We sincerely thank the reviewers and ACs for their valuable time and efforts. Here is our detailed response to the comments.

Response for Reviewer qgVT, Zk96, and tWub@Some minor problems and typos: Thanks for your feedback. We have proofread the paper to remove typos and enhance readability, particularly in Sections 3 and 4, for the camera-ready version. Additionally, we have carefully reviewed the references and added missing citations.

Response for Reviewer ACFL@Comparison with existing Mamba-based restoration model. We provide more comparison with Mamba-based methods in Table 1, and it will be incorporated into the camera-ready version. These results confirmed the excellent performance of our Wave-Mamba. **@Add previous ACMMM papers on Low-Light Image Enhancement (LLIE).** As suggested, we added the mentioned method [1, 3] and follow the discussion: *Wang et al. [1] presented a dual-stage low-light image enhancement network, FourLLIE, which enhances brightness by estimating amplitude transformation mappings in the frequency domain. Feng et al. [3] proposed a learnability enhancement strategy based on noise modeling, which improves the denoising performance of raw images in low-light conditions.*

Table 1: Comparison of quantitative results on LOLv1 dataset.

Method	LOLv1	
	PSNR	SSIM
MambaIR	22.31	0.826
RetinexMamba	24.03	0.827
LLEMamba	23.71	0.901
Wave-Mamba	26.54	0.883

Table 2: The runtime (s) and GPU memory (G) costs of different methods consumed.

Methods	1920×1080(1080P)		LOLv1
	Mem.(G)↓	Time(s)↓	PSNR
URetinex-Net	5.416	0.432	19.842
SNRNet	8.125	0.588	24.610
UHDFour	8.263	0.501	23.093
WeatherDiff	8.344	548.421	17.913
DiffLL	3.873	0.943	26.336
Wave-Mamba	4.256	0.402	26.540

Response for Reviewer qgVT@Computational load and runtime comparison. We compare runtime and GPU memory cost in Table 2. Specifically, we evaluated different LLIE methods on 1920×1080 (1080P) images. Wave-Mamba achieved the best performance and runtime efficiency.

Response for Reviewer Zk96@Discussion for wavelet transform in the image and feature domain. The observation in the image domain provided foundational insights that informed the overall design of our Wave-Mamba. Moreover, this observation is also consistent in the feature domain and has richer high-frequency and low-frequency information as well as multi-scale information. To illustrate this point, we use histograms to separately analyze the input and output in both domains, as shown in Figure 1. The figure shows that although the histogram distribution of the feature domain output differs from the image domain output due to residual connections, the overall trends in histogram changes are similar in both domains. Thus, this indirectly confirms the validity of our method. **@Ablation study on the scales of wavelet transform (WT).** As suggested, we conducted an ablation study on the scales of the wavelet transform. As shown in Table 3, while increasing the scale of the wavelet transform can further separate information, it also significantly raises the computational cost without markedly improving the enhancement effect. The 8× wavelet transform achieves the best balance between information extraction and computational efficiency, which is why we adopted it in the Wave-Mamba framework. **@Clarify designs for Frequency Matching Transformation.** As described in the paper, we first use high-frequency and low-frequency features $F_H, F_L \in \mathbb{R}^{HW\times C}$ to calculate the similarity matrix $\mathbf{M} \in \mathbb{R}^{C\times C}$. Then, using the Top-1 operation for filtering, select the score $D \in \mathbb{R}^{C\times 1}$ that is most similar to the F_H. Then, based on the score D, the C channel features of F_L are indexed through Indices operations to obtain the features $y \in \mathbb{R}^{HW\times 1}$ that are most similar to F_H, thus forming the selection feature $Y_{selected} \in \mathbb{R}^{HW\times C}$. In addition, we updated Figure 5 and the description in the camera-ready version to accurately reflect the details of the selection and matching process.

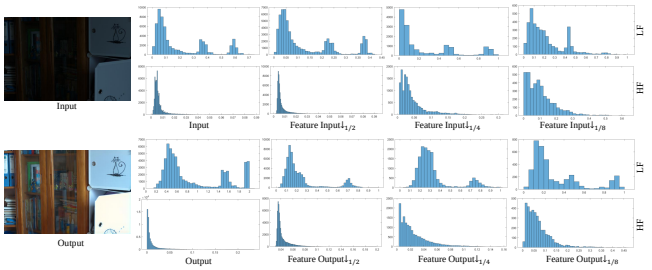

Figure 1: Visualization histogram of the low and high-frequency in the Wave-Mamba.(Zoom in for the best view)

Table 3: Ablation studies on different numbers of LFSSBlocks and HFEBlocks.

Experiment	LFSSBlocks	HFEBlocks	PSNR	SSIM	Params
1	[1,1,4]	[1,1,1]	26.48	0.904	1.2M
2	[1,2,4]	[1,1,1]	27.35	0.913	1.3M
3	[1,1,4]	[1,1,2]	26.67	0.909	2.2M
4	[1,2,4]	[1,1,2]	27.41	0.922	2.3M
5	[1,1,2,4]	[1,1,1,1]	27.38	0.913	2.3M

Response for Reviewer tWub @Reasons for not selecting the UHD-OL8K dataset. Following the previous work [2], the UHD-LOL dataset uses the same data synthesis method for both UHD-LOL8K and UHD-LOL4K, resulting in identical data distributions. In addition, current UHD LLIE methods benchmark the UHD-LOL8K by partitioning each 8K image into four 4K patches. Therefore, we believe that the UHD-LOL4K dataset is more representative. **@Add references and introductions to estimated metrics.** As suggested, we added references for the metrics PSNR, SSIM, and LPIPS. Additionally, we provided a brief introduction: *PSNR measures the quality of reconstructed images by comparing pixel intensity differences, SSIM evaluates image quality based on luminance, contrast, and structure, and LPIPS assesses perceptual similarity using deep learning models to capture human visual similarity.* **@Add a set of ablation experiments with parameters set to LFSSBlocks: [1,1,4], HFEBlocks: [1,1,2].** Thank you for your valuable feedback. We performed more ablation studies on different numbers of LFSSBlocks and HFEBlocks, as shown in Table 3. The results of these experiments will be included in the camera-ready version to provide a more comprehensive analysis of our method.

REFERENCES

[1] Hansen Feng, Lizhi Wang, Yuzhi Wang, and Hua Huang. 2022. Learnability enhancement for low-light raw denoising: Where paired real data meets noise modeling. In *Proceedings of the 30th ACM International Conference on Multimedia*. 1436–1444.

[2] Haoxiang Jie, Xinyi Zuo, Jian Gao, Wei Liu, Jun Hu, and Shuai Cheng. 2023. Llformer: An efficient and real-time lidar lane detection method based on transformer. In *Proceedings of the 2023 5th international conference on pattern recognition and intelligent systems*. 18–23.

[3] Chenxi Wang, Hongjun Wu, and Zhi Jin. 2023. Fourllie: Boosting low-light image enhancement by fourier frequency information. In *Proceedings of the 31st ACM International Conference on Multimedia*. 7459–7469.