PBDL 2024 Low-light Raw Image Enhancement
A Light-weight Aligned Attention for Low-light Raw Enhancement

1. Team details

- Team name: Miers
- Team leader name: Cheng Li
- Team leader email: licheng8@xiaomi.com
- Rest of the team members:
  - Jun Cao, caojun6@xiaomi.com;
  - Shu Chen, chenshu1@xiaomi.com;
  - Zifei Dou, douzifei@xiaomi.com
- Affiliation: Xiaomi Inc., China
- User names on Codalab: lc

2. Methods

The Miers proposed a multi-scale, light-weight transformer model for low-light raw image enhancement. Unlike previous Retinex-based methods that generally decompose the input image into illumination components and reflection components, the proposed method adaptively aligns brightness-induced differences by introducing a learnable guidance vector in the self-attention mechanism.The network architecture called SANet is shown in Fig. 1. The SANet extracts features at different scales sequentially and performs feature fusion through long connections, which can effectively reduce the calculation. At each scale, the proposed method uses concatenated residual blocks and the SABlock as basic modules to obtain non-local views.

The self-attention mechanism in Transformer structure has been proven to have great advantages in low-level image enhancement, and the proposed SABlock (show in Fig.2) is also based on this. The SABlock captures global dependency information by building key-value pairs on feature blocks. In the low-light enhancement task, due to the large differences in input image distribution caused by illumination, the model is not easy to fit for natural state. This team introduced a learnable adaptive vector in SABlock to control the gap between the input RAW and the target. This allows the model to be effectively fitted to the direction that contributes to the correct output.

It is also worth noting that downsampling in SANet is implemented using 4×4 convolution with $stride = 2$ and the UP Block consists of a 3×3 depthwise separable convolution and pixelshuffle.

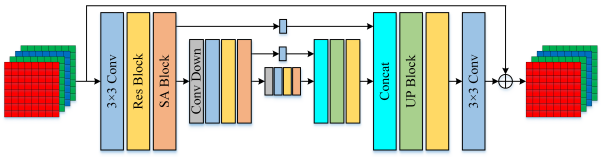

Figure 1. The proposed method.

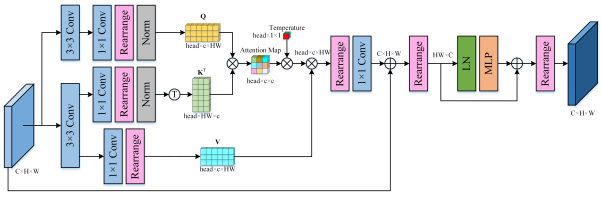

Figure 2. The structure of the SABlock.

3. Training

Our code is based on [1] and [2]. During the training stage, only the data provided by the competition is used. First, the input image is cropped to 128x128, rotation and flipping are added as data augmentation. It should be noted that due to the large difference in image brightness under different ISOs, the input RAW image subtracts the black level and divides by the difference between the white level and the black level, and then multiplies by the ratio for normalization. The ratio is calculated by

$$ratio = \frac{1}{\max(\frac{\text{raw\_image - black\_level}}{\text{white\_level - black\_level}})}$$

.

In the training process, the batch size is 4, total iterations is set to 500,000. This team uses L1 loss as the training loss and MultiStepLR for learning rate decay. In addition, the model weight uses exponential moving average (EMA), and the model with the highest PSNR on the validation set

is finally selected for testing.

References

[1] Lei Sun, Christos Sakaridis, Jingyun Liang, Qi Jiang, Kailun Yang, Peng Sun, Yaozu Ye, Kaiwei Wang, and Luc Van Gool. Event-based fusion for motion deblurring with cross-modal attention. In *European Conference on Computer Vision*, pages 412–428. Springer, 2022.

[2] Xintao Wang, Liangbin Xie, Ke Yu, Kelvin C.K. Chan, Chen Change Loy, and Chao Dong. BasicSR: Open source image and video restoration toolbox. `https://github.com/XPixelGroup/BasicSR`, 2022.