ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-6874-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICASSP49660.2025.10889889

RestorMamba: An Enhanced Synergistic State Space Model for Image Restoration

Zeyu Wang
Computer Science and Technology
Zhejiang Normal University
JinHua, China
14797857499@zjnu.edu.cn

Chen Li
Computer Science and Technology
Zhejiang Normal University
JinHua, China
LilSodaChen@zjnu.edu.cn

Huiying Xu*
Computer Science and Technology
Zhejiang Normal University
JinHua, China
xhy@zjnu.edu.cn

Xinzhong Zhu
Computer Science and Technology
Zhejiang Normal University
JinHua, China
zxz@zjnu.edu.cn

Xiao Huang
College of Education
Zhejiang Normal University
JinHua, China
huangxiao@zjnu.edu.cn

Hongbo Li
Beijing Geekplus Technology
BeiJing, China
Jason.li@geekplus.com

***Abstract*—In this paper, we introduce an image inpainting method based on the State Space Model (SSM), named Restoration Mamba (RestorMamba). This approach incorporates efficient long-range dependency modeling within the network, which is particularly suited for the complexities of high-texture and high-resolution image restoration scenarios. To benefit from a broader context while maintaining global receptive fields, we have designed two pivotal modules: Skip Scan and Enhanced Synergistic Mamba (ESM) Block. Our experimental results demonstrate that RestorMamba achieves state-of-the-art performance in tasks such as image deraining and image denoising, encompassing Gaussian grayscale / color denoising and real image denoising.**

***Index Terms*—State Space Model, Image Restoration, Image Deraining, Image Denoising.**

I. Introduction

Image restoration techniques are pivotal for enhancing the perceptual quality and usability of visual data, encompassing tasks such as super-resolution, denoising, deblurring, and deraining. The advent of deep learning methodologies, particularly **C**onvolutional **N**eural **N**etwork (**CNN**) [1]–[3] and **Vi**sion **T**ransformer (**ViT**) [4]–[6], has propelled image restoration to new heights, achieving unprecedented detail recovery and noise suppression. While CNN-based restoration networks [7]–[10] excel in computational efficiency due to convolution operations, their inherently local nature limits the receptive field, which can be a critical drawback for tasks like denoising and deraining that benefit from broader contextual information. Transformer-based restoration networks [11]–[14], characterized by their extensive global receptive fields, generally outperform CNNs, especially when handling long-range dependencies within images. However, ViTs often struggle with the quadratic growth in computational complexity associated with their self-attention mechanism, leading to significant resource consumption as image resolution increases.

*Corresponding author.

Fig. 1. Visual comparison of image deraining methods on an image from Rain100H [17]. Zoom in for better visualization.

Recent advancements, Mamba [15] have integrated time-varying parameters into **S**tate **S**pace **M**odel (**SSM**), proposing an algorithm with linear complexity that is hardware-aware. This approach facilitates efficient training and inference while having global receptive fields, representing a promising method for efficient long-range dependency modeling in image restoration networks. Nonetheless, Mamba faces limitations: originally designed for 1-dimensional sequential data, its recent adaptation to visual backbones [16] may not adequately address the complexities of high-texture and high-resolution image restoration scenarios. Additionally, the unidirectional nature of Mamba's data modeling can lead to the neglect of local pixel interactions in two-dimensional images, posing certain limitations when processing adjacent pixel sequences. To address the aforementioned issues, we propose a novel scanning method called Skip Scan, which enhances the Mamba model's perception of local information. By maintaining spatial locality during the processing of visual data, we introduce the **E**nhanced **S**ynergistic **M**amba (**ESM**) Block and effectively apply it to image restoration tasks. As shown in Fig 1, we compared RestorMamba with representative image restoration methods; our method effectively preserved more detailed content while generating drop-free images. In summary, our contributions are as follows:

- We propose RestorMamba, a state-space model specifically designed for image restoration, which is capable of multi-scale local-global representation learning for high-resolution images. Additionally, we introduce Skip Scan,

a method that processes adjacent pixel sequences via leapfrog scanning while maintaining linear complexity.
- We design the ESM Block, which leverages the complementary features of Selective Scan and Skip Scan to enhance both local and non-local pixel interactions, making it effective for handling high-resolution images.
- Extensive experiments on multiple datasets for rain removal and image denoising demonstrate that our method achieves **S**tate-**O**f-**T**he-**A**rt (**SOTA**) performance, providing a new baseline for image restoration approaches.

II. Related Work

A. Image Restoration

Since the advent of deep learning, CNNs have risen to prominence due to their strong performance in image restoration tasks. Innovations such as residual blocks in ResNet [1] addressed the vanishing gradient problem, while attention mechanisms [18]–[20] enhanced the model's ability to focus on critical image features, thereby improving restoration [21] [22]. Despite these advances, CNNs' local receptive fields limit their capacity to capture global dependencies. Vision Transformers [4], which use the Self-Attention Mechanism to capture global dependencies, have emerged as a solution. However, the computational complexity of self-attention poses challenges for high-resolution images. Approaches such as IPT [11], which processes image patches separately, SwinIR [12] with its shifted window attention, and Restormer [13], integrating restoration priors and Fourier Transform, aim to mitigate these issues but still contend with high computational demands and large model sizes.

B. State Space Models (SSMs)

SSMs have gained favor as powerful tools for sequence modeling, particularly due to their superior performance in handling long-range dependencies. Mamba [15] is a data-dependent SSM equipped with a selection mechanism and efficient hardware design, which has shown impressive results in various visual tasks such as image classification [16], video understanding [23], and biomedical image segmentation [24]. However, Mamba has not explored the field of image restoration (deraining and denoising) far enough. Leveraging the long-range dependency and computational efficiency advantages of State Space Models, we propose an optimized design of Mamba tailored for the image restoration.

III. Method

A. Preliminaries

State Space Models [25] [15] map an input sequence $x(t) \in \mathbb{R}^L$ to an output sequence $y(t) \in \mathbb{R}^L$ via hidden states $h(t) \in \mathbb{R}^N$. The system dynamics are defined by matrices $A \in \mathbb{R}^{N\times N}$, $B \in \mathbb{R}^{N\times 1}$, and $C \in \mathbb{R}^{N\times 1}$.The update of the hidden state follows the following equation:

$$h'(t) = Ah(t) + Bx(t) \tag{1}$$

$$y(t) = Ch(t) \tag{2}$$

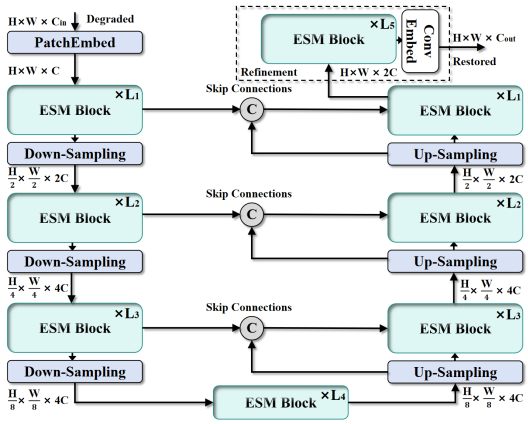

Fig. 2. The overall architecture of our RestorMamba.

In practical applications, under the **Z**ero-**O**rder **H**old (**ZOH**) assumption, a positive real number timescale Δ discretizes the continuous system's state-space matrices A and B into their discrete forms A_d and B_d.

$$A_d = \exp(\Delta A) \tag{3}$$

$$B_d = (\Delta A)^{-1}(e^{\Delta A} - I) \cdot \Delta B \tag{4}$$

Through these transformations, we obtain the discrete model, which can be expressed as:

$$h_t = A_d \cdot h_{t-1} + B_d \cdot x_t \tag{5}$$

$$y_t = C \cdot h_t \tag{6}$$

At each step t, the hidden state h_t is updated based on h_{t-1} and the current input x_t, with the output y_t derived from h_t via C. For high-dimensional input $x \in \mathbb{R}^{L\times D}$, the mapping to $\mathbb{R}^{L\times D\times N}$ uses $A_d, B_d \in \mathbb{R}^{D\times N}$ to project into an N-dimensional state space, and $C \in \mathbb{R}^{D\times N}$ maps $h_t \in \mathbb{R}^{D\times N}$ back to the observation space, yielding $y_t \in \mathbb{R}^D$. According to Mamba [15], this process can be reformulated as parallel convolutions to enhance computational efficiency:

$$y = x \star K \tag{7}$$

where $K = (CB, CAB, \cdots, CA^{L-1}B)$, $\star$ denotes the convolution operation, $\mathbf{K}$ is a convolution kernel consisting of a series of matrix products.

B. Model Architecture

Given the low-quality input image $I_0 \in \mathbb{R}^{H\times W\times C_{in}}$, where H and W denote the height and width of the image, and C_{in} represents the channels. The network initially employs a 3×3 convolutional layer to extract shallow features $X_L \in \mathbb{R}^{H\times W\times C}$. This architecture follows the design of U-Net, allowing the shallow features X_L to pass through a symmetric encoder-decoder structure. The encoder part consists of multiple stages $\{L_1, L_2, L_3, L_4\}$, each including a stack of ESM Blocks and a downsampling layer, with the numbers of blocks being $\{4, 6, 6, 8\}$ across the stages. The ESM Block leverages the inherent global modeling capabilities of state space models while enhancing local spatial and channel

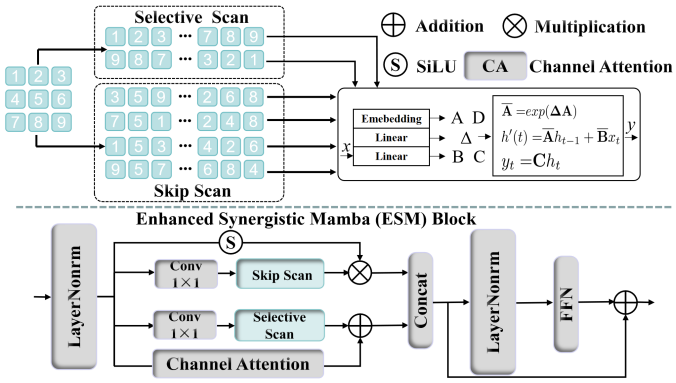

Fig. 3. The architecture of our ESM Block.

information interaction. Similarly, the decoder part utilizes stacked ESM Blocks and upsampling layers at each stage for feature reconstruction. To facilitate the restoration process, each decoding stage concatenates with corresponding low-level features from the encoder via skip connections, aiding in capturing fine structures and texture details in the restored image, resulting in $X_D \in \mathbb{R}^{H\times W\times 2C}$. Finally, during the refinement stage, a convolutional layer is applied to generate the high-quality output image $I \in \mathbb{R}^{H\times W\times C_{out}}$, as shown in Fig 2. Additionally, for a fair comparison, we adopt the following loss function in our training:

$$\mathcal{L}(I,\hat{I}) = \sqrt{||I-\hat{I}||_2 + \epsilon^2} \tag{8}$$

where $\hat{I}$ is the ground-truth image, and $\epsilon = 10^{-3}$ is a constant that remains unchanged across all experiments.

C. Enhanced Synergistic Mamba (ESM) Block

As shown in Fig 3, the ESM Block is designed to address the limitations of SSM and tackle complex challenges in the image restoration field through two scanning branches: skip scan and selective scan. Initially, we apply LayerNorm to the input feature map $\mathcal{F}_0$ to obtain $\mathcal{F}_{LN}$. Given the impact of background variability, which complicates the restoration process due to diverse backgrounds, the selective scan branch incorporates **C**hannel **A**ttention (**CA**) to weight different channels according to the unique data-dependent attributes of SSM. This allows it to dynamically focus on or ignore specific inputs based on their importance, with residual connections aggregating the output of the channel attention branch with the weighted feature maps:

$$\mathcal{F}_{se} = Selective\text{-}SCAN(Conv_{1\times1}(\mathcal{F}_{LN}) \oplus CA(\mathcal{F}_{LN})) \tag{9}$$

The Mamba model's long-range modeling capabilities lend it robust global degradation handling in image restoration tasks; however, its locality can be a weakness. The skip scan branch, on the other hand, refines local content correction, employing SiLU activation functions, and then element-wise multiplication of the outputs from both sub-branches:

$$\mathcal{F}_{sk} = Skip\text{-}SCAN(Conv_{1\times1}(\mathcal{F}_{LN}) \otimes \text{SiLU}(\mathcal{F}_{LN}) \tag{10}$$

The outputs from the enhanced selective scan branch and the skip scan branch are concatenated to form $\mathcal{F}_{Con} = Concat(\mathcal{F}_{se}, \mathcal{F}_{sk})$. Finally, adhering to the transformer block architecture, another LayerNorm is applied followed by a residual connection with the features generated by the **F**eed-**F**orward **N**etwork (**FFN**):

$$\mathcal{F}_D = \text{FFN}\left(LayerNorm(\mathcal{F}_{Con})\right) + \mathcal{F}_{Con}$$

The design of the ESM Block ensures that the distinct characteristics of global and local information are dynamically rebalanced, thus maintaining the integrity of both the global receptive field and the intricate local regions within the context of image restoration.

D. Selective Scan & Skip Scan

We propose an efficient Skip Scan approach that abandons the traditional strategy of cross-scanning entire regions and instead adopts a leapfrog scanning of adjacent pixels, with subsequent passes to revisit the skipped areas. This method enhances the model's capability to capture subtle local differences. Specifically, given an input feature map $X_G \in \mathbb{R}^{H\times W\times C}$, these regions are segmented into selected spatial subsets $\{K\}_i$, which can be formulated as:

$$X_G[:, m::p, n::p] \xrightarrow{scan} \{K\}_i \tag{11}$$

where m and n denote the starting positions along H and W, respectively, and p represents the stride length. Unlike traditional selective scan, Skip Scan is configured to perform four scans, feeding the obtained channel slices into the SSM for individual processing of the image block sequences, followed by merging the outputs to construct the final two-dimensional feature map. This process can be expressed as:

$$\{K\}_i \rightarrow \text{SSM}(\tilde{K}_i) \xrightarrow{mage} X_N \tag{12}$$

Notably, the Skip Scan operation will not to skip certain positions, as shown in Fig 3, it actually traverses the entire feature map, thus establishing local two-dimensional dependencies while preserving global information. Essentially, the features of Skip Scan complement the channel attention branch of selective scan, which aggregates global representations through averaging input features across the spatial dimensions, enhancing the model's representational power in image restoration.

IV. Experiments and Analysis

A. Experiment Setup

We trained using the AdamW optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay 1e-4) and L1 loss with an initial learning rate of 3e-4 , which was gradually reduced to 1e-6 by cosine annealing We trained with an 8 $\times$ GeForce RTX 2080 with a patch size of 128^2 and a batch size of 16 to start training, with a total of 300K iterations.

B. Comparison with Other Methods

We conducted exhaustive comparative experiments on three classical tasks in image restoration to verify the effectiveness of our method. We conducted quantitative comparisons across five **image deraining** datasets, employing the evaluation metrics of **P**eak **S**ignal-to-**N**oise **R**atio (**PSNR**) and **S**tructural

TABLE I
IMAGE DERAINING RESULTS.

Method	Test100 [26]		Rain100H [17]		Rain100L [17]		Test2800 [27]		Test1200 [28]		Average	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
DerainNet [29]	22.77	0.810	14.92	0.592	27.03	0.884	24.31	0.861	23.38	0.835	22.48	0.796
UMRL [30]	24.41	0.829	26.01	0.832	29.18	0.923	29.97	0.905	30.55	0.910	28.02	0.880
RESCAN [31]	25.00	0.835	26.36	0.786	29.80	0.881	31.29	0.904	30.51	0.882	28.59	0.857
MSPFN [7]	27.50	0.876	28.66	0.860	32.40	0.933	32.82	0.930	32.39	0.916	30.75	0.903
MPRNet [9]	30.27	0.897	30.41	0.890	36.40	0.965	33.64	0.938	32.91	0.916	32.73	0.921
SPAIR [32]	30.35	0.909	30.95	0.892	36.93	0.969	33.34	0.936	33.04	0.922	32.91	0.926
Restormer [13]	32.00	0.923	31.46	0.904	38.99	0.978	34.18	0.944	33.19	0.926	33.96	0.935
Fourmer [33]	30.54	0.911	30.76	0.896	37.47	0.970	-	-	33.05	0.921	32.96	0.924
VMambaIR [34]	31.66	0.909	-	-	**39.09**	0.979	34.01	0.944	33.33	0.926	32.27	0.939
RestorMamba	**32.67**	**0.929**	**31.57**	**0.911**	39.07	**0.982**	**34.23**	**0.958**	**33.40**	**0.935**	**34.19**	**0.943**

TABLE II
GAUSSIAN GRAYSCALE AND COLOR IMAGE DENOISING RESULTS.

Method	Set12 [35]			BSD68 [36]			Urban100 [37]			CBSD68 [38]			Kodak24 [39]		
	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$
DnCNN [35]	32.67	30.35	27.18	31.62	29.16	26.23	32.28	29.80	26.35	33.90	31.24	27.95	34.60	32.14	28.95
FFDNet [8]	32.75	30.43	27.32	31.63	29.19	26.29	32.40	29.90	26.50	33.87	31.21	27.96	34.63	32.13	28.98
IRCNN [40]	32.76	30.37	27.12	31.63	29.15	26.19	32.46	29.80	26.22	33.86	31.16	27.86	34.69	32.18	28.93
DRUNet [10]	33.25	30.94	27.90	31.91	29.48	26.59	33.44	31.11	27.96	34.30	31.69	28.51	35.31	32.89	29.86
Restormer [13]	33.42	31.08	28.00	31.96	29.52	26.62	33.79	31.46	28.29	34.40	31.79	28.60	35.47	**33.04**	30.01
SwinIR [12]	33.36	31.01	27.91	31.97	29.50	26.58	33.70	31.30	27.98	34.42	31.78	28.56	35.34	32.89	29.79
RestorMamba	**33.48**	**31.13**	**28.17**	**32.04**	**29.58**	**26.71**	**33.86**	**31.51**	**28.30**	**34.46**	**31.81**	**28.67**	**35.56**	32.98	**30.42**

TABLE III
REAL IMAGE DENOISING RESULTS.

	Method	DAGL [41]	Uformer [14]	MambaIR [42]	**RestorMamba**
SIDD [43]	PSNR ↑	38.94	39.77	39.89	**40.29**
	SSIM ↑	0.953	0.959	0.960	**0.962**
DND [44]	PSNR ↑	39.77	39.96	40.04	**40.12**
	SSIM ↑	0.956	0.956	0.956	**0.959**

TABLE IV
ABLATION EXPERIMENTS FOR THE RESTORMAMBA

Method	Component Details	Kodak24			Test100	
		σ=15	σ=25	σ=50	PSNR ↑	SSIM ↑
Scan operation	Baseline	35.47	33.04	30.01	32.00	0.923
	Selective-2-Direction	35.32	28.89	29.94	32.02	0.910
	Skip-2-Direction	35.45	29.31	30.22	32.41	0.924
	Skip-4-Direction	35.50	29.42	30.39	32.55	0.925
ESM Block	+ ESM	35.43	31.29	30.11	32.36	0.915
	+w/ CA	35.46	32.38	30.15	32.22	0.917
	+w/ CA + ESM	35.51	32.53	30.21	32.39	0.921
Overall	ESM Block+Skip Scan	35.56	32.98	30.42	32.67	0.929

Similarity **I**ndex **M**easure (**SSIM**) [47]. For **Gaussian Image Denoising**, we utilized the Set12 [35], Urban100 [37], BSD68 [36] (Grayscale), CBSD68 [38] and Kodak24 [39] (Color) datasets were employed; and for **Real Image Denoising**, we used the SIDD [43] and DND [44] datasets. The best and second-best performances are highlighted in bold and underlined, respectively.

a) Image Deraining: As shown in Table I, to provide a comprehensive evaluation of the algorithm's performance, we have presented the average values across these five datasets. Our proposed RestorMamba achieves SOTA performance, with an improvement of 0.23 dB in PSNR.

b) Gaussian grayscale and color image denoising: As shown in Table II, In challenging PSNR σ(noise levels)=50, our RestorMamba boosts gain by 0.17dB on the grayscale image Set12 dataset and 0.41dB on the color image Kodak24.

c) Real image denoising: As shown in Table III, On the DND datasets, the PSNR is improved by 0.16 dB compared to the Transformer-based model Uformer, improved by 0.40 dB compared to the SSM-based model MambaIR.

C. Ablation Studies

As shown in Table IV, For the ablation experiments, we tested on the Kodak24 dataset of Gaussian color image denoising by categorizing the PSNR into three noise levels of 15, 25, and 50, while we reported the PSNR and SSIM on the Test100 of Image Deraining.

In the experiment of scanning operation, we analyzed the effect of different directions and number of scans on the model. The proposed Skip Scan achieves superior performance in terms of accuracy. As can be seen from the change in the σ=50 metric, while the improvement in Skip Scan is noticeable, the synergistic effect of Selective Scan brought about by the ESM Block plays an important role. ESM Block is the core module of Restormamba, in the experiments of this module, we default the scanning method used by ESM to Selective-4-Direction, w/ CA stands for the branch that creates Channel Attention. Synergistic effects in employing channel attention can effectively enhance local interactions, and sub-optimal results can be obtained without Skip Scan.

V. CONCLUSION

In this paper, addressing the initial limitation of Mamba in handling flattened 2-dimensional data, we introduce the concept of skip scan to establish localized 2-dimensional dependencies. This enhancement allows the model to better capture spatial relationships within local regions during horizontal scans. Additionally, we propose the ESM Block to strengthen the interaction between local and non-local pixels, thereby enabling more efficient processing of high-resolution images. Extensive experiments conducted across multiple datasets have demonstrated the effectiveness of our proposed methodology.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (62376252); Key Project of Natural Science Foundation of Zhejiang Province (LZ22F030003); Zhejiang Province Leading Geese Plan (2024C02G1123882).

REFERENCES

[1] K. He, et al, "Deep residual learning for image recognition." in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[2] Z. Liu, et al, "A convnet for the 2020s," Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.

[3] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arxiv preprint arxiv:1409.1556 (2014).

[4] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv Preprint arXiv:2010.11929, 2020.

[5] Z. Liu, Y. Lin, Y. Cao, and et al., "Swin transformer: Hierarchical vision transformer using shifted windows." Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.

[6] Shi, Dai. "TransNeXt: Robust Foveal Visual Perception for Vision Transformers." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.

[7] K. Jiang, W. Zuo, and D. Zhang, "Multi-scale progressive fusion network for single image deraining." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.

[8] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising." IEEE Transactions on Image Processing, vol. 27, no. 9, pp. 4608–4622, 2018.

[9] S. W. Zamir, S. Khan, M. M. Riaz, and et al., "Multi-stage progressive image restoration." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.

[10] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Plug-and-play image restoration with deep denoiser prior." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 10, pp. 6360–6376, 2021.

[11] H. Chen, et al, "Pre-trained image processing transformer." in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.

[12] L. **gyun, et al, "SwinIR: Image Restoration Using Swin Transformer." in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.

[13] S. W. Zamir, et al, "Restormer: Efficient transformer for high-resolution image restoration." in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.

[14] Z. Wang, X. Yu, Y. Chen, and et al., "Uformer: A general u-shaped transformer for image restoration." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.

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

[16] Y. Liu, et al, "VMamba: Visual State Space Model." arXiv Preprint arXiv:2401.10166, 2024.

[17] W. Yang, et al, "Deep joint rain detection and removal from a single image. " in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[18] Q. Wang, Z. Wang, Y. Wang, Y. Wang, and T. Huang, "ECA-Net: Efficient channel attention for deep convolutional neural networks." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.

[19] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.

[20] S. Woo, J. Park, J. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module." In Proceedings of the European Conference on Computer Vision (ECCV), 2018.

[21] Y. Tian, Y. Liu, Y. Wang, and Y. Zhang, "CANet: Concatenated attention neural network for image restoration." IEEE Signal Processing Letters, vol. 27, pp. 1615–1619, 2020.

[22] Y. Mei, Y. Zhang, Y. Wang, and Y. Li, "Pyramid attention network for image restoration." International Journal of Computer Vision, vol. 131, no. 12, pp. 3207–3225, 2023.

[23] K. Li, et al, "VideoMamba: State Space Model for Efficient Video Understanding." arXiv Preprint arXiv:2403.06977, 2024.

[24] J. Ruan and S. Xiang, "Vm-unet: Vision mamba unet for medical image segmentation." arXiv preprint arXiv:2402.02491, 2024.

[25] A. Gu, K. Goel, and C. R, "Efficiently modeling long sequences with structured state spaces." arXiv Preprint arXiv:2111.00396, 2021.

[26] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network." IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 11, pp. 3943-3956, 2019.

[27] X. Fu, et al, "Removing rain from single images via a deep detail network. " in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[28] H. Zhang and V. M. Patel, "Density-aware single image de-raining using a multi-stream dense network." in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[29] X. Fu, J. Sun, J. Qin, and et al., "Clearing the skies: A deep network architecture for single-image rain removal." IEEE Transactions on Image Processing, vol. 26, no. 6, pp. 2944–2956, 2017.

[30] R. Yasarla and V. M. Patel, "Uncertainty guided multi-scale residual learning-using a cycle spinning CNN for single image de-raining." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.

[31] L. **a, W. Zuo, D. Zhang, and et al., "Recurrent squeeze-and-excitation context aggregation net for single image deraining." Proceedings of the European Conference on Computer Vision (ECCV), 2018.

[32] K. Purohit, S. K. Sahu, S. Prasad, and et al., "Spatially-adaptive image restoration using distortion-guided networks." Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.

[33] M. Zhou, Y. Xu, Z. Wang, and et al., "Fourmer: An efficient global modeling paradigm for image restoration." International Conference on Machine Learning, PMLR, 2023.

[34] Y. Shi, X. Wang, Y. Li, and et al., "Vmambair: Visual state space model for image restoration." arXiv preprint arXiv:2403.11423, 2024.

[35] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising." IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3142–3155, 2017.

[36] D. Martin, C. Fowlkes, D. Tal, and J. 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, Vol. 2, 2001.

[37] J. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.

[38] D. Martin, C. Fowlkes, D. Tal, and J. 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, Vol. 2, 2001.

[39] R. Franzen, "Kodak lossless true color image suite." http://r0k.us/graphics/kodak, 1999. Online; accessed 24-Oct-2021.

[40] A. Abdelhamed, S. Lin, and M. S. Brown, "A high-quality denoising dataset for smartphone cameras." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[41] C. Mou, J. Zhang, and Z. Wu, "Dynamic attentive graph learning for image restoration." In Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.

[42] Guo, Hang, et al. "Mambair: A simple baseline for image restoration with state-space model." arxiv preprint arxiv:2402.15648 (2024).

[43] A. Abdelhamed, S. Lin, and M. S. Brown, "A high-quality denoising dataset for smartphone cameras." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[44] T. Plotz and S. Roth, "Benchmarking denoising algorithms with real photographs." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[45] Z. Zou, J. Li, Y. Wang, and et al., "FreqMamba: Viewing Mamba from a Frequency Perspective for Image Deraining." ACM Multimedia, 2024.

[46] Y. Zhang, et al. "Kbnet: Kernel basis network for image restoration." arxiv preprint arxiv:2303.02881 (2023).

[47] A. Hore, and Z. Djemel. "Image quality metrics: PSNR vs. SSIM." 2010 20th international conference on pattern recognition. IEEE, 2010.