Low-Light Image Denoising with Residual Learning

Kundan Kumar

***Abstract*—Imaging in low light is challenging due to low photon count and low SNR. Short-exposure images suffer from noise, while long exposure can induce blur and is often impractical. A variety of denoising, deblurring, and enhancement techniques have been proposed, but their effectiveness is limited in extreme conditions, such as video-rate imaging at night. Recently, deep learning based approaches have been presented that have higher objective quality than traditional methods, but they usually have high computational cost which makes them impractical to use in real-time applications or where the processing power is limited. We used a dataset of short-exposure low-light images, with corresponding long-exposure reference images. We used a new residual learning based deep neural network for end-to-end extreme low-light image denoising.**

I. Introduction

Low-light imaging is a critical and challenging task in the fields of image processing and computer vision. Traditional imaging systems struggle in low-light conditions due to the inherent noise and low signal-to-noise ratio (SNR) associated with minimal photon counts. This challenge is exacerbated when high ISO settings amplify noise, and common physical adjustments such as widening the aperture or extending exposure time introduce undesirable effects like reduced depth of field or motion blur. While several techniques have been proposed to address these issues, including denoising and enhancement methods, they often fall short in extreme low-light environments where illumination is severely limited.

Recent advancements have focused on computational approaches to improve low-light imaging. One approach involves the use of deep learning techniques to enhance raw sensor data captured under such conditions. Traditional methods like BM3D perform reasonably well under moderate noise but fail in extremely low-light scenarios. Deep learning models, particularly U-Net architectures, have shown promise in this domain but come with limitations such as loss of image details and slow inference speeds.

In response to these challenges, new methodologies have emerged. For instance, an innovative image processing pipeline utilizes deep neural networks trained end-to-end to handle color transformations, demosaicing, noise reduction, and image enhancement in low-light raw data. Additionally, a novel residual learning-based network has been proposed, incorporating enhancements like LeakyReLU activation functions and Squeeze-and-Excitation blocks, to improve denoising performance and speed up the processing time.

This report delves into using Residual Learning based network along with Leaky ReLU activation function and squeeze and excitation blocks. The dataset includes png images captured with short exposure times in low-light conditions paired with long-exposure high-quality reference images. Our analysis demonstrates significant improvements in noise reduction and color accuracy, showcasing the potential of deep learning-based approaches in overcoming the limitations of conventional low-light imaging techniques.

II. Your Procedure or Your Method

The proposed method for extreme low-light image denoising and enhancement revolves around a residual learning framework. This approach aims to effectively reconstruct images captured in extremely low-light conditions by leveraging deep neural networks. Below is a detailed breakdown of each component (Network Architecture) and step involved in the method.

1. Initial Processing of Raw Sensor Image:

   Direct Use of RGB Channels: Unlike raw images, PNG images are already processed and stored in RGB format. Thus, there is no need to separate the image into RGBG color planes. The network will directly use the RGB channels of the PNG image.

2. Residual Learning Framework:

   Concept of Residual Learning: Instead of learning the entire image, the network focuses on learning the residual (or difference) between the low-light image and the ideal denoised image. This simplifies the learning process, as the network deals with smaller changes rather than the full image data.

2. Network Depth and Structure:

   32 Residual Blocks: The main network consists of 32 residual blocks, each designed to progressively refine the image features. This depth allows the network to learn intricate details and complex patterns in the data.

3. Structure of Residual Blocks:

   First 3x3 Convolution Layer: The input to each residual block is first processed by a 3x3 convolution layer, which applies a set of filters to extract local features from the image.

   Leaky ReLU Layer: Instead of the traditional ReLU (Rectified Linear Unit), which zeroes out all negative

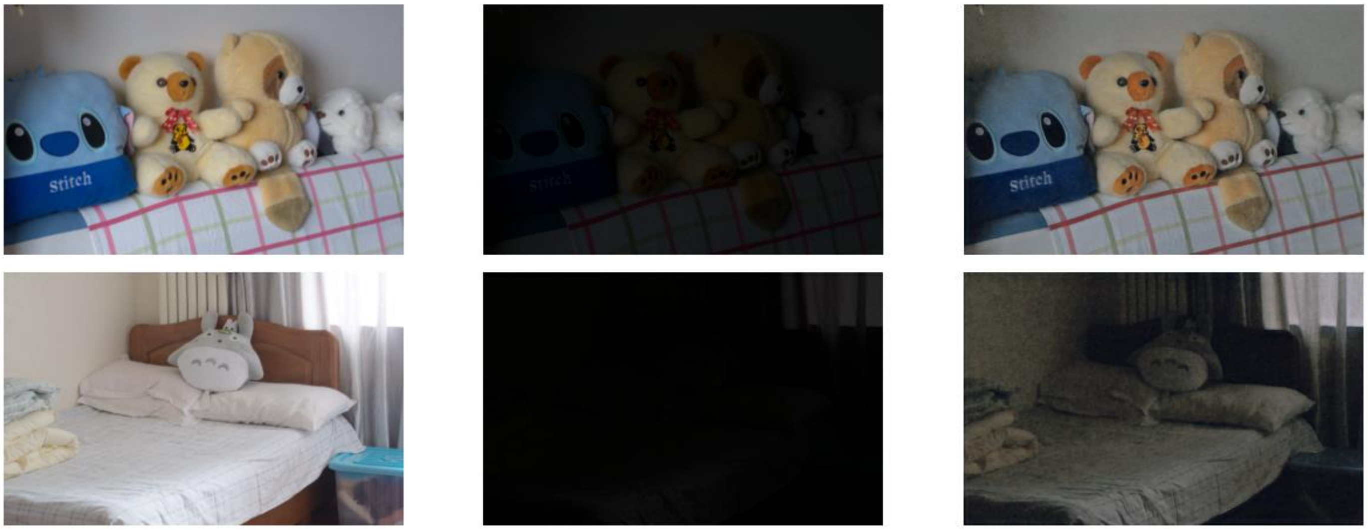

figure 1: {ground_truth image, input image, predicted image}

values, Leaky ReLU is used. Leaky ReLU allows a small, non-zero gradient for negative values, preserving more

Second 3x3 Convolutio Layer: The feature processed by the Leaky ReLU layer are then passed through another 3x3 convolution layer to further refine the extracted features.

Constant Linear Scaling Unit: This unit scales the output linearly, ensuring that the values remain within a manageable range before passing to the next stage.

Squeeze-and-Excitation (SE) Block:

Re-Calibration of Features: The SE block re-calibrates the convolutional features by performing channel-wise feature scaling. This improves the representation quality and helps the network focus on the most important features.

Boosting Network Performance: Integrating the SE block within the residual block not only enhances the network's performance but also speeds up the training process by providing more meaningful gradients.

4. Comparison with U-net Architecture:

Avoid Downscaling Structure: Unlike U-Net, which uses max pooling and downsampling layers, this network maintains a constant feature size throughout the residual learning part. This design choice prevents the loss of important information that typically occurs during downsampling.

Complexity and Depth: By avoiding complex downscaling and upscaling processes, the network remains less complex than its U-Net counterpart. This allows for an increase in the number of residual blocks (to 32), thereby enhancing the depth and learning capability of the network.

5. Training Process:

Input Size and Channel:

400x600 Pixels: as given in the dataset

Three-Channel RGB Input: The network uses a three-channel input (RGB) directly from the PNG images. This format is standard and already contains the necessary color information that is crucial for accurately reconstructing the image.

Efficiency and Speed: Real-time Processing, Despite the increased depth, the network is designed to process 4K resolution images quickly enough for real-time applications. This is crucial for practical usage in low-light video capture and other dynamic scenarios.

By integrating these carefully designed components and steps, the proposed method aims to effectively denoise extreme low-light images, providing significant improvements in both visual quality and processing speed.

Table 1: Complexity Analysis

Experiments	# of parameters	Time(sec)
SID	7.76M	0.235
Ours – 32 ResBlocks	2.5M	0.011

Table 2: Results

Specification	Iterations	PNSR	SSIM
150 patch_size	500	18.544	0.76
300 patch_size*	35	18.08	0.749
300 patch_size	100	17.85	0.775

*After every 10 iterations learning rate was reset to start.
Other experiments were made in one go without resetting learning rate.

IV. Experiments

1. Dataset and Experimental Setup

We used a dataset provided by the VLG IITR for this project. This dataset consists of 485 low-light images and 485 corresponding high-light images. We divided the dataset into a 9:1 ratio for training and testing.

We tried training on two platforms, first is google colab T4 GPU and second is Paramganga supercomputer provided by IITR.

The input to the network is a low-light png image and output is denoised png image. The ground truth is the corresponding high-light png image. During the training process, the input size is 400x600, randomly cropped from input image set with flipping and rotation for data augmentation and the output is 3 channel 400x600 png image. We have experimented with different residual blocks and different number of feature maps. The negative slope parameter of LeakyReLU is set to 0.2. We use L1 loss and Adam optimizer. We tried training for different number of epochs to see the pattern and fine tune the model. We also used Learning Rate scheduler for with factor 0.5 and patience 10 to reduce learning rate on Plateau.

2. Results

2.1 Denoising

Our proposed network reduces the noise of low-light images while preserving the color and texture information. Our best **PSNR** is 18.5.

2.2 Image Details

Since we do not reduce the feature size, we find our approach can better preserve the texture and edge details in the output images. On the contrary, SID produces output with smoother texture and may lose details due to contracting and symmetric expanding structure of the U-Net architecture.

3. Complexity Analysis

Our proposed network architecture has much less model parameters compared to the U-Net architecture used in SID [1]. SID uses 7.76M parameters whereas our network with 32 Residual Blocks uses 2.5M parameters.

V. Conclusions

In this paper we worked on deep residual learning network with Squeeze-and-Excitation block for denoising low-light image. Our network has good PSNR at decently reduced computational cost. With our residual network we can denoise the image under extremely low light condition while preserving most of the color and texture information. This advantage makes our network suitable for fast processing of low light images and videos on resource constrained devices.

VI. References

[1] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun, “Learning to see in the dark,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 3291–3300.

[2] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian, “Image denoising by sparse 3-d transformdomain collaborative filtering,” in IEEE Transactions on Image Processing, Aug 2007, pp. 2080–2095.

[3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, “U-net: Convolutional networks for biomedical image segmentation,” in Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, vol. 9351 of LNCS, pp. 234–241.

[4] Jie Hu, Li Shen, and Gang Sun, “Squeeze-and-excitation networks,” in IEEE Conference on Computer Vision and Pattern Recognition, 2018E.

[5] Tobias Plotz and Stefan Roth, “Benchmarking denoising algorithms with real photographs,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul 2017.

[6] J Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng, “Weighted nuclear norm minimization with application to image denoising,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 2862–2869.

[7] Michal Aharon, Michael Elad, and Alfred Bruckstein, “Ksvd: An algorithm for designing overcomplete dictionaries for sparse representation,” Signal Processing, IEEE Transactions on, vol. 54, pp. 4311–4322, Dec 2006.

[8] Daniel Zoran and Yair Weiss, “From learning models of natural image patches to whole image restoration,” in Proceedings of the IEEE International Conference on Computer Vision, Nov 2011, pp. 479–486.

[9] Stefan Roth and Michael Black, “Fields of experts: A framework for learning image priors.,” in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jan 2005, vol. 2, pp. 860–867.

[10] Weisheng Dong, Lei Zhang, Guangming Shi, and Xin li, “Nonlocally centralized sparse representation for image restoration,” IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, vol. 22, Dec 2012.

[11] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang, “Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising,” IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3142–3155, 2017.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep residual learning for image recognition,” The IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[13] K. Ram Prabhakar, V Sai Srikar, and R. Venkatesh Babu, “Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs,” in The IEEE International Conference on Computer Vision (ICCV), Oct 2017.

[14] Wenzhe Shi, Jose Caballero, Ferenc Huszar, 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,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1874–1883, 2016.

[15] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee, “Enhanced deep residual networks for single image super-resolution,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Jul 2017, pp. 136–144.