Supplementary Material
ShadowDiffusion: When Degradation Prior Meets Diffusion Model for Shadow Removal

Lanqing Guo[1], Chong Wang[1], Wenhan Yang[2], Siyu Huang[3], Yufei Wang[1], Hanspeter Pfister[3], Bihan Wen[1*]
[1]Nanyang Technological University, Singapore
[2]Peng Cheng Laboratory, China [3]Harvard University, USA
{lanqing001, wang1711, yufei001, bihan.wen}@ntu.edu.sg,
yangwh@pcl.ac.cn, huang@seas.harvard.edu, pfister@g.harvard.edu

In this supplementary material, we include more implementation details of the proposed ShadowDiffusion (Section A), more visual comparisons on ISTD+ [19] and SRD [22] datasets (Section B), and the detailed extension experiments on other image enhancement tasks (Section C). The code will be released.

A. Implementation Details

We used the same network architecture for all experiments. The diffusion model related configurations and parameters are summarised in Table A. The network had a U-Net architecture based on [26], which has four-scale resolutions and contains two residual blocks per resolution. We also use group normalization and self-attention blocks at 16×16 feature map resolution. We employed input time step embedding for t through sinusoidal positional encoding [27] and fed these time embedding into each residual block, enabling the model to share parameters across iterations. For the conditional input, we channel-wise concatenate the shadow image $\mathbf{y}$, $\mathbf{x}_t$, and $\mathbf{m}_t$, resulting in seven dimensional input image channels (*i.e.*, RGB for $\mathbf{y}$ and $\mathbf{x}_t$, and gray channel for $\mathbf{m}_t$). We did not perform task-specific or dataset-specific parameter tuning or modifications to the neural network architecture.

	Hyper-parameters		Hyper-parameters
Diffusion steps (T)	1000	Noise schedule (β_t)	linear: $0.0001 \rightarrow 0.02$
Base channels	64	channel multipliers	$\{1, 2, 4, 8\}$
Residual blocks per resolution	2	Attention resolution	16×16

Table A. Diffusion model configurations and parameter choices.

Moreover, according to Eq. (11) & (12) in the main paper, the penalty parameter ρ should be large enough to enforce $\mathbf{x}$ and $\mathbf{z}$ as well as $\mathbf{m}$ and $\mathbf{v}$ are approximately equal to the fixed point. To guarantee the convergence of unrolling-inspired diffusive iteration, following [4], we set ρ to increase linearly with the diffusive sampling step t as $\rho_{t-1} = \gamma\rho_t$, for a constant $\gamma > 1$. The convergence analysis can refer to [4, 14].

B. More Visual Examples

Figure A and Figure B illustrate some visual results on SRD [22] and ISTD+ [19] datasets, respectively. Besides, to verify the effectiveness of our method on different resolution input, we also provide some visual examples on the original resolution over ISTD+ [19] as shown in Figure C.

*Corresponding author: Bihan Wen.
This work was carried out at ROSE Lab, supported in part by the MOE AcRF Tier 1 (RG61/22) and Start-Up Grant.

C. Extension on Other Image Enhancement Tasks

As we have mentioned in the main paper, the shadow degradation model can be written as

$$\mathbf{y} = \mathbf{h} \cdot \mathbf{x} = \mathbf{w} \cdot \mathbf{m} \cdot \mathbf{x} + (\mathbf{1} - \mathbf{m}) \cdot \mathbf{x} \, , \tag{A}$$

where $\mathbf{h}$ denotes the pixel-wise illumination degradation map, which can be decomposed into the shadow mask $\mathbf{m}$ and illumination weight $\mathbf{w}$. If the shadow mask is the all one matrix, the shadow degradation model can be extended to low-light enhancement and exposure correction $\mathbf{y} = \mathbf{h} \cdot \mathbf{x} = \mathbf{w} \cdot \mathbf{m} \cdot \mathbf{x}$. Here the element in degradation map $\mathbf{h}$ is larger than zero.

Method	PSNR ↑	SSIM ↑	LPIPS ↓	Method	PSNR ↑	SSIM ↑	LPIPS ↓
LIME [13]	14.02	0.56	0.35	DRBN [31]	20.08	0.83	0.16
Retinex-Net [5]	16.77	0.56	0.35	Zhao *et al.* [38]	21.71	0.83	0.20
EnlightenGAN [17]	17.48	0.65	0.32	KinD++ [36]	21.30	0.82	0.16
Zero-DCE [12]	14.86	0.54	0.34	Lv *et al.* [21]	20.24	0.79	0.14
LR3M [23]	18.91	0.75	0.28	URetinex-Net [30]	21.33	0.83	0.12
RUAS [25]	18.23	0.72	0.35	MIRNet [33]	24.14	0.84	0.13
KinD [37]	20.38	0.80	0.17	Ours	**27.36**	**0.93**	**0.10**

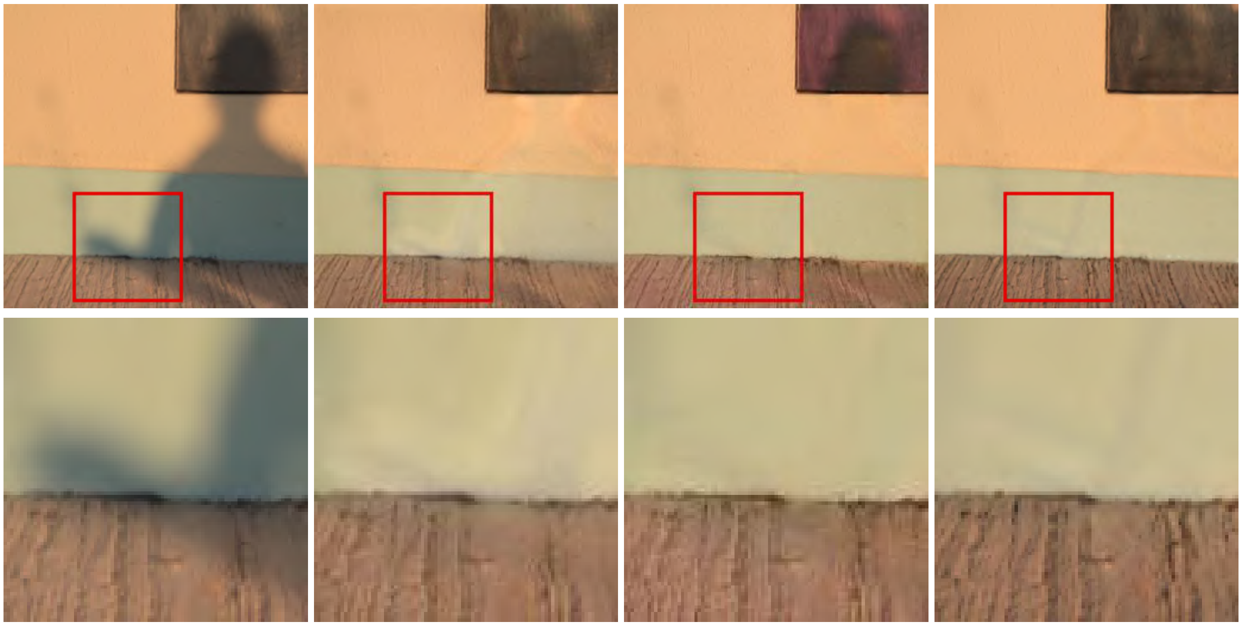
Table B. The quantitative results of low-light enhancement on LOL dataset [5].

Method	Expert A		Expert B		Expert C		Expert D		Expert E		Avg.	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
HE [1]	16.148	0.685	16.283	0.671	16.525	0.696	16.639	0.668	17.305	0.688	16.580	0.682
CLAHE [24]	14.884	0.589	15.669	0.610	15.383	0.599	15.452	0.601	15.737	0.610	15.425	0.602
WVM [11]	14.488	0.665	15.803	0.699	15.117	0.678	15.863	0.693	16.469	0.704	15.548	0.688
LIME [13]	11.154	0.591	11.828	0.610	11.517	0.607	12.638	0.628	13.613	0.653	12.150	0.618
HDR CNN [9] w/ RHT [32]	13.709	0.467	13.921	0.458	13.800	0.474	13.716	0.446	13.558	0.454	13.741	0.460
HDR CNN [9] w/ PS [8]	15.812	0.667	16.970	0.699	16.428	0.681	17.301	0.687	18.650	0.702	17.032	0.687
DPED (iPhone) [16]	15.134	0.609	16.505	0.636	15.907	0.622	16.571	0.627	17.251	0.649	16.274	0.629
DPED (BlackBerry) [16]	16.910	0.642	18.649	0.713	17.606	0.653	18.070	0.679	18.217	0.668	17.890	0.671
DPED (Sony) [16]	17.419	0.675	18.636	0.701	18.020	0.683	17.554	0.660	17.778	0.663	17.881	0.676
DPE (HDR) [6]	15.690	0.614	16.548	0.626	16.305	0.626	16.147	0.615	16.341	0.633	16.206	0.623
DPE (U-FiveK) [6]	16.240	0.653	16.805	0.646	16.837	0.671	16.762	0.654	16.707	0.650	16.670	0.655
DPE (S-FiveK) [6]	16.933	0.678	17.701	0.668	17.741	0.696	17.572	0.674	17.601	0.670	17.510	0.677
HQEC [34]	13.385	0.641	14.470	0.666	13.911	0.656	14.891	0.674	15.777	0.692	14.487	0.666
RetinexNet [5]	10.759	0.585	11.613	0.596	11.135	0.605	11.987	0.615	12.671	0.636	11.633	0.607
Deep UPE [28]	13.161	0.610	13.901	0.642	13.689	0.632	14.806	0.649	15.678	0.667	14.247	0.640
Zero-DCE [12]	11.643	0.536	12.555	0.539	12.058	0.544	12.964	0.548	13.769	0.580	12.598	0.549
Afifi *et al.* [2]	19.158	0.746	20.096	0.734	20.205	0.769	18.975	0.719	18.983	0.727	19.483	0.739
Ours	**22.742**	**0.828**	**24.224**	**0.848**	**22.662**	**0.846**	**21.651**	**0.834**	**20.366**	**0.820**	**22.329**	**0.835**

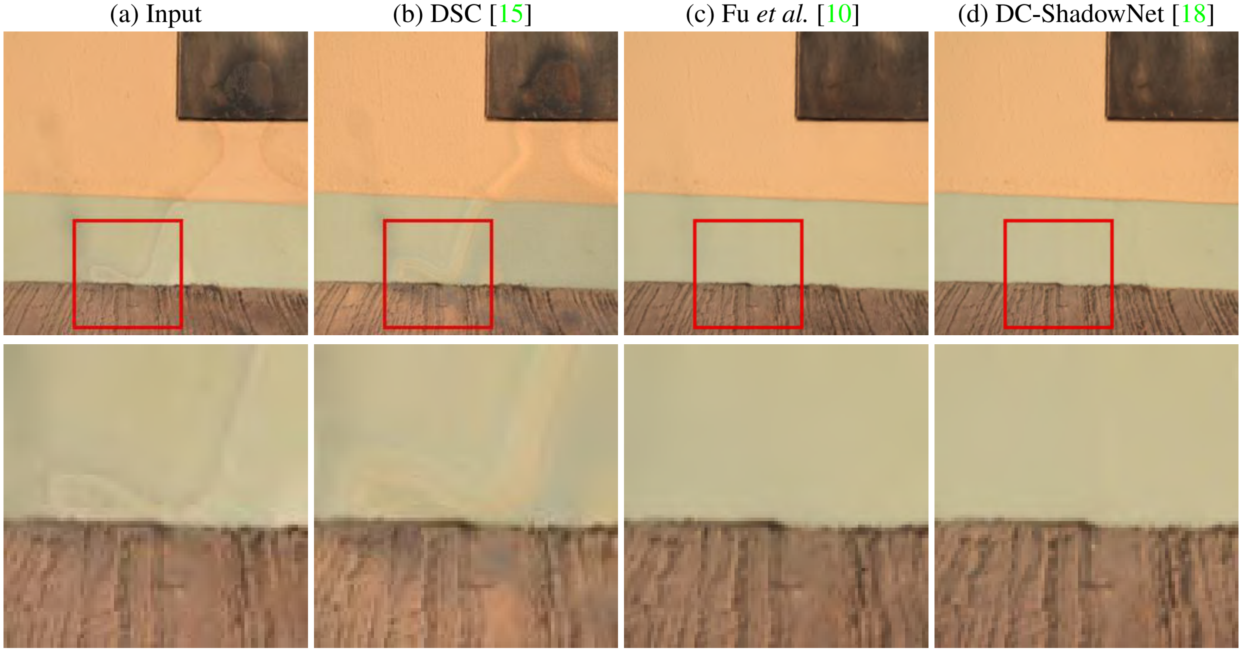
Table C. The quantitative results on the exposure correction dataset [2]. We compare each method with properly exposed reference image sets rendered by five expert photographers [3].

Low-light enhancement. We evaluate our ShadowDiffusion on the widely-used LOL real captured low/normal light images [5], which includes 485 images for training and 15 images for testing. We select the non-learning based method LIME [13], unsupervised methods EnlightenGAN [17], LR3M [23], and Zero-DCE [12], fully supervised methods Retinex-Net [5], RUAS [25], Zhao *et al.* [38], KinD [37], Lv *et al.* [21], MIRNet [33], and URetinex-Net [30], and semi-supervised method DRBN [31] as the competitors. Three metrics are adopted for quantitative comparison including PSNR, SSIM [29], and LPIPS [35]. The numerical results among different methods are reported in Table B. As shown in Table B, we can find that our method significantly outperforms all the other competing methods. The higher PSNR values indicate that the restored images contain fewer artifacts and the color information is accurately recovered. The higher SSIM values demonstrate that the restored images have more complete structural information with richer details. Besides, the LPIPS is designed for human perception, which shows the embedded feature similarity between restored results and ground truth. Figures D & E illustrate some visual results on LOL [30] dataset. In general, most of the previous methods fail to suppress the amplified noise and preserve the structural details, while our method can well restore the underlying structures from the darkness.

Exposure correction. We evaluate our ShadowDiffusion on the recent public available exposure correction dataset [2], which is rendered from the MIT-Adobe FiveK dataset [3] consisting of 17,675 images as training set, 750 images as validation set, and 5,905 images as testing set. We select the non-learning based methods histogram equalization (HE) [1], contrast-limited adaptive histogram equalization (CLANE) [24], the weighted variational model (WVM) [11], the low-light image enhancement method (LIME) [13], HDR CNN [9], DPED models [16], deep photo enhancer (DPE) models [6], the high-quality exposure correction method (HQEC) [34], RetinexNet [5], deep underexposed photo enhancer (UPE) [28], zero-reference deep curve estimation method (Zero-DCE) [12], and Afifi *et al.* [2]. We adopt the PSNR and SSIM metrics for quantitative comparison, where we compare the results against five different expert photographers in the MIT-Adobe FiveK dataset [3] following previous work [2]. Table C summarizes the quantitative results obtained by each method. The qualitative comparison has been demonstrated in Figure F, in which the most competing method, *i.e.*, Afifi *et al.* [2], always produce unnatural colour-distortion results (*e.g.*, the first, second, and fourth rows in Figure F), and lead to severe artifacts for some challenging cases (*e.g.*, the ghosts artifacts in the dark background as shown in the third row in Figure F), and over-exposure artifacts in the fifth row in Figure F).



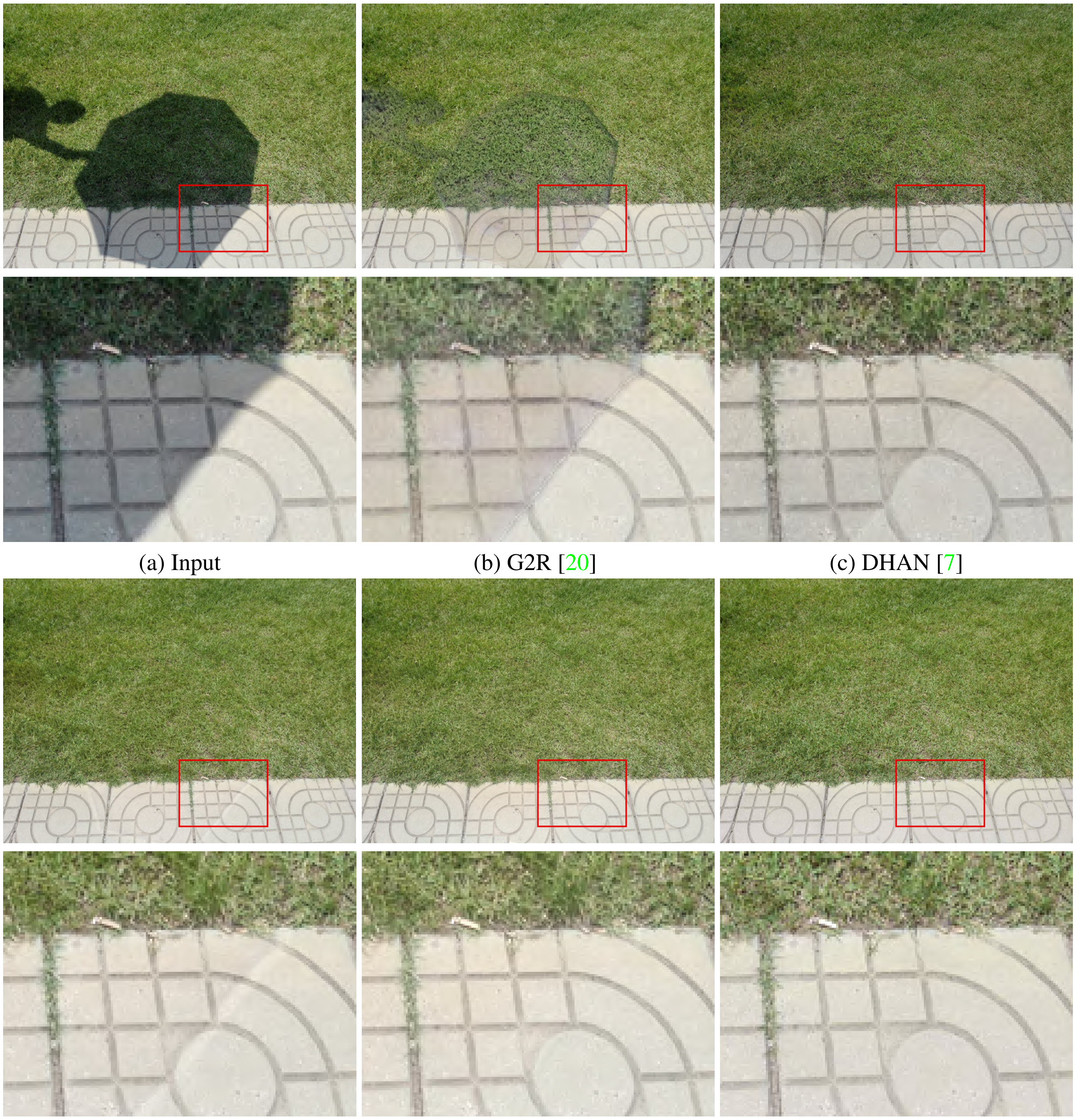
(a) Input (b) DSC [15] (c) Fu *et al.* [10] (d) DC-ShadowNet [18]



(e) Zhu *et al.* [40] (f) BMNet [39] (g) Ours (h)GT

Figure A. One example of shadow removal results on the SRD [22] dataset. The input shadow image (a), the estimated results of DSC [18] (b), Fu *et al.* [10] (c), DC-ShadowNet [18] (d), Zhu *et al.* [40] (e), BMNet [39] (f), Ours (g), and the ground truth (h), respectively, as well as their corresponding zoom-in regions. Please zoom in to see the details.

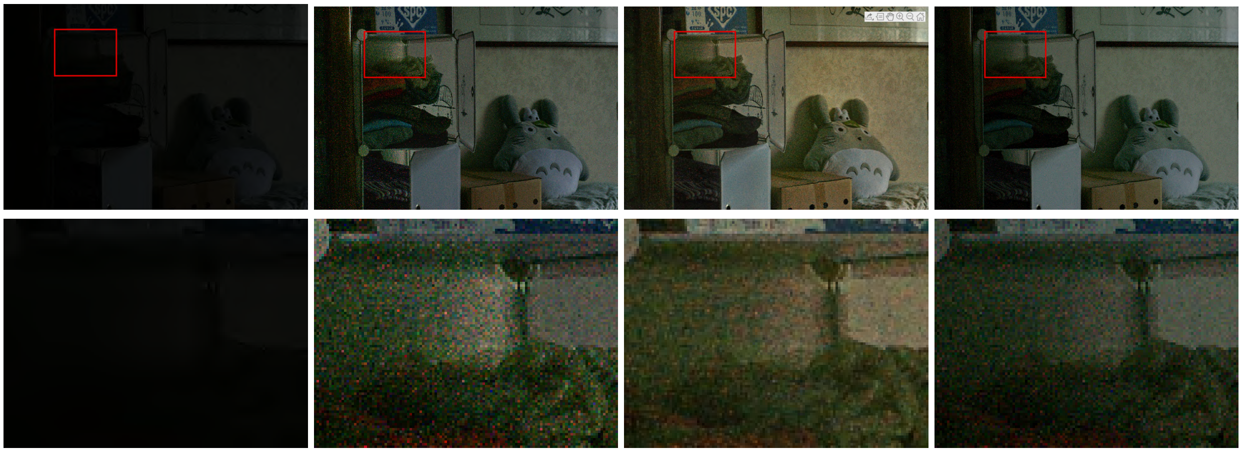
Figure B. One example of shadow removal results on the ISTD+ [19] dataset. The input shadow image (a), the estimated results of DSC [18] (b), G2R [20] (c), Le *et al.* [19] (d), Fu *et al.* [10] (e), BMNet [39] (f), Ours (g), and the ground truth (h), respectively, as well as their corresponding zoom-in regions. Please zoom in to see the details.



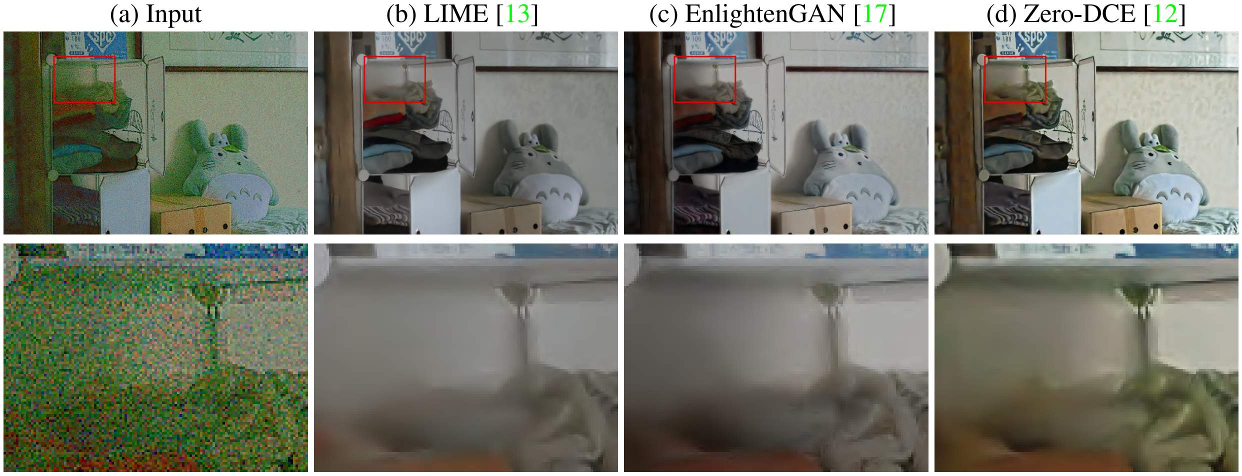
(a) Input (b) G2R [20] (c) DHAN [7]

(d) BMNet [39] (e) Ours (f) GT

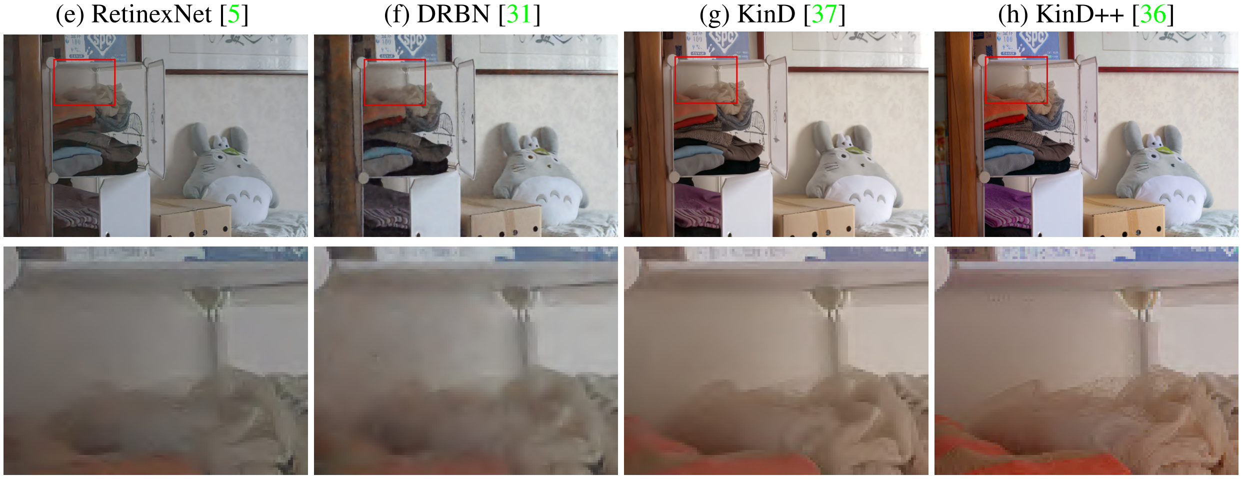
Figure C. One example of shadow removal results on the original resolution of ISTD+ [19] dataset. The input shadow image (a), the estimated results of G2R [20] (b), DHAN [7] (c), BMNet [39] (d), Ours (e), and the ground truth (f), respectively, as well as their corresponding zoom-in regions. Please zoom in to see the details.



(a) Input (b) LIME [13] (c) EnlightenGAN [17] (d) Zero-DCE [12]



(e) RetinexNet [5] (f) DRBN [31] (g) KinD [37] (h) KinD++ [36]



(i) URetinex-Net [30] (j) MIRNet [33] (k) Ours (l) GT

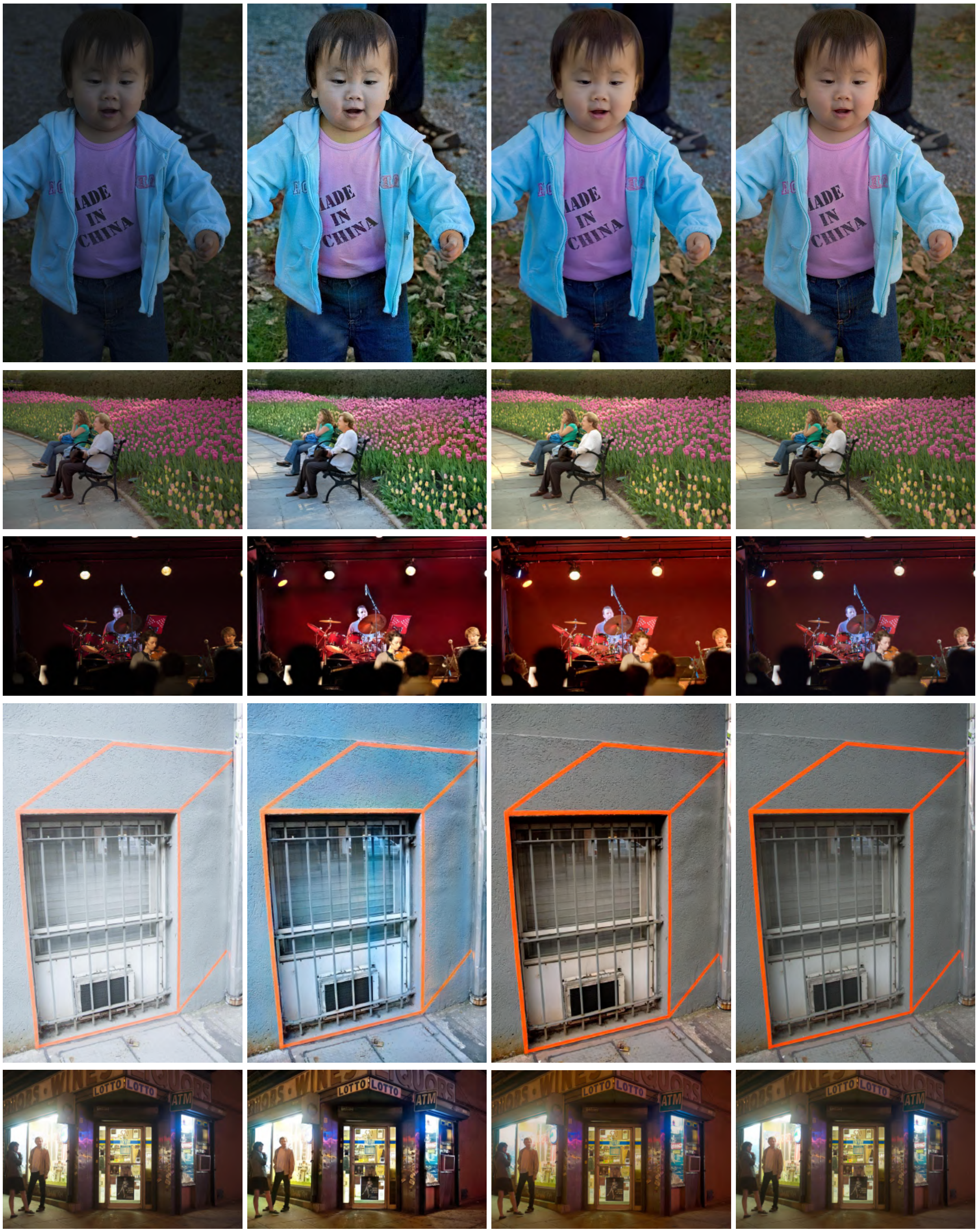
Figure D. One example of low-light enhancement results on the LOL [5] dataset. The input low-light image (a), the estimated results of LIME [13] (b), EnlightenGAN [17] (c), Zero-DCE [12] (d), RetinexNet [5] (e), DRBN [31] (f), KinD [37] (g), KinD++ [36] (h), URetinex-Net [30] (l), MIRNet [33] (i), Ours (j), and the ground truth (h), respectively, as well as their corresponding zoom-in regions. Please zoom in to see the details.

(a) Input (b) LIME [13] (c) EnlightenGAN [17] (d) Zero-DCE [12]

(e)RetinexNet [5] (f) DRBN [31] (g) KinD [37] (h) KinD++ [36]

(i) URetinex-Net [30] (j) MIRNet [33] (k) Ours (l) GT

Figure E. One example of low-light enhancement results on the LOL [5] dataset. The input low-light image (a), the estimated results of LIME [13] (b), EnlightenGAN [17] (c), Zero-DCE [12] (d), RetinexNet [5] (e), DRBN [31] (f), KinD [37] (g), KinD++ [36] (h), URetinex-Net [30] (l), MIRNet [33] (i), Ours (j), and the ground truth (h), respectively, as well as their corresponding zoom-in regions. Please zoom in to see the details.



(a) Input (b) Afifi *et al.* [2] (c) Ours (d) Ref. Img

Figure F. Visual examples of exposure correction results on [2] dataset. The input over/under-exposed image (a), the estimated results of Afifi *et al.* [2] (b), Ours (c), and the reference standard-exposed image, including three examples of under-exposure (rows 1-3) and two examples of over-exposure (rows 4-5). Please zoom in to see the details.

References

[1] Mohammad Abdullah-Al-Wadud, Md Hasanul Kabir, M Ali Akber Dewan, and Oksam Chae. A dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics*, 53(2):593–600, 2007. 2, 3

[2] Mahmoud Afifi, Konstantinos G Derpanis, Bjorn Ommer, and Michael S Brown. Learning multi-scale photo exposure correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9157–9167, 2021. 2, 3, 9

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

[4] Stanley H Chan, Xiran Wang, and Omar A Elgendy. Plug-and-play admm for image restoration: Fixed-point convergence and applications. *IEEE Transactions on Computational Imaging*, 3(1):84–98, 2016. 1

[5] Wei Chen, Wang Wenjing, Yang Wenhan, and Liu Jiaying. Deep retinex decomposition for low-light enhancement. In *British Machine Vision Conference*. British Machine Vision Association, 2018. 2, 3, 7, 8

[6] Yu-Sheng Chen, Yu-Ching Wang, Man-Hsin Kao, and Yung-Yu Chuang. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6306–6314, 2018. 2, 3

[7] Xiaodong Cun, Chi-Man Pun, and Cheng Shi. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan. In *AAAI*, pages 10680–10687, 2020. 6

[8] Lisa DaNae Dayley and Brad Dayley. *Photoshop CS5 Bible*. John Wiley & Sons, 2010. 2

[9] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał K Mantiuk, and Jonas Unger. Hdr image reconstruction from a single exposure using deep cnns. *ACM transactions on graphics (TOG)*, 36(6):1–15, 2017. 2, 3

[10] Lan Fu, Changqing Zhou, Qing Guo, Felix Juefei-Xu, Hongkai Yu, Wei Feng, Yang Liu, and Song Wang. Auto-exposure fusion for single-image shadow removal. In *CVPR*, pages 10571–10580, 2021. 4, 5

[11] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2782–2790, 2016. 2, 3

[12] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1780–1789, 2020. 2, 3, 7, 8

[13] Xiaojie Guo, Yu Li, and Haibin Ling. LIME: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2016. 2, 3, 7, 8

[14] Ruizhi Hou and Fang Li. Idpcnn: Iterative denoising and projecting cnn for mri reconstruction. *Journal of Computational and Applied Mathematics*, 406:113973, 2022. 1

[15] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection and removal. 42(11):2795–2808, 2020. 4, 5

[16] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Dslr-quality photos on mobile devices with deep convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3277–3285, 2017. 2, 3

[17] 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 Transactions on Image Processing*, 30:2340–2349, 2021. 2, 7, 8

[18] Yeying Jin, Aashish Sharma, and Robby T Tan. Dc-shadownet: Single-image hard and soft shadow removal using unsupervised domain-classifier guided network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5027–5036, 2021. 4, 5

[19] Hieu Le and Dimitris Samaras. Shadow removal via shadow image decomposition. In *ICCV*, pages 8578–8587, 2019. 1, 5, 6

[20] Zhihao Liu, Hui Yin, Xinyi Wu, Zhenyao Wu, Yang Mi, and Song Wang. From shadow generation to shadow removal. In *CVPR*, 2021. 5, 6

[21] Feifan Lv, Yu Li, and Feng Lu. Attention guided low-light image enhancement with a large scale low-light simulation dataset. *International Journal of Computer Vision*, 129(7):2175–2193, 2021. 2

[22] Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, and Rynson WH Lau. Deshadownet: A multi-context embedding deep network for shadow removal. In *CVPR*, pages 4067–4075, 2017. 1, 4

[23] Xutong Ren, Wenhan Yang, Wen-Huang Cheng, and Jiaying Liu. Lr3m: Robust low-light enhancement via low-rank regularized retinex model. *IEEE Transactions on Image Processing*, 29:5862–5876, 2020. 2

[24] Ali M Reza. 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*, 38(1):35–44, 2004. 2, 3

[25] Liu Risheng, Ma Long, Zhang Jiaao, Fan Xin, and Luo Zhongxuan. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 2

[26] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 1

[27] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. 1

[28] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6849–6857, 2019. 2, 3

[29] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 2

[30] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5901–5910, 2022. 2, 7, 8

[31] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 2, 7, 8

[32] Xin Yang, Ke Xu, Yibing Song, Qiang Zhang, Xiaopeng Wei, and Rynson WH Lau. Image correction via deep reciprocating hdr transformation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1798–1807, 2018. 2

[33] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *ECCV*, 2020. 2, 7, 8

[34] Qing Zhang, Ganzhao Yuan, Chunxia Xiao, Lei Zhu, and Wei-Shi Zheng. High-quality exposure correction of underexposed photos. In *Proceedings of the 26th AC international conference on Multimedia*, pages 582–590, 2018. 2, 3

[35] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 2

[36] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129(4):1013–1037, 2021. 2, 7, 8

[37] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 1632–1640, 2019. 2, 7, 8

[38] Lin Zhao, Shao-Ping Lu, Tao Chen, Zhenglu Yang, and Ariel Shamir. Deep symmetric network for underexposed image enhancement with recurrent attentional learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12075–12084, October 2021. 2

[39] Yurui Zhu, Jie Huang, Xueyang Fu, Feng Zhao, Qibin Sun, and Zheng-Jun Zha. Bijective mapping network for shadow removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5627–5636, 2022. 4, 5, 6

[40] Yurui Zhu, Zeyu Xiao, Yanchi Fang, Xueyang Fu, Zhiwei Xiong, and Zheng-Jun Zha. Efficient model-driven network for shadow removal. 2022. 4