Seeing in Extra Darkness Using a Deep-Red Flash
Supplementary Material

Jinhui Xiong[1*] Jian Wang[2*] Wolfgang Heidrich[1] Shree Nayar[2]
[1]KAUST [2]Snap Research

1. Lux and $\mathrm{cd/m^2}$

In the main manuscript, we have used two units to describe the ambient light levels: Lux and $\mathrm{cd/m^2}$. Lux is the unit for **illuminance**, which measures the total luminous flux incident on a surface per unit area. $\mathrm{cd/m^2}$ is the unit for **luminance**, which measures the luminous flux per unit area of a surface in a given direction. **Illuminance** is the quantity of light falling on a surface and **luminance** describes the intensity of reflected light. Brightness is a subjective term of the objective measurement of luminance. The luminance and illuminance are related by the reflectance of a surface. For a perfectly diffusely reflecting surface (a Lambertian reflector), the relationship can be expressed by:

$$L = \frac{ER}{\pi},$$

where L is the luminance, E is the received illuminance, R is the reflectance ratio. Full moon typically provides around 0.1 lux illuminance. With an illuminance level of 0.1 lux and a reflectance ratio of 0.5, the luminance of the scene would be $\frac{0.1\times 0.5}{\pi} = 0.016\ \mathrm{cd/m^2}$.

2. Brightness Gain and Signal-to-Noise Gain

We define $C(\lambda)$ as the camera spectral sensitivity, $V_{M,B}(\lambda)$ and $V_{M,R}(\lambda)$ as the human mesopic luminosity function for blue-heavy (e.g. $5000K$ white flash we compared with) and red-heavy (e.g. our proposed deep-red flash) lights, $\Phi_B(\lambda)$ and $\Phi_R(\lambda)$ as the spectral power distribution of corresponding blue-heavy and red-heavy lights, respectively. Brightness gain is computed by:

$$\text{Brightness Gain} = \frac{\int V_{M,R}(\lambda)\Phi_R(\lambda)d\lambda}{\int V_{M,B}(\lambda)\Phi_B(\lambda)d\lambda},$$
$$\text{where} \int C(\lambda)\Phi_R(\lambda)d\lambda = \int C(\lambda)\Phi_B(\lambda)d\lambda.$$

This can be interpreted as the ratio of the luminous flux from the red flash to the white flash, where the total power received by the camera is the same for both flashes.

Signal-to-noise gain is computed by:

$$\text{Signal-to-Noise Gain} = \frac{\int C(\lambda)\Phi_R(\lambda)d\lambda}{\int C(\lambda)\Phi_B(\lambda)d\lambda},$$
$$\text{where} \int V_{M,R}(\lambda)\Phi_R(\lambda)d\lambda = \int V_{M,B}(\lambda)\Phi_B(\lambda)d\lambda.$$

This can be interpreted as the ratio of total received signals by the camera when using the red flash versus using the white flash, which have the same brightness to human eye.

3. Energy Efficiency

We measured that the radiant efficiency (from electric power to optical power) of our used deep-red flash, LZ4-40R208-0000, is half of the white flash CREE XPEBWT-L1-0000-00C51. In specific, the radiant flux of red flash is half of that of white flash with the same electric power.

Notice that radiant flux measures the physical power of light, and luminous flux measures the perceived power of light. To convert from radiant flux to luminous flux, the power at each wavelength is weighted according to the luminosity function.

4. Network Details

4.1. Architecture

Fig. 1 shows an overview of the proposed MFF-Net.

4.2. Modulator

As introduced in the main manuscript, we introduce a specific modulator during training to let network exploit high-frequency features in the guide signal. The modulator can be written as:

$$f(x, y) = \alpha \cdot \sin(\frac{2\pi}{T}\sqrt{(x-\bar{x})^2 + (y-\bar{y})^2}) + \beta,$$

where $x \in \{1, 2, ...W\}$ and $y \in \{1, 2, ..., H\}$ (W and H are the width and height of images, respectively). α is the amplitude, T is the period, $\bar{x}$ and $\bar{y}$ are the phase shift, and β is the offset. All parameters are randomly chosen; α ranges from 0.1 to 0.5, and T ranges from 500 to 1000 pixels,

[*] denotes equal contribution. Part of the work was done while Jinhui Xiong was an intern in Snap Research.

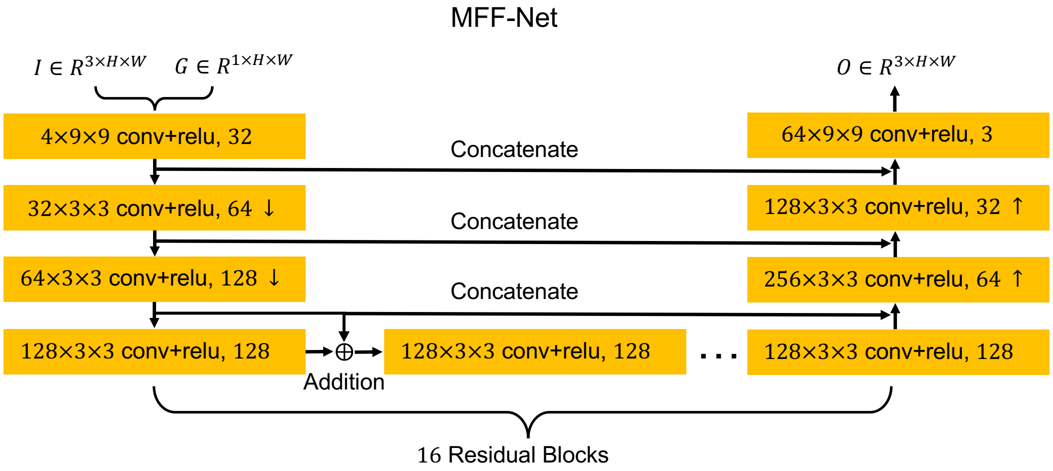

Figure 1: Architecture of the proposed MFF-Net network. The inputs are concatenation of the ambient-light image $I \in \mathbb{R}^{3\times H\times W}$ and its associated guide signal $G \in \mathbb{R}^{1\times H\times W}$. The output is the denoised image $O \in \mathbb{R}^{3\times H\times W}$.

yielding a low-frequency sine function. Notice that $f(x, y)$ should be consistently positive. We establish a random β to ensure that the least value of $f(x, y)$ is no less than 0.2 and the largest value is no more than 1.6. In Sec. 5.1, we show the output images from trained networks using and without using the modulation strategy. Without employing the modulation strategy, $f(x, y)$ can be considered to be 1.

4.3. Training Details

We train MFF-Net on 1000 images. We use a batch size of 16 and a patch size of 128×128. Random image flips and rotations are applied for data augmentation. A combination of VGG perceptual loss and ℓ_2 loss is applied as the loss function, with hyper-parameter 0.01 and 1 for each of them. We use Adam [3] for optimization with a learning rate of 10^{-4} in the first 300 iterations, and reduce it to 10^{-5} in the subsequent 300 iterations. The training takes about 6 hours on a Nvidia Tesla V100 GPU.

4.4. Training on Different Datasets

We trained our MFF-net on different datasets to validate its generality. As shown in Table 1, our network exhibits similar performance on real captured data when it was trained with images from three different sources. *The employment of the modulation strategy demonstrates its generalization ability from synthesized training data to real captured data, and also retains its generality across a broad range of datasets.*

Table 1: Quantitative results on different datasets.

Dataset	PSNR/SSIM
NYU v2 [4]	26.89/0.72
MIT-Adobe FiveK [1]	26.78/0.71
SID SONY [2]	26.83/0.72

5. Visual Comparisons

5.1. With and Without Modulation

In Fig. 2, we show the results from the networks trained with and without our proposed modulation strategy. Using without-modulation network appears to produce wrong color output images. We also visualize red, green and blue channel signals. Without using modulation operation, the intensity of red, green and blue channels in the generated output is correlated to the guide image. As the red-flash guide image has relatively poor signal-to-noise ratio in green and blue channels, signals in those two channels cannot be well recovered, especially in blue channel which has small spectral response to the deep-red wavelength. Applying the modulation function helps the network to learn to exploit the edges rather than rely on the intensity in the guide signal, and transmit the edge information, mainly in red channel, to the green and blue channels.

5.2. Additional Comparisons

In Fig. 3, we show additional qualitative comparisons with SOTA learning-based and model-based image fusion approaches. The images taken under good illuminance are served as reference.

We show additional video reconstruction results in the supplemental video.

References

[1] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input / output image pairs. In *CVPR*, 2011. 2

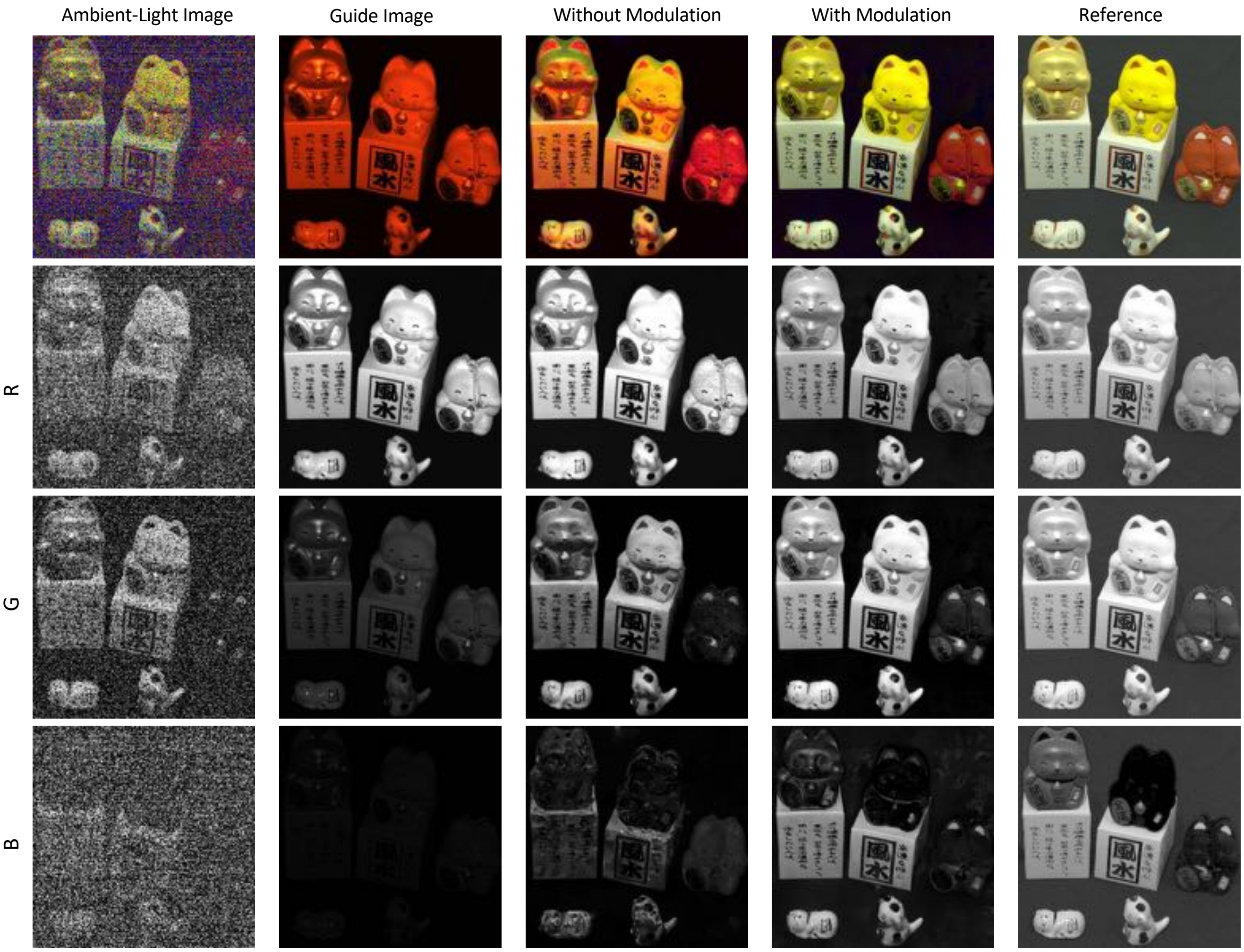

Figure 2: Visual comparisons for the effect of without exploiting and exploiting the modulation strategy when training the MFF-Net. Notice that the modulation operation is only performed during training. At the test phase, the guide signals for both networks are the same – summation of red, green and blue color channels. The signals in red (R), green (G) and blue (B) channels are visualized, respectively (guide image has poor signals in the blue channel, yet it is not completely black). The image taken at good illuminance is shown as reference.

[2] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *CVPR*, 2018. 2

[3] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 2

[4] Pushmeet Kohli Nathan Silberman, Derek Hoiem and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *ECCV*, 2012. 2

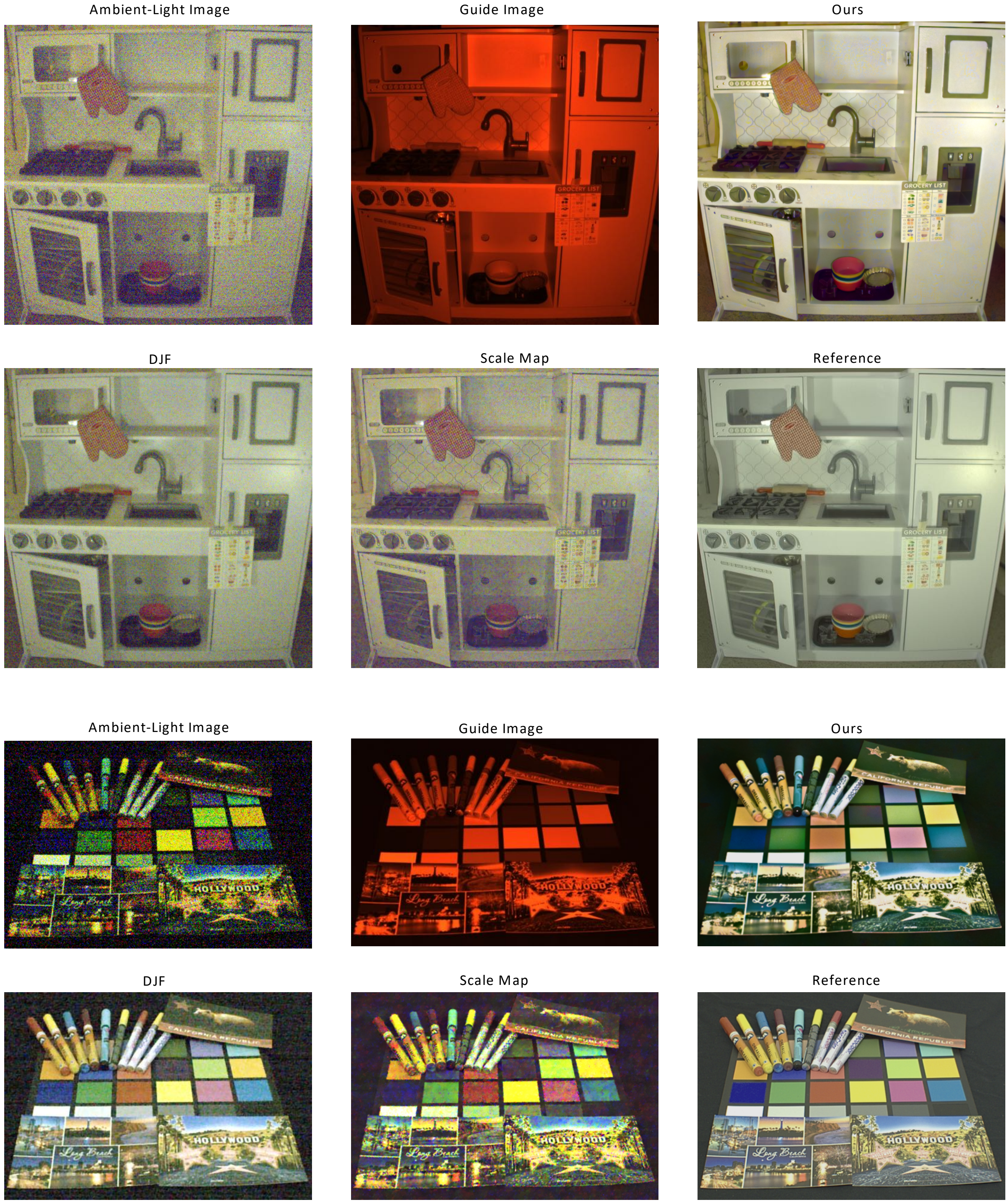

Figure 3: Additional qualitative comparisons. Please zoom in for a better view (see the grocery list in the first example and the texts on the pens in the second example).