Learning an Adaptive Model for Extreme Low-light Raw Image Processing

Authors: Qingxu Fu, Ti Xiaoguang, Zhang Yu 1, Ti Xiaoguang

Date: 2020-04-14T00:00:00+00:00

Abstract

Low-light images suffer from severe noise and low illumination. Current deep learning models that are trained with real-world images have excellent noise reduction, but a ratio parameter must be chosen manually to complete the enhancement pipeline. In this work, we propose an adaptive low-light raw image enhancement network to avoid parameter-handcrafting and to improve image quality. The proposed method can be divided into two sub-models: Brightness Prediction (BP) and Exposure Shifting (ES). The former is designed to control the brightness of the resulting image by estimating a guideline exposure time t 1 . The latter learns to approximate an exposure-shifting operator ES, converting a low-light image with real exposure time t 0 to a noise-free image with guideline exposure time t 1 . Additionally, structural similarity (SSIM) loss and Image Enhancement Vector (IEV) are introduced to promote image quality, and a new Campus Image Dataset (CID) is proposed to overcome the limitations of the existing datasets and to supervise the training of the proposed model. In quantitative tests, it is shown that the proposed method has the lowest Noise Level Estimation (NLE) score compared with BM3D-based low-light algorithms, suggesting a superior denoising performance. Furthermore, those tests illustrate that the proposed method is able to adaptively control the global image brightness according to the content of the image scene. Lastly, the potential application in video processing is briefly discussed.

Full Text

Preamble

Regular Paper: Learning an Adaptive Model for Extreme Low-light Raw Image Processing

Qingxu Fu[1], Xiaoguang Di[1]∗, Yu Zhang[2]
[1] Control and Simulation Center, Harbin Institute of Technology, Harbin,

150080, People' s Republic of China
[2] National Key Laboratory of Tunable Laser Technology, Harbin Institute of Technology, Harbin, 150080, People' s Republic of China
∗ E-mail: dixiaoguang@hit.edu.cn

ISSN 1751-8644
doi: 0000000000

Abstract: Low-light images suffer from severe noise and low illumination. Current deep learning models trained with real-world images exhibit excellent noise reduction capabilities, but require manual selection of a ratio parameter to complete the enhancement pipeline. In this work, we propose an adaptive low-light raw image enhancement network to avoid parameter handcrafting and improve image quality. The proposed method consists of two sub-models: Brightness Prediction (BP) and Exposure Shifting (ES). The former controls the brightness of the resulting image by estimating a guideline exposure time t . The latter learns to approximate an exposure-shifting operator ES, converting a low-light image with real exposure time t to a noise-free image with guideline exposure time t . Additionally, structural similarity (SSIM) loss and Image Enhancement Vector (IEV) are introduced to promote image quality, and a new Campus Image Dataset (CID) is proposed to overcome the limitations of existing datasets and supervise the training of our model. Quantitative tests demonstrate that the proposed method achieves the lowest Noise Level Estimation (NLE) score compared with BM3D-based low-light algorithms, suggesting superior denoising performance. Furthermore, these tests illustrate that our method can adaptively control global image brightness according to scene content. Finally, potential applications in video processing are briefly discussed.

Introduction

Images play an irreplaceable role in industry, military, and entertainment. As the art of light, obtaining better images in low-light environments has been extensively studied in the literature. While advanced photography equipment can help, it comes at considerable cost. Post-processing of low-light images faces fewer limitations, but remains challenging due to problems such as noise and color distortion.

Exposure time, ISO (which measures image sensor sensitivity), and aperture are known as the three pillars of photography. Extended exposure time can easily solve the low-light problem, but this is often impractical because the camera may not be fixed or the scene may contain moving objects. Higher ISO introduces more noise and is not always available on mobile devices. As a flexible solution, low-light image enhancement provides an alternative for imaging in low-light environments.

Current low-light image enhancement methods can be generally classified as follows:

Classic low-light image enhancement methods without convolutional neural networks. Histogram equalization (HE) [1] and gamma correction [2] provide simple solutions for low-light image processing. Based on Retinex theory [3], Single-scale Retinex (SSR) [4], Multi-scale Retinex (MSR) [5], SRIE [6], and LIME [7] were proposed, enhancing low-light images by estimating illumination and reflectance maps. Dehazing methods [8] can also be applied to this problem by treating low-light images as inverted hazed images.

With the rapid development of deep neural networks, researchers have combined classic low-light enhancement methods with Convolutional Neural Networks (CNNs). Chen et al. proposed Retinex-Net [9] to decompose low-light images into illumination and reflectance components, using BM3D [10] for denoising.

These methods assume low-light images are noise-free or that noise has already been removed by an additional denoising process. Consequently, when applied to real-world low-light images, an extra denoising step is required. However, most low-light images suffer from severe noise, and traditional denoising methods like BM3D cannot provide satisfactory results. To address image noise, researchers have also explored deep learning approaches [11–14] for denoising.

Some low-light images are barely visible before post-processing, as they are captured in extremely dark environments or with very short exposure times. In this work, we refer to these as extreme low-light images. While classic methods cannot tackle the severe noise and serious color distortion in these images, we discovered that with deep learning and raw-format image datasets, satisfactory results can still be obtained. Chen et al. proposed an end-to-end solution for raw low-light image enhancement [15], addressing both low-light problems and denoising in a single model. However, this approach requires an extra brightness amplification ratio and manual parameter tuning for different scenes, achieving excellent denoising effects but losing the flexibility and adaptability of classic methods.

Currently, all datasets available for low-light enhancement models select a longer-exposed image as the ground-truth during training, heavily relying on the dataset collector' s experience to choose exposure parameters such as exposure time, ISO, and white balance mode. These parameters are usually not optimal when considering scene content, environmental illumination, etc., which can lead to undesired trained models.

It is difficult for deep CNNs to learn how we determine the best exposure parameters for ground-truth images because this process is highly subjective and lacks specific standards, leading to more problems when multiple collectors are involved. On the other hand, establishing a baseline for ground-truth selection is equally difficult. Nevertheless, it is possible to predict how an image varies when exposure parameters change. We name this process Exposure Shifting (ES). With a learned ES model, it becomes possible to reversely study what good exposure parameters (e.g., exposure time) should be for each low-light

image, utilizing the characteristics of the back-propagation algorithm.

The contributions of our work are summarized as follows:

1) We present the Campus Image Dataset (CID), which contains a variety of scenes captured with multi-level exposure. CID provides powerful support for training our data-driven low-light enhancement model.

2) We propose an adaptive two-stage low-light image enhancement model that provides state-of-the-art low-light noise suppression as well as adaptive global brightness control:

- In stage one, a Brightness Prediction Network (BPN) estimates proper exposure time based on image content and Exif (Exchangeable image file) metadata. BPN is trained to provide adaptive brightness control during enhancement and eliminate external parameters required in the Exposure Shifting stage.
- In stage two, an Exposure Shifting Network (ESN) estimates a longer-exposed version of the image with target exposure time estimated by BPN. Moreover, ESN completes image denoising and produces the final RGB result image.

The rest of the article is organized as follows. The Related Works section reviews previous research. The Dataset section demonstrates the importance and advantages of our CID dataset. The Method section details our model, training procedures, and design concepts. The Experiments section compares our model with other state-of-the-art methods from multiple perspectives, including denoising, adaptive brightness control performance, and computational cost. Potential extensions to video processing are briefly discussed. The Conclusion and Discussion section summarizes the advantages and limitations of our model and presents future research directions.

Related Works

Our proposed method draws inspiration not only from deep neural network applications but also from classic methods based on histogram equalization, dehazing, and Retinex models.

2.1 Histogram Equalization Methods

The Histogram Equalization (HE) method is extensively used for digital image processing. The basic idea is to improve image visibility by transforming the image histogram into a uniform distribution, which effectively increases image entropy for low-light images. Reference [16] argues that applying HE can yield a color-invariant representation. However, HE tends to cause noise amplification, detail disappearance, and color distortion in low-light image enhancement.

Due to these drawbacks, numerous improvements have been proposed. Adaptive Histogram Equalization (AHE) [17] and its variation Contrast-limited Adaptive

Histogram Equalization (CLAHE) [18] perform histogram equalization in regions surrounding each pixel rather than across the entire image. Hue-preserving color image enhancement [19] focuses on contrast enhancement while preserving hue, while [20] and [21] address preserving original image luminance.

2.2 Dehazing and Retinex Model-Based Methods

Many low-light enhancement methods are based on dehazing and Retinex models. It is observed that low-light images and inverted haze images share many similarities. For this reason, [8] presented a method where low-light images are inverted, dehazed, and then inverted back. This method was further studied by [22] and [23].

On the other hand, the Retinex model [3] reveals that a natural image consists of an illumination map and a reflectance map. With Retinex decomposition and the assumption that the illumination map is smooth, researchers have proposed single-scale Retinex [24] and multi-scale Retinex [4]. Based on image fusion, adjustments can be applied to the illumination map to improve performance [25]. Reference [26] focused on preserving natural characteristics with a Bright-Pass filter. In [27], a variational Retinex model was proposed, formulating the illumination estimation problem as a Quadratic Programming optimization problem. Reference [28] estimates the illumination map considering local structure, while [29] focused on revealing structure details from the reflectance map. A recent study shows that the Retinex model can be combined with the atmospheric scattering model [30]. More refined models were presented by [31, 32] based on variational Retinex theory.

However, noise modeling is not well established in these algorithms. Reference [8] assumes noise is insignificant and can be removed before or after the dehazing stage. Retinex-based models rely on classic noise reduction algorithms (e.g., BM3D) to reduce noise in the reflectance map. Therefore, they can achieve good results on mild low-light images but tend to amplify noise when processing severe or extreme low-light images. Besides, the variational Retinex model is very time-consuming as it requires many iterations to solve optimization objectives, as do many denoising algorithms, making real-time processing difficult.

2.3 Data-driven Methods

Image denoising and low-light enhancement can be realized through data-driven approaches, either separately or integrated.

Noise reduction has been extensively studied. Reference [11] discovered that convolutional neural networks can compete with the state-of-the-art classic denoising algorithm BM3D. Data-driven denoising research has been extended by [14, 33, 34]. Although these works were not targeted at low-light image processing, they provide references for other image processing studies.

The first application of the data-driven method is LLNET [12], which utilizes a

deep auto-encoder to learn from synthetic low-light image datasets, integrating denoising into the low-light enhancement process. It was also demonstrated that Retinex-based methods can be further improved by deep learning [9, 35], but these works still have undesirable denoising performance because synthetic datasets cannot reflect the characteristics of real low-light images.

To address the drawbacks of synthetic datasets, a large multi-exposure image dataset was proposed in [36], but [15] found that models utilizing raw-format images can provide much better results for low-light enhancement. Reference [15] suggested that state-of-the-art results can be achieved using raw-format paired images and end-to-end training. However, since the paired images have different exposure times, a ratio was introduced as additional input to bridge the gap. Thus, after training, with no reference to indicate the ratio, [15] requires manual adjustment for every different image to be enhanced, resulting in practical drawbacks.

Dataset

For extreme low-light image enhancement, few optional datasets exist. The Google HDR+ dataset [37] and Darmstadt Noise Dataset [38] avoid the disadvantages of synthetic datasets but were mainly collected in environments with sufficient illumination. The RENOIR [39] and Learning-to-See-In-the-Dark dataset (SID) [15] provide real low-light image pairs captured by carefully selecting ISO and exposure time for each scene. The Exclusively Dark Dataset (EDD) [40] is another low-light dataset with object-level annotations targeting object recognition. However, these datasets cannot satisfy our study' s needs.

In our work, to adaptively enhance images with different noise levels, we must consider how exposure parameters impact low-light images. More specifically, as environmental illumination decreases, low-light images become gradually overwhelmed by noise and invalid pixels. While it is not flexible to manipulate environmental illumination, camera exposure time settings can be easily changed to simulate this illumination shift.

Therefore, we require a dataset containing not just image pairs but multiple multi-exposed image series, where each series contains low-light images at different levels and one reference image captured in the same scene.

Overall, three conditions must be met for a satisfactory dataset: - Real scenes: Unlike synthetic images, photos captured in real scenes reflect much more diverse noise and distortion patterns - Multiple exposure levels: Using numerous multi-exposed image series to train an adaptive model capable of enhancing low-light images at different levels - Raw-format images: Most data captured by the camera sensor is lost in format conversion, and this lost information is essential for extreme low-light image enhancement

Based on these requirements, we propose our Campus Image Dataset (CID). CID contains approximately 200 image groups, each consisting of 8 raw images

with different exposure times shot continuously in the same scene using a tripod (Fig. 3). Specifically, for each group, an upper exposure time limit was selected to ensure the first image in the sequence was slightly overexposed, then a lower limit was chosen to capture an extreme low-light image as the last image in the sequence, and the gap between limits was filled with 6 additional images. Finally, the ground-truth image for each group was manually selected, and invalid images containing only noise due to extremely short exposure were tagged and excluded. A demonstration of scenes in CID is presented in Fig. 1.

Method

4.1 Method Overview

Most existing methods cannot suppress the severe noise present in extreme low-light images. Recent studies such as [15] have focused on extreme low-light enhancement using raw-format images and real-scene datasets in an end-to-end manner. However, unlike classic methods, the brightness of enhanced images cannot be automatically controlled in this framework—specifically, an external ratio must be manually adjusted when the input image lacks a paired reference image.

Our aim is to enhance extreme low-light images while combining the advantages of adaptive brightness control and superior denoising performance. We frame low-light enhancement as a special case of changing exposure time. With t and t representing exposure times, the low-light image X can be interpreted as a normal image Y corrupted by a hypothetical exposure-shifting operator ES, which only alters exposure time while keeping aperture and ISO unchanged:

BX = ES(Y , t)

In this process, more stochastic noise N is generated by operator ES because t < t , corresponding to the low PSNR (Peak Signal to Noise Ratio) value in low-light images. Conversely, the corrupted low-light image can also be restored by the same operator ES:

ˆY = ES(X , t)

The noise N is reduced because t > t , allowing reconstruction of the normal image ˆY .

The latter process, named Exposure Shifting (ES), can enhance low-light images while suppressing noise. More importantly, it can be learned in a data-driven, end-to-end manner, thus achieving superior denoising performance. However, a proper guideline exposure time t must always be provided for reconstruction of ˆY . To make our model adaptive, a Brightness Prediction (BP) procedure is introduced for guideline time estimation.

In brief, the extreme low-light enhancement process is split into two procedures. First, for a raw-format low-light image (with real exposure time t), an optimal guideline exposure time t is estimated via a sub-model called Brightness

Prediction Network (BPN). Second, the final RGB-format enhanced image is obtained by another sub-model, namely Exposure Shifting Network (ESN), using the estimated guideline time t to control result brightness.

In addition to the image itself, extra data are involved in both BPN and ESN, such as white balance index, ISO, t and t . The Image Enhancing Vector (IEV) is proposed to encode this additional data.

Image Enhancing Vector: Raw image array is the original data captured by the camera. Other key parameters, such as camera white balance, ISO, exposure time, and date/time, are stored with the raw image array as Exif metadata. In HDR imaging research [41], exposure time and ISO information have played essential roles in calculating response curves and LDR-to-HDR conversion. Although it is not necessary to explicitly train the network to learn the camera response curve, feeding these parameters as network input avoids underfitting caused by insufficient features.

Image Enhancing Vector (IEV) introduces extra features into convolutional neural networks. In this paper, IEV consists of ISO (u), white balance indices for 4 raw image channels (w , w_g, w_b, w_g), exposure time of the low-light image t , and guideline exposure time t . Moreover, other Exif metadata can be appended to IEV to improve network performance when necessary, e.g., aperture and focal length.

IEV is used as network input in both BPN (for t estimation) and ESN (for exposure shifting), with a difference. As shown in Fig. 2, in the ES procedure the IEV can be written as:

V $\rightarrow$ = V(t , t) = (w , w_g, w_b, w_g , u , ···, t , t)

where (w , w_g, w_b, w_g) are the white balance indices of 4 raw image channels and u is the ISO value (controlling camera sensitivity).

In the BP procedure, t is removed from IEV. It is renamed partial IEV (pIEV) to indicate the difference:

p = V (t) = (w , w_g, w_b, w_g , u , ···, t)

IEV and pIEV are fed into the convolutional neural network as additional channels of the raw image array. For example, the first extra channel provided by IEV is a uniform image filled with pixel value w .

Exposure Shifting: Exposure Shifting (ES) is the second procedure in our model but is trained first. U-NET [42] is selected as the structure for Exposure Shifting Network (ESN). Compared with fully convolutional networks, the U-NET architecture reduces GPU memory usage, enabling easy processing of high-resolution images.

Recall that the basic unit of our Campus Image Dataset (CID) is an image group, consisting of 8 raw-format images with different exposure times captured in the same scene. Let (X_R, Y_R) be paired raw images extracted from one group,

where Y_R is the pre-selected ground-truth image and X_R is a randomly selected low-light image. More specifically, X_R is randomly chosen within the image group, excluding Y_R and over-exposed images (Fig. 3).

Moreover, the exposure time of X_R is denoted as t . The corresponding RGB-format image of Y_R is denoted as Y, with its exposure time as t_g.

ESN is expected to learn the exposure-shifting operator ES from numerous paired images. Specifically, ESN must estimate an image ˆY that resembles Y as closely as possible. In this way, ESN successfully transforms the raw-format low-light image X_R into an RGB-format longer-exposed version of itself. More importantly, the severe noise in X_R can also be removed. This procedure can be written as:

(cid:16) = F_ES ˆY Θ X_R, V → _g V → _g = V(t , t = t_g) (cid:17) (cid:12) (cid:12) (cid:12)Θ

where ESN is denoted as F_ES and its parameters as Θ . Note that since the BP procedure is not yet involved, the guideline exposure time t is replaced by t_g in IEV V → _g. The enhanced result image is represented by ˆY Θ .

Two metrics guide the training process of the ESN model: Mean Square Error (MAE) and multi-scale Structural Similarity (SSIM) [43]. The former serves as a simple but effective indicator evaluating general similarity between the estimated and ground-truth images. The latter is a perceptual metric more sensitive to visible structures. Both are full-reference metrics. Let L_MAE be the MAE loss and L_SSIM be the SSIM loss.

The image width and length are denoted as m and n respectively. With subscript t_g omitted, the MAE loss L_MAE and SSIM loss L_SSIM are defined as:

L_MAE (cid:16) ˆY Θ , Y (cid:17) = M AE (cid:16) ˆY Θ , Y (cid:17) (cid:88) (cid:88) (cid:12) (cid:12) (cid:12) ˆY Θ _ij − Y_ij (cid:12) (cid:12) (cid:12)

L_SSIM (cid:16) ˆY Θ , Y (cid:17) = 1 − SSIM (cid:16) ˆY Θ , Y (cid:17)

The final MAE loss is the average of the channel MAE loss; for simplicity, image channels are not presented in the equation. The calculation of multi-scale structural similarity (SSIM) can be referred to in [43]. With subscript t_g omitted, the loss function is defined as the linear combination of L_MAE and L_SSIM:

L_ES = L_ES (cid:16) ˆY Θ , Y (cid:17) = (1 −) L_MAE (cid:16) ˆY Θ , Y (cid:17) + L_SSIM (cid:16) ˆY Θ , Y (cid:17) (7)

where is a constant and 0 (cid:54) < 1. The influence of will be discussed in the Experiments section.

The ESN network parameters Θ are learned by minimizing L_ES over K pairs of images in the training set:

1 = argmin (cid:88) (cid:16) ˆY Θ _k , Y_k (cid:17)

The obtained sub-model F_ES approximates the exposure-shifting operator ES. However, for a low-light image outside the training set, no ground-truth counterpart exists, making the guideline time t in IEV V → absent. This leads to the Brightness Prediction (BP) sub-model and guideline time estimation.

Brightness Prediction: The Brightness Prediction (BP) procedure provides estimation of guideline exposure time t :

1 = F_BP (cid:16) X_R, V (cid:17) (cid:12) (cid:12) (cid:12)Θ

where F_BP and Θ represent BPN and its parameters respectively. The pIEV is represented by V ^p = V_p(t). This is the first enhancement procedure in our model but is trained after ESN. As illustrated in Fig. 4, the Brightness Prediction Network (BPN) is a relatively small network consisting of multiple convolution layers followed by several fully connected layers. Due to the fully connected layers, the input image X_R width and length should be uniformly resized to 512 × 512.

With the estimated t ^Θ obtained in this BP procedure, the t_g item in IEV from the ES procedure can be replaced by t ^Θ . Different from the derivation in (5), the new final result image can be formulated as:

^Y Θ ^Y Θ |Θ∗_1 = F_ES (cid:18) X_R, V → ^Θ (cid:19) (cid:12) (cid:12) (cid:12) (cid:12)

An approach to train BPN is then required. To begin, the purpose of BPN is to control the brightness of the final result image, achieved by adjusting the guiding exposure time t in the ES procedure. Since image pairs (X_R, Y_R) and their exposure times (t , t_g) are available, it might seem intuitive that BPN could simply learn from (X_R, t_g) in an end-to-end manner, with X_R as input and t_g as label. However, this cannot be achieved, for reasons discussed in the next subsection.

The proposed BPN training approach and loss function L_BP are as follows. We define the brightness of a single pixel B_r as the average pixel value across all color channels. Given a certain dark scene for image capture with fixed ISO, aperture settings, and scene, pixel brightness is determined solely by exposure time. Let R be the pixel irradiance in this scene; when camera exposure time t increases, pixel Exposure E(R, t) increases as well [44]:

E(R, t) = Rt

Thus the pixel becomes brighter. As inspiration, it is feasible to design loss function L_BP based on pixel brightness.

It should be noted that the relationship between pixel brightness B_r(E) and Exposure E = Rt is not linear. Instead, it is typically described by an S-shaped curve because pixel brightness is limited within [0, 1] (or [0, 255] for 8-bit image storage), and the camera sensor is less sensitive to changes in t when pixel values approach 0 or 1. In other words, as exposure time t changes, some pixels are not sensitive and exhibit little brightness change compared to other pixels.

For example, when photographing a street at night, pixels revealing lamp bulbs are always saturated unless t is extremely small, while pixels representing the dark sky are always close to zero unless affected by noise. If these pixels are involved in BPN training, they can cause gradient vanishing in backpropagation because B_r(E)/ t = B_r(E)/ E · R, and as revealed in the S-shaped curve, B_r(E)/ E is close to 0. On the other hand, when pixel brightness B_r is around 0.5, B_r/ E reaches its maximum. Pixels satisfying this condition can be collected into an Area of Interest (AoI), which identifies pixels sensitive to the Exposure Shifting process.

First, in the ground-truth image Y, an AoI weight map W can be obtained by filtering out pixels with brightness near 0.5. Specifically, the ground-truth image Y is converted to a single-channel grayscale image Y_G, then a Gaussian curve with mean _w = 0.5 and variance ²_w = 0.01 is applied to each pixel of Y_G:

W_G = exp (cid:33) (Y_G − _w)²

Then W_G is normalized, resulting in the AoI weight map W. The value at position (i, j) is denoted with subscripts, namely W_G,ij and W_ij:

W_ij = W_G,ij (cid:80)^n_j=1 W_G,ij (cid:80)^m, 1 (cid:54) i (cid:54) m, 1 (cid:54) j (cid:54) n

Combining Equation 12 and 13, they can be simplified by the softmax function:

W = softmax (cid:33) (Y_G − _w)²

where ²_v = 0.04 is another variance constant. The image value at position (i, j) is represented by i, j subscripts. Finally, BPN is trained by optimizing Θ over K pairs of images in the training set:

2 = argmin (cid:88) (cid:16) ˆY Θ |Θ∗ (cid:12) (cid:12)Θ ; Θ∗ (cid:12) (cid:17)

4.5 Method Summary and Discussion

Algorithm 1 summarizes the training process of our model, together with the method to evaluate and apply this model.

Algorithm 1: Model training and evaluation

1: Initialize ESN, BPN parameters Θ and Θ
2: Prepare CID dataset
3: function TRAIN(Dataset)
4: for Epoch_Train_ESN = 1 → E do
5: for Image Pair k = 1 → K do
6: Read raw image pairs (X_R,k, Y_R,k) and V → _g
7: Convert to RGB format. Y_k ← Y_R,k
8: Compute ˆY Θ _k
9: Θ ← argmin_Θ via Equation 5
10: end for

11: end for
12: Θ∗_1 ← Θ
13: for Epoch_Train_BPN = 1 → E do
14: for Image Pair k = 1 → K do
15: Read raw image pair (X_R,k, Y_R,k) and V
16: Compute t ,k^Θ and ^Y Θ |Θ∗_1
17: Convert to grayscale image. Y_G,k ← Y_R,k
18: Convert to grayscale image. ^Y Θ |Θ∗_1_G,k ← ^Y Θ |Θ∗_1
19: Compute W_k via Equation 14
20: Θ ← argmin_Θ
21: end for
22: end for
23: Θ∗_2 ← Θ
24: return Θ∗_1, Θ∗_2
25: end function
26:
27: function EVALUATE(Image)
28: Read raw-format image X_R
29: Read all elements in V ^p from raw-image file
30: BPN: Compute t ^Θ∗_2 ← F_BP (cid:16) X_R, V ^p (cid:12)Θ∗ (cid:12) (cid:17)
31: ESN: Compute ^Y Θ∗_2|Θ∗_1 ← F_ES (cid:18) X_R, V → (cid:19) (cid:12) (cid:12) (cid:12) (cid:12)
32: return ^Y Θ∗_2|Θ∗_1
33: end function

As mentioned earlier, training BPN with input X_R and label t_g is straightforward but not feasible. First, let F∗ be the hypothetical optimal BPN model, then t∗_BP(X_R) is the optimal estimation of guideline exposure time. For an image-time label pair (X_R, t_g), t_g can be considered a sample from a Gaussian distribution: t_g N(t∗_1(X_R), ²_c). The variance ²_c in this process is too large because: - In every CID image group, the ground-truth image is chosen manually and dominated by subjective human judgment - The ground-truth image is selected from only 8 candidate images, but the optimal t∗_1 exists somewhere in the time continuum (0, ∞)

Furthermore, the number of samples is very limited since each image group can only provide one t_g sample. Consequently, this straightforward approach cannot be applied to BPN training due to high label variance and limited samples.

Experiments

Limited by GPU memory, we did not adopt Batch Normalization [45] technique. To accelerate training, before the process started, each channel of the input image and each element of IEV were normalized along the sample dimension. The means and variances used in normalization then became invariant constants. Additionally, training images were randomly cropped into 512 × 512 patches.

Both ESN and BPN were trained with Adam optimizer [46] following Algorithm 1. ESN was trained first with learning rate slowly descending from 2×10 to 1×10 . After approximately 300 epochs, when training set loss became stable, ESN parameters were made untrainable. Then BPN was appended to the training graph and trained with the same learning rate. Considering BPN may require global information to determine proper exposure time, we used larger patches but avoided whole-image training to prevent overfitting. Although loss had to propagate through the deep ESN before reaching BPN, training proceeded smoothly and completed after 100 epochs.

5.2 Content-Based Brightness Control

We first highlight the adaptive brightness control feature of our method. Here we explicitly define brightness as the mean value of an image Y, denoted as Br(Y). We expect dynamic adjustment of image brightness based on image content. More specifically, while extreme low-light images suffer from severe noise, color distortion, and are barely visible before post-processing, mild low-light images only have moderate noise and relatively lower brightness compared to normal images. Thus, enhancement algorithms need to make adaptive adjustments for different inputs. Currently, classic methods such as Retinex- and Dehazing-based algorithms perform poorly on extreme low-light images. Meanwhile, learning-based methods targeting end-to-end denoising can process various low-light images but lack flexibility and adaptability because they either require manual brightness amplification tuning or simply serve as a denoising component in other low-light methods. Our proposed model combines the advantages of both approaches and can process low-light images at different levels.

To evaluate adaptability, we apply the trained model to enhance all images in the test set. Recall that the image group is the basic unit of our CID dataset; accordingly, instead of evaluating single images, we investigate collaborative characteristics across different images within each group. Note that group identity is NOT involved in model training and only serves as a tag to gather resulting images for further analysis.

For a sequence of multi-exposed images (X_R,1, X_R,2, ···) with increasing exposure time, we expect enhanced images to possess similar brightness, namely Br(ˆY) Br(ˆY) ···, where ˆY denotes the enhanced image.

Fig. 5(a) illustrates a busy-road scene. The first row displays RGB images produced by the camera; the second row shows enhanced results using our method on individual raw images. Note that the first image in this group is exposed for 1/20 seconds, causing motion blur on the truck, while the last image is exposed for only 1/100 seconds. All four images have approximately uniform brightness after enhancement, despite changing t and moving vehicle headlights. Fig. 5(b) presents our method' s performance under extreme low-light conditions. The last image in the sequence suffers from both low environmental illumination and short exposure, with object colors and shapes drowned out by noise. How-

ever, the resulting image has reduced noise and maintains acceptable global brightness.

For quantitative analysis, we gathered images belonging to the same groups and calculated brightness for all images to observe brightness shifts. Let (Br , Br , ···)_k be the brightness of each unenhanced low-light image in the k-th image group, and let (Br , Br , ···)_k be that of the enhanced ones. Then the mean value and CV (coefficient of variation) of (Br , Br , ···) are computed and denoted as (, cv)_k; correspondingly, (, cv)_k is calculated from (Br , Br , ···)_k. In Fig. 8, we plotted (, cv) and (, cv) for all groups. To illustrate changes after enhancement, for each group (, cv) (marked with stars) and (, cv) (marked with circles) are linked with dotted lines.

Fig. 8 shows that for different groups captured from diverse scenes, enhanced images all achieve increased brightness and significantly reduced CV. Yet brightness growth between groups can vary considerably, ranging from approximately 60% to over 1,000%. Together with Fig. 5, Fig. 8 indicates that the BP sub-model can adaptively control global image brightness according to not only original image brightness but also scene content. The dramatic CV reduction suggests that enhanced images within each group share almost identical brightness despite varying exposure time t . Specifically, our model can control group brightness CV under 0.29 on 76.3% of test set groups and 98.5% of training set groups, considering the influence of diverse scenes, illumination conditions, capture ISO, etc.

When designing our model, we expect BPN to extract scene content features and estimate a reasonable t accordingly. To clearly understand how this black box works, Fig. 9 illustrates the relationship between real exposure time t , BPN-estimated guideline exposure time t , and enhanced image brightness Br(ˆY). Similar to Fig. 8, we use group averages to simplify the figure, where each point corresponds to an image group. As shown in Fig. 9, the network tends to estimate t positively associated with t . Furthermore, for groups with identical average t , scene content impacts t prediction. Specifically, if a group' s scene contains more relatively bright areas, a smaller t is more likely adopted to prevent overexposure, and vice versa. A few exceptions exist, suggesting the BPN sub-model has learned more sophisticated rules within the neural network black box.

5.3 Denoising Evaluation

We compare our method with four classic and state-of-the-art methods: multi-scale Retinex (MSR) [5], illumination map estimation based (LIME) [7], deep Retinex decomposition (Retinex-Net, R-Net) [9], and Learning-to-See-In-the-Dark (SID) [15]. For fair comparison, 3D transform-domain filtering (BM3D) [10] is applied to MSR, LIME, and R-Net, as these methods do not model noise and require extra denoising. Additionally, we train the SID model on the CID dataset and manually select ratio for SID since it does not provide automatic

brightness estimation.

Recall that BPN utilizes a reduced-reference evaluation index as its training loss. In this way, exposure parameters of the ground-truth image, chosen through highly subjective human judgment, have much less influence on BPN training. However, as a result of this novel technique, PSNR and SSIM evaluation cannot be adopted because no reference image exists. We provide perceptual comparisons in Fig. 6 and evaluate our method with no-reference quality metrics.

Table 1 compares our method with MSR, LIME, and R-Net (Retinex-Net) using several no-reference quality metrics: Statistical Noise Level Estimation (SNLE) [47], Noise Level Estimation (NLE) [48], Noise Variance (NV) [49], and image entropy. SNLE, NLE, and NV metrics can estimate noise level from a single image, with clean images having relatively smaller values than noisy ones. SNLE estimates noise variance by observing image eigenvalues, NLE applies principal component analysis (PCA) on special image patches, and NV is motivated by the fact that clean and noisy images have different sensitivity to Laplacian operations. Image entropy reflects the average information amount in an image; it is the simplest image evaluation metric, and images with proper overall brightness have larger entropy.

We evaluated approximately 400 CID test images, excluding the SID method because it requires external manual ratio selection for every image. As shown in Table 1, our method demonstrates superior performance on Noise Level Estimation. Moreover, no-reference metrics use very different standards to assess image quality. MSR+BM3D, R-Net+BM3D, and LIME+BM3D methods have undesired scores on one or multiple indices, while our method performs well on all metrics.

Fig. 6 illustrates that our method and SID provide more significant noise removal improvements than the classic BM3D method and can adaptively denoise images with diverse noise levels. Furthermore, our results benefit from the white balance index introduced in IEV and avoid abnormal colors in extreme low-light images.

However, CNN-based methods tend to blur objects with sharp edges (e.g., text). By introducing SSIM into the loss function, our method shows more promising results on images containing text and advantages over SID in preserving edge information, as seen in the first row of Fig. 6. On the other hand, controlled experiments in Fig. 7 reveal that the SSIM component ratio in loss function L_ES must be constrained to avoid noise being interpreted as texture and edges by ESN.

5.5 Computational Cost

The proposed algorithm was evaluated on ModelArts cloud service equipped with 8 vCPUs and an Nvidia-P100 GPU (16GB GPU memory). Execution time for a high-resolution image (3968×2976) averages 0.27 seconds, show-

ing only minor increase compared to the SID algorithm (0.21 seconds). More specifically, execution times for BPN and ESN sub-models are 0.05 and 0.22 seconds respectively. For comparison, BM3D (GPU implementation [50]) takes 4.32 seconds to complete denoising, making other low-light methods dependent on BM3D for denoising—including MSR, LIME, R-Net, etc.—much less efficient on GPU-equipped devices.

Conclusion and Discussion

6.1 Conclusion

This paper proposes an adaptive raw image enhancement model for extreme low-light image processing. The two-stage framework provides adaptability to images with different scenes and exposure parameters. We also present the CID dataset for model training and evaluation. Experimental tests show our model can provide state-of-the-art low-light image denoising and adaptive global brightness control.

Our method has many advantages over existing approaches. First, no external parameters are needed after model training—a single raw image and its Exif metadata (e.g., white balance, ISO, exposure time) are sufficient to complete the process. Therefore, our model can process large batches of raw images without parameter handcrafting. Second, our model significantly outperforms methods where denoising works as a separate step, because instead of simply implementing noise-suppression algorithms (e.g., BM3D) before or after the main low-light enhancement procedure, we integrate denoising as part of our model, allowing the ESN module to learn exposure shifting and denoising simultaneously in an end-to-end manner.

Quantitative experiments adopted no-reference image quality metrics including SNLE, NLE, NV, and image entropy to test denoising performance. Our model achieves the smallest noise variance on the NLE metric compared to BM3D-based low-light methods and obtains near-optimal scores on the other three indices. We also reveal potential applications in low-light video processing.

The BPN module enables our model to process images with diverse exposure levels, from extreme low-light to mild low-light images. However, for video processing applications, additional challenges must be addressed—e.g., frame brightness should change continuously and avoid fluctuation. We experimented on continuous raw image series; despite using no adjacent frame information, BPN output changes continuously with illumination conditions and shows no fluctuation.

6.2 Limitations and Future Works

The proposed model is limited by its dataset, a common problem in data-driven methods. On one hand, training with raw-format, multi-exposed images significantly improves image quality. On the other hand, all reference ground-truth

images are captured in real scenes, sometimes containing minor defects such as noise, halos around light sources, local over-exposure, and white balance deviation. Denoising performance may be further improved if these drawbacks can be avoided.

Additionally, CID image groups have not covered enough natural scenes. For example, green grass and trees tend to have low color saturation in resulting images because most training set images were collected in winter when few green plants appear in scenes. For future study, the raw-format CID dataset can be extended using unprocessing [14] techniques to generate raw-format images from other available online datasets.

Another limitation is algorithm memory consumption. Batch size is constrained because training the proposed model requires substantial GPU memory, making model training time-consuming on GPUs with limited memory. We plan to improve network architecture and investigate possibilities for mobile device applications.

As a future direction, the model can be further modified by introducing HDRI (High Dynamic Range Imaging). Since our CID dataset comprises multi-exposed images, we can calculate an HDR image for each scene. These HDR images can then be used as ground truth for ESN training.

References

[1] Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., et al.: 'Adaptive histogram equalization and its variations', Computer vision, graphics, and image processing, 1987, 39, (3), pp. 355–368

[2] Huang, S.C., Cheng, F.C., Chiu, Y.S.: 'Efficient contrast enhancement using adaptive gamma correction with weighting distribution' , IEEE Trans. on image processing, 2012, 22, (3), pp. 1032–1041

[3] Land, E.H.: 'The retinex' , American Scientist, 1964, 52, (2), pp. 247–264

[4] Jobson, D.J., Rahman, Z., Woodell, G.A.: 'Properties and performance of a center/surround retinex' , IEEE Trans. on Image Processing, 1997, 6, (3), pp. 451–462

[5] Jobson, D.J., Rahman, Z.u., Woodell, G.A.: 'A multiscale retinex for bridging the gap between color images and the human observation of scenes' , IEEE Trans. on Image processing, 1997, 6, (7), pp. 965–976

[6] Fu, X., Zeng, D., Huang, Y., Liao, Y., Ding, X., Paisley, J.: 'A fusion-based enhancing method for weakly illuminated images' , Signal Processing, 2016, 129, pp. 82–96

[7] Guo, X., Li, Y., Ling, H.: 'Lime: Low-light image enhancement via illumination map estimation.' , IEEE Trans. Image Processing, 2017, 26, (2), pp. 982–993

[8] Dong, X., Wang, G., Pang, Y., Li, W., Wen, J., Meng, W., et al.: 'Fast efficient algorithm for enhancement of low lighting video' . IEEE Int. Conf. on Multimedia and Expo, Barcelona, Spain, 2011. pp. 1–6

[9] Wei, C., Wang, W., Yang, W., Liu, J.: 'Deep retinex decomposition for low-light enhancement' , arXiv preprint arXiv:180804560, 2018

[10] Kostadin, D., Alessandro, F., Vladimir, K., Karen, E.: 'Image denoising by sparse 3-d transform-domain collaborative filtering' , IEEE Trans. on Image Processing, 2007, 16, (8), pp. 2080–2095

[11] Burger, H.C., Schuler, C.J., Harmeling, S.: 'Image denoising: Can plain neural networks compete with bm3d?' . IEEE Conf. on computer vision and pattern recognition, Providence, RI, USA, 2012. pp. 2392–2399

[12] Lore, K.G., Akintayo, A., Sarkar, S.: 'Llnet: A deep autoencoder approach to natural low-light image enhancement', Pattern Recognition, 2017, 61, pp. 650–662

[13] Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: 'Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising' , IEEE Trans. on Image Processing, 2017, 26, (7), pp. 3142–3153

[14] Brooks, T., Mildenhall, B., Xue, T., Chen, J., Sharlet, D., Barron, J.T.: 'Unprocessing images for learned raw denoising' . Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019. pp. 11036–11045

[15] Chen, C., Chen, Q., Xu, J., Koltun, V.: 'Learning to see in the dark' . Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018. pp. 3291–3300

[16] Finlayson, G., Hordley, S., Schaefer, G.: 'Illuminant and device invariant colour using histogram equalisation' , Pattern Recognition, 2005, 38, pp. 179–190

[17] Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., et al.: 'Adaptive histogram equalization and its variations', Graphical Models and Image Processing, 1987, 39, (3), pp. 355–368

[18] Zuiderveld, K.J.: 'Contrast limited adaptive histogram equalization', Graphics gems, 1994

[19] Naik, S.K., Murthy, C.A.: 'Hue-preserving color image enhancement without gamut problem', IEEE Trans. on Image Processing, 2003, 12, (12), pp. 1591–1598

[20] Wang, Y., Chen, Q., Zhang, B.: 'Image enhancement based on equal area dualistic sub-image histogram equalization method' , IEEE Trans. on Consumer Electronics, 1999, 45, (1), pp. 68–75

[21] Kim, Y.: 'Contrast enhancement using brightness preserving bi-histogram equalization' , IEEE Trans. on Consumer Electronics, 1997, 43, (1), pp. 1–8

[22] Li, L., Wang, R., Wang, W., Gao, W.: 'A low-light image enhancement method for both denoising and contrast enlarging' , , 2015

[23] Malm, H., Oskarsson, M., Warrant, E.J., Clarberg, P., Hasselgren, J., Lejdfors, C.: 'Adaptive enhancement and noise reduction in very low light-level video' , , 2007

[24] Jobson, D.J., Rahman, Z., Woodell, G.A.: 'Properties and performance of a center/surround retinex' , IEEE Trans. on Image Processing, 1997, 6, (3), pp. 451–462

[25] Fu, X., Zeng, D., Huang, Y., Liao, Y., Ding, X., Paisley, J.: 'A fusion-based enhancing method for weakly illuminated images' , Signal Processing, 2016, 129, pp. 82–96

[26] Wang, S., Zheng, J., Hu, H., Li, B.: 'Naturalness preserved enhancement algorithm for non-uniform illumination images' , IEEE Trans. on Image Processing, 2013, 22, (9), pp. 3538–3548

[27] Kimmel, R., Elad, M., Shaked, D., Keshet, R., Sobel, I.: 'A variational framework for retinex' , Int. Journal of Computer Vision, 2003, 52, (1), pp. 7–23

[28] Hao, S., Feng, Z., Guo, Y.: 'Low-light image enhancement with a refined illumination map', Multimedia Tools and Applications, 2018, 77, (22), pp. 29639–29650

[29] Li, M., Liu, J., Yang, W., Sun, X., Guo, Z.: 'Structure-revealing low-light image enhancement via robust retinex model' , IEEE Transactions on Image Processing, 2018, 27, (6), pp. 2828–2841

[30] Gu, Z., Chen, C., Zhang, D.y.: 'A low-light image enhancement method based on image degradation model and pure pixel ratio prior' , Mathematical Problems in Engineering, 2018, 2018, pp. 1–19

[31] Park, S., Moon, B., Ko, S., Yu, S., Paik, J.: 'Low-light image enhancement using variational optimization-based retinex model' , , 2017

[32] Fu, G., Duan, L., Xiao, C.: 'A hybrid l2-lp variational model for single low-light image enhancement with bright channel prior' , , 2019. pp. 1925–1929

[33] Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: 'Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising' , IEEE Trans. on Image Processing, 2017, 26, (7), pp. 3142–3153

[34] Guo, S., Yan, Z., Zhang, K., Zuo, W., Zhang, L.: 'Toward convolutional blind denoising of real photographs' , , 2019

[35] Shen, L., Yue, Z., Feng, F., Chen, Q., Liu, S., Ma, J.: 'Msr-net: Low-light image enhancement using deep convolutional network' , arXiv preprint

arXiv:171102488,

[36] Cai, J., Gu, S., Zhang, L.: 'Learning a deep single image contrast enhancer from multi-exposure images' , IEEE Trans. on Image Processing, 2018, 27, (4), pp. 2049–2062

[37] Hasinoff, S.W., Sharlet, D., Geiss, R., Adams, A., Barron, J.T., Kainz, F., et al.: 'Burst photography for high dynamic range and low-light imaging on mobile cameras' , ACM Trans. on Graphics (TOG), 2016, 35, (6), pp. 192

[38] Xu, J., Li, H., Liang, Z., Zhang, D., Zhang, L.: 'Real-world noisy image denoising: A new benchmark' , arXiv preprint arXiv:180402603, 2018

[39] Anaya, J., Barbu, A.: 'Renoir - a dataset for real low-light noise image reduction' , arXiv preprint arXiv:14098230, 2014

[40] Loh, Y.P., Chan, C.S.: 'Getting to know low-light images with the exclusively dark dataset' , Computer Vision and Image Understanding, 2019, 178, pp. 30–42

[41] Reinhard, E., Ward, G., Pattanaik, S., Debevec, P.: 'High dynamic range imaging: acquisition, display, and image-based lighting' . (Morgan Kaufmann, 2005)

[42] Ronneberger, O., Fischer, P., Brox, T.: 'U-net: Convolutional networks for biomedical image segmentation' . Int. Conf. on Medical image computing and computer-assisted intervention, Munich, Germany: Springer, 2015. pp. 234–241

[43] Wang, Z., Simoncelli, E.P., Bovik, A.C.: 'Multiscale structural similarity for image quality assessment' . The Thirty-Seventh Asilomar Conf. on Signals, Systems & Computers, 2003, vol. 2. Pacific Grove, CA, USA: IEEE, 2003. pp. 1398–1402

[44] Fu, L., Qi, Y.: 'Camera response function estimation and application with a single image' . Yang, D., editor. Informatics in Control, Automation and Robotics, Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. pp. 149–156

[45] Ioffe, S., Szegedy, C.: 'Batch normalization: Accelerating deep network training by reducing internal covariate shift' , arXiv preprint arXiv:150203167, 2015

[46] Kingma, D.P., Ba, J.: 'Adam: A method for stochastic optimization' , arXiv preprint arXiv:14126980, 2014

[47] Chen, G., Zhu, F., Heng, P.A.: 'An efficient statistical method for image noise level estimation' . 2015 IEEE Int. Conf. on Computer Vision (ICCV), , 2015. pp. 477–

[48] Liu, X., Tanaka, M., Okutomi, M.: 'Single-image noise level estimation for blind denoising' , IEEE Trans. on image processing, 2013, 22, (12), pp. 5226–5237

[49] Immerkaer, J.: 'Fast noise variance estimation' , Computer vision and image understanding, 1996, 64, (2), pp. 300–302

[50] Honzátko, D., Kruliš, M.: 'Accelerating block-matching and 3d filtering method for image denoising on gpus' , Journal of Real-Time Image Processing, 2017

Note: Figure translations are in progress. See original paper for figures.