, and low-light image enhancement demonstrate that TAMambaIR achieves state-of-the-art performance with significantly improved efficiency, establishing it as a robust and efficient framework for image restoration.

1 Introduction

Image restoration as a fundamental task in computer vision and image processing, aiming to recover details and improve image contrast from degraded images [Liang *et al.*, 2021a; Tsai and Huang, 1984; Guo *et al.*, 2024a; Zamir *et al.*, 2022; Xiao *et al.*, 2022], which has been widely applied in imaging devices and various vision systems [Zheng *et al.*, 2022; Li *et al.*, 2022]. With the continuous advancements and widespread deployment of modern smartphones and cameras, the demand for higher image quality has significantly increased from the 1K resolutions (1280 × 720) to 4K (4096 × 2160) [Kong *et al.*, 2021]. Therefore, balancing the growing demand for high-quality restoration quality and model efficiency has become a critical challenge in image restoration [Liang *et al.*, 2021b; Guo *et al.*, 2024a].

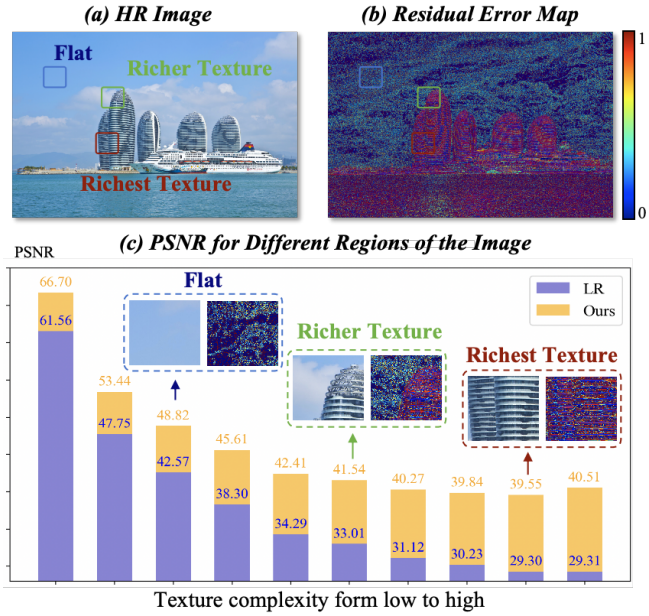

Figure 1: (a-b) We compute the residual error map between the high-resolution and low-resolution images in the DIV2K and Manage109 datasets and find that the degree of degradation varies across different regions of the image. (c) We divide image patches into 10 groups sorted by texture complexity (measured by statistical variance) and calculate the average PSNR for each group in all Manage109 datasets. It can be observed that regions with richer textures suffer more severe degradation, resulting in lower PSNR values.

With the rapid advancement of deep learning, various image restoration methods have been proposed. Early research in image restoration primarily focused on convolutional neural networks (CNNs) [Zhang *et al.*, 2018; Sun *et al.*, 2023], which offer high computational efficiency. However, limited by the local receptive fields of CNNs, these methods struggle

to model long-range dependencies, resulting in suboptimal restoration quality. Therefore, Vision Transformers are introduced to improve image restoration performance due to their ability to model long-range dependencies [Chen *et al.*, 2023a; Zhou *et al.*, 2023; Cai *et al.*, 2023a]. Despite these advancements, the high computational complexity of Transformers remains a major obstacle for real-world deployment, especially for high-resolution images.

To seek efficient frameworks, many previous works have proposed combining the efficiency of CNN with the global modeling capabilities of Transformers, resulting in hybrid CNN-Transformer models that aim to balance performance and complexity [Liang *et al.*, 2021b; Zhou *et al.*, 2023; Fang *et al.*, 2022]. However, these approaches often overlook the spatial characteristics of degradation, where regions with richer textures tend to suffer more severe damage, applying uniform deep representation extraction across the image. This uniform approach makes it challenging to achieve a trade-off between performance and efficiency. Specifically, we argue that regions with richer textures are generally more prone to severe quality degradation. To clarify this observation, we conduct a statistical analysis on image super-resolution (additional analyses of low-light image enhancement are shown in the appendix). First, we divide the low-resolution or low-quality images into patches and measure the texture complexity of each patch using statistical variance. We then evaluate the PSNR between the degraded patches and their ground truth counterparts. The results, as shown in Figure 1, reveal that regions with richer textures tend to suffer more severe degradation, resulting in lower PSNR values. This highlights the importance of paying more attention to regions with complex textures. Although these methods [Kong *et al.*, 2021; Jeong *et al.*, 2025] attempt to address this issue by classifying image patches/pixels and applying different-sized CNNs, they still face challenges in perceiving image textures and modeling contextual information, making it difficult to achieve high-efficiency.

To address these issues, we propose a novel Texture-Aware State Space Model (TA-SSM) to simultaneously achieve efficient contextual modeling and texture awareness, enabling an optimal trade-off between performance and efficiency. Specifically, we modulate the transition matrix of state-space equations to mitigate catastrophic forgetting in regions with richer textures, enhancing texture perception. Furthermore, by focusing more on challenging texture regions, the proposed method improves overall efficiency. Additionally, for the first time, we introduce positional embeddings into SSM, improving their ability to perceive contextual positions. To further reduce computational costs while maintaining multi-directional perception, we design a Multi-Directional Perception Block to enhance efficiency. Extensive experiments on various image restoration benchmarks including image super-resolution, image deraining, and low-light image enhancement, demonstrate that the proposed method significantly improves the efficiency of Mamba-based image restoration. This provides a more powerful and efficient framework for future image restoration tasks.

The contributions of this paper can be summarized:

- We propose a novel and efficient Texture-Aware State Space Model (TA-SSM) that perceives complex textures by modifying the state-space equations and transition matrices and simultaneously focuses on more challenging texture regions to enhance efficiency.
- An efficient Multi-Directional Perception Block is proposed to expand the receptive field in multiple directions while maintaining low computational costs. Furthermore, position embedding is introduced into SSM to enhance its capability of perceiving contextual positions.
- Based on these, a straightforward yet efficient model TAMambaIR is proposed, showcasing superior performance in both efficiency and effectiveness across various image restoration tasks and benchmarks, offering an efficient backbone for image restoration.

2 Related Work

2.1 Image Restoration

Images captured in complex real-world scenarios often suffer from degradations like low resolution, low-light, rain, and haze, resulting in reduced contrast and detail loss [Rim *et al.*, 2020; Wang *et al.*, 2018; Li *et al.*, 2025; Peng *et al.*, 2024a; Peng *et al.*, 2025; Peng *et al.*, 2024c; Peng *et al.*, 2024b]. Image restoration aims to enhance contrast and recover details, improving visual quality. Advances in deep learning have significantly boosted its effectiveness [Liang *et al.*, 2021b]. Dong *et al.* [Dong *et al.*, 2015] introduced SRCNN in 2015, pioneering deep learning for image super-resolution (SR). Since then, numerous CNN-based methods have emerged to address image restoration tasks [Zhang *et al.*, 2018; Dai *et al.*, 2019], though their lightweight designs often limit receptive fields. To address this issue, Vision Transformers (ViTs) have been introduced to leverage long-range modeling for image restoration. Chen *et al.* [Chen *et al.*, 2021] demonstrated their effectiveness in image denoising, while the Swin Transformer [Liang *et al.*, 2021b] captured multi-scale features hierarchically. However, increasing image resolutions pose significant computational challenges for Transformer-based methods [Guo *et al.*, 2024a]. While CNN-Transformer hybrids aim to balance performance and efficiency, they often overlook the spatial characteristics of degradation, where regions with richer textures tend to suffer more severe damage. As a result, they apply uniform enhancements and struggle to effectively focus on the damaged textures. Patch- or pixel-level classification methods [Kong *et al.*, 2021; Jeong *et al.*, 2025] address this partially but fail to effectively model textures and contextual information. To tackle this, we adopt efficient global modeling with State Space Model (SSM) and introduce texture-aware capabilities to better handle complex textures, achieving a balance between performance and efficiency.

2.2 State Space Model

State Space Model (SSM), originally developed in the 1960s for control systems [Kalman, 1960], has recently been extended to applications in computer vision [Zhu *et al.*, 2024; Xiao *et al.*, 2025]. The introduction of SSM into computer vision was pioneered by Visual Mamba, which designed the

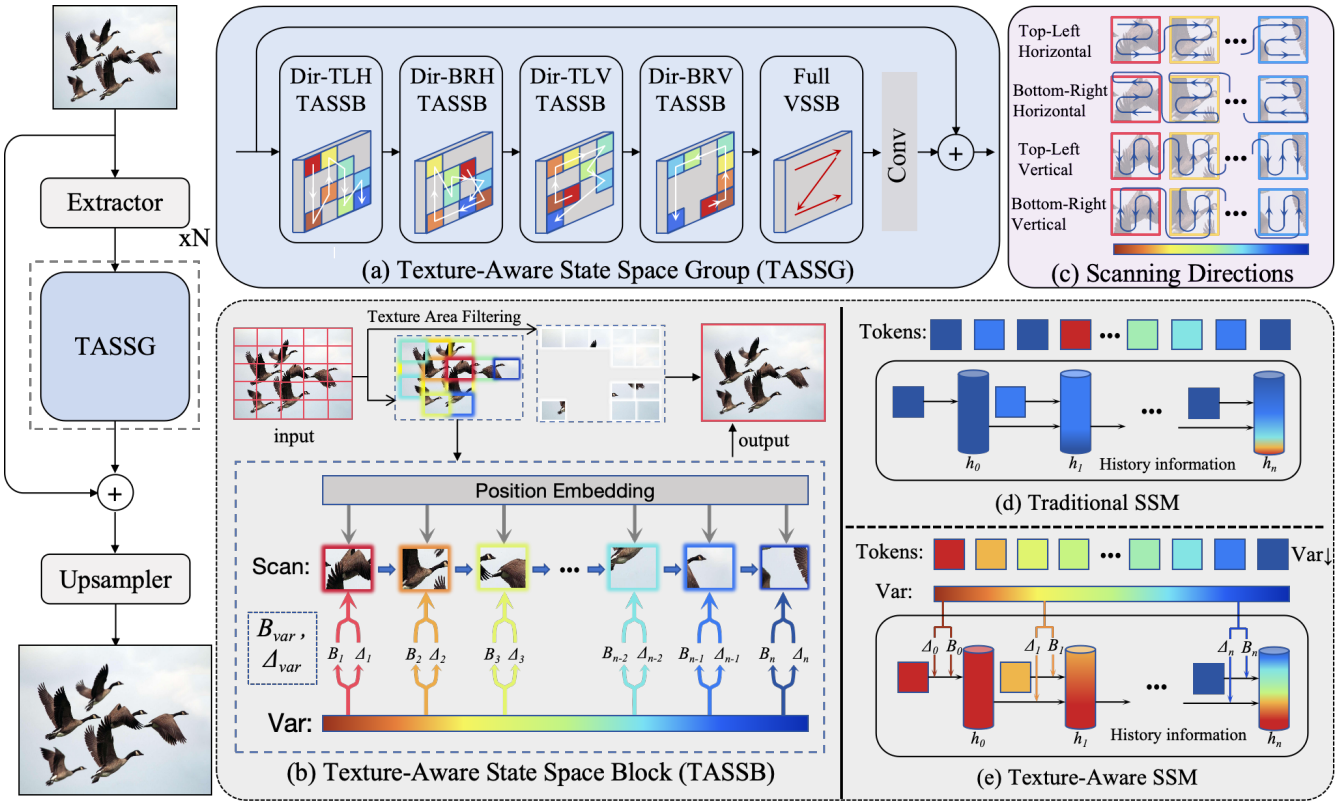

Figure 2: The overall framework of TAMambaIR consists of a feature extractor and several TASSGs, with an upsampler by pixel shuffle.

VSS module to achieve superior performance compared to Vision Transformers (ViTs) [Dosovitskiy *et al.*, 2020], while maintaining lower model complexity. This breakthrough inspired a wave of research utilizing SSM in various tasks [Guo *et al.*, 2024b; Qiao *et al.*, 2024]. Notably, MambaIR [Guo *et al.*, 2024b] is the first to employ SSM for image restoration, demonstrating improved efficiency and enhanced global perceptual capabilities. However, directly applying SSM to image restoration poses challenges in effectively enhancing texture-rich regions, which limits the full potential of the State Space Model in this domain.

3 Proposed Method

3.1 Traditional State Space Model

Let's briefly review the traditional State Space Model (SSM), which maps sequence input $x(t)$ to output $y(t)$ through an implicit latent state $h(t) \in \mathbb{R}^N$ [Guo *et al.*, 2024b] and can be represented as a linear ordinary differential equation:

$$\begin{aligned} \dot{h}(t) &= Ah(t) + Bx(t), \\ y(t) &= Ch(t) + Dx(t), \end{aligned} \tag{1}$$

where N is the state size, $A \in \mathbb{R}^{N\times N}$, $B \in \mathbb{R}^{N\times 1}$, $C \in \mathbb{R}^{1\times N}$, and $D \in \mathbb{R}$. Discretized using a zero-order hold as:

$$\begin{aligned} \overline{A} &= \exp(\Delta A), \\ \overline{B} &= (\Delta A)^{-1}(\exp(\Delta A) - I)\Delta B, \end{aligned} \tag{2}$$

After the discretization, the discretized version of Eq. (1) with step size Δ can be rewritten as:

$$\begin{aligned} h_k &= \overline{A}h_{k-1} + \overline{B}x_k, \\ y_k &= Ch_k + Dx_k, \end{aligned} \tag{3}$$

The advent of SSM has garnered significant attention for its ability to perform global modeling with linear complexity in image restoration [Guo *et al.*, 2024a]. However, traditional SSM still faces two critical challenges: **(1) It consistently handles flat and texture-rich regions, spending too much computation on easily recoverable flat areas (as shown in Figure 1), resulting in inefficiency. (2) It lacks an effective texture-awareness mechanism, which not only leads to catastrophic forgetting of texture information but also hinders the model's ability to focus on challenging texture-rich regions.**

3.2 Texture-Aware State Space Model

To address these limitations, we propose a novel Texture-Aware State Space Model to fully unlock the potential of SSM in image restoration, as shown in Figure 2. We tackle the aforementioned challenges from two perspectives. First, we propose a novel Texture Area Filtering method, which utilizes variance statistics to identify texture-rich and flat regions and pay less/more computation on flat/texture-rich regions, thereby improving efficiency. Second, we propose a novel Texture-Aware Modulation, which ranks regions based on their texture richness in descending order and enhances

SSM's historical retention of texture-rich information by modulating the transition matrix in the state-space equation.

Texture Area Filtering. We first patchify the image/feature $x \in \mathbb{R}^{C\times H\times W}$ into patches $p \in \mathbb{R}^{C\times h\times w}$ to obtain a sequence $\{P_i\}$, as follows:

$$\{P_i\} = Patchify(x), \quad i = 1, 2, \ldots, N. \tag{4}$$

where N represents the number of patches. Then, we sort this patch sequence based on the texture complexity from high to low, which is measured by statistical variance (other measurements of compression are provided in the appendix), resulting in a new ordered patch sequence $\{P_i^{'}\}$, as follows:

$$\begin{aligned} \mu(P_i) &= \frac{1}{|P_i|}\sum_{j\in P_i} P_{ij}, \\ \mathrm{Var}(P_i) &= \frac{1}{|P_i|}\sum_{j\in P_i}\left(P_{ij}-\mu(P_i)\right)^2, \\ \{P_i'\}_{i=1}^{n} &= \mathrm{argsort}_{\downarrow}\left(\mathrm{Var}(P_i)\right), \end{aligned} \tag{5}$$

Then, we process the top $p\%$ patches with the highest texture complexity and skip the easy flat regions (which will be enhanced by the following full-sequence scan and convolution as shown in the (a) of Figure 2), directing SSM focus on the more challenging region as follows:

$$\{P_i\}_{top-p\%} = \left\{P_i' \mid i \le \lceil p\cdot N\rceil\right\}, \tag{6}$$

Texture-Aware Modulation In traditional SSM, the transition matrixes B and Δ are expressed as follows:

$$B = Linear(x), \Delta = Linear(x). \tag{7}$$

This way may lead to two key issues: (a) The transition matrix relies solely on the input itself and lacks the ability to perceive texture information, **resulting in catastrophic forgetting of texture-related representations in the state-space history, thereby weakening the model's ability to enhance fine details**, as shown in the (d) of Figure 2. (b) Flat regions, which are easier to enhance, dominate the state-space history, **leading to an excessive focus on flat areas and hindering the model's ability to address challenging texture-rich regions**. To address these issues and enhance texture awareness, we first sort the patches as described in Eq.6, and then propose modulating the transition matrix within the state-space model to improve its capability of preserving historical texture information, as shown in Figure2 (b) and (e).

$$\begin{aligned} B_{\text{var}} &= Var(x)\cdot Linear(x), \\ \Delta_{\text{var}} &= Var(x)\cdot Linear(x). \end{aligned} \tag{8}$$

Through this method, the transition matrices B_{var} and Δ_{car} are able to capture the varying texture complexities of different patches. Furthermore, they retain the historical information of regions with richer textures, effectively alleviating catastrophic forgetting and enhancing the ability to represent fine details in image restoration, as illustrated in Figure 2 (e). Finally, the state-space function formulation of Texture-Aware State Space Model can be expressed as follows:

$$\begin{aligned} \overline{A}_{\text{var}_k} &= \exp(\Delta_{\text{var}_k} A) \\ \overline{B}_{\text{var}_k} &= (\Delta_{\text{var}_k} A)^{-1}(\exp(A) - I)\cdot \Delta_{\text{var}_k} B_{\text{var}_k} \\ h_k &= \overline{A}_{\text{var}_k} h_{k-1} + \overline{B}_{\text{var}_k} x_k \\ y_k &= Ch_k + Dx_k. \end{aligned} \tag{9}$$

Position Embedding. It is well known that the traditional SSM adopts a linear sequence scanning method, which inevitably limits the contextual receptive field. Although SSM proposes scanning in four directions to partially improve the contextual receptive field, the perception capabilities of contextual positions remain limited. On the other hand, considering that our TA-SSM employs a texture complexity-based scanning method, this further exacerbates the contextual position perception process. To address this, we propose to introduce position embedding, which is widely used in Transformers, into SSM to enhance its contextual position awareness. This method enables SSM to capture the global positional relationship of the current token during each sequence scan, thereby improving its ability to perform global texture arrangement and contextual modeling. Specifically, we add a learnable position embedding Pos to each patch after Eq. 6:

$$\{P_i\}_{top-p\%} = \{P_i + Pos(P_i) \mid P_i \in \{P_i\}_{top-p\%}\}. \tag{10}$$

3.3 Multi-Directional Perception Block

Through Texture-Aware State Space Model can model the global information of different patches based on their texture complexity. However, the contextual perception within a patch remains critical. Existing approaches often adopt multi-directional scanning within a single SSM, which can lead to significant computational overhead. To address this issue, we propose a novel Multi-Directional Perception Block that sequentially connects four different scaning directions (Top-Left Horizontal, Bottom-Right Horizontal, Top-Left Vertical, and Bottom-Right Vertical) to reduce computational cost while enhancing multi-directional perception, as shown in Figure 2 (a) and (c). Additionally, low-texture complexity regions (*e.g.*, flat areas), which are relatively easier to enhance, are also crucial for image restoration. Therefore, we introduce a full-scan SSM and convolution layer to ensure the network can also focus on these regions, facilitating interactions between high-texture complexity patches and flat patches.

3.4 TAMambaIR

Following the previous approach in [Guo *et al.*, 2024a], we adopt a simple yet efficient architecture to build TAMambaIR, as depicted in Figure 2. TAMambaIR comprises a feature extractor, a Texture-Aware State Space Group, and an upsampler. Specifically, convolution layers and pixel shuffle are used in the feature extractor and upsampler. Due to limited space, the implement details of the Texture-Aware State Space Group and Texture-Aware State Space Block are provided in **Appendix Section 8**.

3.5 Loss Function

Following previous works [Zamir *et al.*, 2022; Cui *et al.*, 2023], we utilize the L1 loss $\mathcal{L}_1$ and frequency loss $\mathcal{L}_{fft}$ for training. The total loss is presented as follows:

$$\mathcal{L}_{total} = \lambda_1\mathcal{L}_1(\mathcal{O}, \mathcal{O}_{gt}) + \lambda_2\mathcal{L}_{fft}(\mathcal{O}, \mathcal{O}_{gt}). \tag{11}$$

where $\mathcal{O}$ and $\mathcal{O}_{gt}$ denote the model output and the ground truth, respectively. The parameters λ_1 and λ_2 are balancing factors. We set λ_1 and λ_2 to 1, 0.05, respectively.

Method	Set5		Set14		BSDS100		Manga109		FLOPs (G)	Params (M)
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑		
RCAN	38.27	0.9614	34.12	0.9216	32.41	0.9027	39.44	0.9786	62.75	15.44
SAN	38.31	0.9620	34.07	0.9213	32.42	0.9028	39.32	0.9792	64.11	15.71
ClassSR	38.29	0.9615	34.18	0.9218	32.45	0.9029	39.45	0.9786	40.78	30.10
IPT	38.37	-	34.43	-	32.48	-	-	-	-	115.61
CSNLN	38.28	0.9616	34.12	0.9223	32.40	0.9024	39.37	0.9785	481.97	6.21
NLSA	38.34	0.9618	34.08	0.9231	32.43	0.9027	39.59	0.9789	182.82	41.80
EDT-B	38.45	0.9624	34.57	0.9258	32.52	0.9041	39.93	0.9800	30.22	11.48
EDSR	38.11	0.9602	33.92	0.9195	32.32	0.9013	39.10	0.9773	166.84	40.73
RDN	38.24	0.9614	34.01	0.9212	32.34	0.9017	39.18	0.9780	90.60	22.12
HAN	38.27	0.9614	34.16	0.9217	32.41	0.9027	39.46	0.9785	258.82	63.61
SwinIR	38.42	0.9623	34.46	0.9250	32.53	0.9041	39.92	0.9797	51.33	11.75
SRFormer	38.51	0.9627	34.44	0.9253	32.57	0.9046	40.07	0.9802	62.95	10.40
TAMambaIR-S	**38.53**	**0.9627**	**34.64**	**0.9262**	**32.57**	**0.9046**	**40.23**	**0.9806**	56.88	12.19
MambaIR	38.57	0.9627	34.67	0.9261	32.58	0.9048	40.28	0.9806	110.49	20.42
TAMambaIR	**38.58**	**0.9627**	**34.72**	**0.9265**	**32.58**	**0.9048**	**40.35**	**0.9810**	89.99	16.07

Table 1: Quantitative comparison on $\times 2$ **image super-resolution** with state-of-the-art methods. The best and the second best results are in **bold** and bold.

4 Experiences

4.1 Experimental Settings

Training Details. Following prior works [Liang *et al.*, 2021b; Guo *et al.*, 2024a], the training batch sizes for image super-resolution, image deraining, and low-light image enhancement are set to 32, 16, and 16, respectively. During training, the original images are cropped into 64×64 patches for image super-resolution, and 256×256 patches for image deraining and low-light image enhancement. Top $p\%$ epirically is set at 0.5. Following [Liang *et al.*, 2021b; Guo *et al.*, 2024a], we use the UNet architecture for image deraining and low-light image enhancement. The Adam optimizer [Kingma and Ba, 2014] is used to train our method, with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and an initial learning rate of 2×10^{-4}. The standard model consists of 7 intermediate blocks with a depth of 7, while the small version consists of 6 intermediate blocks with a depth of 6. All experiments are conducted on 8 NVIDIA V100 GPUs. Additional details of the implementation are provided in **Appendix Section 9**.

Dataset. To evaluate our method, we select three popular and representative image restoration tasks, including image super-resolution (SR), image draining (Derain), and low-light image enhancement (LLIE). Specifically, for SR, DIV2K [Timofte *et al.*, 2017] and Flickr2K [Lim *et al.*, 2017] are employed to train the network, while Set5 [Bevilacqua *et al.*, 2012], Set14 [Zeyde *et al.*, 2012], B100 [Martin *et al.*, 2001], and Manga109 [Matsui *et al.*, 2017] are used for evaluation. For Derain, we follow [Chen *et al.*, 2023b] and validate our approach on both popular benchmarks: Rain200H and Rain200L [Yang *et al.*, 2017], which contain heavy and light rain conditions, respectively, for training and evaluation. For LLIE, we follow [Cai *et al.*, 2023b] and validate our approach on the synthetic version of the LOL-V2 dataset [Yang *et al.*, 2021]. More details about the datasets used are provided in the **Appendix Section 10**.

Evaluation Metrics. Following previous work [Chen *et al.*, 2023b; Cai *et al.*, 2023b; Guo *et al.*, 2024a], we use Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) as evaluation metrics. For the image super-resolution and deraining tasks, evaluation is performed on the Y channel of the YCbCr color space, while for the low-light enhancement task, evaluation is conducted in the RGB space.

Comparisons with State-of-the-art Methods. We compare with many existing state-of-the-art methods. **Image super-resolution**: we compare our method against thirteen state-of-the-art methods, including: RCAN, SAN, ClassSR, IPT, CSNLN, NLSA, EDT-B, EDSR, RDN, HAN, SwinIR, SRFormer, and MambaIR. **Image deraining**: we compare our method against fifteen state-of-the-art methods, including: DDN, RESCAN, PReNet, MSPFN, RCDNet, MPRNet, SwinIR, DualGCN, SPDNet, Uformer, Restormer, IDT, DLINet, DRSformer, and MambaIR. **Low-light image enhancement**: we compare our method against fifteen state-of-the-art methods, including: RetinexNet, KinD, ZeroDCE, 3DLUT, DRBN, RUAS, LLFlow, EnlightenGAN, Restormer, LEDNet, SNR-Aware, LLFormer, RetinexFormer, CIDNet, and MambaIR. Due to the limited space, more detail and references are provided in **Appendix Section 11**.

4.2 Ablation Study

Table **??** presents the results of our ablation study on TAMambaIR-S at Manga109 datasets, showing the impact of removing critical components from the model. Excluding positional embeddings (w/o PosEmb) reduces the PSNR to 40.11, indicating their importance in enhancing the model's ability to perceive and utilize spatial context, which is especially beneficial for accurately restoring global structures. Removing the Multi-Directional Perception Block (w/o MDPB) and using a single direction to replace it reduces the PSNR to 40.14, demonstrating its role in capturing features from multiple directions efficiently. Replacing TA-SSM with traditional SSM and keeping the similar complexity causes the PSNR to decline to 39.98. This emphasizes the crucial role of TA-SSM in prioritizing complex textures and effectively modeling challenging regions. The complete model

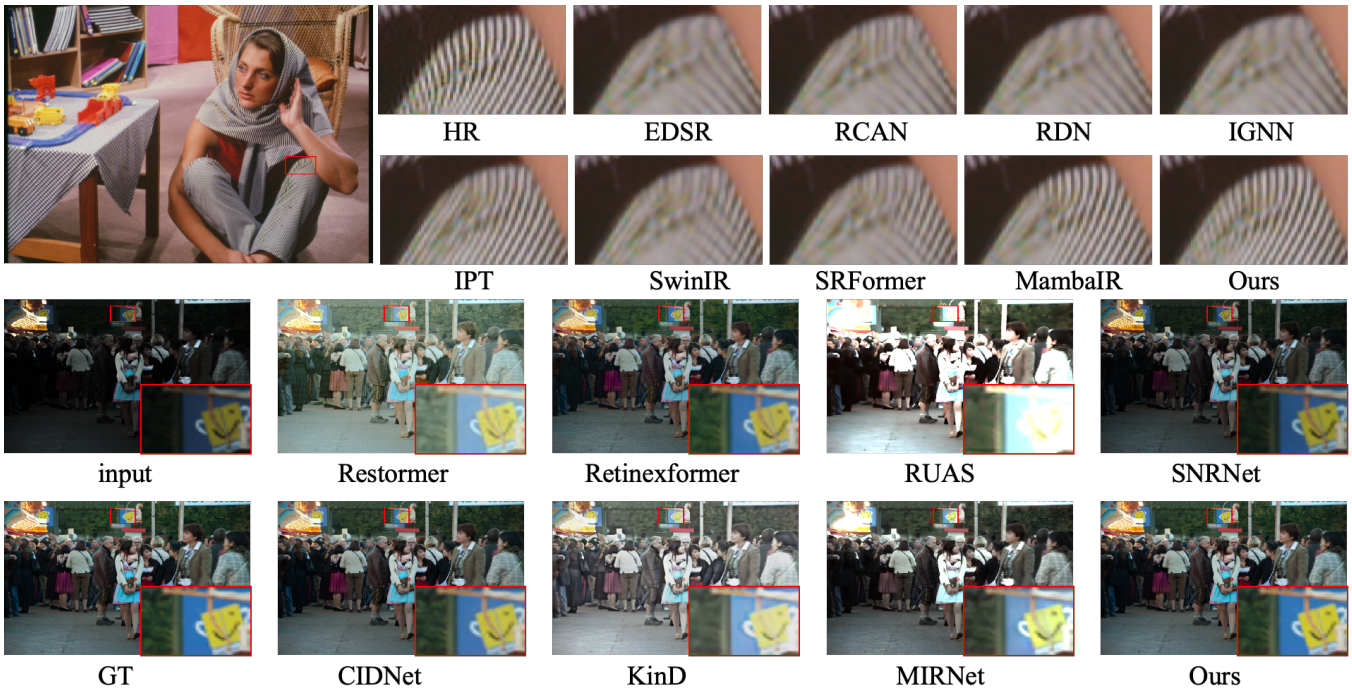

Figure 3: Qualitative comparison on image super-resolution and low-light image enhancement.

Methods	**Normal**		**GT Mean**	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
RetinexNet	17.137	0.762	19.099	0.774
KinD	13.290	0.578	16.259	0.591
ZeroDCE	17.712	0.815	21.463	0.848
3DLUT	18.040	0.800	22.173	0.854
DRBN	23.220	0.927	-	-
RUAS	13.765	0.638	16.584	0.719
LLFlow	24.807	0.919	27.961	0.930
EnlightenGAN	16.570	0.734	-	-
Restormer	21.413	0.830	25.428	0.859
LEDNet	23.709	0.914	27.367	0.928
SNR-Aware	24.140	0.928	27.787	0.941
LLFormer	24.038	0.909	28.006	0.927
RetinexFormer	25.670	0.930	28.992	0.939
CIDNet	25.705	0.942	29.566	0.950
MambaIR	<u>25.830</u>	<u>0.953</u>	<u>30.445</u>	<u>0.957</u>
TAMambaIR	**26.735**	**0.951**	**31.358**	**0.961**

Table 2: Quantitative comparison on **low-light image enhancement** with state-of-the-art methods.

Methods	Rain200L		Rain200H	
	PSNR↑	SSIM ↑	PSNR ↑	SSIM ↑
DDN	34.68	0.9671	26.05	0.8056
RESCAN	36.09	0.9697	26.75	0.8353
PReNet	37.80	0.9814	29.04	0.8991
MSPFN	38.58	0.9827	29.36	0.9034
RCDNet	39.17	0.9885	30.24	0.9048
MPRNet	39.47	0.9825	30.67	0.9110
SwinIR	40.61	0.9871	31.76	0.9151
DualGCN	40.73	0.9886	31.15	0.9125
SPDNet	40.50	0.9875	31.28	0.9207
Uformer	40.20	0.9860	30.80	0.9105
Restormer	40.99	0.9890	32.00	0.9329
IDT	40.74	0.9884	32.10	<u>0.9344</u>
DLINet	40.91	0.9886	31.47	0.9231
DRSformer	<u>41.23</u>	0.9894	32.17	0.9326
MambaIR	41.13	<u>0.9895</u>	<u>32.18</u>	0.9295
TAMambaIR	**41.25**	**0.9896**	**32.19**	**0.9345**

Table 3: Quantitative comparison on **image deraining** with state-of-the-art methods.

achieves the highest PSNR of 40.23, underscoring the necessity and complementary contributions of all components in delivering optimal performance. More ablation studies are provided in **Appendix Section 12**.

4.3 Quantitative Results

Image Super-Resolution. We compare our method against thirteen state-of-the-art CNN-based, transformer-based, and Mamba-based SR approaches. As shown in Table **??**, our method achieves the best performance across all four benchmarks. Specifically, the small version of our method surpasses SRFormer by 0.15 dB and 0.20 dB on the PSNR of Manga109 and Set14 datasets, respectively. Moreover, the standard version of our method outperforms the current state-of-the-art method, MambaIR, with PSNR improvements of 0.07 dB and 0.05 dB on the Manga109 and Set14 datasets, respectively. These results highlight the superior performance and effectiveness of our approach.

Model Complexity Comparison. As shown in Table **??**, benefiting from its strong texture-awareness capability, our

	w/o PosEmb	w/o MDPB	w/o TA-SSM	Ours
PSNR	40.11	40.14	39.98	40.23
SSIM	0.9801	0.9802	0.9796	0.9806

Table 4: Ablation studies on our proposed core module.

method not only achieves the best performance but also demonstrates significant advantages in computational efficiency, including FLOPs (with input size 64×64) and model parameters. Specifically, the small version of our model TAMambaIR-S achieves state-of-the-art performance with relatively low complexity, surpassing SRFormer in terms of performance while reducing FLOPs by 6.07G. Furthermore, the standard version of our network outperforms MambaIR while achieving reductions of 20.5G (18.5%) in FLOPs and 4.35M (21.3%) in parameters. These results highlight the efficiency of our method. More comparisons of inferencing time are provided in **Appendix Section 13**.

Image Deraining. We compare our method against fifteen state-of-the-art deraining approaches, including CNN-based, transformer-based, and Mamba-based methods. As shown in Table **??**, our method achieves the best performance on both Rain200L and Rain200H datasets in terms of PSNR and SSIM. Specifically, our method achieves a PSNR of 41.25 dB and an SSIM of 0.9896 on Rain200L, surpassing DRSformer, the previous best method, by 0.02 dB and 0.002, respectively. Similarly, on Rain200H, our approach attains a PSNR of 32.19 dB and an SSIM of 0.9345, outperforming MambaIR by 0.01 dB and 0.005, respectively.

Low-Light Image Enhancement. We compare our method with thirteen state-of-the-art approaches, including both CNN-, Transformer- and Mamba-based methods. We follow the Normal and GT Mean test settings in [Feng *et al.*, 2024], and the results are shown in Table **??**. Our method achieves the best performance on all settings in terms of PSNR and SSIM. Specifically, under the Normal setting, our method achieves a PSNR of 26.735 dB and an SSIM of 0.951, significantly surpassing RetinexFormer by 1.03 dB in PSNR and 0.021 in SSIM. Under the GT Mean setting, our method achieves a PSNR of 31.358 dB and an SSIM of 0.961, outperforming MambaIR by a notable margin. These results demonstrate that our approach effectively enhances low-light images while preserving details.

4.4 Qualitative Results

We present qualitative comparisons to demonstrate the effectiveness of our method on both image super-resolution (SR) and low-light image enhancement (LLIE), as shown at the top of Figure 3. For SR, our method reconstructs fine-grained textures and sharp details that are highly consistent with the ground truth (GT), in regions with complex textures and high-frequency details. This highlights the superiority of our approach in generating realistic and visually pleasing results. For LLIE, our method achieves significant improvements in both visual clarity and natural color restoration, as shown at the bottom of Figure 3. Compared to existing approaches, our results better preserve structural details and generate more accurate brightness adjustments, making

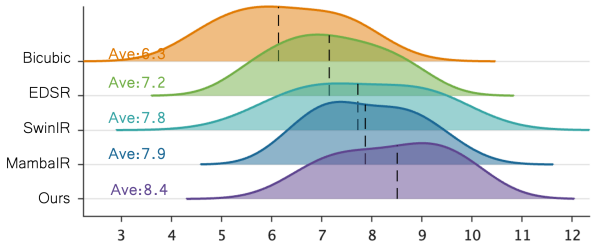

Figure 4: User study on image visual quality.

them closely aligned with the GT. More visual comparisons are provided in **Appendix Section 14**. Furthermore, to evaluate the visual quality, a user study is conducted for the image super-resolution. Specifically, we randomly select 15 images from the test datasets. Fifteen participants were asked to rate the visual quality of each processed image on a scale from 0 (poor quality) to 10 (excellent quality). The aggregated results, as shown in Figure 4, reveal that existing methods often fail to fully restore image quality, leading to lower user satisfaction. In contrast, our method achieves the highest average score of 8.4, demonstrating superior visual performance and generalization capabilities.

5 Limitaion and Future Work

While our TAMambaIR demonstrates strong performance on three representative restoration tasks, its evaluation is currently confined to these specific domains. Future work will aim to extend our method to a broader range of low-level vision tasks, such as image dehazing, debluring, denoising, and inpainting, among others, to further investigate its generalizability. Additionally, the top $p\%$ of different images in the current framework is uniform, which may not fully account for the varying texture complexities of different images. Images with richer textures often pose greater challenges than those with fewer textures. Developing adaptive processing strategies, such as texture-aware top $p\%$ selection mechanisms, could enhance performance by dynamically allocating computational resources based on image characteristics. These adaptations will be a key focus of our future research.

6 Conclusion

In this paper, we proposed TAMambaIR, a novel framework for image restoration that achieves a balance between high restoration quality and computational efficiency. Leveraging the statical characteristics of image degradation, a novel Texture-Aware State Space Model is introduced, which enhances texture awareness and improves performance by modulating the state-space equation and focusing more attention on texture-rich regions. Additionally, the Multi-Directional Perception Block expands the receptive field in multiple directions while maintaining low computational overhead. Extensive experiments on benchmarks for super-resolution, deraining, and low-light image enhancement demonstrate that TAMambaIR achieves state-of-the-art performance with significant improvements in efficiency, providing a robust and effective backbone for image restoration.

Acknowledgments

This work was supported by the Natural Science Foundation of China under Grants 62225207,62436008 and 62206262

Contribution Statement

Renjing Pei and Yang Wang are the corresponding authors. Long Peng and Xin Di contributed equally.

References

[Bevilacqua *et al.*, 2012] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on non-negative neighbor embedding. 2012.

[Cai *et al.*, 2023a] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12504–12513, October 2023.

[Cai *et al.*, 2023b] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *In ICCV*, pages 12504–12513, 2023.

[Chen *et al.*, 2021] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *CVPR*, pages 12299–12310, 2021.

[Chen *et al.*, 2023a] Xiang Chen, Hao Li, Mingqiang Li, and Jinshan Pan. Learning a sparse transformer network for effective image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5896–5905, June 2023.

[Chen *et al.*, 2023b] Xiang Chen, Hao Li, Mingqiang Li, and Jinshan Pan. Learning a sparse transformer network for effective image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5896–5905, June 2023.

[Cui *et al.*, 2023] Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Focal network for image restoration. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 13001–13011, 2023.

[Dai *et al.*, 2019] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11065–11074, 2019.

[Dong *et al.*, 2015] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015.

[Dosovitskiy *et al.*, 2020] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[Fang *et al.*, 2022] Jinsheng Fang, Hanjiang Lin, Xinyu Chen, and Kun Zeng. A hybrid network of cnn and transformer for lightweight image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1103–1112, 2022.

[Feng *et al.*, 2024] Yixu Feng, Cheng Zhang, Pei Wang, Peng Wu, Qingsen Yan, and Yanning Zhang. You only need one color space: An efficient network for low-light image enhancement, 2024.

[Guo *et al.*, 2024a] 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 *ECCV*, 2024.

[Guo *et al.*, 2024b] 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. *arXiv preprint arXiv:2402.15648*, 2024.

[Jeong *et al.*, 2025] Jinho Jeong, Jinwoo Kim, Younghyun Jo, and Seon Joo Kim. Accelerating image super-resolution networks with pixel-level classification. In *European Conference on Computer Vision*, pages 236–251. Springer, 2025.

[Kalman, 1960] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. 1960.

[Kingma and Ba, 2014] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[Kong *et al.*, 2021] Xiangtao Kong, Hengyuan Zhao, Yu Qiao, and Chao Dong. Classsr: A general framework to accelerate super-resolution networks by data characteristic. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12016–12025, June 2021.

[Li *et al.*, 2022] Chongyi Li, Chunle Guo, Linghao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. Low-light image and video enhancement using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):9396–9416, 2022.

[Li *et al.*, 2025] Zhuoyuan Li, Junqi Liao, Chuanbo Tang, Haotian Zhang, Yuqi Li, Yifan Bian, Xihua Sheng, Xinmin Feng, Yao Li, Changsheng Gao, et al. Ustc-td: A test dataset and benchmark for image and video coding in 2020s. *IEEE Transactions on Multimedia*, pages 1–16, 2025.

[Liang *et al.*, 2021a] 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.

[Liang *et al.*, 2021b] 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.

[Lim *et al.*, 2017] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017.

[Martin *et al.*, 2001] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 416–423. IEEE, 2001.

[Matsui *et al.*, 2017] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, 76:21811–21838, 2017.

[Peng *et al.*, 2024a] Long Peng, Yang Cao, Yuejin Sun, and Yang Wang. Lightweight adaptive feature de-drifting for compressed image classification. *IEEE Transactions on Multimedia*, 26:6424–6436, 2024.

[Peng *et al.*, 2024b] Long Peng, Wenbo Li, Jiaming Guo, Xin Di, Haoze Sun, Yong Li, Renjing Pei, Yang Wang, Yang Cao, and Zheng-Jun Zha. Unveiling hidden details: A raw data-enhanced paradigm for real-world super-resolution. *arXiv preprint arXiv:2411.10798*, 2024.

[Peng *et al.*, 2024c] Long Peng, Wenbo Li, Renjing Pei, Jingjing Ren, Jiaqi Xu, Yang Wang, Yang Cao, and Zheng-Jun Zha. Towards realistic data generation for real-world super-resolution. *arXiv preprint arXiv:2406.07255*, 2024.

[Peng *et al.*, 2025] Long Peng, Yang Wang, Xin Di, Xueyang Fu, Yang Cao, Zheng-Jun Zha, et al. Boosting image de-raining via central-surrounding synergistic convolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 6470–6478, 2025.

[Qiao *et al.*, 2024] Yanyuan Qiao, Zheng Yu, Longteng Guo, Sihan Chen, Zijia Zhao, Mingzhen Sun, Qi Wu, and Jing Liu. Vl-mamba: Exploring state space models for multimodal learning. *arXiv preprint arXiv:2403.13600*, 2024.

[Rim *et al.*, 2020] Jaesung Rim, Haeyun Lee, Jucheol Won, and Sunghyun Cho. Real-world blur dataset for learning and benchmarking deblurring algorithms. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.

[Sun *et al.*, 2023] Long Sun, Jiangxin Dong, Jinhui Tang, and Jinshan Pan. Spatially-adaptive feature modulation for efficient image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13190–13199, 2023.

[Timofte *et al.*, 2017] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 114–125, 2017.

[Tsai and Huang, 1984] Roger Y Tsai and Thomas S Huang. Multiframe image restoration and registration. *Multiframe image restoration and registration*, 1:317–339, 1984.

[Wang *et al.*, 2018] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *The European Conference on Computer Vision Workshops (ECCVW)*, September 2018.

[Xiao *et al.*, 2022] Jie Xiao, Xueyang Fu, Aiping Liu, Feng Wu, and Zheng-Jun Zha. Image de-raining transformer. *IEEE TPAMI*, 2022.

[Xiao *et al.*, 2025] Haoke Xiao, Lv Tang, Peng-tao Jiang, Hao Zhang, Jinwei Chen, and Bo Li. Boosting vision state space model with fractal scanning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 8646–8654, 2025.

[Yang *et al.*, 2017] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *CVPR*, pages 1357–1366, 2017.

[Yang *et al.*, 2021] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep retinex network for robust low-light image enhancement. volume 30, pages 2072–2086, 2021.

[Zamir *et al.*, 2022] 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.

[Zeyde *et al.*, 2012] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7*, pages 711–730. Springer, 2012.

[Zhang *et al.*, 2018] 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.

[Zheng *et al.*, 2022] Shen Zheng, Yiling Ma, Jinqian Pan, Changjie Lu, and Gaurav Gupta. Low-light image and video enhancement: A comprehensive survey and beyond. *arXiv preprint arXiv:2212.10772*, 2022.

[Zhou *et al.*, 2023] Yupeng Zhou, Zhen Li, Chun-Le Guo, Song Bai, Ming-Ming Cheng, and Qibin Hou. Srformer: Permuted self-attention for single image super-resolution. *arXiv preprint arXiv:2303.09735*, 2023.

[Zhu *et al.*, 2024] Lianghui Zhu, Bencheng Liao, Qian Zhang, Xinlong Wang, Wenyu Liu, and Xinggang Wang. Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint arXiv:2401.09417*, 2024.