

Efficient Diffusion as Low Light Enhancer

Supplementary Material

1. Proof of Corollary 1

Given \tilde{x}_0 , it can be represented into the noisy states according to the forward process of diffusion models as follows:

$$\tilde{x}_0 = \frac{x_t - \sigma_t \tilde{\epsilon}}{\alpha_t}, \quad (1)$$

Substitute Equation (1) into $\tilde{x}_s = \alpha_s \tilde{x}_0 + \sigma_s \epsilon_\eta$, we get:

$$\begin{aligned} \tilde{x}_s &= \alpha_s \tilde{x}_0 + \sigma_s \epsilon_\eta \\ &= \frac{\alpha_s}{\alpha_t} x_t - \frac{\alpha_s}{\alpha_t} \sigma_t \tilde{\epsilon} + \sigma_s \epsilon_\eta(x_u, y, u), \end{aligned} \quad (2)$$

Given the teacher trajectory $x_{s,u,t}^\eta$ defined in the main paper:

$$\begin{aligned} x_{s,u,t}^\eta &= \frac{\alpha_s}{\alpha_t} x_t + \sigma_s \epsilon_\eta(x_u, y, u) \\ &\quad + \frac{\alpha_s}{\alpha_u} \sigma_u (\epsilon_\eta(x_t, y, t) - \epsilon_\eta(x_u, y, u)) \\ &\quad - \frac{\alpha_s}{\alpha_t} \sigma_t \epsilon_\eta(x_t, y, t), \end{aligned} \quad (3)$$

we can obtain the following equation:

$$\begin{aligned} \omega x_{s,u,t}^\eta + (1 - \omega) \tilde{x}_s &= \frac{\alpha_s}{\alpha_t} x_t + \sigma_s \epsilon_\eta(x_u, y, u) \\ &\quad + \omega \frac{\alpha_s}{\alpha_u} \sigma_u (\epsilon_\eta(x_t, y, t) - \epsilon_\eta(x_u, y, u)) \\ &\quad - \frac{\alpha_s}{\alpha_t} \sigma_t (\omega \epsilon_\eta(x_t, y, t) + (1 - \omega) \tilde{\epsilon}) \\ &= \tilde{x}_{s,u,t}^\eta. \end{aligned} \quad (4)$$

2. Implementation details

We train our teacher models for 2 million iterations with the framework of SR3 on the Gaussian flow [3]. In the distillation stage, and 5000 iterations is required. We employ the Adam optimizer with a learning rate of 1×10^{-4} , applying linear weight decay. To stabilize training, we utilize an exponential moving average (EMA) with a weight of 0.9999 during parameter updates in both the pretrain and distillation stages. We adopt a fine-grained diffusion method with $T = 512$ steps and implement a linear noise schedule with endpoints set at 1×10^{-4} and 2×10^{-2} . The patch size and batch size are set to 96 and 16, respectively. All the experiments are run on a single A100 GPU with 80GB of memory. Training our method with a smaller patch size and batch size on a device with less memory is feasible.

3. Dataset

LOLv1 comprises 485 pairs of low/normal-light images for training and 15 pairs for testing, taken under different exposure conditions. **LOLv2** is divided into LOLv2-real and LOLv2-synthetic subsets. LOLv2-real contains 689 pairs for training and 100 for testing, adjusted for exposure time and ISO. LOLv2-Synthetic is created by analyzing the illumination distribution of low-light images, comprising 900 pairs for training and 100 for testing.

SID. The subset of the SID dataset captured by a Sony camera is adopted for evaluation. There are 2697 short-/long-exposure RAW image pairs. The low-/normal-light RGB images are obtained by using the same in-camera signal processing of SID [1] to transfer RAW to sRGB. 2099 and 598 image pairs are used for training and testing. **SDDSD.** We adopt the static version of SDDSD, captured by a Canon EOS 6D Mark II camera with an ND filter. We use 62:6 and 116:10 low-/normal-light video pairs for training and testing on indoor subsets.

DICM, LIME, MEF, NPE, and VV. We evaluate the generalization capability of our models trained on the LOLv1 and LOLv2 datasets by testing them on the DICM, LIME, MEF, NPE, and VV datasets. Since these datasets lack ground truth images, evaluation relies on visual comparison and no-reference image quality assessment methods.

4. More experiment results

4.1. Visual comparison on SID and SDDSD dataset

Due to space constraints in the main manuscript, we only presented quantitative comparisons. Here, we offer additional visual comparisons on the SID [1] and SDDSD [4] datasets in Figure 1. These datasets pose significant challenges for low-light image enhancement due to both low light levels and severe noise degradation. Our ReDDiT method demonstrates strong performance in improving brightness and reducing noise on these challenging datasets. It can be observed in Figure 1 that our ReDDiT *effectively enhances the visibility, preserves the color, improves noise reduction, and retains finer details in normal light images.*

4.2. Visual comparison on Unpair dataset

Due to space limitations in the main manuscript, we only provided visual comparisons with GSAD [2]. In Figure 2, we offer additional visual comparisons on DICM, LIME, NPE, MEF, and VV datasets. These results demonstrate the generalizability of ReDDiT to real-world low-light im-

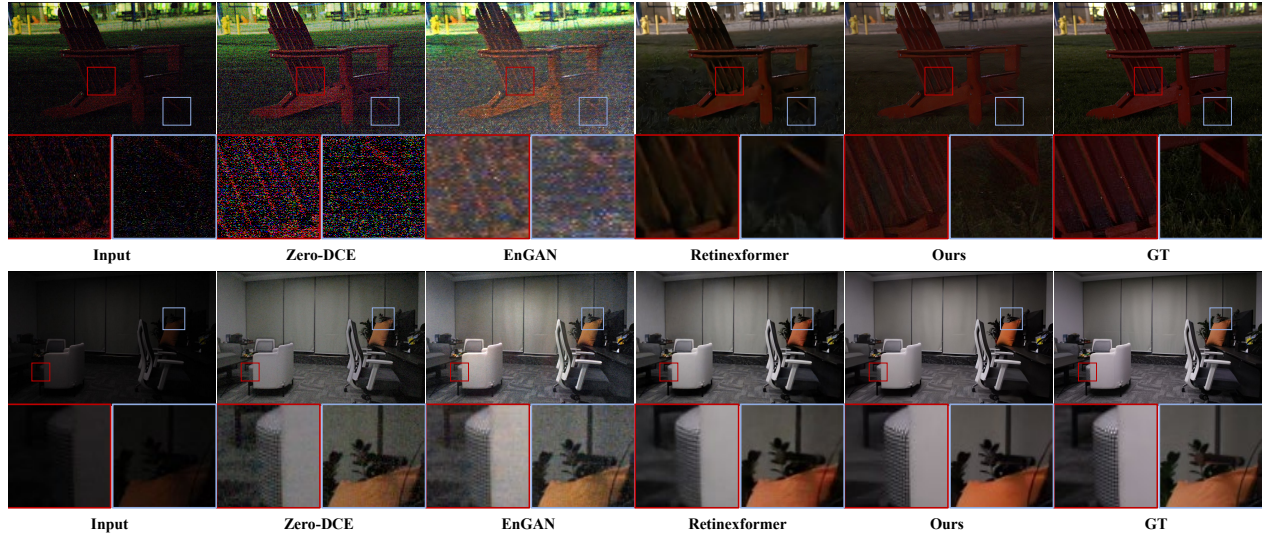


Figure 1. Qualitative results on SID and SDS. Patches highlighted in each image by the red and blue boxes indicate that ReDDiT effectively enhances the visibility, preserves the color, reduces noise, and retains finer details in normal light images. It can be observed that the visual results of our method align more closely with the ground truth. Please zoom in for a clearer view of the image details.

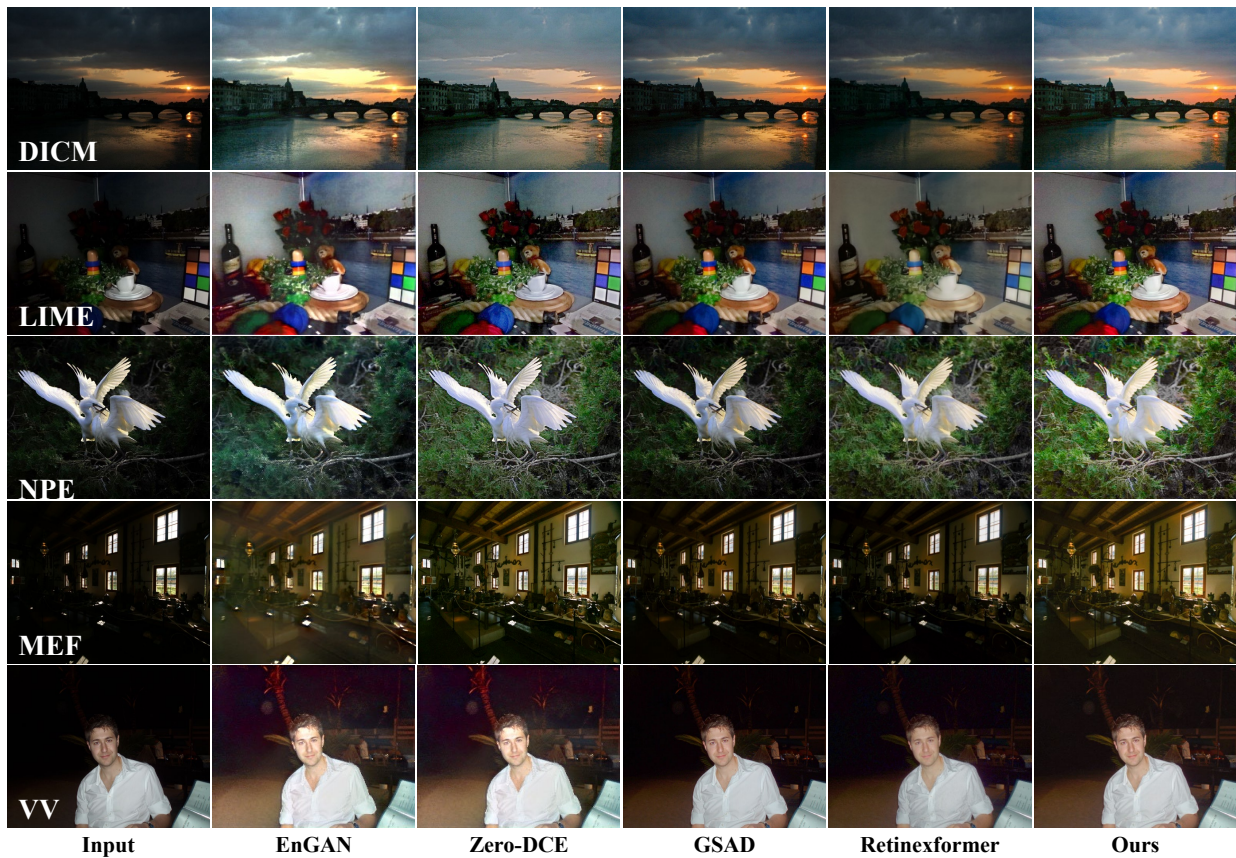


Figure 2. Qualitative results on DICM, LIME, NPE, MEF, and, VV. Our method effectively enhances the visibility and preserves the color. Please zoom in for a clearer view of the image details.

age enhancement scenarios and its consistent performance and superiority in preserving image details and improving visibility, making it a reliable solution for real-world applications.

References

- [1] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. [1](#)
- [2] Jinhui Hou, Zhiyu Zhu, Junhui Hou, Hui Liu, Huanqiang Zeng, and Hui Yuan. Global structure-aware diffusion process for low-light image enhancement. *Advances in Neural Information Processing Systems*, 36, 2024. [1](#)
- [3] 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*, 45(4):4713–4726, 2022. [1](#)
- [4] Ruixing Wang, Xiaogang Xu, Chi-Wing Fu, Jiangbo Lu, Bei Yu, and Jiaya Jia. Seeing dynamic scene in the dark: A high-quality video dataset with mechatronic alignment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9700–9709, 2021. [1](#)