Deep Bilateral Retinex for Low-Light Image Enhancement

Jinxiu Liang, Yong Xu, Yuhui Quan*, Jingwen Wang, Haibin Ling and Hui Ji

***Abstract*—Low-light images, *i.e.* the images captured in low-light conditions, suffer from very poor visibility caused by low contrast, color distortion and significant measurement noise. Low-light image enhancement is about improving the visibility of low-light images. As the measurement noise in low-light images is usually significant yet complex with spatially-varying characteristic, how to handle the noise effectively is an important yet challenging problem in low-light image enhancement. Based on the Retinex decomposition of natural images, this paper proposes a deep learning method for low-light image enhancement with a particular focus on handling the measurement noise. The basic idea is to train a neural network to generate a set of pixel-wise operators for simultaneously predicting the noise and the illumination layer, where the operators are defined in the bilateral space. Such an integrated approach allows us to have an accurate prediction of the reflectance layer in the presence of significant spatially-varying measurement noise. Extensive experiments on several benchmark datasets have shown that the proposed method is very competitive to the state-of-the-art methods, and has significant advantage over others when processing images captured in extremely low lighting conditions.**

***Index Terms*—Low-light image enhancement, deep bilateral learning, robust Retinex model**

I. Introduction

It often occurs in practice that one needs to capture images in low-light conditions, *e.g.* at dawn/twilight and in dimly-lit indoor rooms. Images captured in low-light conditions, *i.e. low-light images*, usually have poor visibility in terms of low contrast, color distortion and low signal-to-noise-ratio (SNR). Low-light image enhancement is then about improving the visual quality of low-light images for better visibility of image details and higher SNR. See Fig. 1 for an illustration. Such a technique not only sees its practical values in digital photography, but also benefits many computer vision applications (*e.g.* surveillance and tracking) in low-light conditions.

There has been an enduring effort on developing effective techniques for low-light image enhancement, *e.g.* histogram equalization and gamma correction. In recent years, the Retinex model of images has been one prominent choice for developing more powerful low-light image enhancement techniques; see *e.g.* [9]–[11], [13], [23], [38], [40], [43], [49]. The Retinex model of images assumes that an image $\boldsymbol{I}$ is composed of two different layers, the reflectance $\boldsymbol{R}$ and the illumination $\boldsymbol{E}$, in the following expression:

$$\boldsymbol{I} = \boldsymbol{R} \odot \boldsymbol{E} + \boldsymbol{N}, \tag{1}$$

where $\odot$ denotes element-wise multiplication, and $\boldsymbol{N}$ denotes the measurement noise. The layer $\boldsymbol{R}$ denotes the reflectance map that encodes inherent image structures, *i.e.* physical characteristics of scenes/objects. The layer $\boldsymbol{E}$ denotes the illumination map which is related to the light intensities of scenes/objects determined by the lighting condition. Once the Retinex decomposition of $\boldsymbol{I}$ is done, one can reconstruct a new image $\widetilde{\boldsymbol{I}}$ with better visibility by replacing $\boldsymbol{E}$ using another illumination layer $\widetilde{\boldsymbol{E}}$:

$$\widetilde{\boldsymbol{I}} = \boldsymbol{R} \odot \widetilde{\boldsymbol{E}}. \tag{2}$$

For instance, $\widetilde{\boldsymbol{E}}$ can be defined using the gamma correction function $\widetilde{\boldsymbol{E}} := \boldsymbol{E}^{\frac{1}{\gamma}}$.

It can be seen that the problem of low-light image enhancement can be recast as the Retinex decomposition problem (1). It is an ill-posed inverse problem, and the low SNR of the input low-light image further aggravates the ill-posedness. Therefore, there are two main challenges for solving (1):

1) How to resolve the ambiguities between the two maps,
2) How to make the estimation robust to noise.

Regarding the first question, the answer from most existing works is to impose certain prior on both the reflectance layer and the illumination layer. In the past, such priors usually are pre-defined based on empirical observations, *e.g.* spatial smoothness prior on the illumination layer [10], [13], [19], [40] and piece-wise smoothness prior on the reflectance layer [11], [27], [32]. More recently, deep learning has become one promising tool of learning the priors for Retinex decomposition. It has been used either for only estimating the illumination layer (*e.g.* [38], [41]) or for estimating both layers (*e.g.* [43], [49]).

The answer to the second question also plays an important role in low-light image enhancement, as the measurement noise will be noticeably amplified when taking a direct inversion. The SNR of a low-light image is usually much lower than its counterparts taken under normal lighting conditions. Recall that, as the light sensors of a camera usually cannot receive adequate light in low-light conditions, the shot noise caused by statistical quantum fluctuations will be much more

Jinxiu Liang, Yong Xu, and Yuhui Quan are with School of Computer Science and Engineering at South China University of Technology, Guangzhou, China. Yong Xu is also with Peng Cheng Laboratory, Shenzhen, China. Yuhui Quan is also with Guangdong Provincial Key Laboratory of Computational Intelligence and Cyberspace Information, China. (Email: cssherryliang@mail.scut.edu.cn; yxu@scut.edu.cn; csyhquan@scut.edu.cn)

Jingwen Wang is with Tencent AI Lab, Shenzhen, China. (Email: jaywong-jaywong@gmail.com)

Haibin Ling is with the Department of Computer Science, Stony Brook University, Strony Brook, NY, USA. (Email: hling@cs.stonybrook.edu)

Hui Ji is with Department of Mathematics at National University of Singapore, Singapore. (Email: matjh@nus.edu.sg)

Asterisk indicates the corresponding author.

(a) Low-light image (b) SRIE [11] (c) RRM [23] (d) KinD [49] (e) Ours

Fig. 1: Demonstration of existing low-light image enhancement methods and the proposed one. Please zoom-in for easier visual inspection.

prominent in a low-light image. Together with the necessity of an amplification of light sensitivity of sensors (*i.e.* a higher ISO) in low-light conditions, low-light images tend to have low SNRs. In other words, an effective denoising mechanism is another key component of a Retinex-model-based low-light image enhancement method with good performance.

A. Discussion on measurement noise of low-light images

As a low-light image often has rather low SNR, the treatment of measurement noise plays an important role for the Retinex decomposition. Many existing methods ignore this issue, leading to noticeable noise magnification in the reflectance layer; see Fig. 1 (b) and Fig. 7 (b) for an illustration. Some other existing solutions deal with the magnified noise in the result by running a denoising post-processing. However, the noise after magnification has much more complex characteristic and is closely related to inherent image structures. As a result, the reflectance layer after post-processing often tends to be over-smoothed with many image details lost; see Fig. 1 (d) and Fig. 7 (d) for an illustration.

There exist profound connections between the noise $\boldsymbol{N}$, the illumination layer $\boldsymbol{E}$ as well as the reflectance layer $\boldsymbol{R}$. The measurement noise $\boldsymbol{N}$ spatially varies over different regions of a low-light image. It is not *i.i.d.* and thus cannot be easily distinguished from the structures of reflectance by off-the-shelf image denoisers or image denoisers with pre-defined regularizations. See Fig. 1 (c) and Fig. 7 (c) for an illustration of the result using a pre-defined regularization model from [23]. Indeed, the noise variance is closely related to the illumination map $\boldsymbol{E}$. In bright regions, $\boldsymbol{N}$ is dominated by the image-dependent shot noise caused by the randomness of light arrival. In dark regions, $\boldsymbol{N}$ is dominated by the image-independent read noise caused by the sensitivity of sensor readout. In addition, there is also other noise from many sources, including dark current noise, thermal noise and quantization noise. Interested readers are referred to [44] for more details.

The treatment of measurement noise also plays a critical role for recovering the reflectance layer $\boldsymbol{R}$. It can be seen from (1) that once the illumination layer $\boldsymbol{E}$ is estimated, one can estimate $\boldsymbol{R}$ via a linear inversion. In a low-light image, many image structures of the reflectance layer, *e.g.* edges and textures, are of weak magnitude. Thus, it is challenging to distinguish noise from these weak structures. To effectively remove the noise during inversion, a powerful denoising scheme needs to be specifically designed for low-light images with low SNRs.

B. Main idea

Deep learning has emerged as a powerful tool in many image processing tasks. Based on the Retinex model (1), this paper aims at developing a powerful low-light image enhancement method with effective treatment on complex measurement noise. The proposed method takes a two-stage approach. Given an input low-light image $\boldsymbol{I}$, we first estimate the illumination layer $\boldsymbol{E}$ and measurement noise $\boldsymbol{N}$:

$$\boldsymbol{I} \rightarrow (\boldsymbol{N}, \boldsymbol{E}). \tag{3}$$

Once $\boldsymbol{N}$ and $\boldsymbol{E}$ are estimated, the reflectance layer $\boldsymbol{R}$ can be obtained by a linear inversion:

$$\boldsymbol{R} := (\boldsymbol{I} - \boldsymbol{N}) \oslash \boldsymbol{E}, \tag{4}$$

where $\oslash$ denotes element-wise division.

In the procedure above, an accurate estimation of noise $\boldsymbol{N}$ with spatially-varying characteristic is critical to the success of low-light image enhancement. As we discussed in Section I-A, there exists profound connection between the illumination layer $\boldsymbol{E}$ and the noise $\boldsymbol{N}$. Thus, we proposed a deep NN, called *Deep Bilateral Retinex* (DBR), which is mainly an NN-based joint estimator of the measurement noise and the illumination layer.

More specifically, in the proposed method, the interaction between the estimation of $\boldsymbol{E}$ and $\boldsymbol{N}$ is done by training a single NN which takes the low-light image as input and outputs a pair of learnable pixel-wise linear transforms for predicting the two layers. The transform for predicting the illumination layer is simply a pixel-wise affine transform. For the noise, motivated by the bilateral filtering for image denoising and its NN extensions [12], [37], the pixel-wise linear transform learned for the noise estimation is defined in the so-called *bilateral space*, *i.e.*, the spatial-range product space with an augmented dimension on pixel color.

Inside such a pair of learned pixel-wise linear transforms, the module for estimating the noise $\boldsymbol{N}$ is based on the pixel-wise *deformable convolution* which uses spatially-varying filtering kernels learned in the bilateral space. The module for estimating the illumination layer $\boldsymbol{E}$ is built on point-wise color transform matrices. Once the illumination layer $\boldsymbol{E}$ and noise $\boldsymbol{N}$ are estimated, the reflectance layer is predicted using (4). Also, a loss function that encourages the focus on image edges is adopted for further refinement on the separation between the noise and the reflectance layer.

C. Contributions

The effective treatment on the measurement noise plays an important role in Retinex-decomposition-based low-light image enhancement. The measurement noise in low-light image is not only significant in comparison to the magnitude of image structures, but also is spatially varying with complex statistical characteristics. This paper proposes a deep-learning-based method for low-light image enhancement with a particular focus on handling the measurement noise.

By exploiting the inherent connections between the spatially-varying noise and the illumination layer, we develop a framework that enables the interaction between noise estimation and illumination layer estimation in the bilateral space. The effectiveness of the proposed method is extensively evaluated on several benchmarks. The experimental results show that the proposed method is very effective at handling measurement noise. For the images captured in very low-light conditions, the proposed method outperforms existing ones by a large margin. For the images captured in better lighting conditions whose measurement noise is relatively low, the proposed method still provides comparable performance to those state-of-the-art (SOTA) methods.

II. Related Works

In the past, there have been extensive studies on low-light image enhancement. In the next, we give a brief discussion on existing low-light image enhancement methods, and focus more on Retinex-model-based methods.

A. Non-Retinex-based methods

Early works tackle the problem of low-light image enhancement by directly modifying the low-light image such that the resulting image has higher contrast. The histogram equalization [1], [4], [22] improves the visibility of a low-light image by balancing its histogram. The Gamma correction (power-law transformation) [15], [48] modifies the brightness of an image by increasing the brightness of dark regions and decreasing the brightness of bright regions. Multi-exposure sequence fusion is also exploited for the contrast enhancement in low-light images [3], [47]. Chen *et al.* [5] tackles the problem by directly modifying the raw data from image sensors using a learnable NN.

Since a direct contrast enhancement will magnify the measurement noise, much effort has been devoted to the noise reduction in contrast enhancement. Loza *et al.* [25] performed wavelet-based noise reduction during contrast enhancement. Based on deep auto-encoder, Lore *et al.* [24] proposed a Low-Light Net (LLNet) to sequentially learn contrast enhancement and noise reduction.

B. Retinex-based non-learning methods

The Retinex image model (1) proposed in [21] has been widely used for image enhancement; see *e.g.* [13], [16], [17], [40], [50]. The majority of existing Retinex-based approaches assume the image being processed contains only negligible noise. The key of these methods is about how to resolve the ambiguities between the illumination and reflectance layers. Most existing non-learning methods resolve such ambiguities by imposing certain prior either on the illumination layer or the reflectance layer, or both.

Several methods proposed different priors on the illumination layer. The smoothness prior is first introduced to variational models by Kimmel *et al.* [19] which minimizes the squared ℓ_2 norm of gradients of illumination layer. Wang *et al.* [40] proposed a bright-pass filter for better preserving the naturalness of the illumination layer. Such an idea is further refined by Fu *et al.* [10] via fusing multiple derivatives of the illumination layer for better performance. Guo *et al.* [13] proposed a structure-aware prior for the illumination layer which is motivated from relative total variation (RTV) [45]. There is also some work imposing the prior only on the reflectance layer. For instance, Ma *et al.* [27] imposed a piece-wise smoothness prior on the reflectance layer.

Another class of methods resolves the solution ambiguity by imposing the priors on both two layers. In Ng *et al.* [32], the TV prior is imposed on both reflectance and illumination layers after applying the logarithmic transformation on the input image. Instead of using logarithmic transform as a pre-processing, Fu *et al.* [9] introduced a probabilistic method for simultaneous illumination and reflectance estimation (SIRE) in the linear space rather than the logarithmic space. Another variation comes form [11] which proposes a weighted variational model to enhance the variation of derivative magnitudes in bright regions.

In addition to resolving the solution ambiguity, some methods are proposed to process low-light images with significant noise. Elad *et al.* [8] proposed to constrain the bilateral smoothness on pixel values of both illumination layer and reflectance layer using two tailored bilateral filters. A robust fidelity term with an explicit noise term is used in Ren *et al.* [36] and Li *et al.* [23] to handle measurement noise. Nevertheless, the complex and spatially-varying characteristic of measurement noise makes these approaches not very effective.

C. Retinex-based learning methods without noise handling

In recent years, deep learning has emerged as one prominent tool in image enhancement, including Retinex-based low-light image enhancement. Wang *et al.* [41] proposed to estimate and adjust the illumination layer of a low-light image by an NN. Gharbi *et al.* [12] proposed a bilateral learning framework for photography enhancement, which trains an NN to predict point-wise color transform coefficients for the color vector at each pixel. The similar idea is used in [38] that learns the image-to-illumination mapping for under-exposure correction. These methods do not take the measurement noise into consideration. In the case of low SNR, the point-wise transform used in these methods is sensitive to noise, especially in the dark regions of low-light images. As a result, the visual quality of the results from these methods is not very satisfactory, especially for low-light images with low SNRs.

D. Retinex-based methods with noise handling

The measurement noise of low-light images is often quite significant. Without appropriate noise treatment, those deep learning methods listed in Section II-C are likely to have erroneous estimations of both layers in dark regions. Recently, several deep learning methods have been proposed with the focus on better robustness to noise. Wei *et al.* [43] proposed to decompose a low/normal-light image into the corresponding reflectance and illumination layers by an NN and then adjust the illumination by another NN. An off-the-shelf denoiser was then used as a post-processing to remove the artifacts of the reflectance layer caused by noise. Zhang *et al.* [49] trained a denoising NN for removing the artifacts of the reflectance layer, which leads to better visual quality of the result. However, as the artifacts caused by noise have complex characteristic and are highly correlated to the reflectance layer, it is difficult to accurately separate artifacts and the truth reflectance layer. Often some details of the reflectance layer are wrongly removed as artifacts in these methods.

III. Deep Bilateral Retinex

In this section, we aim at developing a deep learning method for Retinex-based low-light image enhancement with a built-in powerful denoising module. Recall that the Retinex model of a low-light image $\boldsymbol{I} \in \mathbb{R}^{P_x \times P_y \times 3}$ is expressed as

$$\boldsymbol{I} = \boldsymbol{R} \odot \boldsymbol{E} + \boldsymbol{N},$$

where $\boldsymbol{R}$ denotes the reflectance layer, $\boldsymbol{E}$ denotes the illumination layer and $\boldsymbol{N}$ denotes the measurement noise. Once $\boldsymbol{N}$ and $\boldsymbol{E}$ are estimated, the reflectance layer $\boldsymbol{R}$ is obtained by

$$\widetilde{\boldsymbol{R}} = (\boldsymbol{I} - \boldsymbol{N}) \oslash \boldsymbol{E}. \tag{5}$$

Following [38], for a low-light image, we assume the illumination layer of its counterpart taken in the normal lighting condition has the following illumination layer:

$$\widetilde{\boldsymbol{E}} = \boldsymbol{E}^{\frac{1}{\infty}} = \boldsymbol{1}.$$

In other words, the estimated reflectance layer $\widetilde{\boldsymbol{R}}$ of the input low-light image is considered as the output of the proposed Retinex-based low-light image enhancement.

The focus of the proposed low-light image enhancement is then on how to estimate the noise layer $\boldsymbol{N}$ and the illumination layer $\boldsymbol{E}$. As we discussed in the previous section, the noise characteristic of $\boldsymbol{N}$ is spatially varying and inherently related to the illumination layer $\boldsymbol{E}$. In the next, we introduce an NN architecture that enables a joint prediction of the two layers with a built-in interaction mechanism.

A. Outline of the NN for joint estimation of $\boldsymbol{N}$ and $\boldsymbol{E}$ in bilateral space

Recall that we need to have a joint estimation of $\boldsymbol{N}$ and $\boldsymbol{E}$ for exploiting their inherent correlation. Thus, instead of proposing an end-to-end network that directly maps the input image to these two layers, we propose to train an NN that learns a pair of linear transforms, which maps a low-light image $\boldsymbol{I}$ to the noise layer $\boldsymbol{N}$ and the illumination layer $\boldsymbol{E}$, as shown in Fig. 2. More specifically, the proposed NN, denoted by $\mathcal{F}_{\boldsymbol{\Theta}}$, maps an input image $\boldsymbol{I}$ to a pair of transforms:

$$\mathcal{F}_{\boldsymbol{\Theta}}: \quad \boldsymbol{I} \longrightarrow \{\underbrace{\mathcal{T}_{\boldsymbol{A}}(\boldsymbol{I})}_{\boldsymbol{E}}, \underbrace{\mathcal{T}_{\boldsymbol{W},\Delta}(\boldsymbol{I})}_{\boldsymbol{N}}\}. \tag{6}$$

Then, the noise $\boldsymbol{N}$ and illumination layer $\boldsymbol{E}$ are estimated by

$$\begin{aligned} \mathcal{T}_{\boldsymbol{A}}: &\quad \boldsymbol{I} \longrightarrow \boldsymbol{E} \\ \mathcal{T}_{\boldsymbol{W},\Delta}: &\quad \boldsymbol{I} \longrightarrow \boldsymbol{N}. \end{aligned} \tag{7}$$

It is shown in [6] that many photographic transformations can be locally well-approximated by affine color transforms. Therefore, for the illumination layer $\boldsymbol{E}$, the transform $\mathcal{T}$ is defined by a set of affine transforms $\mathcal{A} = \{\boldsymbol{A}_{\boldsymbol{p}}\}_{\boldsymbol{p}} \subseteq \mathbb{R}^{3,4}$. In other words, the operator $\mathcal{T}_{\boldsymbol{A}}$ in (7) is defined as

$$\mathcal{T}_{\boldsymbol{A}}: \boldsymbol{I}_{\boldsymbol{p}} \longrightarrow \boldsymbol{E}_{\boldsymbol{p}} := \boldsymbol{A}_p[\boldsymbol{I}_{\boldsymbol{p}}, 1]^\top, \quad \text{for each pixel } \boldsymbol{p}, \tag{8}$$

where the set $\mathcal{A} = \{\boldsymbol{A}_{\boldsymbol{p}}\}_{\boldsymbol{p}}$ contain the coefficients predicted by the NN.

For the estimation of noise layer, we also need to learn a pixel-wise transform, as the noise $\boldsymbol{N}$ has the spatially varying characteristic. Furthermore, the low SNR of low-light image makes it challenging to distinguish noise from the image edges with weak magnitude. In other words, we need to learn a pixel-wised transform with edge awareness. Motivated by the computational efficiency and edge adaptivity of bilateral filtering, we propose to learn such a transform in the bilateral space, *i.e.* the space which treats each image pixel $\boldsymbol{p}$ as a point $(\boldsymbol{p}, \boldsymbol{I}_{\boldsymbol{p}})$ in $\mathbb{R}^5$. The resulting transform can be expressed as a spatially-varying convolution:

$$\mathcal{T}_{\boldsymbol{W},\Delta}: \boldsymbol{I}_{\boldsymbol{p}} \longrightarrow \boldsymbol{N}_{\boldsymbol{p}} := \sum_{\boldsymbol{q} \in \mathcal{N}_{\boldsymbol{p}}} \boldsymbol{W}_{\boldsymbol{p},\boldsymbol{q}+\Delta_{\boldsymbol{p},\boldsymbol{q}}} \boldsymbol{I}^{\top}_{\boldsymbol{q}+\Delta_{\boldsymbol{p},\boldsymbol{q}}}, \tag{9}$$

where $\mathcal{N}_{\boldsymbol{p}}$ denotes a $K \times K$ regular neighborhood centered at pixel $\boldsymbol{p}$ in image $\boldsymbol{I}$. Each entry of *kernel*, $\boldsymbol{W}_{\boldsymbol{p},\boldsymbol{q}+\Delta_{\boldsymbol{p},\boldsymbol{q}}} \in \mathbb{R}^{3\times 3}$, is defined on the regular grid in the bilateral space, where $\Delta_{\boldsymbol{p},\boldsymbol{q}} \in \mathbb{R}^2$ denotes the associated *offset*.

It can be seen that the family of coefficients, $\boldsymbol{\Gamma}$, for defining the transform of estimating the noise layer is composed of

$$\boldsymbol{\Gamma} := \quad \{\boldsymbol{W}_{\boldsymbol{p},\boldsymbol{q}+\Delta_{\boldsymbol{p},\boldsymbol{q}}}, \quad \Delta_{\boldsymbol{p},\boldsymbol{q}}\}. \tag{10}$$

Similarly, we propose to train an NN that takes the image $\boldsymbol{I}$ as the input and outputs the prediction of the coefficients

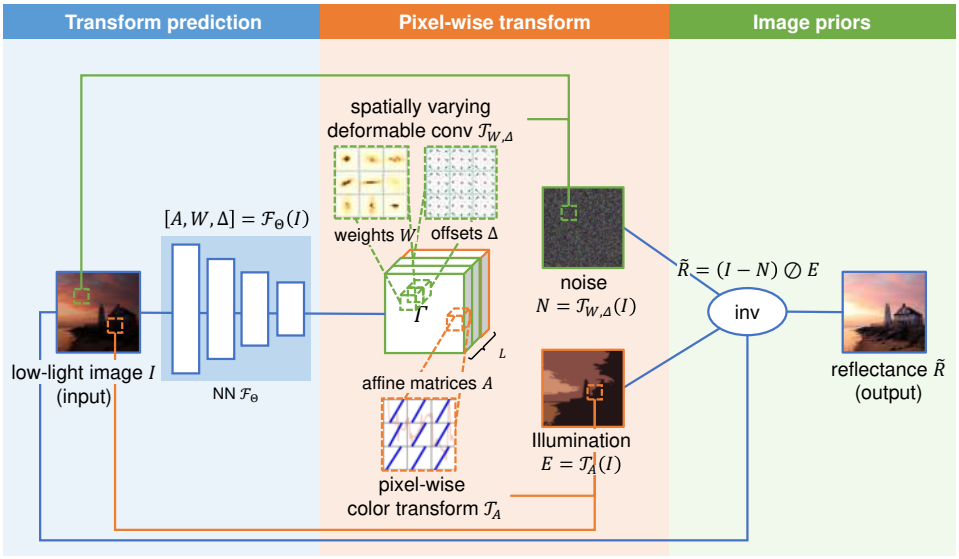


Fig. 2: Framework of the proposed method. The NN $\mathcal{F}_\Theta$ learns pixel-wise linear transforms $\mathbf{\Gamma}$ including color transform $\mathcal{T}_{\boldsymbol{A}}$ and deformable convolution $\mathcal{T}_{\boldsymbol{W},\Delta}$ to transform input $\boldsymbol{I}$ pixel-wise into $\boldsymbol{E}, \boldsymbol{N}$ firstly. Then the output $\widetilde{\boldsymbol{R}}$ is obtained by performing inversion using (5) with estimated $\boldsymbol{E}, \boldsymbol{N}$.

above to obtain the transform (9), which will then be applied to predicting the noise $\boldsymbol{N}$.

B. Detailed discussion on the transform

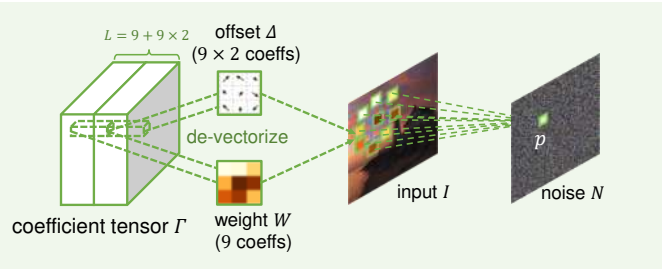


Fig. 3: Illustration of spatially-varying deformable convolution $\mathcal{T}_{\boldsymbol{W},\Delta}$ to estimate the noise $\boldsymbol{N}$ from a low-light image $\boldsymbol{I}$.

We give a more detailed discussion on the spatially-varying convolution defined in (9):

$$\mathcal{T}_{\boldsymbol{W},\Delta} : \boldsymbol{I}_{\boldsymbol{p}} \longrightarrow \boldsymbol{N}_{\boldsymbol{p}} := \sum_{\boldsymbol{q} \in \mathcal{N}_{\boldsymbol{p}}} \boldsymbol{W}_{\boldsymbol{p},\boldsymbol{q}+\Delta_{\boldsymbol{p},\boldsymbol{q}}} \boldsymbol{I}^{\top}_{\boldsymbol{q}+\Delta_{\boldsymbol{p},\boldsymbol{q}}},$$

which is used for predicting the noise $\boldsymbol{N}$. See Fig. 3 for the illustration of the transform. The offsets $\{\Delta_{\boldsymbol{p},\boldsymbol{q}}\}_{\boldsymbol{p},\boldsymbol{q}}$ used in the proposed method are chosen from a larger neighborhood $W \times W$ where scalar $W (W \geq K)$ denotes the window size, which are firstly scaled to $[0, 1]$ by a sigmoid function and then linearly scaled to $[-W, W]$. Notice that these offsets do not form a regular grid, and we use the bi-linear interpolation to generate $\boldsymbol{I}_{\boldsymbol{q}+\Delta_{\boldsymbol{p},\boldsymbol{q}}}$.

For spatially-varying kernels $\{\boldsymbol{W}_{\boldsymbol{p},\boldsymbol{q}+\Delta_{\boldsymbol{p},\boldsymbol{q}}}\}_{\boldsymbol{p},\boldsymbol{q}}$, we need to use them to estimate the noise of a low-light image. As the energy of noise is typically concentrated on high-frequency channels, we impose that $\{\boldsymbol{W}_{\boldsymbol{p},\boldsymbol{q}+\Delta_{\boldsymbol{p},\boldsymbol{q}}}\}_{\boldsymbol{p},\boldsymbol{q}}$ should be high-pass filters, which is done by normalizing the kernels to be zero mean[1].

C. Transform prediction in bilateral space

The prediction of per-pixel kernels is not a new idea. It has been exploited in denoising [2], [28], [46], video interpolation [33], [34] and joint image filtering [18]. All of them are learned in the image space, which is not suitable for separating noise and weak image gradients of a low-light image. In this section, we give a detailed discussion on the NN for predicting the transform coefficients (10) of the spatial varying convolution in the bilateral space. See Fig. 4 for the outline of the NN, where there are three main modules: guidance module $\mathcal{G}$, prediction module $\mathcal{P}$, and slicing module.

The guidance module $\mathcal{G}$ produces a single-channel image $\boldsymbol{J} \in \mathbb{R}^{P_x \times P_y}$ whose edges are likely to be kept in the resulting image,

$$\mathcal{G} : \boldsymbol{I} \longrightarrow \boldsymbol{J}, \tag{11}$$

where $\boldsymbol{J}_{\boldsymbol{p}} \in \{1, \cdots, 2^r\}$ (typically $r = 8$) denotes the gray scale value at pixel $\boldsymbol{p} = (\boldsymbol{p}_x, \boldsymbol{p}_y)$.

Recall that the bilateral space refers to the spatial-range product space with an augmented dimension on pixel color compared to the pixel space. In our method, the transform coefficients are predicted and thus efficiently embedded in the reduced bilateral space such that the produced per-pixel spatial varying convolutions have the edge-aware properties that is critical for our task. Specifically, the prediction module $\mathcal{P}$ predicts a low-resolution bilateral grid of transform coefficients $\mathbf{\Lambda}$:

$$\mathcal{P} : \boldsymbol{I} \longrightarrow \mathbf{\Lambda}, \tag{12}$$

[1]When these kernels are used to estimate the noise-free image rather than noise in ablation study in Sec. IV, we add a softmax layer to ensure that the kernels with positive values and sums to 1.

where $\mathbf{\Lambda}$ indexed by $(i,j,k) \in \mathbb{R}^{[P_x/s_s]\times[P_y/s_s]\times[2^r/s_r]}$ is downsampled by $\mathcal{P}$ with sampling rates s_s, s_r for spatial and range domain respectively. The module $\mathcal{P}$ has a two-stream structure. The one with fully-connected layers encodes non-local information, and the other with only convolutional layers captures local information. The features from the local and non-local streams are fused eventually to generate $\mathbf{\Lambda}$.

The parameter $\mathbf{\Lambda}$ is then rolled into a bilateral grid and sliced into the coefficient tensor $\mathbf{\Gamma}$ of full resolution $P_x \times P_y$:

$$\mathbf{\Gamma}_{\boldsymbol{p}} := \sum_{i,j,k} \delta(s_s \boldsymbol{p}_x - i)\delta(s_s \boldsymbol{p}_y - j)\delta(s_r \boldsymbol{J}_{\boldsymbol{p}} - k)\mathbf{\Lambda}_{i,j,k}, \quad (13)$$

which is a 3D tensor of the same spatial resolution as $\boldsymbol{I}$ with L channels in the third dimension, where $\delta(\cdot) = \max(1 - |\cdot|, 0)$ is a linear interpolation kernel. Thanks to the slicing operation, the resulting coefficient $\mathbf{\Gamma}_{\boldsymbol{p}}$ to define the transforms that will then map the input to output is smooth in the bilateral space and keep the discontinuities of $\boldsymbol{J}$. Such a design regularizes the output towards edge-aware solutions even though edge preservation is not explicitly handled.

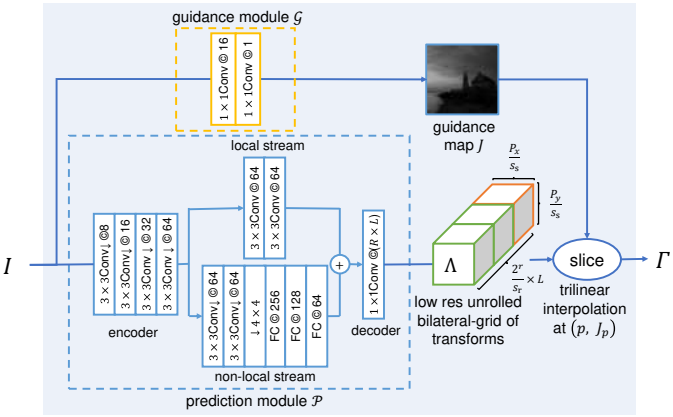


Fig. 4: Architecture of the NN $\mathcal{F}_{\mathbf{\Theta}}$ to predict transform coefficients $\mathbf{\Gamma}$ from a low-light image $\boldsymbol{I}$ in bilateral space.

D. Cost function with regularizations

The NN $\mathcal{F}_{\mathbf{\Theta}}$ is trained for predicting pixel-wise transforms, which will be used to estimate the illumination layer $\boldsymbol{E}$ and the noise $\boldsymbol{N}$ from the input $\boldsymbol{I}$. Then, the reflectance layer $\widetilde{\boldsymbol{R}}$ will be estimated by (5). See Fig. 2 for the pipeline of the method.

Consider a dataset of N image pairs $\{(\boldsymbol{I}_i, \boldsymbol{R}_i)\}_{i=1}^N$, where $\boldsymbol{R}_i$ denotes the ground truth of the reflectance layer of an input image $\boldsymbol{I}_i$. Several regularizations are imposed on the loss function in order to separate the two layers $\boldsymbol{E}$ and $\boldsymbol{R}$, in the presence of significant noise. The loss function $\mathcal{L}$ is defined as the summation of three terms:

$$\mathcal{L} := \mathcal{L}_r(\boldsymbol{R}, \widetilde{\boldsymbol{R}}) + \lambda_n \mathcal{L}_n(\boldsymbol{N}) + \lambda_e \mathcal{L}_e(\boldsymbol{E}, \boldsymbol{I}), \quad (14)$$

where λ_n and λ_e are two regularization parameters. The term $\mathcal{L}_r(\boldsymbol{R}, \widetilde{\boldsymbol{R}})$ measures the fidelity on the estimated reflectance layer, the term $\mathcal{L}_n(\boldsymbol{N})$ denotes the regularization on the estimate of noise, and the term $\mathcal{L}_e(\boldsymbol{E}, \boldsymbol{I})$ denotes the regularization on the estimate of illumination layer.

The fidelity term on the reflectance layer is defined by

$$\mathcal{L}_r = \sum_{\boldsymbol{p}} (\|\widetilde{\boldsymbol{R}}_{\boldsymbol{p}} - \boldsymbol{R}_{\boldsymbol{p}}\|_1 + \lambda_g \|\nabla\widetilde{\boldsymbol{R}}_{\boldsymbol{p}} - \nabla\boldsymbol{R}_{\boldsymbol{p}}\|_1), \quad (15)$$

where ∇ denotes the first order difference operator, and λ_g is a weighting parameter. The fidelity is measured in both intensity and gradient domains using the ℓ_1-norm metric. Such a loss function is helpful to enhance the robustness to noise and keep sharp edges in the estimate of the reflectance layer.

Motivated by relative total variation for separating cartoon structure and textures in [45], we propose the following regularization on the estimate of noise:

$$\mathcal{L}_n = \sum_{\boldsymbol{p}} \| \sum_{\boldsymbol{q}\in\mathcal{N}_{\boldsymbol{p}}} G_\sigma(\|\boldsymbol{p} - \boldsymbol{q}\|_2) \nabla \boldsymbol{N}_{\boldsymbol{q}} \|_1 \quad (16)$$

where G_σ denoted the 2D Gaussian kernel $G_\sigma(x) = \frac{1}{2\pi\sigma^2}\exp(-\frac{x^2}{2\sigma^2})$ ($\sigma = 1$ is used in the implementation). It can be seen that such a regularization alleviates possible attenuation of image edges so as to keep sharp edges in the reflectance layer.

The third term L_e is about the regularization on the illumination layer $\boldsymbol{E}$. In this paper, we consider a piecewise smoothness prior for the illumination layer, which is formulated as a re-weighted ℓ_1-norm on the gradients of $\boldsymbol{E}$:

$$\mathcal{L}_e = \sum_{\boldsymbol{p}} \frac{\|\nabla \boldsymbol{E}_{\boldsymbol{p}}\|_1}{\|\nabla \boldsymbol{I}_{\boldsymbol{p}}\|_1^\theta + \epsilon}, \quad (17)$$

where the weights are inversely proportional to the magnitude of image gradients. In other words, the larger the magnitude of low-light image gradient is, the more likely it indicates the discontinuity of the illumination layer. The exponential parameter $\theta = 1.2$ is to control the likeliness and ϵ is a small constant for avoiding the division by zero. In addition, we impose physical constraints on $\boldsymbol{E}$: $\boldsymbol{I} \le \boldsymbol{E} \le \mathbf{1}$. In training, all input-target images are normalized from original n-bit RGB color channels to $[0, 1]$. We set $\boldsymbol{I}$ as the lower bound of $\boldsymbol{E}$ to ensure that the obtained $\widetilde{\boldsymbol{R}}$ is bounded by $\mathbf{1}$, whereas setting $\mathbf{1}$ as the upper bound of $\boldsymbol{E}$ to avoid mistakenly darkening the low-light images.

IV. Experiments

A. Datasets

The proposed method is trained on the LOL dataset [43], which includes 1500 low/normal-light image pairs. Concretely, there are 500 image pairs of size 400×600 captured in *real* scenes and 1000 image pairs of size 384×384 synthesized from raw data. We use 1000 synthesis pairs and 485 real pairs for training and the remaining 15 real pairs for test as suggested in [43]. Since the images in the test set of LOL are taken in extreme low-light conditions (as shown in the topleft of Fig. 5), the dark regions of the images are full of intensive noise. The results on this dataset reveal the performance in challenging low-light conditions.

In addition to the LOL dataset, we also evaluate the proposed method on other four widely-adopted benchmarks for low-light image enhancement that contain underexposed

or low-light images without corresponding normal-light reference: (i) DICM contains 69 captured images from commercial digital cameras collected by [22]. (ii) MEF contains 17 high-quality image sequences including natural scenarios, indoor and outdoor views, and man-made architectures provided by [26]. Each image sequence has several multi-exposure images, and we select one of poor-exposed images as input to perform evaluation. (iii) LIME contains 10 low-light images used in [13]. (iv) NPE contains 8 outdoor natural scene images which are used in [40].

B. Metric for evaluation

For all datasets, four quality metrics are adopted for evaluation: (i) Lightness Order Error (LOE) [40] is designed for objectively measuring the lightness distortion. The computation requires only the low-light images as references. However, as pointed out in [13], using the low-light input as reference might be problematic. Therefore, for dataset which provides normal-light reference, we additionally measure $\mathrm{LOE}_{\mathrm{ref}}$ which uses the normal-light image as the reference. (ii) Blind Image Spatial Quality Evaluator (BRISQUE) [29] correlates subjective quality scores and can measure the quality of images with common distortion such as compression artifacts, blurring, and noise. (iii) Natural Image Quality Evaluator (NIQE) [30] does not relate to subjective quality scores and can measure the quality of images with arbitrary distortion. (iv) Perception based Image Quality Evaluator (PIQE) [31] measures the block-wise quality of images with arbitrary distortion. Lower values of the four metrics reflect better perceptual quality.

For the LOL dataset which provides reference normal-light images, two extra full-reference metrics are used, *i.e.* Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) Index [42], with higher value for better quality.

C. Implementation details

Our approach is implemented using PyTorch [35] and trained on an Nvidia Titan RTX GPU and Intel i7-7700K 4.20GHz CPU. We use Adam optimizer [20] with a fixed learning rate of 10^{-4} and ℓ_2 weight decay of 10^{-8}. Other hyper parameters are set as default (*i.e.*, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$). Totally 2000 epochs are used for training. During training, each image is normalized to the range $[0, 1]$. The batch size is set to 16. For data augmentation, we randomly cropped 256×256 patches followed by random mirroring, resizing and rotation for all patches. The weights for the convolutional and fully-connected layers are initialized according to [14] and the biases are initialized to 0. In our experiments, the parameter setting for deformable convolutions are: kernel size $K = 3$ and window size $W = 15$. The sampling rates in the prediction module are $s_{\mathrm{s}} = 16$ and $s_{\mathrm{r}} = 32$ for spatial and range domain respectively. As for the loss function, we set $\lambda_g = 0.1$, $\lambda_n = 1$, $\lambda_e = 1$, $\epsilon = 10^{-4}$.

D. Results and comparisons

We compare the proposed method with 3 existing state-of-the-art methods that explicitly consider how to deal with severe noise in low-light images, including Joint Enhancement and Denoising (JED) [36], Robust Retinex Model (RRM) [23], and Kindling the Darkness (KinD) [49]. JED and RRM are variational methods, while KinD is a deep learning-based method which uses an NN separately trained for denoising.

In addition, 12 classic low-light image enhancement methods, though without considering the existence of noise, is also evaluated for comparison. They are histogram equalization (HE), Multi-Scale Retinex with Color Restoration (MSRCR) [16], dehazing based method (Dong) [7], naturalness preserved enhancement algorithm (NPE) [40], SRIE [11], Multi-deviation Fusion method (MF) [10], low-light image enhancement via illumination map estimation (LIME) [13], Bio-Inspired Multi-Exposure Fusion (BIMEF) [47], Naturalness Preserved Image Enhancement Using a Priori Multi-Layer Lightness Statistics (NPIE-MLLS) [39], RetinexNet [43], and Deep Underexposed Photo Enhancement (DeepUPE) [38]. Among them, RetinexNet and DeepUPE are methods based on deep learning. HE is performed by using the MATLAB built-in function *histeq*. The results of other methods are generated by the codes released by the authors, with recommended experiment settings.

1) Results on LOL: The quantitative results on the LOL dataset are listed in Table I. It can be seen that the proposed method significantly outperforms all other compared methods for all seven metrics except for SSIM, of which our score is only slightly lower than KinD-nonblind. It is noted that, KinD are fed with an illumination ratio computed from the ground truth data, which is denoted as "nonblind". In contrast, our method does not require such nonblind information. The results of KinD without a ground truth illumination ratio is also reported for fair comparison. The proposed method outperform this blind version of KinD in terms of all metrics in this setting. All above noticeable performance improvement has demonstrated the effectiveness of the proposed approach.

Please see some visual comparisons in Fig. 5. It can be seen from the first three rows that the methods without noise treatment mechanisms produce noisy results, especially in the dark areas of the original image, although some of them such as LIME and DeepUPE do produce relative vivid colors. For instance, as shown in Fig. 5 and 6, the hands of the white doll are full of noise and artifacts in the enhanced results on first three rows. In contrast, the methods with noise treatment mechanisms, *i.e.*, JED, RRM, KinD and the proposed method, suppress noise well. Among them, the former three over-smooth the image details and textures, while the proposed method not only produces pleasing colors, but also trades off well between noise suppression and preservation of details.

Please see Fig. 7 for the illumination and reflectance layers estimated by several Retinex decomposition methods. Thanks to the edge-aware technique of bilateral learning, the proposed method produced edges of larger multitudes.

2) Results on the other datasets: Table II, III, IV, V summarize the results on the MEF, DICM, LIME, NPE datasets respectively. On these datasets, the methods without denoising mechanisms achieved the best quantitative results. On MEF, BIMEF outperforms other methods in terms of all metrics except for NIQE. As for DICM, LIME, and NPE, there are

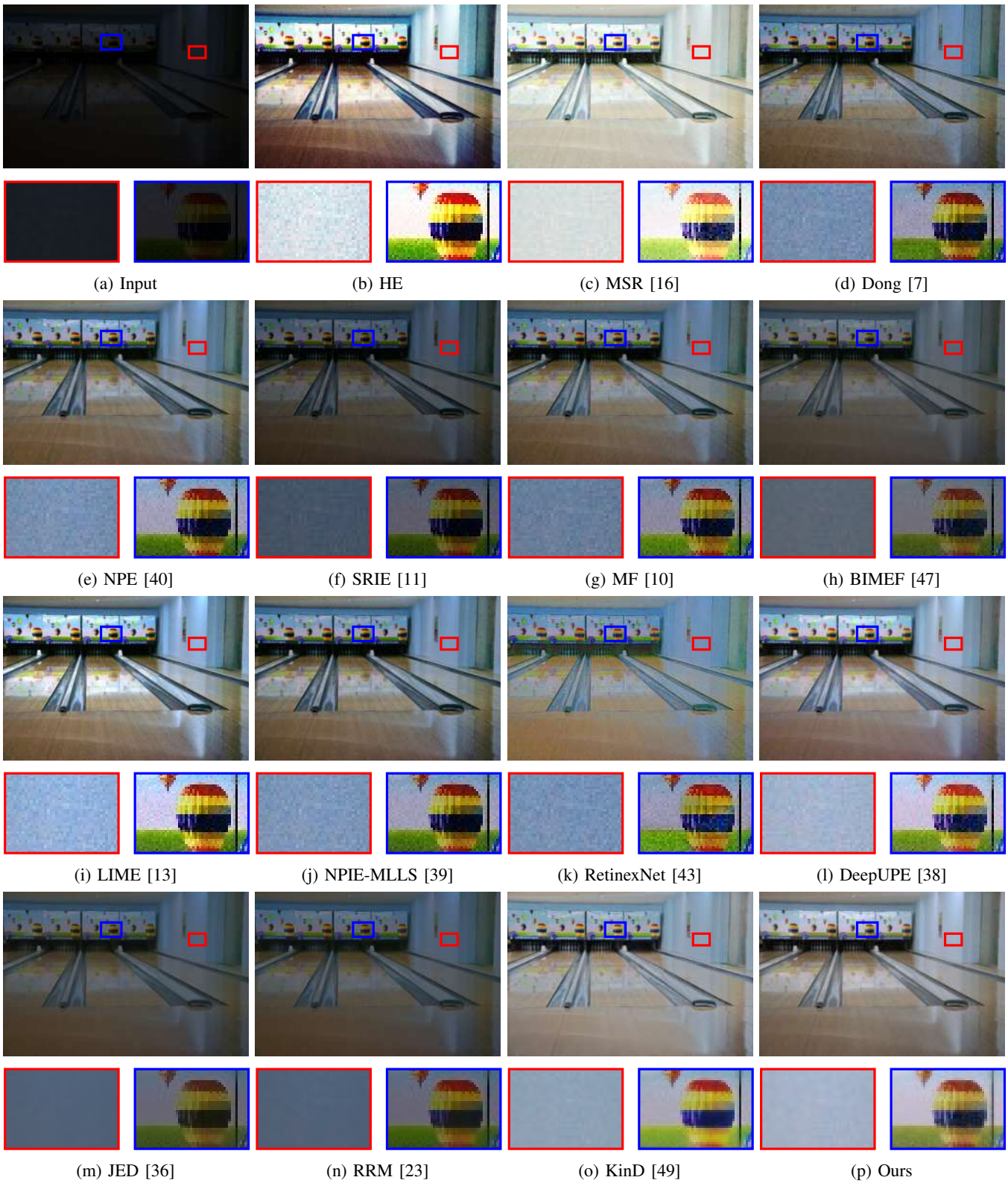
(a) Input (b) HE (c) MSR [16] (d) Dong [7]

(e) NPE [40] (f) SRIE [11] (g) MF [10] (h) BIMEF [47]

(i) LIME [13] (j) NPIE-MLLS [39] (k) RetinexNet [43] (l) DeepUPE [38]

(m) JED [36] (n) RRM [23] (o) KinD [49] (p) Ours

Fig. 5: Visual comparisons of different methods on low-light images from the LOL test set.



(a) Input (b) HE (c) MSR [16] (d) Dong [7]

(e) NPE [40] (f) SRIE [11] (g) MF [10] (h) BIMEF [47]

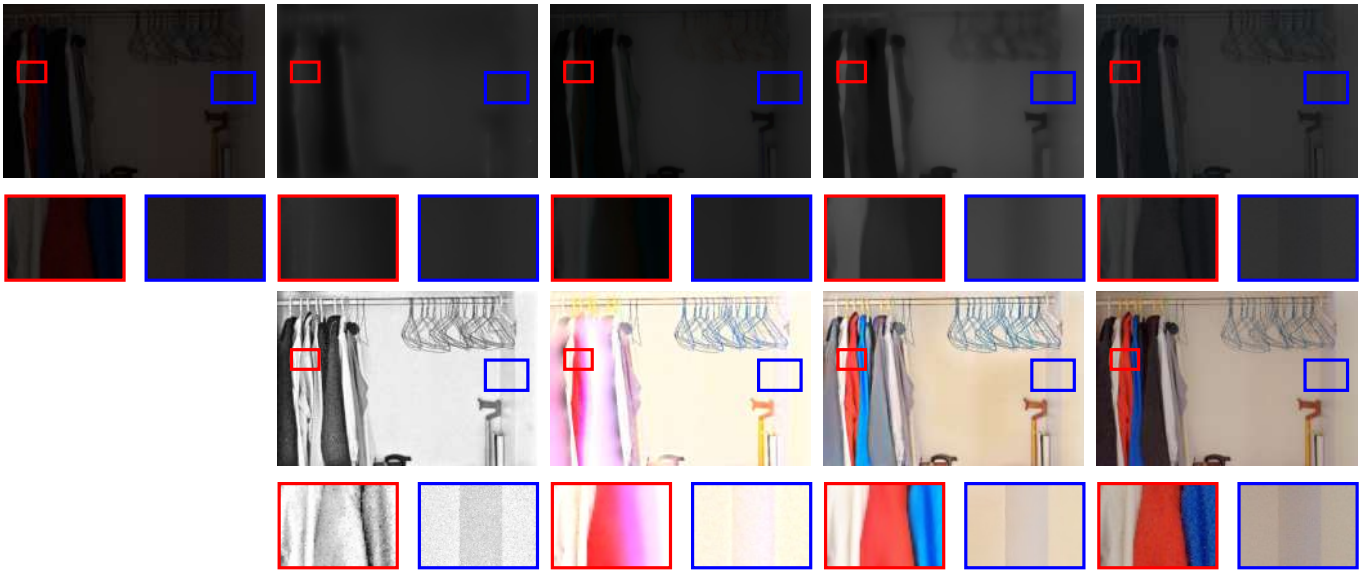
(i) LIME [13] (j) NPIE-MLLS [39] (k) RetinexNet [43] (l) DeepUPE [38]

(m) JED [36] (n) RRM [23] (o) KinD [49] (p) Ours

Fig. 6: Visual comparisons of different methods on low-light images from the LOL test set.

TABLE I: Quantitative results on the LOL [43] test set.

Method	PSNR(dB)↑	SSIM↑	LOE_{ref} ↓	LOE↓	NIQE↓	BRISQUE↓	PIQE↓
HE	14.8006	0.4112	485.67	250.61	8.4763	37.8472	51.7788
MSR [16]	13.1728	0.4787	1428.35	1414.38	8.1136	32.6192	45.9488
Dong [7]	16.7165	0.5824	409.18	89.32	8.3157	34.4683	46.6898
NPE [40]	16.9697	0.5894	681.42	479.72	8.4390	35.1236	47.6651
SRIE [11]	11.8552	0.4979	**370.44**	**83.81**	**7.2869**	**27.6113**	**27.7037**
MF [10]	16.9662	0.6049	429.20	211.19	8.8770	34.7835	47.1974
BIMEF [47]	13.8753	0.5771	387.83	141.16	7.5150	27.6511	28.1861
LIME [13]	16.7586	0.5644	800.34	695.50	8.3777	36.0971	49.9709
NPIE-MLLS [39]	16.6972	0.5945	548.25	317.40	8.1588	33.8586	44.0897
RetinexNet [43]	16.7740	0.5594	1059.27	993.29	8.8792	39.5860	57.6731
DeepUPE [38]	**20.3736**	**0.6379**	444.51	284.92	7.8642	32.8006	40.6565
JED [36]	13.6857	0.6280	398.14	432.89	5.3767	29.0680	40.7568
RRM [23]	13.8765	0.6577	453.30	460.87	5.8100	34.9902	47.3000
KinD [49]	17.6476	0.7601	549.03	493.00	4.7100	26.6443	45.2404
KinD-nonblind [49]	20.3792	**0.8045**	449.01	406.45	5.3546	32.4347	65.6035
Ours	**22.5156**	0.7864	**289.05**	**277.35**	**3.6354**	**21.7781**	**21.0840**



(a) Input (b) SRIE [10] (c) RRM [23] (d) KinD [49] (e) Ours

Fig. 7: Illustration of Retinex decomposition results. (b)-(e): Top: illumination; Bottom: reflectance.

no methods outperforming others in terms of all metrics. Specifically, SRIE outperforms other methods in terms of LOE. It is not surprising that none of methods with denoising mechanisms is among the best ones on these datasets, since images from these datasets are most underexposed with absence of noise. In order to deal with noise, the methods with denoising mechanisms in low-light images inevitably brings smooth artifacts to the images, which have negative effects on the quality evaluation. We present results on these datasets more to evaluate whether the methods designed for severe noise can generalize well for underexposed images.

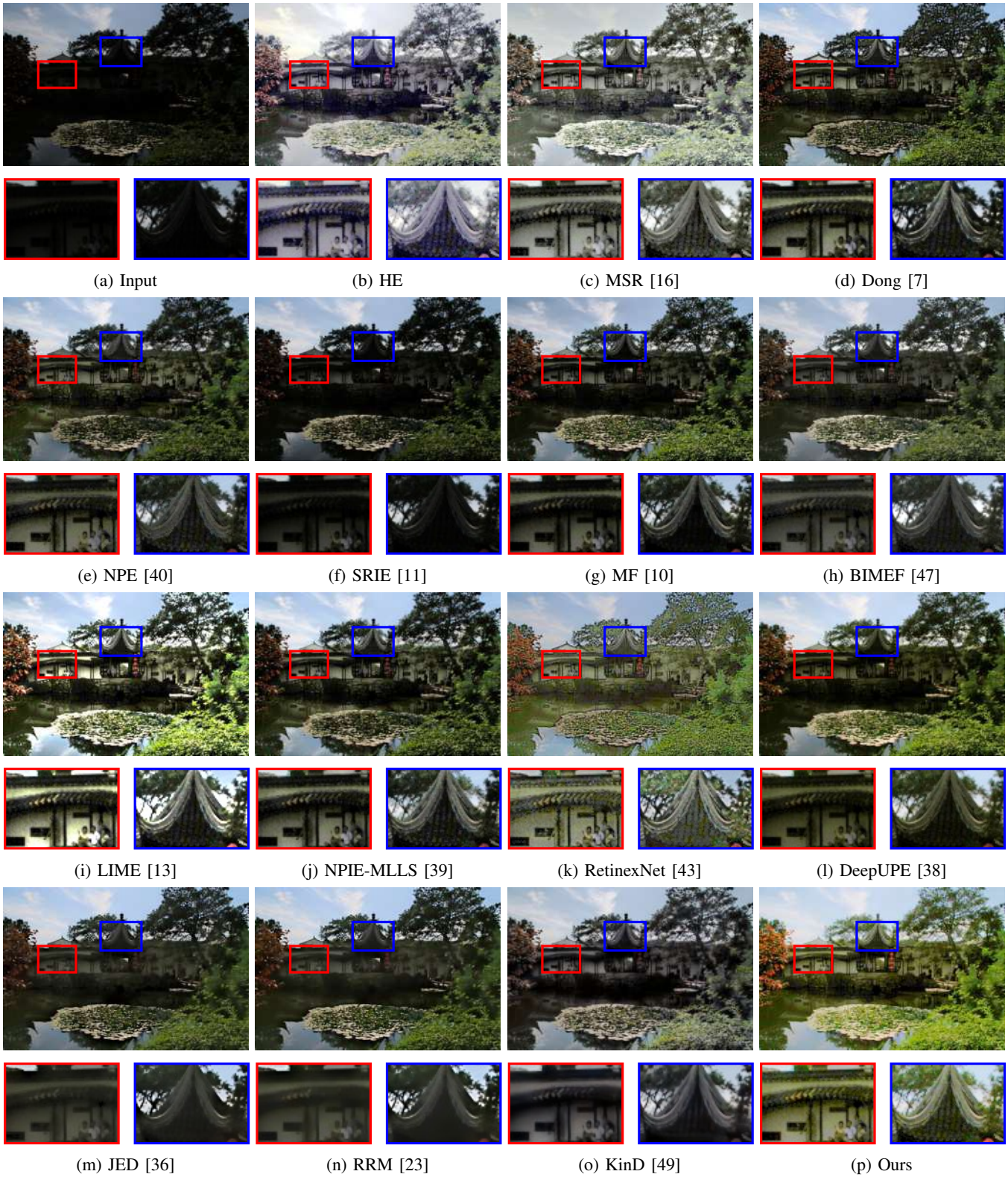
It can be seen that, the proposed method still obtained good results the on MEF and DICM datasets. On the LIME dataset, the score of the proposed method is only lower than JED in terms of NIQE. In the NPE dataset, the score of the proposed method is only lower than KinD. The qualitative results are presented in the supplementary material. Please see Fig. 8 and 9 for visual comparison.

TABLE II: Quantitative results on the MEF [26] dataset.

Method	LOE↓	NIQE↓	BRISQUE↓	PIQE↓
HE	236.15	3.6455	26.3202	42.0736
MSR [16]	1011.88	**3.3090**	23.0138	38.4410
Dong [7]	221.72	4.0989	26.4946	37.0443
NPE [40]	422.32	3.5295	23.7685	35.8246
SRIE [11]	210.26	3.4741	22.0880	37.7866
MF [10]	208.13	3.4923	22.7244	34.1056
BIMEF [47]	**155.62**	3.3290	**20.2203**	**33.7428**
LIME [13]	939.11	3.7023	24.0577	40.0240
NPIE-MLLS [39]	344.95	3.3370	22.3205	34.8772
RetinexNet [43]	708.25	4.4099	26.0365	41.2155
DeepUPE [38]	214.77	3.3717	22.6911	37.2156
JED [36]	294.15	4.5209	30.7195	47.9936
RRM [23]	311.39	5.0621	32.9379	52.8415
KinD [49]	275.47	3.8767	30.4408	53.9012
Ours	**172.80**	**3.4673**	**22.2387**	**35.6976**

E. Ablation study

1) Ablation study on the transforms: One component to distinguish our method from existing ones is the spatially



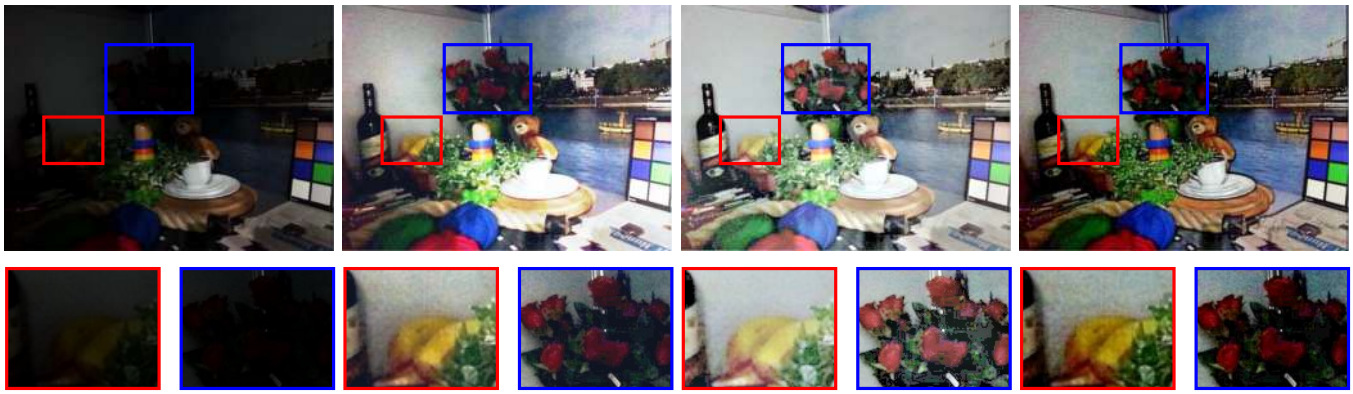
(a) Input (b) HE (c) MSR [16] (d) Dong [7]

(e) NPE [40] (f) SRIE [11] (g) MF [10] (h) BIMEF [47]

(i) LIME [13] (j) NPIE-MLLS [39] (k) RetinexNet [43] (l) DeepUPE [38]

(m) JED [36] (n) RRM [23] (o) KinD [49] (p) Ours

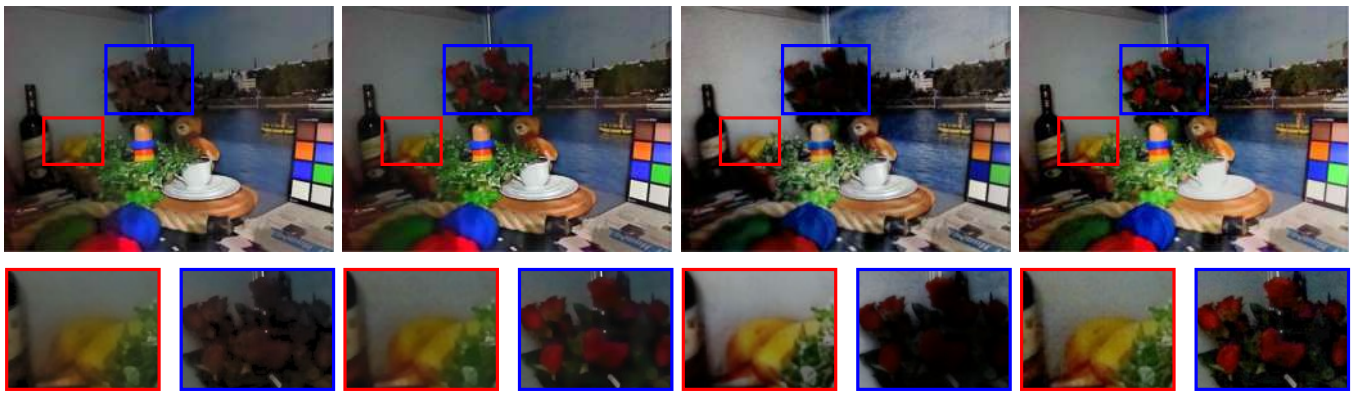
Fig. 8: Visual comparisons of different methods on low-light images from the MEF dataset.



(a) Input (b) HE (c) MSR [16] (d) Dong [7]

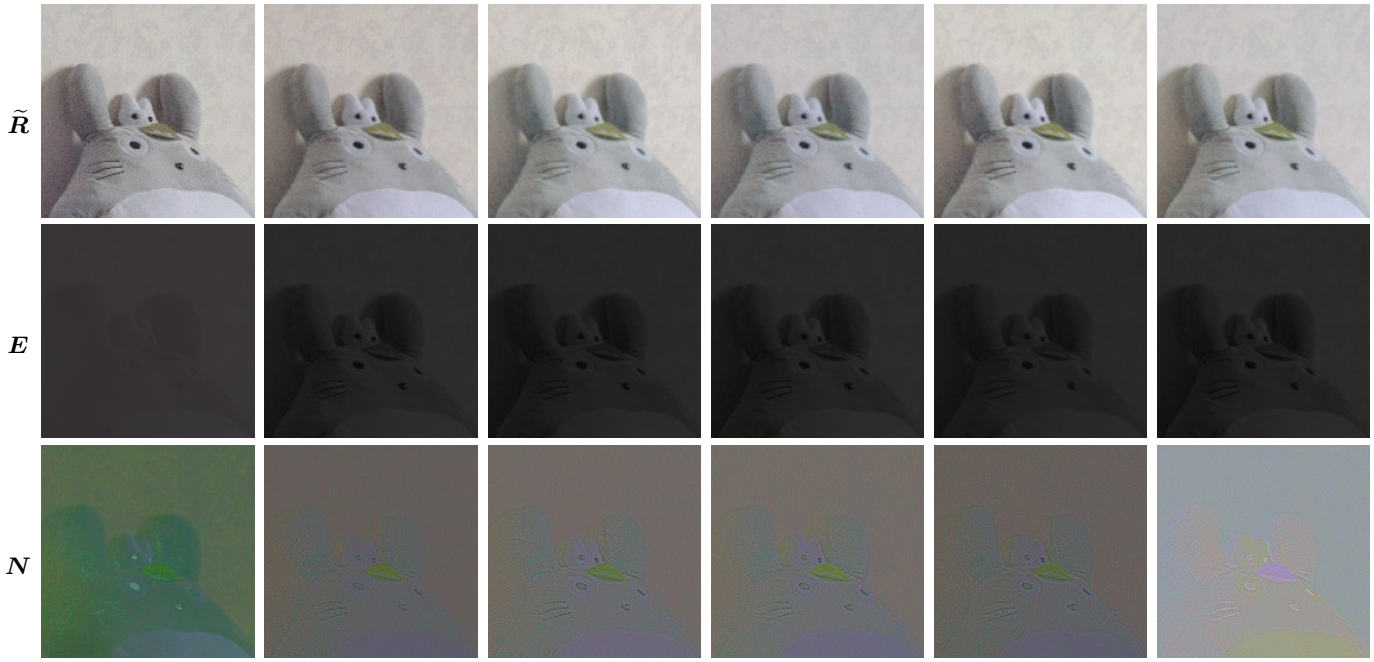
(e) NPE [40] (f) SRIE [11] (g) MF [10] (h) BIMEF [47]

(i) LIME [13] (j) NPIE-MLLS [39] (k) RetinexNet [43] (l) DeepUPE [38]



(m) JED [36] (n) RRM [23] (o) KinD [49] (p) Ours

Fig. 9: Visual comparisons of different methods on low-light images from the LIME dataset.



(a) w/o $\mathcal{T}_{W,\Delta}$ (b) $\mathcal{T}_{W}$ (c) $\mathcal{T}_{W,\Delta}$ $W=7$ (d) $\mathcal{T}_{W,\Delta}$ $W=15$ (e) $\mathcal{T}_{W,\Delta}$ $W=31$ (f) learning $I-N$

Fig. 10: Qualitative comparison of the results using different transforms.

TABLE III: Quantitative results on the DICM [22] dataset.

Method	LOE↓	NIQE↓	BRISQUE↓	PIQE↓
HE	270.31	3.8635	25.4847	41.2639
MSR [16]	1127.82	**3.6766**	26.0095	40.1992
Dong [7]	276.97	4.1191	26.7334	36.0868
NPE [40]	209.00	3.7601	**25.3145**	37.1805
SRIE [11]	**162.22**	3.8987	27.6980	39.3550
MF [10]	321.18	3.8441	25.6823	35.9124
BIMEF [47]	239.27	3.8459	26.8110	36.6763
LIME [13]	1107.77	3.8588	26.8838	41.3392
NPIE-MLLS [39]	264.60	3.7360	25.4938	37.0152
RetinexNet [43]	636.16	4.4154	26.6565	37.9280
DeepUPE [38]	193.38	3.9229	26.9957	**35.4541**
JED [36]	483.89	4.2605	27.4469	41.3614
RRM [23]	518.37	4.5970	30.1769	42.0824
KinD [49]	261.77	4.1505	30.6981	47.2272
Ours	**235.23**	**3.7409**	**26.4639**	**32.1176**

TABLE IV: Quantitative results on the LIME [13] dataset.

Method	LOE↓	NIQE↓	BRISQUE↓	PIQE↓
HE	322.22	4.3966	22.6083	42.6408
MSR [16]	864.68	3.7642	22.6987	40.3416
Dong [7]	241.40	4.0516	26.2226	38.4488
NPE [40]	334.92	3.9048	**22.1569**	36.4812
SRIE [11]	**106.31**	3.7863	24.1806	**34.7993**
MF [10]	183.25	4.0673	22.2979	36.4530
BIMEF [47]	136.90	3.8596	23.1353	35.7954
LIME [13]	793.90	4.1549	22.3094	41.0324
NPIE-MLLS [39]	300.51	**3.5788**	22.5060	37.5379
RetinexNet [43]	539.64	4.5983	26.1007	42.7741
DeepUPE [38]	185.88	3.8982	25.6375	37.4441
JED [36]	270.76	4.7180	28.5736	42.8337
RRM [23]	276.12	4.6426	29.1015	41.3977
KinD [49]	255.80	4.7632	26.7730	45.3084
Ours	**153.99**	**4.0431**	**23.1246**	**37.9949**

TABLE V: Quantitative results on the NPE [40] dataset.

Method	LOE↓	NIQE↓	BRISQUE↓	PIQE↓
HE	212.75	3.9780	24.0995	38.6459
MSR [16]	1270.14	4.3663	24.2051	35.7164
Dong [7]	181.15	4.1263	**23.1679**	**31.9268**
NPE [40]	204.65	**3.9520**	24.4615	31.9680
SRIE [11]	**159.34**	3.9795	25.6351	35.3823
MF [10]	303.82	4.1048	26.2524	34.3280
BIMEF [47]	233.27	4.1328	24.4256	33.4234
LIME [13]	1174.92	4.2629	26.1337	36.8072
NPIE-MLLS [39]	219.79	4.0245	25.1294	32.1868
RetinexNet [43]	653.85	4.5669	26.8732	34.5105
DeepUPE [38]	173.33	4.3201	23.8473	35.1940
JED [36]	595.71	4.9958	26.5272	41.1340
RRM [23]	634.86	4.8452	24.7838	38.6205
KinD [49]	**180.91**	**4.1607**	**24.2792**	43.1331
Ours	275.00	4.3807	24.8661	**37.0634**

TABLE VI: Ablation study on the transforms for estimating N.

transforms	PSNR(dB)	SSIM
another set of affine matrices	20.5639	0.6718
non-deformable (rigid) kernels	21.2876	0.7579
deformable kernels with W=7	21.6337	0.7760
deformable kernels with W=15	**22.5156**	**0.7864**
deformable kernels with W=31	21.2314	0.7741

varying deformable convolution learned in the bilateral space. To verify its effectiveness of the large receptive field and irregular sampling positions, we conduct controlled experiments as listed in Table VI. The corresponding results generated by transforms with different receptive fields are shown in Fig 10. The result in the first row of Table VI is obtained by using the point operation for image-to-noise mapping in the same form of that for image-to-illumination mapping. In this setting, the receptive field of the learned kernels is 1×1. As shown in Fig 10 (a), the produced results with only the point operations

are dominated by amplified noise as it is hard to distinguish those noise without neighborhood information. The second row shows results using learned kernels sampling on a rigid square neighborhood. The next three rows show results using deformable convolution with both learned kernels and learned sampling positions from windows of size $W \times W$, which can better distinguish the noise from texture, as evidenced by the noise component $\boldsymbol{N}$ in Fig 10 (c), (d) and (e). It is noted that the rows from top to bottom show results obtained by transforms with increasingly larger sampling window size. We evaluate the proposed method with exponential increased window size and find that learning deformable convolution with a receptive field of 15×15 yields the best performance.

2) Ablation study on intermediates to be estimated: We verify the effectiveness of the estimation of the noise $\boldsymbol{N}$ by comparing the proposed scheme with the estimation of the noise-free low-light image $\boldsymbol{I} - \boldsymbol{N}$ instead. It can be seen that, estimation of the noise reveals more noise as shown in Fig 10 (d) *v.s.* (f) and yields better performance as shown in Table VII.

TABLE VII: Ablation study on intermediates to be estimated.

Estimated intermediates	PSNR(dB)	SSIM
illumination $\boldsymbol{E}$ and noise-free image $\boldsymbol{I} - \boldsymbol{N}$	21.0485	0.7621
illumination $\boldsymbol{E}$ and noise $\boldsymbol{N}$	**22.5156**	**0.7864**

3) Ablation study on the loss function: We conduct ablation study on the proposed loss function. See Table VIII and Fig. 11 for the quantitative and qualitative results respectively. Firstly, as shown in Table VIII, in all settings, our results are better than the most recent existing work DeepUPE [38] which only estimates illumination map. It indicates the effectiveness of the proposed framework, which handles color and noise simultaneously and performs image-to-noise mapping by edge-aware deformable convolution.

It can be seen from the first three rows in Table VIII that the performance degenerates without ℓ_{grad}. The edge-aware loss $\mathcal{L}_r$ is effective in the proposed framework. The results also demonstrat the superiority in our setting over the perceptually motivated SSIM loss with $\ell_{\text{ssim}} = 1 - \text{SSIM}$.

The last three rows demonstrate the effectiveness of the proposed priors on illumination and noise, without which the results are inferior, especially when there are artifacts in the flat regions of the images and the smoothness of the illumination is hard to preserve, as shown in Fig. 11 (d), (e) and (f). Although with or without $\mathcal{L}_e$ yield comparable performance, without $\mathcal{L}_e$ might produce some artifacts as shown in Fig. 11 (e). Using ℓ_1 norm to regularize the variation of illumination shows better performance over ℓ_2 norm for $\mathcal{L}_e$.

TABLE VIII: Ablation study on the loss function.

$\mathcal{L}_r$	$\mathcal{L}_e$	$\mathcal{L}_n$	PSNR(dB)	SSIM
w/o ℓ_{grad}	default	default	20.9161	0.7570
$\ell_{\text{grad}} \rightarrow \ell_{\text{ssim}}$	default	default	20.7780	0.7614
default	default	default	**22.5156**	**0.7864**
default	default	*None*	21.3178	0.7811
default	*None*	default	22.4539	0.7822
default	ℓ_2	default	21.1428	0.7380

V. Summary

This paper develops a deep learning method for low-light image enhancement, with the focus on the handling of measurement noise. Motivated by the inherently coupled relationship between illumination and measurement noise, we proposed a novel deep bilateral Retinex method, which performs Retinex decomposition in the bilateral space of low-light images. The proposed method is extensively evaluated in several benchmark datasets and compared to several representative related methods. The experiments show that the proposed method outperforms the compared methods, especially in the case of very low lighting conditions. In future, we plan to investigate possible applications of the proposed method in other image processing tasks involving Retinex decomposition.

References

[1] T. Arici, S. Dikbas, and Y. Altunbasak. A Histogram Modification Framework and Its Application for Image Contrast Enhancement. *IEEE Trans. Image Process.*, 18(9):1921–1935, Sept. 2009.

[2] S. Bako, T. Vogels, B. Mcwilliams, M. Meyer, J. NováK, A. Harvill, P. Sen, T. Derose, and F. Rousselle. Kernel-predicting Convolutional Networks for Denoising Monte Carlo Renderings. *ACM Trans. Graph.*, 36(4):97:1–97:14, July 2017.

[3] J. Cai, S. Gu, and L. Zhang. Learning a Deep Single Image Contrast Enhancer from Multi-Exposure Images. *IEEE Trans. Image Process.*, 27(4):2049–2062, Apr. 2018.

[4] T. Celik and T. Tjahjadi. Contextual and Variational Contrast Enhancement. *IEEE Trans. Image Process.*, 20(12):3431–3441, Dec. 2011.

[5] C. Chen, Q. Chen, J. Xu, and V. Koltun. Learning to See in the Dark. In *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 3291–3300, June 2018.

[6] J. Chen, A. Adams, N. Wadhwa, and S. W. Hasinoff. Bilateral Guided Upsampling. *ACM Trans. Graph.*, 35(6):203:1–203:8, Nov. 2016.

[7] X. Dong, Y. A. Pang, and J. G. Wen. Fast Efficient Algorithm for Enhancement of Low Lighting Video. In *ACM SIGGRAPH*, pages 69:1–69:1, 2010.

[8] M. Elad. Retinex by Two Bilateral Filters. In *Proc. Scale Space and PDE Methods in Computer Vision (Scale-Space)*, pages 217–229, 2005.

[9] X. Fu, Y. Liao, D. Zeng, Y. Huang, X. Zhang, and X. Ding. A Probabilistic Method for Image Enhancement With Simultaneous Illumination and Reflectance Estimation. *IEEE Trans. Image Process.*, 24(12):4965–4977, Dec. 2015.

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

[11] 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 Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 2782–2790, 2016.

[12] M. Gharbi, J. Chen, J. T. Barron, S. W. Hasinoff, and F. Durand. Deep Bilateral Learning for Real-time Image Enhancement. *ACM Trans. Graph.*, 36(4):118:1–118:12, July 2017.

[13] X. Guo, Y. Li, and H. Ling. LIME: Low-Light Image Enhancement via Illumination Map Estimation. *IEEE Trans. Image Process.*, 26(2):982–993, Feb. 2017.

[14] 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)*, pages 1026–1034, Dec. 2015.

[15] S. Huang, F. Cheng, and Y. Chiu. Efficient Contrast Enhancement Using Adaptive Gamma Correction With Weighting Distribution. *IEEE Trans. Image Process.*, 22(3):1032–1041, Mar. 2013.

[16] 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.*, 6(7):965–976, July 1997.

[17] D. J. Jobson, Z. Rahman, and G. A. Woodell. Properties and performance of a center/surround retinex. *IEEE Trans. Image Process.*, 6(3):451–462, Mar. 1997.

[18] B. Kim, J. Ponce, and B. Ham. Deformable Kernel Networks for Joint Image Filtering. *arXiv:1910.08373 [cs]*, Oct. 2019.

[19] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel. A Variational Framework for Retinex. *Int. J. Comput. Vision*, 52(1):7–23, Apr. 2003.

(a) $\mathcal{L}_r$ w/o ℓ_{grad} (b) $\ell_{\text{grad}} \rightarrow \ell_{\text{ssim}}$ (c) default (d) w/o $\mathcal{L}_n$ (e) w/o $\mathcal{L}_e$ (f) $\mathcal{L}_e$ w/ ℓ_2

Fig. 11: Ablation study on the loss function.

[20] D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. In *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015.

[21] E. H. Land. The Retinex Theory of Color Vision. *Sci. Amer.*, 237(6):108–129, 1977.

[22] C. Lee, C. Lee, and C.-S. Kim. Contrast Enhancement Based on Layered Difference Representation of 2D Histograms. *IEEE Trans. Image Process.*, 22(12):5372–5384, Dec. 2013.

[23] 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.*, 27(6):2828–2841, June 2018.

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

[25] A. Łoza, D. R. Bull, P. R. Hill, and A. M. Achim. Automatic contrast enhancement of low-light images based on local statistics of wavelet coefficients. *Digital Signal Process.*, 23(6):1856–1866, Dec. 2013.

[26] K. Ma, K. Zeng, and Z. Wang. Perceptual Quality Assessment for Multi-Exposure Image Fusion. *IEEE Trans. Image Process.*, 24(11):3345–3356, Nov. 2015.

[27] W. Ma, J.-M. Morel, S. Osher, and A. Chien. An L 1-based variational model for Retinex theory and its application to medical images. In *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 153–160, June 2011.

[28] B. Mildenhall, J. T. Barron, J. Chen, D. Sharlet, R. Ng, and R. Carroll. Burst Denoising With Kernel Prediction Networks. In *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 2502–2510, 2018.

[29] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Trans. Image Process.*, 21(12):4695–4708, Dec. 2012.

[30] A. Mittal, R. Soundararajan, and A. C. Bovik. Making a "Completely Blind" Image Quality Analyzer. *IEEE Signal. Proc. Let.*, 20(3):209–212, Mar. 2013.

[31] V. N, P. D, M. C. Bh, S. S. Channappayya, and S. S. Medasani. Blind image quality evaluation using perception based features. In *21st Nat. Conf. Commun. (NCC)*, pages 1–6, Feb. 2015.

[32] M. Ng and W. Wang. A Total Variation Model for Retinex. *SIAM J. Imag. Sci.*, 4(1):345–365, Jan. 2011.

[33] S. Niklaus, L. Mai, and F. Liu. Video Frame Interpolation via Adaptive Convolution. In *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 670–679, 2017.

[34] S. Niklaus, L. Mai, and F. Liu. Video Frame Interpolation via Adaptive Separable Convolution. In *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pages 261–270, 2017.

[35] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Proc. Annu. Conf. Neural Inf. Process. Syst. (NeurIPS)*, pages 8026–8037, 2019.

[36] X. Ren, M. Li, W. Cheng, and J. Liu. Joint Enhancement and Denoising Method via Sequential Decomposition. In *IEEE Int. Symp. Circuits Syst. (ISCAS)*, pages 1–5, May 2018.

[37] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pages 839–846, Jan. 1998.

[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 Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, page 9, 2019.

[39] S. Wang and G. Luo. Naturalness Preserved Image Enhancement Using a Priori Multi-Layer Lightness Statistics. *IEEE Trans. Image Process.*, 27(2):938–948, Feb. 2018.

[40] S. Wang, J. Zheng, H. Hu, and B. Li. Naturalness Preserved Enhancement Algorithm for Non-Uniform Illumination Images. *IEEE Trans. Image Process.*, 22(9):3538–3548, Sept. 2013.

[41] W. Wang, C. Wei, W. Yang, and J. Liu. GLADNet: Low-Light Enhancement Network with Global Awareness. In *Proc. IEEE Int. Conf. Automat. Face Gesture Recognit. (FG)*, pages 751–755, May 2018.

[42] 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.*, 13(4):600–612, Apr. 2004.

[43] C. Wei, W. Wang, W. Yang, and J. Liu. Deep Retinex Decomposition for Low-Light Enhancement. In *Br. Mac. Vis. Conf. (BMVC)*, Aug. 2018.

[44] K. Wei, Y. Fu, J. Yang, and H. Huang. A Physics-based Noise Formation Model for Extreme Low-light Raw Denoising. In *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Apr. 2020.

[45] L. Xu, Q. Yan, Y. Xia, and J. Jia. Structure extraction from texture via relative total variation. *ACM Trans. Graph.*, 31(6):139:1–139:10, Nov. 2012.

[46] X. Xu, M. Li, and W. Sun. Learning Deformable Kernels for Image and Video Denoising. *arXiv:1904.06903 [cs]*, Apr. 2019.

[47] Z. Ying, G. Li, and W. Gao. A Bio-Inspired Multi-Exposure Fusion Framework for Low-light Image Enhancement. *arXiv:1711.00591 [cs]*, Nov. 2017.

[48] L. Yuan and J. Sun. Automatic Exposure Correction of Consumer Photographs. In *Proc. IEEE Eur. Conf. Comput. Vis. (ECCV)*, pages 771–785, 2012.

[49] Y. Zhang, J. Zhang, and X. Guo. Kindling the Darkness: A Practical Low-light Image Enhancer. In *Proc. ACM Int. Conf. Multimed. (ACM MM)*, May 2019.

[50] Q. Zhao, P. Tan, Q. Dai, L. Shen, E. Wu, and S. Lin. A Closed-Form Solution to Retinex with Nonlocal Texture Constraints. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(7):1437–1444, July 2012.