# Supplementary for Abandoning the Bayer-Filter to See in the Dark

Xingbo Dong[1,3*†] Wanyan Xu[1,2*†] Zhihui Miao[1,2†] Lan Ma[1]
Chao Zhang[1] Jiewen Yang[1] Zhe Jin[4] Andrew Beng Jin Teoh[3] Jiajun Shen[1]
[1]TCL AI Lab [2]Fuzhou University [3]Yonsei University [4]Anhui University
{xingbo.dong,bjteoh}@yonsei.ac.kr,{208527051,208527090}@fzu.edu.cn,{sjj,rubyma}@tcl.com

This supplementary document provides additional materials in the following aspects: (1) data collection details; (2) the detailed structure of networks employed; (3) more experimental results in terms of visual comparison depicted; (4) qualitative results on smartphone images demonstrated subsequently and (5) challenging cases discussed.

## 1. Dataset Collection

### 1.1. Cameras

The cameras we used in the experiment contain a 1/2-inch CMOS active-pixel digital image sensor with a 1280H x 1024V active imaging pixel array. The cameras can be operated in its default mode or programmed for frame size, exposure, gain setting, and other parameters.

The colored camera (i.e. MT9M001C12STC) uses a Bayer color pattern. Even-numbered rows have green and red color pixels, and odd-numbered rows have blue and green color pixels. On the other hand, Bayer color filter is not equipped on the monochrome camera (i.e. MT9M001C12STM).

The essential difference between the cameras of MT9M001C12STC and MT9M001C12STM is the Bayer color pattern, which leads to different quantum efficiency (QE). Generally, QE is used to measure the effectiveness of an imaging sensor to convert captured photons into electrons. The higher QE at certain photon wavelengths, the better sensing capability. The QE curves of MT9M001C12STC and MT9M001C12STM are shown in Figure S1. As shown by the curve, better QE can be achieved at a broad range of wavelengths on the monochrome camera.

### 1.2. Data collection setup

The cameras are mounted on the sliding platform on sturdy tripods for outdoor scenes. The tripods and two cameras are shown in Figure S2.

### 1.3. Alignment of two cameras

We utilize openCV library to perform homography matrix computation. It consists of six steps: 1) extract ORB features; 2) match features by brute force; 3) sort matches by the score; 4) extract location of matches above threshold; 5) compute homography matrix based on matched locations; 6) align two images using the homography matrix.

## 2. Network Details

The DBF and DBLE network with dimension and channel details are depicted in Figure S3.

The down shuffle operation represents the pack raw operation same to [1]. The RGGB raw image is packed into four channels, reducing the spatial resolution by a factor of two in each dimension.

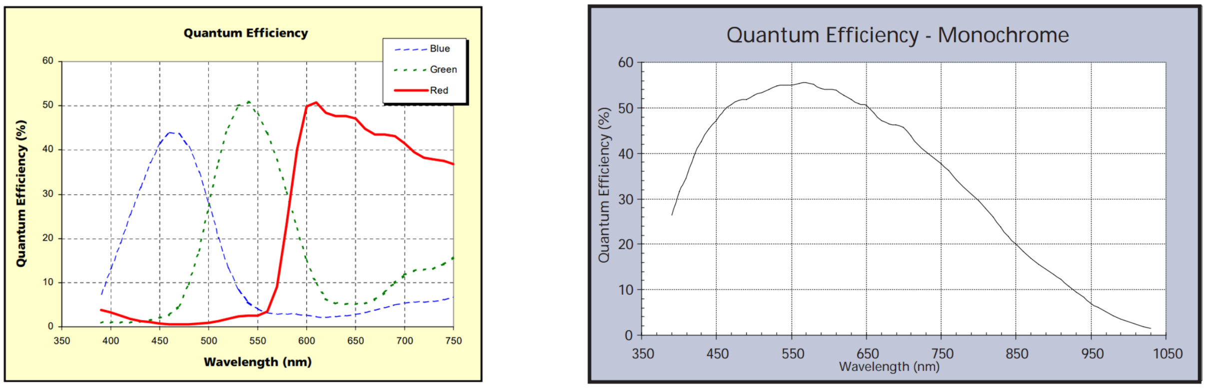
The pixel shuffle operation is taken from [5], which rearranges elements in a tensor of shape $\left(C \times r^2, H, W\right)$ to a tensor of shape $(C, H \times r, W \times r)$, where $r$ is an upscale factor.

## 3. Additional Visual Comparison

The results of the proposed method is visually compared with SID [1], DID [4], SGN [2], LDC [6], RED [3], the traditional histogram equalization (HE) approach and the Commercial Software Automatic Image Enhancement (CSAIE) method based on photo processing software. The CSAIE is performed using automatic tools provided by Adobe Photoshop on the ISP-generated RGB image. Specifically, the “Auto Contrast”, “Auto Tone ”, and “Auto Color” commands are executed sequentially. We provide additional visual comparison results in Figure S4 on two input low-light images.

## 4. Qualitative Results on Smartphone Images

To evaluate the generalizability, images captured by an iPhone 8 smartphone with a 400 ISO and a quick exposure time in outdoor scene are used as the input to evaluate the proposed method. We follow the procedure in SID [1]




(a) Colored camera (b) Monochrome camera

Figure S1. Cameras quantum efficiency. Monochrome camera has better QE performance.

(a) Cameras setup (b) Tripods

Figure S2. Two cameras and tripods setup.

to conduct the traditional pipeline. As for our proposed pipeline, the raw data taken by iPhone are prepossessed before inputting to the network, including subtracting the black-level and normalization to the same scale. Although the above prepossessing has overcome the gaps between sensors, transfer learning is preferred in practical deployment. The model trained on artificial mono-colored 14-bit raw SID dataset is adopted in the evaluation.

Figure S5 demonstrates the visual results of our method and traditional method (see [1]) on input low-light images RAW captured by iPhone 8. The results suggest that our model trained on SID can achieve a good visual performance on iPhone 8.

## 5. Challenging Cases

When some extremely dark cases occur on the images in our MCR Dataset, the existing low-light image enhancement algorithms (SID [1], LDC [6], and ours) show unsatisfying results. The restored images usually lost the high-frequency edge information compared to the ground truth image and became blurred. As shown in Figure S6, the spot on the surface of the white pear could be observed on our restored images but could not be seen on LDC restored image due to over smoothing. Because of inadequate illumination,

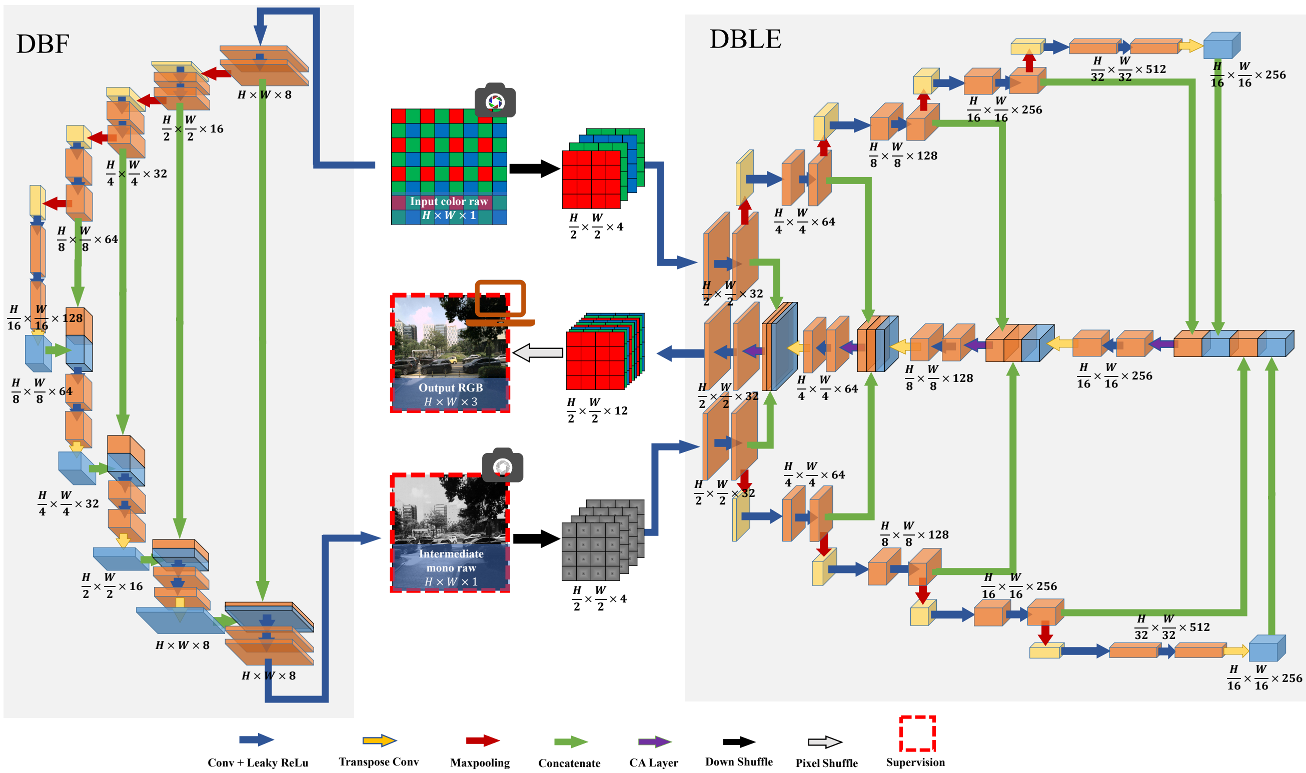



Figure S3. Network architecture with dimension and channel details.

restoring the extremely low-light raw images is still a challenging task.

## 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)*, pages 3291–3300, 2018. 1, 2, 5

[2] Shuhang Gu, Yawei Li, Luc Van Gool, and Radu Timofte. Self-guided network for fast image denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2511–2520, 2019. 1

[3] Mohit Lamba and Kaushik Mitra. Restoring extremely dark images in real time. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3487–3497, 2021. 1

[4] Paras Maharjan, Li Li, Zhu Li, Ning Xu, Chongyang Ma, and Yue Li. Improving extreme low-light image denoising via residual learning. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pages 916–921. IEEE, 2019. 1

[5] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1874–1883, 2016. 1

[6] Ke Xu, Xin Yang, Baocai Yin, and Rynson WH Lau. Learning to restore low-light images via decomposition-and-enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2281–2290, 2020. 1, 2, 5

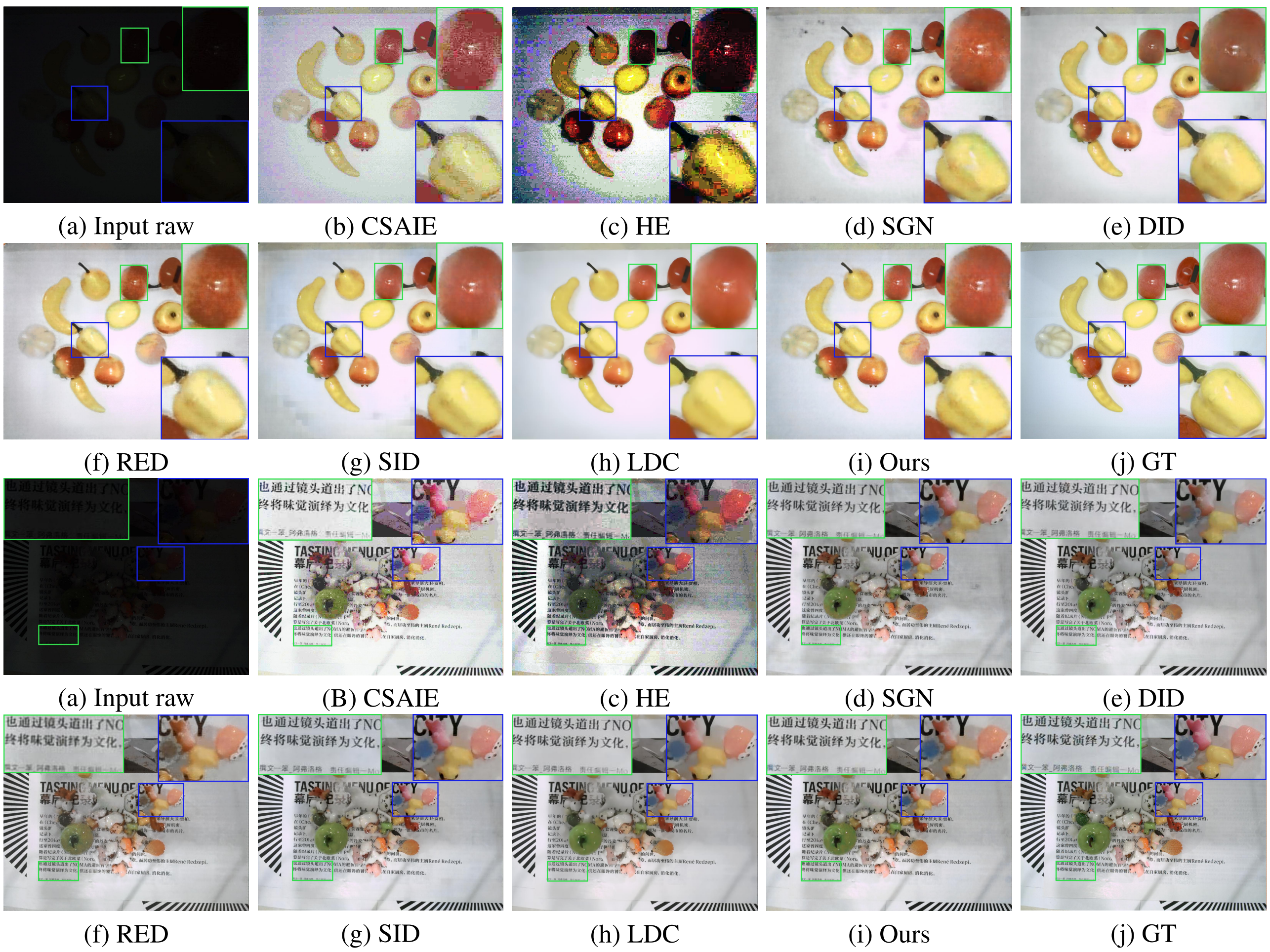
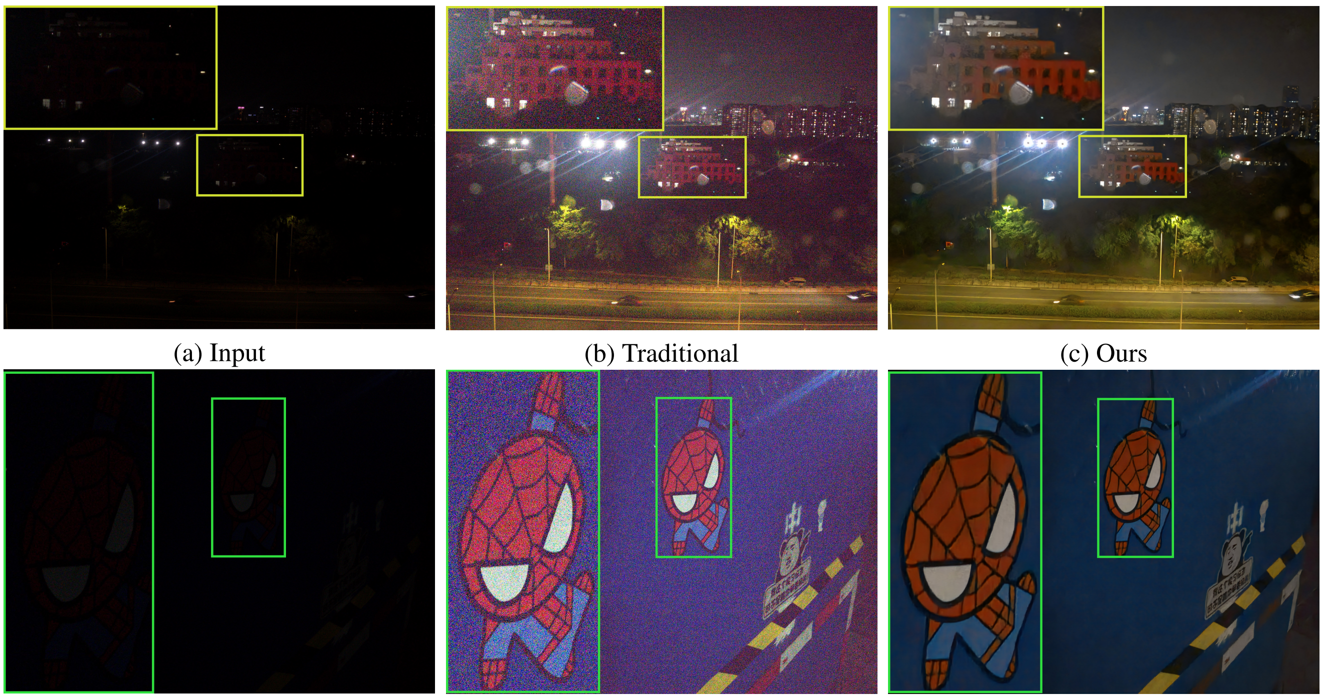


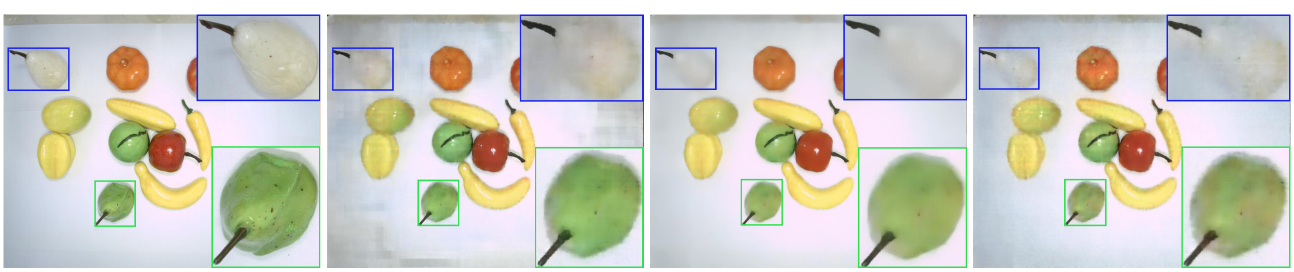

Figure S4. Visual results of state-of-the-art methods and ours on input low-light images RAW on MCR dataset.



(a) Input (b) Traditional (c) Ours

(d) Input (e) Traditional (f) Ours

Figure S5. Qualitative results on smartphone images.



(a) GT (b) SID [1] (c) LDC [6] (d) Ours

Figure S6. Failure cases in our MCR Dataset using different algorithms under extremely low-light situation.