Unveiling Advanced Frequency Disentanglement Paradigm for Low-Light Image Enhancement Supplementary Material

Kun Zhou[1,2*], Xinyu Lin[1,2⋆], Wenbo Li[3], Xiaogang Xu[3]
Yuanhao Cai[4], Zhonghang Liu[5], Xiaoguang Han[1], and Jiangbo Lu[2]

[1]CUHK-Shenzhen, China [2]SmartMore Corporation, China [3]CUHK
[4]Johns Hopkins University, USA [5]SMU, Singapore
hanxiaoguang@cuhk.edu.cn, jiangbo.lu@gmail.com

1 Global branch in ACCA

Inspired by [6], we employ a two-stream framework for efficient coarse adjustment. In our main paper, we have provided comprehensive details regarding our local branch and its primary component: W-CCA. Here, we elucidate the global branch utilized for capturing long-term information. Initially, it employs two strided convolutions followed by GELU activation to obtain feature $F_g \in R^{\frac{H}{4}\times\frac{W}{4}\times 48}$ from an input image $I \in R^{H\times W\times 3}$. Subsequently, we integrate a "Self-Attnetion" layer to extract global information from the last extracted feature F_g and generate the final predictions $A_g \in R, B_g \in R^{3\times 3}$. Specifically, the "Self-Attnetion" mechanism is delineated as follows:

$$
\begin{aligned}
&F_g = \text{Reshape}(F_g), && F_g \in R^{\frac{H*W}{16}\times 48}\\
&K, V = \text{Linear}(F_g, 48, 2*4*16), && K, V \in R^{4\times\frac{H*W}{16}\times 16}\\
&Q = \text{Parameter}(4, 10, 16), && Q \in R^{4\times 10\times 16}\\
&S = \text{Sotfmax}(Q@K.transpose(-2,-1)), && S \in R^{4\times 10\times\frac{H*W}{16}}\\
&Y = S@V, && Y \in R^{4\times 10\times 16}\\
&A_g, B_g = \text{Split}(\text{Linear}(Y, 10, 1), 1, 9), && A_g \in R, B_g \in R^{3\times 3}
\end{aligned}
\tag{1}
$$

where '@' denotes matrix multiplication. 'Linear$(, m, n)$' represents a linear projection layer with 'm' and 'n' representing the input and output dimensions, respectively. Additionally, the function 'Split' divides a 10-channel feature into two parts: one (B_g) with 9 channels and another (A_g) with 1 channel. Following [6], we introduce a global query parameter Q for sparse and representative illumination recovery.

2 Complexity of W-CCA

In our method, the ACCA module enjoys superior computational efficiency in contrast to existing CNN-based methods with an attention-style aggregation

⋆ Equal contribution

scheme from transformer structures. In this part, we will demonstrate the significant reduction in computational resource consumption achieved by our W-CCA, compared with conventional window-based transformer block (W-MSA [26]).

As described in Sec. 3.2, we employ a strided convolution layer to split an input image feature F into $M_h \times M_w$ image patches[1]. To further reduce the parameter count, we implement a group-wise convolution with a group size of s and a kernel size of $s \times s$. Therefore, the complexity of the first convolution layer can be obtained:

$$O(\text{Conv}) = s^2 \times M_h \times M_w \times C^2/s = s \times M_h \times M_w \times C^2. \tag{2}$$

Then, we utilize three strided convolution layers to produce the three separable 1D kernels:

$$\begin{aligned} O(\text{Eq.10}) =& 2 * s^2 \times M_h \times M_w \times C * s/s \\ &+ s^2 \times M_h \times M_w \times C^2/s, \\ =& 2 * s^2 \times M_h \times M_w \times C, \\ &+ s \times M_h \times M_w \times C^2. \end{aligned} \tag{3}$$

Later on, we compose the Onmi aggregation map using Eq.9:

$$O(\text{Eq.9}) = s^2 \times M_h \times M_w \times C. \tag{4}$$

Finally, we perform element-wise Onmi similarity aggregation using Eq.8:

$$O(\text{Eq.8}) = s^2 \times M_h \times M_w \times C. \tag{5}$$

In total, the complexity of a single W-CCA is obtained as:

$$O(\text{W-CFA}) = 4HWC + 2HWC^2/s. \tag{6}$$

Frequency decomposition. We observe that previous uniform optimization methods may suffer from inaccurate illumination corrections or residual noise artifacts (visual examples in Fig. 1,4 and Supp. Fig. 5). To tackle these issues, we present a low-frequency consistent loss for general frequency disentanglement optimization.

Table 1: Frequency decomposition scheme comparison on LOL-v2.

Method	Restormer(Base)	Conv	FFT	DCT	Ours
PSNR (dB)	19.94	23.92	24.05	24.17	**24.56**
SSIM	0.827	0.871	0.862	0.887	**0.893**

[1] $M_h = H/s, M_w = W/s$, $H \times W$ is the feature resolution.

Low-frequency consistency. Here, we assess the low-frequency consistency results (between model prediction and ground truth) in the table below. It suggests that ACCA outperforms other SOTA models[2]. Regarding LDRM, we not only utilize the coarsely enhanced results generated by ACCA as input but also leverage both low-frequency reconstruction and consistent losses (as described in Eq. 5) to refine its low-frequency accuracy.

Table 2: Low-frequency consistency results of different LLIE models on LOL-v2.

Methods	A	B	C	D	E	F	G (ACCA)
MAE	0.77	0.74	0.69	0.54	0.53	0.50	**0.46**
MSE	0.11	0.13	0.09	0.07	0.10	0.07	**0.06**

3 Results on VV and ExDark.

In our research paper, we conduct extensive experiments to validate the efficacy of our approach. Here, we specifically investigate the advantages of our frequency disentanglement learning on the VV [39] and ExDark [27] datasets. VV, which is a non-reference low-light image benchmark, comprises 24 images. We employ the NIQE metric for quantitative evaluation. ExDark, on the other hand, is a dataset for low-light image detection, and we adopt the average precision (AP) scores following the methodology of Retinexformer for object assessment.

To swiftly gauge the impact of our method, we compare two baseline models (Retinexformer [3] and SNR [51]) with their counterparts enhanced by our disentanglement optimization (referred to as Retinexformer-De and SNR-De). As demonstrated in Tab. 3, Retinexformer-De and SNR-De consistently outperform their baseline counterparts, affirming the effectiveness of our techniques in enhancing high-level detection accuracy and achieving superior results on non-reference metrics.

Table 3: Quantitative improvements over SOTA LLIE methods on VV and ExDark benchmarks.

Datasets	Retinexformer	Retinexformer-**De**	SNR	SNR-**De**
VV(NIQE↓)	13.72	10.48(-3.24)	12.33	10.37(-1.96)
ExDark(AP↑)	75.60	76.40(+0.80)	74.00	76.20(+2.20)

Exploring other image restoration tasks. We've examined its potential in tasks such as image HDR, denoising, deblurring, and deraining. As Tab. 4

[2] A-F represent "MIR-Net", "Restormer", "SNR", "Retinexformer", "Star", and "IAT", with G representing "ACCA".

demonstrates, our frequency disentanglement optimization strategy proves advantageous (PSNR/SSIM) compared to the representative baseline model. We will continue exploring other restoration problems in our future work.

Table 4: Quantitative improvements over Restormer for other image restoration tasks.

Task	HDR	Denoising	Deblurring	Deraining
Dataset	MIT-Adobe-5K	SIDD	RealBlur-J	Rain100H
Restormer	24.12/0.882	40.02/0.960	28.96/0.879	31.46/0.904
Restormer-De	**25.90/0.921**	**40.36/0.964**	**29.62/0.903**	**32.39/0.910**

4 More Visual Results

In our main paper, we demonstrate the effectiveness of our disentanglement learning applied to several representative SOTA models for the LLIE task. In this section, we will provide more visual examples for qualitative comparison in Fig. 1, Fig. 2, Fig. 3 and Fig. 4, respectively for MIR-Net [56], Retinexformer [3], Restormer [55] and SNR [51]. In Fig. 5, we also provide visual examples to illustrate the disentanglement enhancement capability of our method. It is clear that our method provides more accurate intensity adjustment in the low-frequency domain and effective denoising in the high-frequency domain.

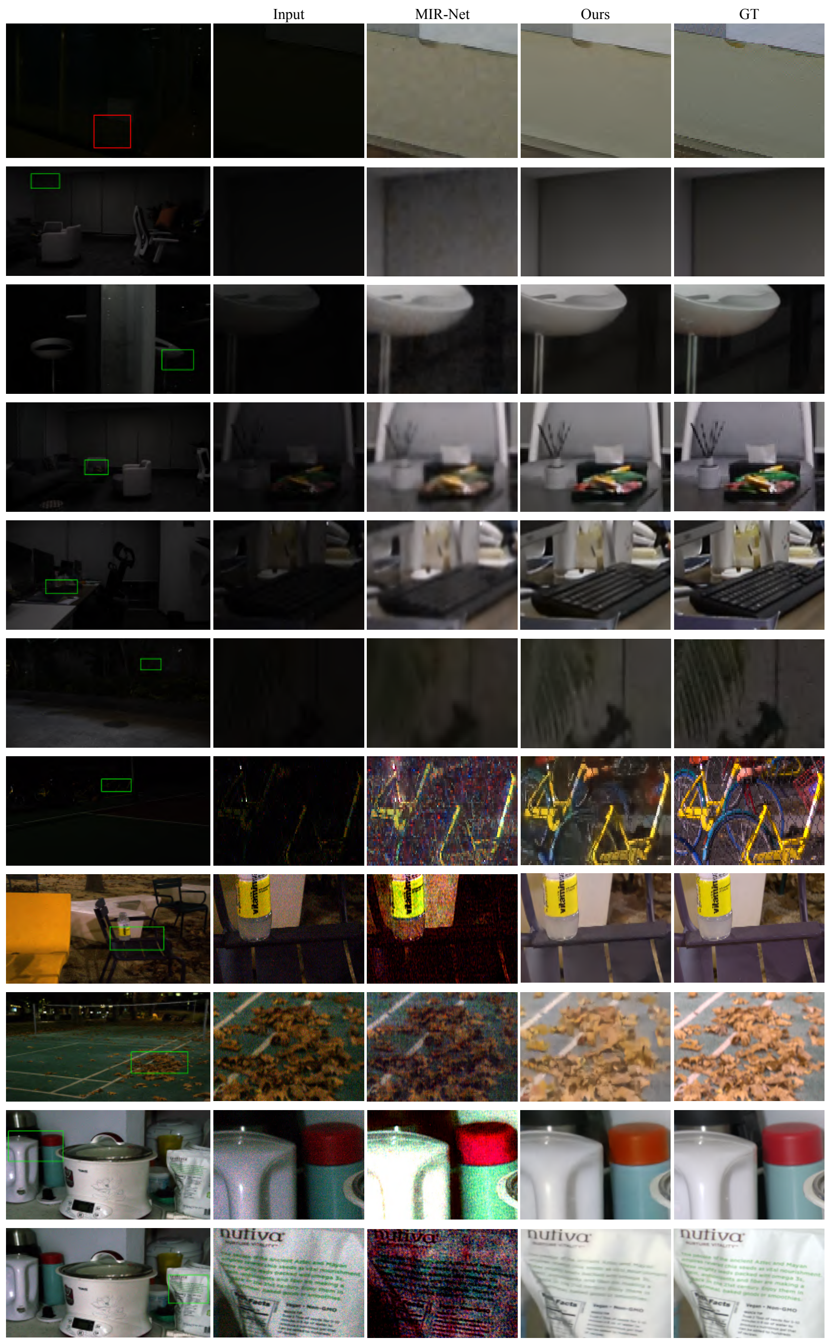

Fig. 1: Qualitative comparison between MIR-Net [56] and the enhanced version by our method. Samples are selected from LOL-v2 [52], SID [4], SDSD [43] (both indoor and outdoor parts) datasets. Our integration provides accurate outcomes for both high-frequency (clearer image detail restoration) and low-frequency (more accurate illumination recovery) areas.

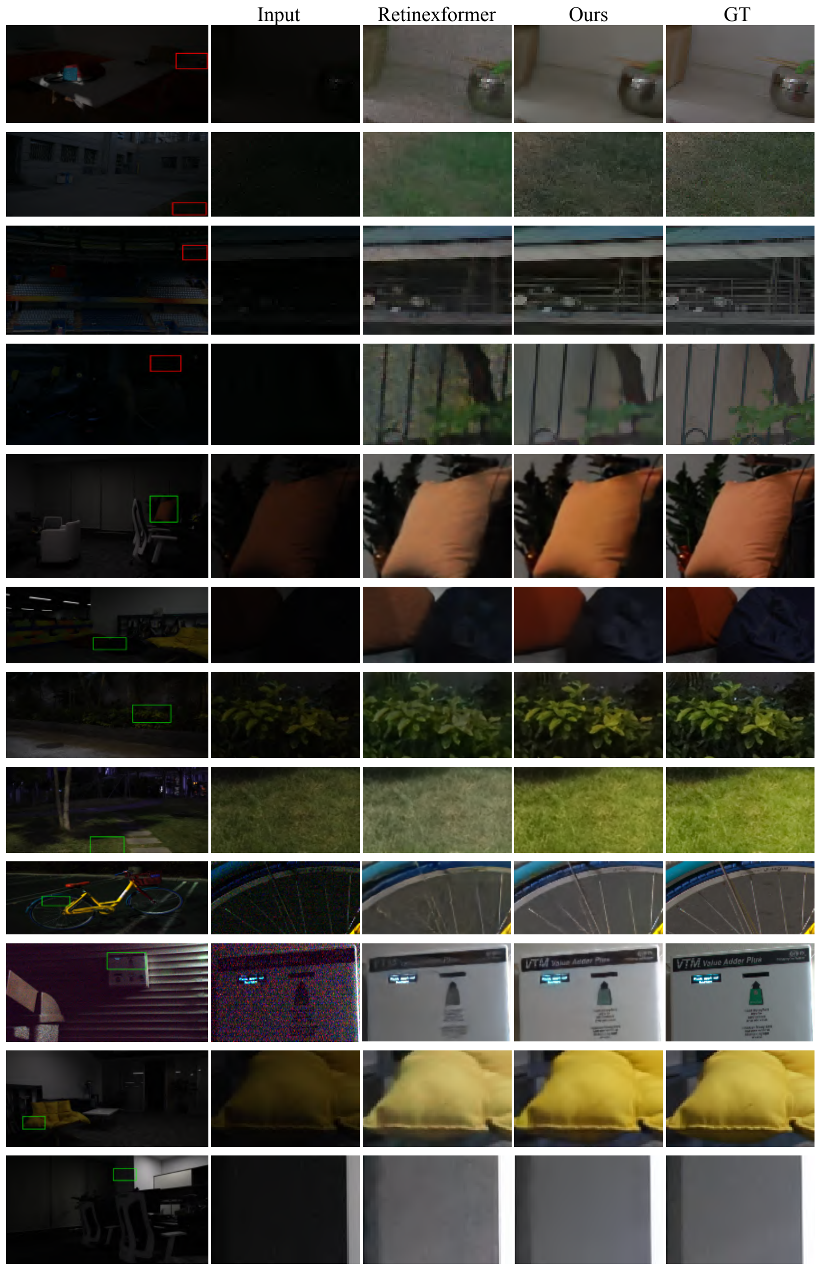

Fig. 2: Qualitative comparison between Retinexformer [3] and the enhanced version by our method. Samples are selected from LOL-v2 [52], SID [4], SDSD [43] (both indoor and outdoor parts) datasets. Our integration provides accurate outcomes for both high-frequency (clearer image detail restoration) and low-frequency (more accurate illumination recovery) areas.

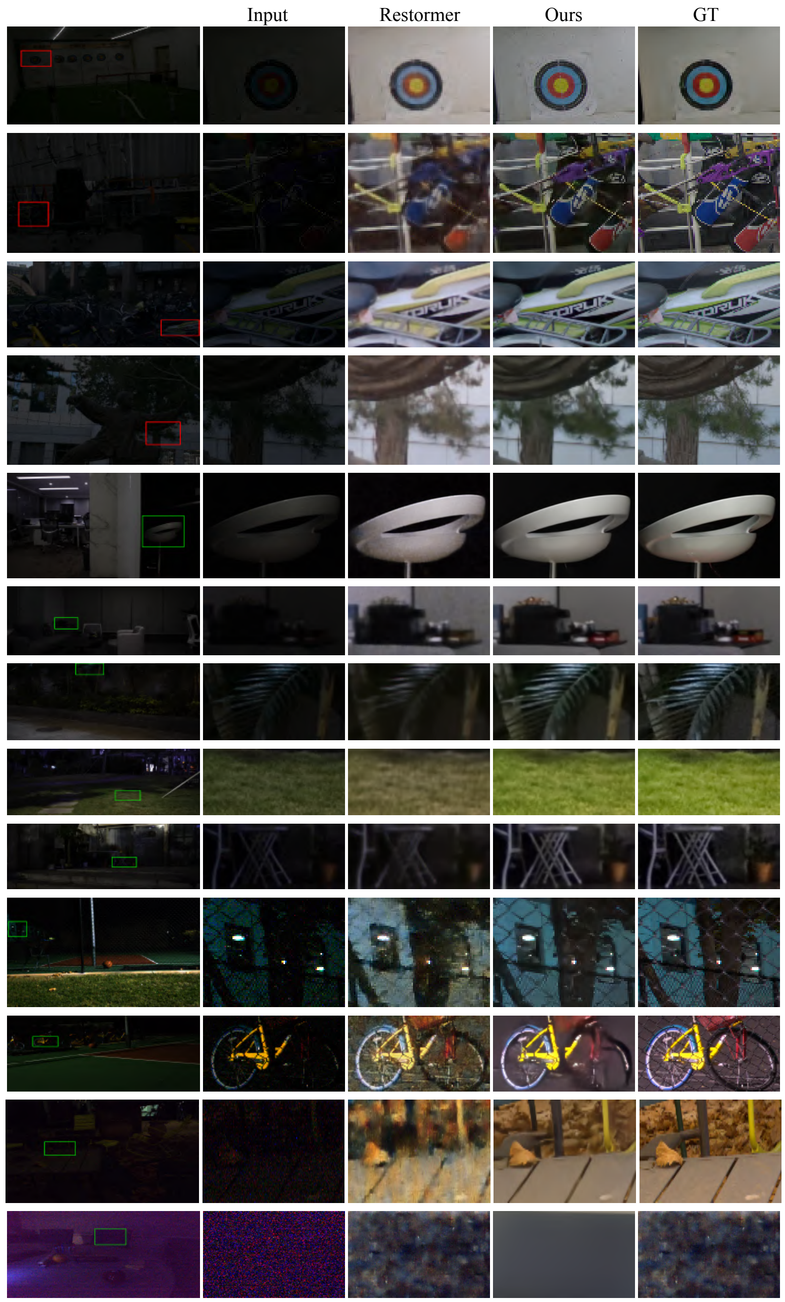

Fig. 3: Qualitative comparison between Restormer [55] and the enhanced version by our method. Samples are selected from LOL-v2 [52], SID [4], SDSD [43] (both indoor and outdoor parts) datasets. Our integration provides accurate outcomes for both high-frequency (clearer image detail restoration) and low-frequency (more accurate illumination recovery) areas.

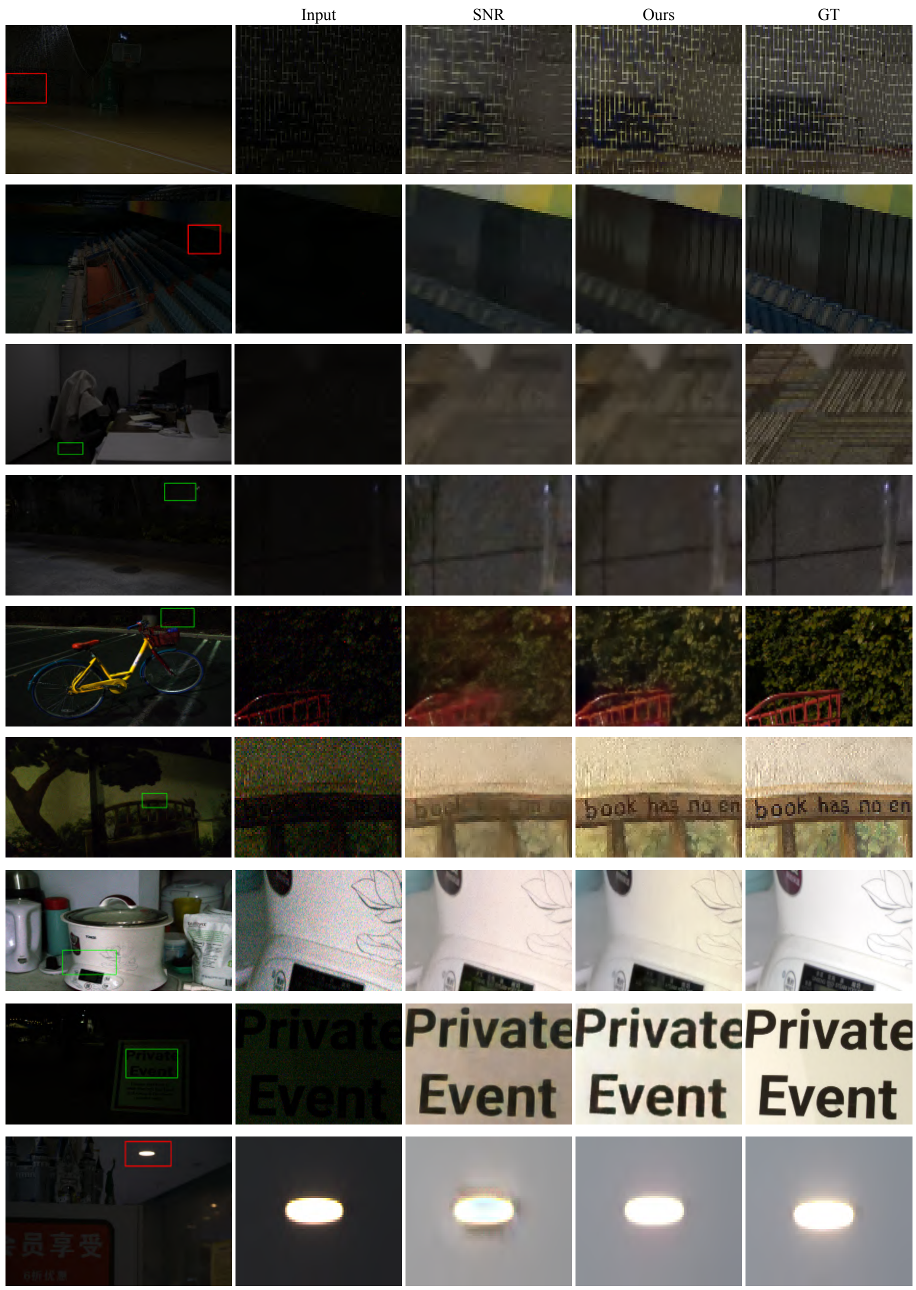

Fig. 4: Qualitative comparison between SNR [51] and the enhanced version by our method. Samples are selected from LOL-v2 [52], SID [4], SDSD [43] (both indoor and outdoor parts) datasets. Our integration provides accurate outcomes for both high-frequency (clearer image detail restoration) and low-frequency (more accurate illumination recovery) areas.

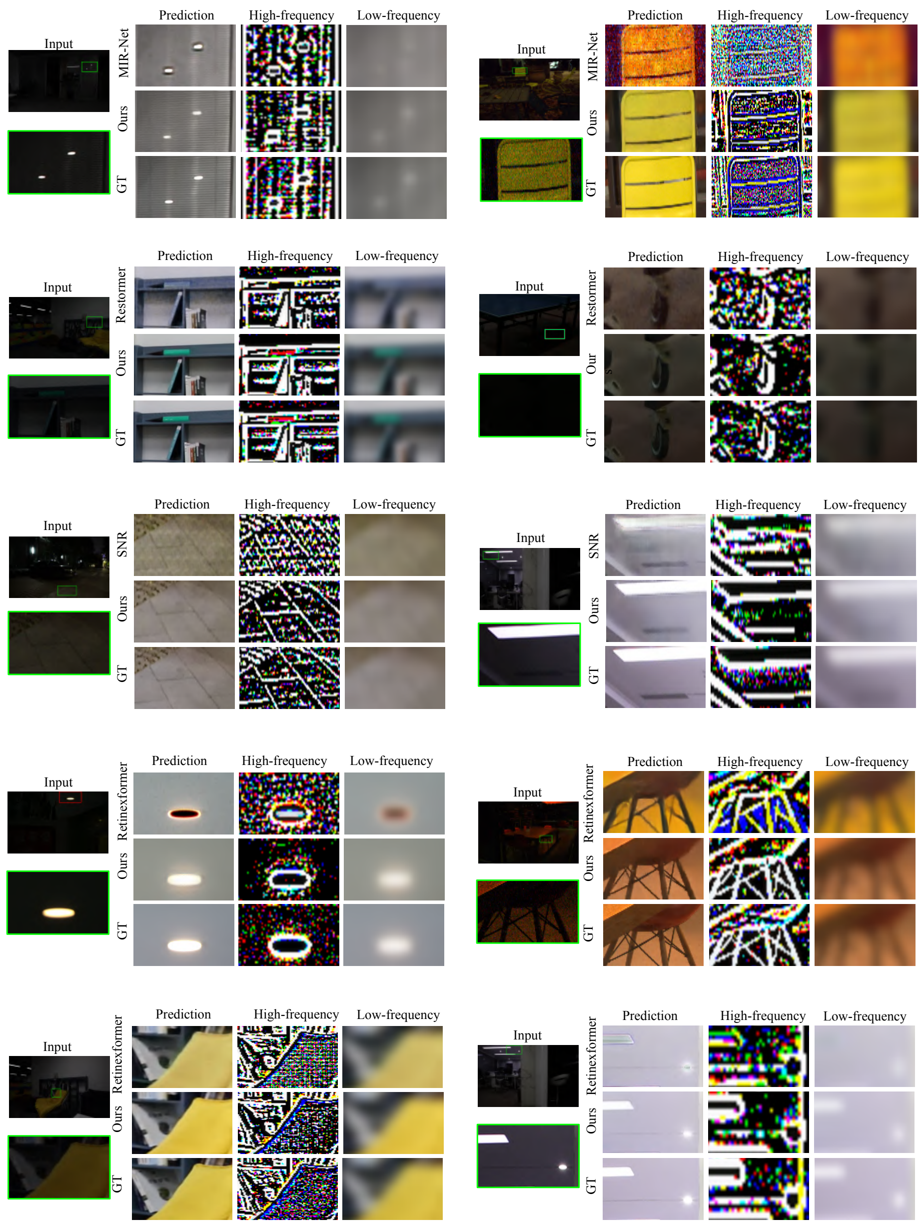

Fig. 5: Visual comparison between SOTA LLIE models and their enhanced version by our method. Our integration provides accurate outcomes for both high-frequency (clearer image detail restoration) and low-frequency (more accurate illumination recovery) areas.

References

1. Adelson, E.H., Anderson, C.H., Bergen, J.R., Burt, P.J., Ogden, J.M.: Pyramid methods in image processing. RCA engineer (1984)
2. Burt, P.J., Adelson, E.H.: The laplacian pyramid as a compact image code. In: Readings in computer vision, pp. 671–679. Elsevier (1987)
3. Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y.: Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In: ICCV (2023)
4. Chen, C., Chen, Q., Do, M.N., Koltun, V.: Seeing motion in the dark. In: ICCV (2019)
5. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. In: CVPR (2018)
6. Cui, Z., Li, K., Gu, L., Su, S., Gao, P., Jiang, Z., Qiao, Y., Harada, T.: You only need 90k parameters to adapt light: a light weight transformer for image enhancement and exposure correction. In: BMVC (2022)
7. Feng, S., Bo, L.: Low-light color image enhancement based on retinex. In: International Conference on Automation, Control and Robotics Engineering (CACRE) (2020)
8. Fu, H., Zheng, W., Meng, X., Wang, X., Wang, C., Ma, H.: You do not need additional priors or regularizers in retinex-based low-light image enhancement. In: CVPR (2023)
9. Fu, Z., Yang, Y., Tu, X., Huang, Y., Ding, X., Ma, K.K.: Learning a simple low-light image enhancer from paired low-light instances. In: CVPR (2023)
10. Gharbi, M., Chen, J., Barron, J.T., Hasinoff, S.W., Durand, F.: Deep bilateral learning for real-time image enhancement. ACM TOG (2017)
11. Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R.: Zero-reference deep curve estimation for low-light image enhancement. In: CVPR (2020)
12. Hao, S., Han, X., Guo, Y., Xu, X., Wang, M.: Low-light image enhancement with semi-decoupled decomposition. IEEE TMM (2020)
13. Hou, J., Zhu, Z., Hou, J., Liu, H., Zeng, H., Yuan, H.: Global structure-aware diffusion process for low-light image enhancement. NeurIPS (2023)
14. Hu, Y., He, H., Xu, C., Wang, B., Lin, S.: Exposure: A white-box photo post-processing framework. ACM TOG (2018)
15. Jiang, H., Luo, A., Fan, H., Han, S., Liu, S.: Low-light image enhancement with wavelet-based diffusion models. ACM TOG (2023)
16. Jobson, D.J., Rahman, Z.u., Woodell, G.A.: A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE TIP (1997)
17. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: ECCV (2016)
18. Kim, H.U., Koh, Y.J., Kim, C.S.: Global and local enhancement networks for paired and unpaired image enhancement. In: ECCV (2020)
19. Kim, H., Choi, S.M., Kim, C.S., Koh, Y.J.: Representative color transform for image enhancement. In: ICCV (2021)
20. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
21. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM review (2009)
22. Kuleshov, V., Chaganty, A., Liang, P.: Tensor factorization via matrix factorization. In: Artificial Intelligence and Statistics (2015)

23. Li, Z., Shu, H., Zheng, C.: Multi-scale single image dehazing using laplacian and gaussian pyramids. IEEE TIP (2021)
24. Liang, J., Zeng, H., Zhang, L.: High-resolution photorealistic image translation in real-time: A laplacian pyramid translation network. In: CVPR (2021)
25. Liang, X., Chen, X., Ren, K., Miao, X., Chen, Z., Jin, Y.: Low-light image enhancement via adaptive frequency decomposition network. Scientific Reports (2023)
26. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: ICCV (2021)
27. Loh, Y.P., Chan, C.S.: Getting to know low-light images with the exclusively dark dataset. Computer Vision and Image Understanding (2019)
28. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. In: ICLR (2017)
29. Moran, S., Marza, P., McDonagh, S., Parisot, S., Slabaugh, G.: Deeplpf: Deep local parametric filters for image enhancement. In: CVPR (2020)
30. Moran, S., McDonagh, S., Slabaugh, Gregorym, M., Ren, L., Tai, Y., Liu, X.: Curl: Neural curve layers for global image enhancement. In: International Conference on Pattern Recognition (ICPR) (2021)
31. Park, J., Lee, J.Y., Yoo, D., Kweon, I.S.: Distort-and-recover: Color enhancement using deep reinforcement learning. In: CVPR (2018)
32. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. NeurIPS (2019)
33. Peng, L., Zhu, C., Bian, L.: U-shape transformer for underwater image enhancement. In: ECCV (2021)
34. Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J.B., Zuiderveld, K.: Adaptive histogram equalization and its variations. Computer vision, graphics, and image processing (1987)
35. Reza, A.M.: Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. Journal of VLSI signal processing systems for signal, image and video technology (2004)
36. Risheng, L., Long, M., Jiaao, Z., Xin, F., Zhongxuan, L.: Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In: CVPR (2021)
37. Song, Y., Qian, H., Du, X.: Starenhancer: Learning real-time and style-aware image enhancement. In: ICCV (2021)
38. Souibgui, M.A., Biswas, S., Jemni, S.K., Kessentini, Y., Fornés, A., Lladós, J., Pal, U.: Docentr: An end-to-end document image enhancement transformer. In: International Conference on Pattern Recognition (ICPR) (2022)
39. Vonikakis, V., Kouskouridas, R., Gasteratos, A.: On the evaluation of illumination compensation algorithms. Multimedia Tools and Applications **77**, 9211–9231 (2018)
40. Wang, C., Pan, J., Wu, X.M.: Structural prior guided generative adversarial transformers for low-light image enhancement. arXiv preprint arXiv:2207.07828 (2022)
41. Wang, H., Chen, X., Ni, B., Liu, Y., Liu, J.: Omni aggregation networks for lightweight image super-resolution. In: CVPR (2023)
42. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: Eca-net: Efficient channel attention for deep convolutional neural networks. In: CVPR (2020)
43. Wang, R., Xu, X., Fu, C.W., Lu, J., Yu, B., Jia, J.: Seeing dynamic scene in the dark: A high-quality video dataset with mechatronic alignment. In: ICCV (2021)

44. Wang, T., Zhang, K., Shen, T., Luo, W., Stenger, B., Lu, T.: Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In: AAAI (2023)
45. Wang, Y., Wan, R., Yang, W., Li, H., Chau, L.P., Kot, A.: Low-light image enhancement with normalizing flow. In: AAAI (2022)
46. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE TIP (2004)
47. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. In: BMVC (2018)
48. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. In: BMVC (2018)
49. Wu, X., Shi, B., Dong, Y., Huang, C., Chawla, N.: Neural tensor factorization for temporal interaction learning. In: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (2018)
50. Xu, K., Yang, X., Yin, B., Lau, R.W.: Learning to restore low-light images via decomposition-and-enhancement. In: CVPR (2020)
51. Xu, X., Wang, R., Fu, C.W., Jia, J.: Snr-aware low-light image enhancement. In: CVPR (2022)
52. Yang, W., Wang, S., Fang, Y., Wang, Y., Liu, J.: From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In: CVPR (2020)
53. Yang, W., Wang, W., Huang, H., Wang, S., Liu, J.: Sparse gradient regularized deep retinex network for robust low-light image enhancement. IEEE TIP (2021)
54. Yuhui, W., Chen, P., Guoqing, W., Yang, Y., Jiwei, W., Chongyi, L., Shen, H.T.: Learning semantic-aware knowledge guidance for low-light image enhancement. In: CVPR (2023)
55. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: CVPR (2022)
56. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Learning enriched features for real image restoration and enhancement. In: ECCV (2020)
57. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Learning enriched features for fast image restoration and enhancement. IEEE TPAMI (2022)
58. Zhang, Z., Jiang, Y., Jiang, J., Wang, X., Luo, P., Gu, J.: Star: A structure-aware lightweight transformer for real-time image enhancement. In: ICCV (2021)
59. Zheng, C., Shi, D., Shi, W.: Adaptive unfolding total variation network for low-light image enhancement. In: ICCV (2021)
60. Zhou, K., Liu, K., Li, W., Han, X., Lu, J.: Mutual guidance and residual integration for image enhancement. arXiv preprint arXiv:2211.13919 (2022)