Raw-Guided Enhancing Reprocess of Low-Light Image via Deep Exposure Adjustment

Haofeng Huang, Wenhan Yang, Yueyu Hu, and Jiaying Liu

Peking University, Beijing, P.R.China
{huang6013,yangwenhan,huyy,liujiaying}@pku.edu.cn

Abstract. Enhancement of images captured in low-light conditions remains to be a challenging problem even with the advanced machine learning techniques. The challenges include the ambiguity of the ground truth for a low-light image and the loss of information during the RAW image processing. To tackle the problems, in this paper, we take a novel view to regard low-light image enhancement as an exposure time adjustment problem and propose a corresponding explicit and mathematical definition. Based on that, we construct a RAW-Guiding exposure time adjustment Network (RGNET), which overcomes RGB images' nonlinearity and RAW images' inaccessibility. That is, RGNET is only trained with RGB images and corresponding RAW images, which helps project nonlinear RGB images into a linear domain, simultaneously without using RAW images in the testing phase. Furthermore, our network consists of three individual sub-modules for unprocessing, reconstruction and processing, respectively. To the best of our knowledge, the proposed sub-net for unprocessing is the first learning-based unprocessing method. After the joint training of three parts, each pre-trained seperately with the RAW image guidance, experimental results demonstrate that RGNET outperforms state-of-the-art low-light image enhancement methods.

1 Introduction

Low-light environments lead to several kinds of degradations in imaging, including low visibility, intensive noise, color cast, *etc.* Modern digital cameras provide options to tackle the problem but all these options have their drawbacks. For example, high ISO leads to amplified noise, and long exposure time introduces blurring effects. Thus, there is a demand for enhancement methods to handle various degradation for images captured in low light conditions. Early attempts [1, 2] focus on the adjustment of global illumination, *e.g.* histogram equalization. Later on, Retinex theory-based methods [3–5] decompose the image into reflectance and illumination layers, and adopt specially designed priors on these two layers for detail preservation and noise suppression.

Recent works [6–9] approach higher-fidelity enhancement of low-light images with learning-based methods, and they show promising performance on the tasks. Despite that, most of the learning-based methods need low/normal light image pairs for supervised training. However, in real applications, a smaller distortion

between the reconstructed image and the only preset ground truth in the training set does not necessarily correspond to the enhancement performance, due to the mapping ambiguity (one normal-light image might correspond to several low-light images, and what "normal-light" is has not been appropriately defined as well).

Besides, there have also been works to improve the quality of images by adjusting the RAW image processing pipeline [10] or to unprocess the images for easier enhancement [11–13]. Because intensities in RAW images maintain the linear relationship with the number of photons captured by the sensor in the physical world, it facilitates the analysis and enhancement of the images. However, RAW representations are sometimes hard to obtain in real application scenarios (need to change the existing ISP systems in many scenarios). And such methods are usually designed for professional image processing on specific devices, *e.g.* digital single-lens reflex (DSLR) and other high-end cameras.

As in practical photography, expanding the exposure time in a steady shot significantly helps improve the Signal-to-Noise Ratio (SNR) and therefore enhances the image quality. Thus, in this paper, we define the low light enhancement task as to simulate the long exposure operation in image capture. It can practically result in adjusting the illumination and restore derivative degradation. We further utilize RAW signals in the training phase which are proportional to the exposure time with fixed settings. For that, we propose the RAW-Guiding exposure time adjustment Network (RGNET). RGNET is trained on processed RGB images and simultaneously takes RAW images as the guidance during the training phase, while it does not rely on any RAW images in real applications considering its unusualness. Experimental results show that our approach outperforms multiple state-of-the-art approaches both quantitatively and qualitatively. The contributions of this work are summarized as follows,

- We take a novel view of the low-light enhancement, and propose an explicit and mathematical definition of the task, *i.e.* a simulation of expanding the exposure time during the capturing. Exposure time adjustment can be formulated as solving the joint problem of illumination adjustment and the derivative degradation restoration. The illumination adjustment is handled by the amplifying RAW signals in linear style, and the enhancement network focuses on tackling the derivative degradation with the learning techniques.
- We propose the RAW-guiding exposure time adjustment network to enhance the quality of images captured in low-light conditions. During the training phase, the network is tuned with the guidance of the corresponding RAW images. And it achieves improved quality in the benchmark where the RAW images are not available.
- To the best of our knowledge, we are the first to adopt a learning-based unprocessing sub-network to facilitate the RAW-guiding low-light image enhancement pipeline. With the proposed sub-network and other enhancing components, the framework achieves state-of-the-art performance in the low-light image enhancement task.

2 Related Work

2.1 Traditional Methods

A traditional way to make dark images visible is stretching the dynamic range of an image via manipulating the corresponding histogram, *i.e.*, HE [1], which applies the adjustment globally. With more side information and constraints, *e.g.* mean intensity preservation, HE-based methods [2, 14, 15] improve local adaptability of the enhancement. Some methods [16] regard the inverted low-light images as haze images, and enhance the visibility by dehazing, then the result is inverted back as the output. A wide range of methods [17, 18] are carefully designed to depict desirable properties of images with statistical models *e.g.* based on imaging or visual perception guided models. Due to the lack of flexibility for the visual property, these methods basically fail in extreme low-light conditions.

Retinex model [19] is proposed to compute visual appearance at first, which is generally adopted in low-light image enhancement afterwards. The methods [20, 3, 4] decompose images into reflectance and illumination, then unique enhancement follows and enhanced results are combined into final outputs in designed style. Retinex-based methods generally assume that the images contain good representations of the scene content, do not model image noise and typically apply denoising as a post-process, but in fact, the severe noise in considered extreme low-light imaging affects decomposing part heavily [10].

2.2 Learning-based Methods.

Adopting deep-learning methods in low-light enhancement starts in 2017, and this branch's excellent performance and flexibility make it mainstream instead. Lore *et al.* [6] made the first efforts in using a deep auto-encoder named LLNet to perform enhancement and denoising. Then, various learning-based methods [8, 9, 21, 22] are proposed. Among them, SICE [8] builds a dataset of low-contrast and good-contrast image pairs, making the discrimination learning of SICE enhancers possible. KinD [9] builds a network to decomposes images into two components inspired by the Retinex theory and adjust light levels arbitrarily. In addition, KinD proposes that no ground-truth real data exists for low light image enhancement. DeepUPE [21] constructs a network for enhancing underexposed photos and introduces intermediate illumination in it to associate the input with expected enhancement result. However, ignored by most methods, there stands a basic problem for low-light supervised learning and fair evaluation, *i.e.* what is the ground truth corresponding to a low-light image. An outstanding method trained to handle images in 10 lux illumination even totally fails to produce a visible result in 0.1 lux condition sometimes.

Moreover, there are RAW-based works for low light enhancement. SID [10] builds the first low-light RAW dataset and proposes a pipeline to directly produce normal-light sRGB images from low-light RAW files. With the similar considering, low light video enhancement methods [23, 24] are constructed, providing low-light RAW videos. Nevertheless, RAW files' unusualness limits their generalization.

2.3 Datasets.

As data-driven methods emerge, some datasets are collected for training and evaluations in low light enhancement. On the one hand, for convenience and to avoid misalignment, synthetic datasets [6, 7] generate corresponding low-light images based on a certain strategy. On the other hand, synthetic style is highly coupled with the adopted strategy and constantly mismatches real-world data, therefore some datasets [8] capture real low/normal-light image pairs, which is accomplished by firmly fixing the camera and taking pictures with different camera settings, *e.g.* exposure. Among them, SID [10], DRV [23], SMOID [24] provide RAW files. Additionally, there are datasets [22, 21] including both synthetic and real data to trade off and unpaired datasets [5] without ground truth.

Specific to exposure time adjustment task, RAW images are required to simulate the unprocess/process stages, whose meta-data records exposure time as a label for our method as well. Because DRV and SMOID serve for video enhancement, we adopt SID as the dataset for training and evaluation.

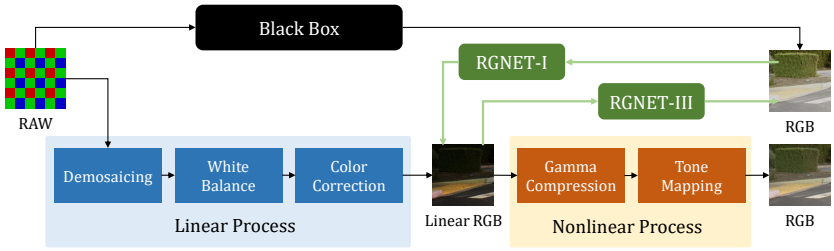
3 RAW-Guiding Exposure Time Adjustment Network

3.1 Motivation

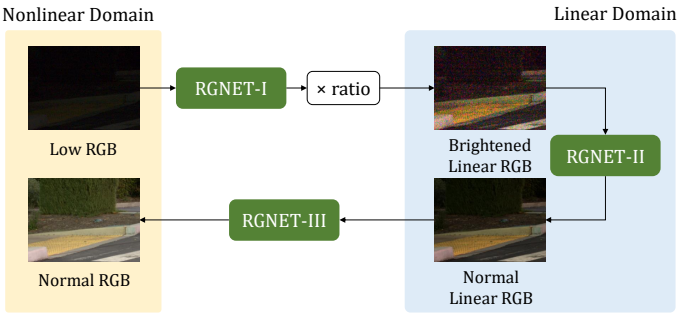
As mentioned in Sec. 1, there is currently not an explicit and mathematical definition for what the corresponding ground truth is, given a low-light image. Naturally, the low light image enhancement problem is highly ill-posed. "There are a thousand Hamlets in a thousand people's eyes." There are different criteria for the best desirable lightness for different people or from different perspectives [9], which results in difficulties to evaluate the generalization performance of low light enhancement methods. To get rid of this dilemma, we treat the low light enhancement as a novel task, *i.e.* exposure time adjustment. That is, given an image and an exposure ratio, the corresponding ground truth is the image shot in the identical static scene with the same settings but the exposure time multiplied by the given ratio. Obviously, from this perspective, the low light enhancement is equivalent to amplifying the exposure time, which, as we will discuss below, has concise mathematical form and yields conveniences for a controllable and deterministic low light image enhancement process. In the following parts, Sec. 3.2 describes a simplified traditional image processing pipeline and Sec. 3.3 explains how RGNET is constructed based on this pipeline, as shown in Fig. 1.

3.2 RAW Image Pipeline

Modern digital cameras are equipped with mature image processing systems, proceeding the raw sensor data into a relatively less noisy and pleasant looking image. Because the specific processing systems are held by camera manufacturers as commercial secrets, we regard the related details, *i.e.* specific system design and parameter selection, as black boxes in our discussion. In spite of that, we describe a conventional image processing pipeline similar to [11], establish a



(a) Image Processing Stages



(b) RAW-Guiding Exposure Time Adjustment Network

Fig. 1. (a) The top path is a real camera image processing system, a black box in our discussion, and the lower path is a simplified conventional pipeline. The proposed RGNET-I and RGNET-III, simulate transformations between actual RGB and linear RGB. (b) The proposed method adopts three subnets to complete transformations and reconstruction in the linear domain.

concise mathematical model, and fully utilize the linear parts to bridge the gap between RAW and RGB. Note that, our designed deep network method follows the paradigm of the conventional image processing pipeline and however do not rely on the detailed knowledge of any specific system.

Shot and Read Noise. The noise model in RAW is undisturbed, without nonlinearity induced by the process. As a simple consensus, sensor noise in RAW comes from two sources, usually named shot noise and read noise [25]. With the fixed settings of aperture and ISO, the noise level λ_{read} and λ_{shot} of two kinds of noise are considered as constants and the mean value x is proportional to the exposure time. Specifically, shooting with the short exposure time in the low light environment, sensor data in RAW is formulated as follows[1],

$$y_s^{raw} \sim \mathcal{N}(\mu_s, \sigma_s^2) \tag{1}$$

[1] The real noise model in low light is clipped and more complex [26] but it does not impact our discussion.

with $\mu_s = x_s\,, \sigma_s^2 = \lambda_{read} + \lambda_{shot} x_s\,$. With the exposure time ratio γ, the corresponding long-exposure data, which is regarded to be noise-free by convention, can be represented as follows,

$$y_l^{raw} = x_l = \gamma x_s\,, \tag{2}$$

x, $y \in R^{m \times n \times 1}$. Therefore, the low light enhancement becomes a denoising task in the RAW domain with the brightened y_s^{raw} as follows,

$$y_b^{raw} = \gamma y_s^{raw} \sim \mathcal{N}(\mu_b, \sigma_b^2) \tag{3}$$

with $\mu_b = \gamma x_s\,, \sigma_b^2 = \gamma^2 \lambda_{read} + \gamma \lambda_{shot}(\gamma x_s)$. The subscript s denotes "with short exposure", l denotes for "with long exposure" and b denotes "brightened".

Demosaicing. Because the sensor only captures photons, imperceptible of the chromatic light, each pixel in a camera is covered by a colored filter which arranges in some patterns, *e.g.* R-G-G-B for Bayer. Demosaicing is the processing stage to reconstruct the full-size image with all three color channels. We simply split the R-G-G-B pattern into RGB channels with green channels averaged and employ bilinear interpolation to demosaic. For concise illustration, we view this stage as detail reconstruction and just modify shapes of x and y into $m \times n \times 3$.

White Balance and Color Correction. Owing to that the color temperature of the ambient light effects filtered sensor data, the camera adopts a white balance to simulate human eyes' adaptation, *i.e.* producing an image that appears under the normal illumination. This stage is proceeded with three channels multiplied by individual weights $w_c\,(c = r, g, b)$, recorded in the RAW file. Note that the light metering process in low light environments is inaccurate, which leads to biases in those weights (denoted by $\widehat{w_c}$), which call for further calibration.

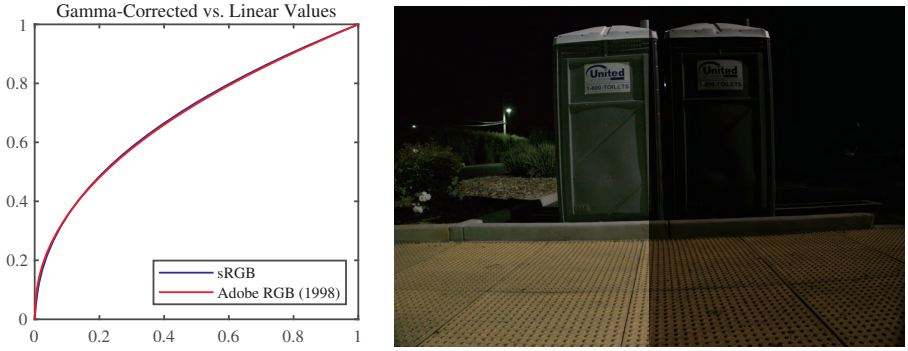
$$\begin{aligned} y_s^{wb} &= y_s^{raw} * \widehat{W}\,, \\ y_l^{wb} &= y_l^{raw} * W\,, \end{aligned} \tag{4}$$

with $\widehat{W} = [[[\widehat{w_r}, \widehat{w_g}, \widehat{w_b}]]]_{1\times1\times3}$ and $W = [[[w_r, w_g, w_b]]]_{1\times1\times3}$.

The following stage is color correction, during which the camera applies a 3×3 color correction matrix (CCM) to convert the camera color space to a required one, *i.e.* sRGB color space. The CCM M_{cc} can be obtained with the meta-data in RAW files.

$$y^{linear} = {y^{cc}}_{3\times(m\times n)} = \begin{pmatrix} y_r^{cc} \\ y_g^{cc} \\ y_b^{cc} \end{pmatrix} = \begin{pmatrix} y_r^{wb} \\ y_g^{wb} \\ y_b^{wb} \end{pmatrix} M_{cc}\,. \tag{5}$$

For convenience, y^{cc} is also named y^{linear} and the procedure to transform y^{raw} into y^{linear} is called "linear process".



(a) Gamma-corrected curves (b) brightened RGB

Fig. 2. (a) The Gamma-corrected curves based on the sRGB standard and Adobe RGB (1998) standard. (b) Brightened RGB images. The original image adopts sRGB-standard Gamma compression, and the left is brightened according to sRGB-standard Gamma function but the other uses Adobe RGB (1998) standard.

Gamma Compression and Tone Mapping. For the sake of human-vision pleasantness, nonlinear procedures follow, which usually consist of Gamma compression and tone mapping. We skip the details of both stages and simply define a function $\sigma(\cdot)$ to express the nonlinearity they introduce, named "nonlinear process".

$$y^{rgb} = \sigma(y^{cc}) \,. \tag{6}$$

As shown in Fig. 2, without using an accurate Gamma compression (the right), it induces a huge gap between the brightened images although the two Gamma-corrected curves approach. [2] Apparently, the wrong brightened one mismatches the normal brightness.

3.3 Network Architecture

The proposed method, with a trade-off between the inaccessibility of RAW images in the testing phase (if we do no change the standard ISP process) and the expedience of adjusting exposure time in the linear domain, adopts RAW-Guiding exposure time adjustment Network (RGNET). RGNET is only trained with RGB images and RAW images, which helps project the nonlinear RGB images into a linear domain, simultaneously without the need of RAW images in the testing phase. Furthermore, considering the difficulty of totally inverting the process, RGNET exploits a linear process defined in Sec. 3.2 with RAW files' meta-data to bridge the gap between RGB images and RAW images, linearity kept. The specific architecture is shown in Fig. 3.

[2] Brightening follows the procedure of inverting the Gamma compression, multiplying by the ratio, applying Gamma compression.

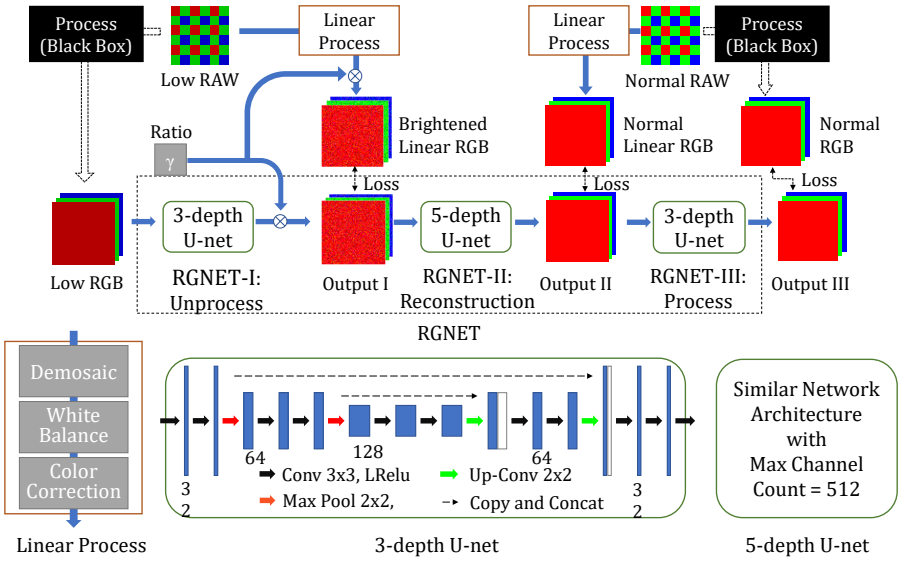


Fig. 3. The architecture of the proposed RAW-Guiding exposure time adjustment Network (RGNET). Fully utilizing the RAW file of low/normal light image, RGNET functionally implements unprocessing, reconstruction and processing, corresponding subnets named RGNET-I, RGNET-II and RGNET-III.

RGNET consists of three sub-modules: RGNET-I for unprocessing RGB images into a linear domain, RGNET-II for dealing with the amplified noise and color casting induced by the inaccurate white balance, RGNET-III for processing the enhanced results back into the nonlinear domain.

Note that as the pipeline introduce in Sec. 3.2 is a simplified one, the real processing systems may be different. With that being said, RGNET-I and RGNET-III not only simulate nonlinear stages but also aim to cope with the gap between linear stages real systems adopt and ours, *e.g.* different interpolation algorithms for demosaicing.

RGNET-I: Nonlinear Unprocess. As RGB images provide no relevant metadata about the pipeline, the traditional method [11] that converts the processed nonlinear RGB images into the linear ones is hand-crafted, which might not be consistent with the real applications. The inconsistency leads to inaccuracy of the estimation, which is further amplified simultaneously when data multiplies by the ratio, as shown in Fig. 2. Consequently, we employ an end-to-end convolutional neural network, *i.e.* a three-layer U-Net [27] to perform the task, which fits the function $\sigma^{'}$.

More exactly, given the processed inputs y_s^{rgb} and linear targets y_s^{linear}, RGNET-I aims to output $\widehat{y}_s^{linear} = \widehat{\sigma}^{'}(y_s^{rgb})$ after training.

RGNET-II: Normal-Light Image Reconstruction. Multiplied by the ratio, the brightened linear RGB images become:

$$y_s^{linear} \times \gamma = \begin{pmatrix} \gamma y_{s,r}^{cc} \\ \gamma y_{s,g}^{cc} \\ \gamma y_{s,b}^{cc} \end{pmatrix} = \begin{pmatrix} \widehat{w_r} \gamma y_{s,r}^{raw} \\ \widehat{w_g} \gamma y_{s,g}^{raw} \\ \widehat{w_b} \gamma y_{s,b}^{raw} \end{pmatrix} M_{cc}. \tag{7}$$

Comparison to the long-exposure linear RGB images

$$y_l^{linear} = \begin{pmatrix} y_{l,r}^{cc} \\ y_{l,g}^{cc} \\ y_{l,b}^{cc} \end{pmatrix} = \begin{pmatrix} w_r y_{l,r}^{raw} \\ w_g y_{l,g}^{raw} \\ w_b y_{l,b}^{raw} \end{pmatrix} M_{cc} = \begin{pmatrix} w_r \gamma x_{s,r} \\ w_g \gamma x_{s,g} \\ w_b \gamma x_{s,b} \end{pmatrix} M_{cc}, \tag{8}$$

and according to Eqn. (3), RGNET-II targets to denoise with a noise level $\gamma^2 \lambda_{read}$ and $\gamma \lambda_{shot}$ and to correct the color casting introduced by inaccurate white balance $\widehat{W}$. Considering the convolution networks' effectivity in image/video denoising [28, 29] and color correction [10, 23], we employ a five-layer U-Net as RGNET-II. The fitted denoising and color restoration function is simply denoted as $f(\cdot)$.

Specifically, given brightened inputs $y_s^{linear} \times \gamma$ and long-exposure linear targets y_l^{linear}, RGNET-II targets to output $\widehat{y_l^{rgb}} = \widehat{f}(y_s^{linear} \times \gamma)$.

RGNET-III: Nonlinear Process. Convolution networks in image processing achieve superior performance and become a common choice [30–32], which provides firm reasons for us to employ an end-to-end CNN *i.e.* three-layer U-Net to model the nonlinear process, identical to RGNET-I.

To be exact, given the long-exposure linear input y_l^{linear} and the corresponding nonlinear RGB target y_l^{rgb}, RGNET-III outputs $\widehat{y_l^{rgb}} = \widehat{\sigma}(y_l^{linear})$.

To summarize, when put into the real application, the joint RGNET outputs the exposure time adjusted result: $\widehat{y_l^{rgb}} = \widehat{\sigma}(\widehat{f}(\widehat{\sigma'}(y_s^{rgb}) \times \gamma))$, RAW files absent.

4 Experiments

In this section, we first provide the details of the adopted dataset and implementation in Sec. 4.1, then demonstrate the superiority of the proposed method through experimental comparisons with traditional and learning-based methods in Sec. 4.2 and Sec. 4.3. The effectiveness of our designs is discussed in ablation studies in Sec. 4.4. More results and analysis (including an additional evaluation of RAW-based methods) are provided in the supplementary material.

Table 1. Quantitative results of traditional methods and ours. The bold value denotes the best result.

Method	PSNR	SSIM	VIF	NIQE
HE	5.90	0.028	0.095	28.27
BPDHE	10.67	0.072	0.051	18.64
Dehazing	12.81	0.103	0.077	25.79
MSR	10.04	0.070	0.116	32.24
MF	13.87	0.111	0.108	33.13
LIME	12.59	0.102	0.118	32.33
BIMEF	13.06	0.110	0.086	26.27
Ours	**28.42**	**0.880**	**0.139**	**15.23**

4.1 Training Details.

We adopt the SID dataset [10] for training and evaluation[3], which includes 409 low/normal-light RAW image pairs, 280 pairs for training, 93 pairs for testing, and 36 pairs for validation.

In the training, we take the low-light and normal-light nonlinear RGB images (processed by Libraw) as inputs and ground truths, respectively, and employ RAW images as the side information guidance. The amplification factor is set to the ratio of exposure time. We use L_1 loss and Adam optimizer [33] for training. The learning rate is set to 10^{-4} and reduced to 10^{-5} after 2000 epochs. Each output of the subnet is clipped to $[0, 1]$. The batch size is set to 1 and patch-size to 512×512. All sub-modules are pre-trained independently and are converged in 2000-3000 epochs and the joint training lasts for 1000 epochs with learning rate 10^{-5}.

The entire network is trained on an Nvidia RTX 2080Ti GPU and Intel(R) Xeon(R) E5-2650 2.20GHz CPU using the Tensorflow framework. Due to the large spatial size of the testing image, cropping is implemented in testing.[4]

4.2 Comparison to Traditional Methods.

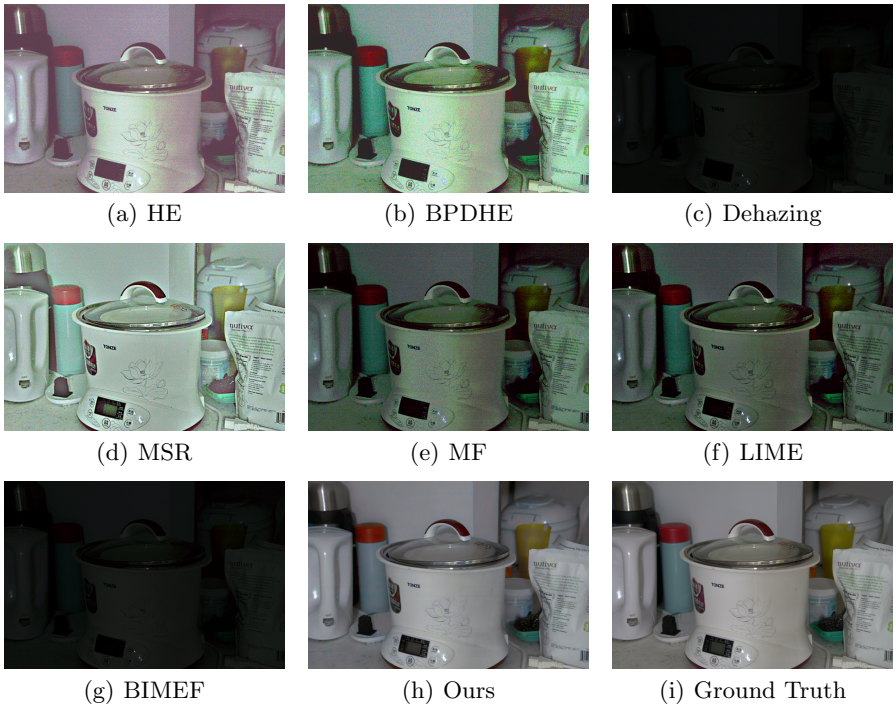
We first compare our methods with several traditional methods including HE [1], Dehazing [16], MF [4], MSR [3], LIME [5], BIMEF [18], BPDHE [14].

The quantitative evaluation adopts full-reference metrics PSNR, SSIM [34], VIF [35], and a no-reference metric NIQE [36], results shown in Table 1. Considering, our RGNET outperforms traditional methods on the SID dataset.

Qualitative results are shown in Fig. 4. Conventional methods tend to brighten images uniformly, which easily leads to the under/over-exposure in some regions. Furthermore, extreme low-light environments of the SID dataset give rise to severe noise and color cast or insufficient illumination in the results. Specifically,

[3] We use the Sony set, whose images are captured by Sony α7S II with a Bayer sensor.

[4] We crop 4256×2848 images into four 2128×1424 patches, and pad 200 pixels to reduce the blocking artifacts.



(a) HE (b) BPDHE (c) Dehazing

(d) MSR (e) MF (f) LIME

(g) BIMEF (h) Ours (i) Ground Truth

Fig. 4. Qualitative results of traditional methods and ours.

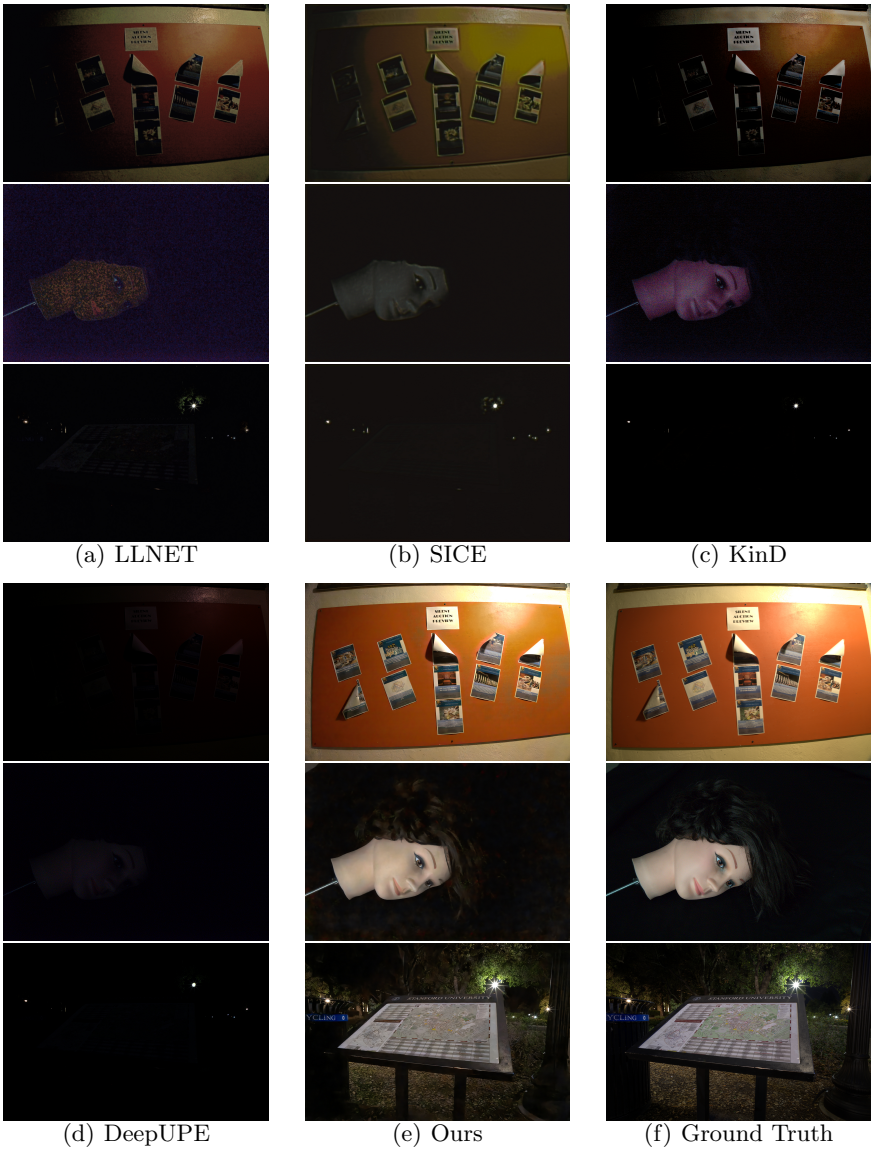
Dehazing and BIMEF totally fail to enhance low-light images with a reasonable illumination level. HE, MF, MSR and LIME produce gray results owing to uniform brightening, and HE and BPDHE handle color biases improperly.

4.3 Comparison to Learning-Based Methods.

We also evaluate the performance of state-of-the-art learning-based RGB low-light enhancement methods on the SID dataset, including LLNET [6], SICE [8], KinD [9] and DeepUPE [21].[5]

Due to extreme low-light imaging with severely limited illumination and short exposure time the SID dataset adopts, without using the ratio, these methods lead to under-exposure, as shown in Fig. 5. That further proves how absence of the definition of ground truths affects the generalization performance of methods in low light enhancement tasks. Excellent methods might totally fail only with the testing set changed. Specifically, DeepUPE fails to produce human visible outputs. KinD, LLNET and SICE's results are still not bright enough and these

[5] We choose the maximum brightening ratio ($\gamma = 5.0$) for KinD. Due to the GPU memory limit, the input resolution of SICE has to be small, so we down-sample the input images, perform SICE and up-sample results back into the original size.



(a) LLNET (b) SICE (c) KinD

(d) DeepUPE (e) Ours (f) Ground Truth

Fig. 5. Qualitative results of learning-based methods and ours.

methods do not generate the bright regions of ground truth as shown at the top panel in Fig. 5. Besides, there still exists much noise in results and color cast is not properly dealt with as well, shown at the middle panel. All other methods reveal their scarcity handling moonlight cases, shown at the bottom panel.

The quantitative evaluation still adopts the same metrics as Sec. 4.2 and BLIINDS-II [37] is added, results shown in Table 2. Comprehensively, only using

Table 2. Quantitative results of learning-based methods and ours. The bold value denotes the best result.

Method	PSNR	SSIM	VIF	NIQE	BLIINDS-II
LLNET	14.21	0.221	0.047	18.02	52.46
SICE	14.26	0.366	0.011	16.89	49.53
KinD	13.35	0.109	0.048	17.73	47.66
DeepUPE	12.10	0.070	0.028	18.30	62.54
Ours	**28.42**	**0.880**	**0.139**	**15.23**	**18.38**

processed RGB images in the testing phase, our RGNET outperforms these learning-based methods on the SID dataset.

4.4 Ablation Study.

We conduct ablation studies to provide a quantitative evaluation of the effectiveness of our network architecture. Summative results are shown as Table 3.

Training without RAW. First, we consider the situation most learning-based low light enhancement methods are designed for, *i.e.* only making use of processed RGB low/normal-light pairs for training. We adopt a U-net architecture for end-to-end deep learning. The results have low SNR and, severely affected by the SID dataset's feature *i.e.* one ground truth for several low-light images with different illumination, enhanced images have serious artifacts. Apparently, without other labels, networks hardly dynamically brighten low-light images at variance. The MS-SSIM loss helps eliminate artifacts but induce a drop on the pixel-level full-reference metric PSNR. We try to stretch the histogram without extra labels *i.e.* multiplying by the ratio of the mean pixel value of the normal/low-light pair. However, the operation conflicts with the nonlinearity of processed RGB images so it doesn't display comparable performance.

Training with original RAW. Then the necessity of linearly preprocessing RAW is revealed by the experiment using original RAW instead. A drop in the

Table 3. Ablation study. This table reports mean PSNR in each condition.

Condition	PSNR
w/o RAW	22.11
w/o RAW, adding MS-SSIM loss	21.88
w/o RAW, adjusting mean	24.27
with original RAW	28.17
Ours	**28.42**

Table 4. Controlled experiments on subnets. Note that different architectures require different inputs. RGNET-III's joint network is identical to the individual one.

Inputs	RGNET			PSNR	
	I	II	III	individual	joint
Low RGB	✓	✓	✓	26.06	28.42
Brightened linear RGB	×	✓	✓	27.91	28.50
Normal linear RGB	×	×	✓	42.68	42.68

score is predictable because the adopted linear process consists of basic stages included in most image processing pipelines therefore it bridges the gap between original RAW and processed RGB images which relieves the pressure to simulate the whole pipeline.

Controlled experiments on subnets. We further evaluate quantitative performance of proposed subnets as shown in Table 4. "individual" means directly testing after seperate pre-training of subnets and "joint" means testing after further joint training. The gap between them is owing to losing information for the following stages and error accumulation. Based on the result, RGNET-I and RGNET-III target nonlinear parts properly, which can be made out with a relatively little drop in PSNR between Table 4. Row. 1 and 2 and excellent PSNR score in Row. 3. The most challenging part is still reconstructing normal-light linear RGB from the brightened one affected by severe noise and biased color.

5 Conclusion

In this paper, we take a novel view to regard low light enhancement as exposure time adjustment and obtain an explicit and mathematical definition of this task. Based on that the proposed RGNET overcome RGB images' nonlinearity and RAW images' unusualness, and outperforms many RGB-based state-of-the-art approaches. Our network consists of three individual sub-modules for unprocessing, reconstruction and processing and it's the first attempt to adopt a learning-based unprocessing method for low light enhancement, which facilitates the end-to-end training of the whole designed pipeline. Our framework only needs access to the RAW image in the training process but can offer better results with that kind of information during the testing phase, therefore our results are guided by the RAW information but not rely on RAW input and the change of ISP process in the real applications.

Our future work is to extend the generalization of RGNET-I which, optimistically, handles the unprocessing task of various RGB images with different illumination and processing systems. Based on that, label-free training and evaluation are accessible, with a ratio of unprocessed linear normal/low-light images' intensity instead.

References

1. Pizer, S.M., Johnston, R.E., Ericksen, J.P., Yankaskas, B.C., Muller, K.E.: Contrast-limited adaptive histogram equalization: speed and effectiveness. In: Proc. Conf. Visualization in Biomedical Computing. (1990) 337–345
2. Abdullah-Al-Wadud, M., Kabir, M.H., Dewan, M.A.A., Chae, O.: A dynamic histogram equalization for image contrast enhancement. IEEE Trans. on Consumer Electronics **53** (2007) 593–600
3. Jobson, D.J., Rahman, Z., Woodell, G.A.: A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE Trans. on Image Processing **6** (1997) 965–976
4. Fu, X., Zeng, D., Huang, Y., Liao, Y., Ding, X., Paisley, J.: A fusion-based enhancing method for weakly illuminated images. Signal Processing **129** (2016) 82 – 96
5. Guo, X., Li, Y., Ling, H.: LIME: Low-light image enhancement via illumination map estimation. IEEE Trans. on Image Processing **26** (2017) 982–993
6. Lore, K.G., Akintayo, A., Sarkar, S.: LLNet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognition **61** (2017) 650 – 662
7. Shen, L., Yue, Z., Feng, F., Chen, Q., Liu, S., Ma, J.: MSR-net: Low-light image enhancement using deep convolutional network. ArXiv e-prints (2017)
8. Cai, J., Gu, S., Zhang, L.: Learning a deep single image contrast enhancer from multi-exposure images. IEEE Trans. on Image Processing **27** (2018) 2049–2062
9. Zhang, Y., Zhang, J., Guo, X.: Kindling the darkness: a practical low-light image enhancer. ACM Int'l Conf. Multimedia (2019)
10. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. In: Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition. (2018) 3291–3300
11. Brooks, T., Mildenhall, B., Xue, T., Chen, J., Sharlet, D., Barron, J.T.: Unprocessing images for learned raw denoising. In: Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition. (2019) 11028–11037
12. Afifi, M., Abdelhamed, A., Abuolaim, A., Punnappurath, A., Brown, M.S.: CIE XYZ Net: Unprocessing images for low-level computer vision tasks. ArXiv e-prints (2020)
13. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: CycleISP: Real image restoration via improved data synthesis. In: Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition. (2020)
14. Ibrahim, H., Pik Kong, N.S.: Brightness preserving dynamic histogram equalization for image contrast enhancement. IEEE Trans. on Consumer Electronics **53** (2007) 1752–1758
15. Arici, T., Dikbas, S., Altunbasak, Y.: A histogram modification framework and its application for image contrast enhancement. IEEE Trans. on Image Processing **18** (2009) 1921–1935
16. Dong, X., Wang, G., Pang, Y., Li, W., Wen, J., Meng, W., Lu, Y.: Fast efficient algorithm for enhancement of low lighting video. In: Proc. IEEE Int'l Conf. Multimedia and Expo. (2011) 1–6
17. Celik, T., Tjahjadi, T.: Contextual and variational contrast enhancement. IEEE Trans. on Image Processing **20** (2011) 3431–3441
18. Ying, Z., Li, G., Gao, W.: A bio-inspired multi-exposure fusion framework for low-light image enhancement. ArXiv e-prints (2017)
19. Land, E., McCann, J.: Lightness and retinex theory. In: Journal of the Optical Society of America. (1971) 1–11

20. Jobson, D.J., Rahman, Z., Woodell, G.A.: Properties and performance of a center/surround retinex. IEEE Trans. on Image Processing **6** (1997) 451–462
21. Wang, R., Zhang, Q., Fu, C.W., Shen, X., Zheng, W.S., Jia, J.: Underexposed photo enhancement using deep illumination estimation. In: Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition. (2019)
22. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. In: British Machine Vision Conference. (2018)
23. Chen, C., Chen, Q., Do, M., Koltun, V.: Seeing motion in the dark. In: Proc. IEEE Int'l Conf. Computer Vision. (2019) 3184–3193
24. Jiang, H., Zheng, Y.: Learning to see moving objects in the dark. In: Proc. IEEE Int'l Conf. Computer Vision. (2019) 7323–7332
25. Hasinoff, S.W. In: Photon, Poisson Noise. Springer US, Boston, MA (2014) 608–610
26. Abdelhamed, A., Lin, S., Brown, M.S.: A high-quality denoising dataset for smartphone cameras. In: Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition. (2018)
27. Ronneberger, O., P.Fischer, Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention. Volume 9351 of LNCS., Springer (2015) 234–241
28. Burger, H., Schuler, C., Harmeling, S.: Image denoising: Can plain neural networks compete with bm3d? In: Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition. (2012)
29. Liu, P., Zhang, H., Zhang, K., Lin, L., Zuo, W.: Multi-level wavelet-cnn for image restoration. In: Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition Workshops. (2018)
30. Gharbi, M., Chaurasia, G., Paris, S., Durand, F.: Deep joint demosaicking and denoising. ACM Trans. on Graphics **35** (2016) 1–12
31. Chen, Q., Xu, J., Koltun, V.: Fast image processing with fully-convolutional networks. In: Proc. IEEE Int'l Conf. Computer Vision. (2017)
32. Yue, H., Cao, C., Liao, L., Chu, R., Yang, J.: Supervised raw video denoising with a benchmark dataset on dynamic scenes. In: Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition. (2020)
33. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In Int'l Conf. Learning Representations. (2015)
34. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. on Image Processing **13** (2004) 600–612
35. Sheikh, H.R., Bovik, A.C.: Image information and visual quality. IEEE Trans. on Image Processing **15** (2006) 430–444
36. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Signal Processing Letters **20** (2013) 209–212
37. Saad, M.A., Bovik, A.C., Charrier, C.: Blind image quality assessment: A natural scene statistics approach in the DCT domain. IEEE Trans. on Image Processing **21** (2012) 3339–3352