Extreme Low-Light Single-Image Denoising Model

Nikhil Parab
Stanford University
nik17@stanford.edu

Abstract

Denoising is challenging under extreme low-light conditions for real-time applications due to low photon count, low Signal-to-Noise (SNR) and single-frame processing constraints. A wide range of denoising solutions have been proposed. Longer exposure time improves the signal-to-noise ratio (SNR) but can result in motion blur artifacts. More recent approaches like burst denoising, which involve capturing and averaging multiple frames show great improvement in image quality but is not a practical solution for real time applications. Also, several other model-driven denoising, deblurring and enhancement techniques have been published but, these are not much effective under photon limited conditions.

Learning based approach have exhibited stunning success in a wide range of image recovery and restoration applications, thereby, gaining a lot of interest. The problem we are trying to address is, using one such data-driven approach, how far can we push a single-image processing under extreme low-normal light conditions to achieve an acceptable image quality for time-constraint systems like autonomous driving. The goal is to design a cross-sensor generalized model which can be conceptualized as a sensor post-processing or an Image Signal Processor (ISP) pre-processing denoiser plugin.

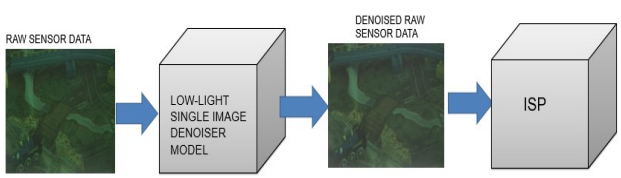

Figure 1. Proposed Model

1. Introduction

In any processing system, we must determine how much of the measured signal is true and how much of it is coming from other background events during acquisition or transmission. These random contributions from various background events along the path of signal detection and measurement can be classified as noise. Noise is an inherent characteristic of any imaging systems which obscures the desired information and is one of the most important consideration that determines the success or failure of any image processing or reconstruction task.

Noise can result from a vast variety of sources, like fix-pattern noise from sensor electronics, pixel defect noise, quantization noise from the Analog-to-Digital conversion (ADC) and shot noise from the discrete nature of light source itself. Denoising is a challenging task as it must tackle against these different noise sources and is an active field of research. Different noise follows different statistical distribution but can be primarily classified into two important ones, signal independent additive Gaussian-distributed noise and signal dependent Poisson-distributed noise. Gaussian noise can be considered as combination of thermal, amplifier and read noise which are independent of the measured signal and remains constant for a specific camera settings. Signal dependent shot noise follows Poisson distribution and is caused by statistical randomness in the light reaching the sensor and its photoelectron conversion.

Camera technology has been advancing, but low-light Poisson noise, also known as shot noise, is a fundamental problem that persists and becomes dominant under low light conditions. Poisson noise appears in many applications in various fields ranging from medical imaging to astronomy. No engineering solution can mitigate this as it is fundamental property of the particle nature of light. Let X denote the noisy image measured by a sensor. The goal of the denoising is to recover the latent clean image Y as observed by the sensor. In low-light given the true value Yij of the (i,j)-th pixel expressed in number of photoelectrons, the corresponding value of the observed pixel Xij is an independent Poisson distributed random variable with mean and variance Yij,

i.e., Xij ~ Poisson (Yij) :

$$\mathrm{P}(X_{ij} = n | Y_{ij} = \lambda) = \begin{cases} \frac{\lambda^n}{n!} e^{-\lambda} & \lambda > 0 \\ \delta_n & \lambda = 0. \end{cases} \quad (1)$$

From the above equation, we can notice that Poisson noise is neither additive nor stationary, its strength is dependent on the image intensity. Lower intensity in the image yields

a stronger noise as the Signal-to-Noise ratio (SNR) in each pixel is $\sqrt{Y_{ij}}$. Thus, in extreme low light imaging, noise removal becomes a critical challenge to produce a high quality, detailed image with low noise.

2. Related Work

Current denoising techniques are often focused on Gaussian noise. High ISO (electronic gain) can be used to boost the brightness, but it also amplifies the noise, arising from read, shot and defective pixel sources. Another naïve approach is to increase the number of incoming photons by increasing the opening the aperture wider or by leaving the shutter open for longer time. This can give good results under controlled static settings and but can cause motion blur artifacts in real world settings.

State-of-art techniques like NLM, BM3D [8] still outperform most of the current solutions out there. However, these are non-blind models and requires noise-level specified extrinsically. These models work exceptionally well compared to most other methods, on images where accurate noise-level can be provided. However, for low-light situations, noise-level prediction can get tricky resulting in over-smoothing or noisy results if the noise-level estimation goes wrong.

Many computational imaging techniques have been studied and proposed for addressing the low-light image reconstruction problems. These can be broadly classified into model-driven and data-driven approach. Model-based methods involve mathematical modelling of image noise by taking advantage of inherent image properties like self-similarity, smoothness and sparsity and using them as the model prior. Most of these model optimization techniques performs well when the image closely matches the model priors like total variation (TV), self-similarity and sparse coding. Also, their performance degrades under photon limited conditions where shot noise is significantly higher than the additive white gaussian noise (AWGN).

A more recent approach to increasing SNR and dynamic range is by increasing the captured photon count using technique called Burst Denoising [4] wherein a burst of images is captured from the same scene and averaged. This indeed is very effective, giving significant denoising performance under low-light conditions. However, it depends on reference image (lucky image) selection which can be tricky under extreme low light. Also, this method requires multi-frame processing which can be challenging in time constrained imaging systems like autonomous driving.

With the advances in Artificial Intelligence (AI), a bunch of Convolutional Neural Network (CNN) based architectures [1] have emerged and gained popularity in image restoration and segmentation domains. Learning to See in the Dark is one of the most recent publications focusing on end-to-end processing of single image under extreme low-light conditions. While it is effective with impressive results, it lacks cross-sensor generalization because of its end-to-end pipeline design. Also, it requires amplification factor as an external input, which can result in saturation for higher amplification factors.

3. Method

In our proposed work, the goal is to design an extreme low-light single-image denoising model which can be generalized across sensors. In a standard traditional camera processing pipeline, the sensor captured raw data is processed through a series of blocks like black level correction, white balancing, demosaicking, denoising, color space conversion, tone mapping and sharpening. Some of these are linear, some are non-linear and some sensor dependent. The traditional pipeline breaks down under extreme low-photon conditions as it is unable to handle low SNR.

Chen Chen proposed an end-to-end pipeline learning approach in his Learning to See in the Dark [2]. The results of this approach are promising but lacks generalization. One of the prime reasons for lack of generalization and poor High Dynamic Range (HDR) performance with this approach is the model trying to learn end-to-end complex multiple processing functions like demosaicking, denoising, color transformation, white balancing and tone-mapping, some of which are sensor-dependent. As a result, the model must be custom trained per camera module and this limits the amount of available data for training. Also, this approach requires new low-light data collection and the model re-training for each new camera sensor, which can involve a significant amount of time and effort. In our proposed approach, we plan to decouple this end-to-end learning method into denoising block and other post-denoising processing blocks.

The key idea is to learn denoising under low-light conditions, as we believe is one of the most critical blocks which drives the success of photon-limited imaging systems. A good denoising architecture not only improves the output image quality but also simplifies the design of rest of the blocks in the image processing pipeline. Also, decoupling denoise learning opens a pool of training dataset which can make it more robust, well-fitted and sensor agnostic. With this approach, not only the model can be well-generalized across camera sensors but also, it provides the flexibility to plugin custom post-denoising blocks to achieve high dynamic range and better color constancy without re-training the model. Figure 2. gives a brief overview of our proposed model.

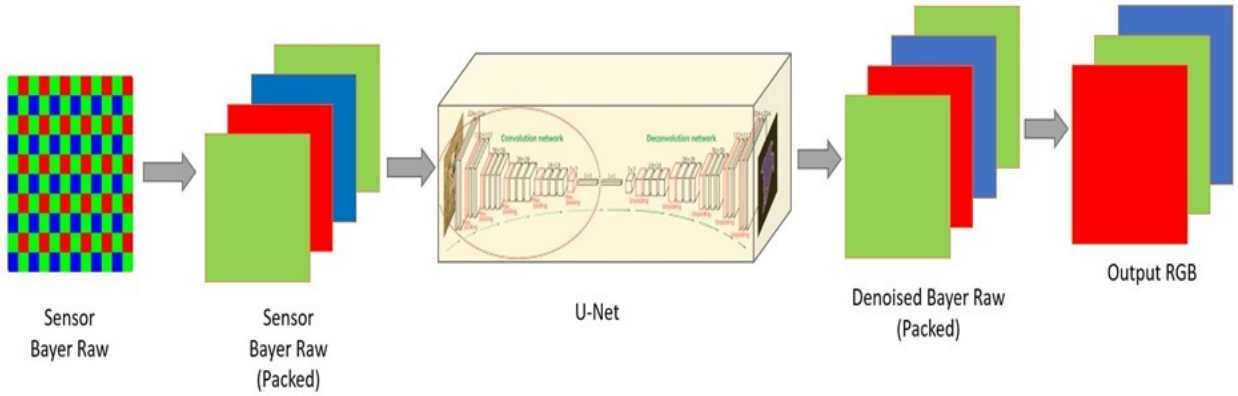

Figure2. Proposed Model

3.1. Network Architecture

Deep learning models, in particular U-Nets, have been widely used for denoising and various types of image reconstruction problems. U-Net [6] is a type of neural network consisting of a contracting CNN followed by an expanding CNN have shown excellent results in image denoising and reconstruction tasks. In the compression network, the model is incentivized to learn finer features in the initial layer, and subsequently coarser information as the image goes through depths due to the down sampling operations. The expanding network focuses on rebuilding features from the lowest depth back up to the first depth. The main intuition behind the success of U-Net in image reconstruction problems is its ability to go from a dense feature representation back towards the sparse one without losing information and the use of image feature mapping it learned for reconstruction preserves the structural integrity.

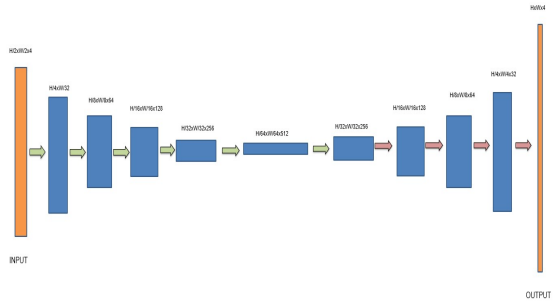

Figure3. U-Net Model

The U-Net network used in our work is shown in Figure 3. The input to the network is pre-processed raw data. The pre-processing step involves black level correction, input bayer data packing into 4-channels followed by data normalization. As a result of data packing, the spatial resolution is reduced by a factor of 2. This pre-processed data is fed into the U-Net. The output is a packed 4-channel denoised raw data. It is further processed through a very naïve minimal ISP consisting of scaling and gamma correction for visualization and qualitative analysis for this project. The output of the network can be processed through a custom processing pipeline to achieve the desired image quality for a specific use-case.

3.2. Dataset

Although there have been lot of studies on low-light denoising, the research community currently lacks a denoising image dataset representative of real-world noisy images with high-quality ground truth. One of the recent papers “A High-Quality Denoising Dataset for Smartphone Cameras" [3] addressed the need for real world data by providing one such dataset ‘Smartphone Image Denoising dataset[SIDD]’ dataset consisting of pairs of noisy and ground-truth images captured under a wide-range of lighting conditions, from low-light to bright-light and covering a wide range of smartphone sensor data like Google Pixel(GP), Iphone7(IP) and Samsung (S6).
The Learning To See In The Dark dataset also has a good collection of real-world extremely low-light images ranging from exposure time of 1/10sec to 1/30sec under 0.2lux to 5lux illumination but, it is limited to only two camera sensors (sony and fuji). Since the goal of our approach is to achieve cross-sensor generalization, we opted to use the SIDD dataset for training, validation our network and reserved SIDD benchmark and See In The Dark dataset for evaluating our learned model.

Figure 4. Two sets (sony-1, sony-2) of extremely low-light (0.1 sec, < 1lux) images from Sony dataset (a) jpeg image from camera (b) scaled (x100) jpeg image from camera (c) scaled(x100) noisy image (d) scaled(x100) denoised image with NLM (e) scaled(x100) denoised image with proposed model.

Figure 5. Example extreme low-light images (fuji-1, sony-3) from Sony and Fuji dataset.
first row: Fuji images (0.1sec exposure, < 1lux illumination) second raw: Sony images (0.04 sec exposure, < 1lux illumination).
first col: scaled noisy images; second col: scaled denoised image using NLM; third col: scaled denoised image using proposed model

3.3. Training

In our work, the model was trained using supervised learning approach. The dataset was Smartphone Image Denoising dataset. We started with a learning rate of 10^{-4} and later reduced it to 10^{-5} after 500 epochs. The model was trained for 1000 epochs. The training was done from scratch with L1 loss and Adam optimizer. The size of each input raw image in the dataset is huge (~24MB per image) and takes significantly loading time. So, the model was modified to run on a patch-size of 512x512 with random flipping and rotation for data augmentation in each iteration.

4. Experiments and Results

The entire model was trained on a Nvidia Tesla K80 GPU using Cuda and PyTorch 1.4.0. Due to GPU/RAM constraints on our testing machine, we could only load approximately 120 full-resolution images from our dataset for training, validation, and testing. We had to make some code-optimization to lower the data loading time in successive epochs which in turn helped increase GPU utilization per epoch and brought down the total training time to approximately 30hours.

Our baseline for comparison is state-of-art Non-Local Means [NLM] denoiser [5]. Due to timeline constraints, we did a comparative study with only one of the standard denoisers. It would be interesting to compare the results with BM3D, other data driven low-light denoiser models and can be topic for future work. Figure 4 and 5 show qualitative results of some extreme low-light images processed using our model and with NLM. Table 1. shows quantitative results for the same. Also, we have used Peak Signal-to-Noise (PSNR) ratio as an image quality metric in our study.

PSNR	Noisy Image	Noisy Image + NLM	Noisy Image + Proposed Model
sony-1 (0.1s)	17.9239	21.11	21.577
sony-2 (0.1s)	14.9193	18.95	24.23
fiji-1 (0.1s)	17..6681	20.49	20.15
sony-3(0.04s)	20.0664	23.11	20.43

Table 1. Quantitative results (PSNR) for denoising using NLM and proposed model.

5. Discussion and Conclusion

Overall, the results from the experiments were quite satisfactory. Though the results are not perfect, we see lot of scope for future work and improvements. Looking at both the qualitative and quantitative (PSNR) results, we can see that the model performs reasonably well under most of the conditions. The testing was mainly done on two sets of extreme low-light datasets (0.1s, and 0.04s). As you can see, both qualitative and quantitative results are quite equivalent, and sometimes better, for exposure time of 0.1s dataset.

Testing on super-low exposure time datasets consisting of 0.04s and 0.033s data, the model performance degrades for images with lot of details i.e. high frequency content since the model tries to strongly denoise the image data because of super-low light condition, resulting in strong smoothing and loss of texture details.

Also, the model fails to recover the chroma details accurately resulting in color bias which can be an interesting topic to study and improve upon. From qualitative point of view, most of the results for exposures up to 0.1s give acceptable results but the performance degrades under extreme low-light conditions (< 0.1s) for challenging textured image data. There is still a scope of improvement by training the model with wider spectrum of low-light data and can be improved in the future work.

6. Acknowledgements

Many thanks to project mentor Chris Metzler and Professor Gordon Wetzstein for advice and guidance.

References

[1] Chunwei Tian and Yong Xu and Lunke Fei and Ke Yan. Deep Learning for Image Denoising A Survey.

[2] Chen, Chen, Koltun & Xu. “Learning to See in the Dark”. May 2018.

[3] Abdelrahman Abdelhamed, Stephen Lin, Michael S. Brown CVPR2018 A High-Quality Denoising Dataset for Smartphone Cameras

[4] Samuel W. Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T. Barron, Florian Kainz, Jiawen Chen, Marc Levoy, Google Research, 2016, Burst photography for high dynamic range and low-light imaging on mobile cameras.

[5] Antoni Buades, Bartomeu Coll, Jean-Michel Morel, A non-local algorithm for image denoising.

[6] Olaf Ronneberger, Philipp Fischer, Thomas Brox, CVPR 2015, F U-Net: Convolutional Networks for Biomedical Image Segmentation

[7] NODE: Extreme Low Light Raw Image Denoising using a Noise Decomposition Network

[8] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian, Senior Member, IEEE Image denoising by sparse 3D transform-domain collaborative ltering

[9] Residual Dense Network for Image Super-Resolution Yulun Zhang and Yapeng Tian and Yu Kong and Bineng Zhong and Yun Fu, 2018 Self-Supervised Deep Image Denoising Samuli Laine and Jaakko Lehtinen and Timo Aila 2019

[10] Or Litany, Raja Giryes, Alex M. Bronstein Deep Convolutional Denoising of Low-Light Images

[11] Tao, Li & Zhu, Chuang & Xiang, Guoqing & Li, Yuan & Jia, Huizhu&Xie,Xiaodong. LLCNN: A convolutional neural network for low-light image enhancement. 1-4. VCIP.2017