

Supplementary Material

Metadata-Based RAW Reconstruction via Implicit Neural Functions

Leyi Li¹ Huijie Qiao² Qi Ye^{1,3} Qinmin Yang¹

¹Zhejiang University ²Chinese Academy of Sciences ³Key Lab of CS&AUS of Zhejiang Province

{lileyi, qi.ye, qmyang}@zju.edu.cn qiaohj@ioz.ac.cn

S1. Quantitative Results on Real Camera ISPs

Method	Samsung NX2000		Olympus E-PL6		Sony SLT-A57	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
RIR [5]	37.62	0.9696	42.19	0.9865	45.22	0.9916
SAM [8]	38.80	0.9725	43.15	0.9881	46.02	0.9921
CAM [4]	41.59	0.9818	47.76	0.9944	49.58	0.9954
Ours	43.05	0.9880	49.35	0.9949	51.36	0.9964

Table S1. Quantitative results on real camera ISPs. The best score for each column is in bold.

We report the quantitative results on the real RAW-JPEG image pairs in the NUS dataset [1]. Since the real ISPs are more complicated, we set the regularisation parameters to $\lambda_s = \lambda_p = \lambda_r = 0$ to ensure the complexity of INF, and use the same patch size as our main experiments. As is shown in Table S1, there is an obvious improvement drop compared with results of our main experiments, due to the complex structure of the real ISPs. However, our method still outperforms the closest competitor CAM [4] about 1.5dB in average PSNR, which indicates that our method generalizes well to the real ISPs. Note that for Samsung NX2000, we only take the coordinates as input, as we find taking the pixel values as input would inevitably lead to a drop in performance. We speculate that the camera ISP of Samsung NX2000 involves operations like resizing and cropping, hence considering the mapping between the pixel values of RAW images and JPEG images would lead to a more inaccurate reconstruction.

S2. Quantitative Results of Low-light Image Enhancement

We report the results on low-light image enhancement (LLIE) task in Figure S1. We first reconstruct the RAW image using our proposed method, and then correct the exposure by enlarging all pixel values 10 times, which is also implemented on the original JPEG image for comparison. After that, a simple camera ISP pipeline is adopted to re-render the RAW image back to sRGB color space. We also provide one classical method [6] and five deep learning based approaches [10, 2, 11, 9, 3] as references. It is obviously that the result of our method achieves equal or even better visual quality compared with other methods, which indicates the effectiveness of executing low-light enhancement on reconstructed RAW images. This has the advantage of simplifying the enhancement operation, as our method only requires a linear conversion. Note that we do not aim to illustrate that our method is better than others, as most LLIE methods are designed not only for enhancement but also for denoising.

S3. Instability from Random Initialization

We point out that if we reconstruct a specific image multiple times, the accuracy values would differ to each other due to the random initialization of INF. Though the random seed can be fixed for reproducibility, this would sometimes mislead the definition of *improvement*. We conduct an experiment that reconstructing an image 300 times, and the result is shown in

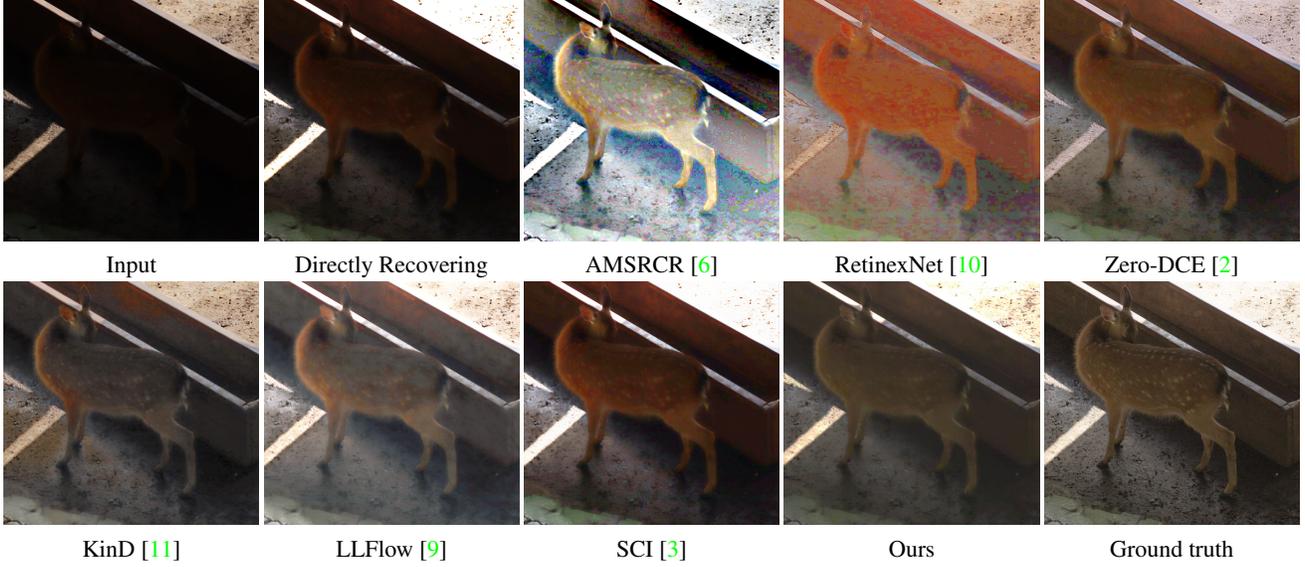


Figure S1. Qualitative comparison on low-light image enhancement task. *Directly Recovering* represents directly processing the original image with the same operation on reconstructed RAW image. For [6], we re-implement their method to produce the result. All deep learning based methods are applied with the pre-trained models. This figure is best viewed in the electronic version.

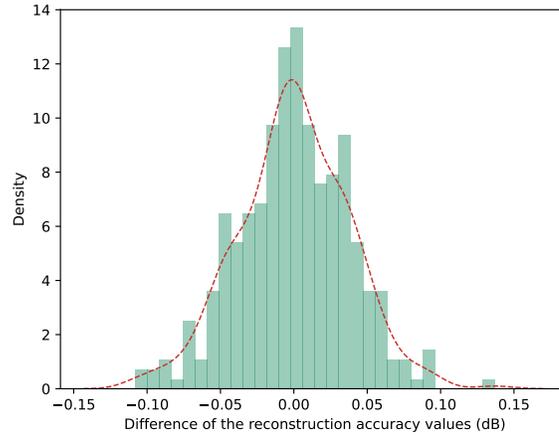


Figure S2. Reconstruction accuracy distribution due to random initialization. We use kernel density estimation (KDE) [7] to measure the error distribution. The abscissa value represents the difference of reconstruction accuracy values to their mean value, and the ordinate value refers to the kernel density of KDE.

Figure S2. As can be seen, the random initialization can lead to a difference about 0.25 dB in the final accuracy, which means an accuracy gap less than 0.25 dB is not properly regarded as an *improvement*.

S4. Different Calculations of Total PSNR

We note that there are two different ways to calculate the total PSNR, which could bring an error about 0.2–0.34 dB. Specifically, since PSNR could be defined as

$$PSNR = 10 \log \left(\frac{MAX_I^2}{MSE} \right) = -10 \log (MSE_{norm}), \quad (1)$$

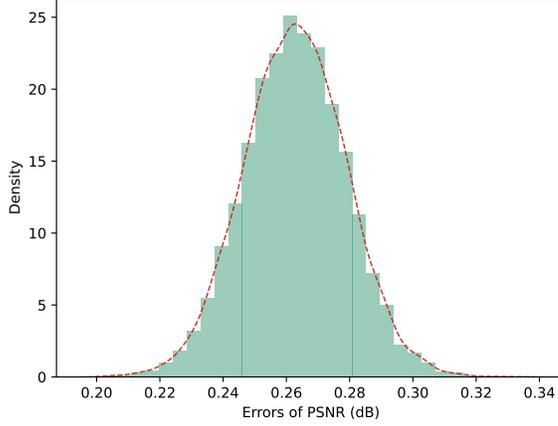


Figure S3. Error distribution for different calculation of total PSNR. We use kernel density estimation (KDE) [7] to measure the error distribution. The abscissa value represents the error of different calculation, and the ordinate value refers to the kernel density of KDE.

where MAX_I denotes the maximum possible pixel value of the image (i.e., $2^B - 1$ if the image is B bit per pixel) and MSE_{norm} represents the MSE after normalization, we could calculate the total PSNR by either

(a) Calculating the PSNR of each image and then averaging them, i.e., $PSNR_{total} = \frac{-10 \sum_{i=1}^n \log(mse_i)}{n}$.

(b) Calculating the average MSE of images and then producing PSNR, i.e., $PSNR_{total} = -10 \log \left(\frac{\sum_{i=1}^n mse_i}{n} \right)$.

We discuss the difference between these two methods by the subtraction,

$$\frac{-10 \sum_{i=1}^n \log(mse_i)}{n} - \left(-10 \log \left(\frac{\sum_{i=1}^n mse_i}{n} \right) \right) \quad (2)$$

$$= 10 \left(\log \left(\frac{\sum_{i=1}^n mse_i}{n} \right) - \frac{\sum_{i=1}^n \log(mse_i)}{n} \right) \quad (3)$$

$$= 10 \left(\log \left(\frac{\sum_{i=1}^n mse_i}{n} \right) - \log \left(\prod_{i=1}^n mse_i \right)^{\frac{1}{n}} \right) \quad (4)$$

$$= 10 \log \left(\frac{\sum_{i=1}^n mse_i}{n \left(\prod_{i=1}^n mse_i \right)^{\frac{1}{n}}} \right). \quad (5)$$

According to *Inequality of Arithmetic and Geometric Means*, $\sum_{i=1}^n mse_i \geq n \left(\prod_{i=1}^n mse_i \right)^{\frac{1}{n}}$, which means the result of (a) is higher than (b).

We further design a numerical experiment. We choose $mse_i \in [0.003, 0.011]$ and $n \in [200, 400]$ based on the previous experiment results, and deploy the Monte Carlo experiments with iterations of 10000. The result is shown in Figure S3. As can be seen, the error of different calculation is normally distributed in (0.2, 0.34). Therefore, jointly considering the error reported in Section S3, we do not take the accuracy gap less than 0.5dB as an improvement.

S5. Visualization

We provide visualization of more results in Figure S4.



Figure S4. Visualization of more results.

References

- [1] Dongliang Cheng, Dilip K Prasad, and Michael S Brown. Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution. *JOSA A*, 31(5):1049–1058, 2014. [1](#)
- [2] Chunle Guo 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, June 2020. [1](#), [2](#)
- [3] 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, 2
- [4] Seonghyeon Nam, Abhijith Punnappurath, Marcus A Brubaker, and Michael S Brown. Learning srgb-to-raw-rgb de-rendering with content-aware metadata. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17704–17713, 2022. 1
- [5] Rang MH Nguyen and Michael S Brown. Raw image reconstruction using a self-contained srgb-jpeg image with only 64 kb overhead. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1655–1663, 2016. 1
- [6] Sudharsan Parthasarathy and Praveen Sankaran. An automated multi scale retinex with color restoration for image enhancement. In *2012 National Conference on Communications (NCC)*, pages 1–5. IEEE, 2012. 1, 2
- [7] Emanuel Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076, 1962. 2, 3
- [8] Abhijith Punnappurath and Michael S Brown. Spatially aware metadata for raw reconstruction. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 218–226, 2021. 1
- [9] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex Kot. Low-light image enhancement with normalizing flow. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 2604–2612, 2022. 1, 2
- [10] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *British Machine Vision Conference*. British Machine Vision Association, 2018. 1, 2
- [11] 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. 1, 2