Towards Low Light Enhancement With RAW Images

Haofeng Huang, *Graduate Student Member, IEEE*, Wenhan Yang◉, *Member, IEEE*,
Yueyu Hu◉, *Graduate Student Member, IEEE*, Jiaying Liu◉, *Senior Member, IEEE*,
and Ling-Yu Duan◉, *Member, IEEE*

Abstract— In this paper, we make the first benchmark effort to elaborate on the superiority of using RAW images in the low light enhancement and develop a novel alternative route to utilize RAW images in a more flexible and practical way. Inspired by a full consideration on the typical image processing pipeline, we are inspired to develop a new evaluation framework, *Factorized Enhancement Model* (*FEM*), which decomposes the properties of RAW images into measurable factors and provides a tool for exploring how properties of RAW images affect the enhancement performance empirically. The empirical benchmark results show that the *Linearity* of data and *Exposure Time* recorded in meta-data play the most critical role, which brings distinct performance gains in various measures over the approaches taking the sRGB images as input. With the insights obtained from the benchmark results in mind, a RAW-guiding Exposure Enhancement Network (REENet) is developed, which makes trade-offs between the advantages and inaccessibility of RAW images in real applications in a way of using RAW images only in the training phase. REENet projects sRGB images into linear RAW domains to apply constraints with corresponding RAW images to reduce the difficulty of modeling training. After that, in the testing phase, our REENet does not rely on RAW images. Experimental results demonstrate not only the superiority of REENet to state-of-the-art sRGB-based methods and but also the effectiveness of the RAW guidance and all components.

Index Terms— Low-light enhancement, benchmark, RAW guidance, deep learning, factorized enhancement model.

I. INTRODUCTION

LOW-LIGHT environments cause a series of degradation in imaging, including intensive noise, low visibility, color cast, *etc*. More sophisticated shooting equipment and advanced specialized photographic systems pay a premium to alleviate

the degradation to some extent. Modern digital cameras make efforts in tackling the problem by adjusting the shooting parameters but also incur accompanying issues. For instance, high ISO introduces amplified noise, and long-exposure time results in blurring. Hence, it is economical and desirable to enhance the low-light images by software.

In most applications, two kinds of images[1] are taken as the input of the enhanced approaches: *RAW images* [1]–[4]; *RGB images* [5]–[10], which are processed from raw images via several procedures, *e.g.* demosaicing, white balance, tone mapping, *etc.*, in consideration of human vision preference and system requirement, *e.g.* the storage limit. As reported in these prevailing works [1], [2], [11], the low-light enhancement methods that take the RAW data as input usually achieve significantly superior performance to those taking sRGB data as their input. On one hand, compared with sRGB images, RAW data possesses two inherent advantages: 1) *Primitive*: RAW data nearly is obtained directly from the sensor, and records the meta-data related to the hardware and shooting settings, whereas sRGB images have been processed for human vision preference and system requirement, which inevitably causes information loss. 2) *Linear*: As RAW data is directly captured by sensors, RAW data's relationship at different exposure levels keeps linear, while that dependency in the sRGB domain is nonlinear as processed by the processing system.

On the other hand, in real applications, it might be more difficult to obtain RAW images from real applications. First, RAW images include abundant information that is stored costly, therefore many devices choose to only store sRGB images. Second, from the user side, a devastating display of RAW images relies on a series of professional processing operations and expert knowledge. Therefore, more casual users prefer a pocket device [12], *e.g.* mobile phone, instead of advanced devices for shooting, *e.g.* digital single-lens reflex (DSLR). Therefore, more user-friendly sRGB image-based applications are becoming a trend. The advantages and disadvantages of using RAW data will be illustrated in detail in Sec. III-B.

Based on the above discussion, two critical issues are revealed:

- What are the properties of RAW files that really contribute to the low-light image enhancement?

---

[1] https://en.wikipedia.org/wiki/Raw_image_format

- Is there an alternative way to utilize RAW files for real applications instead of changing existing commonly used image processing systems? For example, can we make full use of advantages of RAW files but get rid of them in testing?

To address these two issues, we start from a benchmark effort. Centering at the procedures of the image processing pipeline, we describe the low-light enhancement with a newly proposed *Factorized Exposure Model* (*FEM*). FEM decomposes the ambiguity of low-light image enhancement into several measurable factors, *e.g.* a simulation of exposure time adjustment in the image acquisition before processing. With the benefits of this framework, we compare several schemes of using RAW data with different combinations of inputs and guidance to reveal critical properties of RAW data that make real merits to the low-light image enhancement. The benchmark results demonstrate that, among all factors, *Linearity* of data and *Exposure Time* recorded in meta-data play the most important role in quantitative measures. Inspired by this insight, a novel RAW-guiding Exposure Enhancement Network (REENet) is proposed to show an alternative route that not only utilizes the RAW images but also is user-friendly to sRGB-based applications. Different from previous RAW-based approaches, our REENet takes processed sRGB images as the input and only adopts RAW images as the guidance in the training process, while getting rid of them in the testing process. Extensive experimental results demonstrate that our approach outperforms state-of-the-art sRGB-based approaches both quantitatively and qualitatively.

The contributions of this work are summarized as follows,

- To the best of our knowledge, our work is the first benchmark effort to elaborate on the superiority of using RAW images (different inputs/different supervision) quantitatively in the low light enhancement. With a detailed analysis, the benchmark results reveal meaningful insights, which inspire us to explore the new route to fill in the gap between sRGB-based and RAW-based approaches.
- We follow the image processing pipeline and introduce a newly proposed *Factorized Exposure Model* (*FEM*) to describe the low-light enhancement process with several measurable factors that lead to ambiguity, *e.g.* simulating exposure time adjustment in the image acquisition before processing, for benchmarking characteristics of RAW images and the way to utilize them.
- Inspired by the insights from the benchmark, we further propose a novel RAW-guiding Exposure Enhancement Network (REENet) for low-light enhancement that only needs RAW images as input during the training phase. Experimental results show that, the proposed method outperforms state-of-the-art sRGB-based methods when RAW input images are not available.

The rest of this paper is organized as follows. Section II briefly reviews the related sRGB-based and RAW-based work. Section III shows the benchmark results of various approach for the proposed evaluation framework called *Factorized Enhancement Model*. Section IV introduces the proposed RAW-guiding exposure enhancement network and provide experimental results for comparison, and ablation study. Conclusions are summarized in Section V.

II. LITERATURE REVIEW

A. sRGB-Based Methods

The earlier methods mainly take sRGB images as input. The traditional *histogram equalization* methods adjust the illumination via stretching the dynamic range of an image by manipulating its histogram, globally [16], [17] or in a local adaptive way [5], [6], [18]–[20]. These methods can effectively adjust the image contrast, but are incapable of changing visual structures of local regions, which inevitably leads to under/over-exposure and amplified noise.

Inverted dehazing methods [7], [21], [22] invert low-light images to be haze ones, improve the visibility via dehazing algorithms, and then invert the processed result back as the output. Although achieving superior performance in some cases, these methods lack a convincing physical explanation.

Statistical model based methods optimize towards desirable properties of images, *e.g.* perceptual quality measure [23], interpixel relationship [24], physical lighting models [25], and imaging or visual perception guided models [26]. These methods show superior effectiveness in their focused aspects. Because of the absence of flexibility in injecting visual properties, these methods fail to handle extreme low-light environments where images are buried with intensive noise.

Retinex model based methods [8], [27]–[31] separate an image into two representations, *i.e.* reflectance and illumination layers, and then the well-designed enhancement methods follow to enhance these two layers, respectively. The works in [32], [33] enrich Retinex model-based methods with the robust constraint and an explicit noise term, which helps better capture and suppress noise.

Since 2017, the low-light enhancement steps into the deep-learning era [9]. *Deep learning* based methods bring in excellent enhancement performance and flexibility in injecting various kinds of priors and constraints via designing new architectures and training losses [10], [34]–[38], [38], [39]. However, the performance of these methods is dependent on the distribution of the paired training images, which in fact limits the model's generality. Recently, learning-based enhancement methods with unpaired data, *e.g.* Enlighten-GAN [40], Zero-DCE [41] and DRBN [42], partially get rid of the issue with CycleGAN, self-learned curve adjustment, and quality guidance, respectively. Besides, there are many works [43]–[46] dedicated to solving the composite tasks, *e.g.* HDR and Blind Image Restoration, where the low light enhancement just acts as a single component in these pipelines.

However, as the image processing systems introduce non-linearity and discard some fine-grained information when processing RAW images into sRGB ones, the enhancement from sRGB images is highly ill-posed and hard to offer desirable results in the extremely dark condition. Furthermore, most of these methods target to restore both illumination (estimating exposure level) and detailed signals (suppressing noise and revealing details). Comparatively, in our work, we target an image acquired with a longer exposure time, where the

TABLE I

An Overview of Low-Light Image Enhancement Methods. Top: sRGB Image Based Methods. Bottom: RAW Image Based Methods. Gray: Gray Image; sRGB: sRGB Image; RAW: RAW Image; RAWV: RAW Video; γ: Exposure Time Ratio

Category	Method	Input	Highlight
sRGB: - Nonlinearity; - No meta-data; - Coarse-grained quantization level; - Easily stored.	Histogram Equalization	Gray/RGB (Test)	Adjust the illumination via expanding an image's dynamic range by manipulating its histogram globally or adaptively in local regions.
	Invert Dehazing	RGB (Test)	Improve the visibility by dehazing approaches on the inverted versions of the low-light images.
	Statistical Model	–	Optimize statistical structural constraints towards desirable properties of images, *e.g.* gradient, and context priors, and environmental light.
	Retinex Model	–	Improve the visual quality of low-light images via decomposing them into reflectance and illumination representations and use elaborately designed priors on them.
	Deep Learning	RGB (Train) RGB (Test)	Restore normal-light images and pursuit better quality via injecting various kinds of priors and constraints into deep networks with diverse architectures.
RAW: - Linear; - With meta-data; - Fine-grained quantization level; - Costly stored.	SID [1]	RAW, γ (Train) RAW, γ (Test)	The first work that takes an end-to-end learnable structure to act as the image processing pipeline for generating normal-light sRGB images with a dataset with extremely dark RAW data and well-exposed sRGB.
	DeepISP [13]	RAW (Train) RAW (Test)	Propose a new deep network with state-of-the-art results to perform denoising and demosaicing jointly with end-to-end image processing.
	SMD [11]	RAW, γ (Train) RAW, γ (Test)	Construct a static video dataset with the ground truth and proposes a Siamese network to suppress noise while keeping inter-frame stability.
	SMOID [2]	Paired RAWV (Train) RAWV (Test)	Develop a novel optical system that is capable of capturing paired low/normal-light videos and construct a fully convolutional network consisting of 3D and 2D miscellaneous operations for image enhancement.
	ELD [14]	RAW, γ (Train) RAW, γ (Test)	A novel noise formation model to synthesize more realistic extremely dark data for data augmentation that helps the trained models perform better.
	EEMEFN [15]	RAW, γ (Train) RAW, γ (Test)	Take multi-exposed inputs to generate the well-exposed output, which is further enhanced by the edge enhancement module.
	LRDE [4]	RAW (Train) RAW (Test)	Recover objects in low-frequency layers first and enhance high-frequency details based on recovered objects later.

performance in the dimensions except for the exposure level is paid more attention to and the desired exposure level might be not unique and can be given by users at the testing time.

B. RAW-Based Methods

Some works make efforts in improving the image quality by building the learnable RAW image processing pipelines [1], [2], [11] or unprocessing the sRGB images back into the RAW domain for a more effective enhancement process [3], [4], [15], [47], [48]. The signal values in RAW images are totally dependent on the photon number captured by the sensor and have a linear correlation with each other at different exposure levels. This property decreases the difficulty in manipulating the image pixel signal, and facilitates modeling and enhancing low-light images/videos. In [1], Chen *et al.* proposed a novel learnable processing strategy for the RAW data captured in extremely dark indoor/nighttime environments, and constructed the See-in-the-Dark (SID) dataset which is the first to provide short/long exposure RAW pairs. The successive RAW-based methods [4], [13], [15] develop more advanced architectures to further improve the low-light enhancement performance on SID. In [13], a new deep network is designed for image enhancement to offer state-of-the-art results in turning the RAW image into a final high perceptual quality image. In [15], Zhu *et al.* proposed a multi-exposure fusion module to combine the generated multi-exposure images with a set of exposure ratios and then adopted an edge enhancement module to produce high-quality results with sharp edges. The work in [4] also adopts a two-stage framework that introduces attention to context encoding blocks to deal with the restoration of low/high-frequency information at different stages. The work in [14] focuses on noise suppression of RAW images captured in the low-light condition and improving the trained model's capacity via synthesizing more realistic data with the proposed noise model. The works in [2], [11] move one step forward to focus on low-light video enhancement. In [11], a novel Dark Raw Video (DRV) dataset is created including paired low/normal-light RAW images in static scenes and unpaired low-light RAW images in dynamic scenes, and a new deep network fully considering generalization and temporal consistency is built jointly with VBM4D to effectively enhance the low-light videos while suppressing noise. In [2], Jiang and Zeng developed a novel optical system used to capture low/normal-light videos at the same scene, *i.e.* See-Moving-Objects-in-the-Dark (SMOID) dataset, and built a learnable spatial-temporal transformation to turn the RAW videos into normal-light sRGB ones.

Although adopting RAW images in learning-based methods leads to a large performance leap, it is still unclear what properties of RAW data contribute to those gains. Furthermore, the inaccessibility of RAW files limits their application scopes.

In our paper, we aim to benchmark the ways to utilize RAW images quantitatively and, different from RAW-based methods, we explore an alternative way to utilize RAW files for real applications without changing the existing ISP systems.

III. BENCHMARKING RAW DATA UTILIZATION IN LOW-LIGHT IMAGE ENHANCEMENT

A. Motivation

Naturally, the low-light image enhancement problem taking the low-light sRGB image as the input image is highly ill-posed. Comparatively, restoring from RAW images is much less ambiguous especially when the exposure ratio in the meta-data of RAW files has provided much information about the illumination. To compare different methods from the perspective of RAW utilization, we formulate the image processing pipeline and propose a novel view to regard low-light image enhancement as the framework of *Factorized Enhancement Model* (*FEM*), which decomposes that ambiguity into several measurable factors, and facilitates comparing the effects of various properties of RAW files on low-light image enhancement.

B. Characteristics of RAW Files

Modern digital shooting systems with the image processing pipeline proceed the sensor data into a more visually pleasant image with less noise, which is stored as an RGB file (*e.g.* sRGB image in JPEG or PNG format). Compared with the processed sRGB image, the RAW file has the following good properties:

- **Access to meta-data**. During image acquisition, cameras record the shooting parameters as the meta-data d^{meta} for original sensor data d^{sens}. Influenced by the hardware, the sensor data is highly camera-specific, *e.g.* adopting different black levels, saturation, and lens distortion and being modeled by a camera-specific real-world noise model [14]. A RAW file f^{raw} consists of sensor data d^{sens} and meta-data d^{meta}.
- **Linearity of data**. In a linear image, the pixel values are directly related to real-world signal, *i.e.* the number of photons received at that location on the sensor and therefore keep a linear correlation at different exposure levels. To restore *linear RAW data* y^{raw} from the sensor data d^{sens}, the hardware calibration operations $F_{cali}(\cdot)$ such as linearization and lens calibration are applied. A theoretically perfect calibration can decouple sensor data with its capture equipment, making the calibrated signal linearly depend on the real-world signal:

$$y^{raw} = F_{cali}(d^{meta}, d^{sens}, \alpha), \qquad (1)$$

Note that because sensor data d^{sens} is stored discretely, the restored y^{raw} is discrete as well. As the distributions of noise and bias induced by the hardware are quite complex and data-dependent [14], a perfect calibration is hard to obtain.

- **Fine-grained quantization level (*i.e.* more abundant intensities and colors)**. Most RAW files contain much abundant information, due to their high resolution and wide signal range capturing more fine-grained intensities and colors. However, the RAW images are stored costly and unfriendly to be displayed to the human vision (a nonlinear perception system), which limits the application scopes of RAW images. The final output of a processing system is usually an 8-bit sRGB image.

The aforementioned characteristics disappear when the RAW files are processed into final sRGB images. Most image processing systems serve human vision perceptual quality, therefore the successive adjustment stages in processing based on human vision are conducted, *e.g.* white balance, tone mapping and gamma correction. A standard sRGB system produces a nonlinear sRGB image y^{srgb} as follows:

$$f^{srgb} = F_{proc}(d^{meta}, y^{raw}, \beta), \qquad (2)$$

where $F_{proc}(\cdot)$ denotes the processing stage, and β denotes the configuration. In Section III-C, we will provide a more detailed analysis of $F_{proc}(\cdot)$. After the processing, a nonlinear 8-bit image f^{srgb} is obtained. Note that the meta-data might be also available for sRGB files as well, *e.g.* EXIF in JPEG format or a coupled metadata file directly obtained from the digital camera. In more common cases, *e.g.* the images on the Internet and social networks, or the edited images by post-processing or editing, the perfect meta-data is hardly available. Therefore, in our paper, the final version of the proposed method does not rely on access to the meta-data in the final sRGB files. However, to make our paper more comprehensive, we also discuss situations where sRGB files are coupled with the perfect metadata recorded or not to see how it benefits the enhancement.

To summarize, characteristics of RAW files include the access to meta-data, linearity of the data, as well as fine-grained quantization level (*i.e.* more abundant intensities and colors). These properties disappear when the RAW files are projected into the final sRGB images via the image processing systems.

C. Image Processing Pipeline

In this section, we describe the image processing pipeline $F_{proc}(\cdot)$ in Eqn. (2). As the specific pipelines and configurations of the processing systems in each kind of camera are kept as commercial secrets, in our discussion, we treat these details as black boxes. Despite this, the conventional image processing system [47] also helps establish a concise mathematical model as shown in Fig. 1 (a), which we can make use of as the framework to evaluate the properties of RAW that benefit low-light image enhancement as shown in Fig. 1 (b). Note that all sRGB images used in benchmark as the final targets are processed by Libraw, which is regarded as a black box in our discussion. Our defined simplified processing pipeline, including a simplified demosaicking module, only provides the intermediate supervision in the RAW domain and does NOT actually influence benchmark results.

1) Shot and Read Noise: The real-world signal recorded in RAW files is mixed with physically caused noise. Compared with sRGB, the noise model in the RAW domain is seldom
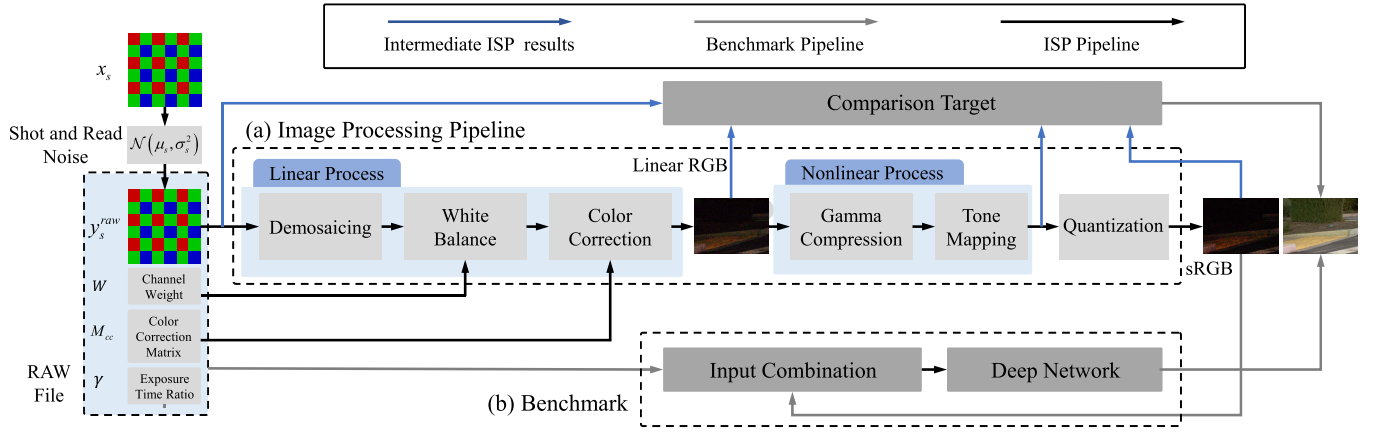
Fig. 1. An overview of the image processing pipeline and benchmark framework. (a) We use a simplified image processing conventional pipeline to model the process. The RAW file includes the noisy RAW image y_s^{raw}, channel weight ω_c for white balance, color correction matrix M_{cc} for color correction, and exposure time ratio γ. The workflow of the image processing pipeline is **denoted by black lines**. (b) In our benchmark, we aim to compare four output targets from the pipeline, *i.e.* the RAW image y_s^{raw}, linear RGB image y^{lin}, the image before the quantization y^{srgb}, and sRGB f^{srgb} (**denoted by blue lines**). However, in the benchmark, we cannot access the image processing pipeline and therefore take the RAW image y_s^{raw}, linear RGB image y^{lin}, channel weight ω_c for white balance, color correction matrix M_{cc} for color correction, and exposure time ratio γ as the input in our benchmark. These factors are combined for comparing the effects of different properties of RAW files on the low-light image enhancement. The related workflow is (**denoted by gray lines**).
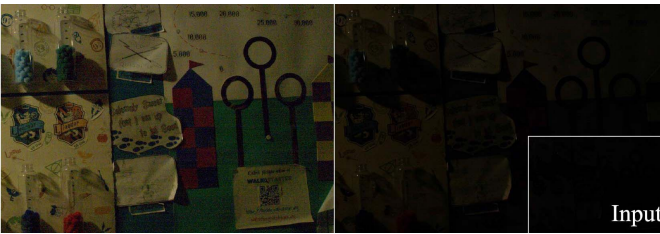

Fig. 2. Visual comparison results of Gamma correction results with different settings. The sRGB-standard Gamma compression is utilized in the original image. **Left Panel**: The results brightened by inverting the sRGB-standard Gamma function. **Right Panel**: The results brightened by the inversion of Adobe RGB (1998) standard.

disturbed by the nonlinearity in the processing pipeline. Sensor noise in the RAW domain consists of two parts: shot noise and read noise [49]. By using the fixed aperture and ISO, the value of noise-free signal x is linearly dependent on the exposure time. Specifically, to simulate shooting in the low-light conditions, we utilize a short-exposure time, then the sensor data y_s^{raw} in the RAW domain can be formulated:

$$y_s^{raw} = x_s + n_{shot}(x_s) + n_{read}, \tag{3}$$

where x_s is the noise-free short-exposure RAW image and y_s^{raw} is the noisy one, n_{shot} and n_{read} are shot noise and read noise. Subscript s denotes *short-exposure* here. As simplified in [50],

$$
\begin{aligned}
y_s^{raw} &\simeq x_s + n(x_s), \\
n(x_s)[i] &\sim \mathcal{N}(0, \sigma_s^2[i]), \\
\sigma_s^2[i] &= \lambda_{shot} x_s[i] + \lambda_{read},
\end{aligned}
\tag{4}
$$

where λ_{shot} and λ_{read} denote the noise levels for a camera, i denotes location and $[\cdot]$ returns the value at location i.

2) Demosaicing: Since the sensor is only capable of capturing photons, not aware of the chromatic light, to precept the chroma information, in the camera the pixels are covered by colored filters that are arranged with a certain pattern, *e.g.* the R-G-G-B Bayer pattern. Demosaicing is one of the processing stages that helps reconstruct the full-size color image. In our implementation, the R-G-G-B pattern is converted into RGB channels via averaging green channels and adopting Bilinear interpolation to upsample the resolution to $m \times n \times 3$.

3) White Balance and Color Correction: Since the filtered sensor data is affected by the color temperature of the ambient light, the camera applies the white balance to generate images under the normal illumination with the colors visually pleasing to human eyes. In this stage, three channels are multiplied with the weights w_c ($c = r, g, b$), which are obtained from the RAW file. Note that, the light metering obtained from the low-light conditions might be inaccurate [11], those weights (denoted by $\widehat{w_c}$) are usually biased and need additional calibration. This module is formalized:

$$
\begin{aligned}
y_s^{wb} &= y_s^{raw} \circ \widehat{W}, \\
y_l^{wb} &= y_l^{raw} \circ W, \\
\widehat{W} &= \left[\left[\left[\widehat{w_r}, \widehat{w_g}, \widehat{w_b}\right]\right]\right]_{1 \times 1 \times 3}, \\
W &= \left[\left[\left[w_r, w_g, w_b\right]\right]\right]_{1 \times 1 \times 3},
\end{aligned}
\tag{5}
$$

where $\circ$ means element-wise product and subscripts s and l denote *short-exposure* and *long-exposure*, respectively. A color correction follows to adopt a 3×3 color correction matrix (CCM) to transform the color space of the camera to the output one, namely sRGB. We obtain the CCM M_{cc} from the meta-data of RAW files. To be specific, the matrix converting the camera color space into XYZ color space is usually recorded in RAW files or a configure file in the processing systems, *e.g.* being stored in EXIF, and the matrix parameter converting the XYZ color space into sRGB color

space is fixed. This module is formalized:

$$y^{lin} = y^{cc}{}_{3\times(m\times n)} = \begin{pmatrix} y_r^{cc} \\ y_g^{cc} \\ y_b^{cc} \end{pmatrix} = \begin{pmatrix} y_r^{wb} \\ y_g^{wb} \\ y_b^{wb} \end{pmatrix} M_{cc}. \qquad (6)$$

For convenience, y^{cc} is represented equivalently as y^{lin}. We call the procedure that converts y^{raw} into y^{lin} as *linear process*.

4) Gamma Compression and Tone Mapping: To make the images better perceived by humans, nonlinear procedures are further conducted, including Gamma compression as well as tone mapping [47]. For simplicity, more details about these two stages are skipped. We use a function $\sigma(\cdot)$ to denote the *nonlinear process* consisting of these two stages as follows:

$$y^{srgb} = \sigma(y^{cc}). \qquad (7)$$

These nonlinear procedures introduce considerable ambiguity for creating the inverse mapping of low-light image enhancement. For example, as shown in Fig. 2, if we cannot obtain the Gamma compression function accurately, a huge gap between the brightened images[2] by inverting two Gamma functions is incurred. It is demonstrated that, for different low-light images, the proper inverse Gamma functions should be adopted adaptively.

5) Quantization: Finally, the quantization $Q(\cdot)$ comes to turn the data with more fine-grained quantization levels into 8bit to obtain a more compact representation for saving storage as follows:

$$f^{srgb} = Q(y^{srgb}). \qquad (8)$$

D. Evaluation Framework: Factorized Enhancement Model

For the benchmark, we regard the low-light enhancement as a simulation of amplifying the exposure time during capturing, which has a concise mathematical form and yields conveniences for an accurate and controllable enhancement process of low-light images. With exposure amplified γ times, a corresponding long-exposure data y_l^{raw}, which is usually approximated as noise-free because of high Signal-Noise Ratio (SNR), can be represented as follows:

$$y_l^{raw} \simeq x_l = \gamma x_s, \qquad (9)$$

where x_s, x_l and $y_l^{raw} \in R^{m\times n\times 1}$ are the latent radiance value without any noise with a short-exposure shot, that with a long-exposure shot, and the measured noisy value captured with the long-exposure time, respectively. Therefore, if the exposure ratio in the normal-light environment is given, the low-light enhancement is intrinsically close to denoising on already properly brightened RAW images y_b^{raw}:

$$y_b^{raw} = \gamma y_s^{raw} \simeq \gamma x_s + \gamma n(x_s),$$
$$\gamma n(x_s)[i] \sim \mathcal{N}(0, \sigma_b^2[i]), \qquad (10)$$

---

[2]The brightened images are generated by sequential operations of inverted Gamma compression, multiplication with the ratio of exposure time, and Gamma compression.
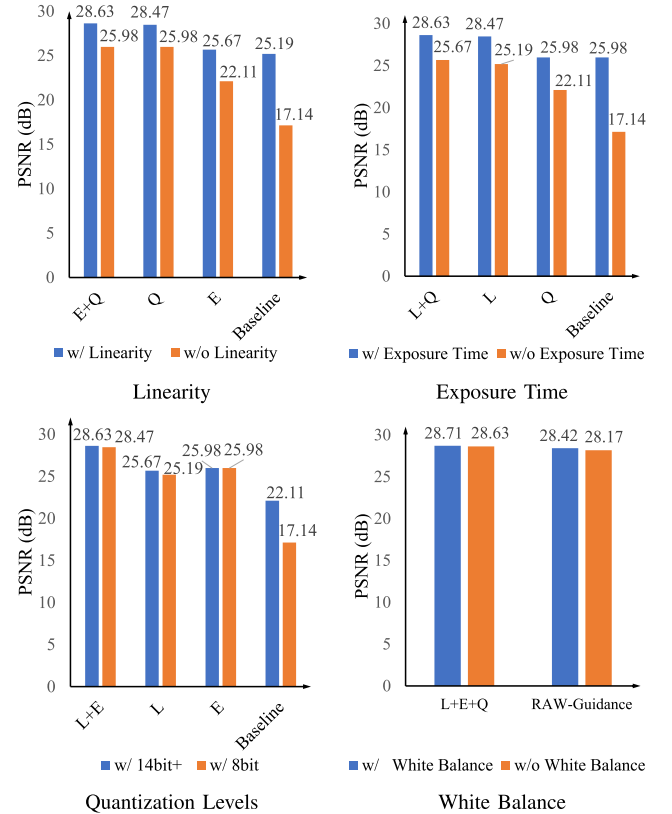

Fig. 3. Evaluation of Effects of RAW properties on low-light image enhancement. E denotes *exposure time*. Q denotes *dense quantizication levels*. L denotes *linearity*. Baseline denotes the method taking 8bit (sRGB) as the input. RAW-guidance denotes the proposed method in Sec. IV.

with $\sigma_b^2[i] = \gamma\,\lambda_{shot}(\gamma\,x_s[i]) + \gamma^2\lambda_{read}$, where subscript s signifies *short-exposure*, l represents *long-exposure*, and b means *brightened*. Therefore, the enhancement model $F_{enhance}(\cdot)$ can be represented as follows,

$$\hat{f}^{srgb} = F_{enhance}\left(y_b^{raw}\right). \qquad (11)$$

where $\hat{f}^{srgb}$ is the prediction of the enhancement model $F_{enhance}(\cdot)$. Eqn. (11) provides a flexible way to benchmark RAW utilization as shown in Fig. 1 (b). That is, y_b^{raw} can be replaced with any reasonable combination of images (RAW/sRGB images) and meta-data in the input combination module. After that, the input is feed-forwarded into a deep network for low-light image enhancement. From the performances of deep networks with different inputs, we can infer the importance of properties of RAW files for the enhancement.

E. RAW Benchmarking

In this benchmark, we compare several schemes of RAW data utilization with different inputs and guidance to explore how many contributions the characteristics of RAW data can bring in to the low-light enhancement task. The effects of different characteristics including linearity, exposure time and white balance parameters recorded in metadata, and quantization levels, denoted by L, E, W and Q, are analyzed with experimental results.

TABLE II
EVALUATION ON EFFECTS OF *linearity*. Z DENOTES ZERO-DCE, WHICH AIMS TO SIMULATE THE ILLUMINATION ADJUSTMENT GUIDED BY THE EXPOSURE TIME RATIO. $\hat{\gamma}$ DENOTES TO THE RATIO CALCULATED WITH THE MEAN PIXEL VALUE. $8bit(\cdot)$ DENOTES STORING DATA WITH 255 QUANTIFICATION LEVELS

Methods	Input	Model	Characteristics			PSNR↑	SSIM↑
			L	E	Q		
L+E+Q	RAW×γ	U-Net	✓	✓	✓	28.63	0.890
E+Q	[sRGB, γ]	Z+U-Net		✓	✓	25.98	0.821
L+E	8bit(RAW×γ)	U-Net	✓	✓		28.47	0.889
E	[sRGB, γ]	Z+8bit+U-Net		✓		25.98	0.820
L+Q	RAW×$\hat{\gamma}$	U-Net	✓		✓	25.67	0.855
Q	sRGB	U-Net			✓	22.11	0.772
L	8bit(RAW×$\hat{\gamma}$)	U-Net	✓			25.19	0.846
Baseline	8bit(sRGB)	U-Net				17.14	0.374

TABLE III
EVALUATION ON EFFECTS OF *exposure time*. Z DENOTES ZERO-DCE, WHICH AIMS TO SIMULATE THE ILLUMINATION ADJUSTMENT GUIDED BY THE EXPOSURE TIME RATIO. $\hat{\gamma}$ DENOTES TO THE RATIO CALCULATED WITH THE MEAN PIXEL VALUE. $8bit(\cdot)$ DENOTES STORING DATA WITH 255 QUANTIFICATION LEVELS

Methods	Input	Model	Characteristics			PSNR↑	SSIM↑
			L	E	Q		
L+E+Q	RAW×γ	U-Net	✓	✓	✓	28.63	0.890
L+Q	RAW×$\hat{\gamma}$	U-Net	✓		✓	25.67	0.855
L+E	8bit(RAW×γ)	U-Net	✓	✓		28.47	0.889
L	8bit(RAW×$\hat{\gamma}$)	U-Net	✓			25.19	0.846
E+Q	[sRGB, γ]	Z+U-Net		✓	✓	25.98	0.821
Q	sRGB	U-Net			✓	22.11	0.772
E	[sRGB, γ]	Z+8bit+U-Net		✓		25.98	0.820
Baseline	8bit(sRGB)	U-Net				17.14	0.374

TABLE IV
EVALUATION ON EFFECTS OF *quantization*. Z DENOTES ZERO-DCE, WHICH AIMS TO SIMULATE THE ILLUMINATION ADJUSTMENT GUIDED BY THE EXPOSURE TIME RATIO. $\hat{\gamma}$ DENOTES TO THE RATIO CALCULATED WITH THE MEAN PIXEL VALUE. $8bit(\cdot)$ DENOTES STORING DATA WITH 255 QUANTIFICATION LEVELS

Methods	Input	Model	Characteristics			PSNR↑	SSIM↑
			L	E	Q		
L+E+Q	RAW×γ	U-Net	✓	✓	✓	28.63	0.890
L+E	8bit(RAW×γ)	U-Net	✓	✓		28.47	0.889
L+Q	RAW×$\hat{\gamma}$	U-Net	✓		✓	25.67	0.855
L	8bit(RAW×$\hat{\gamma}$)	U-Net	✓			25.19	0.846
E+Q	[sRGB, γ]	Z+U-Net		✓	✓	25.98	0.821
E	[sRGB, γ]	Z+8bit+U-Net		✓		25.98	0.820
Q	sRGB	U-Net			✓	22.11	0.772
Baseline	8bit(sRGB)	U-Net				17.14	0.374

TABLE V
EVALUATION ON THE PERFORMANCE GAP BETWEEN *Quantify then Brighten* AND *Brighten then Quantify*. XX/XX MEANS PSNR↑/SSIM↑

Low-Light Input	RAW	RAW	sRGB
Brightening Method	×γ	×$\hat{\gamma}$	Zero-DCE
Brighten	28.63/0.890	25.67/0.855	25.98/0.821
Brighten then Quantify	28.47/0.889	25.19/0.846	25.98/0.820
Quantify then Brighten	12.36/0.137	—	16.89/0.285

TABLE VI
EFFECTS OF *white balance*. WHITE BALANCE PARAMETERS (W) ARE RECORDED IN META-DATA OF RAW FILES AND LINEAR PROCESS (P) IS DEFINED IN SEC. III-C. THE REENet$_{raw}$ AND REENet ADOPT PROPOSED RAW GUIDING STRATEGY AND RELATED DETAILS ARE ILLUSTRATED IN SEC. IV-B. IMPLEMENTING WHITE BALANCE ON RAW DATA HELPS BRIDGE THE GAP BETWEEN RAW AND RGB

Methods	Input	Guidance	PSNR↑	SSIM↑
L+E+Q+W	RAW × γ × W	–	28.71	0.890
L+E+Q	RAW × γ	–	28.63	0.890
REENet	sRGB	P(RAW × γ, W, M_{cc})	28.42	0.883
REENet$_{raw}$	sRGB	P(RAW × γ, 1, M_{cc})	28.17	0.882

TABLE VII
EVALUATION ON EFFECTS OF USING DIFFERENT WHITE BALANCE PARAMETERS FOR PROCESSING. L+E+Q ADOPTS AN END-TO-END RAW-TO-SRGB ARCHITECTURE. R2R$_l$ AND R2R$_s$ LEARN TO TURN LOW-LIGHT RAW IMAGES INTO NORMAL-LIGHT RAW IMAGES, VIA PERFORMING DENOISING ON BRIGHTENED RAW IMAGES, AND THEN PROCESSING THE RAW IMAGES. R2R$_s$ UTILIZES THE WHITE BALANCE PARAMETERS OF SHORT-EXPOSURE RAW AND R2R$_l$ USES THOSE OF LONG-EXPOSURE ONE. WB DENOTES WHITE BALANCE

Methods	Input	Target	WB Parameters	PSNR↑	SSIM↑
L+E+Q	RAW×γ	sRGB	—	28.63	0.890
R2R$_s$	RAW×γ	RAW	Short Exposure	28.20	0.882
R2R$_l$	RAW×γ	RAW	Long Exposure	29.50	0.889

1) Experimental Settings: SID dataset [1] is adopted for training and evaluation. We use Sony sub-dataset, constructed with a Sony α7S II equipped with a Bayer sensor. The subset contains 409 paired low/normal-light RAW images. The training, testing, and validation sets include 280, 93, and 36 paired images. Based on characteristics of RAW files mentioned in Section III-B, we employ different operations on input/target pairs and feed-forward them into the similar architecture *i.e.* U-Net [51] for performance comparisons. All approaches are trained from scratch on SID. For RAW based approaches, the training settings follow the paradigm of [1] *i.e.*, unpacking the RAW data with Bayer pattern into 4 channels, linearizing the data, and normalizing it into [0, 1]. Then, the data is fed into a U-Net [51]. For sRGB-based approaches, corresponding sRGB images are processed by Libraw, where the histogram stretching [1] is not adopted because it will brighten images during processing, which is far away from our both targets in benchmarking and developing a novel RAW utilization paradigm. The network is trained with an L_1 loss with normal-light sRGB images as ground truths. The benchmark results in PSNR ans SSIM are shown in Table II-VII. The extended tables with more metrics are provided in Table I-V of the supplementary material due to the limited space.

2) Linearity: We compare several groups of versions that pre-process the signal in the linear and nonlinear domains, respectively, as shown in Fig. 3 (a) and Table II (corresponding to Table I in the supplementary material). For the methods working in the nonlinear sRGB domain, Zero-DCE [41] is adopted to adjust the illumination in the sRGB domain guided by the exposure time ratio, *i.e.* taking the low-light input and the ratio as the input. It is demonstrated that, the methods working in the linear domain, where the illumination can be directly adjusted via multiplication with the ratio, significantly outperform the ones working in the nonlinear domain with performance gains over 2.49 dB in PSNR, over 0.069 in SSIM, about 0.7 in NIQE and 0.03 in LPIPS. The results illustrate the critical role of linearity of RAW data in low-light image enhancement.

3) Exposure Time: We also compare the methods that are assumed to obtain the exposure time or not, as shown in Fig. 3 (b) and Table III (corresponding to Table II in the

supplementary material). For the methods that do not have the ground truth exposure time ratio γ, we use the estimated $\hat{\gamma}$ with the mean pixel values of short/long-exposure linear data. It is observed that, using the ground truth exposure time ratio γ leads to significant performance improvement with performance gains over 2.96 dB in PSNR, over 0.035 in SSIM, about 0.2 in NIQE and over 0.03 in LPIPS, which demonstrates the exposure time recorded in meta-data as another dominating factor. Apparently, the estimated ratio $\hat{\gamma}$ makes the adjusted low-light images biased, with over-exposed and dark regions, which increases ambiguity in the low-light enhancement.

4) Quantization: We compare the methods taking the images with different quantization levels as their input as shown in Fig. 3-(c) and Table IV (corresponding to Table III in the supplementary material). It is observed that, more fine-grained quantization levels only lead to relatively small gains if the compression is implemented after brightening, called *Brighten then Quantize* strategy, as shown in the top three comparisons of Table IV (corresponding to Table III in the supplementary material) with a performance gain under 0.48 dB in PSNR and competitive performances in other metrics. However, if quantizing the data into 8-bit format before brightening, called *Quantize then Brighten* strategy, tremendous performance drops are observed in Table V. The performance gap originates from the dynamic range stretching that makes the brightened dark region have more fine-grained quantization levels and preserve more detailed signals. These results demonstrate the importance of compressing low-light images following *Brighten then Quantize* strategy.

5) White Balance: The white balance parameters recorded in the meta-data of RAW files also can contribute to low-light image enhancement. The experimental results in Table VI (corresponding to Table V in the supplementary material) and Fig. 3 (d) show the potential to improve the performance of RAW-based and the proposed RAW-guiding methods REENet by pre-processing RAW images with white balance parameters. A gain over 0.08 dB in PSNR is observed, meanwhile SSIM and NIQE improve slightly. In our comparisons, to utilize these parameters during training, we amplify the unpacked 4-channel linear data with the parameters.

We also study the related utilization in the RAW-to-RAW approaches and figure out how much the pre-processing can help bridge the gap between RAW and sRGB in Table VII. R2R$_l$ and R2R$_s$ are two RAW-to-RAW based methods. They are both end-to-end trained to target the ground-truth RAW data and then process them into sRGB images with Libraw, which uses short and long-exposure white balance parameters, respectively. It is observed that, short-exposure white balance parameters lead to a performance drop while the long-exposure ones improve the low-light enhancement performance. The performance gap comes from two reasons: 1) R2R$_l$ takes the same white balance parameters as the ground truth, which leads to similar reconstructed results to the ground truth; 2) the exposure time will affect the accuracy of the light metering in a camera, and the light metering with short exposure might be inaccurate *w.r.t.* ground truth [11].

6) Comparisons to State-of-the-Art Methods: the above-mentioned baselines, we also evaluate several state-of-the-art sRGB-based methods including HE [16], Dehazing [7], MF [31], MSR [8], LIME [52], BIMEF [26], BPDHE [5], LLNet [9], SICE [37], KinD [10], DeepUPE [38] and Zero-DCE [41], and RAW-based methods including EEMEFN [15] and ELD [14] on SID-Sony dataset and provide systematic benchmark results using various metrics including PSNR, SSIM [53], VIF [54], NIQE [55] and LPIPS [56], shown in Table VIII.

Apparently, there is still a huge performance gap between RAW-based and sRGB based approaches, mainly caused by the absence of *linearity*. Among RAW-based methods, the one equipped with the ground truth meta-data shows better performance, and when the ground truth *exposure time* label is absent, the performance drops a lot because it is quite difficult for the enhancement model to predict the illumination level accurately. The effect of *white balance* and *quantization* with *Brighten then Qunatify* strategy is relatively small but still benefits the enhancement. Among the sRGB-based methods, proposed REENet with RAW guiding strategy shows superior performance, and there are also performance drops in PSNR when some characteristics are absent but their measures are still higher than other methods that also takes sRGB images as input. Note that REENet$_{8bit}$ follows the traditional route – *Quantize then Brighten*, and E adopts *Brighten then Quantize* strategy for better quality. If *Quantize then Brighten* is adopted for E, *i.e.* the same settings as REENet$_{8bit}$, the PSNR will drop to 16.89 dB as shown in Table V.

7) Qualitative Evaluation: The corresponding qualitative results are shown in Fig. 4, where we only provide relatively reasonable results. Apparently, the methods utilizing white balance parameters show accurate colors, *e.g.* L+E+Q+W and L+E+Q. Note that all sRGB-based methods have applied white balance as a part of processing, which corrects the weight of RGB channels. Linearity and explicit exposure time induce the correct illumination, *e.g.* L+E+Q and L+Q, and suppress artifacts in Q and baseline. More fine-grained quantization levels make the results change little, *e.g.* L+E+Q and L+E. As shown in the bottom panel, EEMEFN's results have an obvious color bias. ELD achieves better visual quality by using additional synthetic data. Comparatively, our REENet restores visually pleasant colors and details. Note that, REENet does not need RAW files during the testing phase.

IV. REENet: RAW-Guiding Exposure Enhancement Network

From the benchmark results, the critical roles that the properties of RAW files play in the low-light enhancement are obviously observed, especially for linearity and exposure time. Therefore, we are inspired to construct a RAW-guiding Exposure Enhancement Network (REENet) to fully utilize characteristics of RAW files, fully considering the advantages of RAW files as well as its inaccessibility of RAW images if we do not hope to rebuild the ISP process during the testing process. To achieve this, REENet only gets access to RAW images in the training process. With the guidance of RAW

TABLE VIII

AN OVERVIEW OF BENCHMARK RESULTS ON SID. L DENOTES LINEARITY, E DENOTES EXPOSURE TIME, Q DENOTES DENSE QUANTIZATION LEVEL AND W DENOTES WHITE BALANCE. γ IS THE RATIO OF EXPOSURE TIME AND WHEN META-DATA IS ABSENT, IT IS REPLACED BY $\hat{\gamma}$, THE RATIO OF MEAN PIXEL VALUE. THE ZERO-DCE (Z) IMPLEMENTED AS PRE-PROCESSING IS TRAINED FROM SCRATCH WITH INPUTS CONCATENATED WITH RATIO AND L_1 LOSS WITH THE TARGET. R2R MEANS RAW-TO-RAW APPROACHES. KIND∗ IS TRAINED FROM SCRATCH ON SID. THE BOLD VALUE DENOTES THE BEST RESULT IN EACH CATEGORY

Category	Methods	Input	Model	PSNR↑	SSIM↑	VIF↑	NIQE↓	LPIPS↓
RAW based	L+E+Q+W	RAW $\times \gamma \times W$	U-Net	28.71	0.890	0.137	4.99	0.276
	R2R$_l$	RAW $\times \gamma$	U-Net	**29.50**	0.889	0.140	5.10	0.265
	R2R$_s$	RAW $\times \gamma$	U-Net	28.20	0.882	0.139	5.10	0.278
	L+E+Q	RAW $\times \gamma$	U-Net	28.63	0.890	0.137	5.04	0.272
	L+E	8bit(RAW $\times \gamma$)	U-Net	28.47	0.889	0.137	5.03	0.273
	L+Q	RAW $\times \hat{\gamma}$	U-Net	25.67	0.855	0.136	4.81	0.303
	L	8bit(RAW $\times \hat{\gamma}$)	U-Net	25.19	0.846	0.135	4.85	0.310
	EEMEFN [15]	RAW $\times 1, \times \gamma/2$	EEMEFN	29.21	**0.910**	0.144	5.52	0.280
	ELD [14]	RAW $\times \gamma$	U-Net	28.72	0.873	**0.149**	**4.48**	**0.255**
sRGB-based	REENet	[sRGB, γ]	REENet	**28.42**	**0.883**	0.137	**5.60**	0.322
	REENet$_{raw}$	[sRGB, γ]	REENet	28.17	0.882	0.141	5.88	**0.286**
	E+Q	[sRGB, γ]	Z+U-Net	25.98	0.821	**0.143**	5.70	0.315
	REENet$_{8bit}$	[8bit(sRGB), γ]	REENet	25.75	0.808	0.135	6.23	0.424
	E	[sRGB, γ]	Z+8bit+U-Net	25.98	0.820	**0.143**	5.71	0.314
	REENet$_{\hat{\gamma}}$	[sRGB, $\hat{\gamma}$]	REENet	24.18	0.811	0.138	6.82	0.410
	Q	sRGB	U-Net	22.11	0.772	0.139	5.71	0.372
	REENet$_{base}$	[8bit(sRGB), $\hat{\gamma}$]	REENet	22.42	0.802	0.135	6.21	0.422
	baseline	8bit(sRGB)	U-Net	17.14	0.374	0.078	6.87	0.533
	HE [16]	sRGB	–	5.90	0.028	0.095	7.71	0.968
	BPDHE [5]	sRGB	–	10.67	0.072	0.051	16.65	0.969
	Dehazing [7]	sRGB	–	12.81	0.103	0.077	8.09	0.784
	MSR [8]	sRGB	–	10.04	0.070	0.116	6.33	1.031
	MF [31]	sRGB	–	13.87	0.111	0.108	6.34	0.950
	LIME [52]	sRGB	–	12.59	0.102	0.118	6.06	0.980
	BIMEF [26]	sRGB	–	13.06	0.110	0.086	7.67	0.798
	LLNet [9]	sRGB	LLNet	14.21	0.221	0.047	7.65	0.693
	SICE [37]	sRGB	SICE	14.26	0.366	0.011	13.25	0.766
	KinD [10]	sRGB	KinD	13.50	0.109	0.048	9.68	0.718
	KinD∗ [10]	sRGB	KinD	17.87	0.611	0.081	9.71	0.629
	DeepUPE [38]	sRGB	DeepUPE	12.10	0.070	0.028	11.28	0.772
	Zero-DCE [41]	sRGB	Zero-DCE	12.38	0.137	0.050	10.82	0.768

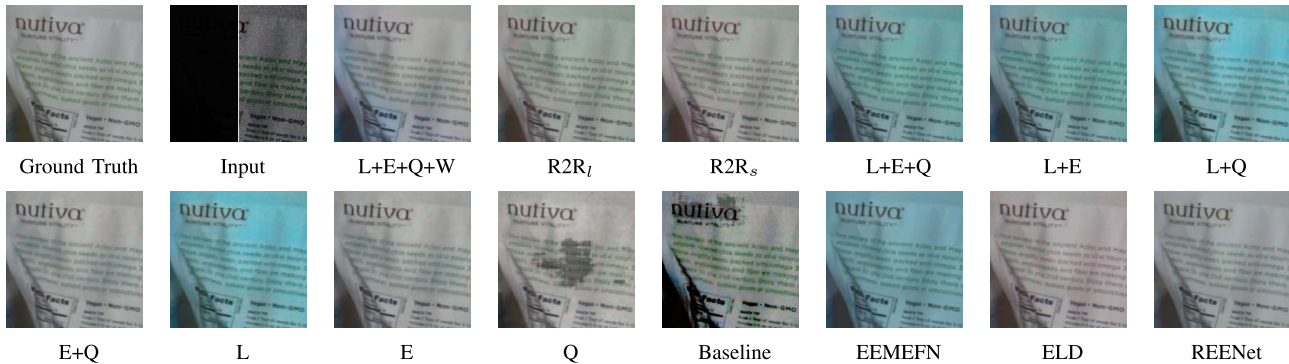| Ground Truth | Input | L+E+Q+W | R2R$_l$ | R2R$_s$ | L+E+Q | L+E | L+Q |
| E+Q | L | E | Q | Baseline | EEMEFN | ELD | REENet |

Fig. 4. Qualitative results of benchmark. Apparently utilizing more characteristics of RAW improves the final quality of results with less noise and artifacts. The right part of the input is brightened.

images in the training, REENet learns to project the nonlinear sRGB images into the linear domain, which is proven to be a better paradigm than directly learning to enhance images in the nonlinear domain. Furthermore, with the difficulty in reversing the total process in mind, REENet performs the enhancement in the linear RGB domain. We adopt the linear process to produce the linear RGB images, which are defined in Sec. III-C. Then, with the wealth of the meta-data of RAW files and the linearity, the gap between sRGB images and RAW images can be largely bridged.

It is noted that, because of the ill-posed nature of the low-light enhancement task, the perfect ground truth is quite hard to define. In *FEM*, the low-light image enhancement mainly focuses on suppressing noise and revealing detailed signals with a target (or given) exposure level. Although the ground truth images may not be perfect in providing a golden exposure level, the abundant information in the RAW image captured with a long exposure time also provides useful guidance for deriving a more effective enhancement model.

A. Model Architecture

As shown in Fig. 5 (a). REENet consists of three sub-modules:
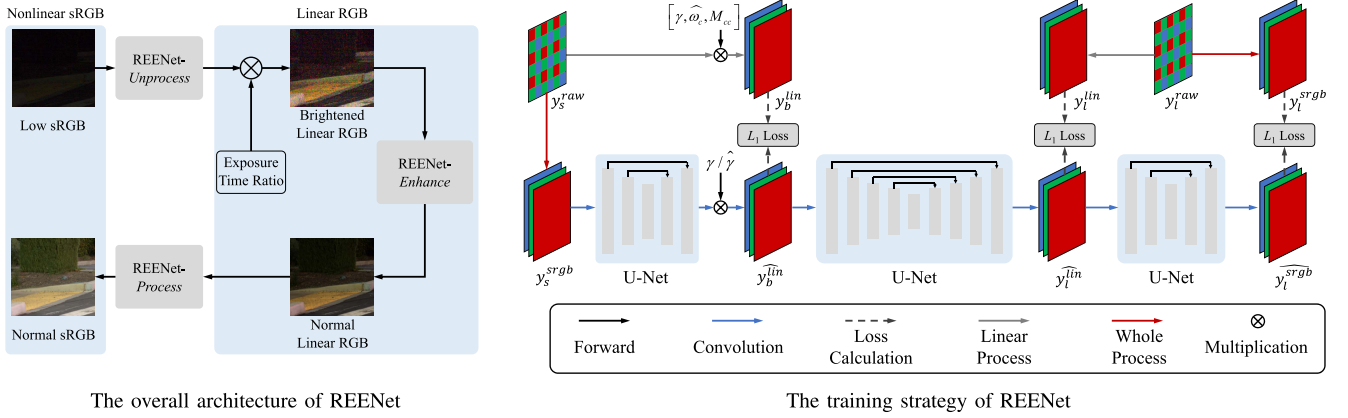
Fig. 5.   Illustration for the architecture of our RAW-guiding Exposure Enhancement Network (REENet). REENet makes full use of low/normal light RAW images in the training. Functionally, the three parts of the REENet, *i.e. Unprocess*, *Enhance* and *Process*, performs unprocessing, enhancement and processing, respectively.

- *Unprocess* to project sRGB images into the linear RGB domain;
- *Enhance* to suppress the amplified noise and color bias in the brightened images, which are adjusted by being multiplied with the exposure time ratio (ground truth or estimated);
- *Process* to project the enhanced results back into the nonlinear sRGB domain.

Note that, although the pipeline introduced in Section III-C is simplified, our developed *Unprocess* and *Process* are flexible and general frameworks to transform the signals between linear/nonlinear domains, which helps bridge the gap between the linear domain in image processing systems and the sRGB domain.

1) Unprocess: Transfer Nonlinear Data Into Linear Domain: Since sRGB images do not include meta-data, the conventional enhancement method [47] projects the processed nonlinear sRGB images back into the linear domain via hand-crafted approaches. Hence, the designed inversion process inevitably has a gap with the real processing approaches in various real applications, leading to inaccurate estimation. The gap might be further magnified especially, as shown in Fig. 2, when the exposure time ratio is multiplied. Therefore, an end-to-end convolutional neural network, *i.e.* a U-Net [51] $\sigma'(\cdot)$ is adopted for that. More exactly, given the processed input f_s^{srgb} and linear target y_s^{lin}, *Unprocess* aims to predict a brightened linear RGB image:

$$\widehat{y}_b^{lin} = \gamma' \widehat{y}_s^{lin} = \gamma' \widehat{\sigma}'(f_s^{srgb}), \qquad (12)$$

where $\gamma' = \gamma$ if the meta-data of the short-exposure RAW image is available during the testing phase, or $\gamma' = \hat{\gamma}$ if not, and f_s^{srgb} is quantized from y_s^{srgb} with 8 bits or 16 bits per pixel. The gap of f_s^{srgb} and y_s^{srgb} depends on the number of quantization levels, whose impact has been explored in our experiments.

2) Enhance: Normal-Light Image Reconstruction: The brightened linear RGB images, amplified by the exposure time

ratio, are:

$$y_b^{lin} = y_s^{lin} \times \gamma = \begin{pmatrix} \widehat{w}_r \gamma y_{s,r}^{raw} \\ \widehat{w}_g \gamma y_{s,g}^{raw} \\ \widehat{w}_b \gamma y_{s,b}^{raw} \end{pmatrix} M_{cc}. \qquad (13)$$

Compared to the long-exposed linear RGB images:

$$y_l^{lin} = \begin{pmatrix} w_r y_{l,r}^{raw} \\ w_g y_{l,g}^{raw} \\ w_b y_{l,b}^{raw} \end{pmatrix} M_{cc} = \begin{pmatrix} w_r \gamma x_{s,r} \\ w_g \gamma x_{s,g} \\ w_b \gamma x_{s,b} \end{pmatrix} M_{cc}, \qquad (14)$$

and according to Eqn. (10), *Enhance* targets to suppress the noise, whose noise levels are $\gamma'^2 \lambda_{read}$ and $\gamma' \lambda_{shot}$, respectively, and aims to compensate for the color casting caused by the inaccurate white balance $\widehat{W}$. Keeping the excellent modeling capacities of convolutional networks for image/video denoising [57], [58] and color correction [1], [11] in mind, a U-Net is adopted to build the architecture of *Enhance*. For simplicity, we use $g(\cdot)$ to denote the fitted denoising and color restoration processes. Given the brightened inputs $y_s^{lin} \times \gamma'$ and long-exposed linear targets y_l^{lin}, *Enhance* aims to estimate $\widehat{y}_l^{lin} = \widehat{g}(y_s^{lin} \times \gamma')$.

3) Process: Transfer Linear Data Into Nonlinear Domain: Similar to *Unprocess*, another U-Net is utilized for modeling the nonlinear process. To be exact, given the long-exposure linear input y_l^{lin} and the corresponding nonlinear sRGB target y_l^{srgb}, *Process* outputs $\widehat{y}_l^{srgb} = \widehat{\sigma}(y_l^{lin})$.

To summarize, our REENet predict the exposure time adjusted result of the input in real applications via:

$$\widehat{y}_l^{srgb} = \widehat{\sigma}(\widehat{g}(\widehat{\sigma}'(y_s^{srgb}) \times \gamma')). \qquad (15)$$

Note that, the testing phase can work without RAW files as input.

B. Experimental Results

1) Training Details: During the training process, the non-linear low/normal-light sRGB images processed by Libraw

TABLE IX
QUANTITATIVE EVALUATION COMPARING TRADITIONAL METHODS AND THE PROPOSED METHOD. THE BEST RESULT IS DENOTED IN BOLD

| Method	PSNR↑	PSNR∗↑	SSIM↑	SSIM∗↑	VIF↑	VIF∗↑	NIQE↓	NIQE∗↓	LPIPS↓	LPIPS∗↓
HE [16]	5.90	5.90	0.028	0.028	0.095	0.095	7.71	7.71	0.968	0.968
BPDHE [5]	10.67	10.79	0.072	0.188	0.051	0.052	16.65	16.66	0.969	0.970
Dehazing [7]	12.81	15.01	0.103	0.404	0.077	0.103	8.09	6.37	0.784	0.889
MSR [8]	10.04	10.04	0.070	0.327	0.116	0.116	6.33	6.33	1.031	1.031
MF [31]	13.87	14.17	0.111	0.387	0.108	0.108	6.34	6.39	0.950	0.960
LIME [52]	12.59	12.79	0.102	0.372	0.118	0.119	6.06	6.10	0.980	0.983
BIMEF [26]	13.06	14.95	0.110	0.410	0.086	0.104	7.67	9.30	0.798	0.890
REENet$_{8bit}$	25.75	25.75	0.808	0.808	0.135	0.135	6.23	6.23	0.424	0.424
REENet	**28.42**	**28.42**	**0.880**	**0.880**	**0.139**	**0.139**	**5.60**	**5.60**	**0.322**	**0.322**
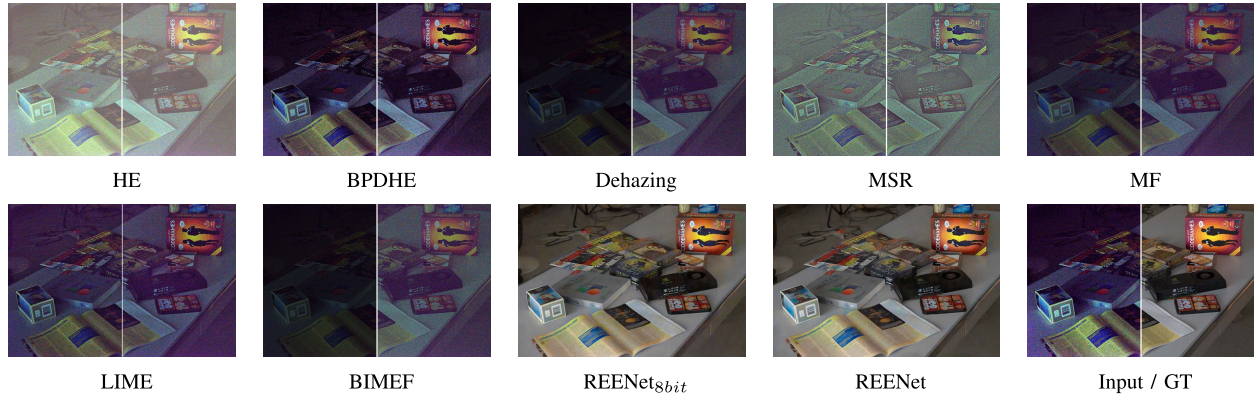
Fig. 6. Qualitative evaluation comparing traditional methods and the proposed method. **Left Panel**: Original results of competing methods. **Right Panel**: Gamma corrected results with aligned brightness. The last image is composed of the brightened input at the left and the *Ground Truth* at the right. Note that the input is almost totally invisible without brightening..
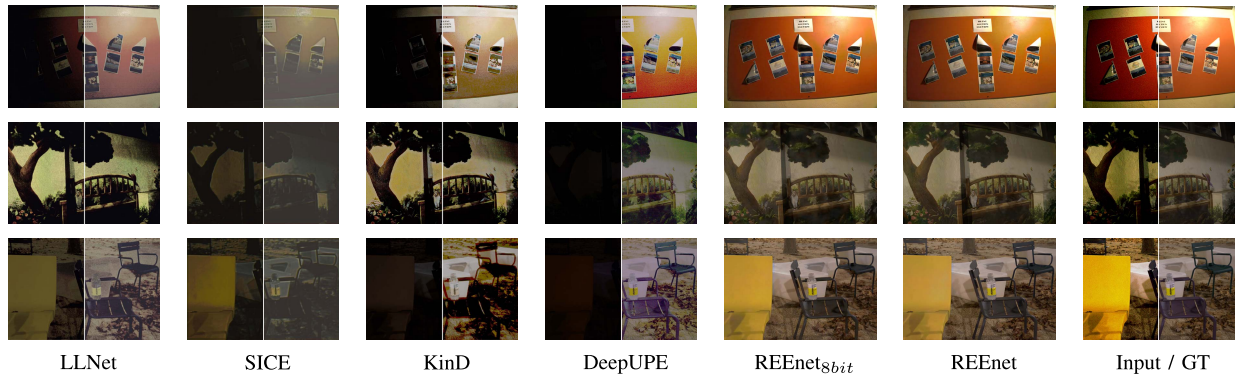

Fig. 7. Qualitative evaluation comparing learning-based methods and the proposed method. **Left Panel**: Original results of competing methods. **Right Panel**: Gamma corrected results with aligned brightness. The last image is composed of the brightened input at the left and the *Ground Truth* at the right. Note that the input is almost totally invisible without brightening..

are taken as inputs and ground truths, respectively. The RAW images are also employed as training guidance. Adam optimizer [59] and L_1 loss are adopted for training. The patch size and batch size are set to 512×512 and 1, respectively. The output results of all three sub-networks are clipped to [0, 1]. All sub-networks of REENet are pre-trained for 3000 epochs independently. We set the learning rate to 10^{-4} at the beginning and 10^{-5} after 2,000 epochs. After that, all sub-networks are trained jointly with the learning rate 10^{-5} for 1,000 epochs. REENet is trained on Intel(R) Xeon(R) E5-2650 2.20GHz CPU and an Nvidia RTX 2080Ti GPU in Python and Tensorflow. Because the input images have very large resolutions, we crop 4256×2848 images into

2128×1424 patches during the testing if needed. To avoid the blocking artifacts, we pad 200 pixels in the patch cropping for each patch. Because of the extremely dark settings of SID, we adopt 16-bit sRGB images as input to produce higher quality results, and also provide results of the proposed method trained and tested on 8-bit sRGB, named REENet$_{8bit}$.

2) Comparison to Conventional Methods: Our methods are compared with conventional methods: Dehazing [7], HE [16], MSR [8], MF [31], BIMEF [26], LIME [52], BPDHE [5]. In the quantitative evaluation, we adopt PSNR, VIF [54], SSIM [53] and LPIPS [56] as the full-reference metrics, and NIQE [55] as the no-reference metric. Considering that some methods do not aim to produce the targeted

TABLE X

QUANTITATIVE EVALUATION COMPARING LEARNING-BASED METHODS AND THE PROPOSED METHOD. THE BEST RESULT IS DENOTED IN BOLD

Method	PSNR↑	PSNR∗↑	SSIM↑	SSIM∗↑	VIF↑	VIF∗↑	NIQE↓	NIQE∗↓	LPIPS↓	LPIPS∗↓
LLNet [9]	14.21	17.91	0.221	0.547	0.047	0.064	7.65	6.64	0.693	0.667
SICE [37]	14.26	17.31	0.366	0.621	0.011	0.049	13.25	15.27	0.766	0.715
KinD [10]	13.50	16.71	0.109	0.361	0.048	0.076	9.68	7.11	0.718	0.703
DeepUPE [38]	12.10	15.16	0.070	0.455	0.028	0.116	11.28	7.99	0.772	0.887
REENet$_{8bit}$	25.75	25.75	0.808	0.808	0.135	0.135	6.23	6.23	0.424	0.424
REENet	**28.42**	**28.42**	**0.880**	**0.880**	**0.139**	**0.139**	**5.60**	**5.60**	**0.322**	**0.322**
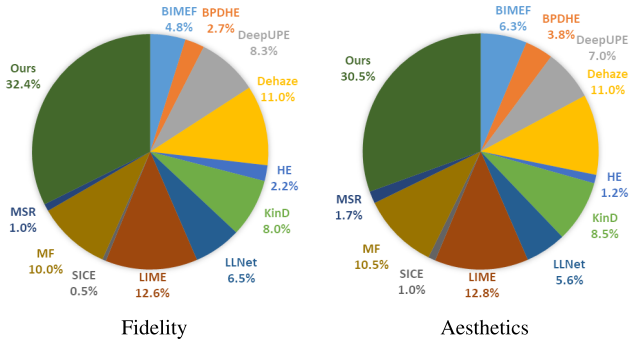
Fig. 8. The preference ratio on fidelity and aesthetics of different methods.
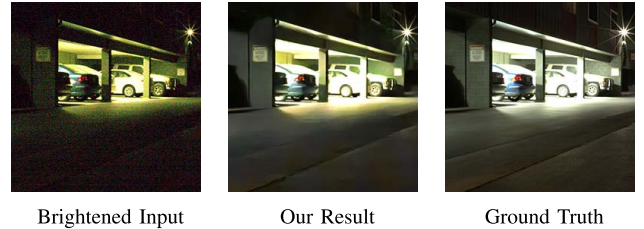

Brightened Input   Our Result   Ground Truth

Fig. 9. A failure case of the proposed method. There is obvious color casting in the input which looks more yellow, and in our result, this degradation is not handled perfectly.

TABLE XI

ABLATION STUDY OF OUR NETWORK ARCHITECTURE DESIGN

Architecture	PSNR↑
w/o RAW	22.11
w/o RAW, with MS-SSIM loss	21.88
w/o RAW, with estimated ratio $\hat{\gamma}$	24.27
w/ estimated ratio $\hat{\gamma}$	25.23
w/o *linear process*	28.17
w/ handcrafted inverse Gamma	23.90
REENet	**28.42**

illumination, we adjust the brightness of these results with Gamma correction, where each image chooses the Gamma curve with the best PSNR to produce the final result. Scores with brightness-aligned results are signified with ∗. The comparison results are presented in Table IX.

It is demonstrated that, our REENet achieves better performances than conventional methods on SID dataset in all metrics. We also show qualitative results in Fig. 6. It is showed that, conventional methods might brighten images uniformly with observed under/over-exposure regions. Besides, SID dataset's images are extremely under-exposed. Hence, the enhanced results might include intensive noise and severe color casting or insufficient illumination. After using a Gamma correction to adjust the brightness, there is still obvious noise and color casting in the results of other methods.

3) Comparison to Learning-Based Methods: The performances of different learning-based methods taking sRGB low-light images are compared on SID dataset, including SICE [37], LLNet [9], DeepUPE [38], and KinD [10]. When testing KinD, the brightening parameter is set to maximum allowed 5.0. We rescale the resolution of input images when testing SICE because of the GPU memory limit. The quantitative results are illustrated in Table X, with the same setting of metrics. It is demonstrated that, if all methods take the processed sRGB images as input during the testing phase, our REENet achieves better results on SID dataset. The qualitative results are presented in Fig. 7. It can be seen from Fig. 7 that, without the help of the ratio, the results of these methods are severely under-exposed on SID. Comparatively, we obtain visually pleasant results. The results further demonstrate the superiority of our method and show the role that the exposure time adjustment play in helping improve the generalized enhancement performance.

4) User Study: Besides full-reference image quality metrics PSNR, SSIM, VIF, LPIPS, and non-reference image quality metric NIQE, we further perform the user study to evaluate the image quality of enhanced results by different methods. Besides the four cases shown in Fig. 6 and Fig. 7, extra six cases are selected from the testing set to add up to 10 cases shown to the participants. Each subject is asked to select 3 from the 12 results that best match the target image (Fidelity) and have the best visual quality (Aesthetics). A total of 20 volunteers participate in this study and 400 selections are tallied. As shown in Fig. 8, the proposed REENet obtains the best average preference ratio of 32.4% and 30.5% for both the fidelity and aesthetics, respectively, outperforming other methods. Note that, for a fair comparison, we use REENet$_{8bit}$ here and align the brightness of results of other methods. The user study quantitatively verifies the superiority of our method.

5) Ablation Studies: We first perform the ablation study to evaluate the effectiveness of our architecture design in Table XI and Fig. 10. Firstly, we consider several versions *i.e.* (the top three methods) only making use of processed sRGB low/normal-light pairs in both training and testing. It is observed that, very low PSNRs are obtained, even with the advanced loss, *i.e.* MS-SSIM, and the estimated exposure time ratio $\hat{\gamma}$ via the mean pixel values.

The experiment with the estimated exposure time ratio $\hat{\gamma}$ instead of the ground truth demonstrates the effect of utilizing the meta-data. The results using original RAW images reflect the importance of adopting linearly preprocessing RAW

Fig. 10. Visual comparison of different settings described in Table XI. The obvious subjective superiority of the proposed REENet proves the efficiency of our design.

TABLE XII
ABLATION STUDY FOR SUB-NETWORKS TAKING DIFFERENT INPUTS

Inputs	REENet			PSNR↑
	Unprocess	*Enhance*	*Process*	
Low sRGB	✓	✓	✓	28.42
Brightened Linear RGB	✗	✓	✓	28.50
Normal Linear RGB	✗	✗	✓	42.68

TABLE XIII
RUNNING TIME (RT) EVALUATION COMPARING THE PROPOSED METHOD AND OTHERS. DUE TO THE HIGH RESOLUTION OF THE SID DATASET, INFERENCE TIME IS LONGER THAN USUAL AND THE PROPOSED REENET HAS A MIDDLE-LEVEL VELOCITY

Method	RT (Sec.)↓	Platform
HE [16]	5.04	MATLAB (CPU)
BPDHE [5]	7.92	MATLAB (CPU)
Dehazing [7]	19.74	MATLAB (CPU)
MSR [8]	39.76	MATLAB (CPU)
MF [31]	32.27	MATLAB (CPU)
LIME [52]	8.49	MATLAB (CPU)
BIMEF [26]	23.5	Matlab (CPU)
LLNet [9]	326.98	Tensorflow (CPU)
SICE [37]	53.21	Tensorflow (CPU)
DeepUPE [38]	52.67	Tensorflow (GPU)
KinD [10]	6.31	Tensorflow (GPU)
REENet	8.31	Tensorflow (GPU)

images. We can see the drop in the measures, as the adopted *linear process* is effective in filling in the gap between RAW image and processed sRGB ones, which helps reduce the difficulty in simulating the whole processing system. Using handcrafted inverse Gamma algorithm [60] instead of U-Net as *Unprocess* module will also cause a performance drop because unprocessing becomes less flexible and effective to deal with extremely dark conditions.

The quantitative results of each subnet are also provided in Table XII. The gaps among these subnets show how different stages in our design affect the enhancement performance. It is observed from the results, *Unprocess* and *Process* make efforts in an accurate nonlinear mapping, which leads to a small performance drop in measures. The large gap between the Row. 1-2 and Row. 3 demonstrates that, it is quite challenging to predict the normal-light linear RGB images from the brightened images, which are severely degraded by intensive noise and color casting.

6) Failure Case: A failure case of our method is shown in Fig. 9. Once the input image is heavily degraded with color casting and proposed REENet can effectively enhance the illumination and suppress noise but still with obvious color casting.

7) Limitations: We have compared the running time of different methods as shown in Table XIII. Note that we adopt GPU to accelerate the running method if the code supports it. Due to the high resolution of the SID dataset, inference time is longer and the proposed REENet has a middle-level running time consumption.

Besides, due to the highly camera-specific intensive noise in images captured in extremely dark environments and the diversity of ISP, our REENet cannot guarantee a promising performance when directly being applied to low-light images captured from another camera whose noise model is far away

from our training set. We will address the issue in our future work.

V. CONCLUSION

In this paper, we make the first benchmarking effort to investigate the superiority of RAW for low light enhancement in detail. The characteristics of RAW files *i.e.* linearity, the access to meta-data, fine-grained information (more abundant intensities and colors), inconvenience for display and lack of efficiency are detailed and their effects on the low-light enhancement are illustrated with quantitative results. For a fair evaluation, we take a novel view to regard low-light enhancement in a Factorized Enhancement Model (FEM) and obtain a precise and explicit description to decompose the ambiguities of this task into several measurable factors. Based on useful insights obtained from the benchmarking results, the proposed REENet adopts RAW-guiding strategy, overcomes the issues brought by the nonlinearity of the sRGB images and the unavailability of RAW images in many applications, and outperforms many state-of-the-art sRGB-based approaches. Our framework only needs to use RAW images during the training phase and offers better results with only sRGB inputs in testing, hence our results absorb the RAW's information as much as possible in the training but do not rely on the RAW input and adjust the ISP process in real applications.

Experimental results show the superior performance of our method and the rationality of our model design.
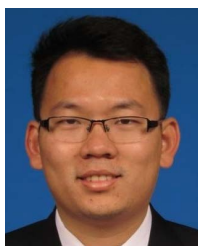
REFERENCES

[1] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3291–3300.

[2] H. Jiang and Y. Zheng, "Learning to see moving objects in the dark," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 7323–7332.

[3] M. Afifi, A. Abdelhamed, A. Abuolaim, A. Punnappurath, and M. S. Brown, "CIE XYZ Net: Unprocessing images for low-level computer vision tasks,"Ã,Âž 2021, *arXiv:2006.12709.*

[4] K. Xu, X. Yang, B. Yin, and R. W. H. Lau, "Learning to restore low-light images via decomposition- and-enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 2278–2287.

[5] H. Ibrahim and N. S. P. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 53, no. 4, pp. 1752–1758, Nov. 2007.

[6] T. Arici, S. Dikbas, and Y. Altunbasak, "A histogram modification framework and its application for image contrast enhancement," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 1921–1935, Sep. 2009.

[7] X. Dong *et al.*, "Fast efficient algorithm for enhancement of low lighting video," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2011, pp. 1–6.

[8] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell, "A multiscale Retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, Jul. 1997.

[9] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognit.*, vol. 61, pp. 650–662, Jan. 2017.

[10] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 1632–1640.

[11] C. Chen, Q. Chen, M. Do, and V. Koltun, "Seeing motion in the dark," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 3184–3193.

[12] M. S. Brown, "Understanding color and the in-camera image processing pipeline for computer vision," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Tutorial*, Oct. 2019, pp. 1–247.

[13] E. Schwartz, R. Giryes, and A. M. Bronstein, "DeepISP: Toward learning an end-to-end image processing pipeline," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 912–923, Feb. 2019.

[14] K. Wei, Y. Fu, J. Yang, and H. Huang, "A physics-based noise formation model for extreme low-light raw denoising," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 2755–2764.

[15] M. Zhu, P. Pan, W. Chen, and Y. Yang, "EEMEFN: Low-light image enhancement via edge-enhanced multi-exposure fusion network," in *Proc. AAAI Conf. Artif. Intell.*, Feb. 2020, pp. 13106–13113.

[16] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Müller, "Contrast-limited adaptive histogram equalization: Speed and effectiveness," in *Proc. Conf. Vis. Biomed. Comput.*, May 1990, pp. 337–345.

[17] M. Abdullah-Al-Wadud, M. H. Kabir, M. A. A. Dewan, and O. Chae, "A dynamic histogram equalization for image contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 53, no. 2, pp. 593–600, May 2007.

[18] C. Lee, J.-H. Kim, C. Lee, and C.-S. Kim, "Optimized brightness compensation and contrast enhancement for transmissive liquid crystal displays," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 4, pp. 576–590, Apr. 2014.

[19] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation of 2D histograms," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5372–5384, Dec. 2013.

[20] K. Nakai, Y. Hoshi, and A. Taguchi, "Color image contrast enhacement method based on differential intensity/saturation gray-levels histograms," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, Nov. 2013, pp. 445–449.

[21] X. Zhang, P. Shen, L. Luo, L. Zhang, and J. Song, "Enhancement and noise reduction of very low light level images," in *Proc. 21st Int. Conf. Pattern Recognit.*, Nov. 2012, pp. 2034–2037.

[22] L. Li, R. Wang, W. Wang, and W. Gao, "A low-light image enhancement method for both denoising and contrast enlarging," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 3730–3734.

[23] Q. Zhang, Y. Nie, L. Zhang, and C. Xiao, "Underexposed video enhancement via perception-driven progressive fusion," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 6, pp. 1773–1785, Jun. 2016.

[24] T. Celik and T. Tjahjadi, "Contextual and variational contrast enhancement," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3431–3441, Dec. 2011.

[25] S.-Y. Yu and H. Zhu, "Low-illumination image enhancement algorithm based on a physical lighting model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 1, pp. 28–37, Jan. 2019.

[26] Z. Ying, G. Li, and W. Gao, "A bio-inspired multi-exposure fusion framework for low-light image enhancement,"Ã,Âž2017, *arXiv:1711.00591.*

[27] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell, "Properties and performance of a center/surround Retinex," *IEEE Trans. Image Process.*, vol. 6, no. 3, pp. 451–462, Mar. 1997.

[28] C. Lee, J. Shih, C. Lien, and C. Han, "Adaptive multiscale retinex for image contrast enhancement," in *Proc. Int. Conf. Signal-Image Technol. Internet-Based Syst.*, Dec. 2013, pp. 43–50.

[29] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3538–3548, Sep. 2013.

[30] X. Fu, Y. Sun, M. LiWang, Y. Huang, X.-P. Zhang, and X. Ding, "A novel retinex based approach for image enhancement with illumination adjustment," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2014, pp. 1190–1194.

[31] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Process.*, vol. 129, pp. 82–96, Dec. 2016.

[32] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust Retinex model," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2828–2841, Jun. 2018.

[33] X. Ren, M. Li, W.-H. Cheng, and J. Liu, "Joint enhancement and denoising method via sequential decomposition," in *Proc. IEEE Int. Symp. Circuits Syst.*, Apr. 2018, pp. 1–5.

[34] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma, "MSR-Net: Low-light image enhancement using deep convolutional network," 2017, *arXiv:1711.02488.*

[35] L. Tao, C. Zhu, G. Xiang, Y. Li, H. Jia, and X. Xie, "LLCNN: A convolutional neural network for low-light image enhancement," in *Proc. IEEE Vis. Commun. Image Process.*, Dec. 2017, pp. 1–4.

[36] F. Lv, F. Lu, J. Wu, and C. Lim, "MBLLEN: Low-light image/video enhancement using cnns," in *Proc. Brit. Mach. Vis. Conf.*, Sep. 2018, p. 220.

[37] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 2049–2062, Apr. 2018.

[38] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 6842–6850.

[39] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *Proc. Brit. Mach. Vis. Conf.*, Sep. 2018, p. 155.

[40] Y. Jiang, X. Gong, D. Liu, Y. Cheng, and C. Fang, "EnlightenGAN: Deep light enhancement without paired supervision," *IEEE Trans. Image Process.*, vol. 30, pp. 2340–2349, 2021.

[41] C. Guo *et al.*, "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 1777–1786.

[42] W. Yang, S. Wang, Y. Fang, Y. Wang, and J. Liu, "From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 3060–3069.

[43] M. Gharbi, J. Chen, J. T. Barron, S. W. Hasinoff, and F. Durand, "Deep bilateral learning for real-time image enhancement," *ACM Trans. Graph.*, vol. 36, no. 4, p. 118, Jul. 2017.

[44] X. Yang, K. Xu, Y. Song, Q. Zhang, X. Wei, and R. W. Lau, "Image correction via deep reciprocating HDR transformation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1798–1807.

[45] M. Afifi, K. G. Derpanis, B. Ommer, and M. S. Brown, "Learning multi-scale photo exposure correction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 9157–9167.

[46] V. Wolf, A. Lugmayr, M. Danelljan, L. Van Gool, and R. Timofte, "DeFlow: Learning complex image degradations from unpaired data with conditional flows," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 94–103.

[47] T. Brooks, B. Mildenhall, T. Xue, J. Chen, D. Sharlet, and J. T. Barron, "Unprocessing images for learned raw denoising," in *Proc. CVPR*, Jun. 2019, pp. 11036–11045.

[48] S. W. Zamir et al., "CycleISP: Real image restoration via improved data synthesis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 2693–2702.

[49] S. W. Hasinoff, *Photon, Poisson Noise*. Boston, MA, USA: Springer, 2014, pp. 608–610.

[50] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1737–1754, Oct. 2008.

[51] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351, Oct. 2015, pp. 234–241.

[52] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2017.

[53] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncell, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2004.

[54] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[55] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[56] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.

[57] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3D?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2392–2399.

[58] P. Liu, H. Zhang, K. Zhang, L. Lin, and W. Zuo, "Multi-level wavelet-CNN for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 773–782.

[59] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, May 2015, pp. 1–15.

[60] S. Lin, J. Gu, S. Yamazaki, and H.-Y. Shum, "Radiometric calibration from a single image," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2004, pp. 938–945.

Yueyu Hu (Graduate Student Member, IEEE) received the B.S. and M.S. degrees in computer science from Peking University, Beijing, China, in 2018 and 2021, respectively. He is currently pursuing the Ph.D. degree with New York University, New York, NY, USA. His current research interest includes machine learning inspired 2D and 3D image compression and processing. He received the Best Paper Award at IEEE ICME-2020.

Jiaying Liu (Senior Member, IEEE) received the Ph.D. degree (Hons.) in computer science from Peking University, Beijing, China, 2010. She was a Visiting Scholar with the University of Southern California, Los Angeles, CA, USA, from 2007 to 2008. In 2015, she was a Visiting Researcher with Microsoft Research Asia, supported by the Star Track Young Faculties Award. She is currently an Associate Professor and a Boya Young Fellow with the Wangxuan Institute of Computer Technology, Peking University. She has authored more than 100 technical articles in refereed journals and proceedings. She holds 60 granted patents. Her current research interests include multimedia signal processing, compression, and computer vision. She is a Senior Member of CSIG and CCF. She has served as a member for Multimedia Systems and Applications Technical Committee (MSA TC) and Visual Signal Processing and Communications Technical Committee (VSPC TC) in IEEE Circuits and Systems Society. She has received the IEEE ICME 2020 Best Paper Award and IEEE MMSP 2015 Top10% Paper Award. She has served as the Technical Program Chair for IEEE ICME-2021/ACM ICMR-2021, the Area Chair for CVPR-2021/ECCV-2020/ICCV-2019, and the CAS Representative for the ICME Steering Committee. She has served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and *Journal of Visual Communication and Image Representation*. She was the APSIPA Distinguished Lecturer (2016–2017).

Haofeng Huang (Graduate Student Member, IEEE) received the B.S. degree in computer science from Peking University, Beijing, China, in 2021, where he is currently pursuing the Ph.D. degree with the Wangxuan Institute of Computer Technology. His current research interests include deep-learning-based image/video compression, image/video coding for machines, and intelligent visual enhancement.

Ling-Yu Duan (Member, IEEE) received the Ph.D. degree in information technology from The University of Newcastle, Callaghan, Australia, in 2008. He has been serving as the Associate Director for the Rapid-Rich Object Search Laboratory (ROSE), a joint lab between Nanyang Technological University (NTU), Singapore, and Peking University (PKU), China, since 2012. He has been with the Peng Cheng Laboratory, Shenzhen, China, since 2019. He is currently a Full Professor with the National Engineering Laboratory of Video Technology (NELVT), School of Electronics Engineering and Computer Science, PKU. He has published about 200 research papers. His research interests include multimedia indexing, search, and retrieval, mobile visual search, visual feature coding, and video analytics. He has received the IEEE ICME Best Paper Award in 2019/2020, the IEEE VCIP Best Paper Award in 2019, *EURASIP Journal on Image and Video Processing* Best Paper Award in 2015, the Ministry of Education Technology Invention Award (First Prize) in 2016, the National Technology Invention Award (Second Prize) in 2017, the China Patent Award for Excellence (2017), and the National Information Technology Standardization Technical Committee "Standardization Work Outstanding Person" Award in 2015. He serves as the Area Chair for ACM MM and IEEE ICME. He was a Co-Editor of MPEG Compact Descriptors for Visual Search (CDVS) Standard (ISO/IEC 15938-13) and MPEG Compact Descriptors for Video Analytics (CDVA) Standard (ISO/IEC 15938-15). He is an Associate Editor of IEEE TRANSACTIONS ON MULTIMEDIA, *ACM Transactions on Intelligent Systems and Technology*, and *ACM Transactions on Multimedia Computing, Communications, and Applications*.

Wenhan Yang (Member, IEEE) received the B.S. and Ph.D. (Hons.) degrees in computer science from Peking University, Beijing, China, in 2012 and 2018, respectively. He is currently a Postdoctoral Research Fellow with the Department of Computer Science, City University of Hong Kong. He has authored over 100 technical articles in refereed journals and proceedings. He holds nine granted patents. His current research interests include image/video processing/restoration, bad weather restoration, and human–machine collaborative coding. He has received the IEEE ICME-2020 Best Paper Award, the IFTC 2017 Best Paper Award, and the IEEE CVPR-2018 UG2 Challenge First Runner-Up Award. He was the Candidate of the CSIG Best Doctoral Dissertation Award in 2019. He has served as the Area Chair for IEEE ICME-2021 and the Organizer for IEEE CVPR-2019/2020/2021 UG2+ Challenge and Workshop.