Lighting up NeRF via Unsupervised Decomposition and Enhancement: Supplementary Material

Anonymous ICCV submission

Paper ID 5161

In this supplementary material, we present more evaluation results of our proposed method for visual low-light enhancement, comparing it to existing state-of-the-art techniques. We also provide more analysis of the implementation details. For a more comprehensive assessment, we refer the reader to the supplementary videos, which showcase comparisons of the rendering results of our method, as well as those of the baseline models and NeRF.

1. Visual Comparisons

We show more comparison results of our methods and the existing SOTA low-light enhancement methods (LLFlow [9], ZeroDCE [3], SNR [11], URetinex [10], SCI [4]), and also with the denoised results of the best one among these methods (NAFNet [2] + URetinex) in Fig. 4. Our experimental results demonstrate superior performance in terms of image quality, as illustrated in Fig. 4. Specifically, our approach achieves a notable reduction in noise, while preserving vivid colors and fine details.

Furthermore, we provide additional comparisons between our proposed method and the baseline approach, which combines LLE with NeRF, as depicted in the **supplementary videos**. Specifically, we adopt URetinex-Net [10] as the LLE method for the baseline approach, and we also explore other LLE methods in combination with NeRF for a more thorough analysis. Our supplementary videos present the results of these additional comparisons, further showing the superior visual quality of our method when compared to all other LLE-NeRF combinations.

2. Implementation Details

We implement our model using Jax [1]. In training, the resolution of images is four times downsampled from the original full resolution to reduce computation. To suppress the floater artifact in the depth map, we use the weights regularization loss as in [5] for more stable convergence. In the color loss, we use the channel-pair-wise difference to implement dynamic variance for the larger color penalty. See our supplemental for more implementation details.

Our dataset contains 15 scenes with 371 images (for reference, the RawNeRF dataset has 14 scenes, and the LLFF dataset has 8 scenes). Our dataset includes both outdoor and indoor scenes, *e.g.*, road, office, campus, and lab. Image resolution is 1156×868.

2.1. Network Structure of ϕ

Our proposed method involves the learning of two types of coefficients, namely α and γ, which are utilized to enhance the $\mathbf{v}$ along the sampled rays. To achieve this, we employ the neural network ϕ, which consists of two branches, as depicted in Figure 1.

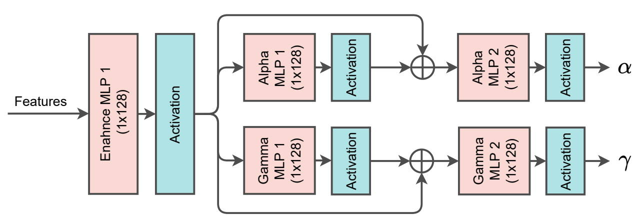

Figure 1. The architecture of our proposed neural network ϕ. The network is composed of two branches, through which the coefficients α and γ are learned. Notably, this network is lightweight (only 5×128) and our approach achieves superior performance due to our effective decomposition strategy and priors, rather than relying solely on the network.

2.2. Loss functions

Besides data loss, we use color loss and smooth loss to constrain the learning of the enhanced $\mathbf{v}$. Our color loss is:

$$L_c = \mathbb{E}[(\hat{\mathbf{c}}_r - e)^2] + \lambda_1 \mathbb{E}\left[\frac{\mathrm{var}_c(\hat{\mathbf{c}}_r)}{\beta_1 + \mathrm{var}_c(\mathbf{r}_r)}\right] + \lambda_2 ||\gamma||_2, \quad (1)$$

where the second term in the loss function serves to constrain the channel-wise variance of the predicted pixel colors and the integrated intrinsic color. Specifically, in our implementation, we utilize the channel-wise squared difference of $\hat{\mathbf{c}}_r$, denoted as $d(\hat{\mathbf{c}}_r)$, to replace the variance term

$\mathrm{var}_c(\hat{\mathbf{c}}_r)$, as follows:

$$d(\hat{\mathbf{c}}_r) = \sum_{i=1}^{3} \mathbf{d}_i, \text{where } \mathbf{d} = \left(\begin{bmatrix} \hat{\mathbf{c}}_{r1} \\ \hat{\mathbf{c}}_{r2} \\ \hat{\mathbf{c}}_{r3} \end{bmatrix} - \begin{bmatrix} \hat{\mathbf{c}}_{r3} \\ \hat{\mathbf{c}}_{r1} \\ \hat{\mathbf{c}}_{r2} \end{bmatrix} \right)^2. \quad (2)$$

Mathematically, Eq. 2 and the channel-wise variance are linearly dependent. This is to further amplify the penalty of the distorted color and correct the biased results introduced by the non-zero noises. The experiments also show that the model trained under this constraint can generate better visual results.

2.3. Neighboring Sampling

During the optimization of our proposed model, we adopt a sampling strategy wherein we randomly select a set of neighboring rays, rather than individual rays from the training pixels. This is motivated by the need to constrain local smoothness in our loss functions. The details of our sampling strategy and a comparison with the sampling process utilized by NeRF are illustrated in Figure 2.

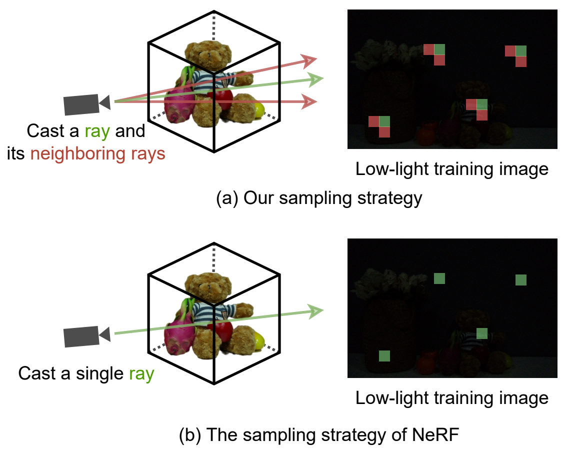

Figure 2. The illustration of the sampling strategy of our model (a) and NeRF (b). The pixels are denoted as small blocks in green (rays) or red (neighboring rays).

3. Limitations

While our proposed method demonstrates notable performance improvements over existing techniques, it is not without limitations. Firstly, our approach is scene-dependent, as it is based on the NeRF model. To address this limitation, future research could explore the combination of our method with more generalized NeRF models, such as those proposed in [8], or accelerated NeRF techniques [6, 7]. Secondly, our method may not perform optimally when lens flare effects are present in the captured images, leading to the production of artifacts in the enhanced results, as shown in Figure 3. A possible solution to this issue is to incorporate a flare detection and removal process into our method, which could help mitigate this problem.

Figure 3. An example failure case of our proposed method. In scenarios where the light source is situated in too close proximity to the camera, internal reflections within the camera lens can introduce unwanted flare effects, leading to the production of artifacts in our enhanced results.

References

[1] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: composable transformations of Python+NumPy programs, 2018. 1

[2] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. *arXiv preprint arXiv:2204.04676*, 2022. 1, 3

[3] 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*, 2020. 1, 3

[4] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5637–5646, 2022. 1, 3

[5] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P. Srinivasan, and Jonathan T. Barron. Nerf in the dark: High dynamic range view synthesis from noisy raw images. In *CVPR*, 2022. 1

[6] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4):102:1–102:15, July 2022. 2

[7] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5459–5469, 2022. 2

[8] Qianqian Wang, Zhicheng Wang, Kyle Genova, Pratul Srinivasan, Howard Zhou, Jonathan T. Barron, Ricardo Martin-Brualla, Noah Snavely, and Thomas Funkhouser. Ibrnet: Learning multi-view image-based rendering. In *CVPR*, 2021. 2

[9] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex C Kot. Low-light image enhancement

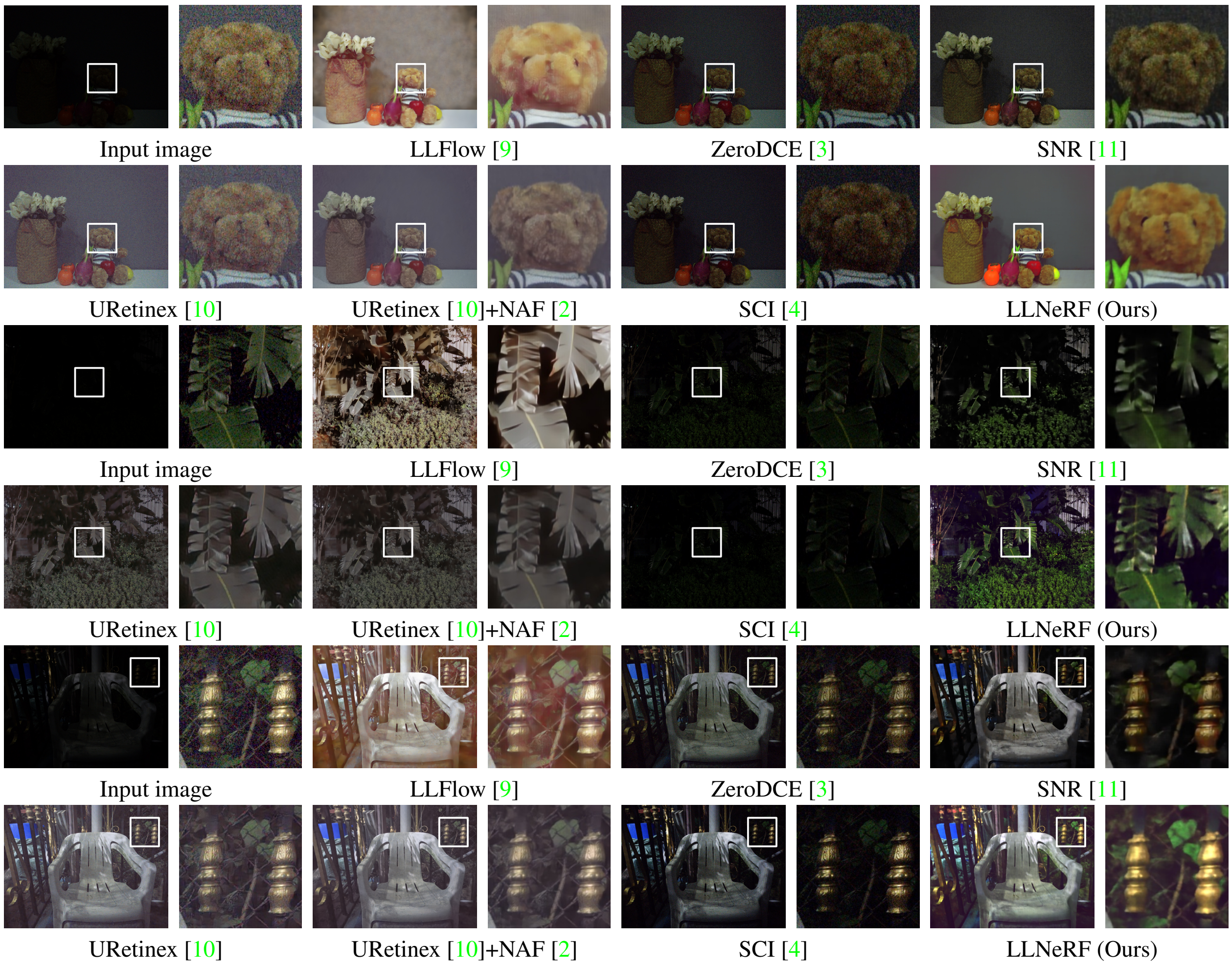

Figure 4. The visual comparison between the results of our model and the existing low-light enhancement methods. Our results have the best quality, with realistic color and fine details.

with normalizing flow. *arXiv preprint arXiv:2109.05923*, 2021. 1, 3

[10] 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 *CVPR*, 2022. 1, 3

[11] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *CVPR*, 2022. 1, 3