Supplementary Materials for Paper: ZeroIDIR: Zero-Reference Illumination Degradation Image Restoration with Perturbed Consistency Diffusion Models

Hai Jiang[1], Zhen Liu[2], Yinjie Lei[3], Songchen Han[1], Bing Zeng[2], Shuaicheng Liu[2,†]
[1]School of Aeronautics and Astronautics, Sichuan University
[2]University of Electronic Science and Technology of China
[3]College of Electronics and Information Engineering, Sichuan University
jianghai@stu.scu.edu.cn, {liuzhen03@std., liushuaicheng@}uestc.edu.cn

This supplementary material is organized as follows:

- Sec. 1 conducts more experiments, including more detailed ablation studies of our proposed method, quantitative comparisons on five real-world unpaired benchmarks, and computational efficiency analysis.
- Sec. 2 conducts a user study to assess the subjective visual quality and human preference of our restored images.
- Sec. 3 provides more qualitative results on the paired datasets as well as the real-world unpaired datasets.
- Sec. 4 presents the limitation of our method.

1. More Experiments

1.1. More Detailed Results of the Ablation Study

In this section, we conduct more detailed experiments to evaluate the effectiveness of our different component choices. We use the implementation details described in Sec. 4.1 of the main paper for training, and quantitative results on the LOL [29] and the over-exposed correction part of the SICE [3] datasets are illustrated in Table S1 and Table S2. Detailed experiment settings are discussed below.

Different Retinex Decomposition Strategy. To validate the effectiveness of the framework design of our AGCM, we first adopt two learning-based Retinex decomposition networks, including RetinexNet [29] and PairLIE [6], to replace our adopted traditional strategy [9]. Moreover, we utilize the illumination-corrected images obtained from these variants to serve as the intermediate noisy state of our PCDM to validate their impact on the overall performance. As reported in rows 1-4 of Table S1, adopting a more effective Retinex decomposition that accurately separates illumination and reflectance components indeed improves the overall performance, and better illumination-corrected images also facilitate the subsequent reconstruction process.

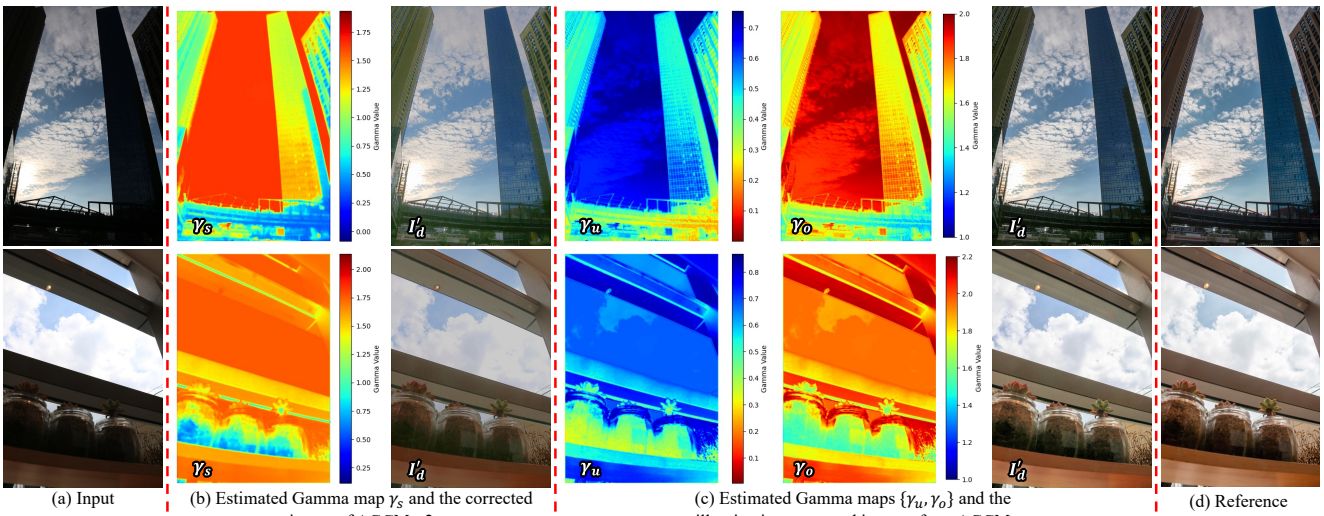
Effectiveness of AGCM. Moreover, we design a variant of the AGCM, denoted as AGCM-v2, which estimates a single spatially varying gamma map γ_s to perform exposure correction. As shown in Fig. S1, the proposed AGCM generates two complementary gamma maps $\{\gamma_u, \gamma_o\}$ that collectively capture both global and local illumination tendencies, enabling more adaptive exposure refinement for under-/over-exposed areas. In contrast, AGCM-v2 with a single gamma map struggles to model complex spatial illumination distributions, thus unable to achieve satisfactory results as reported in rows 5-6 of Table S1.

Table S1. Quantitative results of ablation studies about the proposed AGCM. Please refer to the text for details.

Method	LOL [29]			SICE-over [3]		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
RetinexNet [29]	19.223	0.610	0.329	14.611	0.586	0.282
+PCDM	20.465	0.809	0.187	16.983	0.669	0.261
PairLIE [6]	19.458	0.749	0.298	16.443	0.635	0.259
+PCDM	20.913	0.813	0.161	17.152	0.676	0.248
AGCM-v2	18.010	0.518	0.368	13.889	0.534	0.311
+PCDM	20.080	0.793	0.194	16.234	0.632	0.268
AGCM	19.599	0.535	0.351	14.506	0.574	0.293
+PCDM	20.874	0.811	0.167	16.975	0.661	0.259

Impact of t^*. We further analyze the influence of the predefined range of t^* on the performance of our PCDM. As shown in rows 1-4 of Table S2, using a smaller range restricts the ability of the diffusion model to cover the diverse degradation levels presented in illumination-corrected images, leading to sub-optimal performance. In contrast, adopting a larger range pushes the illumination-corrected inputs toward overly noisy states, where excessive perturbation weakens the correlation between the input and its underlying clean version, causing the diffusion model to struggle to infer accurate denoising trajectories and present inconsistent reconstructions. To this end, we set t^* within $[0, 50]$ to provide a balanced compromise between degradation simulation and reconstruction stability.

†Corresponding author.



(a) Input (b) Estimated Gamma map γ_s and the corrected image of AGCM-v2 (c) Estimated Gamma maps $\{\gamma_u, \gamma_o\}$ and the illumination-corrected image of our AGCM (d) Reference

Figure S1. Illustration of the estimated Gamma maps and the illumination-corrected images of AGCM v2 and our AGCM.

Table S2. Quantitative results of ablation studies about the proposed PCDM. Please refer to the text for details.

Method	LOL [29]			SICE-over [3]		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
$t^* \in [0, 10]$	20.164	0.795	0.184	16.431	0.634	0.271
$t^* \in [0, 30]$	20.660	0.808	0.173	16.782	0.648	0.264
$t^* \in [0, 100]$	20.129	0.804	0.196	16.105	0.632	0.277
$t^* \in [0, 200]$	19.742	0.787	0.201	15.832	0.596	0.288
$\mathbf{x}_{t^*} = \mathbf{R}_d$	20.374	0.781	0.221	16.245	0.610	0.277
Default	20.874	0.811	0.167	16.975	0.661	0.259

Different Perturbed State. We further conduct an ablation study to investigate the effect of different inputs for the diffusion stage. Following previous Retinex-based methods [29, 32, 33], we separately optimize the illuminance map $\mathbf{L}_d$ and reflectance map $\mathbf{R}_d$, where the illuminance map undergoes exposure correction via AGCM to prepare an illumination-corrected representation $\mathbf{L}'_d$ while the reflectance map serves as intermediate noisy state $\mathbf{x}_{t^*}$ to PCDM for noise suppression and detail reconstruction, resulting in the restored reflectance map $\mathbf{R}'_d$. The final restored image thus can be obtained as $\hat{I}_d = \mathbf{L}'_d \odot \mathbf{R}'_d$. However, the degradation in illumination-degraded images is not confined solely to the reflectance domain since information loss occurs across both illumination and reflectance components. Moreover, Retinex decomposition itself introduces additional decomposition errors and information distortion, further hampering the overall performance as reported in row 5 of Table S2. Therefore, employing the illumination-corrected image as the intermediate state allows the diffusion model to fully exploit its generative capability to compensate for overall information degradation.

1.2. More Quantitative Results

To explore the applicability of our method in real-world illumination degradation image restoration tasks, we compare the proposed method with competitive methods that perform well on each configuration on five real-world benchmarks: MEF [23], DICM [16], NPE [27], VV [26], and LIME [9]. We adopt two non-reference perceptual metrics NIQE [24] and MUSIQ [15] to measure the visual quality of the restored images. As shown in Table S3, unsupervised methods present better generalization ability than supervised ones on these unseen datasets, where our method obtains the best MUSIQ scores on all five datasets. For the NIQE scores, our method also achieves the best scores on three datasets, as well as the second-best results on the remaining two datasets. It indicates that our method is able to generate visually satisfactory images and can generalize well to various illumination degradation scenes.

1.3. Computational Efficiency

In this section, we report the model parameters, inference time, and memory consumption of our method and competitive unsupervised diffusion-based methods evaluated on an NVIDIA A100 GPU across three image resolutions: 256×256, 512×512, and 1024×1024. As summarized in Table S4, our model comprises approximately 134.86M parameters, with corresponding inference times of 1.07s, 4.06s, and 17.70s at the respective resolutions. Compared with existing zero-shot diffusion-based methods such as FourierDiff [22] and AGLLDiff [19], our method demonstrates superior computational efficiency. However, it remains less efficient than LightenDiff [12] and QuadPrior [28], which benefit from conducting the diffusion process in the latent space, thereby substantially reducing com-

Table S3. Quantitative comparisons on unpaired datasets MEF [23], DICM [16], NPE [27], VV [26], and LIME [9] for real-world illumination degradation image restoration. The best results are highlighted in **bold**, and the second-best are in underlined.

Methods	MEF [23]		DICM [16]		NPE [27]		VV [26]		LIME [9]	
	NIQE ↓	MUSIQ ↑	NIQE ↓	MUSIQ ↑	NIQE ↓	MUSIQ ↑	NIQE ↓	MUSIQ ↑	NIQE ↓	MUSIQ ↑
PyDiff [34]	4.290	0.494	4.476	0.527	4.091	0.532	4.493	0.516	4.883	0.521
MSLT [35]	4.481	0.404	4.310	0.511	4.001	0.522	4.654	0.520	4.710	0.478
AnlightenDiff [4]	4.471	0.392	4.603	0.437	4.611	0.439	5.579	0.421	4.286	0.458
CoTF [17]	4.228	0.448	4.027	0.516	**3.467**	0.520	3.661	0.482	4.143	0.512
Reti-Diff [11]	4.292	0.481	4.151	0.545	4.095	0.536	4.209	0.494	4.740	0.526
SLOT [14]	4.367	0.486	4.360	0.570	3.913	0.572	4.555	0.552	4.718	0.542
Zero-DCE [8]	**3.500**	0.496	3.951	0.475	3.826	0.524	5.080	0.562	4.379	0.510
EnlightenGAN [13]	4.099	0.438	4.097	0.502	3.895	0.501	3.759	0.518	4.670	0.482
NeRCo [31]	4.336	0.438	4.435	0.506	4.193	0.499	4.182	0.495	4.859	0.478
CLIP-LIT [18]	3.653	0.492	4.119	0.538	4.048	0.533	5.376	0.536	4.609	0.523
PSENet [25]	3.839	0.488	4.234	0.551	3.888	0.549	4.811	0.557	4.612	0.517
FourierDiff [22]	4.309	0.457	4.528	0.506	4.350	0.522	4.864	0.521	4.648	0.504
UEC [5]	3.994	0.456	4.852	0.502	3.963	0.543	5.725	0.458	4.721	0.546
QuadPrior [28]	4.762	0.463	4.910	0.535	4.284	0.523	4.051	0.478	5.442	0.518
RAVE [7]	4.023	0.507	4.177	0.566	3.933	0.557	4.902	0.573	4.810	0.557
LightenDiff [12]	4.153	0.463	4.122	0.530	3.618	0.519	3.094	0.470	4.406	0.484
AGLLDiff [19]	7.660	0.388	7.788	0.414	7.564	0.415	7.375	0.341	7.941	0.430
Ours	3.516	**0.529**	**3.882**	**0.607**	3.593	**0.576**	**3.067**	**0.605**	**3.862**	**0.563**

Table S4. The model parameters (Para.), inference time (Time), and memory consumption (Mem.) of our method and competitive unsupervised diffusion-based methods [12, 19, 22, 28] evaluated on an NVIDIA A100 GPU across three image resolutions of 256×256, 512×512, and 1024×1024.

Method	256×256		512×512		1024×1024		Para. (M)
	Time (s)	Mem. (GB)	Time (s)	Mem. (GB)	Time (s)	Mem. (GB)	
FourierDiff [22]	8.60	4.99	20.03	5.67	87.78	15.04	522.19
QuadPrior [28]	2.54	8.13	2.67	9.51	11.94	30.64	1244.02
LightenDiff [12]	0.49	1.01	0.56	1.27	0.61	2.51	27.84
AGLLDiff [19]	4.75	4.49	13.17	8.95	65.69	30.00	527.20
Ours	1.07	2.39	4.06	3.83	17.70	18.24	134.86

putational overhead and accelerating inference.

2. User Study

In this section, we conduct a user study to compare our method with five competitive supervised and unsupervised methods, including CoTF [17], Reti-Diff [11], SLOT [14], LightenDiff [12], and RAVE [7]. We randomly select 25 images from MEF [23], DICM [16], NPE [27], VV [26], and LIME [9] datasets, and invite 10 professional participants to rank the results of the above methods from 1 to 6 (the higher the better) based on illumination, color, and detail. As shown in Table S5, our method is consistently preferred by participants and receives a higher average score compared to other methods, proving that the restored images of our method are more visually pleasant.

Table S5. Average score (U.S.) of the user study. The best result is highlighted in **bold** and the second best is underlined.

Method	SLOT [14]	RetiDiff [11]	COTF [17]
U.S. (↑)	2.78	2.55	3.10
Method	LightenDiff [12]	RAVE [7]	ZeroIDIR
U.S (↑)	3.78	4.00	**4.67**

3. More Qualitative Results

In this section, we provide more qualitative results of our method and other competitive methods on the low-light image enhancement datasets LOL [29], LSRW [10], and MIT5K [2], the backlit image enhancement datasets BAID [21] and Backlit300 [18], the multiple exposure correction datasets MSEC [1] and SICE [3], as well as the unpaired real-world illumination degradation image restoration benchmarks MEF [23], DICM [16], NPE [27], VV [26], and LIME [9], as shown in Fig. S3-S8. We can see that previous methods appear to have incorrect exposure, color distortion, noise amplification, or artifacts, thereby undermining the overall visual quality. In contrast, our method effectively corrects the illumination and presents vivid color without introducing artifacts.

4. Limitations

Although our method demonstrates strong capability in restoring images captured under various challenging illumination conditions, it still struggles in extremely adverse

(a) Input (b) Our results

Figure S2. Failure cases of our method in scenes with extreme illumination degradation.

lighting scenarios, as shown in Fig. S2. Images obtained from such scenes often suffer from severe information loss and strong environmental noise, making the restoration process considerably more difficult. Furthermore, as reported in Table S4, our method performs iterative diffusion processes in the image space, thus showing inferior inference efficiency compared with the diffusion-based methods that perform in the latent space and previous lightweight methods. In future work, we plan to explore more effective sampling strategies, such as DPM-Solver++ [20] and the one-step sampling strategy [30], to improve inference efficiency and to further investigate the applicability of our method to the illumination degradation video restoration task.

References

[1] Mahmoud Afifi, Konstantinos G Derpanis, Bjorn Ommer, and Michael S Brown. Learning multi-scale photo exposure correction. In *CVPR*, pages 9157–9167, 2021. 3, 8

[2] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input/output image pairs. In *CVPR*, pages 97–104, 2011. 3, 5

[3] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE TIP*, 27(4):2049–2062, 2018. 1, 2, 3, 9

[4] Cheuk-Yiu Chan, Wan-Chi Siu, Yuk-Hee Chan, and H Anthony Chan. Anlightendiff: Anchoring diffusion probabilistic model on low light image enhancement. *IEEE TIP*, 2024. 3

[5] Ruodai Cui, Li Niu, and Guosheng Hu. Unsupervised exposure correction. In *ECCV*, pages 252–268, 2024. 3

[6] Zhenqi Fu, Yan Yang, Xiaotong Tu, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Learning a simple low-light image enhancer from paired low-light instances. In *CVPR*, pages 22252–22261, 2023. 1

[7] Tatiana Gaintseva, Martin Benning, and Gregory Slabaugh. Rave: Residual vector embedding for clip-guided backlit image enhancement. In *ECCV*, pages 412–428, 2024. 3

[8] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *CVPR*, pages 1780–1789, 2020. 3

[9] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE TIP*, 26(2):982–993, 2016. 1, 2, 3, 10

[10] Jiang Hai, Zhu Xuan, Ren Yang, Yutong Hao, Fengzhu Zou, Fang Lin, and Songchen Han. R2rnet: Low-light image enhancement via real-low to real-normal network. *Journal of Visual Communication and Image Representation*, 90: 103712, 2023. 3, 5

[11] Chunming He, Chengyu Fang, Yulun Zhang, Kai Li, Longxiang Tang, Chenyu You, Fengyang Xiao, Zhenhua Guo, and Xiu Li. Reti-diff: Illumination degradation image restoration with retinex-based latent diffusion model. In *ICLR*, 2025. 3

[12] Hai Jiang, Ao Luo, Xiaohong Liu, Songchen Han, and Shuaicheng Liu. Lightendiffusion: Unsupervised low-light image enhancement with latent-retinex diffusion models. In *ECCV*, pages 161–179, 2024. 2, 3

[13] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE TIP*, 30:2340–2349, 2021. 3

[14] Donggo Jung, Daehyun Kim, Guanghui Wang, and Tae Hyun Kim. Exposure-slot: Exposure-centric representations learning with slot-in-slot attention for region-aware exposure correction. In *CVPR*, 2025. 3

[15] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *ICCV*, pages 5148–5157, 2021. 2

[16] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE TIP*, 22(12):5372–5384, 2013. 2, 3, 10

[17] Ziwen Li, Feng Zhang, Meng Cao, Jinpu Zhang, Yuanjie Shao, Yuehuan Wang, and Nong Sang. Real-time exposure correction via collaborative transformations and adaptive sampling. In *CVPR*, pages 2984–2994, 2024. 3

[18] Zhexin Liang, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Iterative prompt learning for unsupervised backlit image enhancement. In *ICCV*, pages 8094–8103, 2023. 3, 7

[19] Yunlong Lin, Tian Ye, Sixiang Chen, Zhenqi Fu, Yingying Wang, Wenhao Chai, Zhaohu Xing, Wenxue Li, Lei Zhu, and Xinghao Ding. Aglldiff: Guiding diffusion models towards unsupervised training-free real-world low-light image enhancement. In *AAAI*, pages 5307–5315, 2025. 2, 3

[20] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver++: Fast solver for guided sampling of diffusion probabilistic models. *Machine Intelligence Research*, pages 1–22, 2025. 4

[21] Xiaoqian Lv, Shengping Zhang, Qinglin Liu, Haozhe Xie, Bineng Zhong, and Huiyu Zhou. Backlitnet: A dataset and network for backlit image enhancement. *Computer Vision and Image Understanding*, 218:103403, 2022. 3, 6

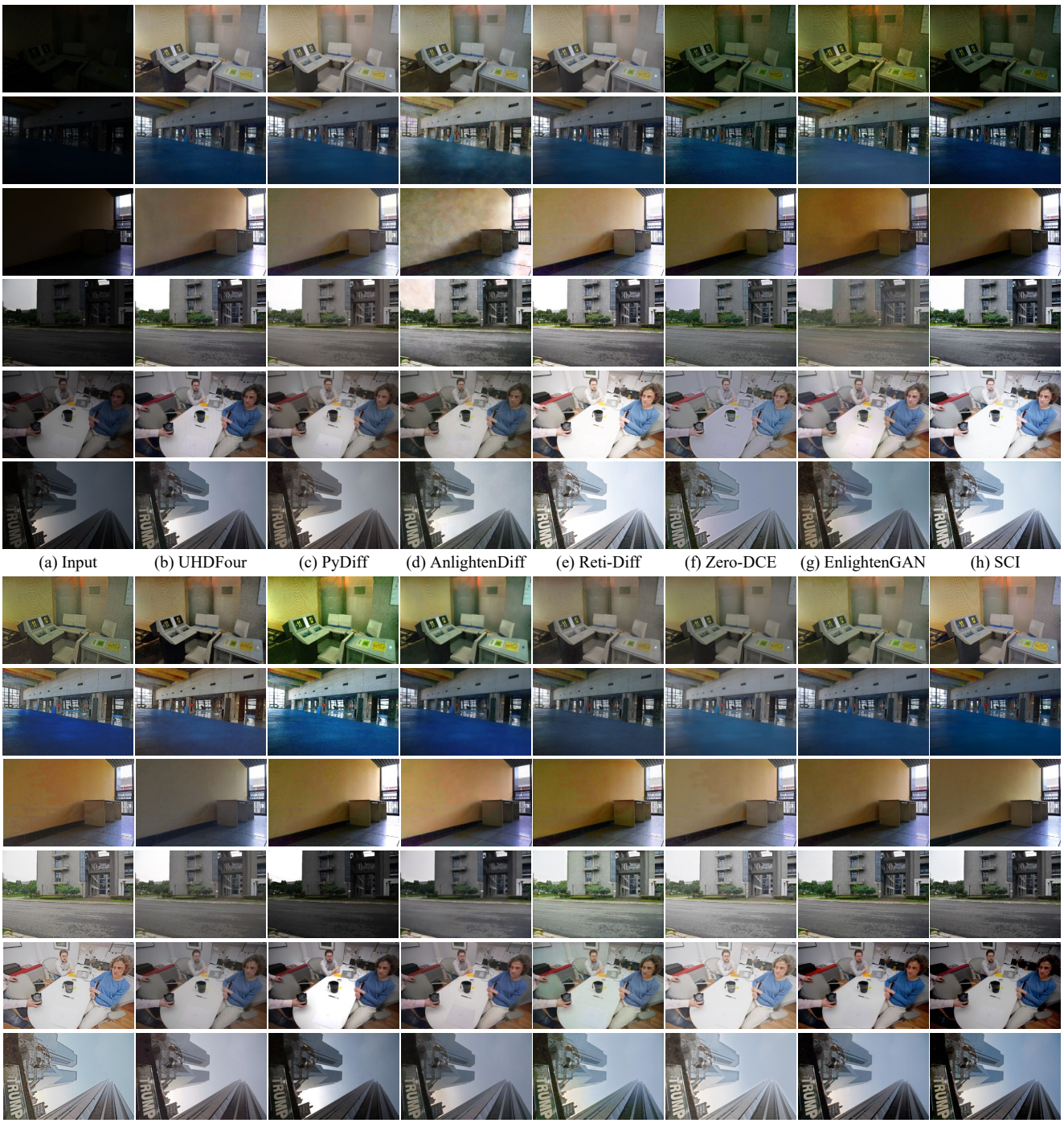


Figure S3. Qualitative comparison of our method and competitive methods on the LOL [29] (rows 1-2), LSRW [10] (rows 3-4), and MIT5K [2] (rows 5-6) test sets for low-light image enhancement (LLIE). Best viewed by zooming in.

[22] Xiaoqian Lv, Shengping Zhang, Chenyang Wang, Yichen Zheng, Bineng Zhong, Chongyi Li, and Liqiang Nie. Fourier priors-guided diffusion for zero-shot joint low-light enhancement and deblurring. In *CVPR*, pages 25378–25388, 2024. 2, 3

[23] Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE TIP*, 24 (11):3345–3356, 2015. 2, 3, 10

[24] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Sign.*

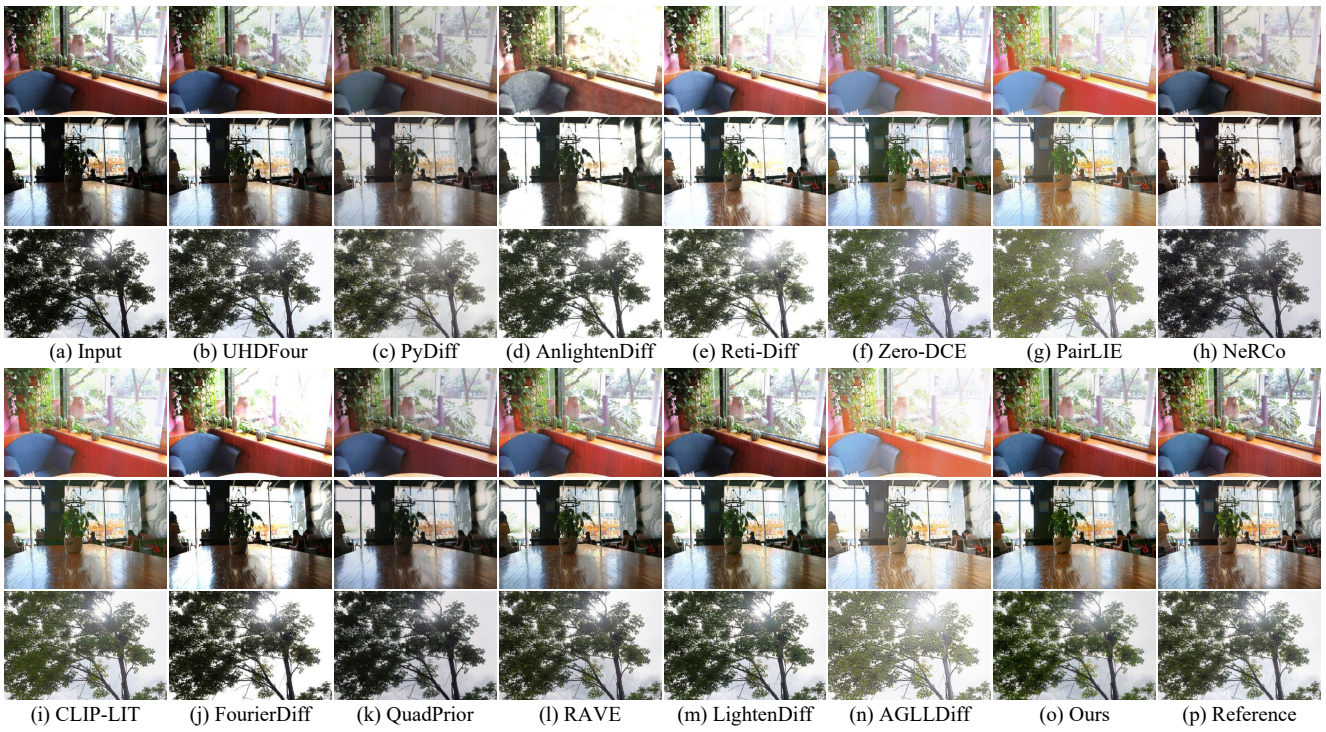


Figure S4. Qualitative comparison of our method and competitive methods on the BAID [21] test set for backlit image enhancement (BIE). Best viewed by zooming in.

Process. Letters, 20(3):209–212, 2012. 2

[25] Hue Nguyen, Diep Tran, Khoi Nguyen, and Rang Nguyen. Psenet: Progressive self-enhancement network for unsupervised extreme-light image enhancement. In *WACV*, pages 1756–1765, 2023. 3

[26] Vassilios Vonikakis, Rigas Kouskouridas, and Antonios Gasteratos. On the evaluation of illumination compensation algorithms. *Multimedia Tools and Applications*, 77:9211–9231, 2018. 2, 3, 10

[27] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE TIP*, 22(9):3538–3548, 2013. 2, 3, 10

[28] Wenjing Wang, Huan Yang, Jianlong Fu, and Jiaying Liu. Zero-reference low-light enhancement via physical quadruple priors. In *CVPR*, pages 26057–26066, 2024. 2, 3

[29] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *BMVC*, 2018. 1, 2, 3, 5

[30] Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-step effective diffusion network for real-world image super-resolution. *NeurIPS*, 37:92529–92553, 2024. 4

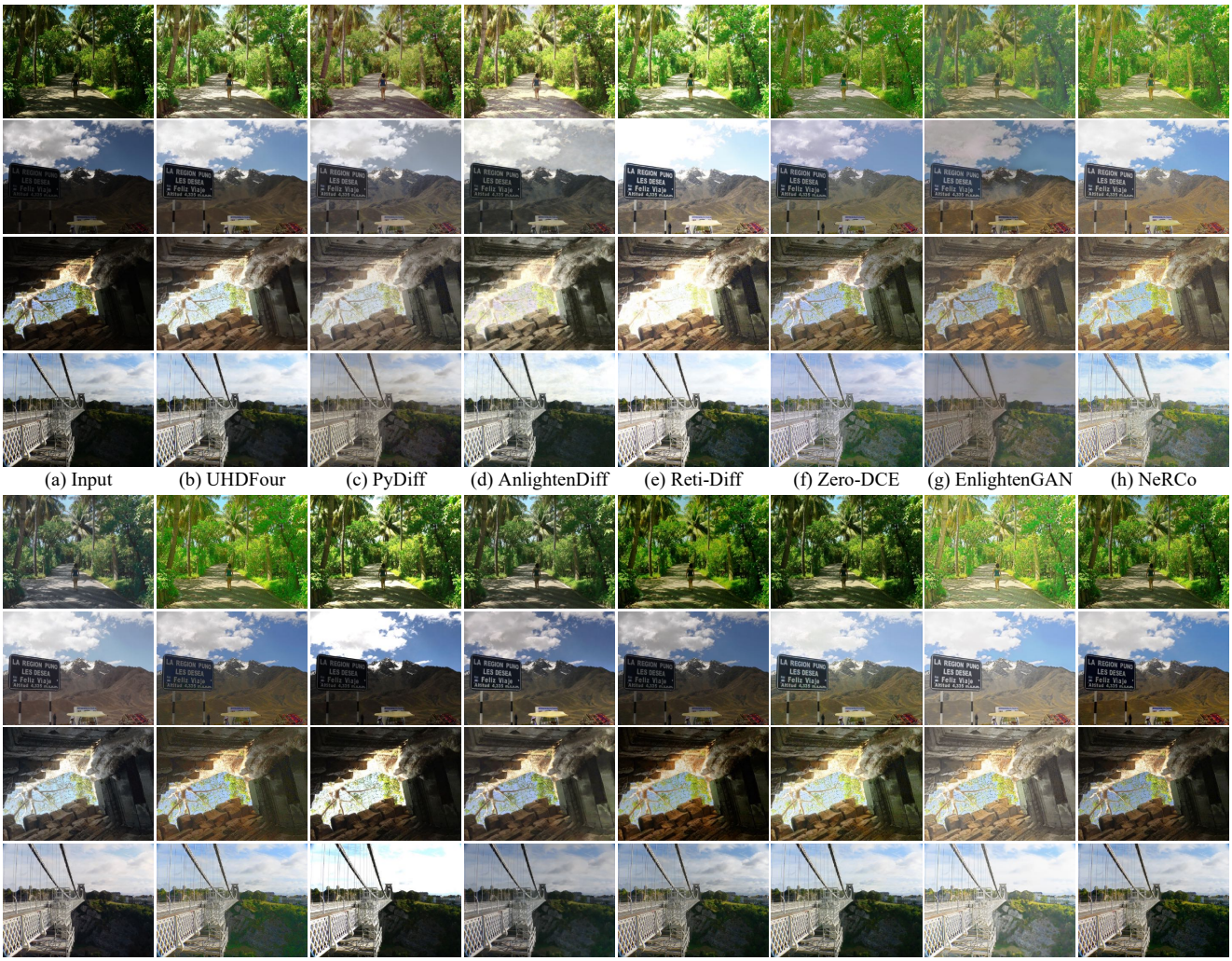
[31] Shuzhou Yang, Moxuan Ding, Yanmin Wu, Zihan Li, and Jian Zhang. Implicit neural representation for cooperative low-light image enhancement. In *ICCV*, pages 12918–12927, 2023. 3

[32] Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *ICCV*, pages 12302–12311, 2023. 2

[33] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *IJCV*, 129:1013–1037, 2021. 2

[34] Dewei Zhou, Zongxin Yang, and Yi Yang. Pyramid diffusion models for low-light image enhancement. In *IJCAI*, 2023. 3

[35] Yijie Zhou, Chao Li, Jin Liang, Tianyi Xu, Xin Liu, and Jun Xu. 4k-resolution photo exposure correction at 125 fps with˜ 8k parameters. In *WACV*, pages 1587–1597, 2024. 3



(a) Input (b) UHDFour (c) PyDiff (d) AnlightenDiff (e) Reti-Diff (f) Zero-DCE (g) EnlightenGAN (h) NeRCo

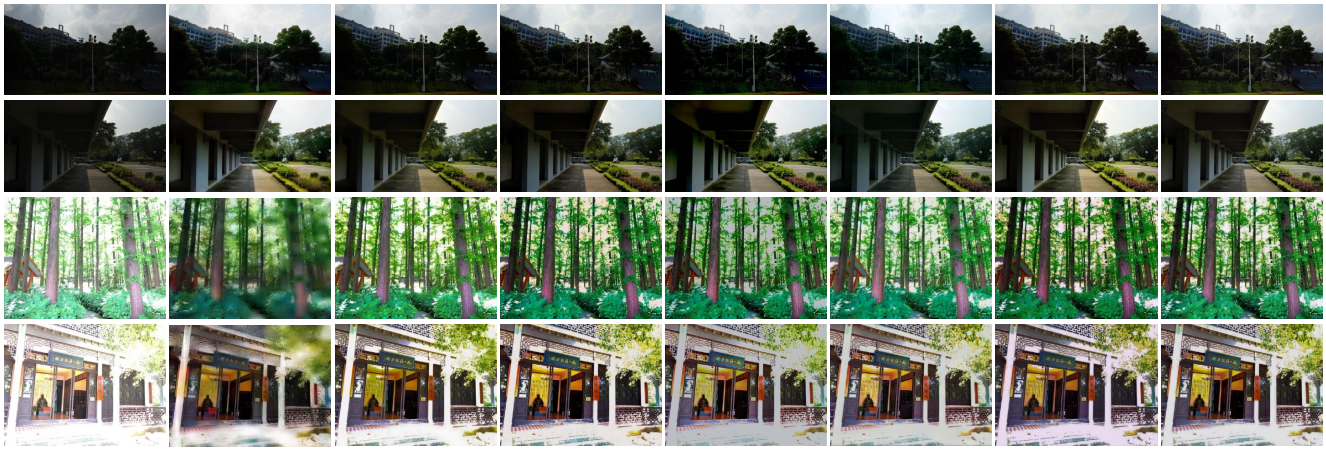
(i) NeRCo (j) CLIP-LIT (k) FourierDiff (l) QuadPrior (m) RAVE (n) LightenDiff (o) AGLLDiff (p) Ours

Figure S5. Qualitative comparison of our method and competitive methods on the Backlit300 [18] dataset for backlit image enhancement (BIE). Best viewed by zooming in.

(a) Input (b) MSEC (c) CMEC (d) FECNet (e) IAT (f) MSLT (g) CoTF (h) SLOT

(i) Exposure (j) DUAL (k) ExCNet (l) PEC (m) PSENet (n) UEC (o) Ours (p) Reference

Figure S6. Qualitative comparison of our method and competitive methods on the MSEC [1] test set for multiple exposure correction (MEC). Best viewed by zooming in.



(a) Input (b) MSEC (c) CMEC (d) FECNet (e) IAT (f) MSLT (g) CoTF (h) SLOT

(i) Exposure (j) DUAL (k) ExCNet (l) PEC (m) PSENet (n) UEC (o) Ours (p) Reference

Figure S7. Qualitative comparison of our method and competitive methods on the SICE [3] dataset for multiple exposure correction (MEC). Best viewed by zooming in.

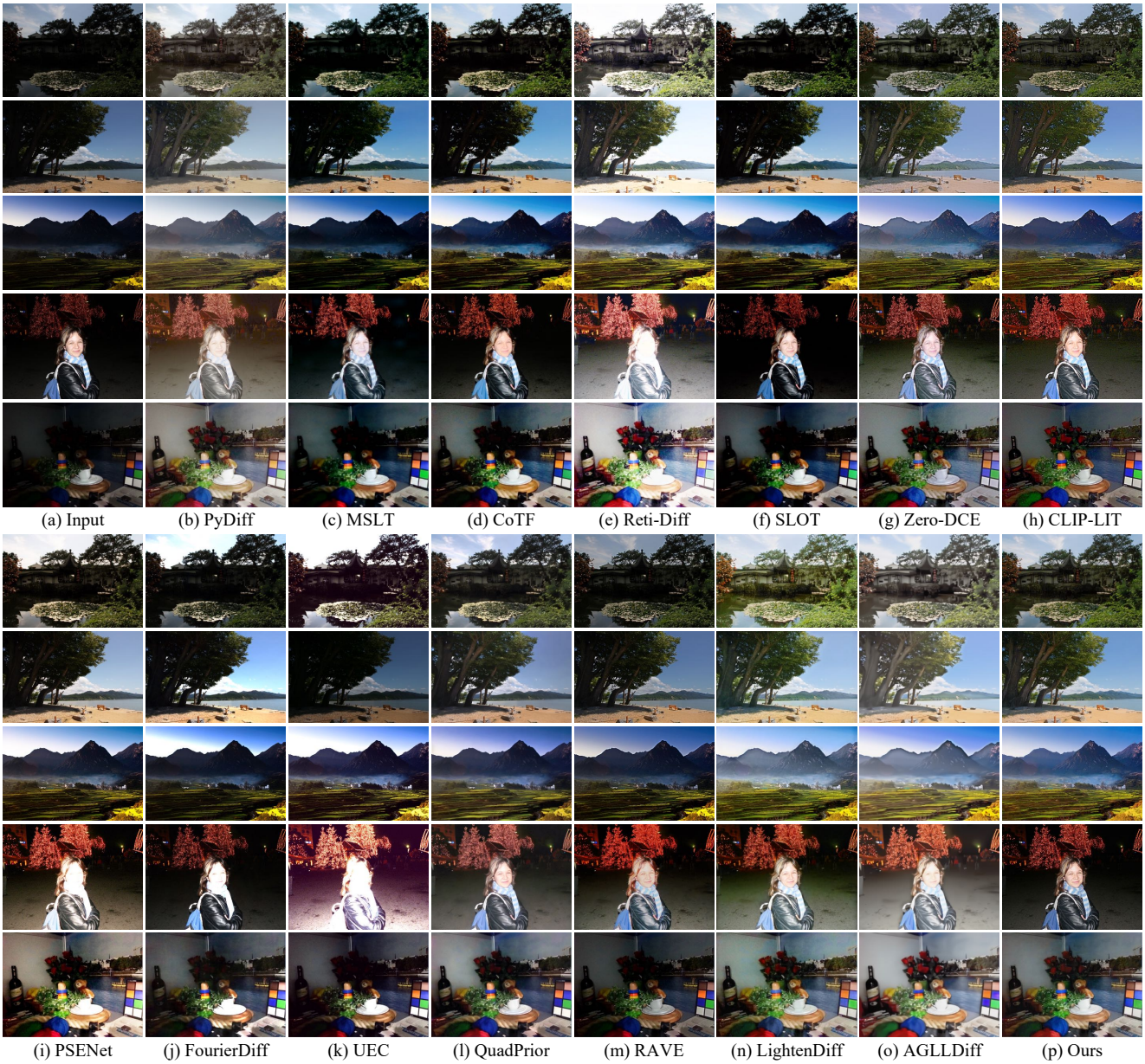


Figure S8. Qualitative comparison of our method and competitive methods on the MEF [23] (row 1), DICM [16] (row 2), NPE [27] (row 3), VV [26] (row 4), and LIME [9] (row 5) datasets for real-world illumination degradation image restoration. Best viewed by zooming in.