Unsupervised Low-light Image Enhancement via Spectral Consistency Supplementary Materials

Bing Li[1 [0009−0005−8863−5450]], Wei Yu[1 [0009−0003−7046−9832]], Naishan Zheng[1 [0000−0002−7451−8780]], Jie Huang[1 [0000−0002−3518−3404]], and Feng Zhao[1 [0000−0001−6767−8105]]

University of Science and Technology of China, Hefei 230027, China
{bing0123, patrick914y, nszheng, hj0117}@mail.ustc.edu.cn
fzhao956@ustc.edu.cn

1 More details on Reflectance Equivariance

In this section, we give more details on the *Reflectance Equivariance* of low-light images to enhance comprehension of this inherent consistency characteristic. We first derive the expression of Reflectance Equivariance based on the Illumination Invariance and Retinex theory. Subsequently, we present an in-depth investigation of the impact of reflectance equivariance on pre-trained networks, which is mentioned in paper Sec.3.2.2, shedding light on its practical implications.

1.1 The derivation of Reflectance Equivariance

The Retinex theory assumes that an image I can be decomposed into the element-wise multiplication of the corresponding illumination component $l(I)$ and reflectance component $r(I)$. This decomposition can be denoted as $I = l(I) \cdot r(I)$, where $l(\cdot)$ and $r(\cdot)$ are illumination and reflectance estimator, respectively, and $\cdot$ is element-wise multiplication. Additionally, the Retinex theory indicates that different color spectra share the same illumination map. Therefore, the illumination remains unchanged if we apply the perturbation on the estimated results, *i.e.*, $t \circ l(I) = l(I)$. Note that $\cdot$ is the element-wise multiplication; we have:

$$t(I) = t \circ (l(I) \cdot r(I)) = (t \circ l(I)) \cdot (t \circ r(I)) = l(I) \cdot (t \circ r(I)). \tag{1}$$

The Illumination Invariance, shown in paper Sec.3.2.1, implies that the illumination map is an invariant attribute of low-light images under spectral shuffle perturbation $t(\cdot)$, which means $l \circ t(I) = l(I)$. Considering that $t(I)$ is the image perturbed by spectral shuffle, it also obeys the Retinex theory, which means:

$$t(I) = (l \circ t(I)) \cdot (r \circ t(I)) = l(I) \cdot (r \circ t(I)). \tag{2}$$

Based on Eq. 1 and Eq. 2, we have:

$$t \circ r(I) = t(I)/l(I) = r \circ t(I), \tag{3}$$

where / is element-wise division. Eq. 3 is what we refer to as Reflectance Equivariance in Sec.3.2.2 of the paper.

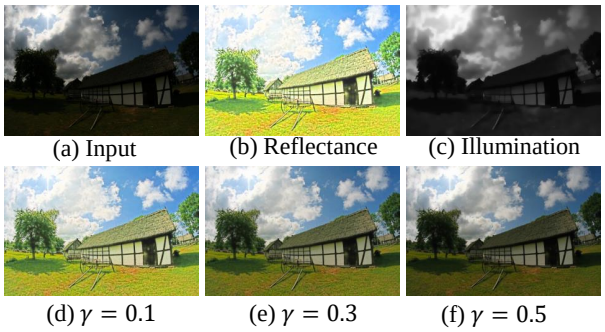

Fig. 1. Visualization of the Retinex decomposition. The enhanced results are obtained using various correction factors.

1.2 The investigation on pre-trained LLIE methods

As mentioned in paper Sec.3.2.2, we directly apply the perturbed input to pre-trained enhancement networks to assess the potential color shift caused by the spectral shuffle. We select two well-known methods [1, 8] as backbones, both well-trained on the LOL dataset [7] without spectral shuffle perturbations. We directly enhance the perturbed input of the LOL dataset without fine-tuning, and the results are shown in Fig. 3. Interestingly, there is no severe color shift of all five other spectral order inputs, which indicates spectral shuffle perturbation is a mild augmentation for low-light image enhancement, leading to minimal influence on color preservation.

Table 1. Quantitative results of ablation studies on the LOL dataset. The best results are marked in bold.

Setting	PSNR↑	SSIM↑
O	17.80	0.6299
A	18.44	0.6945
B	18.85	0.7579
C	19.20	0.7618
D	19.31	0.7633
baseline	**19.72**	**0.7776**

2 The architecture of networks

Our aim is to establish an unsupervised training framework for low-light image enhancement. Therefore, we utilize two simple networks, LNet and RNet, to estimate illumination and reflectance, respectively. The LNet consists of five convolutional layers, with the first four layers utilizing the ReLU activation function,

while the last layer employs the sigmoid activation function. The RNet is similar to LNet, except that RNet has two additional convolutional layers, followed by the ReLU activation function, placed in the middle. The RNet is slightly more complex than the LNet, as the RNet predicts more information. According to the Retinex theory, the three color spectra share the same illumination map. Consequently, we set the output channel of LNet to 1 while that of RNet to 3.

3 Decomposition results

We visualize the decomposed reflectance and illumination results in Fig. 1. The reflectance component exhibits rich textures and intricate details, while the illumination component appears smooth and devoid of texture. We further present the enhancement results obtained with various correction factors. It can be observed that as the value of correction factor γ increases, the overall brightness of the image gradually decreases. Users can adjust the parameter γ according to their preferences.

4 Additional ablation studies

In this section, we conducted additional ablation experiments in order to demonstrate the rationality of the cross multi-scale denoise constraint. We rewrite the constraint as follows:

$$\mathcal{L}_D = \alpha \sum_{s \in \{2^1, 2^2, \cdots, 2^k\}} ||t \circ r(I_i) \downarrow^s - r(I_p) \downarrow^s ||_2^2 + \beta \sum_{j \in \{i,p\}} ||\nabla r(I_j)||_1. \quad (4)$$

If we set $\alpha = 1, \beta = 0.2, k = 2$, it is aligned with the configuration described in the paper. We list our additional ablation configurations as follows: We set A) $\alpha = 1, \beta = 0, k = 2$; B) $\alpha = 0, \beta = 0.2, k = 2$; C) $\alpha = 1, \beta = 0.2, k = 1$; D) $\alpha = 1, \beta = 0.2, k = 3$ to validate the effectiveness of smoothness part (A), multi-scale constraint part (B) and the scale of asymmetrical sub-samplers (C and D). The visual comparisons are presented in Fig. 2, and the quantitative results are shown in Tab. 1. We further present the results of the full framework (denoted as baseline) and the framework without denoise constraint (denoted as O) to assist comparisons. Based on the above results, we can draw the following conclusions: 1) Both the multi-scale constraint part (A outperforms O) and the smoothness part (B outperforms O) exhibit the capability to perform denoising independently; 2) Better results can be achieved by jointly using the multi-scale constraint and smoothness parts (C, D, and baseline outperform A and B); 3) We empirically find that the optimal value for K is 2. We argue that excessively small values of the k may reduce the denoising ability, while large values may lead to over-smoothed outcomes.

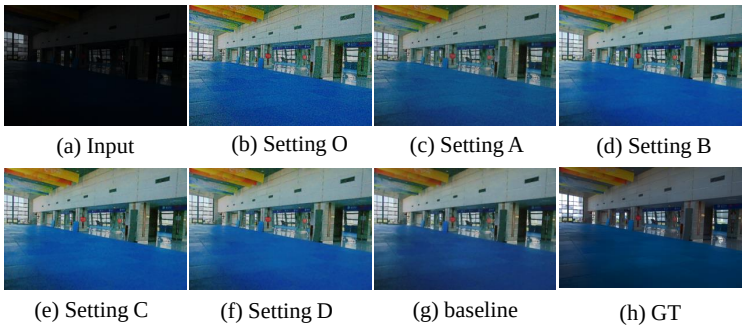

Fig. 2. Visual results of ablation studies on the LOL dataset.

5 Additional results

In this section, we provide more qualitative results. Fig. 4, 5, and 6 are visual results on the LOL and LSRW [3] datasets. Fig. 7 and Fig. 8 are visual results on the DICM [4], NPE [6], MEF [5], and LIME [2] datasets. We can observe that other methods tend to exhibit color shifts or artifacts, while our methods can achieve consistently satisfactory visual results. We also show the enhanced results with only a single low-light image training in Fig. 9. The result indicates that our method is general enough to be applied effectively even when only a single low-light image is accessible for training. What's more, we present the NIQE results on no-reference datasets in Tab. 2. Overall, our method consistently produces the highest-quality visual results compared to other methods.

Table 2. NIQE scores on the no-reference datasets, including MEF, LIME, NPE, and DICM. Smaller values indicate more perceptually appealing results. The best and the second results are marked in bold and underlined, respectively.

Dataset	MEF	LIME	NPE	DICM	All
Input	4.265	4.438	4.319	4.255	4.280
RetinexNet	4.215	5.109	4.283	4.096	4.234
EnGAN	**3.232**	**3.719**	<u>4.113</u>	<u>3.570</u>	<u>3.571</u>
ZeroDCE	<u>3.467</u>	4.130	4.233	3.652	3.715
SCI	3.619	4.097	4.166	3.902	3.894
PairLIE	4.302	4.560	4.215	4.239	4.280
Ours	3.497	<u>3.810</u>	**4.065**	**3.487**	**3.568**

References

1. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3291–3300 (2018)

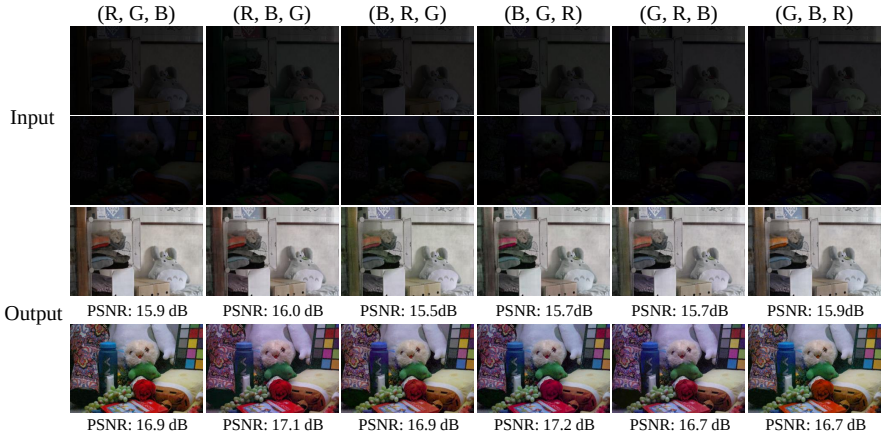

Fig. 3. Quantitative and qualitative results on the LOL dataset. The third and fourth rows are the output corresponding to the input of the first and second rows by SID and Restormer, respectively. We rearrange the spectra of outputs in the RGB order for better visualization.

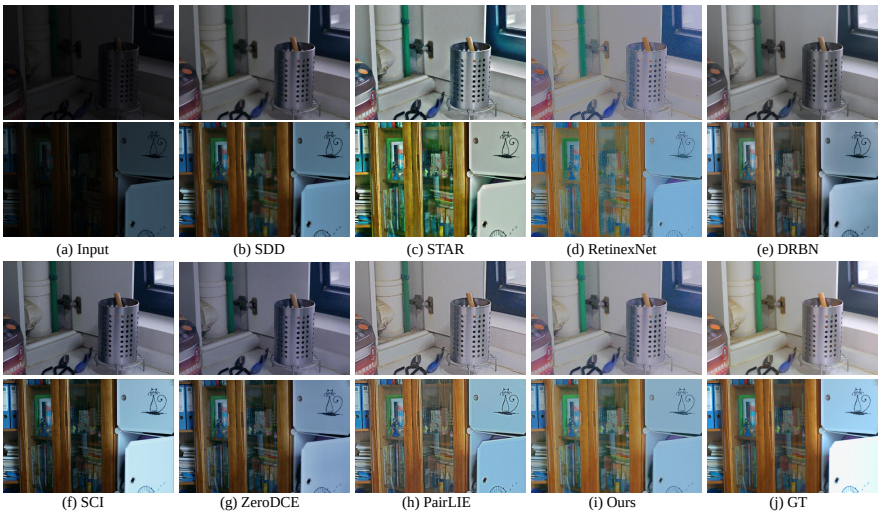

Fig. 4. Visual comparison of different LLIE methods on the LOL dataset.

Fig. 5. Visual comparison of different LLIE methods on the LOL dataset.

Fig. 6. Visual comparison of different LLIE methods on the LSRW dataset.

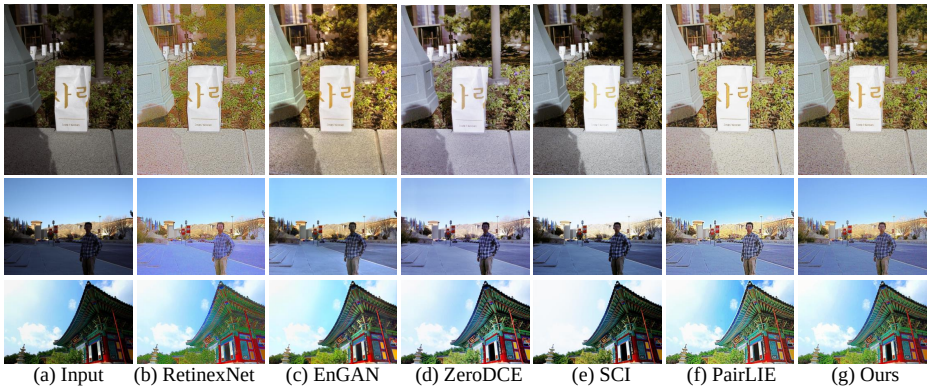

Fig. 7. Visual comparison of different LLIE methods on the DICM dataset.

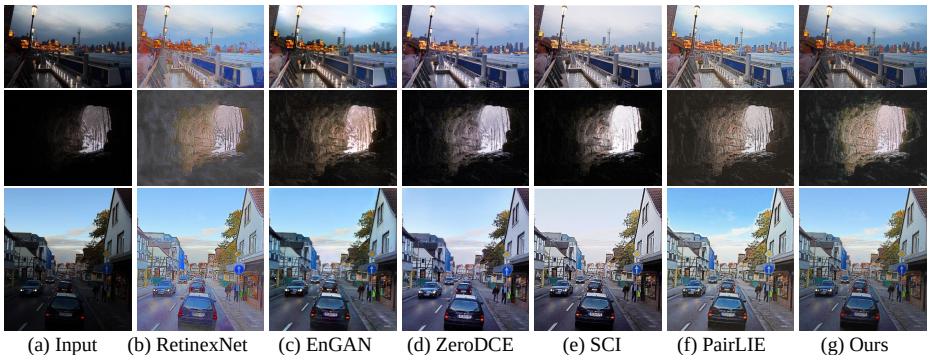

Fig. 8. Visual comparison of different LLIE methods on unpaired datasets, where the first row is the results on the NPE dataset, the second row is the results on the MEF dataset, and the third row is the results on the LIME dataset. Please zoom in for details

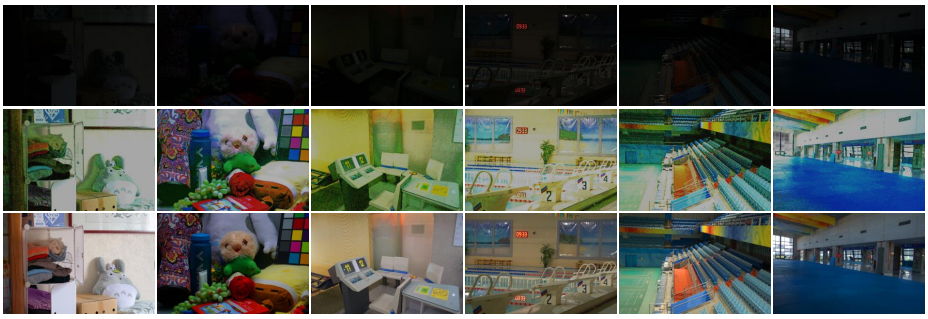

Fig. 9. The enhanced results with only a single low-light image training. The first, second, and third rows are input images, enhanced results, and reference images, respectively.

2. Guo, X., Li, Y., Ling, H.: Lime: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing **26**(2), 982–993 (2016)
3. Hai, J., Xuan, Z., Yang, R., Hao, Y., Zou, F., Lin, F., Han, S.: R2rnet: Low-light image enhancement via real-low to real-normal network. Journal of Visual Communication and Image Representation **90**, 103712 (2023)
4. Lee, C., Lee, C., Kim, C.S.: Contrast enhancement based on layered difference representation. In: 2012 19th IEEE International Conference on Image Processing. pp. 965–968. IEEE (2012)
5. Ma, K., Zeng, K., Wang, Z.: Perceptual quality assessment for multi-exposure image fusion. IEEE Transactions on Image Processing **24**(11), 3345–3356 (2015)
6. Wang, S., Zheng, J., Hu, H.M., Li, B.: Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Transactions on Image Processing **22**(9), 3538–3548 (2013)
7. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560 (2018)
8. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5728–5739 (2022)