Supplementary Materials for
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

1. Introduction

In this material, we present details of the experimental settings and qualitative results.

1.1. Experimental Settings

In this section, we will introduce the evaluation metrics and training details of experiments.

Evaluation metrics: Following [2, 5], two popular metrics are employed for quantitative evaluations, *i.e.*, Peak Signal-to-Noise Ratio (PSNR) [3] and Structure Similarity (SSIM) [6]. Higher values of them indicate better performance of the methods. We calculate PSNR and SSIM on the Y channel for image super-resolution, and RGB channel for image denoising, deblurring and dehazing.

Training details: For image super-resolution, we use 64×64 low-resolution patches for training. The total training iterations and mini-batch size are 500K and 32, respectively. We adopt Adam as optimizer [4] with $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The learning rate is initialized to 2×10^{-4} and reduce by half at 250K, 400K, 450K, and 450K followed [2, 5]. For $\times 3$ and $\times 4$ classic SR, we initialize the model with weights of $\times 2$ SR, and halve the learning rate and total training iterations.

For synthetic image denoising, following [2], the patch size and batch size are set to 128×128 and 8, respectively. The training process contains 3200K iterations, with the Adam optimizer configured to $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The learning rate is initially set to 2×10^{-4} and is halved at 1600K, 2100K, 2500K, 2800K, and 3000K iterations. To reduce computational costs, models for $\sigma = 25$ and $\sigma = 50$ are fine-tuned from the weights of $\sigma = 15$, using halved training iterations.

For deblurring, dehazing and real-world denoising, we use the AdamW ($\beta_1 = 0.9$, $\beta_2 = 0.999$, and weight decay is $1e^{-4}$) as optimizer. Our GPUs don't have enough memory to train with large patch sizes. So, we adjust our training settings to make MaIR train with the same number of pixels as when using large patches. Notably, for these tasks, we adopt the U-Net version of MaIR following [2], which empirically outperforms plain structures on above tasks [1, 7]. All Experiments are conducted through Pytorch on NVIDIA GeForce RTX 3090 GPUs.

*Corresponding author

1.2. Qualitative Results

In this section, we present the qualitative results of MaIR as illustrated in Figs. 1 to 4. Taking Fig. 2 as examples, MaIR can effectively preserve fine textures in the restored images, yielding results that close to the ground truth. Similarly, in Fig. 3, MaIR demonstrates its ability to retain intricate details while minimizing distortion, further highlighting its robustness and fidelity in image restoration tasks.

References

[1] 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. 1

[2] 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. 1

[3] Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of psnr in image/video quality assessment. *Electronics letters*, 44(13):800–801, 2008. 1

[4] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015. 1

[5] 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. 1

[6] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to

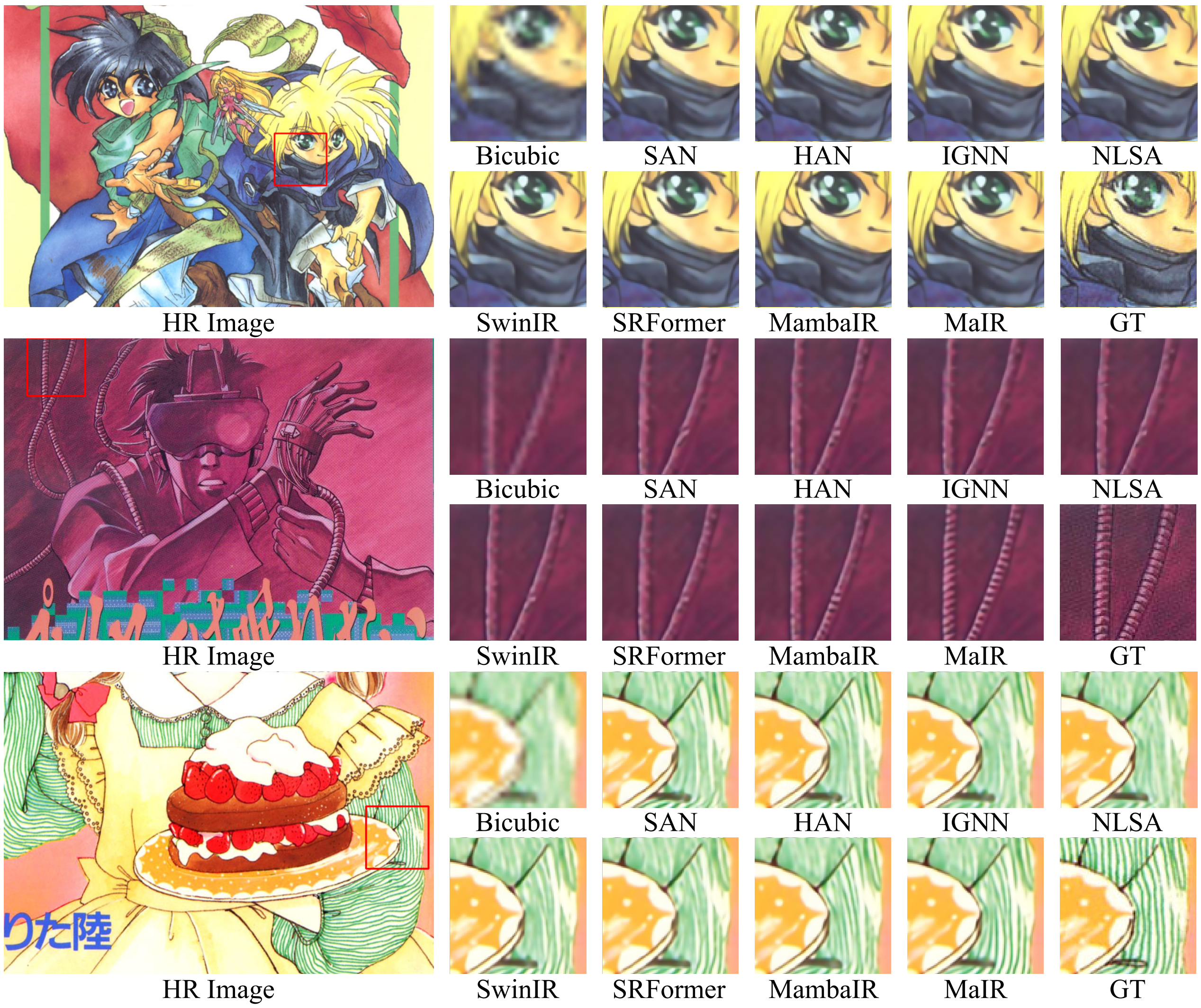

Figure 1. Visual comparison of $\times 4$ image super-resolution results on the Manga109 dataset. MaIR demonstrates superior visual quality, particularly in preserving fine details and textures.

structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 1

[7] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient Transformer for High-Resolution Image Restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5718–5729, New Orleans, LA, 2022. 1

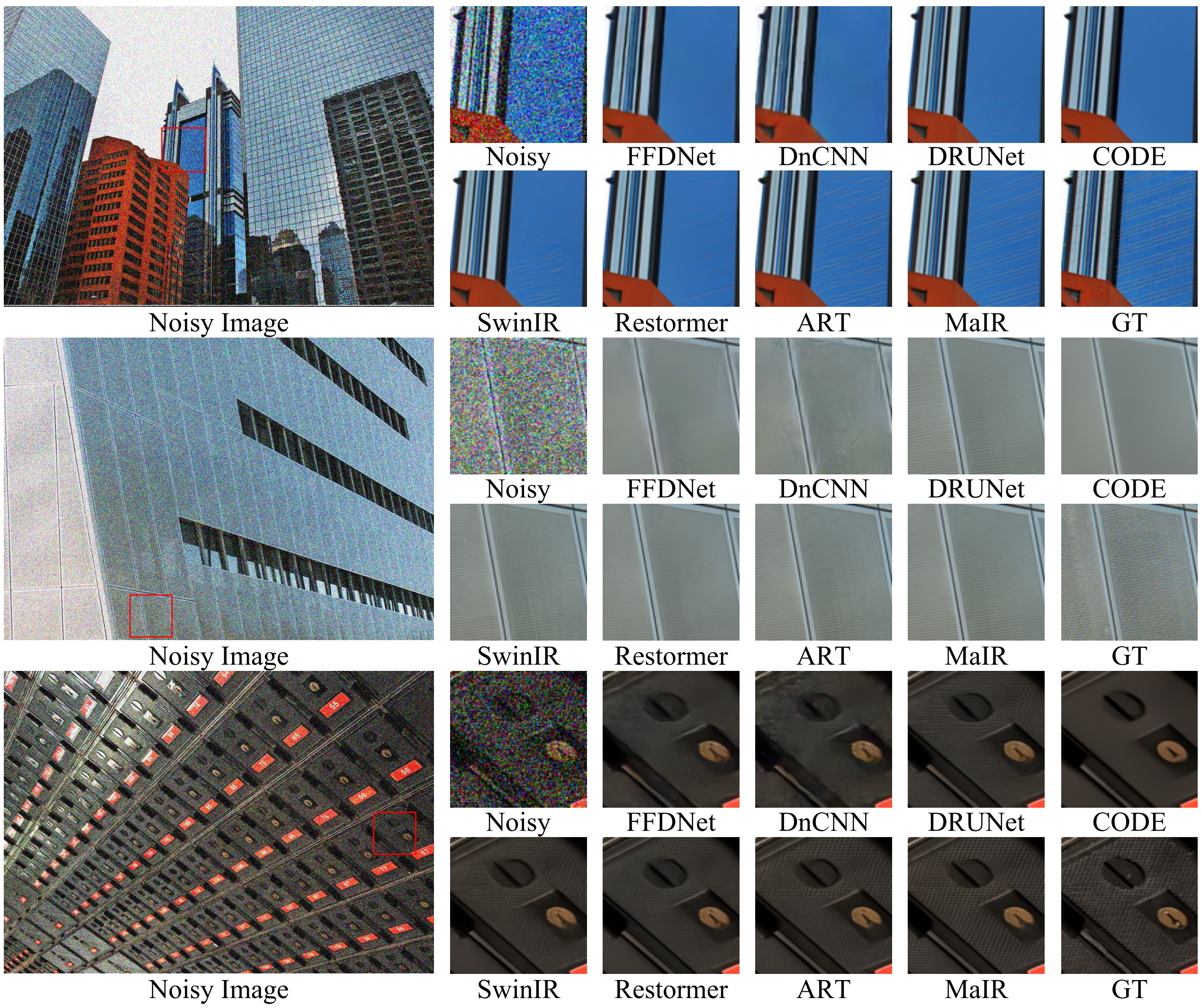

Figure 2. Visual comparison of image denoising results on the Urban100 dataset. MaIR effectively removes noise in the images and produces detailed textures that closely match the ground truth.

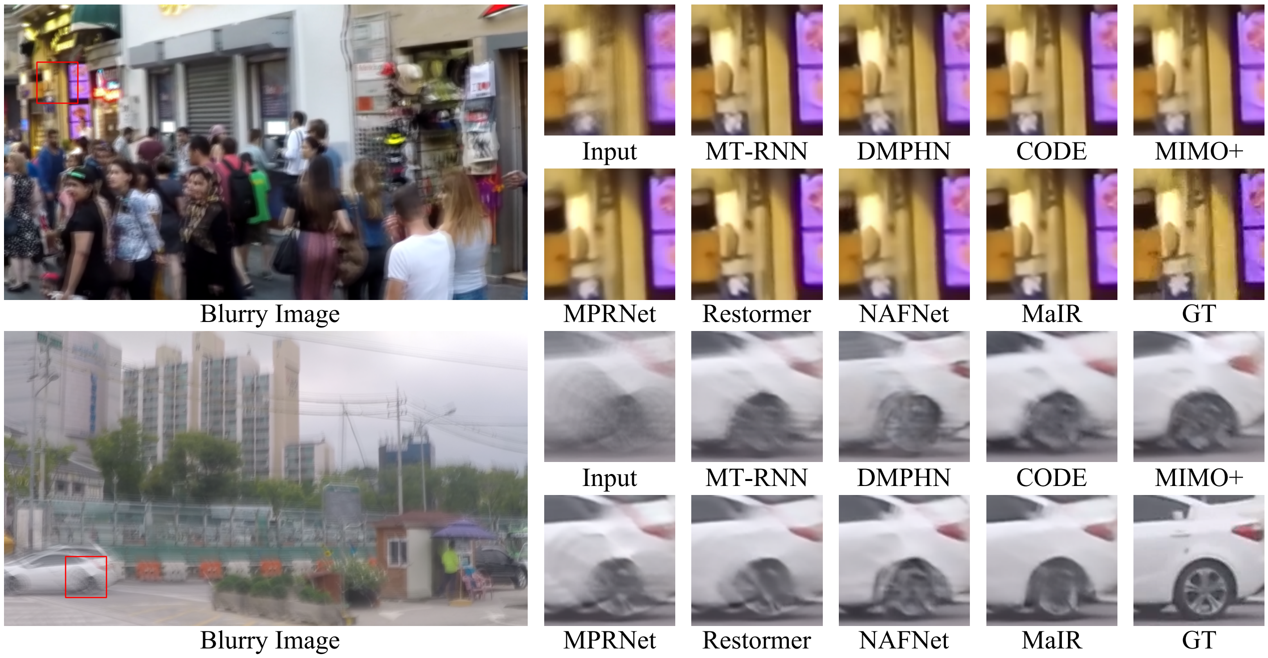

Figure 3. Visual comparison of motion deblurring results on the GoPro dataset. MaIR demonstrates superior performance in effectively removing motion blur while preserving precise fine details and textures, closely matching the ground truth and avoiding distorations.

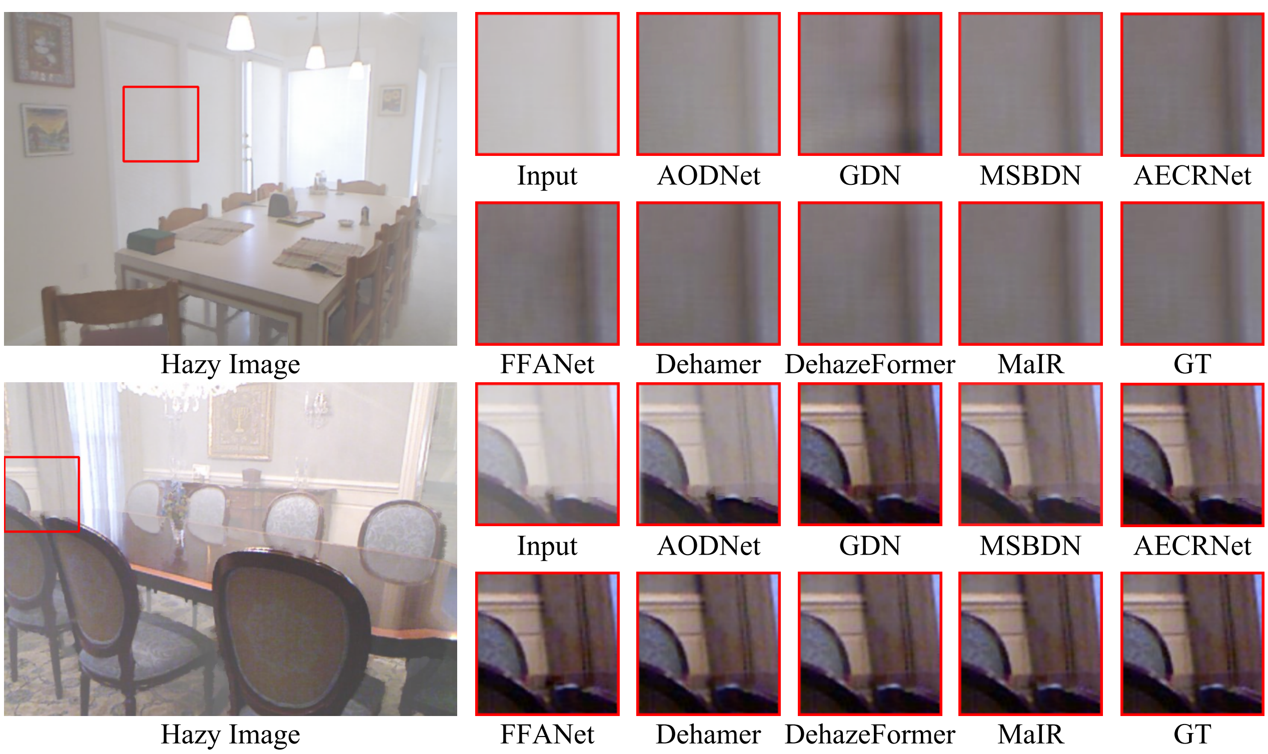

Figure 4. Visual comparison of image dehazing results on the SOTS dataset. MaIR can effectively remove haze and restore contents with colors that closely match the ground truth.