ExpoMamba: Exploiting Frequency SSM Blocks for Efficient and Effective Image Enhancement

Eashan Adhikarla[1,2]; Kai Zhang[1]; John Nicholson[2]; Brian D. Davison[1]
[1]Lehigh University, Bethlehem, PA, USA, [2]Lenovo Research, USA

motorola
a Lenovo company

Lenovo Research

ABSTRACT

Low-light image enhancement remains a challenging task in computer vision, with existing state-of-the-art models often limited by hardware constraints and computational inefficiencies, particularly in handling high-resolution images. Recent foundation models, such as transformers and diffusion models, despite their efficacy in various domains, are limited in use on edge devices due to their computational complexity and slow inference times. We introduce ExpoMamba, a novel architecture that integrates components of the frequency state space within a modified U-Net, offering a blend of efficiency and effectiveness. This model is specifically optimized to address mixed exposure challenges—a common issue in low-light image enhancement—while ensuring computational efficiency. Our experiments demonstrate that ExpoMamba enhances low-light images up to **2-3x faster** than traditional models with an inference time of **36.6 ms** and achieves a PSNR **improvement of ~15-20%** over competing models, making it highly suitable for real-time image processing applications. Model code is open sourced: **github.com/eashanadhikarla/ExpoMamba**

INTRODUCTION

Enhancing low-light images is critical for applications ranging from mobile photography to sophisticated surveillance systems. transformers and diffusion models, have made significant strides in image enhancement but face challenges in computational efficiency, particularly on edge devices with limited processing power.

We introduce **ExpoMamba** (**Fig-2**), a novel architecture that integrates frequency state-space components within a modified U-Net. This model is optimized for mixed exposure challenges, ensuring computational efficiency while enhancing image quality.

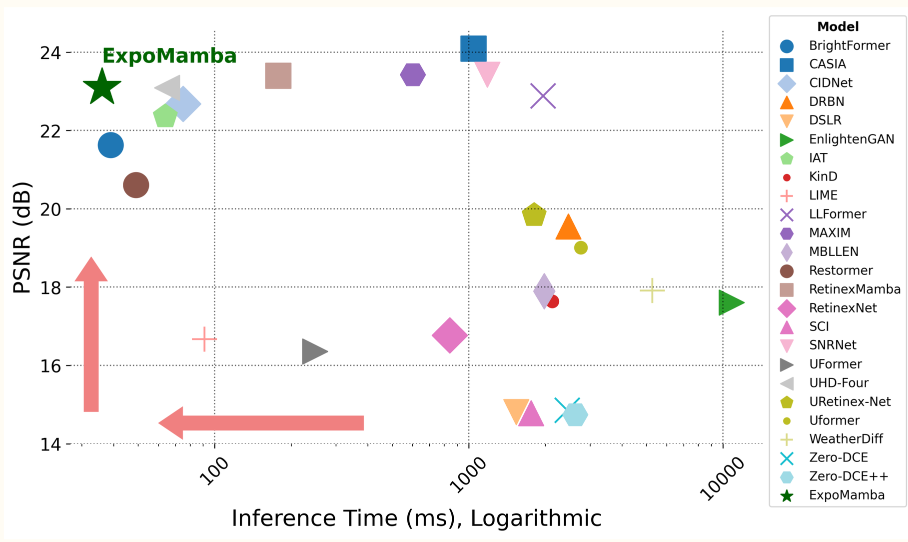


Figure 1.a. PSNR vs Inference time (ms); 400x600

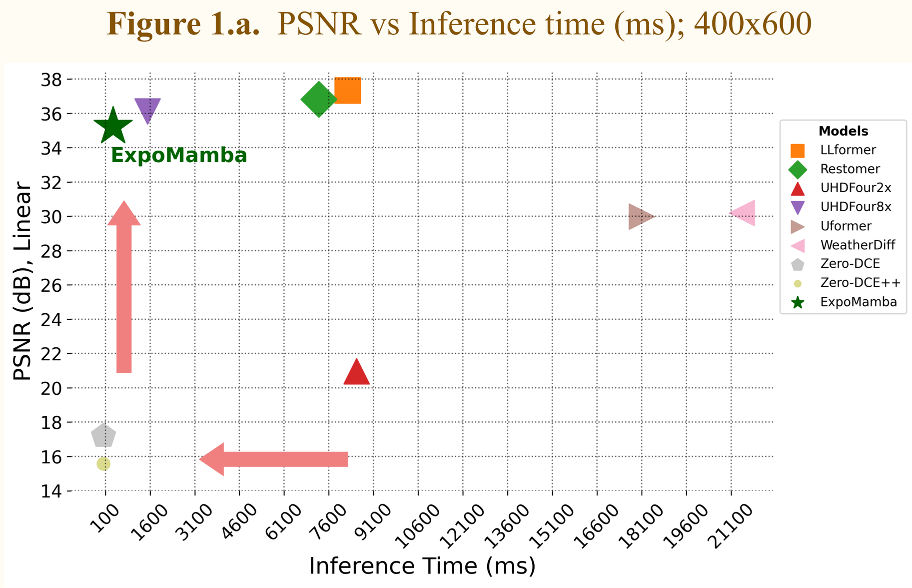


Figure 1.b. PSNR vs Inference time (ms); 4K

PURPOSE

With the rising need for better images, advanced small camera sensors in edge devices have made it more common for customers to capture high-quality images and use them in real-time applications like mobile, laptop, and tablet cameras. However, they all struggle with non-ideal and low lighting conditions. Our goal is to develop an approach that has high image quality (e.g., like CIDNet[1]) for enhancement but also at high speed (e.g., such as IAT[2] and Zero-DCE++[3]).

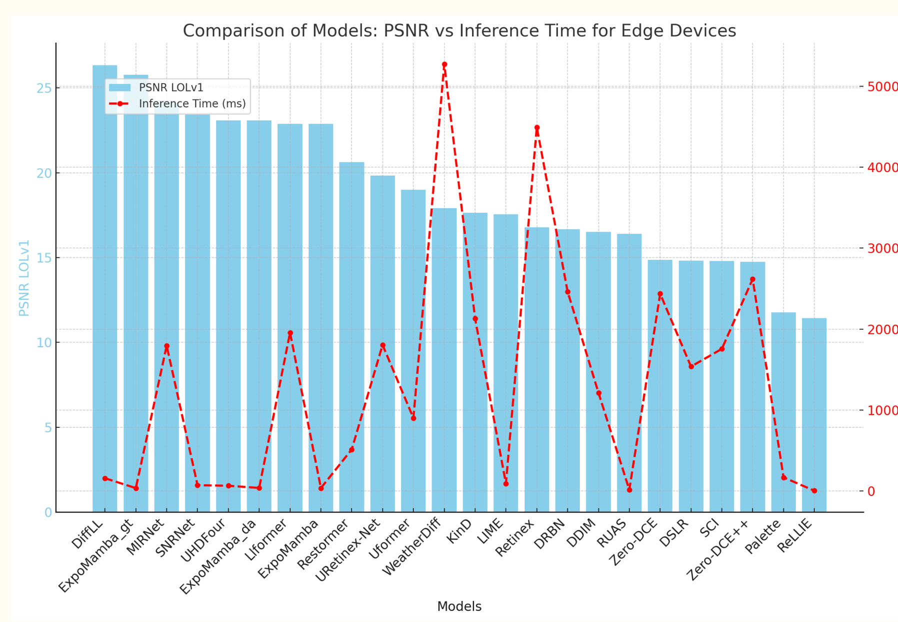


Figure 4. Results on LOLv1 showing the tradeoff b/w PSNR and inference speed (ms)

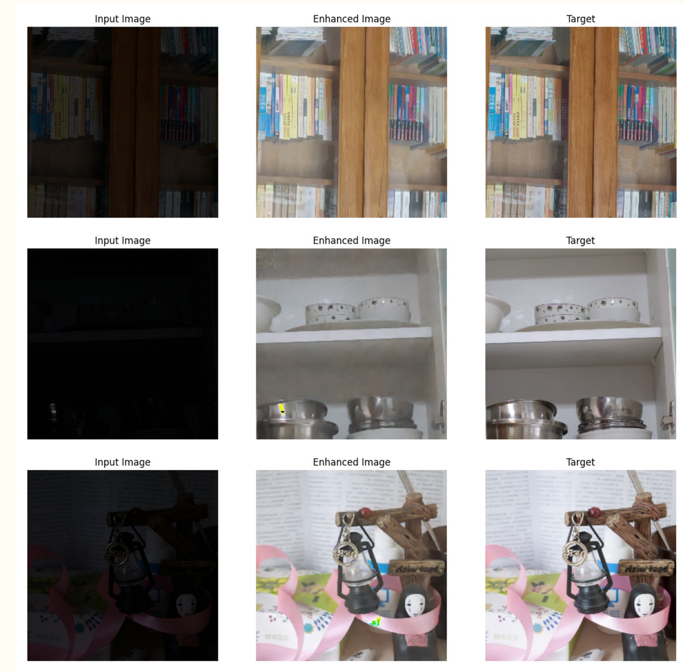
Contributions.

- We introduce the use of Mamba for efficient **low-light image enhancement (LLIE)**, focusing on **mixed exposure** challenges with *underlit* and *overlit* areas in the same frame.
- We propose a novel **Frequency State Space Block (FSSB)** combining two distinct 2D-Mamba blocks to capture and enhance subtle textural details.
- We describe a **dynamic batch training** scheme to improve robustness in multi-resolution inference.
- We implement dynamic processing of the amplitude component to highlight distortion and the phase component for smoothing and noise reduction.

RESULTS

(**Fig-1.a**) **[top: 400x600]** and (**Fig-1.b**) **[bottom: 3840x2160]** Scatter plot of model inference time vs. PSNR. We found that ExpoMamba is faster than majority models. Making it perfectly suitable for edge devices. Important observation is that as we grow in input sequence from **400x600** to **4K**, the **quadratic completexity of transformers** make the time and memory drastically grow. Whereas, for **Expomamba** we notice a **linear growth in time with fixed memory**.

Figure 5. Shown images are input, enhanced output and the ground-truth image.



DISCUSSION

ExpoMamba achieves an inference time of 36 ms, faster than most baselines (**Fig-1.a**) and the fastest among comparable models. Models like DiffLL, CIDNet, and LLformer have comparable results but much longer inference times.

Despite being a **41 million** parameter model, ExpoMamba demonstrates remarkable storage efficiency, consuming **~1/4th memory** (2923 Mb) compared to CIDNet, which, despite its smaller size of **1.9 million** parameters, consumes **8249 Mb**.

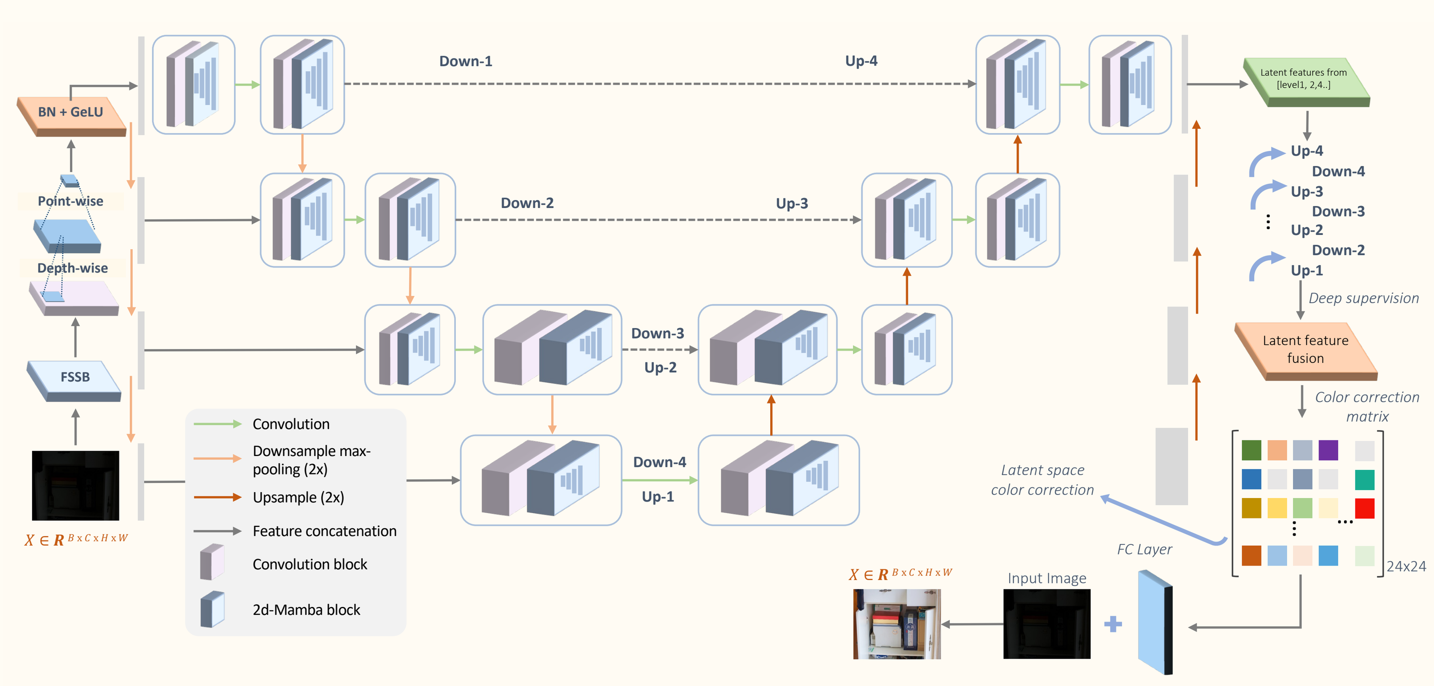


Figure 2. ExpoMamba model pipeline showing the overall work flow.

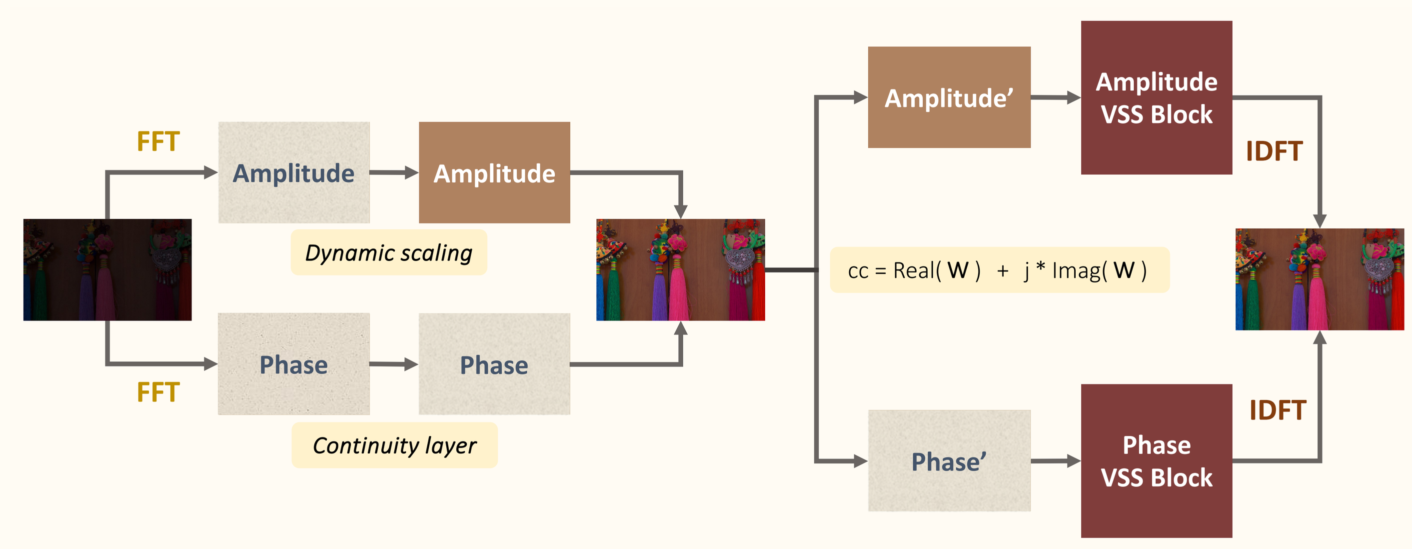


Figure 6. Frequency State-Space Block (FSSB) Processing.

METHODS AND MATERIALS

- ExpoMamba integrates Frequency State-Space Blocks (**FSSB**) (**Fig-6**) within a modified U-Net architecture to enhance image quality efficiently. The model leverages both spatial and frequency domains to address distortions in low-light conditions. FSSB processes **amplitude** and **phase** components separately, ensuring detailed enhancement without introducing artifacts.
- To address **HDR-related issues**, ExpoMamba includes an overexposed regularizer that detects and penalizes overexposed regions, reducing artifacts and preserving details in high dynamic range scenarios. Additionally, The HDR layer is consistently applied as the final layer within each FSSB block and culminates as the ulti- mate layer in the ExpoMamba model, providing a coherent enhancement across all processed images.
- We evaluate ExpoMamba on multiple datasets (names; train/test split; resolution sizes):

LOLv1	LOLv2	LOL4K	SICEv2
485/15	689/100	8099/5999	4,800
3x400x600	3x400x600	3x3840x2160	3x900x600

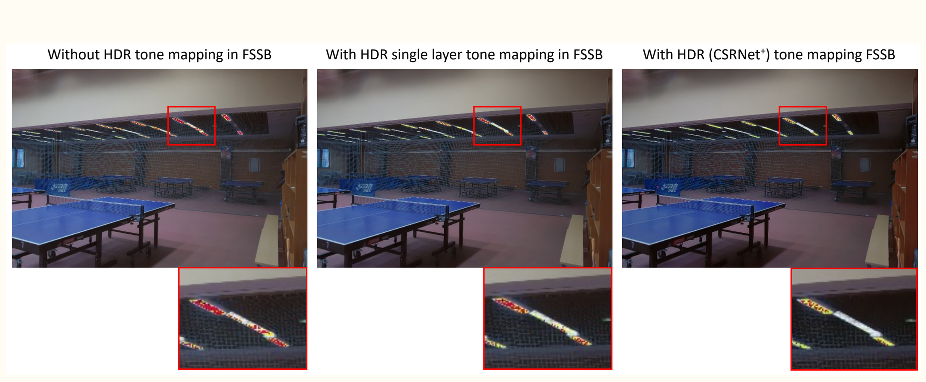


Figure 3. HDR tone mapping layer inside FSS block.

EXPERIMENTS

Table 1. Results for our Exposure Mamba approach over SICE-v2 (Cai et al., 2018) datasets.

	Underexposure		Overexposure	
	PSNR	SSIM	PSNR	SSIM
SID-L	19.43	0.6644	17.00	0.6495
RUAS	16.63	0.5589	4.54	0.3196
MSEC	19.62	0.6512	17.59	0.6560
LCDPNet	17.45	0.5622	17.04	0.6463
DRBN+ERL	18.09	0.6735	17.93	0.6866
ELCNet+ERL	22.14	0.6908	19.47	0.6982
FECNet	22.01	0.6737	19.91	0.6961
IAT	21.41	0.6601	22.29	0.6813
ExpoMamba	**22.59**	**0.7161**	**20.62**	**0.7392**

Loss functions.

$$\mathbf{L} = \mathbf{L}\,(l1) + \mathbf{L}\,(vgg) + \mathbf{L}\,(ssim) + \mathbf{L}\,(lpips) + \lambda \times \mathbf{L}\,(overexposed)$$

The combined loss function as shown above is designed to enhance image quality by addressing different aspects of image reconstruction.

Table 2. We describe two variants of our model, s' and l' represent small and large model configurations.

Model type	Configuration				
	Base channel	Patch size/depth	Param	Inference	Memory
ExpoMamba (small)	48	4/1	41 M	36 ms	2923 Mb
ExpoMamba (Large)	96	6/4	166 M	95.6 ms	5690 Mb

CONCLUSIONS

We introduced *ExpoMamba*, **2-3x** faster & efficient low-light image enhancement. Integrating FSSB component within a U-Net variant, *ExpoMamba* addresses computational inefficiencies and high-resolution challenges by leveraging spatial and frequency domain processing. Our dynamic patch training strategy improves robustness to real-world hardware constraints, making it ideal for real-time edge device applications.

REFERENCES

[1] Feng, Y., Zhang, C., Wang, P., *et al.*, You only need one color space: An efficient network for low-light image enhancement, 2024.
[2] Cui, Z., Li, K., Gu, L., *et al.*, You only need 90k parameters to adapt light: a light weight transformer for image enhancement and exposure correction, BMVC 2022.
[3] Li, C., Guo, C. G. *et al.*, Learning to enhance low-light image via zero-reference deep curve estimation, IEEE TPAMI 2021.

CONTACT

Eashan Adhikarla
113 Research Drive
Lehigh University, Bethlehem, PA
Email: eaa418@lehigh.edu
Phone: +1(650) 695-9272
Website: eashanadhikarla.github.io