Low-light Raw Image Enhancement Technical solution

Xinyu Liu, Jing Zhang, Kexin Zhang, Yuting Yang, Licheng Jiao, Shuyuan Yang
Intelligent Perception and Image Understanding Lab, Xidian University
jingzhang_work @163.com

Abstract

This article outlines a successful solution / approach adopted in addressing the Low-light Raw Image Enhancement challenge held at the CVPR 2023 Conference. Since the noise in the original image (RAW image) does not pass the complex nonlinear mapping of the camera ISP and is easier to model, this low light enhancement challenge is based on the raw image. We used the basic framework of unet, fine tuning on 12 randomly selected images, and got the results of PSNR 27.09 and SSIM 0.84 in the test set. In this paper, this challenge is introduced, clarifying its aims and objectives. The report then delves into the solutions employed in this paper, with many of the techniques inspired by cutting-edge approaches presented during prestigious meetings in the fields of computer vision and machine learning.

1. Introduction

The quality of images taken in low-light environments is significantly worse than those captured under normal lighting, owing to environmental and technical limitations that are beyond our control. This results in a range of visual issues, such as loss of detail, color distortion, and increased noise. To address this issue, enhancing images taken in low-light conditions has emerged as a key area of focus within the field of image processing, aimed at improving visual quality and restoring image details. Beyond enhancements in the sRGB domain, there is a growing interest in RAW image enhancement, which offers higher bit depth and linearity compared to sRGB images. This type of enhancement not only caters to challenging lighting conditions but also offers benefits for a variety of visual tasks.

2. Extreme Low-Light Image Denoising challenge

2.1. The dataset

The dataset utilized in the challenge comprises 832 pairs of images, spread across 208 different scenes, indicating that four distinct exposure ratios (8, 16, 32, 64) for each scene. The dataset is categorized into four groups: indoor normal lighting, indoor low lighting, outdoor normal lighting, and outdoor low lighting. This low-light enhancement dataset features two key attributes: high resolution and the use of RAW images as input.
High Resolution: The dataset is notable for its high resolution of 6720x4480, surpassing the more common resolutions found in other datasets (below 1920x1080). This enhanced resolution captures finer details, providing a more comprehensive analysis for low-light enhancement.
RAW Image Input: The use of unprocessed raw sensor data as the input format capitalizes on the higher bit depth and superior intensity tolerance of raw data, effectively addressing the common issue of insufficient scene information in low-light enhancement scenarios.

2.2. The evaluation protocols

We employ the standard Peak Signal To Noise Ratio (PSNR) and the Structural Similarity Index (SSIM) in grayscale, as is commonly used in the literature. Peak signal-to-noise ratio (PSNR) is the ratio between the maximum possible power of an image and the power of corrupting noise that affects the quality of its representation. To estimate the PSNR of an image, it is necessary to compare that image to an ideal clean image with the maximum possible power. SSIM is a metric of comparison to check the similarity between the cover image and stego-image. It measures the perceptual difference between the two images.

3. Methodology

We used the algorithm proposed in the Lighting Every Darkness in Two Pairs: A Calibration-Free Pipeline for RAW Denoising, which can adapt to the target camera without calibrating noise parameters and repeated training, requiring only a small amount of lens pairing data and fine-tuning, eliminating the complicated calibration steps, and achieved good performance.

As shown in Figure 1, the whole network adopts the macro architecture of Unet, in which the convolution blocks

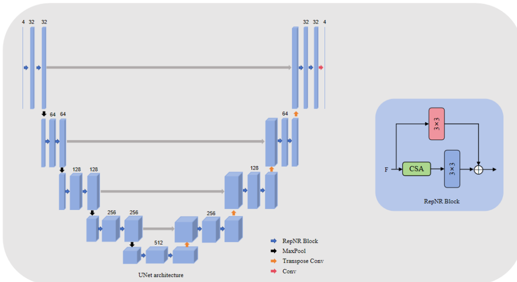

Figure 1. Overview of network architecture.

of the Unet network itself is replaced with the reparameterized noise removal (RepNR) block. In the Pre-train stage, In RepNR Block has k branches of Camera-Specific Alignment (CSA) module, Each of these branches is fitted to a class of camera noise, In the Fine-tune phase, By averaging the k CSA module, Equivalent to the model integration of noise from multiple classes of cameras, At this time the RepNR block consists of two branches, Where the upper 3x3 convolution is designed to fit the out-of-model noise, Lower Camera-Specific Alignment (CSA) module, The main role is to adjust the distribution of the input features.

4. Experiment

By utilizing LED methods, we ultimately reduced the number of scenes to just four groups, randomly selecting three distinct images for each scene to pair with their respective ground truth images. These image pairs were rapidly deployed to a new camera using the provided pre-trained weights.

Subsequently, fine-tuning was conducted using a small amount of real data, with the RepNR block replacing the convolutional layers in the UNet architecture. During the fine-tuning process, we initially iterated the CSA from the pre-trained model for 5000 iterations until convergence. The optimizer used was Adam with a learning rate of 0.0001, and the training strategy employed a cosine annealing approach. An additional branch was then fine-tuned for an additional 3000 iterations, with the optimizer and training strategy consistent with the main branch. The loss function chosen was L1 loss. During testing, a ratio of 10 was used for inference, resulting in a PSNR of 27.09 and an SSIM of 0.84 on the test set.

5. Conclusion

This work leverages LED-based techniques for signal processing within the RAW domain, bypassing the intricate steps such as Gamma correction typically found in traditional Image Signal Processors (ISPs). This streamlined approach contributes to a reduction in processing latency. Additionally, by incorporating channel attention mechanisms and residual learning, the receptive field for feature extraction is expanded, enabling a more comprehensive capture of detail and salient features within low-light images. The model was trained and tested on the provided dataset, achieving objective metrics of Peak Signal-to-Noise Ratio (PSNR) 27.09 and Structural Similarity Index (SSIM) 0.84. Subjectively, the generated images demonstrate reduced noise and color artifacts, thereby enhancing visual quality.