EnlightenGAN: Deep Light Enhancement without Paired Supervision
Supplementary Material

December 12, 2022

1 Data Collection

We collect an unpaired low/normal-light dataset from [7, 5, 1, 2], which contains 914 low-light and 1016 normal-light images respectively. Specifically, the low-light image set consists of 274 images from the LOL dataset [7], 566 lowest exposure photos from [1], and 74 lowest exposure photos from [5]. The normal-light image set includes 690 images from the RAISE dataset [2] and 326 images from the LOL dataset [7]. It is noteworthy that based on this setting there is no similar scene appears in the two part of our proposed dataset.

2 Detailed Network Structure

Our network consists of an attention-based U-Net (AU-Net) as the generator and global-local discriminators. In the generator, we use a *ConvBlock* to denote the convolutional-LeakyReLu-BatchNorm block. In the discriminator, each convolutional layer is followed by a LeakyReLu but we do not adopt batch normalization layer here. The detailed structures of the generator and discriminators are shown in Table. 1 and Table. 2. Furthermore, we also present the generation process of attention map, as shown in the Fig. 1.

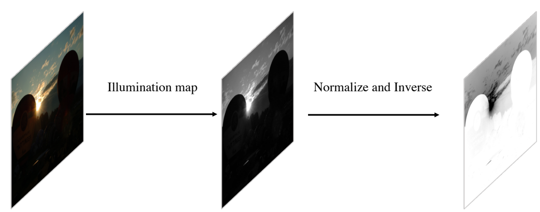

Figure 1: Detailed generation process of attention map.

Table 1: Generator Configuration.

Encoder				
Layer	Kernel	Stride	C_{in}	C_{out}
ConvBlock	3×3	1	4	32
ConvBlock	3×3	1	32	32
MaxPooling	2×2	2	-	-
ConvBlock	3×3	1	32	64
ConvBlock	3×3	1	64	64
MaxPooling	2×2	2	-	-
ConvBlock	3×3	1	64	128
ConvBlock	3×3	1	128	128
MaxPooling	2×2	2	-	-
ConvBlock	3×3	1	128	256
ConvBlock	3×3	1	256	256
MaxPooling	2×2	2	-	-
ConvBlock	3×3	1	256	512

Decoder				
Layer	Kernel	Stride	C_{in}	C_{out}
ConvBlock	3×3	1	512	512
Upsampling	3×3	2	-	-
ConvBlock	3×3	1	512	256
ConvBlock	3×3	1	256	256
Upsampling	3×3	2	-	-
ConvBlock	3×3	1	256	128
ConvBlock	3×3	1	128	128
Upsampling	3×3	2	-	-
ConvBlock	3×3	1	128	64
ConvBlock	3×3	1	64	64
Upsampling	3×3	2	-	-
ConvBlock	3×3	1	64	32
ConvBlock	3×3	1	32	32
Conv	3×3	1	32	3

Table 2: Discriminators Configuration.

Global Discriminator				
Layer	Kernel	Stride	C_{in}	C_{out}
Conv	4×4	2	3	64
Conv	4×4	2	64	128
Conv	4×4	2	128	256
Conv	4×4	2	256	512
Conv	4×4	2	512	512
Conv	4×4	1	512	512
Conv	4×4	1	512	1

Local Discriminator				
Layer	Kernel	Stride	C_{in}	C_{out}
Conv	4×4	2	3	64
Conv	4×4	2	64	128
Conv	4×4	2	128	256
Conv	4×4	2	256	512
Conv	4×4	1	512	512
Conv	4×4	1	512	1

3 Human Subjective Evaluation

We recruit 9 subjects from different age groups, genders, cultural and educational backgrounds. Despite the relatively small number of raters, we observed small interjudge variances among raters on the same pairs comparison results. The average ranking of each images and each methods are shown in the Table. 3

4 Visual Results

We compare our method with several recent competing methods: a vanilla CycleGAN [8] trained using our unpaired training set, RetinexNet [7], SRIE [3], LIME [4], and NPE [6], and show more visual results to demonstrate the effectiveness of our method, including the results of ablation study, comparison and real-world image adaptation. The setting follows what we mentioned in the paper (Sec. 4.2, Sec. 4.3.1, and Sec. 4.4). Images are shown in Fig. 2 to Fig. 6 (Better viewed in electronic version).

Table 3: Detail scores of human subjective evaluation. Red is the best and blue is the second best results.

Image Id	LIME	NPE	RetinexNet	SRIE	EnlightenGAN
1	4	3	5	2	1
2	4	3	5	1	2
3	4	3	5	1	2
3	3	4	5	1	2
4	4	3	5	2	1
5	4	1	5	2	3
7	5	2	4	3	1
8	1	4	5	2	3
9	5	2	4	3	1
10	4	2	5	3	1
11	4	2	5	3	1
12	3	4	5	2	1
13	1	2	5	4	3
13	4	1	5	2	3
14	4	1	5	3	2
15	5	3	4	1	2
17	5	3	4	1	2
18	5	2	4	1	3
19	4	3	5	2	1
20	3	4	5	1	2
21	5	3	4	1	2
22	5	3	4	2	1
23	4	3	5	2	1

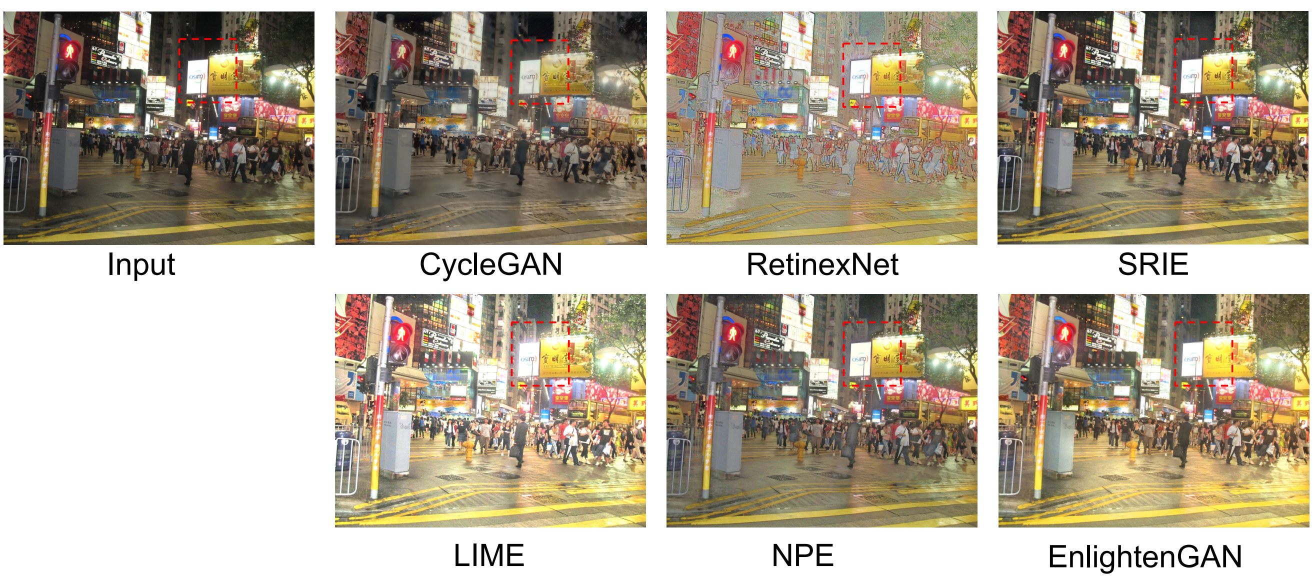

Figure 2: Visual results compared with current state-of-the-art methods.

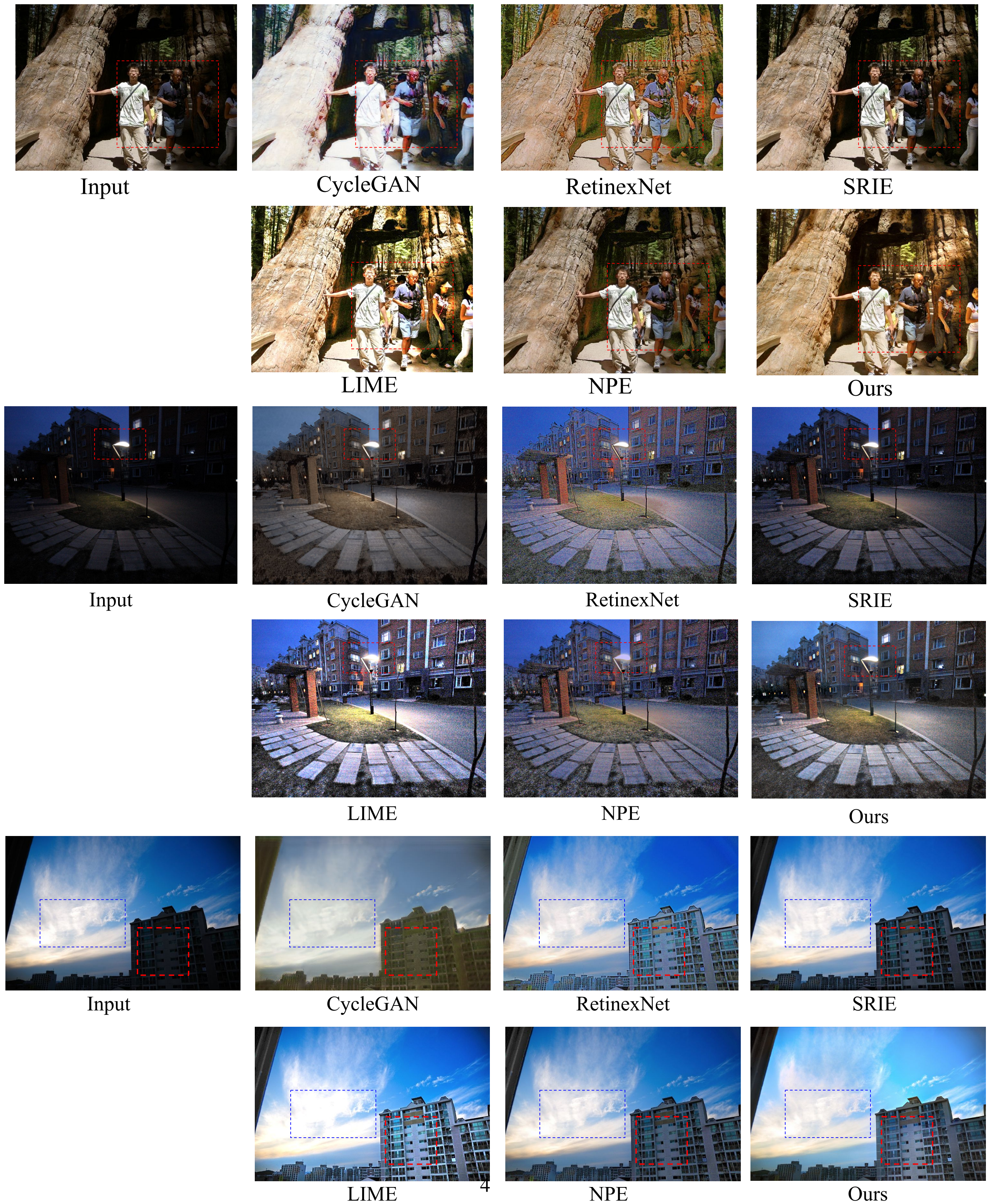

Figure 3: Visual results compared with current state-of-the-art methods.

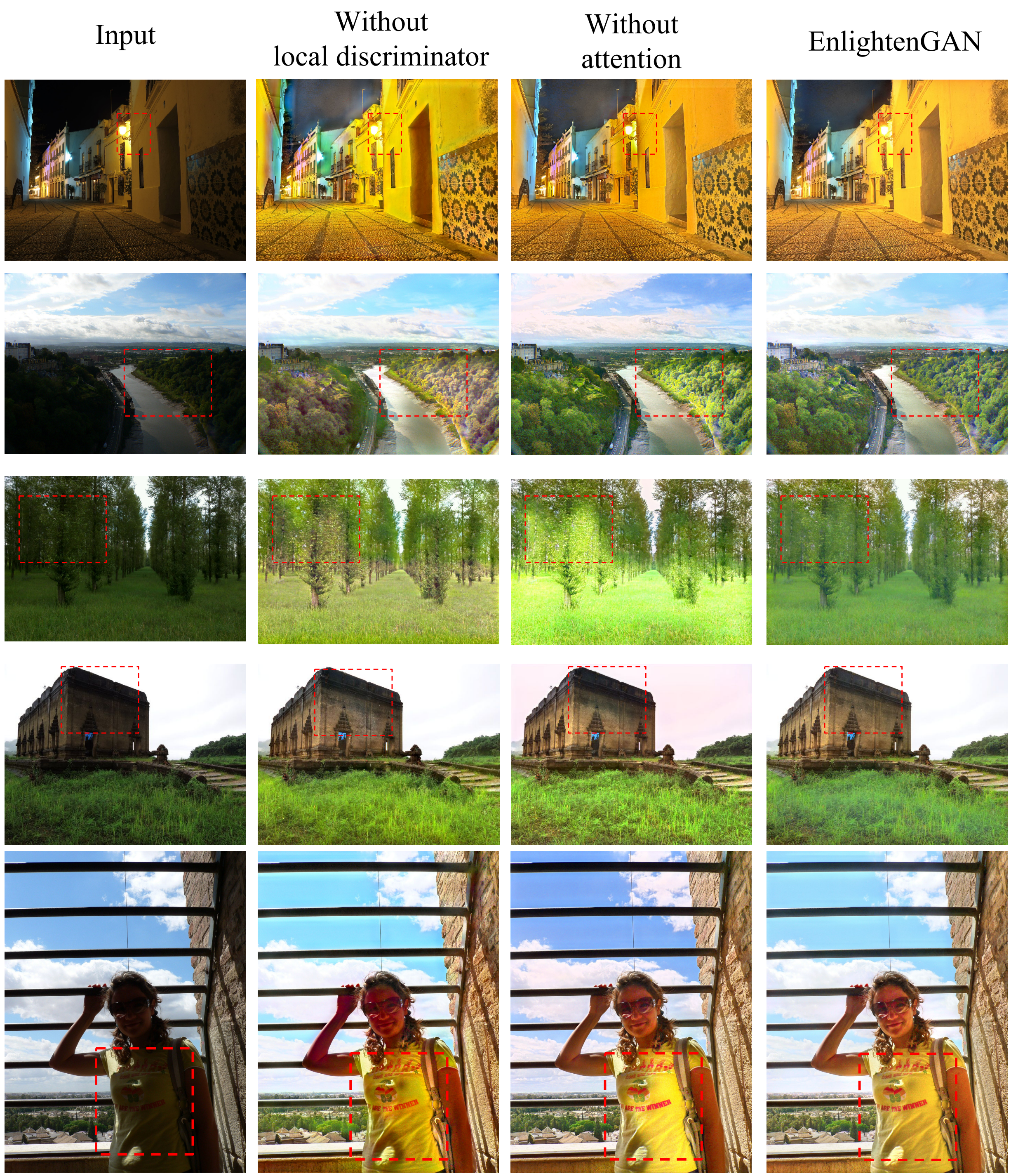

Figure 4: Ablation study.

Figure 5: Results on BDD-100k dataset.

Figure 6: Results on BDD-100k dataset.

References

[1] J. Cai, S. Gu, and L. Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018. 1

[2] D.-T. Dang-Nguyen, C. Pasquini, V. Conotter, and G. Boato. Raise: a raw images dataset for digital image forensics. In *Proceedings of the 6th ACM Multimedia Systems Conference*, pages 219–224. ACM, 2015. 1

[3] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *CVPR*, pages 2782–2790, 2016. 2

[4] X. Guo, Y. Li, and H. Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2017. 2

[5] N. K. Kalantari and R. Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *ACM Trans. Graph*, 36(4):144, 2017. 1

[6] S. Wang, J. Zheng, H.-M. Hu, and B. Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9):3538–3548, 2013. 2

[7] C. Wei, W. Wang, W. Yang, and J. Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 1, 2

[8] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, pages 2223–2232, 2017. 2