Sparse Gradient Regularized Deep Retinex Network for Robust Low-Light Image Enhancement

Wenhan Yang, *Member, IEEE*, Wenjing Wang, *Graduate Student Member, IEEE*, Haofeng Huang, Shiqi Wang, *Member, IEEE*, and Jiaying Liu, *Senior Member, IEEE*

***Abstract*— Due to the absence of a desirable objective for low-light image enhancement, previous data-driven methods may provide undesirable enhanced results including amplified noise, degraded contrast and biased colors. In this work, inspired by Retinex theory, we design an end-to-end signal prior-guided layer separation and data-driven mapping network with layer-specified constraints for single-image low-light enhancement. A Sparse Gradient Minimization sub-Network (SGM-Net) is constructed to remove the low-amplitude structures and preserve major edge information, which facilitates extracting paired illumination maps of low/normal-light images. After the learned decomposition, two sub-networks (Enhance-Net and Restore-Net) are utilized to predict the enhanced illumination and reflectance maps, respectively, which helps stretch the contrast of the illumination map and remove intensive noise in the reflectance map. The effects of all these configured constraints, including the signal structure regularization and losses, combine together reciprocally, which leads to good reconstruction results in overall visual quality. The evaluation on both synthetic and real images, particularly on those containing intensive noise, compression artifacts and their interleaved artifacts, shows the effectiveness of our novel models, which significantly outperforms the state-of-the-art methods.**

***Index Terms*— Low-light enhancement, Retinex model, sparse gradient regularization, residual dense network, denoising.**

I. Introduction

IN LOW-LIGHT conditions, the captured images face several kinds of degradations, including low visibility, low contrast and intensive noise. With more advanced shooting devices and specialized photographic techniques, some of these degradations may be mitigated. However, the presence of noise can be hardly avoided by upgrading hardware facilities. In low-light conditions, there is insufficient amount of light reaching camera sensors, thus their output usually is contaminated by the intrinsic noise in the system. Using longer exposure time helps remove noise to some extent and leads to a higher signal-to-noise ratio (SNR). However, it also causes new problems, such as over-exposure and motion blur. Therefore, low-light enhancement in the sight of software is expected. It can effectively enhance visibility and visual quality of the input image. Furthermore, the enhanced visibility makes the scenes and objects more highlighted and provides a better starting point for high-level computer vision tasks (*i.e.* object detection and recognition).

It is non-trivial to enhance low-light images, especially in the presence of intensive noise, which makes it difficult to adjust the normal image signals. Furthermore, once low-light images are compressed, even with a high compression ratio, large quantization errors and the related visual artifacts, including blockiness and truncated random noise, will appear. Therefore, low-light image enhancement needs to improve the low visibility, suppress the intensive nosies and compression artifacts, and stretch the low contrast.

In the past decades, many researchers have been dedicated to low-light image enhancement. The earliest methods directly amplify the illumination globally. Later on, histogram equalization (HE) methods stretch the dynamic range of an image [1], [2]. Their results may present undesirable illumination with amplified intensive noise. Retinex theory-based methods [3] decompose an image into two components – reflectance and illumination, process these two components separately, and recombine them to obtain the final enhanced results. Several filters [4]–[6] are designed for that decomposition. Many works [7]–[9] impose priors on the decomposed illumination and reflectance. In [10], [11], the Retinex model is further extended to a robust one with an explicit defined noise term, which benefits developing joint low-light enhancement and noise suppression methods. Simultaneously, in [12], a new theoretical model, *i.e.* the absorption light scattering model, is proposed to improve the quality of low-light images using atmospheric light and transmittance. Recently, there are some new approaches using deep networks to facilitate low-light enhancement. In [13], Chen *et al.* proposed a learning-based pipeline to produce the processed normal-light images from low-light RAW format images. In [14]–[17], deep networks are built to enhance

Manuscript received June 6, 2020; revised October 19, 2020 and December 16, 2020; accepted January 5, 2021. Date of publication January 18, 2021; date of current version January 22, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2018AAA0102702, in part by the National Natural Science Foundation of China under Grant 62022002, in part by the Fundamental Research Funds for the Central Universities, in part by the Hong Kong RGC GRF under Grant 11203220, and in part by the National Natural Science Foundation of China under Contract 61772043. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Nikolaos Mitianoudis. *(Corresponding author: Jiaying Liu.)*

Wenhan Yang and Shiqi Wang are with Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: wyang34@cityu.edu.hk; shiqwang@cityu.edu.hk).

Wenjing Wang, Haofeng Huang, and Jiaying Liu are with the Wangxuan Institute of Computer Technology, Peking University, Beijing 100080, China (e-mail: daooshee@pku.edu.cn; huang6013@pku.edu.cn; liujiaying@pku.edu.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/TIP.2021.3050850, provided by the authors.

Digital Object Identifier 10.1109/TIP.2021.3050850

processed low-light images. Lore *et al.* [14] proposed a deep auto-encoder to perform contrast enhancement and denoising. In [15], local and global information is jointly captured to adjust the illumination of the input low-light images. In [16], a CNN-based single image contrast enhancer is designed to learn a mapping function between the low contrast input image and and its corresponding normal-image one. In [17], deep networks are designed to perform the Retinex decomposition and enhancement guided with the reflectance consistency constraint. In [18], the inherent connections between the spatially-varying noise and the illumination layer is explored and a framework is built to capture the interaction between noise estimation and illumination layer estimation in the bilateral space.

Although these previous methods can obtain good performance in some cases , there still exists some important issues as follows:

- In recent learning-based methods [14], [19], the synthesized low-light images used for training are usually generated by applying random Gamma transformation on normal light images. This synthesis process may not fully characterize the formation of natural low-light images. The real captured paired dataset is either in a small scale or in RAW format, which may not meet the requirement of most real applications.
- Some recent data-driven methods learn a direct mapping between low-light and normal light images [14], [19]. Due to the ill-posed nature, the enhanced results generated by the learned mapping function may present blurred details with biased and visually inauthentic colors.
- Most previous methods only consider parts of degradations in low-light enhancement. Specifically, few works pay attentions to suppressing the artifacts caused by the intensive noise after the quantization, which usually appear in real applications.
- In most current Retinex-based methods [8]–[11], [17], the decomposition process and constraints are either handcrafted signal-prior guided or learned freely from data. This leads to two drawbacks: 1) the illumination and reflectance decomposed from the low-light image maybe not desirable for the successive enhancement process; 2) the handcrafted signal-prior guided constraints are not adaptive enough to handle complex cases.

Considering these limitations of existing works, we explore possible deep learning architectures to effectively enhance low-light images suffering from low visibility, low contrast, intensive noise, and compression artifacts.

Specifically, we construct a large-scale **LO**w **L**ight (**LOL**) paired image dataset, including both real photos and synthesized data. The first type of images capture the degradation features and properties in real cases. The second one plays a role in data augmentation to synthesize paired images with diversified scenes and objects. Then, we build a deep network based on Retinex model (Retinex-Net) for robust low-light enhancement, which leads to visually pleasing and realistic enhanced results. For normal light images, we develop a trainable **sparse gradient minimization network** (**SGM-Net**), namely Decom-Net, to remove their low-amplitude structures and preserve major edges to obtain the illumination map. Then, an **Enhance-Net** is employed to predict the illumination of the normal light image. SGM-Net(Decom-Net), and Enhance-Net as well as Restore-Net (introduced later) are jointly trained. Therefore, the decomposed representations (illumination and reflectance layers) can better serve for the successive enhancement process. After the decomposition, the structure details are mainly preserved in reflectance. Thus, we build a **Restore-Net** to map the low-light reflectance to the normal one. This design benefits removing noise and compression artifacts, and stretching the contrast.

This article is an extension of our previous conference paper [20]. Based on our well developed LOL image dataset in the preliminary work, we choose an improved technical route to borrow from Retinex model to develop an end-to-end decomposition and mapping network for single-image low-light enhancement. Comparing with our previous work [20], the proposed method has four new features: 1) We add the sparse gradient constraint to regularize the output of the decomposition sub-network, which makes the learned decomposition results more reasonable; 2) After decomposition, the illumination of the normal-light image is directly predicted from the low-light image, which simplifies the learning paradigm and makes the prediction of the normal-light illumination map and the low-light image preciser; 3) A more expressive deep network – multi-scale residual dense network – is used as the backbone model of each sub-network; 4) Besides random noise, JPEG artifacts and the interleaved noise are further considered in low-light degradation. In summary, the contributions of our work are summarized as follows:

- We construct a new large-scale **LOL** paired image dataset consisting of two categories: real photography pairs and synthesized pairs from raw images. They together capture the degradation in real cases and have diversified scenes and objects.
- We develop a deep network based on Retinex model for low-light image enhancement. Combining the advantages of deep learning and Retinex model, our method can effectively enhance the illumination, and generate more visually pleasing and realistic enhanced results.
- A Restore-Net is built to restore the reflectance. It is capable of learning to suppress the intensive nosies and artifacts that appear in real cases.
- In the proposed deep Retinex framework, the sub-networks for image decomposition and illumination enhancement are jointly trained. Thus, the decomposed representations (illumination and reflectance) better serve for the successive enhancement process. The whole framework is flexible to handle complex cases.

The rest of this article is organized as follows. Section II briefly reviews the related work. Section III presents the proposed Retinex-Net for robust low-light image enhancement. Section III illustrates the built LOL dataset. Experimental results and concluding remarks are presented in Sections V and VI, respectively.

II. Related Work

A. Hand-Crafted Methods

Different image priors have been explored for single image low-light enhancement. A preliminary method is to directly amplify the illumination uniformly. Nevertheless, the regions that are already bright may be over-exposed and lose details. Histogram equalization (HE) can mitigate the problem and make dark image visible by stretching the dynamic range of an image [1], [2]. However, HE aims to enhance the contrast instead of adjusting the illumination. Thus, the results may present undesirable illumination and are easily contaminated by intensive noise in low-light images.

Some methods [21], [22] regard the inverted low-light images as haze images, and enhance the visibility by applying dehazing. Then, the dehazing result is inverted as the enhancement result. These methods also consider the noise suppression. In [22], a joint-bilateral filter is applied after enhancement. In [21], adaptive BM3D denoising operations [23] are conducted on segmented superpixels. These methods obtain reasonable results. However, a convincing physical explanation on their basic model is missing, and applying denoising as postprocessings may lead to blurred details.

B. Retinex Theory Based Methods

To simultaneously suppress the noise and preserve high-frequency details, a series of methods built on Retinex theory [3]. The theory assumes that, images can be decomposed into two components: reflectance and illumination. Then, the enhanced results are obtained by further processing and combining these two parts. Single-scale Retinex [4] is the seminal work. It defines a practical implementation of Retinex – center/surround Retinex, and treats the reflectance as the final enhanced result. Multi-scale Retinex [5] creates the enhanced results by fusing different single-scale Retinex outputs. Wang *et al.* [6] constructed a bright-pass filter for Retinex decomposition, and tried to preserve the naturalness while enhancing details in low-light images. Fu *et al.* [7] proposed an improved version by fusing different merits into a single one based on multiple derivatives of the estimated illumination. Guo *et al.* [8] proposed to refine an initial illumination map with a structure aware prior. In [9], a weighted variational model is proposed to impose better prior representation in the regularization terms. In [24], a novel Retinex-based fractional-order variational model is proposed. The image is decomposed, and the fractional-order gradient total variation regularization is adopted on both the reflectance component and the illumination component to obtain better estimated results. These methods consider less on the constraints on the reflectance, and the latent intensive noise in the low-light regions are usually amplified.

Li *et al.* [10] proposed to extend the traditional Retinex model to a robust one with an explicit noise term, and made the first attempt to estimate a noise map out of that model via an alternating direction minimization algorithm. Ren *et al.* [11] also aimed to enhance low-light images based on that robust Retinex model, and developed a sequential algorithm to estimate a piecewise smoothed illumination and a noise-suppressed reflectance. These methods show impressive results in removing small random noise. However, with only hand-crafted constraints, these methods fail to remove heavy noise with large gradients and structural noise including quantized errors and quantized random noise, which are usually presented in compressed images. Comparatively, in our work, we use a Retinex model-driven deep network for image decomposition, and use two deep networks for illumination adjustment and reflectance restoration, respectively. Due to learning from large collections of training data, our method can not only better enhance the visibility of dark details but also remove large structural noise.

C. Data-Driven Methods and Low-Light Image Datasets

In recent years, deep learning (DL)-based image processing applications have emerged with promising performance. These applications include denoising [25], completion [26], inpainting [27], super-resolution [28], deblurring [29], deconvolution [30], and style transfer [31], *etc*. There are also some recent works on bad weather restoration or image enhancement, such as dehazing [32], raindrop and dirt removal [33], and rain removal [28]. Besides, with the superior modeling capacity than shallow models, DL-based methods begin to solve harder problems, such as blind image denoising [34], and image compression [35].

Learning based low-light image enhancement methods have also been studied. Yang *et al.* [36] proposed to enhance low-light images by coupled dictionary learning. Lore *et al.* [14] used a deep auto-encoder named Low-Light Net (LLNet) to perform contrast enhancement and denoising. In [19], deeply root in multi-scale Retinex representation, a feed-forward convolutional neural network with different Gaussian convolution kernels is proposed to learn an end-to-end mapping between dark and bright images. In these works, low-light images used for training is synthesized by applying random gamma transformation on natural normal light images. This synthesis process may not fully characterize the formation of natural low-light images, which may lead to unnatural results.

Some recent works aim to build paired training data from real scenes. In [13], Chen *et al.* introduced a dataset See-in-the-Dark (SID) of short-exposure low-light raw images with corresponding long-exposure reference raw images. However, there is one limitation for this database, because in most applications, such as video surveillance, it is more preferred to process coded images instead of RAW ones. Because many equipments do not support to output RAW format images and it is too expensive to directly enhance them before encoding in the terminal platform. These RAW images can be converted as PNG or JPEG images. However, the setting of capturing RAW images makes the converted images not conform the character of natural low-light 8-bit images, which sets barriers to network learning. Cai *et al.* [16] built a dataset of under/over-contrast and normal-contrast encoded image pairs, in which the reference normal-contrast images are generated by Multi-Exposure image Fusion (MEF) or High Dynamic Range (HDR) algorithm. However, MEF or HDR results are different from images captured in real scenes.

Using these images as references may lead to inappropriate enhanced results. Comparatively, in our work, we build a large scale dataset with both synthesized and real captured paired low/normal-light images. The images are encoded in PNG format. The shooting parameters are adjusted to capture moderate noise to simulate the commonly seen scenes. The background categories of real captured data are limited. To make our dataset more diversified, we also synthesize low/normal-light pairs. We go beyond the simple random Gamma transformation scheme by analyzing the illumination distribution of low-light images and adjusting synthetic images according to the analysis results. Our models are trained on both synthesized and real data in Retinex domain, and generate illumination enhanced, noise-suppressed and natural looking results.

III. Retinex-Net for Robust Low-Light Image Enhancement

A. Retinex Model-Based Low-Light Image Enhancement

The classic Retinex theory models the human color perception. It assumes that the observed image S can be decomposed into two components, reflectance R and illumination I:

$$S = R \cdot I, \tag{1}$$

where R represents reflectance; I represents illumination; and $\cdot$ represents element-wise multiplication. Reflectance describes the intrinsic property of captured objects, which is considered to be consistent under any lightness conditions. The illumination represents the various lightness on objects.

For Retinex theory-based low-light enhancement, the low-light image S_{low} is first decomposed into illumination $\widetilde{I}_{\text{low}}$ and reflectance $\widetilde{R}_{\text{low}}$:

$$\left[\widetilde{I}_{\text{low}}, \widetilde{R}_{\text{low}}\right] = f_{\text{decom}}\left(S_{\text{low}}\right), \tag{2}$$

$$\left[\widetilde{I}_{\text{normal}}, \widetilde{R}_{\text{normal}}\right] = f_{\text{decom}}\left(S_{\text{normal}}\right), \tag{3}$$

where $f_{\text{decom}}(\cdot)$ signifies a corresponding learnable process (the same below for all $f(\cdot)$) to perform illumination and reflectance separation with a convolutional l_0 gradient minimization process which will be illustrated in Section III-C. Then, the enhanced illumination $\widehat{I}_{\text{normal}}$ and reflectance $\widehat{R}_{\text{normal}}$ are inferred based on $\widetilde{I}_{\text{low}}$ and $\widetilde{R}_{\text{low}}$:

$$\widehat{I}_{\text{normal}} = f_{\text{enhance}}\left(\widetilde{I}_{\text{low}}\right), \tag{4}$$

$$\widehat{R}_{\text{normal}} = f_{\text{restore}}\left(\widetilde{R}_{\text{low}}\right), \tag{5}$$

The final enhanced result $\widehat{S}_{\text{normal}}$ is reconstructed by

$$\widehat{S}_{\text{normal}} = \widehat{R}_{\text{normal}} \cdot \widehat{I}_{\text{normal}}. \tag{6}$$

where $\widehat{\ }$ signifies the predicted results and $\widetilde{\ }$ denotes the decomposition results. In our approach, $f_{\text{decom}}(\cdot)$ is not a hand-crafted filter but a learnable process that is optimized jointly with $f_{\text{enhance}}(\cdot)$ and $f_{\text{restore}}(\cdot)$ in a data-driven way. Furthermore, it might cause accumulated inaccuracy by predicting $\widehat{I}_{\text{normal}}$ based the predicted $\widetilde{I}_{\text{low}}$ first. Therefore, in our scheme, we turn Eq. (4) to:

$$\widehat{I}_{\text{normal}} = f_{\text{enhance}}\left(S_{\text{low}}\right). \tag{7}$$

TABLE I
The Functionality of Each Sub-Network in Retinex-Net

Sub-Network	SGM-Net (Decom-Net)	Enhance-Net	Restore-Net
Input	$S_{\text{normal}}(S_{\text{low}})$	S_{low}	$\widetilde{R}_{\text{low}}$
Output	$\widetilde{R}_{\text{normal}}, \widetilde{I}_{\text{normal}}(\widetilde{R}_{\text{low}}, \widetilde{I}_{\text{low}})$	$\widehat{I}_{\text{low}}$	$\widehat{R}_{\text{normal}}$
Process	$f_{\text{decom}}(\cdot)$	$f_{\text{enhance}}(\cdot)$	$f_{\text{restore}}(\cdot)$
Functionality	Decomposition	Illumination Enhancement	Reflectance Restoration

That is, the illumination layer of the normal-light image will be directly predicted based on the low-light image instead of the estimated low-light illumination.

B. Overall Network Architecture

Motivated by Retinex model, we design a deep Retinex-Net to perform the illumination/reflectance decomposition, illumination enhancement and reflectance restoration jointly. The whole network architecture is shown in Fig. 1. It consists of three stages: decomposition, adjustment, and reconstruction. Three sub-networks are built to model $f_{\text{decom}}(\cdot)$, $f_{\text{enhance}}(\cdot)$ and $f_{\text{restore}}(\cdot)$, respectively.

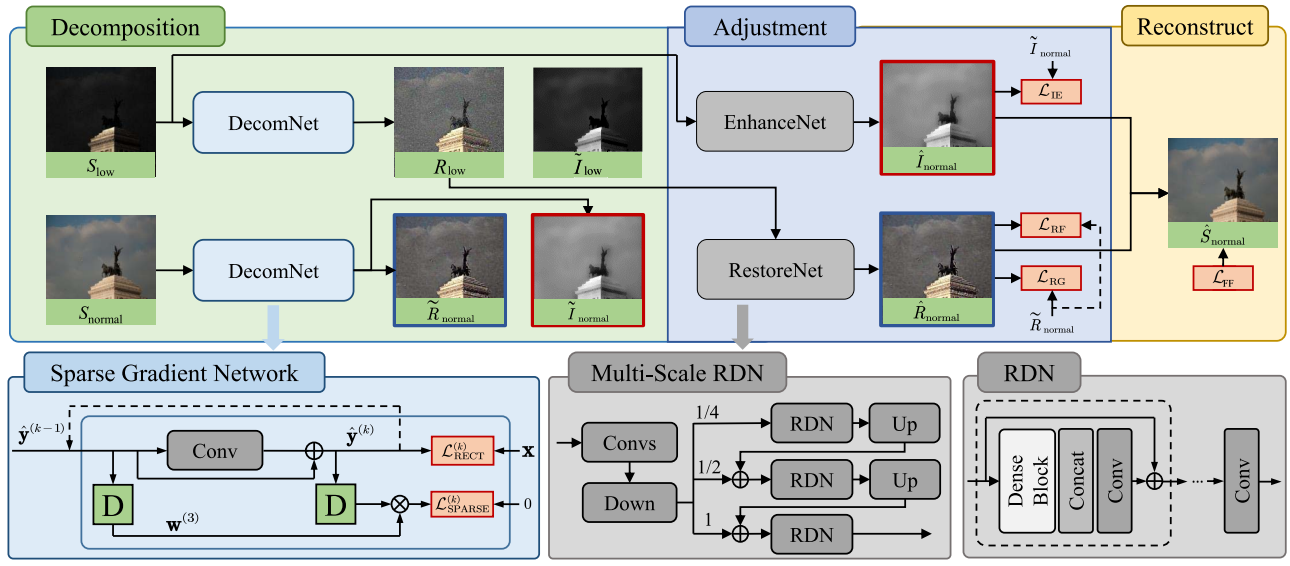
At the decomposition stage, a **sparse gradient minimization network** (**SGM-Net**) is designed (denoted as $f_{\text{decom}}(\cdot)$) to generate the illumination map $\widetilde{I}_{\text{normal}}$ and $\widetilde{I}_{\text{low}}$ which preserve major edges and remove low-amplitude structures, and its counterpart reflectance map $\widetilde{R}_{\text{normal}}$ and $\widetilde{R}_{\text{low}}$. The network design is inspired by an iterative solution of the l_0 gradient minimization problem – re-weighted l_1 minimization – and is embedded into the whole Retinex-Net for an end-to-end training.

At the adjustment stage, an **Enhance-Net** is used to predict $\widehat{I}_{\text{normal}}$, denoted by $f_{\text{enhance}}(\cdot)$. Furthermore, the noise in the reflectance $\widetilde{R}_{\text{low}}$ are suppressed by a **Restore-Net** to generate $\widehat{R}_{\text{normal}}$, denoted by $f_{\text{restore}}(\cdot)$. Finally, we combine the adjusted illumination and reflectance by element-wise multiplication at the reconstruction stage. The functionality of each sub-network is summarized in Table I. For Decom-Net, Enhance-Net and Restore-Net, their basic network structures are our proposed multi-scale residual dense network (MS-RDN) [37].

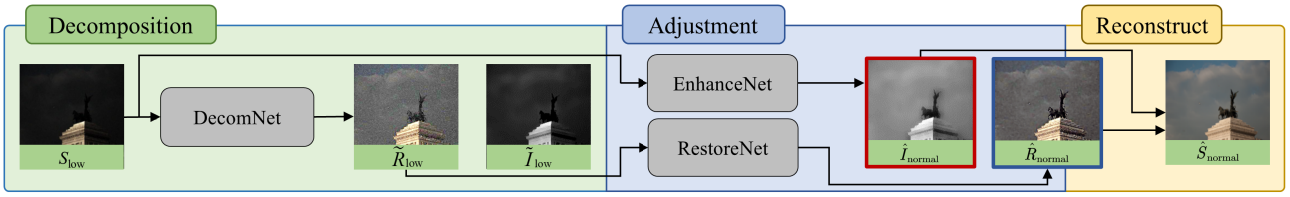
In the following, we will illustrate: 1) the detailed design of SGM-Net; 2) how SGM-Net (Decom-Net) and Enhance-Net work jointly to capture the illumination of low-light and normal light images simultaneously; 3) how Restore-Net is trained to remove noise from reflectance and reveal the structural details of low-light images; 4) the implementation details of our MS-RDN.

C. Sparse Gradient Minimization Network (SGM-Net)

In our work, we aim to decompose an input image into two parts: illumination and reflectance. We hope the former map preserves major edges, increases the steepness of transition and removes low-amplitude structures. Thus, we regard the illumination extraction as solving the problem of l_0 gradient minimization (l_0GM) [38] on an initially extracted illumination map. Compared to [38], we provide a learnable solution, which can be integrated into the whole Retinex-Net for an end-to-end training.



(a) Training Phase



(b) Testing Phase

Fig. 1. The proposed Retinex-Net for robust low-light image enhancement consists of three sub-networks. A trainable **sparse gradient minimization network** (**SGM-Net**), namely Decom-Net, is developed to remove low-amplitude structures and preserve major edges of normal and light images to obtain their illumination maps. An **Enhance-Net** is employed to predict the illumination layers of normal light images. SGM-Net (Decom-Net) and Enhance-Net are jointly trained. A **Restore-Net** is built to map the low-light reflectance to the normal-light one. Finally, the enhanced results are obtained by recombining the enhanced illumination and restored reflectance. The loss functions are denoted by orange blocks in subfigure (a).

Specifically, suppose that $\mathbf{x}$ is an image, $\hat{\mathbf{y}}$ is the smoothed result obtained by solving the following equation:

$$\hat{\mathbf{y}} = \arg\min_{\mathbf{y}} \|\mathbf{x}-\mathbf{y}\|_2^2 + \lambda \|\mathbf{D}\mathbf{y}\|_0, \tag{8}$$

where $\mathbf{D}$ denotes the Prewitt gradient operation.

Inspired by re-weighted l_1 minimization [39] that approximates l_0 minimization, we use re-weighted l_1 regularization to promote the gradient sparsity of an output image by CNN. Thus, we turn Eq. (8) to an iterative optimization process:

$$\begin{aligned}\hat{\mathbf{y}}^{(l)} &= \arg\min_{\mathbf{y}} \|\mathbf{x}-\mathbf{y}\|_2^2 + \lambda \left\|\mathbf{w}^{(l)}\mathbf{D}\mathbf{y}\right\|_1,\\ \mathbf{w}_{i,j}^{(l)} &= \frac{1}{\left(\mathbf{D}\mathbf{y}^{(l-1)}\right)_{i,j}+\epsilon},\\ \mathbf{w}_{i,j}^{(1)} &= 1,\end{aligned} \tag{9}$$

where i and j denote the spatial locations of pixels. ϵ is the parameter related to the sparsity. The iteration is terminated on convergence or l attains a specified maximum number of iterations max. After converting l_0 minimization to re-weighted l_1 minimization, all components are derivable. Thus, we can easily build an iterative network to solve Eq. (9) progressively. Here, we further make a slight change by setting a step-wise increasing $\epsilon^{(l)}$, which simulates the coarse-to-fine process to promote the sparsity:

$$\mathbf{w}_{i,j}^{(l)} = \frac{1}{\left(\mathbf{D}\mathbf{y}^{(l-1)}\right)_{i,j}+\epsilon^{(l)}}. \tag{10}$$

Following Eq. (9) and (10), we construct a fully convolutional network to perform the layer decomposition of illumination and reflectance maps, as shown in Fig. 2. The input $\mathbf{x}$ is transformed into a series of $\{\hat{\mathbf{y}}^{(k)}\}$ whose gradients are sparser along with an increased k via cascaded residual networks. Each network is trained guided by the step-wise losses in Eqn. (9). Specifically, in the stage k, the signal $\hat{\mathbf{y}}^{(k-1)}$ outputted from the previous stage $(k-1)$ is transformed into a sparser signal $\hat{\mathbf{y}}^{(k)}$. $\hat{\mathbf{y}}^{(k)}$ is constrained by the signal reconstruction loss:

$$\mathcal{L}_{\text{RECT}}^{(k)} = \left|\hat{\mathbf{y}}^{(k)} - \mathbf{x}\right|, \tag{11}$$

and its gradients are regularized by re-weighted l_1 total variation loss:

$$\mathcal{L}_{\text{SPARSE}}^{(k)} = \left|\mathbf{w}^{k}\mathbf{D}\hat{\mathbf{y}}^{(k)}\right|, \tag{12}$$

where $\mathbf{w}^k$ is determined by the gradient estimation $\left|\mathbf{D}\hat{\mathbf{y}}^{(k-1)}\right|$ from the stage $(k-1)$. The network will generate $\{\hat{\mathbf{y}}^{(k)}\}$ whose gradients are sparser along with an increasing k.

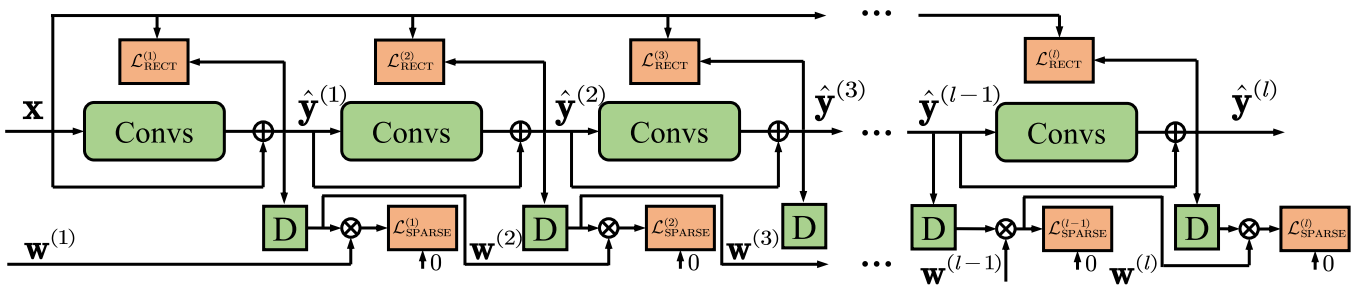


Fig. 2. The framework of our proposed l_0 gradient minimization network (SGM-Net). Inspired by the reweighed l_1 minimization, the network learns to promote l_0 gradient sparsity progressively. In the stage k, the signal $\hat{\mathbf{y}}^{(k-1)}$ outputted from the previous stage $(k-1)$ is transformed into a sparser signal $\hat{\mathbf{y}}^{(k)}$. $\hat{\mathbf{y}}^{(k)}$ is constrained by the signal reconstruction loss $\left|\hat{\mathbf{y}}^{(k)}-\mathbf{x}\right|$ and its gradients are regularized by re-weighted l_1 total variation loss $\left|\mathbf{w}^k\mathbf{D}\hat{\mathbf{y}}^{(k)}\right|$, where $\mathbf{w}^k$ is determined by the gradient estimation $\left|\mathbf{D}\hat{\mathbf{y}}^{(k-1)}\right|$ from the stage $(k-1)$. The input $\mathbf{x}$ is transformed into a series of $\{\hat{\mathbf{y}}^{(k)}\}$ whose gradients are sparser along with an increased k via cascaded residual networks. Each network is trained guided by the step-wise losses in Eqn. (9), namely reconstruction and sparse losses .

D. Coupled Illumination Estimation

To make the structural boundaries of low-light illumination more significant and those of both low-light and normal light illumination better corresponded, which benefit the successive low-light enhancement process, our method trains SGM-Net (Decom-Net), Enhance-Net and Restore-Net (which will be illustrated in Sec. III-E) jointly. The training loss consists of three terms:

$$\begin{aligned}\mathcal{L}_I &= \mathcal{L}_{\text{IE}} + \lambda_{\text{RECT}}\mathcal{L}_{\text{RECT}} + \lambda_{\text{SPARSE}}\mathcal{L}_{\text{SPARSE}},\\ \mathcal{L}_{\text{RECT}} &= \sum_i \mathcal{L}^{(i)}_{\text{RECT}},\\ \mathcal{L}_{\text{SPARSE}} &= \sum_i \mathcal{L}^{(i)}_{\text{SPARSE}},\end{aligned} \tag{13}$$

where λ_{RECT} and λ_{SPARSE} are the coefficients that control the importance of different terms. $\mathcal{L}_{\text{RECT}}$ and $\mathcal{L}_{\text{SPARSE}}$ jointly ensure that $\widetilde{I}_{\text{normal}}(\widetilde{I}_{\text{low}})$ are piecewise-smooth, preserve the main structure of $S_{\text{normal}}(S_{\text{low}})$ and remove low-amplitude structures. Illumination enhancement term $\mathcal{L}_{\text{IE}}$ maximizes Structural SIMilarity (SSIM) index distance [40] between the enhanced illumination $\hat{I}_{\text{normal}}$ and the decomposed illumination $\widetilde{I}_{\text{normal}}$ from normal light images,

$$\mathcal{L}_{\text{IE}} = -\text{SSIM}\left(\hat{I}_{\text{normal}}, \widetilde{I}_{\text{normal}}\right). \tag{14}$$

After obtaining $\widetilde{I}_{\text{normal}}$ and $\widetilde{I}_{\text{low}}$, $\widetilde{R}_{\text{normal}}$ and $\widetilde{R}_{\text{low}}$ are inferred by an element-wise division,

$$\widetilde{R}_{\text{normal}} = \frac{S_{\text{normal}}}{\widetilde{I}_{\text{normal}}}, \tag{15}$$

$$\widetilde{R}_{\text{low}} = \frac{S_{\text{low}}}{\widetilde{I}_{\text{low}}}. \tag{16}$$

Before division, the small values (smaller than 0.01) in $\widetilde{I}_{\text{normal}}$ and $\widetilde{I}_{\text{low}}$ are replaced by 0.01 to avoid too large values appeared in $\widetilde{R}_{\text{normal}}$ and $\widetilde{R}_{\text{low}}$.

E. Structure Revealing Reflectance Restoration

Low-light images usually suffer from low visibility, high-level noise and low contrast. As for the reflectance, it is usually contaminated by the latter two factors. Therefore, in Retinex-Net, we employ a Restore-Net to alleviate noise from reflectance and stretch its contrast.

Due to the ill-posedness nature, directly end-to-end learning a mapping between the reflectance may suffer from two drawbacks: 1) blurred details; 2) biased and visually inauthentic colors. Therefore, we propose a structure revealing reflectance restoration method. Considering that the characteristics of noise and edges are mainly reflected in the gradient domain, we try to fit both the signal and gradient of the reflectance of the normal-light images. The training loss consists of three terms:

$$\mathcal{L}_R = \mathcal{L}_{\text{RF}} + \lambda_{\text{FF}}\mathcal{L}_{\text{FF}} + \lambda_{\text{RG}}\mathcal{L}_{\text{RG}}, \tag{17}$$

where λ_{FF} and λ_{RG} is the coefficients that control the importance of different terms. The role of each term in the Eq. (17) is interpreted below:

- Reflectance fidelity term $\mathcal{L}_{\text{RF}}$ constrains the fidelity between the restored reflectance $\hat{R}_{\text{normal}}$ and the ground truth $\widetilde{R}_{\text{normal}}$,

$$\mathcal{L}_{\text{RF}} = -\text{SSIM}\left(\hat{R}_{\text{normal}}, \widetilde{R}_{\text{normal}}\right). \tag{18}$$

- Full fidelity term $\mathcal{L}_{\text{FF}}$ constrains the fidelity between the restored signal $\hat{I}_{\text{normal}}\hat{R}_{\text{normal}}$ and the ground truth signal S_{normal},

$$\mathcal{L}_{\text{FF}} = -\text{SSIM}\left(\hat{I}_{\text{normal}}\hat{R}_{\text{normal}}, S_{\text{normal}}\right). \tag{19}$$

- Reflectance gradient term $\mathcal{L}_{\text{RG}}$ minimizes the distance between the gradient of the restored reflectance $\hat{R}_{\text{normal}}$ and that of the normal-light one $\widetilde{R}_{\text{normal}}$,

$$\mathcal{L}_{\text{RG}} = \left\|\Delta\hat{R}_{\text{normal}} - \Phi(\Delta\widetilde{R}_{\text{normal}})\right\|_F^2, \tag{20}$$

where $\Phi(\cdot)$ will cuts-off the signal when the value is less than 0.15, which will benefit the noise suppression.

F. Network Structures: Residual Dense Network

We here briefly present our network structures of Decom-Net, Enhance-Net and Restore-Net. We propose a multi-scale residual dense network as our basic backbone because of its distinguished modeling capacity and moderate complexity.

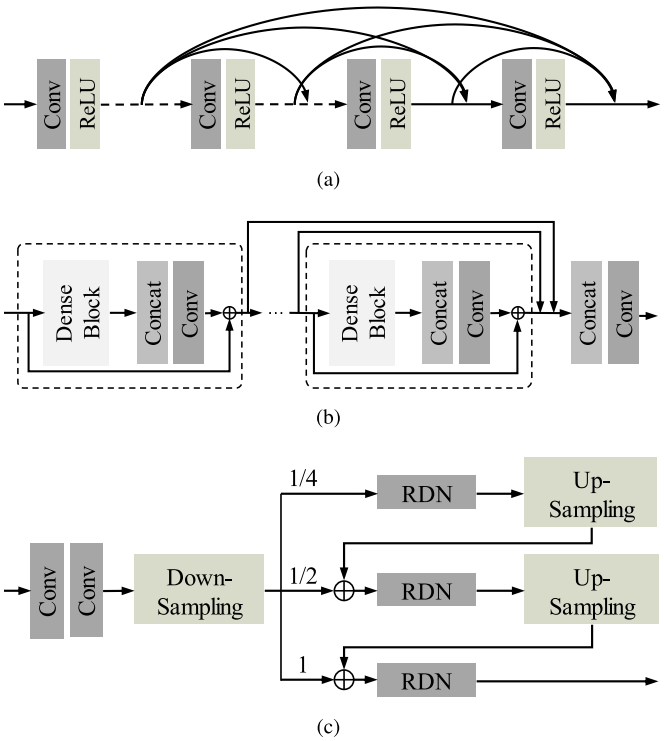


Fig. 3. Network architecture: (a) Dense block. (b) Residual dense network: dense blocks are cascaded by progressive channel compressions and residual skip connections. (c) Our proposed multi-scale residual dense network. The features are down-sampled to different scales (1/4, 1/2 and 1) and then are processed, un-sampled and fused progressively from small scales to the larger ones.

More details about residual dense network (RDN) can be found in [37].

The overall architecture is presented in Fig. 3. MS-RDN can be considered as the chained RDN with progressive down-sampling and up-sampling, which fuses information at different scales to infer the restored results.

Dense Network (Block). A preliminary effective backbone is the dense network as shown in Fig. 3 (a). A convolution layer passes its output feature to all preceding convolutional layers, which preserves the feed-forward nature and integrates the dense features at both local and global levels.

Residual Dense Block. The network in Fig. 3 (a) may face the problem of parameter explosion, due to the progressive increase of the channel number. To make the generated feature maps more compact, we add channel compression mechanism in the network, as shown in Fig. 3 (b). The whole network is splited into several dense blocks, each of which is fed into a concatenation layer and a 1×1 convolution layer for channel compression. Then, to make gradients better back-propagate to previous layers, we add residual skip connections to connect all these dense blocks as well as the successive concatenation and convolution layer.

Multi-Scale Residual Dense Block. Our proposed multi-scale residual dense network in Fig. 3 (c) can aggregate multi-scale information and obtain information of a larger region. The features are down-sampled to different scales (1/4, 1/2 and 1) and then are processed, un-sampled and fused progressively from a small scale to the larger ones.

IV. LOw Light (LOL) Dataset

In this section, we illustrate our LOL dataset used for training and testing.[1] To make it tractable to learn a low-light enhancement network, the constructed LOL dataset consisting of two categories: real photography pairs and synthesized pairs from raw images. The first one captures the degradation features and properties in real cases. The second plays a role in data augmentation, diversified scenes and objects. Because of the space limit, more detail information is provided in the supplementary material.

A. Dataset Captured in Real Scenes

The real captured dataset of LOL contains 500 low/normal-light image pairs. Most low-light images are collected by changing exposure time and ISO, while other configurations of the cameras are fixed. We capture images from a variety of scenes, *e.g.*, houses, campuses, clubs, streets.

Since camera shaking, object movement, and lightness changing may cause misalignment between the image pairs, inspired by [41], a three-step shooting strategy is used to eliminate such misalignments between the image pairs in our dataset. For one scene, we first shoot two normal-light images N_1 and N_2. Then, we change the exposure time and ISO to capture a series of low-light images. Finally, we set the exposure time and ISO back to shoot another two normal-light images N_3 and N_4. The average of N_i (i = 1,2,3,4) is treated as the ground-truth $G = \frac{1}{4}\sum_{i=1}^{4} N_i$. Then, we check whether there is object or camera movement. Specifically, the misalignment for these normal-light images is measured by $M = \frac{1}{4}\sum_{i=1}^{4} \mathrm{MSE}(N_i, G)$. If $M > 0.1$, we abandon the corresponding pair.

These raw images are resized to 400×600 and converted to Portable Network Graphics format. The dataset is publicly available. Fig. 6 shows a subset of the scenes.

B. Synthetic Image Pairs From Raw Images

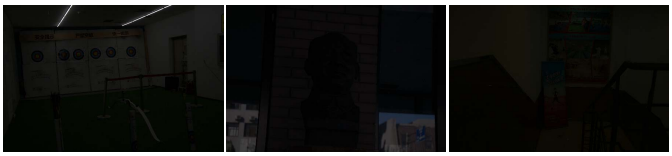
To make synthetic images match the property of real dark photography, we analyze the illumination distribution of low-light images. We collect 270 low-light images from public MEF [42], NPE [6], LIME [8], DICM [43], VV,[2] and Fusion [44] dataset, transform the imagesT into YCbCr channel and calculate the histogram of Y channel. We also collect 1000 raw images from RAISE [45] as normal-light images and calculate the histogram of Y channel in YCbCr. Fig. 4 shows the result.

Raw images contain more information than the converted results. For raw images, all operations used to generate pixel values are performed in one step on the base data, making the result more accurate. 1000 raw images in RAISE [45] are used to synthesize low-light images. Interface provided by Adobe Lightroom is used and we try different kinds of parameters to make the histogram of Y channel fit the result in low-light images. Final parameter configuration can be found in the supplementary material. As shown in Fig. 4,

[1] LOL dataset link: https://daooshee.github.io/BMVC2018website
[2] https://sites.google.com/site/vonikakis/datasets

TABLE II
QUANTITATIVE MEASUREMENT RESULTS ON SYNTHESIZED TEST IMAGES

Metrics	Input	BIMEF [46]	BPDHE [47]	CRM [48]	DHECI [49]	Dong [50]	EFF [51]	CLAHE [52]	LLNet [53]	DeepUPE [15]
PSNR	11.22	17.20	12.49	18.91	17.75	16.90	17.20	15.37	20.15	15.08
SSIM	0.3944	0.7127	0.5290	0.7864	0.7800	0.7487	0.7127	0.5527	0.8081	0.6225
UQI	0.5170	0.8647	0.6568	0.9383	0.8910	0.8914	0.8647	0.8880	0.9377	0.6343
OSS-PSNR	11.20	17.93	12.50	19.83	18.14	17.02	17.93	16.68	20.87	15.09
OSS-SSIM	0.2099	0.7541	0.4957	0.7189	0.4653	0.5901	0.7512	0.4476	0.8503	0.6695
Metrics	JED [11]	LIME [8]	MF [7]	MR [5]	RRM [10]	SRIE [9]	DRD [20]	SRCNN	SICE [16]	Proposed
PSNR	17.48	16.88	17.50	14.98	16.60	17.15	14.50	19.56	18.50	**22.05**
SSIM	0.7444	0.7762	0.7514	0.6248	0.7673	0.7277	0.6163	0.7922	0.7631	**0.9054**
UQI	0.8916	0.9298	0.8570	0.8832	0.8676	0.8857	0.7321	0.9238	0.7193	**0.9624**
OSS-PSNR	17.76	17.04	17.75	16.31	16.80	17.32	14.54	20.08	19.37	**22.55**
OSS-SSIM	0.7496	0.6978	0.7023	0.4132	0.7276	0.6531	0.7936	0.8397	0.8432	**0.9337**



(a) Low-light images.

(b) Normal light images.

Fig. 4. Several examples for low/normal light image pairs in LOL dataset. Objects and scenes captured in this dataset are diverse.

the illumination distribution of synthetic images matches that of low-light images. Finally, we resize these raw images to 400×600 and convert them to Portable Network Graphics format.

C. Features

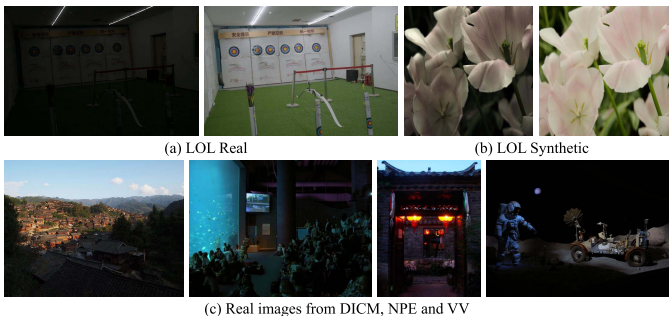
The built LOL dataset is advantageous compared with previous datasets in the following aspects:

- **Well-designed synthesis.** Instead of using the random gamma transform, we analyze the illumination distribution of low-light images and adjusting synthetic images based our analysis results.
- **Real captured data.** Compared with the datasets generated by Multi-Exposure image Fusion (MEF) [42] or High Dynamic Range (HDR) algorithm [16], our real captured paired data is more representative and can provide more useful guidance to train a low-light image enhancer.
- **Encoded format.** Compared with SID datasets in RAW format, our shooting parameters are adjusted to capture moderate noise to simulate the commonly seen scenes of encoded images.

Thus, based on these three properties, our LOL dataset includes commonly seen low-light degradation factors in real scenes with diversified objects and backgrounds. It is suitable for training a model with input and output images in the encoded format, which is beneficial to most applications.

V. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of RetinexNet for low-light enhancement. Due to the space limit, extensive



(a) LOL Real (b) LOL Synthetic

(c) Real images from DICM, NPE and VV

Fig. 5. Samples of testing images used in our experiments.

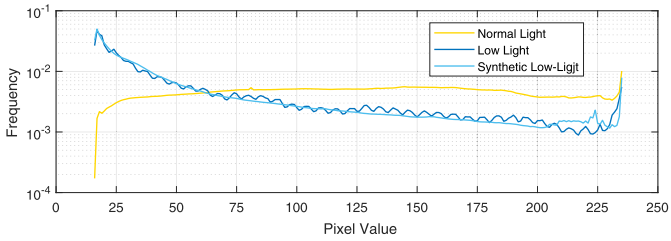


Fig. 6. Fitting results based on the histogram of Y channel in YCbCr. For clarity, the histogram is depicted in the form of curve graphs and the vertical axis is scaled in logarithmic domain. The horizontal axis represents the pixel value, noticing that Y channel ranges from 16 to 240.

experiments on a deeper analysis on our method are presented in the supplementary material.

A. Experiment Settings

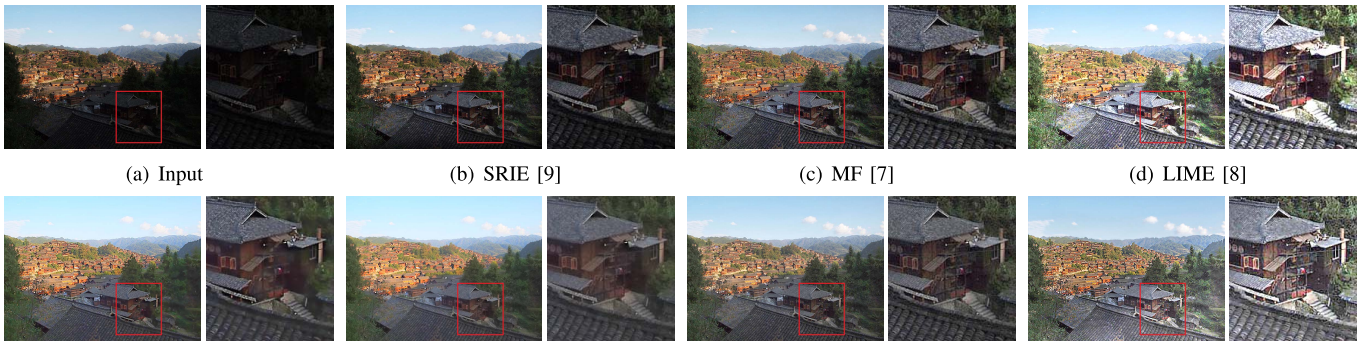
To fully evaluate the proposed method, we test our method on images from various scenes. Our LOL synthetic and real captured low\normal light images are used for objective evaluation. Several images from NPE [6], DICM dataset [43] and VV are used for subjective evaluation. Fig. 5 shows some testing images in our experiments. Their capturing devices are listed in Table IV.

All experiments are conducted in Pytorch-0.4.0 on a PC running Ubuntu 16.04.4 LTS OS with 64G RAM and Intel(R) Core(TM) i7-6850K 3.60GHz CPU.

The training images include both synthetic and captured paired data. We initialize the weights using the initialization proposed by [54]. Adam is used as the optimization method. The whole model, SGM-Net (Decom-Net), Enhance-Net and Restore-Net are trained jointly in an end-to-end manner. λ_{SPARSE}, λ_{RECT}, λ_{RG}, and λ_{FF} are set to 10^{-7}, 10^{-2}, 10, and

TABLE III
QUANTITATIVE MEASUREMENT RESULTS ON REAL TEST IMAGES

Metrics	Input	BIMEF [46]	BPDHE [47]	CRM [48]	DHECI [49]	Dong [50]	EFF [51]	CLAHE [52]	LLNet [14]	DeepUPE [15]
PSNR	9.72	17.85	13.84	19.65	14.64	17.26	17.85	13.13	17.56	13.27
SSIM	0.1752	0.6526	0.4254	0.6623	0.4450	0.5270	0.6526	0.3709	0.5481	0.4521
UQI	0.2937	0.8590	0.6887	0.9202	0.9182	0.8842	0.859	0.8800	0.8693	0.4420
OSS-PSNR	9.77	18.56	14.17	**20.78**	15.97	18.35	18.56	13.80	18.70	13.45
OSS-SSIM	0.4471	0.7233	0.4384	0.7341	0.4794	0.5781	0.7233	0.4504	0.7510	0.5263
Metrics	JED [11]	LIME [8]	MF [7]	MR [5]	RRM [10]	SRIE [9]	DRD [20]	SRCNN	SICE [16]	Proposed
PSNR	17.33	15.24	18.73	11.67	17.33	17.34	14.45	17.31	19.40	**20.06**
SSIM	0.6654	0.4702	0.5590	0.4269	0.5144	0.6859	0.5421	0.5173	0.6906	**0.8158**
UQI	0.8743	0.9245	0.9369	0.8051	0.9387	0.8735	0.7136	0.8492	0.5312	**0.9484**
OSS-PSNR	18.08	16.61	20.46	12.08	19.16	18.09	14.76	18.44	20.20	20.50
OSS-SSIM	0.7618	0.5108	0.6194	0.4784	0.7709	0.6090	0.5120	0.7220	0.7806	**0.8722**



(a) Input (b) SRIE [9] (c) MF [7] (d) LIME [8]

(e) RRM [10] (f) JED [11] (g) BIMEF [46] (h) RetinexNet

Fig. 7. Comparisons of low-light image enhancement results on a real image from NPE [6] with little noise.

100, respectively. The maximum of iteration l is set to 4. $\epsilon^{(l)}$ is set to [0.03, 0.06, 0.12, 0.18] for processing low-light images, and set to [0.09, 0.18, 0.27, 0.36] for processing normal-light images. Each mini-batch contains 8 image pairs. We set the learning rate to 0.0001. The model is trained for 200 epochs.

B. Quantitative Evaluation

We perform quantitative measurements to evaluate the performance of the proposed method. Since assessing the quality of enhanced images is not a trivial task, we adopt four evaluation metrics: Peak Signal-to-Noise Ratio (PSNR), Structural SIMilarity (SSIM) [55], Universal Quality Index (UQI) [56] and Optimal Scale Selection (OSS) [57].

PSNR is the ratio between the maximum possible power of the normal light image and the power of the enhanced image and measures the *fidelity* of between them. SSIM considers more on image structures, takes the image degradation as perceived change in *structural information* and incorporates luminance masking and contrast masking terms into the metric. UQI models image distortion as a combination of three different factors: loss of *correlation*, *luminance distortion*, and *contrast distortion*, thus it measures the image quality from these three perspectives. OSS models in spatial and discrete wavelet transform domains, and measures image quality by resizing filtered images to the optimal scale estimated based on *human eyes' physiological mechanism*.

1) Low-Light Enhancement for Noise-Free Inputs: For noise-free evaluation, we test all methods on our LOL synthetic dataset. For PSNR, SSIM, UQI, OSS-PSNR and OSS-SSIM, larger values represent better image qualities. It is observed from Table II that, our Retinex-Net achieves the best results among all methods on all metrics. It demonstrates the superiority of our Retinex-Net in fidelity, image structures, signal correlation, luminance, contrast and human perception.

2) Low-Light Enhancement for Noisy Inputs: For noisy case, we test all methods on our LOL real dataset. It is observed from Table III that, our Retinex-Net also achieves the best results among all methods almost on all metrics. It witnesses superiority and robustness of our Retinex-Net.

TABLE IV
THE SOURCE AND CAPTURING DEVICES OF TESTING IMAGES USED IN OUR WORK

Testing Images	Dataset	Source
Table II	LOL Synthesized	Nikon D7000
Table III	LOL Real	Nikon D5000
Fig. 8	VV	Panasonic DMC-FZ5
Fig. 7,9,11	NPE	Cannon Digital Camera, NASA and Google Websites
Fig. 10 and Fig. 12-16	DICM	Konica Minolta Photo Imaging DiMAGE G600

C. Qualitative Evaluation

1) Low-Light Enhancement Without Noise Suppression: Figs. 7-12 show several comparisons between enhancement results generated by different methods. Figs. 7-9 contain little noise to solely compare the illumination enhancement ability of different methods while Fig. 10-12 include intensive noise. In this part, no additional post-processing (noise suppression)

(a) Input (b) SRIE [9] (c) CRM [48] (d) LIME [8]

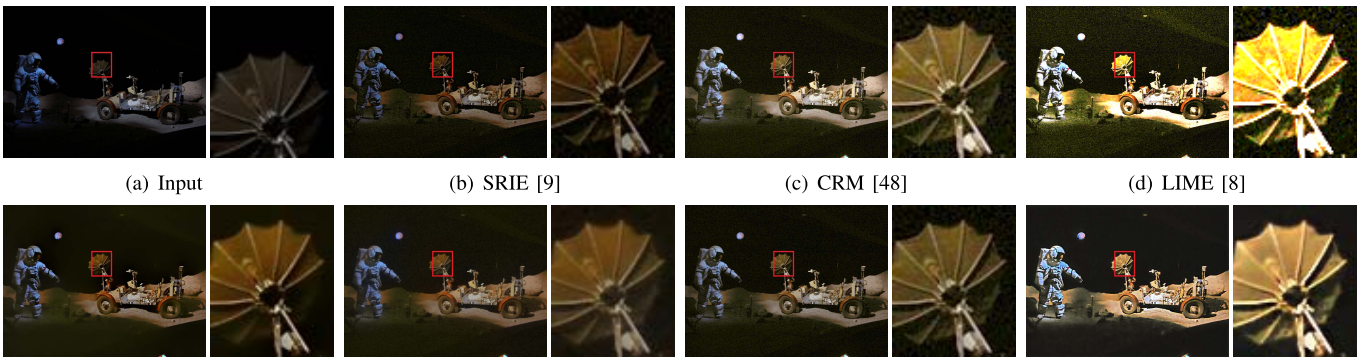
(e) RRM [10] (f) JED [11] (g) BIMEF [46] (h) RetinexNet

Fig. 8. Comparisons of low-light image enhancement results on a real image from VV with little noise.

(a) Input (b) SRIE [9] (c) CRM [48] (d) LIME [8]

(e) RRM [10] (f) JED [11] (g) BIMEF [46] (h) RetinexNet

Fig. 9. Comparisons of low-light image enhancement results on a real image from NPE [6] with little noise.



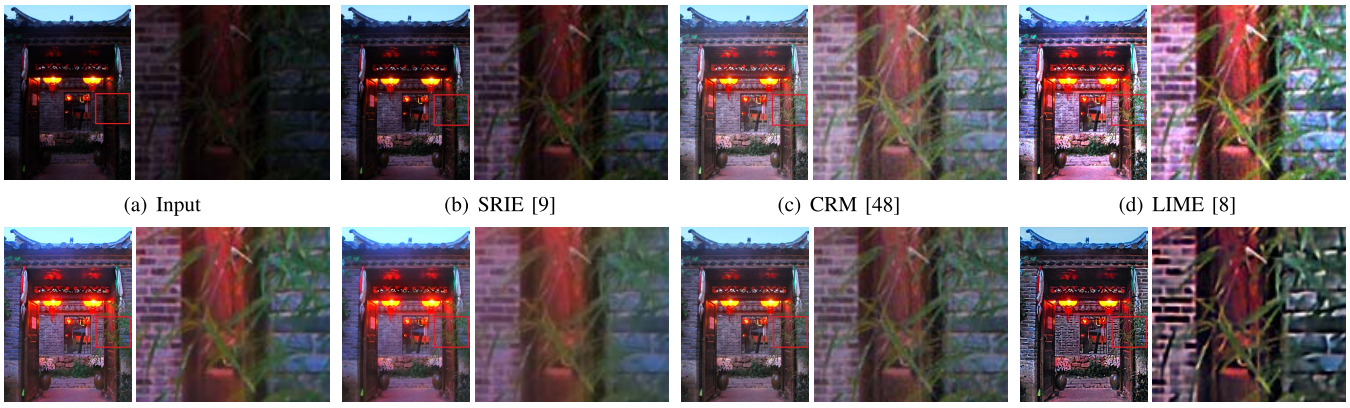
(a) Input (b) SRIE [9] (c) CRM [48] (d) LIME [8]

(e) RRM [10] (f) JED [11] (g) BIMEF [46] (h) RetinexNet

Fig. 10. Comparisons of noisy low-light image enhancement results on a real image from DICM [43]. All methods take no post-processing techniques.

operations are adopted. As seen in Fig. 7-9, our method can better capture the global illumination distribution of the whole image, enhance different regions properly, and generate rich textures, such as "shelf" in Fig. 7, "traffic line" in Fig. 8 and "wall" in Fig. 9 while avoiding darkness and over-exposure.

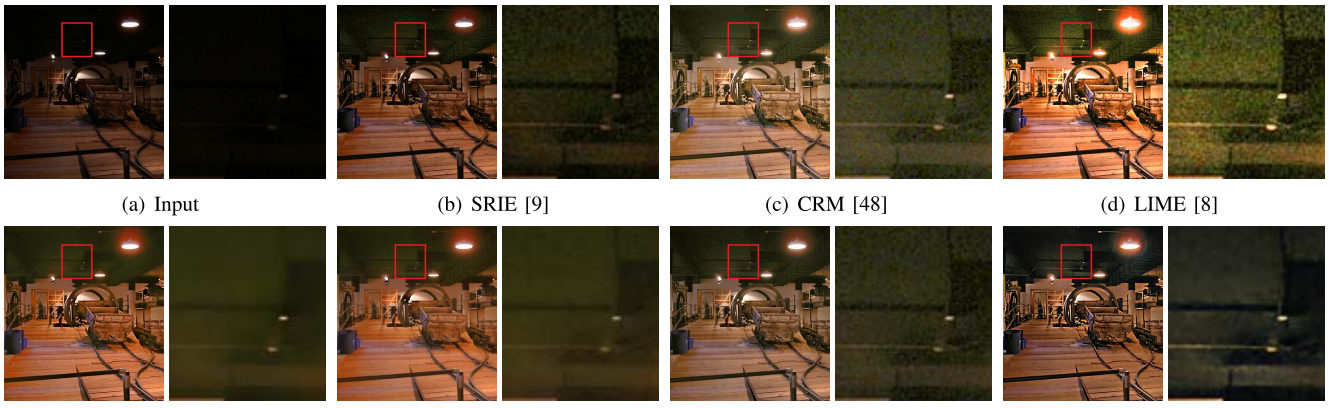
As can be observed in the Fig. 10-12, the noise hidden in very low-light condition becomes intense. SRIE cannot light up some parts of the input images, and its results also contain noticeable noise, *e.g.* the blocking artifacts in Fig. 10. CRM lights up images, however, noise is obviously observable and contrasts are degraded, *e.g.* the results in Fig. 12 and 11. LIME can sufficiently enhance the visibility of low-light images. It also amplifies the intensive noise and sometimes over-enhances local illuminations, as shown in Fig.10-12. RRM and JED can remove some noise to extent, however, some structure details are removed and the contrast is weakened. BIMEF sometimes generates dark results, such as the results shown in Fig. 11 and 12. Our method presents satisfying performance in handling low-light images with intensive noise. The noise, including compression artifacts, are successfully



(a) Input (b) SRIE [9] (c) CRM [48] (d) LIME [8]

(e) RRM [10] (f) JED [11] (g) BIMEF [46] (h) RetinexNet

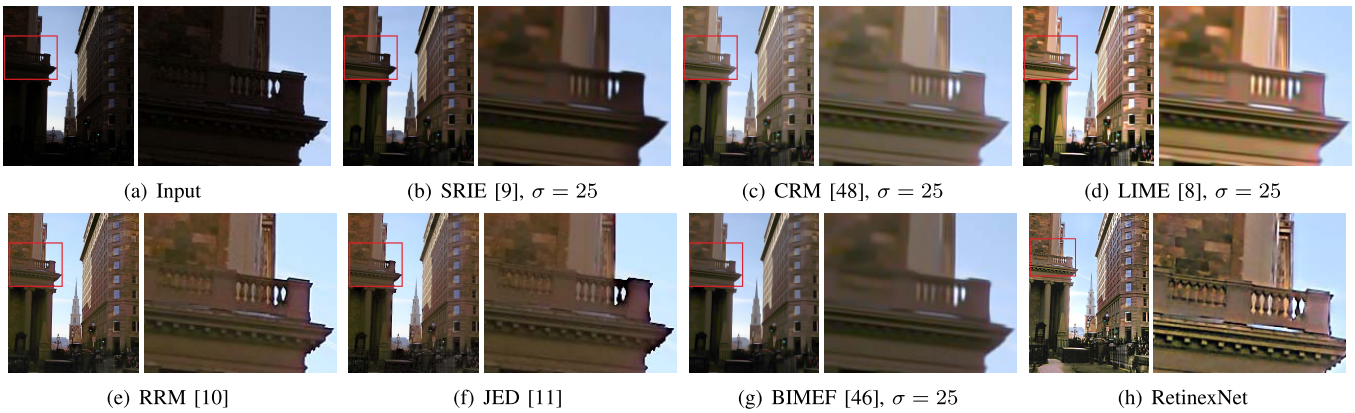
Fig. 11. Comparisons of noisy low-light image enhancement results on a real image from NPE [6]. All methods take no post-processing techniques.



(a) Input (b) SRIE [9] (c) CRM [48] (d) LIME [8]

(e) RRM [10] (f) JED [11] (g) BIMEF [46] (h) RetinexNet

Fig. 12. Comparisons of noisy low-light image enhancement results on a real image from DICM [43]. All methods take no post-processing techniques.



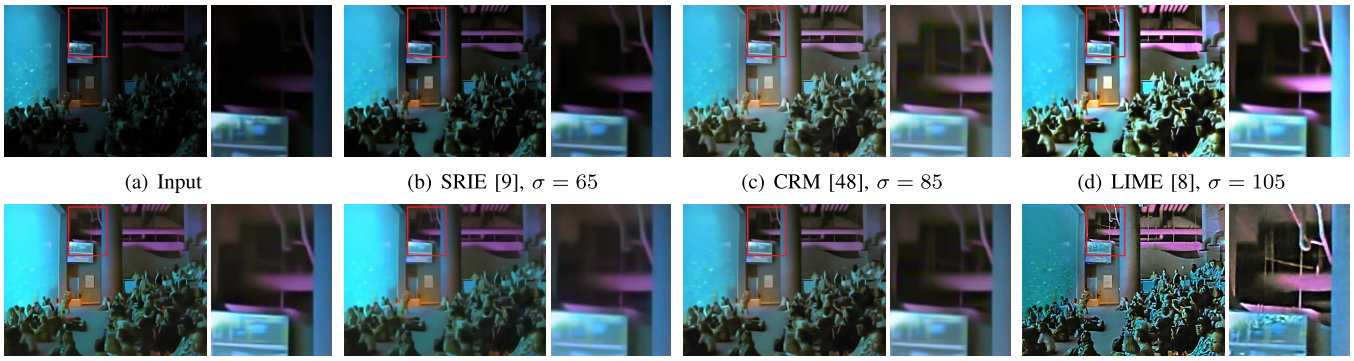
(a) Input (b) SRIE [9], $\sigma = 25$ (c) CRM [48], $\sigma = 25$ (d) LIME [8], $\sigma = 25$

(e) RRM [10] (f) JED [11] (g) BIMEF [46], $\sigma = 25$ (h) RetinexNet

Fig. 13. Comparisons of noisy low-light image enhancement results on a real image from DICM [43]. Besides RRM [10], JED [11] and RetinexNet, other methods are pre-processed by BM3D [23] with the noise level estimated by [58]. The estimated noise level σ is listed after the method name.

removed. Image structure details are preserved and contrasts of images are enhanced.

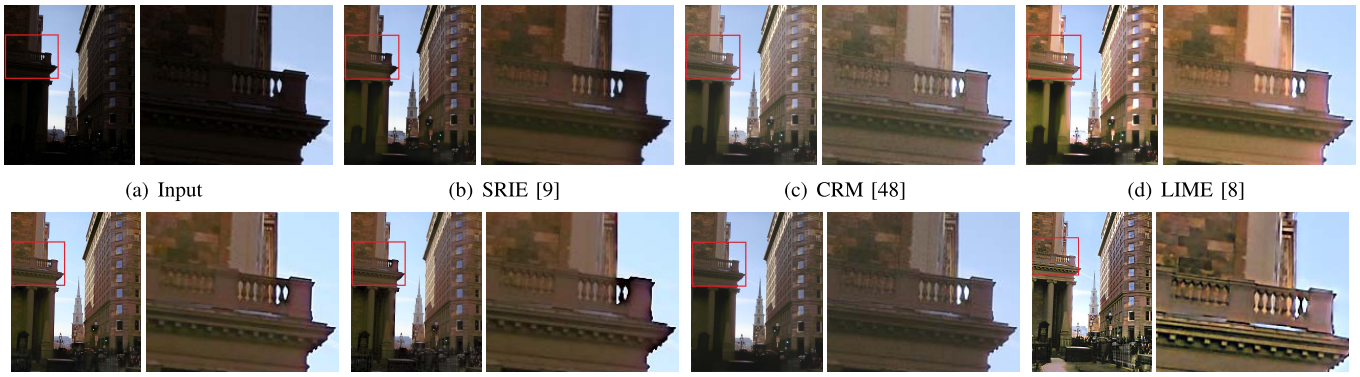
2) Low-Light Enhancement With Noise Suppression: We also provide the comparison of the proposed method with other methods pre/post-processed by BM3D [44] with the denoising parameter σ estimated by [58] and those methods jointly with JPEG artifact removal [59]. The BM3D is performed as the post-processing, following the operations in LIME [8]. For JPEG artifact removal method [59], the structure component enhancement operation is replaced by each low-light enhancement method and the final enhancement results are compared. As shown in Figs. 13-14, the noise amplified by CRM, LIME and BIMEF is handled. However, many tiny structures are removed inevitably. LIME over-enhances the input image,



(a) Input (b) SRIE [9], $\sigma = 65$ (c) CRM [48], $\sigma = 85$ (d) LIME [8], $\sigma = 105$

(e) RRM [10] (f) JED [11] (g) BIMEF [46], $\sigma = 75$ (h) RetinexNet

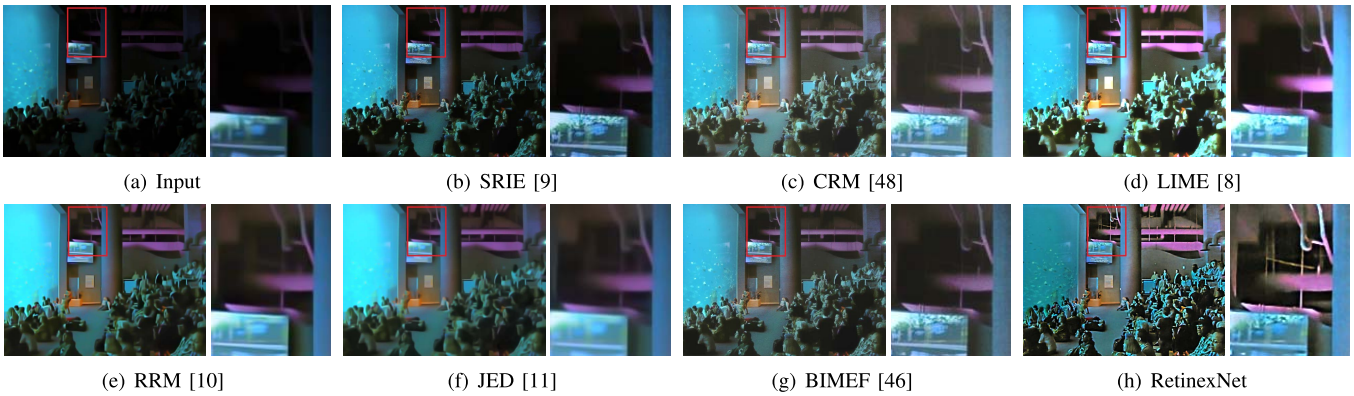
Fig. 14. Comparisons of noisy low-light image enhancement results on a real image from DICM [43]. Besides RRM [10], JED [11] and RetinexNet, other methods are post-processed by BM3D [23] with the noise level estimated by [58]. The estimated noise level σ is listed after the method name.



(a) Input (b) SRIE [9] (c) CRM [48] (d) LIME [8]

(e) RRM [10] (f) JED [11] (g) BIMEF [46] (h) RetinexNet

Fig. 15. Comparisons of noisy low-light image enhancement results on a real image from DICM [43]. Besides RRM [10], JED [11] and RetinexNet, other methods are processed by low-light enhancement methods in the framework proposed by [59].



(a) Input (b) SRIE [9] (c) CRM [48] (d) LIME [8]

(e) RRM [10] (f) JED [11] (g) BIMEF [46] (h) RetinexNet

Fig. 16. Comparisons of noisy low-light image enhancement results on a real image from DICM [43]. Besides RRM [10], JED [11] and RetinexNet, other methods are processed by low-light enhancement methods in the framework proposed by [59].

especially in regions with higher illumination. Similarly, JED and RRM remove details and weaken the contrast. When considering low-light image enhancement and JPEG artifact removal, as shown in Figs. 15-16, the results of other methods jointly with [59] are still unsatisfied. It is clearly demonstrated that, the results of SRIE, CRM, BIMEF, and LIME include blurry details and residual noise. By contrast, in both two cases, our result looks sharper and more natural. The result images are clean with fine-grained texture details.

3) Subjective Evaluation: We also evaluate the subjective visual quality of different methods using the Mean Opinion Score (MOS) for subjective evaluation. 15 images are manually selected from LOL dataset with a thoughtful consideration of content diversity, illumination variety, and randomness of testing images for evaluation. These images are enhanced by different low-light enhancement methods, including DeepUPE [15], EnlightenGAN [60], LIME [8], LLNet [14], DRD [20], RRM [10], SICE [16], and the proposed method.

TABLE V

THE RESULTS OF PAIRWISE COMPARISON IN A USER STUDY. EACH VALUE REPRESENTS THE NUMBER OF TIMES THE METHOD IN EACH ROW HAS OUTPERFORMED THE METHOD IN THE RESPECTIVE COLUMN

	LIME	DRD	LLNet	DeepUPE	EnlightenGAN	SICE	RRM	Ours
LIME	-	258	236	250	162	263	164	57
DRD	42	-	182	187	50	188	87	24
LLNet	64	118	-	172	69	163	66	10
DeepUPE	49	113	128	-	39	143	29	12
EnlightenGAN	138	250	231	261	-	256	148	49
SICE	37	112	137	157	44	-	65	4
RRM	136	213	234	271	152	235	-	31
Ours	243	276	290	288	251	296	269	-

TABLE VI

THE MOS SCORE OF DIFFERENT METHODS

Method	LIME	DRD	LLNet	DeepUPE	EnlightenGAN	SICE	RRM	Ours
MOS	2.21	0.58	0.47	0.33	1.95	0.37	1.70	4.00

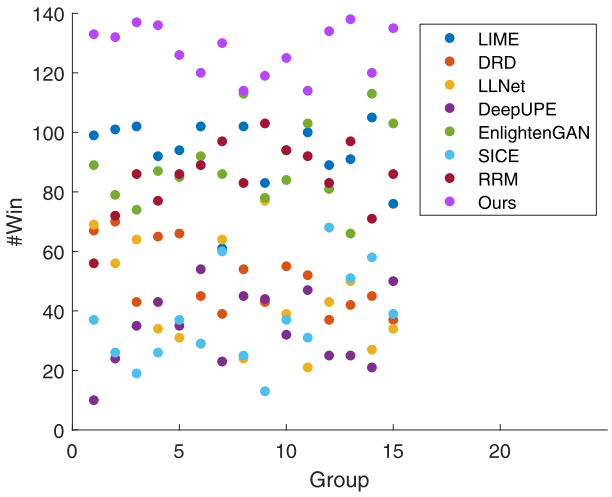


Fig. 17. Visualization of all paired comparisons. The horizontal axis denotes the comparison group ID, while the vertical axis indicates the winning time in the comparison.

Their results are evaluated by human annotators. 30 participants are invited to join the subjective experiment. There are 15 test images comparing the results of 8 methods. For each image, there are 28 paired comparisons. All testing images are divided into 3 groups. Each individual is required to provide subjective results for 2 of 3 groups, including $10 \times 28 = 280$ image pairs. The comparison results are illustrated in Fig. 17 and Table V. Based on the compared pairs, we also fit a Bradley-Terry model [61] to estimate the MOS score for each method so that they can be ranked. The inferred average MOS score is presented in Table VI. It is observed that, the proposed method achieves superior visual quality than other methods.

VI. CONCLUSION

In our work, we build a large-scale paired dataset with real photography and construct a deep network (Retinex-Net) based on Retinex model for robust low-light enhancement. The Retinex-Net enhances the illumination and reflectance maps in the Retinex domain. With the help of integration of l_0 gradient minimization, Retinex-Net learns a coupled representation in low-light and normal-light spaces for decomposition and enhancement. The evaluation on both synthetic and real images, particularly on those containing intensive noise and compression artifacts, shows the effectiveness of our novel model, which significantly outperforms the state-of-the-art methods.

REFERENCES

[1] 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. 1st Conf. Vis. Biomed. Comput.*, May 1990, pp. 337–345.

[2] M. Abdullah-Al-Wadud, M. Hasanul Kabir, M. Ali Akber 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.

[3] E. H. Land, "The retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108–128, Dec. 1977.

[4] D. J. Jobson, Z. 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.

[5] D. J. Jobson, Z. 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.

[6] 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.

[7] 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.

[8] 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.

[9] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2782–2790.

[10] 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.

[11] 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. (ISCAS)*, Florence, Italy, May 2018.

[12] Y.-F. Wang, H.-M. Liu, and Z.-W. Fu, "Low-light image enhancement via the absorption light scattering model," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5679–5690, Nov. 2019.

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

[14] 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.

[15] 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. (CVPR)*, Jun. 2019, pp. 6842–6850.

[16] 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.

[17] 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.

[18] J. Liang, Y. Xu, Y. Quan, J. Wang, H. Ling, and H. Ji, "Deep bilateral retinex for low-light image enhancement," 2020, *arXiv:2007.02018*. [Online]. Available: http://arxiv.org/abs/2007.02018

[19] 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*. [Online]. Available: http://arxiv.org/abs/1711.02488

[20] W. Chen, W. Wenjing, Y. Wenhan, and L. Jiaying, "Deep Retinex decomposition for low-light enhancement," in *Proc. Brit. Mach. Vis. Conf.*, Sep. 2018, pp. 1–12.

[21] 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. (ICIP)*, Sep. 2015, pp. 3730–3734.

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

[23] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[24] Z. Gu, F. Li, F. Fang, and G. Zhang, "A novel retinex-based fractional-order variational model for images with severely low light," *IEEE Trans. Image Process.*, vol. 29, pp. 3239–3253, 2020.

[25] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[26] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and locally consistent image completion," *ACM Trans. Graph.*, vol. 36, no. 4, p. 107, 2017.

[27] R. Köhler, C. Schuler, B. Schölkopf, and S. Harmeling, "Mask-specific inpainting with deep neural networks," in *Proc. Pattern Recognit.*, Oct. 2014, pp. 523–534.

[28] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5505–5514.

[29] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, "Learning to deblur," 2014, *arXiv:1406.7444*. [Online]. Available: http://arxiv.org/abs/1406.7444

[30] L. Xu, J. S. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2014, pp. 1790–1798.

[31] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," 2015, *arXiv:1508.06576*. [Online]. Available: http://arxiv.org/abs/1508.06576

[32] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187–5198, Nov. 2016.

[33] D. Eigen, D. Krishnan, and R. Fergus, "Restoring an image taken through a window covered with dirt or rain," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 633–640.

[34] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," 2016, *arXiv:1608.03981*. [Online]. Available: http://arxiv.org/abs/1608.03981

[35] F. Jiang, W. Tao, S. Liu, J. Ren, X. Guo, and D. Zhao, "An end-to-end compression framework based on convolutional neural networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 10, pp. 3007–3018, Oct. 2018.

[36] J. Yang, X. Jiang, C. Pan, and C.-L. Liu, "Enhancement of low light level images with coupled dictionary learning," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 751–756.

[37] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.

[38] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via L0 gradient minimization," *ACM Trans. Graph.*, vol. 30, no. 6, pp. 174:1–174:12, 2011.

[39] E. Candes, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted L1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, 2008.

[40] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[41] J. Anaya and A. Barbu, "RENOIR—A dataset for real low-light image noise reduction," *J. Vis. Commun. Image Represent.*, vol. 51, pp. 144–154, Feb. 2018.

[42] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015.

[43] 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.

[44] Q. Chen and D. Wu, "Image denoising by bounded block matching and 3D filtering," *Signal Process.*, vol. 90, no. 9, pp. 2778–2783, Sep. 2010.

[45] D.-T. Dang-Nguyen, C. Pasquini, V. Conotter, and G. Boato, "RAISE: A raw images dataset for digital image forensics," in *Proc. 6th ACM Multimedia Syst. Conf.*, Mar. 2015, pp. 219–224.

[46] Z. Ying, G. Li, and W. Gao, "A bio-inspired multi-exposure fusion framework for low-light image enhancement," 2017, *arXiv:1711.00591*. [Online]. Available: http://arxiv.org/abs/1711.00591

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

[48] Z. Ying, G. Li, Y. Ren, R. Wang, and W. Wang, "A new low-light image enhancement algorithm using camera response model," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 3015–3022.

[49] 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.

[50] X. Dong, Y. Pang, and J. Wen, "Fast efficient algorithm for enhancement of low lighting video," in *Proc. ACM SIGGRAPH Posters (SIGGRAPH)*, 2010, pp. 1–6.

[51] Z. Ying, G. Li, Y. Ren, R. Wang, and W. Wang, "A new image contrast enhancement algorithm using exposure fusion framework," in *Int. Conf. Comput. Anal. Images Patterns*. Cham, Switzerland: Springer, 2017, pp. 36–46.

[52] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems IV*. San Diego, CA, USA: Academic, 1994, pp. 474–485.

[53] K. Gwn Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," 2015, *arXiv:1511.03995*. [Online]. Available: http://arxiv.org/abs/1511.03995

[54] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Washington, DC, USA, Dec. 2015, pp. 1026–1034.

[55] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[56] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

[57] K. Gu, M. Liu, G. Zhai, X. Yang, and W. Zhang, "Quality assessment considering viewing distance and image resolution," *IEEE Trans. Broadcast.*, vol. 61, no. 3, pp. 520–531, Sep. 2015.

[58] X. Liu, M. Tanaka, and M. Okutomi, "Single-image noise level estimation for blind denoising," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5226–5237, Dec. 2013.

[59] Y. Li, F. Guo, R. T. Tan, and M. S. Brown, "A contrast enhancement framework with jpeg artifacts suppression," in *Proc. IEEE Eur. Conf. Computer Vis.*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., 2014, pp. 174–188.

[60] Y. Jiang *et al.*, "EnlightenGAN: Deep light enhancement without paired supervision," 2019, *arXiv:1906.06972*. [Online]. Available: http://arxiv.org/abs/1906.06972

[61] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: The method of paired comparisons," *Biometrika*, vol. 39, nos. 3–4, pp. 324–345, Dec. 1952.

Wenhan Yang (Member, IEEE) received the B.S. and Ph.D. degrees (Hons.) 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 more than 100 technical articles in refereed journals and proceedings, and holds 9 granted patents. His current research interests include image/video processing/restoration, bad weather restoration, and human–machine collaborative coding.

He 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 CSIG Best Doctoral Dissertation Award in 2019. He served as the Area Chair for IEEE ICME-2021 and the Organizer for IEEE CVPR-2019/2020/2021 UG2+ Challenge and Workshop.

Wenjing Wang (Graduate Student Member, IEEE) received the B.S. degree in data science from Peking University, Beijing, China, in 2019, where she is currently pursuing the master's degree with the Wangxuan Institute of Computer Technology. Her current research interests include image stylization, image synthesis, and deep learning.

Haofeng Huang is currently pursuing the B.S. degree in computer science from Peking University, Beijing, China. His current research interests include low-level computer vision and deep-learning based image processing.

Shiqi Wang (Member, IEEE) received the B.S. degree in computer science from the Harbin Institute of Technology in 2008 and the Ph.D. degree in computer application technology from Peking University in 2014. He was a Postdoctoral Fellow with the University of Waterloo, Canada, and also with the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore. He is currently an Assistant Professor with the Department of Computer Science, City University of Hong Kong. He has proposed more than 40 technical proposals to ISO/MPEG, ITU-T, and AVS standards. He has authored or coauthored more than 200 refereed journal/conference articles. He has coauthored the article that received the Best Student Paper Award in the IEEE ICIP 2018. His research interests include video compression, and quality assessment and analysis. He received the Best Paper Award from IEEE ICME 2019, VCIP 2019, IEEE Multimedia 2018, and PCM 2017.

Jiaying Liu (Senior Member, IEEE) received the Ph.D. degree (Hons.) in computer science from Peking University, Beijing, China, in 2010. From 2007 to 2008, she was a Visiting Scholar with the University of Southern California, Los Angeles. In 2015, she was also a Visiting Researcher with the Microsoft Research Asia supported by the Star Track Young Faculties Award. She is currently an Associate Professor with Peking University, where she is also a Boya Young Fellow with the Wangxuan Institute of Computer Technology. She has authored more than 100 technical articles in refereed journals and proceedings, and holds 50 granted patents. Her current research interests include multimedia signal processing, compression, and computer vision.

Dr. Liu is a Senior Member of CSIG and CCF. She has served as a member for the Multimedia Systems & Applications Technical Committee (MSA TC) and the Visual Signal Processing and Communications Technical Committee (VSPC TC) in the IEEE Circuits and Systems Society. She received the IEEE ICME-2020 Best Paper Award and IEEE MMSP-2015 Top10% Paper Award. She has also served the Technical Program Chair for the IEEE ICME-2021/ACM ICMR-2021, the Publicity Chair for the IEEE ICME-2020/ICIP-2019, and the Area Chair for CVPR-2021/ECCV-2020/ICCV-2019. She has also served as the Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUIT SYSTEM FOR VIDEO TECHNOLOGY, and *Journal of Visual Communication and Image Representation* (JVCI) (ELSEVIER). She was the APSIPA Distinguished Lecturer from 2016 to 2017.