Self-Supervised Low-Light Image Enhancement Using Discrepant Untrained Network Priors

Jinxiu Liang†, Yong Xu, Yuhui Quan*, Boxin Shi, and Hui Ji

***Abstract*—This paper proposes a deep learning method for low-light image enhancement, which exploits the generation capability of Neural Networks (NNs) while requiring no training samples except the input image itself. Based on the Retinex decomposition model, the reflectance and illumination of a low-light image are parameterized by two untrained NNs. The ambiguity between the two layers is resolved by the discrepancy between the two NNs in terms of architecture and capacity, while the complex noise with spatially-varying characteristics is handled by an illumination-adaptive self-supervised denoising module. The enhancement is done by jointly optimizing the Retinex decomposition and the illumination adjustment. Extensive experiments show that the proposed method not only outperforms existing non-learning-based and unsupervised-learning-based methods, but also competes favorably with some supervised-learning-based methods in extreme low-light conditions.**

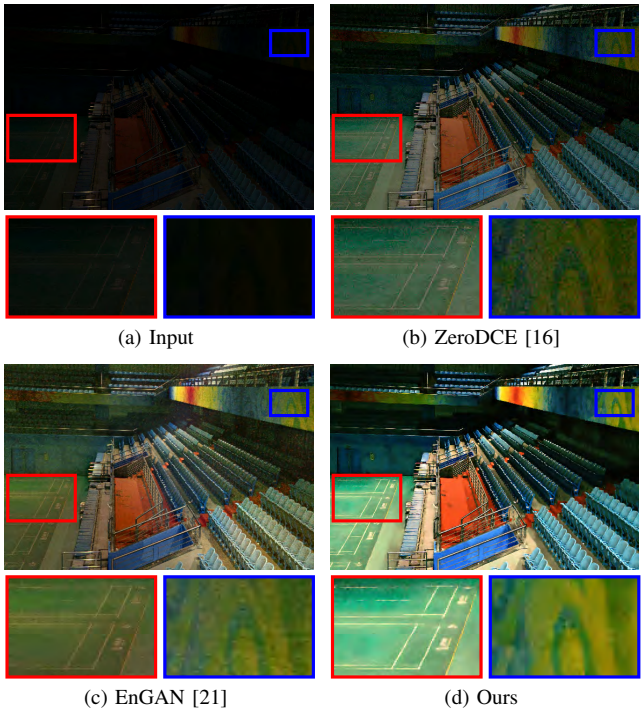
***Index Terms*—Low-light image enhancement, Retinex decomposition model, Untrained network priors**

I. Introduction

Low-light images refer to the images captured in low-light conditions, which often suffer from poor visibility with low contrast and low signal-to-noise-ratio (SNR). Low-light image enhancement (LIE) is about improving the visual quality of a low-light image to have the one with better visibility and higher SNR. Such a technique not only demonstrates its practical value in digital photography, but also benefits many downstream computer vision applications such as surveillance and tracking in low-light conditions.

There has been an enduring effort on developing effective techniques for LIE. In recent years, the Retinex decomposition model (RDM) that assumes an image could be decomposed as the element-wise product of a reflectance layer and an illumination layer has been one prominent choice for developing powerful LIE techniques [10], [17], [47], [56], [27]. The performance of existing RDM-based LIE methods heavily relies on hand-crafted priors which might be inaccurate for characterizing the reflectance and illumination layers on real-world images.

More recently, supervised deep learning has been widely used for LIE with impressive performance [4], [42], [50], [52], [27]. However, their success largely depends on a large quantity of paired training samples with statistical characteristics aligned with test images. The collection of such data is often costly or technically challenging, *e.g.*, collecting paired normal/low-light images of the same scene with moving objects in outdoor environments. A few unsupervised learning approaches [21], [16] that avoid using paired data have been proposed. However, these approaches pose higher



(a) Input (b) ZeroDCE [16] (c) EnGAN [21] (d) Ours

Fig. 1. Visual results of two external-dataset-based unsupervised methods and the proposed dataset-free one. It can be seen that ZeroDCE fails to recover textures from the low-light image with severe noise as it is not present in training data, and EnGAN produces undesired distorted color which is biased by its training data. In contrast, the proposed method restores the global illumination well while suppressing the noise.

†Part of work was finished at South China University of Technology.
*Corresponding author: Yuhui Quan (Email: csyhquan@scut.edu.cn)
Jinxiu Liang and Boxin Shi are with National Engineering Research Center of Visual Technology, School of Computer Science, Peking University, Beijing, China. Boxin Shi is also with Institute for Artificial Intelligence, Peking University, Beijing, China, Beijing Academy of Artificial Intelligence, Beijing, China, and Peng Cheng Laboratory, Shenzhen, China. (Email: cssherryliang@pku.edu.cn; shiboxin@pku.edu.cn)
Yong Xu and Yuhui Quan are with School of Computer Science and Engineering at South China University of Technology, Guangzhou, China, and Pazhou Laboratory, Guangzhou, China. Yong Xu is also with Peng Cheng Laboratory, Shenzhen, China, and Communication and Computer Network Laboratory of Guangdong, Guangzhou, China. (Email: csyhquan@scut.edu.cn; yxu@scut.edu.cn)
Hui Ji is with Department of Mathematics at National University of Singapore, Singapore. (Email: matjh@nus.edu.sg)
This work was supported in part by National Natural Science Foundation of China under Grants 62072188, 61872151, 62136001, and 62088102, in part by Natural Science Foundation of Guangdong Province under Grants 2020A1515011128, and 2022A1515011755, and in part by Science and Technology Program of Guangdong Province under Grant 2019A050510010.

requirements on the distribution of training data to satisfy their assumptions, and therefore they still take considerable efforts and costs to collect and organize the data.

In many scenarios, training data is insufficient or biased. An NN trained on a specific dataset may not perform consistently well across all test images. Often, novel patterns not present in training data are hardly recovered from test samples by the trained NN (Fig. 1 (b)). Similarly, undesired artifacts may appear in an enhanced image (Fig. 1 (c)). Such issues are critical for many applications such as forensics and clinical diagnosis [7], [1]. It is certainly valuable to have an LIE method that leverages the power of deep learning for good performance while requiring no training data.

To achieve this goal, our main idea is to model illumination and reflectance in RDM by the recently developed *untrained network priors* [41], which show that the structure of an NN (rather than the weights learned from training data) is able to capture low-level image statistics. That is, while optimizing an NN with randomly initialized weights to generate a given noisy image, regular image structures and textures will be fitted before random noise. Inspired by this, we propose to solve RDM by optimizing two untrained NNs for modeling the two layers such that their element-wise product yields the given low-light image. Then its normal-light counterpart can be obtained by re-illuminating the image using an adjusted illumination subsequently.

1) Challenges: The above process does not require data for pre-training, which satisfies our need; however, implementing it for LIE is highly non-trivial due to the following issues:

- *Ambiguity between two layers.* Illumination expresses the light intensities striking surfaces of scene/objects, while reflectance encodes the physical characteristics of surfaces of scene/objects, *e.g*., textures and other fine details. As light intensities striking over a surface often vary slowly, illumination is usually assumed to be smoother than reflectance [17], [2], [42], [49]. Appropriate discrepancy between the two untrained NNs is needed for accurately determining the attribution of image gradients.
- *Difficulty on noise handling.* Retinex decomposition is sensitive to noise. Consider a simplified setting where we already have an accurate estimation of illumination. As many entries of the illumination are close to zero for a low-light image, a direct inversion for estimating the reflectance will significantly magnify the measurement noise. To make it worse, the measurement noise for a low-light image usually has complex spatially-varying statistical characteristics [48]. Also, many details (edges) in low-light images are of very small magnitude, which are hard to preserve during noise removal.
- *Flexibility on illumination adjustment.* Without looking at any normal-light image, it is not easy to determine the best hyper-parameter of illumination adjustment for each input image. Therefore, developing an adaptive illumination adjustment mechanism for our case is challenging.

2) Solutions and Contributions: In this paper, we tackle the three challenges by proposing the following strategies: *i)* Motivated by the empirical observation that an untrained NN with lower capacity tends to fit smoother structure, we propose to resolve the ambiguity between illumination and reflectance by introducing discrepancy on both the architecture and model capacity between the two NNs. *ii)* Since pixels with different SNRs should be processed adaptively, we propose a self-supervised denoising scheme with a spatially-varying characteristic driven by illumination. It is motivated by the observation that SNR is related to illumination intensity. *iii)* We introduce a differentiable histogram balance loss such that the parameterized illumination adjustment module can be jointly optimized with the two NNs for maximizing the quality of the enhanced image in terms of entropy.

The proposed dataset-free unsupervised method produces competitive performance to the dataset-based ones for LIE from only the test image itself (Fig. 1 (d)). To summarize, this paper proposes a training-data-free LIE method with following contributions:

- Integrating discrepant untrained NN priors into RDM, which successfully exploits untrained NN priors for LIE.
- Proposing an illumination-adaptive self-supervised denoising scheme for handling spatial-variant real noise.
- Unifying the workflows of Retinex decomposition and illumination adjustment for LIE with better adaptivity.

II. Related Work

A. Conventional methods

Earlier works of LIE directly adjust image intensity to improve the contrast, *e.g*., the classic gamma correction. The adaptivity of adjustment is later improved by using certain parametric S-shape tone curves with parameters estimated from camera response models [53], [38] or estimated under designed criteria of good exposure [5]. Histogram equalization is another intensity/color adjustment technique, which modifies pixel values to fit certain distribution [24]. These methods focus on contrast enhancement, which often result in unnatural visual appearance.

LIE can be recasted as the Retinex decomposition problem, which requires priors to resolve the ambiguity between the two layers. Local smoothness is a prominent prior for illumination with various implementations, such as bilateral filtering [8], ℓ_2-penalty on gradients [23], [9], [11], weighted ℓ_1-penalty on gradients [33], [15], [17], [2], [49], and some others [40]. Statistical priors [44], [10], [43] and physical models [54], [45] on lighting are also used in many methods. In comparison to illumination, reflectance is more challenging to characterize. A widely-used assumption is that a reflectance layer contains fine textures [2], and thus the piece-wise continuity prior is often used for reflectance layers in existing work *e.g*., [30], [31], [33], [23], [55], [2], [49]. Note that texture of a reflectance map is easily confused with noise. Thus, there have been extensive studies on the robust Retinex decomposition [8], [26], [37].

B. Supervised Learning on Paired Data

There are some studies treating LIE as an image-to-image mapping obtained via supervised learning on a dataset with a large number of paired samples. Lore *et al*. [28] trained a stacked denoising auto-encoder to fit the mapping. Chen *et*

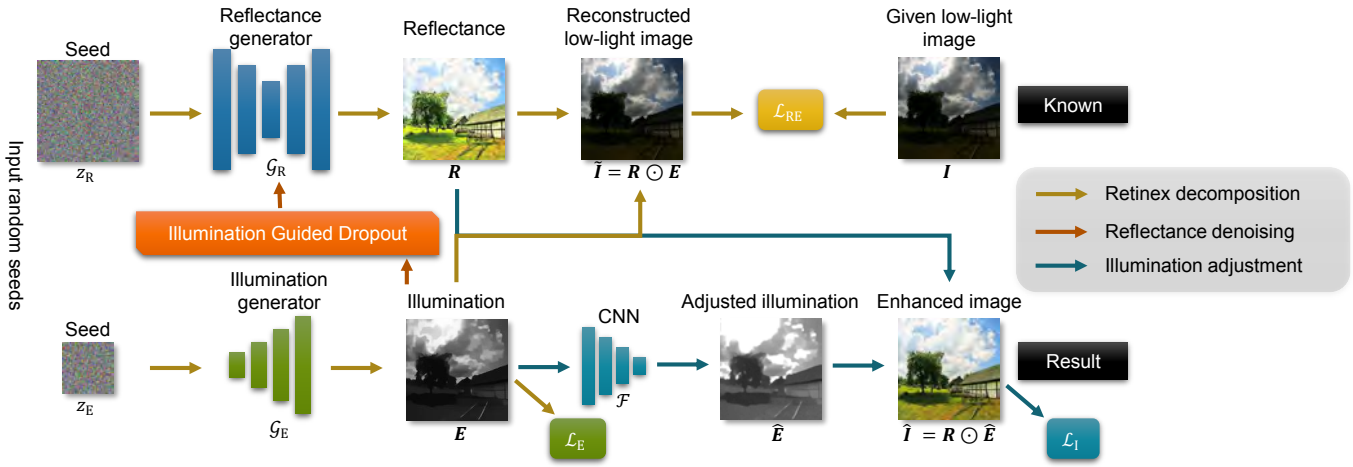


Fig. 2. Overview of the proposed method. The two untrained NNs $\mathcal{G}_E, \mathcal{G}_R$ are randomly initialized to map fixed random seeds z_E, z_R onto illumination $\boldsymbol{E}$ and reflectance $\boldsymbol{R}$, whose element-wise product is optimized towards the known low-light image $\boldsymbol{I}$. For reflectance denoising, an illumination-guided dropout module is introduced inside $\mathcal{G}_R$. Then a CNN $\mathcal{F}$ is adopted to map $\boldsymbol{E}$ onto $\widehat{\boldsymbol{E}}$ for illumination adjustment. At last, the enhanced image $\hat{\boldsymbol{I}}$ is obtained from the adjusted illumination $\widehat{\boldsymbol{E}}$ and the denoised $\boldsymbol{R}$. Our framework is optimized in an end-to-end manner by minimizing three loss functions $\mathcal{L}_{RE}, \mathcal{L}_E, \mathcal{L}_I$.

al. [4] used a U-Net to learn the mapping with a focus on raw input instead of RGB, which was further improved by the idea of multi-exposure fusion in [58]. Li *et al*. [25] trained a recursive NN to enhance low-light images in a progressive way. Some methods learn the mapping via frequency decomposition. Cai *et al*. [3] proposed to learn separately for the low/high-frequency parts of a low-contrast image and its high-contrast counterpart. Similarly, Xu *et al*. [50] proposed to learn the mapping from low-light image to low/high-frequency parts of the normal-light reference sequentially. Ren *et al*. [36] proposed a hybrid NN that performs layer decomposition and then recovers global content and local details separately.

The RDM has also been exploited with supervised deep learning. Shen *et al*. [39] presented a deep NN for LIE which is built on the traditional multi-scale Retinex algorithm. Wei *et al*. [47] proposed to perform Retinex decomposition by an NN and then adjust illumination by another NN. Its off-the-shelf denoiser for reflectance refinement is replaced by a separately-trained NN in [56]. Focusing on underexposed images with negligible noise, Wang *et al*. [42] proposed to estimate the illumination by per-pixel affine transform learned by bilateral upsampling. To bridge the gap between fidelity and perceptual quality, Yang *et al*. [52] combined unpaired training to supervised learning on a recursive banded NN. All these methods require paired data for training, which is critical toward the success of supervised learning methods. Some work studies how to synthesize paired real-world normal/low-light images [39], [28] or how to collect them in an economic manner [3], [47], [4]. Even with such efforts, collecting sufficient real-world training data for real applications remains costly and troublesome. This issue is avoided in our method.

C. Unsupervised Learning on an Unorganized Dataset

There are a few works attempting to relax the prerequisite on paired training images for deep-learning-based LIE. Jiang *et al*. [21] leveraged adversarial learning with dual discriminators to effectively exploit unpaired data as positive and negative samples. Guo *et al*. [16] proposed to learn a set of parameterized curves for light enhancement, using a dataset containing images of different exposures. This is achieved through a set of well-designed non-reference loss functions. Liu *et al*. [27] proposed a lightweight and efficient optimization-inspired NN with searched architecture. While relaxing the prerequisite on paired data, these methods still require training samples highly related to the test image in terms of both image content and noise statistics for good performance.

D. Untrained NN Priors

In recent years, there is a rapid progress on studying untrained NN priors for image recovery [41], [19], [12], [35], [20], so as to avoid using training datasets. In these works, the images or image layers are modeled as being lying in the range of an untrained NN fed with a fixed seed. It is shown in [41] that optimizing the output of an untrained NN to fit a corrupted observation, instead of training the NN with massive input-target pairs, can capture local image correlation and act as a powerful image prior. Such untrained NN priors have been extensively explored in the context of denoising [41], deblurring [35], compression [19], layer decomposition [12], compressive sensing [20], background matting [51], *etc*.

We note a recently concurrent work RetinexDIP proposed in [57], which explores untrained NN priors for LIE; however it is remarkably different from the proposed method. The keys for applying untrained NN priors for LIE are: how to address a) the ambiguity between two NNs and b) the spatially-varying noise. For a), RetinexDIP combines hand-crafted priors with both NNs, while ours additionally exploits discrepancy between NN architectures. For b), RetinexDIP does not have an explicit mechanism, while ours uses a self-supervised illumination-guided denoising module. Furthermore, instead of taking a two-stage approach in RetinexDIP, we propose a unified illumination adjustment scheme. Thanks to these techniques developed in this work, the proposed method demonstrates superior performance over RetinexDIP in the experiments presented in Sec. IV.

III. Proposed Method

To restore a normal-light image $\widehat{\boldsymbol{I}} \in [0,1]^{M\times N\times C}$ from a given low-light image $\boldsymbol{I} \in [0,1]^{M\times N\times C}$ with spatial size $M \times N$ and channel number C ($C = 1$ or 3), the proposed approach optimizes three NNs over only a single input image $\boldsymbol{I}$, with the workflow shown in Fig. 2.

Based on RDM, an image $\boldsymbol{I}$ can be decomposed into a reflectance $\boldsymbol{R}$ and an illumination $\boldsymbol{E}$ as follows,

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

where $\odot$ denotes element-wise multiplication, and $\boldsymbol{N}$ denotes image noise. LIE built on the decomposition (1) is an ill-posed inverse problem. In this paper, we solve it by re-parameterizing $\boldsymbol{R}$ and $\boldsymbol{E}$ with two NNs denoted by

$$\boldsymbol{E} := \mathcal{G}_{\mathrm{E}}(\boldsymbol{z}_{\mathrm{E}};\boldsymbol{\omega}_{\mathrm{E}}), \boldsymbol{R} := \mathcal{G}_{\mathrm{R}}(\boldsymbol{z}_{\mathrm{R}};\boldsymbol{\omega}_{\mathrm{R}}) \in [0,1]^{M\times N\times C}, \tag{2}$$

where $\mathcal{G}_{\mathrm{E}}(\boldsymbol{z}_{\mathrm{E}};\boldsymbol{\omega}_{\mathrm{E}})$ ($\mathcal{G}_{\mathrm{R}}(\boldsymbol{z}_{\mathrm{R}};\boldsymbol{\omega}_{\mathrm{R}})$) is an NN parameterized by $\boldsymbol{\omega}_{\mathrm{E}}$ ($\boldsymbol{\omega}_{\mathrm{R}}$) with a fixed random seed $\boldsymbol{z}_{\mathrm{E}}$ ($\boldsymbol{z}_{\mathrm{R}}$) as its input. The seeds are independently drawn from the same Gaussian distribution. Given an $M \times N$ image, z_{R} is also $M \times N$, and z_{E} is $M/2^L \times N/2^L$, where L is the number of upsampling layers in $\mathcal{G}_{\mathrm{E}}$. The two NNs act as natural illumination/reflectance models that incorporate the priors on their intermediate layers.

Once the decomposition is done, we can obtain a new image $\widehat{\boldsymbol{I}}$ with better visibility by re-illuminating the image using a new illumination map $\widehat{\boldsymbol{E}}$:

$$\widehat{\boldsymbol{I}} := \widehat{\boldsymbol{E}} \odot \boldsymbol{R}, \tag{3}$$

where the adjusted version $\widehat{\boldsymbol{E}}$ is produced by another NN $\mathcal{F}(\cdot;\boldsymbol{\theta})$ from $\boldsymbol{E}$ by

$$\widehat{\boldsymbol{E}} := \mathcal{F}(\boldsymbol{E};\boldsymbol{\theta}) \in [0,1]^{M\times N\times C}. \tag{4}$$

The Retinex decomposition and the illumination adjustment are jointly optimized in an end-to-end manner, via optimizing the parameters

$$\min_{\boldsymbol{\omega}_{\mathrm{E}},\boldsymbol{\omega}_{\mathrm{R}},\boldsymbol{\theta}} \mathcal{L}_{\mathrm{RE}}(\boldsymbol{R}\odot\boldsymbol{E},\boldsymbol{I}) + \lambda_{\mathrm{E}}\mathcal{L}_{\mathrm{E}}(\boldsymbol{E}) + \lambda_{\mathrm{I}}\mathcal{L}_{\mathrm{I}}(\widehat{\boldsymbol{I}}), \tag{5}$$

where $\mathcal{L}_{\mathrm{RE}}$ is the reconstruction loss, $\mathcal{L}_{\mathrm{E}}, \mathcal{L}_{\mathrm{I}}$ are the regularization on $\boldsymbol{E}$ and $\widehat{\boldsymbol{I}}$ respectively, and $\lambda_{\mathrm{E}}, \lambda_{\mathrm{I}}$ are predefined weights. During inference, the sufficiently optimized NNs $\mathcal{G}_{\mathrm{E}}(\cdot;\boldsymbol{\omega}_{\mathrm{E}}^\star), \mathcal{G}_{\mathrm{R}}(\cdot;\boldsymbol{\omega}_{\mathrm{R}}^\star), \mathcal{F}(\cdot;\boldsymbol{\theta}^\star)$ generate $\boldsymbol{E}^\star, \boldsymbol{R}^\star, \widehat{\boldsymbol{E}}^\star$ respectively. Then the result $\widehat{\boldsymbol{I}}^\star$ is obtained in analogous to (3). See Algorithm 1 for the pseudo-code of the proposed method.

A. Retinex Decomposition via Untrained NN Priors

Illumination is often assumed smoother than reflectance. We also observe that while $\mathcal{G}_{\mathrm{E}}(\cdot) \odot \mathcal{G}_{\mathrm{R}}(\cdot)$ is optimized towards an image, the NN with lower capacity tends to fit smoother structure. Motivated by this, we propose to resolve the ambiguity between $\boldsymbol{E}$ and $\boldsymbol{R}$ by introducing certain discrepancy on both the NN architecture and model capacity between $\mathcal{G}_{\mathrm{E}}$ and $\mathcal{G}_{\mathrm{R}}$. Briefly, $\mathcal{G}_{\mathrm{E}}$ is set to a small under-parameterized CNN while $\mathcal{G}_{\mathrm{R}}$ a large over-parameterized one. Together with simple regularizations on illumination, the layer ambiguity is well addressed. The NN architectures and loss functions are detailed as follows.

Algorithm 1 LIE using discrepant untrained NN priors

Input: Low-light image $\boldsymbol{I}$; parameters $\lambda_{\mathrm{E}}, \lambda_{\mathrm{I}}, \tau$; maximum iterations S for optimization of NN parameters $\boldsymbol{\omega}_{\mathrm{E}}, \boldsymbol{\omega}_{\mathrm{R}}, \boldsymbol{\theta}$; iterations T for dropout ensemble; update iterations K for dropout probability maps $\{\boldsymbol{P}^{(l)}\}_{l=1}^{L}$

Output: Normal-light image $\widehat{\boldsymbol{I}}$

for $s = 0$ to S **do**
  Sample $z_{\mathrm{E}}^{(s)}$ and $z_{\mathrm{R}}^{(s)}$ from $\mathcal{N}(0,\sigma^2)$
  Sample $\boldsymbol{\omega}_{\mathrm{R}}^{(s)}$ according to $\{\boldsymbol{P}^{(l)}\}_{l=1}^{L}$ ▷ Eq. (8)
  $\boldsymbol{E}^{(s)} \leftarrow \mathcal{G}_{\mathrm{E}}(\boldsymbol{z}_{\mathrm{E}}^{(s)};\boldsymbol{\omega}_{\mathrm{E}}^{(s)})$,
  $\boldsymbol{R}^{(s)} \leftarrow \mathcal{G}_{\mathrm{R}}(\boldsymbol{z}_{\mathrm{R}}^{(s)};\boldsymbol{\omega}_{\mathrm{R}}^{(s)})$ ▷ Eq. (2)
  $\widehat{\boldsymbol{E}}^{(s)} \leftarrow \mathcal{F}(\boldsymbol{E}^{(s)};\boldsymbol{\theta}^{(s)})$ ▷ Eq. (4)
  $\widehat{\boldsymbol{I}}^{(s)} \leftarrow \widehat{\boldsymbol{E}}^{(s)} \odot \boldsymbol{R}^{(s)}$ ▷ Eq. (3)
  Compute the gradients *w.r.t.* $\boldsymbol{\omega}_{\mathrm{E}}, \boldsymbol{\omega}_{\mathrm{R}}, \boldsymbol{\theta}$ ▷ Eq. (5-7,14)
  Update $\boldsymbol{\omega}_{\mathrm{E}}^{(s+1)}, \boldsymbol{\omega}_{\mathrm{R}}^{(s+1)}, \boldsymbol{\theta}^{(s+1)}$ using the Adam
  if $s = (2n+1)K$ **and** $n \in \mathbb{N}$ **then**
    update $\{\boldsymbol{P}^{(l)}\}_{l=1}^{L}$ according to $\boldsymbol{E}^{(s)}$ ▷ Eq. (9-10)
  end if
end for
$\boldsymbol{\omega}_{\mathrm{E}}^\star \leftarrow \boldsymbol{\omega}_{\mathrm{E}}^{(S)}, \boldsymbol{\omega}_{\mathrm{R}}^\star \leftarrow \boldsymbol{\omega}_{\mathrm{R}}^{(S)}, \boldsymbol{\theta}^\star \leftarrow \boldsymbol{\theta}^{(S)}$
Sample $z_{\mathrm{E}}^\star$ and $z_{\mathrm{R}}^\star$ from $\mathcal{N}(0,\sigma^2)$
$\boldsymbol{E}^t \leftarrow \mathcal{G}_{\mathrm{E}}(\boldsymbol{z}_{\mathrm{E}}^\star, \boldsymbol{\omega}_{\mathrm{E}}^\star)$
for $t = 0$ to T **do**
  Sample $\boldsymbol{\omega}_{\mathrm{R}}^{\star(t)}$ according to $\{\boldsymbol{P}^{(l)}\}_{l=1}^{L}$ ▷ Eq. (8)
  $\boldsymbol{R}^{\star(t)} \leftarrow \mathcal{G}_{\mathrm{R}}(\boldsymbol{z}_{\mathrm{R}}^\star;\boldsymbol{\omega}_{\mathrm{R}}^{\star(t)})$ ▷ Eq. (11)
end for
$\boldsymbol{R}^\star \leftarrow \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{R}^{\star(t)}$ ▷ Eq. (12)
$\widehat{\boldsymbol{E}}^\star \leftarrow \mathcal{F}(\boldsymbol{E}^\star;\boldsymbol{\theta}^\star)$
$\widehat{\boldsymbol{I}} \leftarrow \widehat{\boldsymbol{E}}^\star \odot \boldsymbol{R}^\star$

1) Architecture of Reflectance Generator $\mathcal{G}_{\mathrm{R}}$: The EncoderDecoder [41] is adopted for $\mathcal{G}_{\mathrm{R}}$, which contains five encoder blocks and five decoder blocks with following structures respectively:

$$\text{Encoder: } \mathrm{Conv}_{\downarrow 2} \rightarrow \mathrm{BN} \rightarrow \mathrm{LR} \rightarrow \mathrm{Conv} \rightarrow \mathrm{BN} \rightarrow \mathrm{LR},$$
$$\text{Decoder: } \uparrow 2 \rightarrow \mathrm{Conv} \rightarrow \mathrm{BN} \rightarrow \mathrm{LR} \rightarrow \mathrm{Conv} \rightarrow \mathrm{BN} \rightarrow \mathrm{LR},$$

where $\mathrm{Conv}, \mathrm{Conv}_{\downarrow 2}$ denote convolutional layer with kernel size 3×3 and stride $1, 2$ respectively; and $\mathrm{BN}, \mathrm{LR}, \uparrow 2$ denote batch normalization layer, leaky rectified linear activation function, and bilinear upsampling operation, respectively. Please see more details in [41]. The number of output channels for all convolutional layers is set to 128, which leads to strong expressibility. To perform dropout ensemble for improving noise robustness of the prediction, we configure dropout before the last convolutional layer in all blocks. To handle spatially-varying noise, the dropout probability is varied spatially according to the illumination intensity estimated by $\mathcal{G}_{\mathrm{E}}$.

2) Architecture of Illumination Generator $\mathcal{G}_{\mathrm{E}}$: Considering the illumination is usually assumed to be smoother than the reflectance, we adopt a simple under-parameterized architecture for $\mathcal{G}_{\mathrm{E}}$. It contains five decoder blocks of following structures:

$$\text{Decoder: } \mathrm{Conv} \rightarrow \uparrow 2 \rightarrow \mathrm{ReLU} \rightarrow \mathrm{BN},$$

where ReLU denotes rectified linear activation function. It is very similar to DeepDecoder [19], except that the 1×1

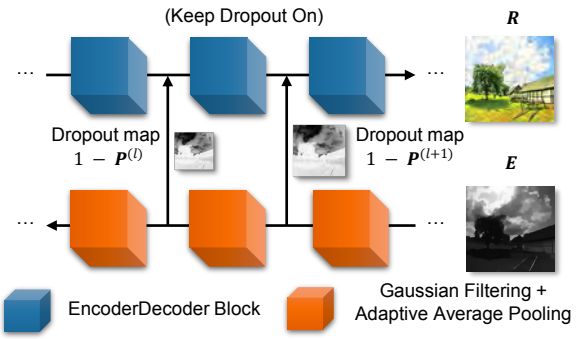


Fig. 3. The illumination guided dropout module in $\mathcal{G}_{\mathrm{R}}$. The dropout probability map $1-\boldsymbol{P}^{(l)}$ for the l-th layer is generated from the illumination map $\boldsymbol{E}$ recursively by (9) and (10).

convolution is replaced with a 3×3 one to pursue more smoothness. The successive bilinear upsampling operations implicitly induce piece-wise smoothness to the illumination layer $\boldsymbol{E}$. The number of output channels for all linear transform layers is set to 16, resulting in significant discrepancy (about $200:1$ parameters) between $\mathcal{G}_{\mathrm{R}}$ and $\mathcal{G}_{\mathrm{E}}$ in terms of model capacity. This helps to reduce the ambiguity between $\boldsymbol{E}$ and $\boldsymbol{R}$. Regarding the other configuration, we use the same setting as that in the experiment of [19]. In practice, we set $\boldsymbol{E}$ consistent on all color channels by duplicating the single-channel output of $\mathcal{G}_{\mathrm{E}}$.

3) Weighted Reconstruction Loss: The loss $\mathcal{L}_{\mathrm{RE}}$ measures the reconstruction error between $\boldsymbol{R} \odot \boldsymbol{E}$ and $\boldsymbol{I}$. Let $\mathbb{P} = \mathbb{Z}_{[1,M]} \times \mathbb{Z}_{[1,N]}$ denote a set of spatial indices. Considering Poisson noise has higher variance on brighter pixels, we use the following weighted ℓ_2 loss:

$$\mathcal{L}_{\mathrm{RE}} := \sum_{c=1}^{C} \sum_{\boldsymbol{p} \in \mathbb{P}} \frac{2}{\max\left(\boldsymbol{R}_{\boldsymbol{p}}^{c} \boldsymbol{E}_{\boldsymbol{p}}^{c} + \boldsymbol{I}_{\boldsymbol{p}}^{c}, \epsilon_r\right)} (\boldsymbol{R}_{\boldsymbol{p}}^{c} \boldsymbol{E}_{\boldsymbol{p}}^{c} - \boldsymbol{I}_{\boldsymbol{p}}^{c})^2, \quad (6)$$

where $\boldsymbol{R}_{\boldsymbol{p}}^{c}, \boldsymbol{E}_{\boldsymbol{p}}^{c}, \boldsymbol{I}_{\boldsymbol{p}}^{c}$ denote the element of $\boldsymbol{R}, \boldsymbol{E}, \boldsymbol{I}$ at $\boldsymbol{p} \in \mathbb{P}$ of the c-th channel, and ϵ_r is a stabilizer set to 10^{-3}. It uses $(\boldsymbol{R}_{\boldsymbol{p}}^{c} \boldsymbol{E}_{\boldsymbol{p}}^{c} + \boldsymbol{I}_{\boldsymbol{p}}^{c})/2$ to estimate noise-less pixels for weighting.

4) Illumination Regularization Loss: While the discrepant NN priors regularize gradient distribution, the loss $\mathcal{L}_{\mathrm{E}}$ aims at regularizing the intensity distribution in $\boldsymbol{E}$. Following the commonly-used white-patch prior that there is a patch with perfect reflectance in the image causing the maximum response across color channels and reflecting the intensity of the illumination [44], [10], [17], [2], we define $\mathcal{L}_{\mathrm{E}}$ by

$$\mathcal{L}_{\mathrm{E}} := \sum_{c=1}^{C} \|(\boldsymbol{E}^{c})\downarrow_{16} - (\widetilde{\boldsymbol{E}})\downarrow_{16}\|_2^2, \quad (7)$$

where $\widetilde{\boldsymbol{E}}_{\boldsymbol{p}} := \max_c \boldsymbol{I}_{\boldsymbol{p}}^{c}$ and $(\cdot)\downarrow_{16}$ denotes average pooling with kernel size 32×32 and stride 16. It is noted that the white-patch prior may not always hold on a fine image scale but can be more accurate for a coarser scale. Thus, we impose it on a coarse scale by via downsampling. We empirically found that the downsampling factor of 16 could lead to higher accuracy than the factors of 2 and 4.

B. Reflectance Denoising via Dropout Ensemble

The estimation of $\boldsymbol{R}$ is sensitive to noise, due to the existence of many zero entries of $\boldsymbol{E}$. Moreover, the over-parameterized nature of $\mathcal{G}_{\mathrm{R}}$ makes it vulnerable to overfitting, *i.e.*, the prediction fits both image and noise. To make the estimation of $\boldsymbol{R}$ robust to noise, we integrate a self-supervised denoising mechanism to our method, which is inspired by the dropout ensemble for dealing with *i.i.d* noise [34]. Briefly, $\mathcal{G}_{\mathrm{R}}$ is optimized with dropout switched on. Then by performing multiple inference of the optimized $\mathcal{G}_{\mathrm{R}}$ with dropout remained switched on, diverse predictions on reflectance with statistical independence are produced, whose average with reduced noise and artifacts is used as a denoised $\boldsymbol{R}$. We observe that the dropout probability is closely related to the denoising strength. In addition, pixels with different SNRs should be processed with different denoising strengths. Thus, the dropout probability of a feature point should be set according to the SNR of the pixels associated to the feature point.

Recall that a point in a feature of a CNN only affects some pixels in the output, which is determined by the receptive field size. Then, the dropout probability of a feature point should be set according to the noise variances of the pixels it can affect. Usually, the SNR is higher in brighter regions for low-light images. Thus, we propose to guide the generation of a spatially-variant dropout probability map with the estimated illumination with the receptive field size.

1) Optimization: Let $\boldsymbol{A}^{(\ell)} \in \mathbb{R}^{M_\ell \times N_\ell}$ denote a feature map output by the ℓ-th convolutional layer after activation. The dropout can be formulated as

$$\widehat{\boldsymbol{A}}^{(\ell)} = \boldsymbol{M}^{(\ell)} \odot \boldsymbol{A}^{(\ell)}, \ \ell = 1, \cdots, L, \quad (8)$$

where $\boldsymbol{M}_{\boldsymbol{q}}^{(\ell)} \sim \mathcal{B}(\boldsymbol{P}_{\boldsymbol{q}}^{(\ell)})$ at location $\boldsymbol{q}$, $\mathcal{B}(p)$ denotes the Bernoulli distribution with success probability of p, and the *keep probability* map $\boldsymbol{P}^{(\ell)}$ is a matrix storing the probabilities of retaining the elements of $\boldsymbol{A}^{(\ell)}$ at different locations during dropout, which is the exact opposite of the *dropout probability* $1-\boldsymbol{P}^{(\ell)}$. We configure dropout before the last convolutional layer in all blocks of $\mathcal{G}_{\mathrm{R}}$.

The proposed *illumination guided dropout* (IGD) module is shown in Fig. 3. We define the guidance map $\boldsymbol{U}$ by the following recursive rule:

$$\boldsymbol{U}^{(L)} = \boldsymbol{E}, \quad \boldsymbol{U}^{(\ell)} = (\boldsymbol{G}_{K_\ell} * \boldsymbol{U}^{(\ell+1)})\downarrow_{M_\ell \times N_\ell} \quad (9)$$

for layer $\ell = L-1, \cdots, 1$, where $*$ denotes convolution, $K_\ell \times K_\ell$ is the size of receptive field of the ℓ-th convolutional layer, $\boldsymbol{G}_{K_\ell}$ is a $K_\ell \times K_\ell$ Gaussian kernel, and $\downarrow_{M_\ell \times N_\ell}$ denotes resizing to $M_\ell \times N_\ell$ by adaptive average pooling. To align the keep probability to $[0.5, 1]$, let

$$\boldsymbol{P}_{\boldsymbol{q}} = (1 + \exp(-\boldsymbol{U}_{\boldsymbol{q}}^{(\ell)}/\tau))^{-1} \in [0.5, 1], \quad (10)$$

where $\tau > 0$. $\boldsymbol{U}$ is initialized at the K-th iteration, then updated every $2K$ iterations during optimization and kept unchanged in later iterations.

2) Inference: In inference, the nodes of the sufficiently optimized $\mathcal{G}_{\mathrm{R}}$ are kept randomly dropped out so that T instances of $\mathcal{G}_{\mathrm{R}}$ are generated to make inferences with certain degree of statistical independence, whose average leads to

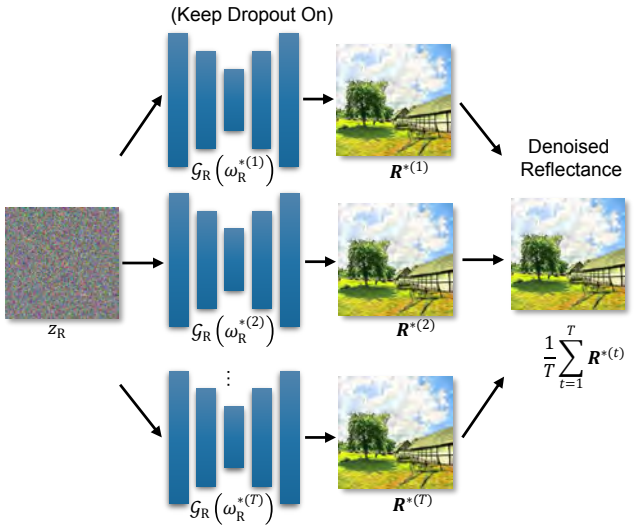
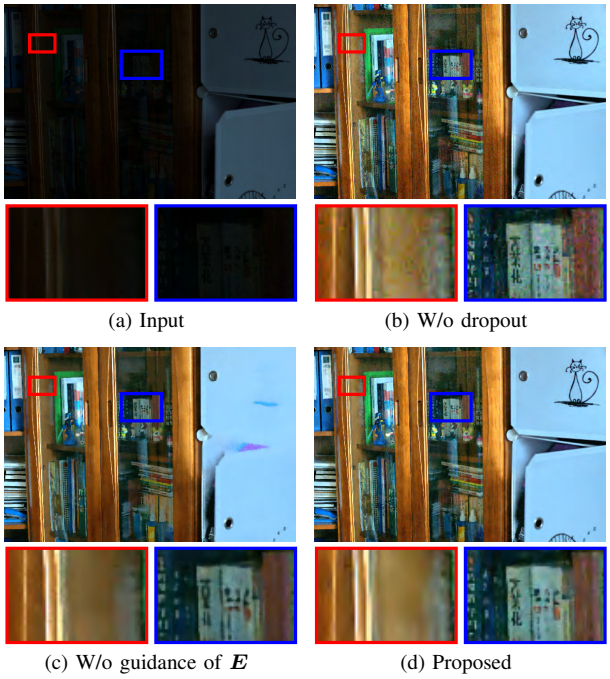


Fig. 4. Dropout ensemble for inference of $\mathcal{G}_R$. Once $\mathcal{G}_R$ is sufficiently optimized, the denoised $\boldsymbol{R}$ is obtained by averaging over multiple $\boldsymbol{R}$s produced by running $\mathcal{G}_R(\boldsymbol{\omega}_R^\star)$ T times with its nodes kept randomly dropped out.



(a) Input (b) W/o dropout

(c) W/o guidance of $\boldsymbol{E}$ (d) Proposed

Fig. 5. Demonstration of the effectiveness of the proposed illumination-guided dropout module.

better denoising result. As shown in Fig. 4, T predictions of reflectance are generated first ($T = 100$ by default),

$$\boldsymbol{R}^{\star(t)} = \mathcal{G}_R(\boldsymbol{z}_R; \boldsymbol{\omega}_R^{\star(t)}), t = 1, ..., T. \tag{11}$$

Then they are averaged to generate the estimated reflectance

$$\boldsymbol{R}^\star = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{R}^{\star(t)}. \tag{12}$$

In Fig. 5, we show the advantage of the proposed module over [34] for handling noise in low-light image.

C. Illumination Adjustment via Histogram Balancing

1) Architecture of $\mathcal{F}$ for Illumination Adjustment: For illumination adjustment, a point-wise mapping $f_\gamma(x) = \max(x, \epsilon_f)^\gamma$ is applied on $\boldsymbol{E}$, where the parameter $\gamma > 0$ controls the curve shape, and ϵ_f is a stabilizer set to 10^{-6}. We employ a lightweight CNN $\mathcal{F}$ to generate the parameter γ of the point-wise mapping function for illumination adjustment on $\boldsymbol{E}$. $\mathcal{F}$ concatenates the given low-light image $\boldsymbol{I}$ and the illumination map $\boldsymbol{E}$ estimated from $\mathcal{G}_E$ as input, which is sequentially passed to four convolutional layers with kernel size 3×3 and channels $8, 8, 8, 4$ respectively, each of which is followed by a max pooling layer with stride 2 and an ReLU activation function. Then the resulting feature maps are sent to two fully-connected layers with number of hidden unit 8 and 1 respectively. The output scalar is passed to a sigmoid activation function to obtain γ.

2) Histogram balance loss: The visual quality of the recovered normal-light image $\widehat{\boldsymbol{I}}$ is measured by $\mathcal{L}_I$ in terms of contrast. Inspired by histogram equalization [24], we define *histogram balance loss* $\mathcal{L}_I$ using the entropy of a soft histogram on $\widehat{\boldsymbol{I}}$. Let $\boldsymbol{H}_1, \cdots, \boldsymbol{H}_J \in \mathbb{R}$ denote the histogram bins and δ denote the bin size. Let $\mathcal{S} : \mathbb{R} \to [0, 1]$ denote the sigmoid function. The soft histogram $\boldsymbol{h} \in \mathbb{R}^J$ is defined by

$$\boldsymbol{h}_j = \sum_{\boldsymbol{p}\in\mathbb{P}} \mathcal{S}(\bar{\boldsymbol{I}}_{\boldsymbol{p}} - \boldsymbol{H}_j + \frac{\delta}{2}) - \mathcal{S}(\bar{\boldsymbol{I}}_{\boldsymbol{p}} - \boldsymbol{H}_j - \frac{\delta}{2}), \forall j, \tag{13}$$

where $\bar{\boldsymbol{I}} \in \mathbb{Z}_{[0,1]}^{M\times N}$ denotes the mean of $\widehat{\boldsymbol{I}}$ along the channel dimension. Then $\mathcal{L}_I$ is given by

$$\mathcal{L}_I := -\sum_j \boldsymbol{h}_j \log \boldsymbol{h}_j. \tag{14}$$

Minimizing $\mathcal{L}_I$ indeed equalizes the histogram. In practice, we simply set $\delta = 1/256$, $J = 256$ and $\boldsymbol{H}_j = (j-1)/255$.

IV. Experiments

A. Benchmark Datasets and Experimental Details

Five benchmark datasets covering a wide range of lighting conditions are used for evaluation: *(i)* LOL [47] contains 15 low/normal-light image pairs of size 400×600 captured in *real* scenes. *(ii)* LIME [17] contains 10 low-light images. *(iii)* NPE [44] contains 8 outdoor natural scene images. *(iv)* MEF [29] contains 17 high-quality image sequences including natural scenarios, indoor and outdoor views, and man-made architectures. Each image sequence has several multi-exposure images, and we select one of poor-exposed images as input to perform evaluation. *(vi)* DICM [24] contains 69 captured images from commercial digital cameras. It is noted that LOL is collected under extreme low-light conditions with normal-light references provided, while the other four are collected under moderate low-light conditions without ground truths.

The scores of three no-reference metrics that are widely used in the studies of LIE [11], [2], [26], [45], [37] are reported: *(i)* Natural Image Quality Evaluator (NIQE) [32] does not relate to subjective quality scores and can measure the image quality with arbitrary distortion. *(ii)* AutoRegressive-based Image Sharpness Metric (ARISM) [14] estimates the

TABLE I. PSNR/SSIM/NIQE/ARISM/NIQMC scores on five datasets. The best and second best are highlighted in red and yellow respectively. The overall 'Rank' is calculated by averaging the per-dataset average rankings among different methods, while 'RoR' denotes the rank of the overall 'Rank'. ↑ (↓) means higher (lower) is better.

Setting	Method	LOL					LIME			NPE			MEF			DICM			Rank	RoR
		P↑	S↑	N↓	A↓	C↑	N↓	A↓	C↑	N↓	A↓	C↑	N↓	A↓	C↑	N↓	A↓	C↑		
Non-learning	MSR	13.17	0.479	8.114	4.194	5.350	3.764	3.621	5.464	4.366	3.047	5.328	3.309	3.859	**5.489**	3.677	3.436	**5.480**	12.8	12
	Dong *et al.*	16.72	0.582	8.316	4.036	4.525	4.052	3.197	4.885	4.126	2.951	5.366	4.099	3.470	5.115	4.119	3.229	5.056	13.7	17
	NPE	16.97	0.589	8.439	4.058	4.692	3.905	3.217	4.617	3.952	2.992	5.174	3.529	3.530	4.861	3.760	3.211	5.034	14.4	19
	PIE	12.28	0.515	7.506	3.958	3.532	4.050	3.018	4.592	4.137	2.941	5.148	3.451	3.146	4.803	3.978	3.071	5.008	14.2	18
	SRIE	11.86	0.498	7.287	3.967	3.489	3.786	3.115	4.503	3.979	2.923	5.185	3.474	3.304	4.704	3.899	3.161	4.985	15.1	21
	MF	16.97	0.605	8.877	3.977	4.502	4.067	3.118	4.869	4.105	2.944	5.284	3.492	3.270	5.042	3.844	3.141	5.115	11.8	8
	BIMEF	13.88	0.577	7.515	3.908	3.711	3.860	3.103	4.721	4.133	2.959	5.227	3.329	3.236	4.879	3.846	3.144	5.047	13.0	15
	JIEP	12.05	0.512	6.872	3.985	3.527	3.719	3.049	4.545	4.226	2.920	5.207	3.391	3.195	4.772	3.569	2.813	4.940	12.9	14
	LIME	16.76	0.564	8.378	4.063	**5.487**	4.155	3.292	**5.496**	4.263	3.056	**5.448**	3.702	3.523	5.417	3.859	3.248	5.295	13.1	16
	NPIE	16.70	0.594	8.159	4.042	4.557	**3.579**	3.322	4.890	4.025	2.947	5.202	3.337	3.548	5.188	3.736	3.210	5.098	11.5	6
	RRM	13.88	0.658	5.810	3.465	3.318	4.643	2.903	4.816	4.845	2.989	5.159	5.062	2.862	4.914	4.597	2.876	4.924	15.1	22
	STAR	12.64	0.538	6.205	3.720	3.651	3.684	3.045	4.580	4.052	2.950	5.184	3.296	3.171	4.824	**3.512**	3.075	4.902	12.1	11
	LR3M	10.22	0.399	7.522	**2.663**	2.143	5.180	**2.716**	4.742	4.641	2.845	5.134	5.508	2.705	4.806	4.568	2.845	4.975	15.0	20
Dataset w/ paired data	RetinexNet	16.77	0.559	8.879	3.911	4.225	4.598	3.458	4.697	4.567	2.968	4.967	4.410	3.551	4.747	4.415	3.204	4.763	19.4	24
	DeepUPE	11.04	0.412	7.366	3.821	3.477	3.959	3.035	4.894	3.994	2.943	5.221	3.527	3.053	4.989	3.884	2.988	5.136	11.0	5
	KinD	17.65	**0.760**	4.710	3.050	4.504	4.763	2.781	4.942	4.161	2.886	5.207	3.877	2.759	5.093	4.151	2.762	5.009	9.3	4
	HybridNet	16.60	0.668	**3.391**	3.037	3.983	4.801	2.767	4.491	**3.728**	2.761	4.931	4.000	2.762	4.711	4.097	2.724	4.776	12.0	9
	FIDE	**19.41**	0.732	4.366	2.745	4.362	4.424	2.775	4.628	4.747	2.772	4.926	4.220	2.759	4.578	4.155	2.736	4.709	12.8	13
	DRBN	16.75	0.659	4.724	2.720	4.583	4.433	2.717	4.844	4.211	**2.739**	5.344	4.253	**2.689**	5.077	4.333	**2.701**	5.063	8.5	3
Dataset w/o paired data	EnGAN	17.49	0.658	4.684	3.445	4.682	3.657	3.035	5.082	4.125	2.914	5.153	3.221	3.068	5.115	3.562	3.007	5.095	6.8	2
	ZeroDCE	14.86	0.585	7.767	3.964	4.015	3.769	3.150	4.839	4.275	2.889	5.166	3.283	3.310	4.944	3.560	2.814	4.934	12.0	10
	RUAS	16.40	0.583	6.340	3.635	5.009	4.262	3.091	4.965	5.512	3.003	4.717	3.830	3.186	4.734	5.115	3.125	4.432	16.3	23
Dataset-free	RetinexDIP	9.442	0.322	7.070	3.838	2.836	3.674	3.115	4.766	4.096	2.932	5.347	3.288	3.287	4.979	3.719	3.103	5.020	11.7	7
	Ours	15.49	0.654	3.731	3.487	5.476	4.123	2.949	5.453	3.973	2.888	5.417	**3.070**	3.091	5.486	3.649	3.003	5.388	**5.6**	**1**

sharpness/blurriness of images based on analysis in the autoregressive (AR) parameter space. *(iii)* No-reference Image Quality Metric for Contrast distortion (NIQMC) [13] utilize the concept of information maximization to access the image quality of contrast-changed images. On LOL, we also report the full-reference metrics Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) Index [46], although they are not appropriate for unsupervised methods without the notion of reference. It is known that different metrics assess different aspects of image quality. Following [37], we use non-parametric metrics for a more comprehensive and fair comparison, average ranking (denoted by 'Rank') as well as the rank of rank (denoted by 'RoR'). Specifically, average rankings of different metrics on each dataset are computed first and then averaged to obtain the overall rank, and RoR is the rank of the overall Rank.

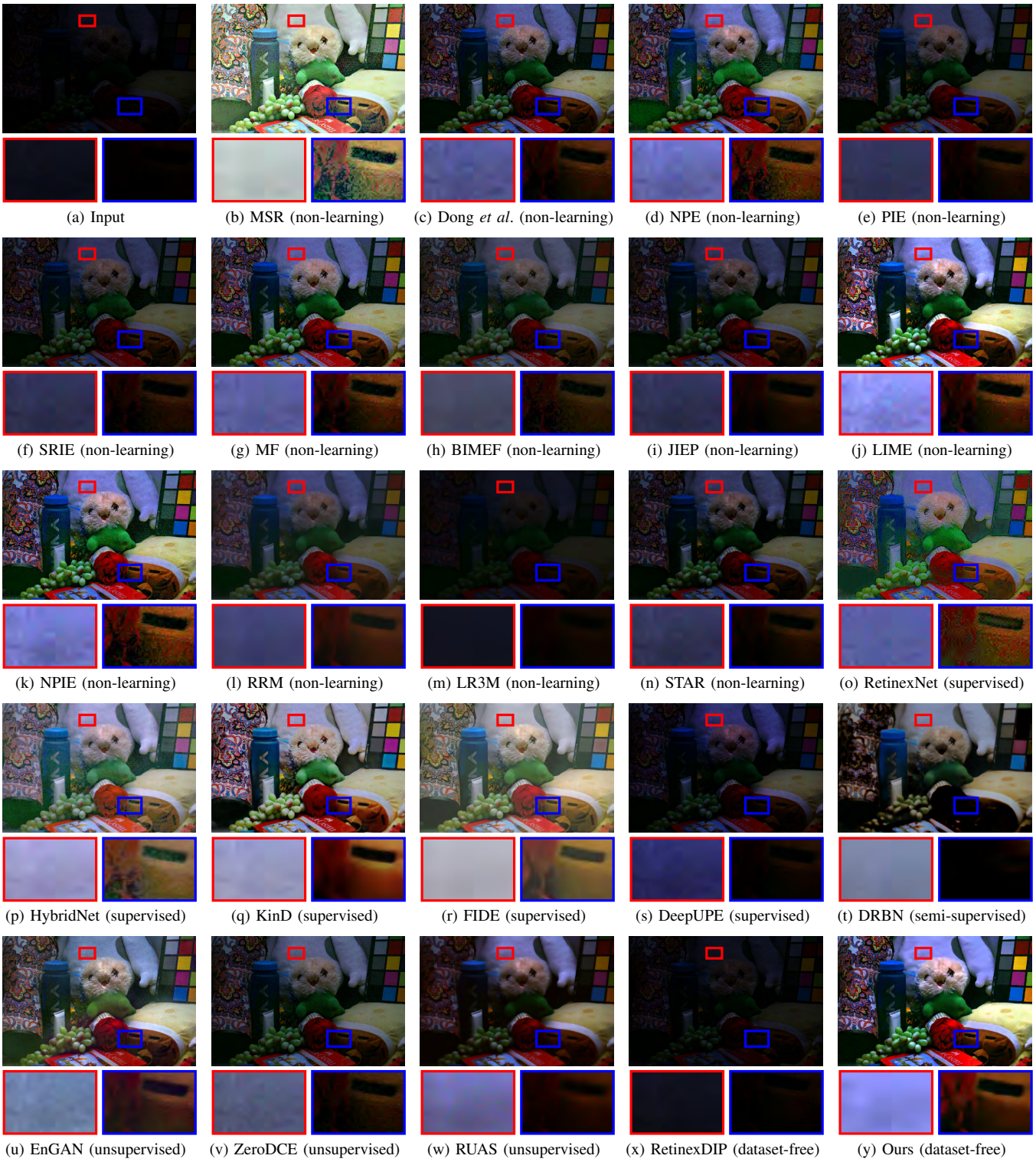
We implemented the proposed approach using PyTorch and set its hyper-parameters as follows: $\lambda_{\mathrm{E}} = 0.01, \lambda_{\mathrm{I}} = 10^{-5}$ through all datasets, and $\tau = 0.1$ for LOL and $\tau = 0.01$ for other four datasets. We use Adam with a fixed learning rate of 10^{-3} for optimization. The optimization is stopped after 1×10^5 iterations. The weights for all convolutional and fully-connected layers are initialized by [18] and all biases are initialized to 0.

B. Comparison with State-of-the-Art Methods

Twenty-three methods including the state-of-the-art ones are selected for comparison, including *(i) non-learning methods*: MSR [22], Dong *et al.* [6], NPE [44], PIE [9], SRIE [11], MF [10], BIMEF [53], JIEP [2], LIME [17], NPIE [43], RRM [26], STAR [49] and LR3M [37]; *(ii) models trained on paired data*: RetinexNet [47], DeepUPE [42], KinD [56], HybridNet [36], FIDE [50], and DRBN [52]; *(iii) models trained on an unorganized dataset*: EnlightenGAN (EnGAN for short) [21], ZeroDCE [16], and RUAS [27]; and *(iv) model trained without training data*: RetinexDIP [57]. The results of these methods are produced by their released codes with recommended parameter setting.

The results of different methods on the five datasets are summarized in Table I. As shown in the table, the rankings of different IQA metrics for a method can vary a lot, *e.g.*, LR3M performs quite well in terms of ARISM, but almost the worst for NIQMC; LIME performs quite well in terms of NIQMC, but almost the worst for NIQE. RetinexNet performs well on a few metrics (*e.g.*, PSNR) but much worse on others. In comparison, our method achieved the best rank. In particular, it achieved very competitive NIQE and NIQMC on the LOL dataset. This is impressive as our model never saw other images including normal-light and low-light ones. Noted that it is very challenging for an untrained-NN-based method to perform fully better than supervised methods in terms of all metrics and all test images. Our aim is to develop an untrained-NN-based method that performs on a par with the supervised methods so as to address the difficulty in the case of limited training data. Indeed, it is also very difficult to have a supervised method with the best performance across all images in all metrics. A good LIE method should win the trade-off among different aspects. As shown in the experimental results, the proposed method can achieve a very good balance among different aspects of image quality, while being competitive to state-of-the-art supervised methods.

The visual comparison results on LOL are shown in Fig. 6 and Fig. 7. The images of LOL are taken under extreme low-light conditions, with very low SNR and severe noises. Most



(a) Input (b) MSR (non-learning) (c) Dong *et al*. (non-learning) (d) NPE (non-learning) (e) PIE (non-learning)

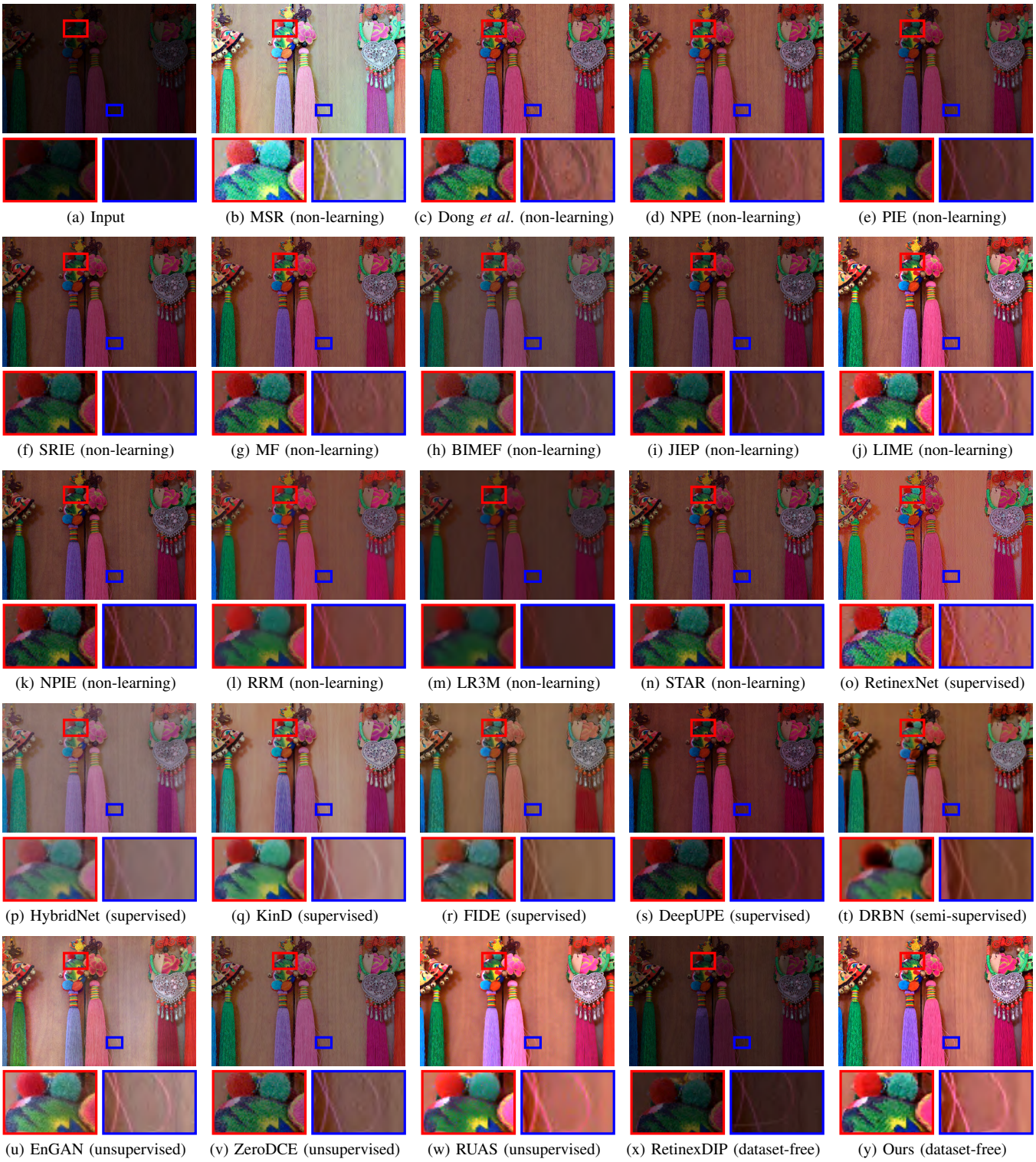
(f) SRIE (non-learning) (g) MF (non-learning) (h) BIMEF (non-learning) (i) JIEP (non-learning) (j) LIME (non-learning)

(k) NPIE (non-learning) (l) RRM (non-learning) (m) LR3M (non-learning) (n) STAR (non-learning) (o) RetinexNet (supervised)

(p) HybridNet (supervised) (q) KinD (supervised) (r) FIDE (supervised) (s) DeepUPE (supervised) (t) DRBN (semi-supervised)

(u) EnGAN (unsupervised) (v) ZeroDCE (unsupervised) (w) RUAS (unsupervised) (x) RetinexDIP (dataset-free) (y) Ours (dataset-free)

Fig. 6. Visual quality comparison of enhancement results. LIME and KinD can restore vivid color; however, it reveals significant noise (see *e.g*., red boxes). KinD produces over-smoothed results in some regions (see *e.g*., the ball of red wool in the middle), and the color of its result is less vivid in comparison to that of the proposed method and LIME. In contrast, the proposed method can enhance low-SNR regions while preserving color, even though it never sees any well/over-exposed images.



(a) Input (b) MSR (non-learning) (c) Dong *et al*. (non-learning) (d) NPE (non-learning) (e) PIE (non-learning)

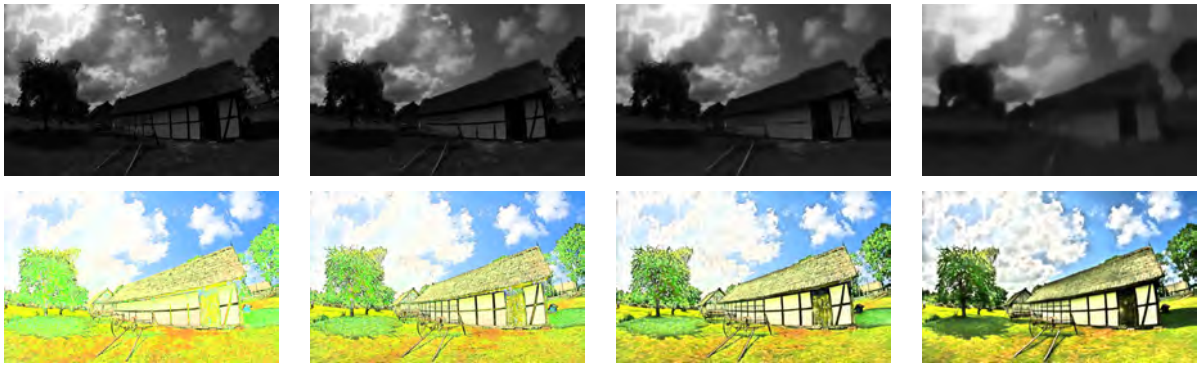
(f) SRIE (non-learning) (g) MF (non-learning) (h) BIMEF (non-learning) (i) JIEP (non-learning) (j) LIME (non-learning)

(k) NPIE (non-learning) (l) RRM (non-learning) (m) LR3M (non-learning) (n) STAR (non-learning) (o) RetinexNet (supervised)

(p) HybridNet (supervised) (q) KinD (supervised) (r) FIDE (supervised) (s) DeepUPE (supervised) (t) DRBN (semi-supervised)

(u) EnGAN (unsupervised) (v) ZeroDCE (unsupervised) (w) RUAS (unsupervised) (x) RetinexDIP (dataset-free) (y) Ours (dataset-free)

Fig. 7. Visual quality comparison of enhancement results on a low-light image. The global illumination can be well restored by LIME, RUAS and the proposed method. However, LIME and RUAS reveals significant noise. Although HybridNet, KinD can suppress noise while restoring color, they tend to over-smooth the input image, which leads to loss of rich details. The proposed method is able to preserve vivid color as well as fine details.



(a) $A(\mathcal{G}_{\mathrm{E}}) = A(\mathcal{G}_{\mathrm{R}})$ (b) $N_{\mathrm{c}}(\mathcal{G}_{\mathrm{E}}) = 64$ (c) $N_{\mathrm{c}}(\mathcal{G}_{\mathrm{E}}) = 32$ (d) $N_{\mathrm{c}}(\mathcal{G}_{\mathrm{E}}) = 16$

Fig. 8. Ablation study on the discrepant priors. Top: Illumination; Bottom: Reflectance. The input low-light image is shown in Fig. 9 (a). $A(\cdot)$ denotes the architecture of the NN, while $N_{\mathrm{c}}(\cdot)$ denotes the number of output channel for each convolutional layers. It can be seen that less discrepancy between $\mathcal{G}_{\mathrm{E}}$ and $\mathcal{G}_{\mathrm{R}}$, more edges are wrongly assigned to the illumination map.

(a) Input (b) W/o $\mathcal{L}_{\mathrm{I}}$ (c) W/o $\mathcal{L}_{\mathrm{E}}$ (d) Ours

Fig. 9. Ablation study on the regularization loss. (b)-(d): Top: Illumination; Bottom: Output. As shown in (b), using a fixed γ instead of an adaptive one guided by $\mathcal{L}_{\mathrm{I}}$ produces unsatisfactory result and contrast. As shown in (c), without the intensity distribution regularization $\mathcal{L}_{\mathrm{E}}$, $\mathcal{G}_{\mathrm{E}}$ outputs a trivial constant illumination map.

compared methods either fail to remove noise or to recover high contrast. For instance, RetinexNet produces results with vivid color and pleasing appearances in most cases; however, it also produces severe artifacts and cartoonish effects in the results and cannot well handle the severe noise. This is the main reason why its performance rank is not very high. In comparison, the proposed method can enhance low-SNR regions while preserving color, even though it never sees any well/over-exposed images. Such good performance is attributed to the accurate Retinex decomposition by untrained NN priors, while the noise robustness is mainly attributed to the spatially-variant dropout ensemble for handling noise.

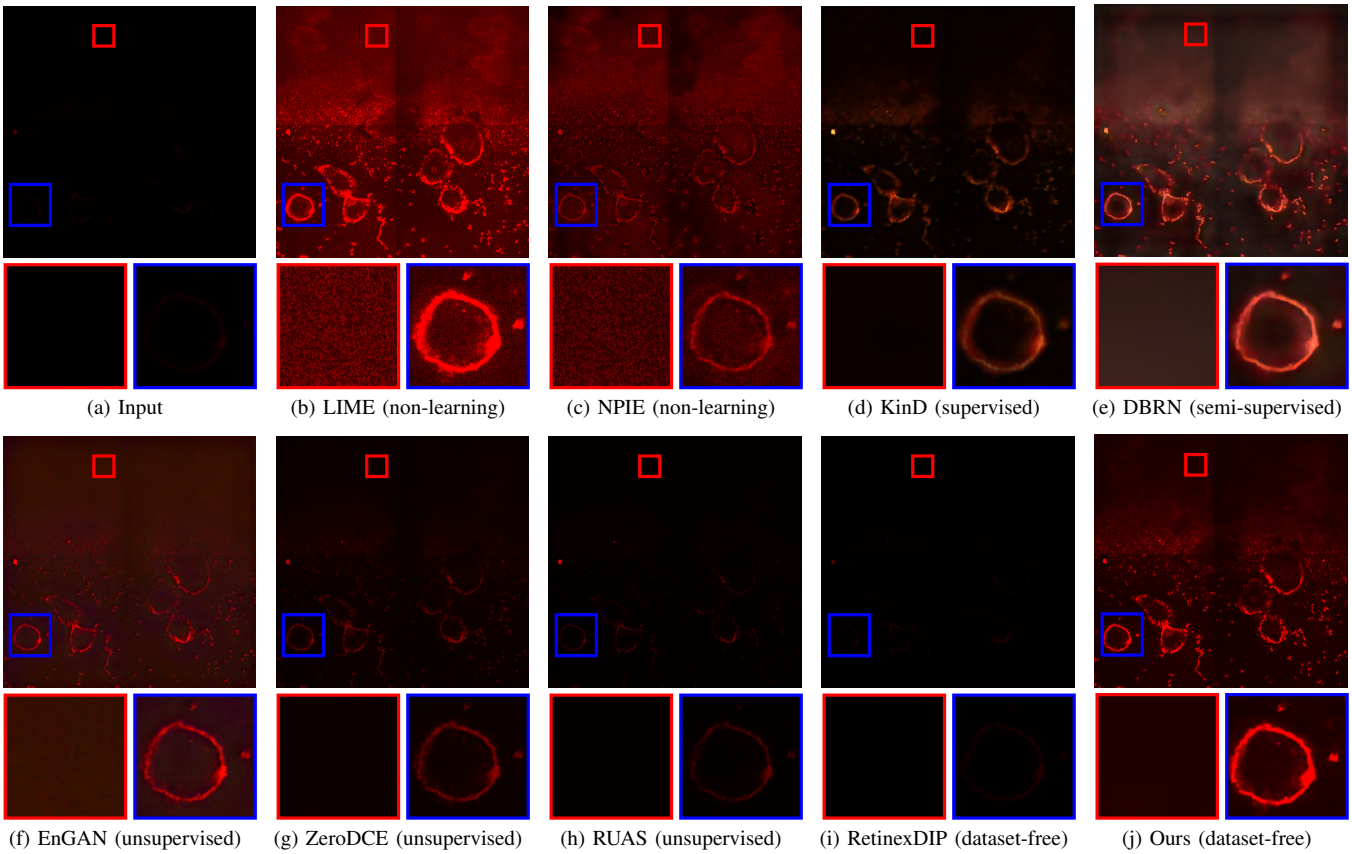
TABLE II. PSNR/SSIM/NIQE/ARISM/NIQMC scores for ablation studies on the LOL dataset. The best and second best are highlighted in red and yellow respectively. ↑ (↓) means higher (lower) is better.

Configuration	P↑	S↑	N↓	A↓	C↑	Rank
Proposed	15.49	0.654	**3.731**	3.487	5.476	**2.2**
$N_{\mathrm{c}}(\mathcal{G}_{\mathrm{E}}) = 32$	14.33	0.617	3.833	3.409	5.376	3.0
$N_{\mathrm{c}}(\mathcal{G}_{\mathrm{E}}) = 64$	12.99	0.569	4.338	3.542	4.823	5.8
$A(\mathcal{G}_{\mathrm{E}}) = A(\mathcal{G}_{\mathrm{R}})$	14.16	0.582	4.306	3.643	3.525	5.4
w/o dropout	**15.77**	0.579	3.837	4.031	**5.594**	4.4
w/o IGD	15.53	**0.661**	4.580	3.391	5.387	3.0
w/o $\mathcal{L}_{\mathrm{E}}$	13.79	0.515	4.023	3.716	4.059	6.2
w/o $\mathcal{L}_{\mathrm{I}}$	11.61	0.503	4.187	**3.310**	3.513	6.0

C. Ablation Study

1) Effectiveness of the Discrepant NN Priors: In the proposed method, discrepant network architectures and model capacity are designed to resolve the ambiguity between the two layers. To examine its effectiveness, we set the architecture $A(\mathcal{G}_{\mathrm{E}})$ of $\mathcal{G}_{\mathrm{E}}$ the same as $\mathcal{G}_{\mathrm{R}}$ except the dropout module. See Table II for result of such a variant. Its performance drops noticeably in terms of SSIM, NIQE and NIQMC. As shown in Fig. 8 (a), textures and fine details are more likely to be wrongly assigned to the illumination map in such a configuration. Furthermore, we compare our default setting with increased number of output channels $N_{\mathrm{c}}(\mathcal{G}_{\mathrm{E}})$ of $\mathcal{G}_{\mathrm{E}}$, which eliminates the discrepancy between $\mathcal{G}_{\mathrm{E}}$ and $\mathcal{G}_{\mathrm{R}}$. The performance also drops for these setting as shown in Table II. Less discrepancy between $\mathcal{G}_{\mathrm{E}}$ and $\mathcal{G}_{\mathrm{R}}$, more edges are wrongly assigned to the illumination map. This validates the design of discrepant NN architecture in the proposed method. See also Fig. 8 (b)-(d) for the visual comparison.

2) Effectiveness of the IGD Module: We evaluate the effectiveness of the proposed IGD module by comparing it with two variants: *(i) w/o dropout:* removing dropout in optimization and inference, *i.e.* no dropout ensemble is used; and *(ii) w/o*



(a) Input (b) LIME (non-learning) (c) NPIE (non-learning) (d) KinD (supervised) (e) DBRN (semi-supervised)

(f) EnGAN (unsupervised) (g) ZeroDCE (unsupervised) (h) RUAS (unsupervised) (i) RetinexDIP (dataset-free) (j) Ours (dataset-free)

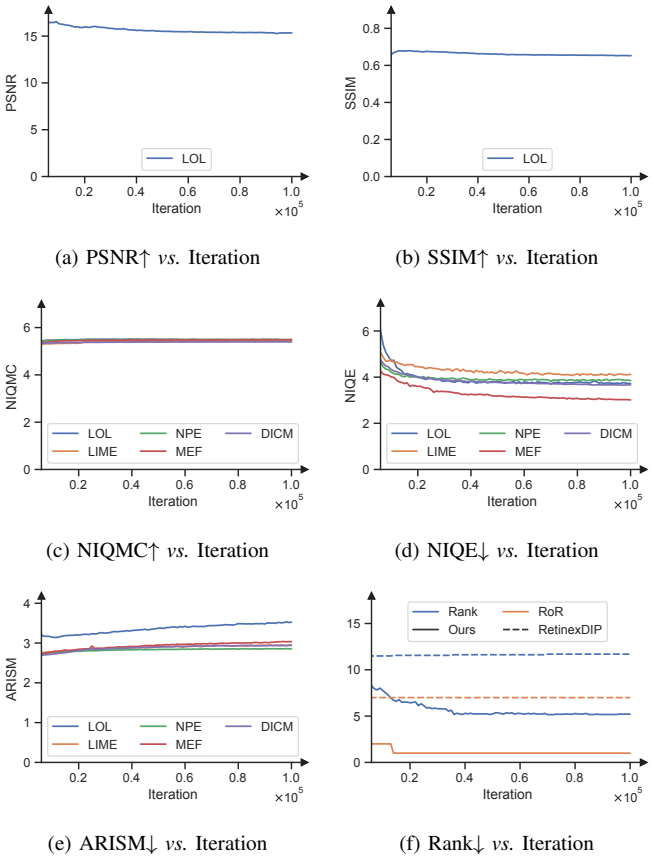
Fig. 10. Visual comparison of enhancement results on a fluorescent image. It can be seen that the performance of ZeroDCE, RUAS, and RetinexDIP is relatively poor. Although NPIE and EnGAN improve the contrast, they also greatly amplify the noise. KinD and DBRN can improve contrast while suppressing noise, but they also blur out some important details. In contrast, the proposed method can improve the contrast while preserving details during denoising.

IGD: fixing dropout probabilities according to [34], instead of using the illumination guidance. See Table II for the results on LOL. It can be seen that dropout ensemble is very helpful for the performance improvement in terms of SSIM and ARISM. However, it results in a high NIQE, which can be improved by the proposed IGD module. For further analysis, a visual comparison is given in Fig. 5. It can be seen that without dropout ensemble, the generator for reflectance is prone to overfitting and revealing significant noise in low-SNR regions, *e.g.*, that in the red box. Using the dropout ensemble with fixed setting across different regions, the generator produces cleaner results. However, since it processes regions of different SNRs uniformly, when noise is well removed in low-SNR regions (see *e.g.*, red box), the high-SNR regions may be over-smoothed (see *e.g.*, blue box). In contrast, with spatially-varying dropout probabilities in the IGD module, our method can handle regions with different SNR better.

3) Effectiveness of the Regularization Loss: We examine the effectiveness of $\mathcal{L}_{\mathrm{E}}$ by removing it from the loss function. This configuration is denoted by 'w/o $\mathcal{L}_{\mathrm{E}}$' and its result on LOL is listed in Table II. The performance drops noticeably without the use of regularization loss on the illumination. This is because only the discrepancy between $\mathcal{G}_{\mathrm{E}}$ and $\mathcal{G}_{\mathrm{R}}$ is not sufficient to address the ambiguity in terms of the intensity distribution. We also validate the effectiveness of $\mathcal{L}_{\mathrm{I}}$ for adaptive illumination adjustment by comparing it to gamma correction (a common practice) with $\gamma = 2.2$, which is denoted by 'w/o $\mathcal{L}_{\mathrm{I}}$'. See also Fig. 9 for the visual comparison.

D. Analysis on Computational Complexity and Iterations

While the dropout ensemble brings some additional time cost, the main computational burden in the the proposed method is caused by the large iteration number, which is used for achieving as high performance as possible for the case where image quality is main focus. Fortunately, it can be largely reduced for the case where processing time is taken into consideration. Please see Table III for the performance rank and computational complexity of the proposed method with different iteration numbers. The computational complexity is measured by the average running time per image calculated on 100 test images with a size 600×400 from the LOL dataset. The performance is measured by the aforementioned metrics, the overall Rank as well as the rank of rank (RoR). It can be seen that the proposed method outperformed RetinexDIP after 1,000 iterations with acceptable additional time cost (2.4 times the running time of RetinexDIP). In addition, it already achieved the second best and the best ranking after 10,000 and 14,000 iterations, respectively, which are much less than 100,000 iterations. In such cases, the running time of the proposed method is less than 12 minutes per image.



(a) PSNR↑ *vs.* Iteration (b) SSIM↑ *vs.* Iteration

(c) NIQMC↑ *vs.* Iteration (d) NIQE↓ *vs.* Iteration

(e) ARISM↓ *vs.* Iteration (f) Rank↓ *vs.* Iteration

Fig. 11. Performance-iteration curves of the proposed method.

The performance-iteration curves of the proposed method are shown in Fig. 11. Such curves are consistent with analysis above. That is, using $10,000$ iterations rather than $100,000$ iterations can have a very competitive performance, and using around $1,000$ iterations can result in better performance than RetinexDIP (see Fig. 11 (f)). In addition, it is worth mentioning an interesting phenomenon in the figure. That is, unlike other DIP or DIP-based methods whose performance will decrease noticeably without early stopping, the proposed method *does not suffer from such an overfitting issue* and it performs stably after a long time of learning.

Same as other untrained-NN-prior based image recovery methods, the online manner of the proposed approach makes it slower than a pre-trained NN model. However, owing to its training-data-free nature, it is of value to applications where sufficient unbiased training data are hard to collect. Indeed, such a method is not against the dataset-based learning methods for LIE, but provides a complement to address the cases where training data is insufficient or of low quality for dataset-based learning. It can also inspire further studies on exploiting untrained NN priors for solving other bi-linear inverse problems involving noise in low-level vision. In future work, we will investigate the combination of NN trained on dataset and the proposed untrained NN prior to build up an efficient and unbiased learning system.

TABLE III. Comparison of running time, which is calculated on 100 test images with size 600×400. Our method is ranked with different iteration numbers respectively.

Method	Time (seconds)	Rank	RoR
RetinexNet	0.119	19.4	24
KinD	0.181	9.3	4
FIDE	0.594	12.8	13
DBRN	0.053	8.5	3
EnGAN	0.010	6.8	2
ZeroDCE	0.003	12.0	10
RUAS	0.016	16.3	23
RetinexDIP	21.288	11.7	7
Ours ($100,000$ iterations)	5926.836	5.6	1
Ours ($1,000$ iterations)	53.265	10.2	4
Ours ($10,000$ iterations)	506.627	7.7	2
Ours ($14,000$ iterations)	665.728	6.7	1

E. An Example of Application to Scientific Imaging

We provide an example of application to scientific imaging in Fig. 10, which shows a fluorescent image of tumor cells with low visibility and the corresponding image enhancement results produced by different methods. We show the results of two non-learning method (b) LIME and (c) NPIE, two supervised method (d) KinD, a semi-supervised method (e) DBRN, two dataset-based unsupervised methods (f) EnGAN and (g) ZeroDCE, and (h) RUAS, an existing dataset-free method (i) RetinexDIP and (j) the proposed dataset-free method. Thanks to the training-data-free nature, the proposed method exhibits strong generalization performance, which is of value to applications where sufficient unbiased training data are hard to obtain, such as clinical diagnosis.

V. Conclusion

In this paper, we successfully exploited untrained NN priors for enhancing low-light images with noise. This led to an effective unsupervised approach that provided state-of-the-art performance while requiring no data for training. The key ingredients of our method include the discrepant NN architectures and capacity for addressing layer ambiguity, the illumination-guided self-supervised denoising scheme for handling noise with spatially-varying variance, and the joint optimization of NN-based Retinex decomposition and illumination adjustment. The effectiveness of the proposed approach has been demonstrated by extensive experiments with diverse lighting conditions, particularly extreme low-light conditions.

References

[1] S. Abbasi-Sureshjani, R. Raumanns, B. E. Michels, G. Schouten, and V. Cheplygina. Risk of Training Diagnostic Algorithms on Data with Demographic Bias. In *Interpretable and Annotation-Efficient Learning for Medical Image Computing*, pages 183–192, 2020.
[2] B. Cai, X. Xu, K. Guo, K. Jia, B. Hu, and D. Tao. A Joint Intrinsic-Extrinsic Prior Model for Retinex. In *Proc. ICCV*, 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] C. Chen, Q. Chen, J. Xu, and V. Koltun. Learning to See in the Dark. In *Proc. CVPR*, 2018.
[5] S. K. Dhara and D. Sen. Exposedness based Noise-Suppressing Low-Light Image Enhancement. *IEEE Trans. Circuits Syst. Video Technol.*, pages 1–1, 2021.
[6] X. Dong, Y. A. Pang, and J. G. Wen. Fast Efficient Algorithm for Enhancement of Low Lighting Video. In *Proc. ACM SIGGRAPH*, 2010.

[7] P. Drozdowski, C. Rathgeb, A. Dantcheva, N. Damer, and C. Busch. Demographic Bias in Biometrics: A Survey on an Emerging Challenge. *IEEE Trans. Tech. Soc.*, 1(2):89–103, June 2020.

[8] M. Elad. Retinex by two bilateral filters. In *Scale-Space*, 2005.

[9] X. Fu, Y. Liao, D. Zeng, Y. Huang, X.-P. 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. CVPR*, 2016.

[12] Y. Gandelsman, A. Shocher, and M. Irani. "Double-DIP": Unsupervised Image Decomposition via Coupled Deep-Image-Priors. In *Proc. CVPR*, 2019.

[13] K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, and C. W. Chen. No-Reference Quality Metric of Contrast-Distorted Images Based on Information Maximization. *IEEE Trans. Cybern.*, 47(12):4559–4565, Dec. 2017.

[14] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang. No-Reference Image Sharpness Assessment in Autoregressive Parameter Space. *IEEE Trans. Image Process.*, 24(10):3218–3231, Oct. 2015.

[15] 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.*, 29:3239–3253, 2020.

[16] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In *Proc. CVPR*, 2020.

[17] 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.

[18] K. He, X. Zhang, S. Ren, and J. Sun. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In *Proc. ICCV*, 2015.

[19] R. Heckel and P. Hand. Deep Decoder: Concise Image Representations from Untrained Non-convolutional Networks. In *Proc. ICLR*, 2018.

[20] G. Jagatap and C. Hegde. Algorithmic Guarantees for Inverse Imaging with Untrained Network Priors. In *Proc. NeurIPS*, 2019.

[21] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang. EnlightenGAN: Deep Light Enhancement Without Paired Supervision. *IEEE Trans. Image Process.*, 30:2340–2349, 2021.

[22] D. Jobson, Z. Rahman, and G. 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.

[23] 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.

[24] 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.

[25] J. Li, X. Feng, and Z. Hua. Low-Light Image Enhancement via Progressive-Recursive Network. *IEEE Trans. Circuits Syst. Video Technol.*, 31(11):4227–4240, Nov. 2021.

[26] 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.

[27] R. Liu, L. Ma, J. Zhang, X. Fan, and Z. Luo. Retinex-Inspired Unrolling With Cooperative Prior Architecture Search for Low-Light Image Enhancement. In *Proc. CVPR*, 2021.

[28] 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.

[29] 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.

[30] 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. CVPR*, 2011.

[31] W. Ma and S. Osher. A TV Bregman iterative model of Retinex theory. *Inverse Problems & Imaging*, 6(4):697, 2012.

[32] 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.

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

[34] Y. Quan, M. Chen, T. Pang, and H. Ji. Self2Self With Dropout: Learning Self-Supervised Denoising From Single Image. In *Proc. CVPR*, 2020.

[35] D. Ren, K. Zhang, Q. Wang, Q. Hu, and W. Zuo. Neural Blind Deconvolution Using Deep Priors. In *Proc. CVPR*, 2020.

[36] W. Ren, S. Liu, L. Ma, Q. Xu, X. Xu, X. Cao, J. Du, and M.-H. Yang. Low-Light Image Enhancement via a Deep Hybrid Network. *IEEE Trans. Image Process.*, 28(9):4364–4375, Sept. 2019.

[37] X. Ren, W. Yang, W.-H. Cheng, and J. Liu. LR3M: Robust Low-Light Enhancement via Low-Rank Regularized Retinex Model. *IEEE Trans. Image Process.*, 29:5862–5876, 2020.

[38] Y. Ren, Z. Ying, T. H. Li, and G. Li. LECARM: Low-Light Image Enhancement Using the Camera Response Model. *IEEE Trans. Circuits Syst. Video Technol.*, 29(4):968–981, Apr. 2019.

[39] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma. MSR-net:Low-light Image Enhancement Using Deep Convolutional Network. *arXiv:1711.02488 [cs]*, Nov. 2017.

[40] M. Tang, F. Xie, R. Zhang, Z. Jiang, and A. C. Bovik. A Local Flatness Based Variational Approach to Retinex. *IEEE Trans. Image Process.*, 29:7217–7232, 2020.

[41] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep Image Prior. In *Proc. CVPR*, 2018.

[42] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia. Underexposed Photo Enhancement using Deep Illumination Estimation. In *Proc. CVPR*, 2019.

[43] 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.

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

[45] Y.-F. Wang, H.-M. Liu, and Z.-W. Fu. Low-Light Image Enhancement via the Absorption Light Scattering Model. *IEEE Trans. Image Process.*, 28(11):5679–5690, Nov. 2019.

[46] 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.

[47] C. Wei, W. Wang, W. Yang, and J. Liu. Deep Retinex Decomposition for Low-Light Enhancement. In *Proc. BMVC*, 2018.

[48] K. Wei, Y. Fu, J. Yang, and H. Huang. A Physics-based Noise Formation Model for Extreme Low-light Raw Denoising. In *Proc. CVPR*, Apr. 2020.

[49] J. Xu, Y. Hou, D. Ren, L. Liu, F. Zhu, M. Yu, H. Wang, and L. Shao. STAR: A Structure and Texture Aware Retinex Model. *IEEE Trans. Image Process.*, 29:5022–5037, 2020.

[50] K. Xu, X. Yang, B. Yin, and R. W. H. Lau. Learning to Restore Low-Light Images via Decomposition-and-Enhancement. In *Proc. CVPR*, 2020.

[51] Y. Xu, B. Liu, Y. Quan, and H. Ji. Unsupervised Deep Background Matting Using Deep Matte Prior. *IEEE Trans. Circuits Syst. Video Technol.*, pages 1–1, 2021.

[52] W. Yang, S. Wang, Y. Fang, Y. Wang, and J. Liu. From Fidelity to Perceptual Quality: A Semi-Supervised Approach for Low-Light Image Enhancement. In *Proc. CVPR*, 2020.

[53] 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.

[54] S.-Y. Yu and H. Zhu. Low-Illumination Image Enhancement Algorithm Based on a Physical Lighting Model. *IEEE Trans. Circuits Syst. Video Technol.*, 29(1):28–37, Jan. 2019.

[55] H. Yue, J. Yang, X. Sun, F. Wu, and C. Hou. Contrast Enhancement Based on Intrinsic Image Decomposition. *IEEE Trans. Image Process.*, 26(8):3981–3994, Aug. 2017.

[56] Y. Zhang, J. Zhang, and X. Guo. Kindling the Darkness: A Practical Low-light Image Enhancer. In *Proc. ACM MM*, 2019.

[57] Z. Zhao, B. Xiong, L. Wang, Q. Ou, L. Yu, and F. Kuang. RetinexDIP: A Unified Deep Framework for Low-light Image Enhancement. *IEEE Trans. Circuits Syst. Video Technol.*, pages 1–1, 2021.

[58] M. Zhu, P. Pan, W. Chen, and Y. Yang. EEMEFN: Low-Light Image Enhancement via Edge-Enhanced Multi-Exposure Fusion Network. In *Proc. AAAI*, 2020.

Jinxiu Liang received the Ph.D. and B.E. degrees from South China University of Technology, Guangzhou, China, in 2021 and 2016, respectively. She is currently a postdoctoral research fellow of computer science with the Peking University, Beijing, China. Her research interests include computer vision, computational photography, and machine learning.

Yong Xu received the B.S., M.S., and Ph.D. degrees in mathematics from Nanjing University, Nanjing, China, in 1993, 1996, and 1999, respectively. He was a Postdoctoral Research Fellow of computer science with the South China University of Technology, Guangzhou, China, from 1999 to 2001, where he became a Faculty Member and is currently a professor with the School of Computer Science and Engineering. He is the Dean of Guangdong Big Data Analysis and Processing Engineering & Technology Research Center. His current research interests include computer vision, pattern recognition, image processing, and big data. He is a senior member of the IEEE Computer Society and the ACM. He has received the New Century Excellent Talent Program of MOE Award.

Yuhui Quan received the Ph.D. degree in Computer Science from South China University of Technology in 2013. He worked as the postdoctoral research fellow in Mathematics at National University of Singapore from 2013 to 2016. He is currently the associate professor at School of Computer Science and Engineering in South China University of Technology. His research interests include computer vision, image processing and sparse representation.

Boxin Shi received the BE degree from the Beijing University of Posts and Telecommunications, the ME degree from Peking University, and the PhD degree from the University of Tokyo, in 2007, 2010, and 2013. He is currently a Boya Young Fellow Assistant Professor and Research Professor at Peking University, where he leads the Camera Intelligence Lab. Before joining PKU, he did postdoctoral research with MIT Media Lab, Singapore University of Technology and Design, Nanyang Technological University from 2013 to 2016, and worked as a researcher in the National Institute of Advanced Industrial Science and Technology from 2016 to 2017. His papers were awarded as Best Paper Runner-Up at International Conference on Computational Photography 2015 and selected as Best Papers from ICCV 2015 for IJCV Special Issue. He has served as an editorial board member of IJCV and an area chair of CVPR/ICCV. He is a senior member of IEEE.

Hui Ji received the B.Sc. degree in Mathematics from Nanjing University in China, the M.Sc. degree in Mathematics from National University of Singapore and the Ph.D. degree in Computer Science from the University of Maryland, College Park. In 2006, he joined National University of Singapore as an assistant professor in Mathematics. Currently, he is an associate professor in mathematics at National University of Singapore. His research interests include computational harmonic analysis, optimization, image processing and machine learning.