

Introduction

Bottleneck in Low-Light Vision

The Challenge

- Enhancing low-light images is critical for edge applications like mobile cameras and real-time surveillance.
- Transformer-based foundation models achieve high quality but are bottlenecked by $O(n^2)$ and designs to reduce bottleneck compromise with performance.
- This complexity makes them unsuitable for real-time, high-resolution processing on power-constrained edge devices.

The ExpoMamba Solution

- ExpoMamba establishes a new state-of-the-art by balancing perceptual realism with runtime efficiency.
- The architecture leverages the linear-time efficiency $O(n)$ of Mamba models while operating directly in the frequency domain.

→ Higher resolution, more input tokens, quadratically more complex.

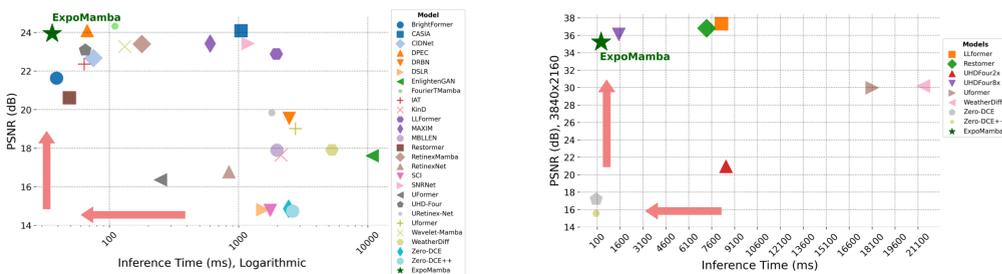


Fig. 1 [Left: 400x600; Right: 3840x2160] Logarithmic Scatter Plot of Inference Time vs. PSNR. Baselines that used ground-truth mean information to produce metrics were reproduced without such information for fairness.

Frequency State Space Block (FSSB)

ExpoMamba performs a Fourier decomposition of the input image, isolating amplitude (illumination) and phase (structure) for independent processing. Dual 2D-Mamba (Visual-SSM) blocks process these frequency streams in parallel.

Amplitude Stream: Targets illumination consistency and noise reduction.

Phase Stream: Preserves fine-grained structural details, edges, and object boundaries.

$$h[t+1] = A[t] \cdot h[t] + B[t] \cdot x[t] \quad (1) \quad y[t+1] = C[t] \cdot h[t] \quad (2)$$

ExpoMamba Framework

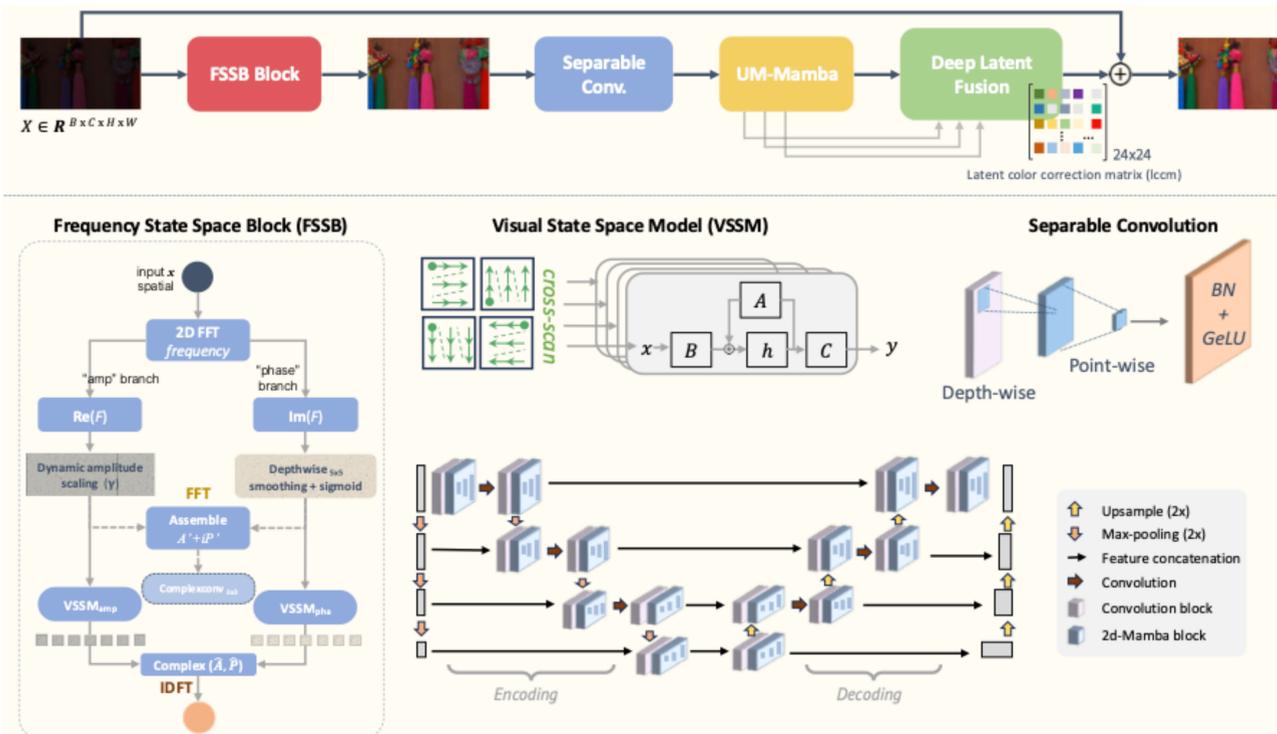


Fig. 2 The diagram illustrates the information flow through the ExpoMamba model.

Latent Space Color Correction (LCCM)

- Color correction is applied in the latent feature space rather than standard 8-to-10 bit image space.
- This avoids rounding and clipping, ensuring small corrections are learned without staircasing or banding artifacts.

$$\hat{I}(x, y) = \cos(\Delta\theta) + i \sin(\Delta\theta) I(x, y) \quad (4)$$

$$\hat{I}(x, y) = \iint A(u, v) e^{i\phi(u, v) + \Delta\theta} e^{i2\pi(ux+vy)} dudv \quad (3)$$

Phase Manipulation for Structural Fidelity

- Because human vision is highly sensitive to phase variations, the model adaptively modifies phase to refine scene geometry.
- Applying a uniform learned phase shift $\Delta\theta$ across the frequency spectrum mathematically transforms the spatial structure without destroying illumination cues

Latent HDR Feature Recovery

- To combat *mixed-exposure* scenarios, an HDR gate is implemented directly in the latent feature space.
- Overexposed activations are detected via a soft mask $M = \sigma(k(Y - \tau))$ and corrected using a learned tone curve $T(x)$.
- The corrected features are gently blended to attenuate halos and blooming while preserving detail:

$$\tilde{x} = (1 - M) \cdot x + M \cdot T(x) \quad (5)$$



ExpoMamba is 2-3x faster with 6.8% PSNR improvement. Consumes roughly 4x less memory than models like CIDNet (82.49 Mb)

Methods	LOLv1			LOLv2 (Real Captured)			Inference time (ms)
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	
Retinex [†]	16.774	0.462	0.417	17.715	0.652	0.436	4493
SCI [†]	14.784	0.525	0.366	17.304	0.54	0.345	1755
URetinex-Net [†]	19.842	0.824	0.237	21.093	0.858	0.208	1804
SNRNet [†]	23.432	0.843	0.234	21.480	0.849	0.237	72.16
Uformer [*]	19.001	0.741	0.354	18.442	0.759	0.347	901.2
Restormer [*]	20.614	0.797	0.288	24.910	0.851	0.264	513.1
Palette [*]	11.771	0.561	0.498	14.703	0.692	0.333	168.5
UHDFour [†]	23.093	0.821	0.259	21.785	0.854	0.292	64.92
WeatherDiff [*]	17.913	0.811	0.272	20.009	0.829	0.253	5271
GDP [*]	15.896	0.542	0.421	14.290	0.493	0.435	1052
DiffLL [*]	26.336	0.845	0.217	28.857	0.876	0.207	157.9
CIDNet [†]	23.090	0.851	0.085	23.220	0.863	0.103	149
LLformer [*]	22.890	0.816	0.202	23.128	0.855	0.153	1956
RetinexMamba [†]	24.02	0.827	0.250	22.45	0.844	0.236	91
FourierTMamba [†]	24.33	0.845	0.289	22.41	0.860	0.242	102
Wavelet-Mamba [†]	23.27	0.851	0.248	22.49	0.869	0.218	83
ExpoMamba[†]_{gt}	25.770	0.860	0.212	28.040	0.885	0.232	36.00

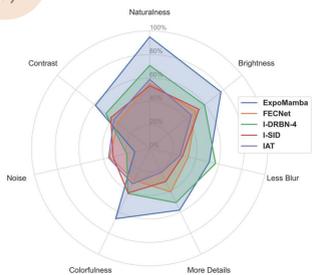


Fig. 3 Comparing ExpoMamba outputs with SOTA light-weight models on perceptual realism.

Best Efficiency

Tab. 1 Comparing three popular metrics such that every column showcases for LOLv1 and LOLv2 datasets. Last column shows the inference time over NVIDIA A6000.

Qualitative Comparisons

ExpoMamba demonstrates superior structural fidelity and natural color balance under challenging mixed-lighting conditions.

Method	SICE-v2				#params
	Underexposure		Overexposure		
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	
LCDPNet [71]	17.45	0.5622	17.04	0.6463	0.96M
DRBN [86]	17.96	0.6767	17.33	0.6828	0.53M
DRBN-ERL+ENC [31]	22.06	0.7053	19.50	0.7205	0.58M
ELCNet [32]	22.05	0.6893	19.25	0.6872	0.018M
ELCNet+ERL [31]	22.14	0.6908	19.47	0.6982	0.018M
FECNet [28]	22.01	0.6737	19.91	0.6961	0.15M
FECNet+ERL [31]	22.35	0.6671	20.10	0.6891	0.15M
IAT [12]	21.41	0.6601	22.29	0.6813	0.090M
ExpoMamba_s	22.59	0.7161	20.62	0.7392	41M

Tab. 2 ExpoMamba evaluated on multi-exposure dataset for mixed-exposure task.

Methods	ExDark Detection		ACDC Segmentation	
	mAP ↑	time (s) ↓	mIOU ↑	time (s) ↓
YOLOv3 [63]	76.4	0.033	63.3	0.249
MBLLEN [54]	76.3	0.086	63.0	0.332
DeepLPF [57]	76.3	0.138	61.9	0.807
Zero-DCE [24]	76.9	0.042	61.9	0.300
MAET (w ort) [14]	74.0	0.123	-	-
IAT [13]	77.2	0.040	62.1	0.280
ILE-YOLO [81]	78.5	0.047	64.3	0.306
ExpoMamba	79.8	0.076	64.3	0.297

Tab. 3 ExpoMamba evaluated on downstream autonomous vision tasks.

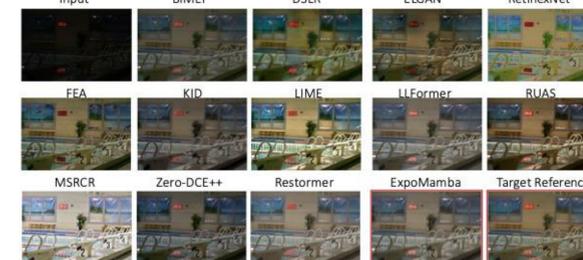


Fig. 4 Qualitative comparison of ExpoMamba and baselines on the LOLv1 dataset.

Conclusion

By uniting frequency-domain processing with linear-time state-space models, ExpoMamba breaks the quadratic bottleneck to deliver state-of-the-art, real-time low-light enhancement for resource-constrained edge devices.

