Extreme Low-Light Image Denoising

Bingyi Song, Zhuoyu An, Linyan Jiang, Haibo Lei, Qing Luo, Jie Song, Ru Wang

May 24, 2024

Abstract

The project on Physics Based Vision meets Deep Learning (PBDL) for Extreme Low-Light Image Denoising

In our project focusing on extreme low-light image denoising, we built up models based on training data from ELD and SID datasets.

The training data pairs are pre-processed by

1) patch-based phase-correlation alignment,
2) exposure ratio turbulation,
3) Bayer aligned random crop and rotate data argumentation,
4) an approach for normalizing/de-normalizing Bayer raw.

Then we use the modified U-net structure with residual block and attention units, our method including 2-stage training strategy. First, we use all raw datasets for extracting as much as possible low level features. Second, the ELD datasets input to fine-tune the accurate camera model noise properties.

1 Datasets processing

The Bayer pattern raw sensor specialty lies in each individual pixel receiving only one spectral wavelength of light at a time. Given the limited amount of training data available, it becomes imperative to consider the spectral properties of every training pairs. There are four channels R, Gr, B, Gb in the sensor, Gr and Gb pixels have slightly different intensity response even they are both capturing the green color wavelength due to the Color Filter Array lens are not perfect(usually compensated by Image Signal Processing GbGr balance module). As well as the R and B channels offset impacts the denoise performance.

Our method first using the accurate patch-based registration while the images are captured on the tripod.

The registration is done using phase-correlation [Fig.1] which is consider to against the strong noise and brightness change. First, we transform the images into frequency domain using Fast-Fourier transform, then calculate the cross-power spectrum by taking the complex conjugate multiplication with elementwise normalization. $\mathbf{G}_a = \mathcal{F}\{I_a\},\ \mathbf{G}_b = \mathcal{F}\{I_b\}$

$$R = \frac{\mathbf{G}_a \circ \mathbf{G}_b^*}{|\mathbf{G}_a \circ \mathbf{G}_b^*|}$$

Then we find the maximum response phase as the image patch offset $(\Delta x, \Delta y)$ in the inverse Fast-Fourier transform result.

$$r = \mathcal{F}^{-1}\{R\}$$

$$(\Delta x, \Delta y) = \arg\max_{(x,y)}\{r\}$$

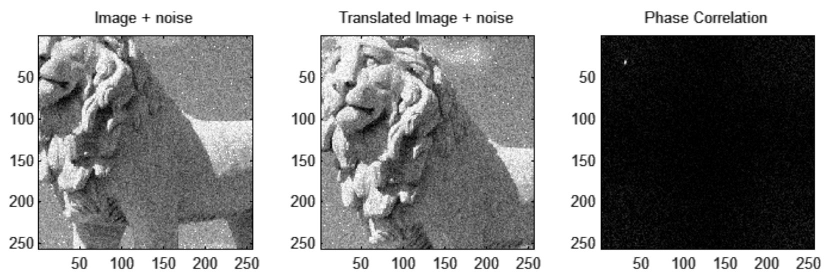

Figure 1: phase correlation. The white point corresponded phase position is the image offset.

Due to the settings and exposure time of different brand of sensors, as well as the expect exposure value (EV) is very sensitive for the final result, we introduce a variable λ, ranging from 0.1 to 10 in our settings to reduce reliance on accurate exposure precision. The digital gain (ISO) and exposure time in second is extract from EXIF metadata and calculate into exposure value (EV) for ratio estimation.

$$ratio = \lambda \frac{EV_{gt}}{EV_{in}} = \lambda \frac{ISO_{gt} \times TIME_{gt}}{ISO_{in} \times TIME_{in}}$$

The data argumentation is carefully done by random size crop and rotation while keeping Bayer pixels alignment. We didn't do any scale-like re-sampling argumentation for keeping the sensor noise properties.

2 Network Architecture

We use a 2-stage training strategy for Bayer raw denoising with slightly different networks for each stage. In the first stage, we use a U-Net with residual blocks as the denoising network, and the L1 loss function is used for faster convergence. In the second stage, we add attention blocks after residual blocks and freeze the weights of first-stage network, enhancing the denoising capability by minimizing the mean square error (MSE).

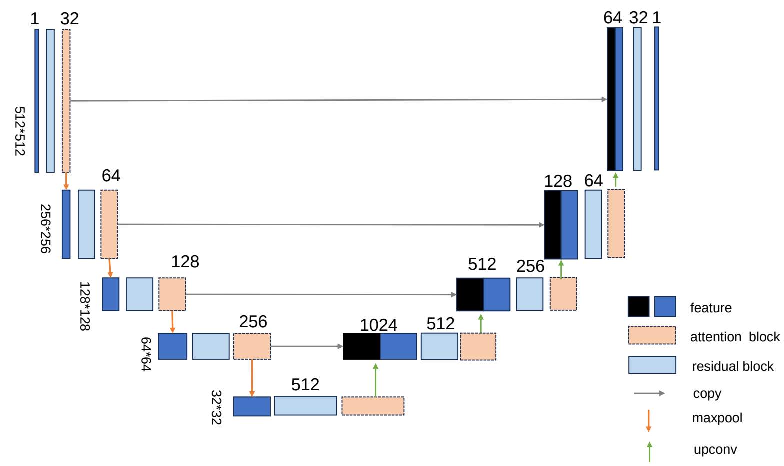

Figure 2: U-Net architecture.

3 Training strategy

First, our model is trained on all brand of camera raw files. Each image patch pairs for training input is fully-argumented with crop, phase alignment with ground truth, rotate and λ ratio scale.

Second, follow the first pretrained model, we fine-tune it by testing camera brand for better noise model estimation.

In practice, different sensors have non-equal black levels and white points per-channel, our other contribution is find a way to normalize raw Bayer data value I_n through different images taken by variant brand of cameras and exposure settings. While an image is taken, the raw file also save black levels L_b and white points per-channel L_w; After multiplying the EV ratio (the ratio is relatively 10 times larger than camera EV change) to get normal brightness, the noisy input may be clip by sensor max bit value (usually 14-bits), introducing signal missing. Our method solve this problem by normalize using a large denominator L_m (above 16-bits) after subtracting the black levels. Replacing L_w by L_m make image detail region preserved. Also the variable λ in training results better denoised image while keeping using GPU capable precision.

$$I_n = \frac{I - L_b}{L_m - L_b}$$

The following equation shows a normalized Bayer raw multiply exposure ratio, then $f(.)$ donates our network function, finally de-normalize to the expect denoised bright image I_{nr}.

$$I_{nr} = f(I_n * ratio)(L_m - L_b) + L_b$$

4 Reference

1. https://en.wikipedia.org/wiki/Phase_correlation

2. K. Wei, Y. Fu, Y. Zheng and J. Yang, Physics-Based Noise Modeling for Extreme Low-Light Photography, in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 11, pp. 8520-8537, 1 Nov. 2022, doi: 10.1109/TPAMI.2021.3103114.

3. Jin, Xin and Xiao, Jia-Wen and Han, Ling-Hao and Guo, Chunle and Zhang, Ruixun and Liu, Xialei and Li, Chongyi, Lighting Every Darkness in Two Pairs: A Calibration-Free Pipeline for RAW Denoising, Proceedings of the IEEE/CVF International Conference on Computer Vision 2023

4. Feng H, Wang L, Wang Y, et al. Learnability enhancement for low-light raw denoising: Where paired real data meets noise modeling[C]//Proceedings of the 30th ACM International Conference on Multimedia. 2022: 1436-1444.

5. Chen C, Chen Q, Xu J, et al. Learning to see in the dark[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 3291-3300.

6. Wei K, Fu Y, Yang J, et al. A physics-based noise formation model for extreme low-light raw denoising[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 2758-2767.

7. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer International Publishing, 2015: 234-241.

8. Ding X, Zhang X, Ma N, et al. Repvgg: Making vgg-style convnets great again[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13733-13742.