Wavelet-based Mamba with Fourier Adjustment for Low-light Image Enhancement (Supplementary Materials)

Junhao Tan†[1], Songwen Pei†∗[1], Wei Qin[1], Bo Fu[2], Ximing Li[3], and Libo Huang[4]

†Contribute equally ∗Corresponding author. Email address: swpei@usst.edu.cn

[1] School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, 200093, China
223330941@st.usst.edu.cn, swpei@usst.edu.cn, 201440056@st.usst.edu.cn
[2] School of computer science and artificial intelligence, Liaoning Normal University, Liaoning, 116081, China
[3] College of computer science and technology, Jilin university, Jilin, 134000, China
[4] School of Computer, National University of Dense Technology, Changsha, 410073, China

1 Additional Ablation Studies

1.1 Width and Depth Ablation.

The width and depth of the model refer to embedding dimension and the number of iterations for each stage module, respectively. $[D_1]$, $[D_2]$, $[D_3]$ respectively indicates number of iterations for the WMB. The depth of [2, 3, 4] used for WalMaFa achieves the best performance as well as fewer parameters.

Why larger models seem to perform worst? LOL-v1 dataset only consists of 485 train images and 15 test images, which inevitably leads to overfitting. Besides, we speculate that the deeper model will greatly overfit the global brightness due to D_1, D_2, D_3 indicating the number of iteration for WMB in the encoder-decoder, which will undermine the global and local balance.

1.2 Why Encoder-Latent-Decoder?

In this work, Encoder mainly aims at the coarse-grained global multi-scale brightness extraction (thanks to the low-frequency component of the WMB). Then, Latent fine-tines the fine-grained local details (thanks to the Phase component of the FFAB). However, we found that this coarse-to-fine pipeline exists a local overexposure problem (*i.e.*, color distortion) caused by local texture smoothing, as shown in Figure 1. So the extra coarse-grained Decoder is adopted to further balance the global brightness.

Code is available at: https://github.com/mcpaulgeorge/WalMaFa

Table 1: Width and depth ablation on LOL-v1 dataset.

W	D_1	D_2	D_3	Params (M)	PSNR/SSIM
16	1	1	2	8.92	22.15/0.825
16	2	3	4	11.09	**23.27/0.851**
16	4	4	4	12.49	22.60/0.831
16	4	6	8	20.16	22.99/0.850
32	2	3	4	41.86	22.12/0.842


Input Coarse-to-Fine **Ours** GT

Fig. 1: The visual comparisons with coarse-to-fine pipeline.

Table 2: Structure ablation on LOL datasets.

Model	LOLv1	LOLv2-real	LOLv2-syn	Flops(G)
Unet	21.18/0.833	20.80/0.821	23.18/0.898	11.94
Unet-skip-connection	21.92/0.825	21.85/0.812	23.76/0.925	4.24
Channel-wise Self-Attention	21.71/0.832	22.02/0.851	24.61/0.927	6.52
Simplified Channel Attention	22.16/0.843	22.32/0.863	25.02/0.935	5.39
Ours	**23.27/0.851**	**22.49/0.869**	**25.56/0.945**	14.41

1.3 Supplementary Structure Abaltion.

As shown in Table 2, we have experimented the Unet (Encoder with WMB and Decoder with FFAB) to verify the efficiency of Encoder-Latent-Decoder. We replace SSM with Unet-skip-connection between Encoder and Decoder to verify the efficiency of SSM. We also replace Channel-wise Mamba with Channel-wise Self-Attention (Restormer [2]) and Simplified Channel Attention (NAFNet [1]) to verify the efficiency of Channel-wise Mamba.

References

1. Chu, X., Chen, L., Yu, W.: Nafssr: Stereo image super-resolution using nafnet. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 1239–1248 (June 2022)

2. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.: Restormer: Efficient transformer for high-resolution image restoration. In: CVPR. pp. 5718–5729 (2022). https://doi.org/10.1109/CVPR52688.2022.00564