# MaIR: A Locality- and Continuity-Preserving Mamba for Image Restoration


Boyun Li[1], Haiyu Zhao[1], Wenxin Wang[1], Peng Hu[1], Yuanbiao Gou[1*], Xi Peng[1,2*]

[1] College of Computer Science, Sichuan University, China.

[2] National Key Laboratory of Fundamental Algorithms and Models for Engineering Numerical Simulation, Sichuan University, China.

{liboyun.gm, haiyuzhao.gm, wangwenxin.gm, penghu.ml, gouyuanbiao, pengx.gm}@gmail.com



## Abstract

*Recent advancements in Mamba have shown promising results in image restoration. These methods typically flatten 2D images into multiple distinct 1D sequences along rows and columns, process each sequence independently using selective scan operation, and recombine them to form the outputs. However, such a paradigm overlooks two vital aspects: i) the local relationships and spatial continuity inherent in natural images, and ii) the discrepancies among sequences unfolded through totally different ways. To overcome the drawbacks, we explore two problems in Mamba-based restoration methods: i) how to design a scanning strategy preserving both locality and continuity while facilitating restoration, and ii) how to aggregate the distinct sequences unfolded in totally different ways. To address these problems, we propose a novel Mamba-based Image Restoration model (MaIR), which consists of Nested S-shaped Scanning strategy (NSS) and Sequence Shuffle Attention block (SSA). Specifically, NSS preserves locality and continuity of the input images through the stripe-based scanning region and the S-shaped scanning path, respectively. SSA aggregates sequences through calculating attention weights within the corresponding channels of different sequences. Thanks to NSS and SSA, MaIR surpasses 40 baselines across 14 challenging datasets, achieving state-of-the-art performance on the tasks of image super-resolution, denoising, deblurring and dehazing. The code is available at https://github.com/XLearning-SCU/2025-CVPR-MaIR.*


## 1. Introduction

Image restoration aims to recover visually appealing high-quality images from given degraded correspondences, *e.g.*, noisy, blurry, and hazy images. In recent years, the methods based on Convolutional Neural Networks (CNNs) and Transformers have significantly advanced image restoration by effectively capturing locality (*i.e.*, fine-grained patterns and correlations in small regions) and continuity (*i.e.*, smooth, gradual transitions across larger areas) inherent in 2D natural images. To be specific, CNNs capture locality and continuity through the elaborately designed small kernels and sliding strides, respectively. Transformers capture them through local window partitions and adjacent window communications (*e.g.*, window shifts and window expansions). However, like a coin with two sides, the success of CNNs and Transformers in preserving locality and continu-

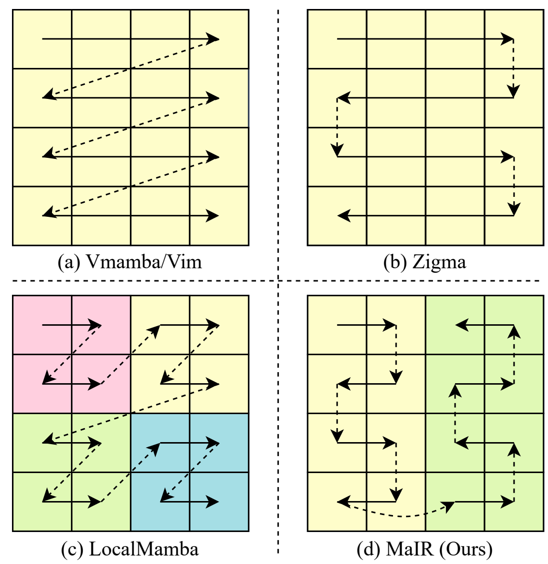


Figure 1. The scanning strategies in existing Mamba-based methods and our proposed method. (a) **Vmamba/Vim** uses Z-shaped scan path to flatten 2D image into 1D sequences, in which both the locality and continuity of 2D image are disrupted. (b) **Zigma** utilizes S-shaped path to maintain spatial continuity, while ignores the locality. (c) **LocalMamba** leverages window-based scanning region to preserve locality. However, the Z-shaped scanning path within and across the windows disrupts the spatial continuity. In contrast, (d) **MaIR** divides images into multiple non-overlapping stripes, and adopts S-shaped scanning path within and across the stripes, thus simultaneously preserves both locality and continuity.


*Corresponding author

ity comes at the cost of their ability to capture long-range dependencies. Both of them only consider a limited region of the input image at a time due to their localized kernels or windows, making them challenging to model relationships that span across larger sections of the image. Therefore, it is highly expected to develop a method that is able to capture long-range dependencies while well preserving locality and continuity inherent in 2D natural images.

Mamba [11, 15], a novel selective State Space Model [16], has garnered significant attention due to its promising performance in long sequence modeling while maintaining nearly linear complexity. As Mamba's core algorithm, Selective Scan Operation (SSO), is inherently designed for 1D sequences, it can not be directly applicable to processing 2D images. To address the problem, Mamba-based restoration methods typically involve a 3-step pipeline: i) flattening 2D image into multiple 1D sequences along rows and columns; ii) processing each sequence independently using SSO; and iii) aggregating the processed sequences to form the output 2D image. However, such a paradigm still faces two demerits when processing images. First, when transforming image into sequences, it disrupts the locality and continuity inherent in image, as illustrated in Fig. 1(a)-(c). Second, it generally aggregates processed sequences via pixel-wise summation, overlooking the distinct contexts among sequences unfolded through totally different ways.

In this work, we present a novel locality- and continuity-preserving Mamba for Image Restoration (MaIR), which consists of Nested S-shaped Scanning strategy (NSS) and Sequence Shuffle Attention block (SSA). Specifically, NSS preserves the locality through stripe-based scanning region, and the continuity via the S-shaped scanning path with shift-stripe mechanism. SSA aggregates the processed sequences by calculating attention weights within corresponding channels of sequences. Thanks to corporation of NSS and SSA, MaIR enjoys the following merits. Firstly, MaIR involves a cost-free solution to preserve the locality and continuity inherent in natural images, ensuring structural coherence and avoiding computational overhead. Secondly, MaIR captures complex dependencies across distinct sequences, facilitating to leverage complementary information from both forward and reversed rows and columns.

To summarize, the contributions and innovations of this work are as below:

- In this work, we present MaIR, an approach that efficiently captures long-range dependencies while preserving the locality and continuity inherent in natural images.
- For Mamba, we introduce NSS, a cost-free solution to preserve locality and continuity, and SSA, a module to capture dependencies across distinct sequences.
- MaIR obtains state-of-the-art performance on four tasks across 14 benchmarks comparing with 40 baselines.

## 2. Related Works

In this section, we will briefly review related works in image restoration and vision Mamba.

### 2.1. Image Restoration

According to the focus of this paper, existing methods can be classified into three categories, *i.e.*, CNN-, Transformer- and Mamba-based methods. We will introduce the first two categories here, while the last one is detailed in Sec. 2.2.

**CNN-based Method:** Benefiting from the ability of capturing locality and continuity in natural images, CNN-based methods have achieved promising results in various tasks of image restoration, such as image super-resolution [10, 26, 28, 40, 64], image denoising [14, 25, 44, 57, 65] and image deblurring [39, 39, 48, 60]. However, since their localized receptive fields, CNNs are inherently limited in capturing long-range dependencies.

**Transformer-based Method:** Transformers are theoretically capable of capturing the global dependencies [53, 66]. However, to avoid impractical quadratic complexity on images, existing methods [7, 27, 32] tend to partition the local regions of input image into different windows, and calculate attentions within or across the windows. For instance, SwinIR [27] computes attentions within local windows and shifts these windows between layers. HAT [7] divides images into overlapping windows to enhance the interaction between neighbor windows. Although these methods have ensured structural coherence (*i.e.*, locality and continuity) of natural images and avoided computational overhead, they fell into another dilemma of failing to fully capture long-range dependencies due to their limited window sizes.

### 2.2. Vision Mamba

Due to Mamba's demonstrated superiority in long-sequence modeling [9, 16, 46], some studies have introduced it into high- [21, 31, 70] and low-level [12, 18, 67] vision tasks. To enable SSO to process images, these methods [31, 70] tend to flatten 2D images into multiple 1D sequences along the different directions. For instance, Vmamba [31] proposes cross-scan strategy which flattens input images along rows and columns. However, existing scanning strategies disrupt structure coherence which is essential for image restoration. Recently, some Mamba-based restoration methods have begun to recognize the importance of structure coherence, and tend to introduce extra coherence-preserving modules. For instance, MambaIR [18] and UVM-Net [67] enhances locality through additional CNN layers, but introduces extra computational costs. Although some other studies [19, 21] devote to designing scanning strategy to preserve locality and continuity, most of them can only preserve one of them. In contrast, MaIR provides a cost-free solution to preserve both locality and continuity.

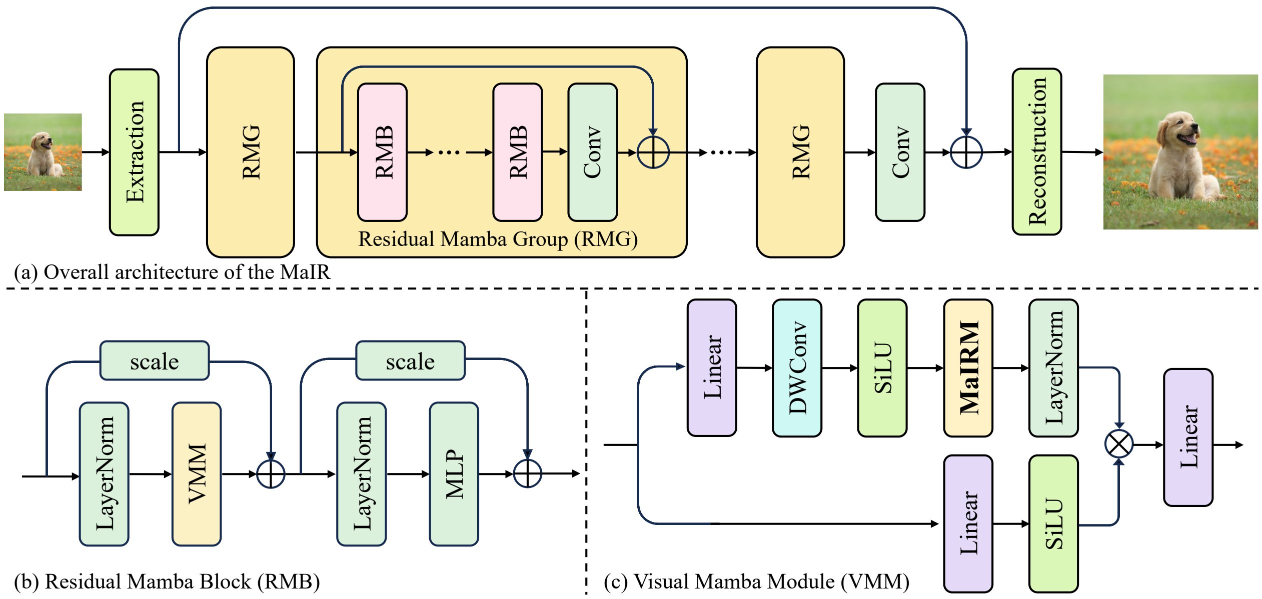


Figure 2. Illustrations of MaIR. (a) The overall architecture of MaIR, highlighting its core component, Residual Mamba Group (RMG). RMG is primarily composed of (b) Residual Mamba Block (RMB), in which (c) Visual Mamba Module (VMM) plays a pivotal role.

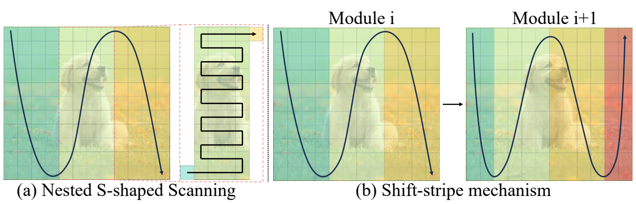


Figure 3. Illustrations of (a) Nested S-shaped Scanning strategy (NSS) and (b) shift-stripe mechanism.

# 3. Methods

In this section, we first introduce the overall architecture of our MaIR, and then elaborate on NSS and SSA assembled in MaIR Module (MaIRM).

## 3.1. Overall Architecture

**Network Structure:** Following previous works [18, 27], MaIR is built up with three stages, namely, shallow feature extraction stage, deep feature extraction stage and reconstruction stage. Specifically, in the shallow feature extraction stage, for a given degraded image $x \in \mathcal{R}^{3\times H\times W}$, we first employ a convolution layer to extract shallow feature $F_S \in \mathcal{R}^{C\times H\times W}$, where $H$ and $W$ represent the height and width of $x$, and $C$ is the number of channels. After that, $F_S$ is fed to the deep feature extraction stage to produce deep feature $F_D \in \mathcal{R}^{C\times H\times W}$. As illustrated in Fig. 2, the deep feature extraction stage is stacked by multiple Residual Mamba Groups (RMGs), where each RMG consists of several Residual Mamba Blocks (RMBs). Within each RMB, a Visual Mamba Module (VMM) is introduced to capture long-range dependencies, which is further composed of our proposed MaIRM. Finally, we reconstruct the high-quality image based on $F_S$ and $F_D$. Specifically, for image super-resolution, we introduce a pixel-shuffle layer $U_{ps}(\cdot)$ and a $3 \times 3$ convolution layer $\Phi_{3\times 3}(\cdot)$ to reconstruct the high-resolution image $y' = \Phi_{3\times 3}(U_{ps}(F_S + F_D))$. For tasks that do not require upsampling (*e.g.*, denoising, deblurring and dehazing), we employ single convolution layer with residual connection to construct high-quality result, which can be formulated as $y' = \Phi_{3\times 3}(F_S + F_D) + x$.

**Loss Function:** For image super-resolution, we use $L_1$ loss to optimize the network following [18, 27, 69], which can be formulated as

$$\mathcal{L} = \|y - y'\|_1,$$

where $y$ is the target image. For image denoising, deblurring and dehazing, we adopt Charbonnier loss, *i.e.*,

$$\mathcal{L} = \sqrt{\|y - y'\|^2 + \epsilon^2},$$

where $\epsilon$ is a hyper-parameter and set to $10^{-3}$ empirically.

## 3.2. MaIR Module

As elaborated above, MaIRM serves as the core module of MaIR, which involves a three-step pipeline. To be specific, MaIRM first flattens 2D features into four 1D sequences through NSS along four distinct directions following [31]. Then, MaIRM employs SSO to capture long-range dependencies. Finally, MaIRM aggregates processed sequences through SSA to form outputs. Mathematically, for input feature $F_{i,j}$, output feature $F^M_{i,j}$ can be formulated as

$$\begin{aligned} F^M_{i,j} &= M_{i,j}(F_{i,j}), \\ &= \Phi^{SSA}_{i,j}(\Phi^{SSO}_{i,j}(\Phi^{NSS}_{i,j}(F_{i,j}))), \end{aligned}$$

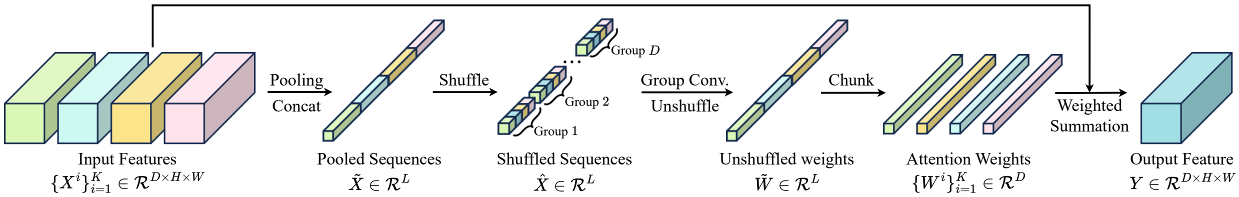


Figure 4. Illustration of the Sequence Shuffle Attention (SSA). The input features $\{X^i\}_{i=1}^{K} \in \mathcal{R}^{D\times H\times W}$ are first pooled and concatenated to form $\tilde{X} \in \mathcal{R}^L$, where $L = K \times D$. This sequence undergoes the sequence shuffle operation and results in shuffled sequences $\hat{X} \in \mathcal{R}^L$, whose channels are split by $D$ group. Then, group convolution and sequence unshuffle operation are applied, producing unshuffled weights $\tilde{W} \in \mathcal{R}^L$, which are further chunked and reshaped into attention weights $\{W^i\}_{i=1}^{K} \in \mathcal{R}^D$. Finally, the output feature $Y \in \mathcal{R}^{D\times H\times W}$ is computed by performing a weighted summation of the input features using the attention weights.

where $M_{i,j}(\cdot)$, $\Phi_{i,j}^{NSS}(\cdot)$, $\Phi_{i,j}^{SSO}(\cdot)$ and $\Phi_{i,j}^{SSA}(\cdot)$ are MaIRM, NSS, SSO and SSA in the $j$-th RMB of the $i$-th RMG, respectively.

**NSS:** NSS is designed to extract locality- and continuity-preserving sequences from input features. Motivated by the observation illustrated in Fig. 1, one could find that i) LocalMamba [21] preserves locality through restricted scanning region, and ii) Zigma [19] preserves continuity through S-shaped scanning path. Thus, as shown in Fig. 3(a), we design the nested S-shaped scanning strategy, which divides features into multiple non-overlapping stripes and uses S-shaped scanning path within and across stripes to maintain both locality and continuity. To better leverage spatial information, we extract sequences with four different scanning directions: top-left to bottom-right, bottom-right to top-left, top-right to bottom-left, and bottom-left to top-right, following previous works [18, 31].

Besides, NSS includes shift-stipe mechanism to preserve locality and continuity on the boundary regions between adjacent stripes. As depicted in Fig. 3(b), for two successive modules, the first module partitions features into multiple non-overlapping stripes with stripe width $w_s$. For the second module, we employ the shift-stripe operation, and set the first and last stripe widths as $\frac{w_s}{2}$ and others' width as $w_s$. Consequently, the boundary regions in the previous module will be fully covered by a single stripe in this module.

**SSA:** SSA aggregates the processed sequences by calculating attentions within corresponding channels. This design enables it to capture complex dependencies across distinct sequences, thus better leveraging complementary information from different scanning directions. As shown in Fig. 4, supposing sequence number $K = 4$, for SSO-processed sequences $\{X^i\}_{i=1}^{4}$, we first apply spatial average pooling $\Phi_{AP}(\cdot)$ to reduce the computational cost, and then concatenate as

$$\begin{aligned}\tilde{X} &= \Phi_{cat}(\Phi_{AP}(\{X^i\}_{i=1}^{4})) \\ &= [x_1^1, \cdots, x_D^1, x_1^2, \cdots, x_D^2, x_1^3, \cdots, x_D^3, x_1^4, \cdots, x_D^4],\end{aligned}$$

where $x_d^k$ is the pooled feature in $d$-th channel of $k$-th sequence, and $D$ is the number of channel in MaIRM. Then, we employ sequence shuffle operation $\Phi_{ss}(\cdot)$ to rearrange features into

$$\begin{aligned}\hat{X} &= \Phi_{ss}(X) \\ &= [x_1^1, x_1^2, x_1^3, x_1^4, x_2^1, x_2^2, x_2^3, x_2^4, \cdots, x_D^1, x_D^2, x_D^3, x_D^4].\end{aligned}$$

After that, we employ group convolution $\Phi_g(\cdot)$ with group size four to obtain the channel-wise attention weights and unshuffle the weights back to their original order, *i.e.*,

$$\begin{aligned}\tilde{W} &= \Phi_{su}(\Phi_g(\hat{X})) \\ &= [w_1^1, \cdots, w_D^1, w_1^2, \cdots, w_D^2, w_1^3, \cdots, w_D^3, w_1^4, \cdots, w_D^4],\end{aligned}$$

where $\Phi_{su}(\cdot)$ is sequence unshuffle operation. The unshuffled weights $\tilde{W}$ are chunked as $\{W^i\}_{i=1}^{4} = \Phi_{chunk}(\tilde{W})$, where $\Phi_{chunk}(\cdot)$ refers to the chunk operation. Finally, we adopt weight summation based on $\{W^i\}_{i=1}^{4}$ to generate the output, which can be formulated as:

$$Y = \sum_{i=1}^{K=4} W^i * X^i,$$

and $Y$ is the output sequence of SSA.

# 4. Experiments

In this section, we evaluate our MaIR on four representative image restoration tasks, *i.e.*, image super-resolution, image denoising, image deblurring, and image dehazing. Experimental settings and visual results will be presented in the supplementary materials.

## 4.1. Results on Image Super-Resolution

**Datasets:** Following [18, 27], we employ DF2K (DIV2K [49]+Flickr2K [29]) and DIV2K as training sets for classic and lightweight image super-resolution, respectively. For evaluation, we employ Set5 [3], Set14 [54], B100 [35], Urban100 [20] and Manga109 [36] as test sets. Following existing works [10, 40, 58, 64], the low-resolution images are downsampled from the corresponding high-resolution images via bicubic interpolation.

Table 1. Quantitative results on classic image super-resolution. The best and second best results are in red and blue.

| Methods | Scale | Set5 | | Set14 | | B100 | | Urban100 | | Manga109 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| SAN [10] | ×2 | 38.31 | 0.9620 | 34.07 | 0.9213 | 32.42 | 0.9028 | 33.10 | 0.9370 | 39.32 | 0.9792 |
| HAN [40] | ×2 | 38.27 | 0.9614 | 34.16 | 0.9217 | 32.41 | 0.9027 | 33.35 | 0.9385 | 39.46 | 0.9785 |
| IGNN [68] | ×2 | 38.24 | 0.9613 | 34.07 | 0.9217 | 32.41 | 0.9025 | 33.23 | 0.9383 | 39.35 | 0.9786 |
| NLSA [37] | ×2 | 38.34 | 0.9618 | 34.08 | 0.9231 | 32.43 | 0.9027 | 33.42 | 0.9394 | 39.59 | 0.9789 |
| ELAN [63] | ×2 | 38.36 | 0.9620 | 34.20 | 0.9228 | 32.45 | 0.9030 | 33.44 | 0.9391 | 39.62 | 0.9793 |
| IPT [4] | ×2 | 38.37 | - | 34.43 | - | 32.48 | - | 33.76 | - | - | - |
| SwinIR [27] | ×2 | 38.42 | 0.9623 | 34.46 | 0.9250 | 32.53 | 0.9041 | 33.81 | 0.9427 | 39.92 | 0.9797 |
| SRFormer [69] | ×2 | 38.51 | 0.9627 | 34.44 | 0.9253 | 32.57 | 0.9046 | 34.09 | 0.9449 | 40.07 | 0.9802 |
| MambaIR [18] | ×2 | 38.57 | 0.9627 | 34.67 | 0.9261 | 32.58 | 0.9048 | 34.15 | 0.9446 | 40.28 | 0.9806 |
| MaIR | ×2 | 38.56 | 0.9628 | 34.75 | 0.9268 | 32.59 | 0.9049 | 34.19 | 0.9451 | 40.30 | 0.9807 |
| MaIR+ | ×2 | 38.62 | 0.9630 | 34.82 | 0.9272 | 32.62 | 0.9053 | 34.38 | 0.9462 | 40.48 | 0.9811 |
| SAN [10] | ×3 | 34.75 | 0.9300 | 30.59 | 0.8476 | 29.33 | 0.8112 | 28.93 | 0.8671 | 34.30 | 0.9494 |
| HAN [40] | ×3 | 34.75 | 0.9299 | 30.67 | 0.8483 | 29.32 | 0.8110 | 29.10 | 0.8705 | 34.48 | 0.9500 |
| IGNN [68] | ×3 | 34.72 | 0.9298 | 30.66 | 0.8484 | 29.31 | 0.8105 | 29.03 | 0.8696 | 34.39 | 0.9496 |
| NLSA [37] | ×3 | 34.85 | 0.9306 | 30.70 | 0.8485 | 29.34 | 0.8117 | 29.25 | 0.8726 | 34.57 | 0.9508 |
| ELAN [63] | ×3 | 34.90 | 0.9313 | 30.80 | 0.8504 | 29.38 | 0.8124 | 29.32 | 0.8745 | 34.73 | 0.9517 |
| IPT [4] | ×3 | 34.81 | - | 30.85 | - | 29.38 | - | 29.49 | - | - | - |
| SwinIR [27] | ×3 | 34.97 | 0.9318 | 30.93 | 0.8534 | 29.46 | 0.8145 | 29.45 | 0.8826 | 35.12 | 0.9537 |
| SRFormer [69] | ×3 | 35.02 | 0.9323 | 30.94 | 0.8540 | 29.48 | 0.8156 | 30.04 | 0.8865 | 35.26 | 0.9543 |
| MambaIR [18] | ×3 | 35.08 | 0.9323 | 30.99 | 0.8536 | 29.51 | 0.8157 | 29.93 | 0.8841 | 35.43 | 0.9546 |
| MaIR | ×3 | 35.10 | 0.9324 | 31.05 | 0.8541 | 29.51 | 0.8160 | 30.05 | 0.8863 | 35.44 | 0.9547 |
| MaIR+ | ×3 | 35.15 | 0.9328 | 31.12 | 0.8550 | 29.56 | 0.8167 | 30.24 | 0.8881 | 35.67 | 0.9556 |
| SAN [10] | ×4 | 32.64 | 0.9003 | 28.92 | 0.7888 | 27.78 | 0.7436 | 26.79 | 0.8068 | 31.18 | 0.9169 |
| HAN [40] | ×4 | 32.64 | 0.9002 | 28.90 | 0.7890 | 27.80 | 0.7442 | 26.85 | 0.8094 | 31.42 | 0.9177 |
| IGNN [68] | ×4 | 32.57 | 0.8998 | 28.85 | 0.7891 | 27.77 | 0.7434 | 26.84 | 0.8090 | 31.28 | 0.9182 |
| NLSA [37] | ×4 | 32.59 | 0.9000 | 28.87 | 0.7891 | 27.78 | 0.7444 | 26.96 | 0.8109 | 31.27 | 0.9184 |
| ELAN [63] | ×4 | 32.75 | 0.9022 | 28.96 | 0.7914 | 27.83 | 0.7459 | 27.13 | 0.8167 | 31.68 | 0.9226 |
| IPT [4] | ×4 | 32.64 | - | 29.01 | - | 27.82 | - | 27.26 | - | - | - |
| SwinIR [27] | ×4 | 32.92 | 0.9044 | 29.09 | 0.7950 | 27.92 | 0.7489 | 27.45 | 0.8254 | 32.03 | 0.9260 |
| SRFormer [69] | ×4 | 32.93 | 0.9041 | 29.08 | 0.7953 | 27.94 | 0.7502 | 27.68 | 0.8311 | 32.21 | 0.9271 |
| MambaIR [18] | ×4 | 33.03 | 0.9046 | 29.20 | 0.7961 | 27.98 | 0.7503 | 27.68 | 0.8287 | 32.32 | 0.9272 |
| MaIR | ×4 | 32.93 | 0.9045 | 29.20 | 0.7958 | 27.98 | 0.7507 | 27.71 | 0.8305 | 32.46 | 0.9284 |
| MaIR+ | ×4 | 33.14 | 0.9058 | 29.28 | 0.7974 | 28.02 | 0.7516 | 27.89 | 0.8336 | 32.66 | 0.9297 |

**Baselines:** We compare our method with 15 competitive baselines. Specifically, we adopt four CNN-based methods (*i.e.*, SAN [10], HAN [40], IGNN [68], and NLSA [37]), four transformer-based methods (*i.e.*, ELAN [63], IPT [4], SwinIR [27] and SRFormer [69]) and one Mamba-based method (*i.e.,* MambaIR [18]) as the baselines for classic super-resolution. For lightweight super-resolution, four CNN-based methods (*i.e.*, CARN [2], IMDN [22], LAPAR [26], LatticeNet [33]), two transformer-based methods (*i.e.*, SwinIR [27] and SRFormer [69]) and one Mamba-based method (*i.e.,* MambaIR [18]) are introduced in comparisons. Similar to MambaIR, which offers two versions for lightweight super-resolution, MaIR is also available in two configurations: MaIR-Tiny and MaIR-Small.

**Results:** For classic super-resolution, as shown in Tab. 1, MaIR achieves the best result in almost all quantitative comparisons. For instance, our method surpasses MambaIR [18] with 0.03dB∼0.12dB in terms of PSNR on Urban100, and SRFormer with at most 0.04dB, 0.10dB and 0.25dB in terms of PSNR on B100, Urban100, and Manga109, respectively, which demonstrates the superiority of MaIR. For lightweight SR, MaIR also exhibits its advancement compared to baselines as reported in Tab. 2. Taking ×4 scale as examples, MaIR-Small surpasses MambaIR-Small by 0.08dB in terms of PSNR on Manga109 with fewer parameters and MACs. MaIR-Tiny outperforms MambaIR-Tiny and SwinIR by 0.08dB and 0.12dB in terms of PSNR on Urban100 with fewer parame-

Table 2. Quantitative results on lightweight image super-resolution. The best and second best results are in red and blue.

| Methods | Scale | Params | MACs | Set5 | | Set14 | | B100 | | Urban100 | | Manga109 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| CARN [2] | ×2 | 1,592K | 222.8G | 37.76 | 0.9590 | 33.52 | 0.9166 | 32.09 | 0.8978 | 31.92 | 0.9256 | 38.36 | 0.9765 |
| IMDN [22] | ×2 | 694K | 158.8G | 38.00 | 0.9605 | 33.63 | 0.9177 | 32.19 | 0.8996 | 32.17 | 0.9283 | 38.88 | 0.9774 |
| LAPAR-A [26] | ×2 | 548K | 171.0G | 38.01 | 0.9605 | 33.62 | 0.9183 | 32.19 | 0.8999 | 32.10 | 0.9283 | 38.67 | 0.9772 |
| LatticeNet [33] | ×2 | 756K | 169.5G | 38.15 | 0.9610 | 33.78 | 0.9193 | 32.25 | 0.9005 | 32.43 | 0.9302 | - | - |
| SwinIR [27] | ×2 | 910K | 122.2G | 38.14 | 0.9611 | 33.86 | 0.9206 | 32.31 | 0.9012 | 32.76 | 0.9340 | 39.12 | 0.9783 |
| MambaIR-Tiny [18] | ×2 | 905K | 167.1G | 38.13 | 0.9610 | 33.95 | 0.9208 | 32.31 | 0.9013 | 32.85 | 0.9349 | 39.20 | 0.9782 |
| MaIR-Tiny | ×2 | 878K | 207.8G | 38.18 | 0.9610 | 33.89 | 0.9209 | 32.31 | 0.9013 | 32.89 | 0.9346 | 39.22 | 0.9778 |
| MambaIR-Small [18] | ×2 | 1,363K | 567.5G | 38.16 | 0.9610 | 34.00 | 0.9212 | 32.34 | 0.9017 | 32.92 | 0.9356 | 39.31 | 0.9779 |
| MaIR-Small | ×2 | 1,355K | 542.0G | 38.20 | 0.9611 | 33.91 | 0.9209 | 32.34 | 0.9016 | 32.97 | 0.9359 | 39.32 | 0.9779 |
| CARN [2] | ×3 | 1,592K | 111.8G | 34.29 | 0.9255 | 30.29 | 0.8407 | 29.06 | 0.8034 | 28.06 | 0.8493 | 33.50 | 0.9440 |
| IMDN [22] | ×3 | 703K | 71.5G | 34.36 | 0.9270 | 30.32 | 0.8417 | 29.09 | 0.8046 | 28.17 | 0.8519 | 33.61 | 0.9445 |
| LAPAR-A [26] | ×3 | 544K | 144.0G | 34.36 | 0.9267 | 30.34 | 0.8421 | 29.11 | 0.8054 | 28.15 | 0.8523 | 33.51 | 0.9441 |
| LatticeNet [33] | ×3 | 765K | 76.3G | 34.53 | 0.9281 | 30.39 | 0.8424 | 29.15 | 0.8059 | 28.33 | 0.8538 | - | - |
| SwinIR [27] | ×3 | 918K | 55.4G | 34.62 | 0.9289 | 30.54 | 0.8463 | 29.20 | 0.8082 | 28.66 | 0.8624 | 33.98 | 0.9478 |
| MambaIR-Tiny [18] | ×3 | 913K | 74.5G | 34.63 | 0.9288 | 30.54 | 0.8459 | 29.23 | 0.8084 | 28.70 | 0.8631 | 34.12 | 0.9479 |
| MaIR-Tiny | ×3 | 886K | 93.0G | 34.68 | 0.9292 | 30.54 | 0.8461 | 29.25 | 0.8088 | 28.83 | 0.8651 | 34.21 | 0.9484 |
| MambaIR-Small [18] | ×3 | 1,371K | 252.7G | 34.72 | 0.9296 | 30.63 | 0.8475 | 29.29 | 0.8099 | 29.00 | 0.8689 | 34.39 | 0.9495 |
| MaIR-Small | ×3 | 1,363K | 241.4G | 34.75 | 0.9300 | 30.63 | 0.8479 | 29.29 | 0.8103 | 28.92 | 0.8676 | 34.46 | 0.9497 |
| CARN [2] | ×4 | 1,592K | 90.9G | 32.13 | 0.8937 | 28.60 | 0.7806 | 27.58 | 0.7349 | 26.07 | 0.7837 | 30.47 | 0.9084 |
| IMDN [22] | ×4 | 715K | 40.9G | 32.21 | 0.8948 | 28.58 | 0.7811 | 27.56 | 0.7353 | 26.04 | 0.7838 | 30.45 | 0.9075 |
| LAPAR-A [26] | ×4 | 659K | 94.0G | 32.15 | 0.8944 | 28.61 | 0.7818 | 27.61 | 0.7366 | 26.14 | 0.7871 | 30.42 | 0.9074 |
| LatticeNet [33] | ×4 | 777K | 43.6G | 32.30 | 0.8962 | 28.68 | 0.7830 | 27.62 | 0.7367 | 26.25 | 0.7873 | - | - |
| SwinIR [27] | ×4 | 930K | 31.8G | 32.44 | 0.8976 | 28.77 | 0.7858 | 27.69 | 0.7406 | 26.48 | 0.7980 | 30.92 | 0.9151 |
| MambaIR-Tiny [18] | ×4 | 924K | 42.3G | 32.42 | 0.8977 | 28.74 | 0.7847 | 27.68 | 0.7400 | 26.52 | 0.7983 | 30.94 | 0.9135 |
| MaIR-Tiny | ×4 | 897K | 53.1G | 32.48 | 0.8985 | 28.81 | 0.7864 | 27.71 | 0.7414 | 26.60 | 0.8013 | 31.13 | 0.9161 |
| MambaIR-Small [18] | ×4 | 1,383K | 143.0G | 32.51 | 0.8993 | 28.85 | 0.7876 | 27.75 | 0.7423 | 26.75 | 0.8051 | 31.26 | 0.9175 |
| MaIR-Small | ×4 | 1,374K | 136.6G | 32.62 | 0.8998 | 28.90 | 0.7882 | 27.77 | 0.7431 | 26.73 | 0.8049 | 31.34 | 0.9183 |

ters, verifying both efficiency and effectiveness of MaIR.

## 4.2. Results on Image Denoising

**Datasets:** For synthetic noise removal, we train MaIR on DFWB, which consists of DIV2K, Flickr2K, Waterloo Exploration Dataset (WED) [34] and BSD400 [35]. For evaluation, we utilize BSD68 [35], Kodak24, McMaster [62], and Urban100 as test set. Following [27, 57–59], we generate noisy images by manually adding white Gaussian noise to the clean images with three distinct noise levels, *i.e.*, $\sigma = 15, 25, 50$. For real-world image denoising, our model is trained and tested on the SIDD-Medium [1] dataset, which provides 320 high-resolution noisy-clean image pairs for training and additional 40 image pairs for test.

**Baselines:** We compare our MaIR with 14 representative methods. To be specific, we adopt four CNN-based methods (*i.e.*, IRCNN [58], FFDNet [59], DnCNN [57] and DRUNet [61]), four transformer-based methods (*i.e.*, SwinIR [27], Restormer [53], CODE [66] and ART [56]) and one Mamba-based method (*i.e.,* MambaIR [18]) as the baselines for synthetic noise removal. For real-world image denoising, four CNN-based methods (*i.e.*, DeamNet [43], MPRNet [52], NBNet [8] and DAGL [38]), two transformer-based methods (*i.e.*, Uformer [50] and Restormer [53]) and one Mamba-based method (*i.e.,* MambaIR [18]) are introduced for comparisons.

**Results:** As depicted in the Tab. 3-4, MaIR demonstrates superior performance on both synthetic and real-world image denoising compared to baselines. Taking results on Urban100 as examples, MaIR averagely outperforms MambaIR by 0.21dB in terms of PSNR, indicating its superiority on image denoising.

## 4.3. Results on Image Deblurring

**Datasets:** Following previous works [52, 53], we employ GoPro dataset [39] for training which consists of 2,103 blurry-clean image pairs. For evaluation, we use two common datasets, *i.e.*, GoPro test set and HIDE [45], which consist of 1,111 and 2,025 blurry-clean pairs, respectively.

**Baselines:** We adopt 11 competitive image deblurring baselines for comparisons, including six CNN-based deblurring methods (*i.e.*, SRN [48], DBGAN [60], DMPHN [55], MIMO [9], MPRNet [52], and NAFNet [6]), three transformer-based methods (*i.e.*, CODE [66], Restormer [53] and Uformer [50]), one RNN-based method (*i.e.,* MT-RNN [41]) and one Mamba-based method (*i.e.*, CU-Mamba [12]).

**Results:** As shown in Tab. 5, MaIR surpasses other baselines by PSNR on both GoPro and HIDE. For instance, MaIR outperforms Restormer [53] by 0.77dB on the GoPro dataset and by 0.35dB on the HIDE dataset in terms of PSNR. Although NAFNet achieves similar quantitative results on GoPro, MaIR surpasses NAFNet on HIDE dataset by 0.25dB in terms of PSNR.

Table 3. Quantitative results on gaussian color image denoising. The best and second best results are in red and blue.

| Methods | BSD68 | | | Kodak24 | | | McMaster | | | Urban100 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\sigma$=15 | $\sigma$=25 | $\sigma$=50 | $\sigma$=15 | $\sigma$=25 | $\sigma$=50 | $\sigma$=15 | $\sigma$=25 | $\sigma$=50 | $\sigma$=15 | $\sigma$=25 | $\sigma$=50 |
| IRCNN [58] | 33.86 | 31.16 | 27.86 | 34.69 | 32.18 | 28.93 | 34.58 | 32.18 | 28.91 | 33.78 | 31.20 | 27.70 |
| FFDNet [59] | 33.87 | 31.21 | 27.96 | 34.63 | 32.13 | 28.98 | 34.66 | 32.35 | 29.18 | 33.83 | 31.40 | 28.05 |
| DnCNN [57] | 33.90 | 31.24 | 27.95 | 34.60 | 32.14 | 28.95 | 33.45 | 31.52 | 28.62 | 32.98 | 30.81 | 27.59 |
| DRUNet [61] | 34.30 | 31.69 | 28.51 | 35.31 | 32.89 | 29.86 | 35.40 | 33.14 | 30.08 | 34.81 | 32.60 | 29.61 |
| SwinIR [27] | 34.42 | 31.78 | 28.56 | 35.34 | 32.89 | 29.79 | 35.61 | 33.20 | 30.22 | 35.13 | 32.90 | 29.82 |
| Restormer [53] | 34.40 | 31.79 | 28.60 | 35.47 | 33.04 | 30.01 | 35.61 | 33.34 | 30.30 | 35.13 | 32.96 | 30.02 |
| CODE [66] | 34.33 | 31.69 | 28.47 | 35.32 | 32.88 | 29.82 | 35.38 | 33.11 | 30.03 | - | - | - |
| ART [56] | 34.46 | 31.84 | 28.63 | 35.39 | 32.95 | 29.87 | 35.68 | 33.41 | 30.31 | 35.29 | 33.14 | 30.19 |
| MambaIR [18] | 34.43 | 31.80 | 28.61 | 35.34 | 32.91 | 29.85 | 35.62 | 33.35 | 30.31 | 35.17 | 32.99 | 30.07 |
| MaIR | 34.48 | 31.86 | 28.66 | 35.53 | 33.09 | 30.04 | 35.71 | 33.44 | 30.35 | 35.35 | 33.22 | 30.30 |
| MaIR+ | 34.50 | 31.88 | 28.69 | 35.56 | 33.13 | 30.08 | 35.74 | 33.48 | 30.39 | 35.42 | 33.30 | 30.41 |

Table 4. Quantitative results on real image denoising. The best and second best results are in red and blue.

| | DeamNet [43] | MPRNet [52] | NBNet [8] | DAGL [38] | Uformer [50] | MambaIR [18] | MaIR |
|---|---|---|---|---|---|---|---|
| PSNR | 39.47 | 39.71 | 39.75 | 38.94 | 39.89 | 39.89 | 39.92 |
| SSIM | 0.957 | 0.958 | 0.959 | 0.953 | 0.960 | 0.960 | 0.960 |

Table 5. Quantitative results on image motion deblurring. The best and second best results are in red and blue. MACs in this table are evaluated on 128×128 patches followed [66].

| Method | Params | MACs | GoPro | HIDE |
|---|---|---|---|---|
| SRN [48] | 3.76M | 35.87G | 30.26 | 28.36 |
| DBGAN [60] | 11.59M | 379.92G | 31.10 | 28.94 |
| MT-RNN [41] | 2.64M | 13.72G | 31.15 | 29.15 |
| DMPHN [55] | 86.80M | - | 31.20 | 29.09 |
| CODE [66] | 12.18M | 22.52G | 31.94 | 29.67 |
| MIMO+ [9] | 16.10M | 38.64G | 32.45 | 29.99 |
| MPRNet [52] | 20.13M | 194.42G | 32.66 | 30.96 |
| Restormer [53] | 26.13M | 35.31G | 32.92 | 31.22 |
| Uformer [50] | 50.88M | 22.36G | 33.06 | 30.90 |
| CU-Mamba [12] | 19.7M | - | 33.53 | 31.47 |
| NAFNet [6] | 67.89M | 15.85G | 33.69 | 31.32 |
| MaIR | 26.29M | 49.29G | 33.69 | 31.57 |

## 4.4. Results on Image Dehazing

**Datasets:** Following [47], we employ RESIDE [24] for training and testing. For indoor scenes, we train MaIR on Indoor Training Set (ITS) which consists of 13,990 hazy-clean pairs, and test it on indoor synthetic objective testing set (SOTS-Indoor) involving 500 pairs. For outdoor scenes, we train MaIR on Outdoor Training Set (OTS), which contains 313,950 image pairs, and evaluate it on outdoor synthetic objective testing set (SOTS-Outdoor) involving 500 images. To verify MaIR on more general cases, we also train the model on RESIDE-6K and test it on the SOTS-mix, which contain both indoor and outdoor images.

**Baselines:** We compare MaIR with eight baselines, including five CNNs (AODNet [23], GDN [30], MSBDN [13], FFANet [42] and AECRNet [51]), two transformers (Dehamer [17], Dehazeformer [47]) and one Mamba (UVM-Net [67]).

**Results:** As shown in Tab. 6, MaIR surpasses most baselines on quantitative comparisons. Specifically, MaIR outperforms DehazeFormer and UVM-Net by 2.67dB and 2.04dB in terms of PSNR on the outdoor scenes. Although UVM-Net is slightly higher on PSNR in the indoor scenes, MaIR only takes 0.3% and 4.8% params and MACs of the UVM-Net, verifying both effectiveness and efficiency.

## 4.5. Analysis Experiments

### 4.5.1. Ablation Studies

We first verify the effectiveness of NSS under five configurations, i) replacing NSS with Z-shaped scanning strategy (denoted as w/o NSS), ii) removing shift stripe (denoted as w/o SS), iii) replacing NSS by Window-based scanning strategy [21] (denoted as LM), iv) replacing NSS by Z-shaped scanning strategy [19] (denoted as ZigMa) and v) replacing NSS by the Peano-Hilbert curve (denoted as PH). As illustrated in Tab. 7, NSS is important for MaIR.

To verify the effectiveness of SSA, we remove SSA and aggregate sequences through: i) sequences-wise addition (termed as w/o SSA), ii) SSM [67] (termed as UVM), iii) sequence-wise gating [5] (termed as SeqGat), iv) channel-wise gating (termed as CAGat), v) pixel-wise gating through fully connected convolution (termed as FPixGat). vi) pixel-wise gating through depth-wise convolution (termed as DWPixGat). The size of different models are set be similar for fair comparisons. As shown in Tab. 8, SSA is more effective than others.

### 4.5.2. Verification of Observations

To verify observations shown in Fig. 1, we conduct visual comparisons among different scanning strategies. As shown

Table 6. Quantitative results on image dehazing. The best and second best results are in red and blue. MACs in this table are evaluated on 256×256 patches followed [47, 67].

| Method | | AODNet [23] | GDN [30] | MSBDN [13] | FFANet [42] | AECRNet [51] | Dehamer [17] | DehazeFormer [47] | UVM-Net [67] | MaIR (Ours) |
|---|---|---|---|---|---|---|---|---|---|---|
| Params | | 0.002M | 0.96M | 31.35M | 4.46M | 2.61M | 132.45M | 4.63M | 1,003.94M | 3.40M |
| MACs | | 0.115G | 21.49G | 41.54G | 287.8G | 52.20G | 48.93G | 48.64G | 501.91G | 24.03G |
| SOTS-Indoor | PSNR | 20.51 | 32.16 | 33.67 | 36.39 | 37.17 | 36.63 | 38.46 | 40.17 | 39.45 |
| | SSIM | 0.816 | 0.984 | 0.985 | 0.989 | 0.990 | 0.988 | 0.994 | 0.996 | 0.997 |
| SOTS-Outdoor | PSNR | 24.14 | 30.86 | 33.48 | 33.57 | - | 35.18 | 34.29 | 34.92 | 36.96 |
| | SSIM | 0.920 | 0.982 | 0.982 | 0.984 | - | 0.986 | 0.983 | 0.984 | 0.991 |
| SOTS-Mix | PSNR | 20.27 | 25.86 | 28.56 | 29.96 | 28.52 | - | 30.89 | 31.92 | 31.52 |
| | SSIM | 0.855 | 0.944 | 0.966 | 0.973 | 0.964 | - | 0.977 | 0.982 | 0.980 |

Input | MambaIR (Z-shaped) | LocalMamba (Window-based) | ZigMa (S-shaped) | MaIR (Nested S-shaped) | HR

Figure 5. Visual comparisons of different scanning strategies, illustrating that i) windows-based scanning path overlooks the continuity between different regions (*e.g.*, relationship between different layers of the scarf), resulting in wrong textures, ii) S-shaped scanning path leads to distortion in local regions, causing the scarf's texture to appear warped. iii) Z-shaped scanning path suffers from both of them. In contrast, MaIR avoids aforementioned problems and achieves visually appealing results.

Table 7. Ablation study on the proposed NSS scheme, tested on lightweight super-resolution tasks with scale factor ×2. The results on the Urban100 dataset are presented, which demonstrates the effectiveness of the NSS.

| | Baseline | w/o NSS | w/o SS | LM | ZigMa | PH |
|---|---|---|---|---|---|---|
| PSNR | 32.97 | 32.94 | 32.93 | 32.93 | 32.88 | 32.95 |
| SSIM | 0.9359 | 0.9355 | 0.9351 | 0.9357 | 0.9354 | 0.9356 |

Table 8. Ablation study on SSA for lightweight super-resolution with scale factor ×2. The results on Urban100 dataset demonstrate the effectiveness of SSA.

| | MaIR | w/o SSA | UVM | SeqGat |
|---|---|---|---|---|
| PSNR | 32.97 | 32.90 | 32.59 | 32.92 |
| SSIM | 0.9359 | 0.9350 | 0.9324 | 0.9352 |
| | CAGat | FPixGat | DWPixGat | - |
| PSNR | 32.71 | 31.90 | 32.74 | - |
| SSIM | 0.9335 | 0.9250 | 0.9342 | - |

in Fig. 5, MaIR can maintain both locality and continuity and produce more visual pleasant results.

#### 4.5.3. Results on Different Stripe Width

To investigate influence of stripe width, we train and evaluate models with stripe width $w_s = \{2, 4, 8, 16, 32\}$ on Urban100 dataset. As presented in Tab. 9, the PSNR and SSIM values are quite similar under different settings, except for the cases with the largest and the smallest stripe widths. It indicates that the proposed method exhibits robustness against changes in stripe width, maintaining high-quality image restoration across a range of stripe widths.

Table 9. Analyses on stripe widths. Experiment was conducted on the Urban100 dataset with a scale factor ×2 for lightweight super-resolution tasks, which illustrates how changes in stripe width affect the restored image quality.

| | 2 | 4 | 8 | 16 | 32 |
|---|---|---|---|---|---|
| PSNR | 32.95 | 32.97 | 32.97 | 32.97 | 32.92 |
| SSIM | 0.9356 | 0.9359 | 0.9355 | 0.9357 | 0.9355 |

## 5. Conclusion

In this paper, we propose MaIR, a novel state space model for image restoration that can preserve both local dependencies and spatial continuity of input images. To this end, we propose two designs: Nested S-shaped Scanning strategy (NSS) and Sequences Shuffle Attention (SSA). NSS is designed to extract locality- and continuity-preserving sequences from images, and SSA adaptively aggregates these sequences. Thanks to their cooperation, MaIR not only addresses the limitations of existing Mamba-based restoration methods but also improves image quality without introducing extra computations. Extensive experiments across four tasks on 14 benchmarks comparing with 40 baselines, validate the superiority of MaIR, demonstrating its robustness and effectiveness in various image restoration tasks.

## Acknowledgments

This work was supported in part by NSFC under Grant 62176171, U21B2040, 62472295; in part by the Fundamental Research Funds for the Central Universities under Grant CJ202303; and in part by Sichuan Science and Technology Planning Project under Grant 24NSFTD0130.


## References


[1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1692–1700, 2018. 6

[2] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *Proceedings of the European conference on computer vision (ECCV)*, pages 252–268, 2018. 5, 6

[3] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. 2012. 4

[4] 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 *IEEE Conference on Computer Vision and Pattern Recognition*, pages 12299–12310, Virtual, 2021. 5

[5] Keyan Chen, Bowen Chen, Chenyang Liu, Wenyuan Li, Zhengxia Zou, and Zhenwei Shi. Rsmamba: Remote sensing image classification with state space model. *IEEE Geoscience and Remote Sensing Letters*, 2024. 7

[6] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European conference on computer vision*, pages 17–33. Springer, 2022. 6, 7

[7] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22367–22377, 2023. 2

[8] Shen Cheng, Yuzhi Wang, Haibin Huang, Donghao Liu, Haoqiang Fan, and Shuaicheng Liu. Nbnet: Noise basis learning for image denoising with subspace projection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4896–4906, 2021. 6, 7

[9] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4641–4650, 2021. 2, 6, 7

[10] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 11065–11074, 2019. 2, 4, 5

[11] Tri Dao and Albert Gu. Transformers are SSMs: Generalized models and efficient algorithms through structured state space duality. In *International Conference on Machine Learning (ICML)*, 2024. 2

[12] Rui Deng and Tianpei Gu. Cu-mamba: Selective state space models with channel learning for image restoration. *arXiv preprint arXiv:2404.11778*, 2024. 2, 6, 7

[13] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-Scale Boosted Dehazing Network with Dense Feature Fusion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2154–2164, Seattle, WA, 2020. 7, 8

[14] Yuanbiao Gou, Boyun Li, Zitao Liu, Songfan Yang, and Xi Peng. Clearer: Multi-scale neural architecture search for image restoration. *Advances in Neural Information Processing Systems*, 33, 2020. 2

[15] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023. 2

[16] Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*, 2021. 2

[17] Chun-Le Guo, Qixin Yan, Saeed Anwar, Runmin Cong, Wenqi Ren, and Li Chongyi. Image dehazing transformer with transmission-aware 3d position embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 7, 8

[18] 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. 2, 3, 4, 5, 6, 7

[19] Vincent Tao Hu, Stefan Andreas Baumann, Ming Gui, Olga Grebenkova, Pingchuan Ma, Johannes S Fischer, and Björn Ommer. Zigma: A dit-style zigzag mamba diffusion model. *arXiv preprint arXiv:2403.13802*, 2024. 2, 4, 7

[20] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5197–5206, 2015. 4

[21] Tao Huang, Xiaohuan Pei, Shan You, Fei Wang, Chen Qian, and Chang Xu. Localmamba: Visual state space model with windowed selective scan. *arXiv preprint arXiv:2403.09338*, 2024. 2, 4, 7

[22] Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. Lightweight image super-resolution with information multi-distillation network. In *Proceedings of the 27th acm international conference on multimedia*, pages 2024–2032, 2019. 5, 6

[23] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. AOD-Net: All-in-One Dehazing Network. In *IEEE International Conference on Computer Vision*, pages 4780–4788, Venice, Italy, 2017. 7, 8

[24] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking Single Image Dehazing and Beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2019. 7

[25] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-In-One Image Restoration for Unknown Corruption. In *IEEE Conference on Computer Vision and*

*Pattern Recognition*, pages 17431–17441, New Orleans, LA, 2022. 2

[26] Wenbo Li, Kun Zhou, Lu Qi, Nianjuan Jiang, Jiangbo Lu, and Jiaya Jia. Lapar: Linearly-assembled pixel-adaptive regression network for single image super-resolution and beyond. *Advances in Neural Information Processing Systems*, 33:20343–20355, 2020. 2, 5, 6

[27] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image Restoration Using Swin Transformer. In *International Conference on Computer Vision Workshops*, Virtual, 2021. 2, 3, 4, 5, 6, 7

[28] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition WorkShop*, pages 1132–1140, 2017. 2

[29] 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. 4

[30] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing. In *International Conference on Computer Vision*, pages 7313–7322, Seoul, Korea, 2019. 7, 8

[31] Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, and Yunfan Liu. Vmamba: Visual state space model. *arXiv preprint arXiv:2401.10166*, 2024. 2, 3, 4

[32] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021. 2

[33] Xiaotong Luo, Yuan Xie, Yulun Zhang, Yanyun Qu, Cuihua Li, and Yun Fu. Latticenet: Towards lightweight image super-resolution with lattice block. In *European Conference on Computer Vision*, pages 272–289, 2020. 5, 6

[34] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. Waterloo Exploration Database: New Challenges for Image Quality Assessment Models. *IEEE Transactions on Image Processing*, 26(2):1004–1016, 2017. 6

[35] 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 *International Conference on Computer Vision*, pages 416–425, Vancouver, Canada, 2001. 4, 6

[36] 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. 4

[37] Yiqun Mei, Yuchen Fan, and Yuqian Zhou. Image super-resolution with non-local sparse attention. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3517–3526, 2021. 5

[38] Chong Mou, Jian Zhang, and Zhuoyuan Wu. Dynamic attentive graph learning for image restoration. In *IEEE International Conference on Computer Vision*, 2021. 6, 7

[39] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 257–265, Honolulu, HI, 2017. 2, 6

[40] Ben Niu, Weiwei Wen, Wenqi Ren, Xiangde Zhang, Lianping Yang, Shuzhen Wang, Kaihao Zhang, Xiaochun Cao, and Haifeng Shen. Single image super-resolution via a holistic attention network. In *European Conference on Computer Vision*, pages 191–207, 2020. 2, 4, 5

[41] Dongwon Park, Dong Un Kang, Jisoo Kim, and Se Young Chun. Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training. In *European Conference on Computer Vision*, pages 327–343. Springer, 2020. 6, 7

[42] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. FFA-Net: Feature Fusion Attention Network for Single Image Dehazing. In *AAAI Conference on Artificial Intelligence*, pages 11908–11915, New York, NY, 2020. 7, 8

[43] Chao Ren, Xiaohai He, Chuncheng Wang, and Zhibo Zhao. Adaptive consistency prior based deep network for image denoising. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8596–8606, 2021. 6, 7

[44] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126(9):973–992, 2018. 2

[45] Ziyi Shen, Wenguan Wang, Xiankai Lu, Jianbing Shen, Haibin Ling, Tingfa Xu, and Ling Shao. Human-aware motion deblurring. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5572–5581, 2019. 6

[46] Jimmy TH Smith, Andrew Warrington, and Scott Linderman. Simplified state space layers for sequence modeling. In *The Eleventh International Conference on Learning Representations*, 2022. 2

[47] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32:1927–1941, 2023. 7, 8

[48] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8174–8182, 2018. 2, 6, 7

[49] 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. 4

[50] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17683–17693, 2022. 6, 7

[51] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive Learning for Compact Single Image Dehazing. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 10551–10560, Virtual, 2021. 7, 8

[52] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-Stage Progressive Image Restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 14821–14831, Virtual, 2021. 6, 7

[53] 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 *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5718–5729, New Orleans, LA, 2022. 2, 6, 7

[54] 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. 4

[55] Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5978–5986, 2019. 6, 7

[56] Jiale Zhang, Yulun Zhang, Jinjin Gu, Yongbing Zhang, Linghe Kong, and Xin Yuan. Accurate image restoration with attention retractable transformer. In *ICLR*, 2023. 6, 7

[57] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. 2, 6, 7

[58] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3929–3938, 2017. 4, 6, 7

[59] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for CNN based image denoising. *IEEE Transactions on Image Processing*, 2018. 6, 7

[60] Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Bjorn Stenger, Wei Liu, and Hongdong Li. Deblurring by realistic blurring. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2737–2746, 2020. 2, 6, 7

[61] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6360–6376, 2021. 6, 7

[62] Lei Zhang, Xiaolin Wu, Antoni Buades, and Xin Li. Color demosaicking by local directional interpolation and nonlocal adaptive thresholding. *Journal of Electronic imaging*, 20(2): 023016–023016, 2011. 6

[63] Xindong Zhang, Hui Zeng, Shi Guo, and Lei Zhang. Efficient long-range attention network for image super-resolution. In *European Conference on Computer Vision*, pages 649–667. Springer, 2022. 5

[64] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *European Conference on Computer Vision*, pages 294–310, 2018. To appear in ECCV 2018. 2, 4

[65] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual Dense Network for Image Restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7):2480–2495, 2018. 2

[66] Haiyu Zhao, Yuanbiao Gou, Boyun Li, Dezhong Peng, Jiancheng Lv, and Xi Peng. Comprehensive and delicate: An efficient transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14122–14132, 2023. 2, 6, 7

[67] Zhuoran Zheng and Chen Wu. U-shaped vision mamba for single image dehazing. *arXiv preprint arXiv:2402.04139*, 2024. 2, 7, 8

[68] Shangchen Zhou, Jiawei Zhang, Wangmeng Zuo, and Chen Change Loy. Cross-scale internal graph neural network for image super-resolution. In *Neural Information Processing Systems*, 2020. NeurIPS 2020. 5

[69] Yupeng Zhou, Zhen Li, Chun-Le Guo, Song Bai, Ming-Ming Cheng, and Qibin Hou. Srformer: Permuted self-attention for single image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 12780–12791, 2023. 3, 5

[70] 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. 2