# Retinex Image Enhancement Based on Sequential Decomposition With a Plug-and-Play Framework

Tingting Wu, Wenna Wu, Ying Yang, Feng-Lei Fan, *Member, IEEE*, and Tieyong Zeng

*Abstract*— The Retinex model is one of the most representative and effective methods for low-light image enhancement. However, the Retinex model does not explicitly tackle the noise problem and shows unsatisfactory enhancing results. In recent years, due to the excellent performance, deep learning models have been widely used in low-light image enhancement. However, these methods have two limitations. First, the desirable performance can only be achieved by deep learning when a large number of labeled data are available. However, it is not easy to curate massive low-/normal-light paired data. Second, deep learning is notoriously a black-box model. It is difficult to explain their inner working mechanism and understand their behaviors. In this article, using a sequential Retinex decomposition strategy, we design a plug-and-play framework based on the Retinex theory for simultaneous image enhancement and noise removal. Meanwhile, we develop a convolutional neural network-based (CNN-based) denoiser into our proposed plug-and-play framework to generate a reflectance component. The final image is enhanced by integrating the illumination and reflectance with gamma correction. The proposed plug-and-play framework can facilitate both post hoc and ad hoc interpretability. Extensive experiments on different datasets demonstrate that our framework outcompetes the state-of-the-art methods in both image enhancement and denoising.

*Index Terms*— Image enhancement, image restoration, plug-and-play, Retinex theory.

## I. INTRODUCTION

IMAGES usually exhibit low contrast and unexpected noise distortions when the photographic environment suffers from low illumination. Such drawbacks not only affect the visual quality of captured images but also hinder the effectiveness of downstream vision tasks such as image classification [1], [2] and segmentation [3]. Image enhancement aims to reconstruct a visually pleasing and clear image from its low-light counterpart. During the past decades, various methods have been proposed for low-light image enhancement [4], [5], [6], [7], [8]. Among them, the Retinex model is one of the most representative and significant models. However, most existing optimization-based Retinex models still suffer from image quality degradation such as heavy noise, inadequate details, and low contrast.

The Retinex theory, proposed by Land and McCann [9] in 1971, shows an impressive agreement with the color perception of the human visual system (HVS) and inspires numerous image enhancement algorithms [6], [10], [11]. According to the Retinex theory, an observation $S$ can be represented by the pixelwise product of two different components: a reflectance layer $R$ and an illumination layer $L$

$$S = R \cdot L \tag{1}$$

where $R$ denotes the inherent property of the scene surface and contains details and color information of the original image, while $L$ represents the intensity and distribution of the environmental illumination. Note that $L$ is spatially determined by the darkened regions of the image.

Simultaneously, estimating $R$ and $L$ from $S$ according to (1) is an ill-posed task. To address this issue, priors are needed to be incorporated. Ng and Wang [6] adopted a log transform to overcome the ill-posedness of the Retinex model. Then, the total variation regularization and the $L_2$-norm regularization were employed to estimate the log-reflectance component and log-illumination component, respectively. However, the log transform may seriously distort gradient information, which is problematic for noise in image enhancement tasks [6], [11], [13]. To circumvent the abovementioned limitations, numerous decomposition algorithms have been investigated without the log transform. Gu et al. [14] predicted the illumination and the reflectance directly in the image domain by $L_2$ and $L_1$ norms, respectively. Following the work in [14], Gu et al. [12] proposed a Retinex-based fractional-order total variational (FOTV) model by employing the fractional-order gradient total variation regularization $[\nabla^\alpha, \alpha \in (1, 2)]$ on both the reflectance and the illumination. However, such a decomposition method is mainly carried out in the V-channel

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

2

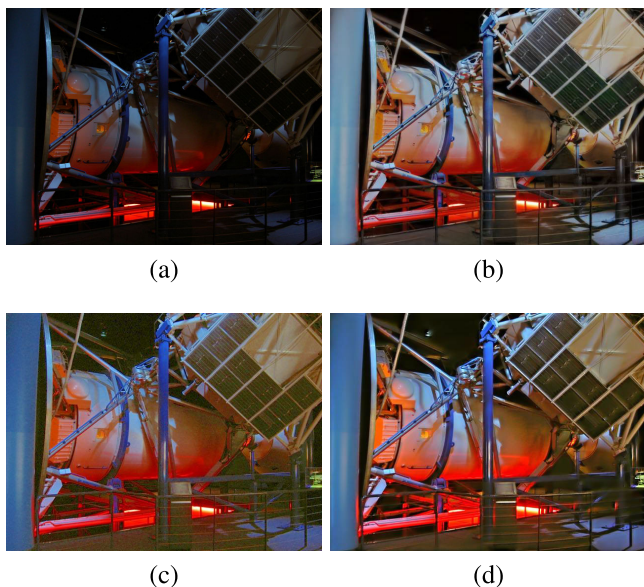IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS



Fig. 1. Exemplary comparison with two existing methods on Img3. The images generated by KinD [8] and FOTV [12] still have low contrast. Instead, the image recovered by ours achieves a desirable result. (a) Img3. (b) KinD [8]. (c) FOTV [12]. (d) Ours.
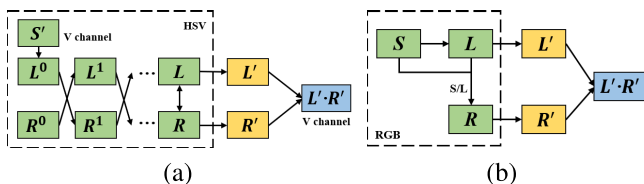


Fig. 2. Comparison between the common and our sequential decomposition strategies. The parts in the dotted boxes represent the Retinex decomposition operations. (a) Alternating decomposition strategy. (b) Our Retinex sequential decomposition method. $L'$ and $R'$ demonstrate the illumination and reflectance obtained by gamma correction, respectively.

of the HSV color space, which ignores the noise in the other two channels. Thus, although these methods can remove noise to a certain extent while preserving finer details, their denoising ability is still limited.

It is increasingly noticed that noise in the dark channel is a key issue in image enhancement, which must be treated appropriately [7], [8], [15], [16], [17], [18], [19]. To desirably enhance the low-light illuminated images, the model should be endowed with denoising capability. In recent years, due to its excellent performance, deep learning has been widely used in low-light image enhancement. Realizing the criticality of denoising, a plethora of deep models [7], [8], [15], [16], [17], [18], [19] take the denoising as a separate task. However, these methods have two limitations.

1) The desirable performance can only be achieved by deep learning when a large number of labeled data are available. However, it is not easy to curate massive low-/normal-light paired data due to the following reasons. First, there is no convincing synthetic method to simulate realistic low-light images from normal-light ones because illumination conditions are highly varied. Second, the existing datasets are either small or collected in a too specific context such as the extremely underexposed condition to fulfill the general

need. Then, the model trained over such datasets may not be translatable.
2) Deep learning is notoriously a black-box model [20]. Despite that deep learning delivers great performance in real-world tasks, it is difficult to explain their inner working mechanism and understand their behaviors.

Overall, existing methods still have some drawbacks as shown in Fig. 1. To address these issues, inspired by the plug-and-play image restoration method [21], we propose a framework that uses sequential decomposition strategy to solve $R$ and $L$ sequentially, thereby avoiding the alternating iteration and canceling the mutual interference between solving $R$ and $L$. Fig. 2 shows the differences between the common Retinex decomposition and the sequential decomposition strategies. This framework circumvents the reliance on large paired low-/high-light data, a key problem encountered in image enhancement, by transforming the problem of learning how to enhance an image into the problem of learning how to do image denoising in solving $R$. The training of a convolutional neural network-based (CNN-based) denoiser can be done reliably and effectively over synthetic noisy/clean data because synthesizing noise is much easier and more faithful than synthesizing light illumination. Furthermore, the plug-and-play framework promotes interpretability. On the one hand, due to the modularized structure of the plug-and-play framework, it is relatively easier to apply post hoc analysis to identify the mechanism of the framework. Thus, we can make the most use of the performance of advanced deep learning denoisers and use a post hoc analysis method to gain interpretability. Along this direction, instead of translating the existing CNN denoisers into this task, we independently design a simple yet efficient and effective CNN-based denoiser. On the other hand, we can directly seek an explainable CNN-based denoiser. The employment of an interpretable denoiser can further enhance the entire framework's interpretability. Specifically, we use a wavelet-inspired autoencoder [22], which is essentially a learnable wavelet shrinkage model. Overall, our contributions are threefold.

1) We present an efficient sequential decomposition Retinex algorithm to solve the illumination and reflectance functions. To the best of our knowledge, it is the first time to investigate the problem of low-light image enhancement by a novel plug-and-play framework. This framework is an organic fusion of model- and data-driven modalities, which can circumvent the reliance on large low-/normal-light data.
2) Through a post hoc analysis and applying an explainable denoiser, we show how the modularized structure in the proposed plug-and-play framework facilitates interpretability in terms of both post hoc analysis and ad hoc interpretable modeling.
3) Extensive and systematic low-light image enhancement experiments are conducted on different datasets to demonstrate the outperformance of our proposed method from both quantitative and qualitative aspects.

## II. RELATED WORK

### A. Low-Light Image Enhancement Methods

During the past decades, a large number of algorithms have been developed for low-light image enhancement. These

algorithms can be reasonably categorized as model- and learning-based algorithms.

*1) Model-Based Methods:* In this section, we mainly introduce two model-based methods: the histogram equalization (HE)-based methods and the decomposition-based methods.

Intuitively yet effectively, HE enhances image contrast by changing the histogram distribution. The basic principle of HE and its variants is to expand the dynamic range of pixels in an image. Different from the standard HE algorithm, the adaptive HE (AHE) [4] adjusts image contrast by calculating the local histogram and redistributing the brightness. Although AHE enhances local details in extremely dark or bright areas, it also amplifies noise. The work in [5] can alleviate overexposure and noise amplification by limiting the distribution of gray levels in different regions (CLAHE). Subsequently, various HE-based methods are devised to improve the overall visual quality based on different constraints [23], [24].

Unfortunately, these HE-based methods are handicapped to images with nonuniform illumination, which motivates the invention of decomposition-based methods. This thread of work assumes that an image can be decomposed into the reflectance and the illumination components, and an image is enhanced by further processing and integrating these two components. The existing decomposition-based approaches can be categorized into the variational methods [6], [10], [11], [12], [13], [14], [25], [26], [27], [28], path-based algorithms [29], [30], recursive algorithms [31], [32], partial differential equation (PDE)-based methods [33], [34], and learning-based methods [7], [8], [18], [35], [36]. However, these techniques tend to induce sketchy details, prominent noise, and unknown artifacts.

*2) Learning-Based Methods:* Recently, many learning-based methods have emerged with promising performance for underexposed image correction. The key bottleneck therein is the lack of paired low-/high-light datasets. To solve this problem, Wei et al. [7] collected a dataset named LOL, which was widely applied in many works. Meanwhile, they introduced an end-to-end trainable network called RetinexNet, which is the first work that combines the Retinex theory and deep learning. Chen et al. [18] built the SID dataset and trained an end-to-end network with good noise reduction and image enhancement, particularly for the extremely underexposed images. Considering the poor generalization performance of these methods in other test sets, some work attempts to adopt semisupervised or unsupervised learning, such as RetinexDIP [37]. However, their results sometimes suffer from unexpected artifacts and improper exposure.

Another key problem in the image enhancement task is the noise hidden in the dark. Lore et al. [38] exploited an autoencoder to extract the image information and performed the enhancement. Wang et al. [39] designed the DeepUPE, which establishes the loss function according to the prior information of illumination. Since these methods are not designed for noise, the enhanced results still suffer visible noise. Some learning-based methods treat denoising as an independent module. Lv et al. [15] proposed a multibranch decomposition-and-fusion enhancement network to cope well with color distortion and noise. Xu et al. [16] designed a frequency-based network for denoising and enhancement simultaneously. They adopted a cross domain transformation (CDT) module to eliminate noise and protect details. Zhu et al. [17] designed a two-stage network to effectively remove noise through the multiexposure fusion module. In our study, we also regard noise suppression as a nonnegligible factor. Progressive Retinex [19] uses two fully pointwise convolutional neural networks to simulate the statistical regularities of ambient light and image noise and leverage them as constraints to facilitate the mutual learning process, which not only avoids the ambiguity between tiny textures and image noise but also enhances the computational efficiency. Liu et al. [40] focused on constructing a lightweight yet effective network, referred to as RUAS [40] to enhance low-light images in real-world scenarios. Yang et al. [41] noticed that there lacks a desirable objective for low-light image enhancement, and they designed an end-to-end signal prior-guided layer separation network with layer-specified constraints.

Despite the promising performance, these methods either are trained over synthetic data and data collected in a specific condition or suffer from the lack of interpretability. In contrast, by synergizing the model- and data-driven modalities, our plug-and-play framework delivers superior performance without dependence on paired low-/normal-light data and enjoys interpretability.

## B. Plug-and-Play Framework

Earlier, the plug-and-play framework was proposed to solve the denoising problem [21], which allows the insertion of different denoisers for prior knowledge learning. The plug-and-play framework mainly contains two steps. First, the objective function is decoupled into a fidelity subproblem and a prior subproblem via the variable splitting algorithms. These two subproblems are combined into an iterative scheme solved alternately. Second, the existing state-of-the-art denoising techniques can be employed directly to solve the prior subproblem, including regularization denoising methods [42], [43] and learning-based denoising methods [44].

Recently, various image reconstruction approaches have been explored based on the plug-and-play framework, such as image super-resolution [45], image deblurring [46], and image denoising [46], [47]. Zhang et al. [46] introduced the IRCNN model for nonblind image deblurring and image denoising by plugging a deep denoiser prior into the half quadratic splitting (HQS) algorithm [48]. Sun et al. [47] developed a block-coordinate regularized denoising algorithm, which decomposes the large-scale estimation problem into a series of updates covering a small part of unknown variables. These methods have revealed the surprising potential of the plug-and-play framework in different image restoration tasks.

One of the major advantages of the plug-and-play framework is that a pretrained denoiser can be used if there is no enough data for end-to-end training. Since the available paired data for image enhancement tasks are very limited, the plug-and-play framework is suitable for our task. At the same time, the model can gain interpretability from a mathematical perspective since there is a close tie between the plug-and-play framework and the traditional restoration methods.

## III. MODEL AND ALGORITHM

Our proposed method mainly consists of three components: the sequential decomposition module, the denoising module, and the adjustment module. Because of the sequential decomposition, $L$ and $R$ are solved independently. For the illumination layer $L$, we first compute the initial illumination $\hat{L}$ from the input $S$ via the meanRGB operation; then, we derive $L$ by minimizing a loss function through iterative optimization. For the reflectance layer $R$, since $L$ is done, we can get the initial map of $R$ through the simple relationship, i.e., $R^{(0)} = S/L$. Then, the reflectance map $R$ is processed by a CNN-based denoiser. Finally, the gamma correction is employed to combine the illumination and reflectance maps to yield the final enhanced image. Fig. 3 shows the flowchart of our proposed method based on the sequential decomposition within a plug-and-play framework.

### A. Retinex Model With the Plug-and-Play Scheme

We propose a plug-and-play scheme for low-light image enhancement with the sequential decomposition-based Retinex model. We construct the following optimization function to alleviate the ill-posedness of (1):

$$\min_{R,L} \|S - L \cdot R\|_F^2 + \lambda\Phi(R) + \beta\|\nabla R - G\|_F^2 + \alpha\|\nabla L\|_1 \tag{2}$$

where $S$, $L$, and $R$ denote the low-light image, estimated illumination, and reflectance, respectively; $G$ is the adjustment to $\nabla S$; $\lambda$, $\alpha$, and $\beta$ are the regularization parameters; $\|\cdot\|_F$ represents the Frobenius norm; and $\|\cdot\|_1$ denotes the $\ell_1$ norm. Next, we explain in detail each term in the objective function:

1) The term $\|S - L \cdot R\|_F^2$ minimizes the Frobenius distance between $S$ and $L \cdot R$, which is the fidelity term.
2) $\Phi(R)$ is the regularization term of $R$, which is used to denoise the reflectance part. If $\Phi(R) = \|\nabla R\|_1$, the proposed model (2) is a hybrid model about total variation. If $\Phi(R) = \sum_i \|NN_i(R)\|_*$, the proposed model (2) is a low-rank hybrid model, and here, $\|\cdot\|_*$ represents the nuclear norm and $NN_i(R) = [R_{i1}, R_{i2}, \ldots, R_{ik}]$ is the similar patch group of the reference patch $R_{i1}$. In our proposed framework, $\Phi(R)$ is an implicit prior whose subproblem is denoising.
3) The third term $\|\nabla R - G\|_F^2$ minimizes the distance between $\nabla R$ and $G$. Since $G$ is obtained from $\nabla \hat{S}$ (adjusted by the gradient of $S$), the details of edges can be preserved. $\nabla \hat{S}$ and $G$ [49] are given by

$$\nabla\hat{S} = \begin{cases} 0, & \text{if } \nabla\hat{S} < \varepsilon \\ \nabla S, & \text{otherwise} \end{cases} \tag{3}$$

$$G = \left(1 + \kappa e^{-|\nabla\hat{S}|/\sigma}\right) \circ \nabla\hat{S}. \tag{4}$$

Here, $\epsilon$ can filter out small gradients; $\sigma$ and $\kappa$ are the parameters that control the enlarged level. Equation (3) can suppress the possible noise before the amplification.
4) $\|\nabla L\|_1$ ensures the smoothness of the illumination layer.

Typically, to estimate the illumination map $L$ and the reflectance map $R$, we can solve (2) in an alternating fashion (solving $L$ and $R$ alternatively in the iteration). However,

this is time-consuming and complex. To address this issue, we adopt a novel sequential decomposition strategy to generate the illumination part independently. Specifically, we first estimate the illumination $L$ from the initial illumination $\hat{L}$ [see (5)] and then estimate the reflectance map $R$ [see (6)]

$$\min_L \|L - \hat{L}\|_F^2 + \alpha\|\nabla L\|_1 \tag{5}$$

$$\min_R \|S - L \cdot R\|_F^2 + \lambda\Phi(R) + \beta\|\nabla R - G\|_F^2. \tag{6}$$

*1) Illumination Map (L) Estimation:* Recently, numerous approaches have been proposed to directly estimate the initial illumination map $\hat{L}$, such as the maxRGB [50] or meanRGB [25] operator. These approaches preassume that each channel of an RGB image has a common illumination. We adopt the meanRGB operator to ensure the consistency of illumination, which is defined as

$$\hat{L}(x) = \frac{1}{3}\sum_{c\in\{R,G,B\}} S^c(x) \tag{7}$$

where $x$ denotes the pixel of the image. Then, an alternating direction minimization method (ADMM) [51] is employed to solve (5)

$$\min_L \|L - \hat{L}\|_F^2 + \alpha\|v\|_1, \quad \text{s.t. } \nabla L = v \tag{8}$$

where $v$ is an auxiliary variable. Then, we can create a corresponding augmented Lagrangian function

$$\mathcal{L}_\theta(L, v) = \|L - \hat{L}\|_F^2 + \alpha\|v\|_1 + \frac{\theta}{2}\|\nabla L - v\|_F^2 + \langle Z, \nabla L - v\rangle \tag{9}$$

where $\theta$ is the penalty parameter and $Z$ is the Lagrangian multiplier. The ADMM algorithm is derived by minimizing $\mathcal{L}$ with respect to $L$ and $v$ (one at a time while fixing the other at its most recent value). By direct computation, we can get the updating formula of $L$

$$L^{(k+1)} = \left(2 + \theta^{(k)}\nabla^T\nabla\right)^{-1}\left(2\hat{L} + \theta^{(k)}\nabla^T v^{(k)} - \nabla^T Z^{(k)}\right) \tag{10}$$

which is implemented by fast Fourier transform (FFT) and inverse FFT (IFFT). Meanwhile, the $v$-subproblem can be quickly solved by the soft shrinkage

$$v^{(k+1)} = \eta\left(\nabla L^{(k+1)} + \frac{Z^{(k)}}{\theta^{(k)}}, \frac{\alpha}{\theta^{(k)}}\right) \tag{11}$$

where $\eta$ is a soft shrinkage function defined as

$$\eta(x_{ij}, c) := \frac{x_{ij}}{|x_{i,j}|} \cdot \max(|x_{i,j}| - c, 0).$$

Finally, the Lagrangian multiplier $Z$ and the parameter $\theta$ are updated through

$$\begin{aligned} Z^{(k+1)} &\leftarrow Z^{(k)} + \theta^{(k)}\left(\nabla L^{(k+1)} - v^{(k+1)}\right), \\ \theta^{(k+1)} &\leftarrow \theta^{(k)}\rho, \quad \rho > 1. \end{aligned} \tag{12}$$

Here, $\rho$ is the step size. The iteration will terminate when $\|\nabla L - v\|_F \le \iota\|\hat{L}\|_F$ with $\iota = 10^{-5}$ or $k$ reaches the maximum value.
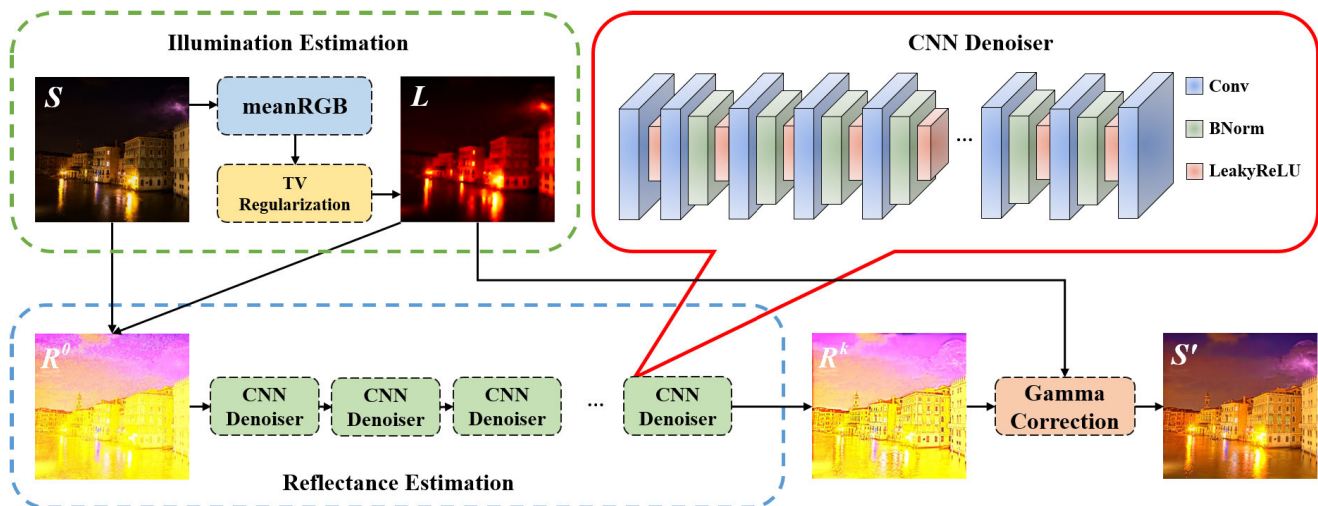
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

WU et al.: RETINEX IMAGE ENHANCEMENT BASED ON SEQUENTIAL DECOMPOSITION 5



Fig. 3. Flowchart of our method. We first obtain the illumination $L$ via refining the initial illumination $\hat{L}$ generated by the meanRGB operation. Then, we obtain the initial value of $R$ via $R^{(0)} = S/L$. The noise-free reflectance map is obtained by inserting the denoiser with the designed plug-and-play framework. Finally, gamma correction is applied to the decomposed component to adjust the image.

*2) Solution of R-Subproblem:* Following the HQS method [48], (6) can be reformulated as a constrained optimization problem by introducing an auxiliary variable $z$:

$$\min_{R} \|S - L \cdot R\|_F^2 + \lambda \Phi(z) + \beta \|\nabla R - G\|_F^2$$
$$\text{s.t. } R = z. \tag{13}$$

Then, (13) is solved by minimizing the following problem:

$$\mathcal{L}_\mu(R, z) = \|S - L \cdot R\|_F^2 + \lambda \Phi(R)$$
$$+ \beta \|\nabla z - G\|_F^2 + \frac{\mu}{2}\|z - R\|_F^2 \tag{14}$$

where $\mu$ is a positive penalty scalar.

*z-Subproblem (Contrast Enhancement):* The fidelity term and regularization term are decoupled into two individual subproblems. Collecting the $z$-involved terms from (14) gives the problem as follows:

$$z^{(k+1)} = \arg\min_{z} \beta \|\nabla z - G\|_F^2 + \frac{\mu}{2}\|R - z^{(k)}\|_F^2. \tag{15}$$

Then, the updating formula of $z^{k+1}$ is given by

$$z^{(k+1)} = \frac{2\beta \nabla^T G + \mu R^{(k)}}{2\beta \nabla^T \nabla + \mu I}. \tag{16}$$

Note that the update of $z^{(k+1)}$ can be implemented by FFT and IFFT.

*R-Subproblem (Noise Suppression):* Neglecting the terms unrelated to $R$, the $R$-subproblem can be solved by the following iterative scheme:

$$R^{(k+1)} = \arg\min_{R} \|S - L \cdot R\|_F^2 + \frac{\mu}{2}\|z^{(k+1)} - R\|_F^2 + \lambda \Phi(R). \tag{17}$$

We rewrite (17) as follows:

$$R^{(k+1)} = \arg\min_{R} \left\| \sqrt{2L^2 + \mu I} R - \frac{2S \cdot L + \mu z^{k+1}}{\sqrt{2L^2 + \mu I}} \right\|_F^2 + \lambda \Phi(R). \tag{18}$$

**Algorithm 1** Single-Image Low-Light Enhancement

**Input:** The input image $S$, $k = 0$, $\alpha = 0.1$, $\theta = 0.0045$, $\rho = 1.08$, $Z^{(0)} = 0$, $\epsilon = 1$, $\kappa = 2.5$, $\sigma = 10$;
1: Compute $G$ via Eqs. (3) and (4);
2: Estimate $\hat{L}$ via Eq. (7);
3: **while** not converged **do**
4:     Update $L^{(k+1)}$ via Eq. (10);
5:     Update $v^{(k+1)}$ via Eq. (11);
6:     Update $Z^{(k+1)}$ via Eq. (12);
7: **end while**
**Input:** $R^{(0)} = S/L$, $k = 0$, $\mu = 0.001$, $\beta = 0.001$, noise level $\omega$, gamma correction coefficient $\gamma_1, \gamma_2$;
8: **while** not converged **do**
9:     Update $z^{(k+1)}$ via Eq. (16);
10:     Update $R^{(k+1)}$ via Eq. (19);
11: **end while**
12: Estimate $S'$ via Eq. (22).
**Output:** The estimated result $S'$.

Solving (18) is essentially equivalent to solving an adaptive denoising problem, where different regions of an image are configured with different weights. Such an adaptation is plausible because the noise level varies in different regions. Usually, darker regions are subjected to stronger noise in image enhancement. Due to the equivalence, we can use a Gaussian denoiser to solve (18) as

$$R^{(k+1)} = \text{Denoiser}\left(\frac{2S \cdot L + \mu z^{k+1}}{2L^2 + \mu I}, \omega\right) \tag{19}$$

where $\omega$ is a hyperparameter of the denoiser related to the noise level. For example, for different levels of noise, the denoising effect of the Gaussian denoiser should be different. In the experiment, the selection of $\omega$ depends on images. We will discuss how $\omega$ affects the overall image enhancement performance in Section VI-B2.

*Summary:* According to the Retinex theory, an observed image $S$ can be represented by the pixelwise product of two different components: a reflectance layer $R$ and an illumination

layer $L$: $S = R \cdot L$, where $R$ denotes the inherent property of the scene surface and contains details and color information of the original image, while $L$ represents the intensity and distribution of the environmental illumination. Our framework uses a sequential decomposition strategy to solve $R$ and $L$ sequentially, thereby avoiding the alternating iteration and canceling the mutual interference between solving $R$ and solving $L$. Fig. 2 shows the differences between the common Retinex decomposition strategy and the sequential decomposition strategy. Furthermore, the important prior subproblem in solving $R$ is reduced to a denoising problem, amendable to the advanced denoisers to be employed in a plug-and-play manner. This is why our framework is plug-and-play.

### B. CNN-Based Denoisers

Since our framework is plug-and-play, denoisers can be selected as appropriate. In recent years, deep learning has been widely used in various fields [52], [53]. Considering that learning-based methods significantly outperform traditional variational models, we design a CNN-based denoiser to solve the $R$ subproblem in the image enhancement task. In addition, we also use an explainable denoiser.

*Ours:* Currently, the activation unit ReLU is the most widely used nonlinear activation function in deep learning because of its good properties in preventing gradients vanishing or explosion. However, ReLU also tends to block information transmission because it sets zero to all negative parts of the input, which may block reasonable information circulation because those inhibited neurons will not be updated in the backpropagation. To address this issue, a thread of the activation functions, e.g., ON/OFF ReLU [54], Concatenated-ReLU [55], and Leaky-ReLU [56], which consider the information flow of negative parts, were proposed to supplement ReLU in deep networks. Compared to other activation functions, Leaky-ReLU has a smaller number of parameters and a reasonably effective generalization performance. Therefore, we design the CNN-based denoiser with Leaky-ReLU, as shown in Fig. 3. The new denoiser mainly includes three parts: "Conv + LeakyReLU" for the first layer, "Conv + BactchNorm + LeakyReLU" for the 2–18 layers, and "Conv" for the last layer. The convolutional layer in the first part uses 64 filters to generate 64 feature maps and the size is $3 \times 3 \times 3$. Also, we use the filters of size $3 \times 3 \times 64$ for the second and last parts. The LeakyReLU function is defined as follows:

$$\text{LeakyReLU}(x) = \begin{cases} x, & x > 0 \\ ax, & x \geq 0 \end{cases}$$

where $a$ denotes a small constant and is usually set to $-0.02$. We also add batch normalization between the convolutional layer and the activation function. The application of batch normalization can also improve training efficiency and denoising performance [57].

*Soft-AE (An Explainable Denoiser):* Fan et al. [22] proposed the so-called soft autoencoder (Soft-AE), which is interpretable based on the wavelet shrinkage theory. As shown in Fig. 4, the activation functions in the encoding layers are set to soft-thresholding units $\eta_{b<0}(x) = \text{sgn}(x) \max\{|x| + b, 0\}$,
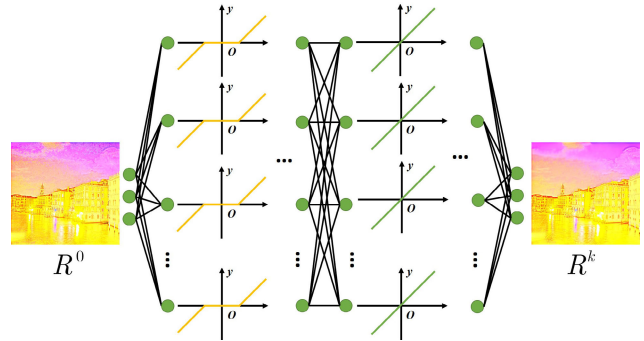


Fig. 4. Architecture of Soft-AE. The Soft-AE is a learnable wavelet shrinkage algorithm.

where $b$ is a threshold, $\text{sgn}(\cdot)$ is the sign function, and activation functions of the decoding layers are linear. As a result, convolutional layers in the encoder part conduct wavelet transform, while convolutional layers in the decoder part conduct inverse wavelet transform. The Soft-AE can be regarded as unrolling the cascade wavelet shrinkage algorithm into a network, and the wavelet transform is learnable.

Mathematically, the wavelet shrinkage algorithm consists of three steps. Suppose that we have the following additive noise model: $Y(t) = S(t) + N(t)$, where $Y(t)$ and $S(t)$ are the measured and the authentic signals, respectively: 1) perform the wavelet transform over the noise signal to derive wavelet coefficients: $\hat{Y} = W[Y]$; 2) apply an elementwise soft-thresholding activation to the wavelet coefficients: $Z = \eta_{-\sigma_N (2 \log n)^{1/2}}(\hat{Y})$, where is the noise variance and $n$ is the number of pixels; and 3) perform the inverse wavelet transform: $\hat{S} = W^{-1}[Z]$.

---

**Algorithm 2** Wavelet Shrinkage Algorithm
---
**Input:** $Y(t) = S(t) + N(t)$, wavelet $\phi$
  1: Wavelet transform by $\phi$: $\hat{Y} = W_\phi[Y(t)]$;
  2: Soft thresholding: $Z = \eta_b(\hat{Y})$;
  3: Inverse wavelet transform by $\phi^{-1}$: $\hat{S} = W_\phi^{-1}[Z]$;
**Output:** $\hat{S}(t)$

---

*Proposition 1:* Let $\hat{S}(t)$ and $S(t)$ be the recovered signal and the noisy signal in Algorithm 2. Given the Besov norm $B$ that is a measure for smoothness, there is a universal constant $\pi_n$, where $n$ is the number of elements in $\hat{S}(t)$ and $S(t)$, with $\pi_n \to 1$ as $n \to \infty$, and constant $C(B)$ depending on the Besov norm such that

$$\text{Prob}\{\|f_{LR}\|_B \leq C(B)\|\hat{f}_{HR}\|_B\} \geq \pi_n. \tag{20}$$

Proposition 1 reveals the important smoothness relationship between the original noisy signal and their wavelet recovery. With the overwhelming likelihood and in a common smoothness measure: Besov norm, the recovered signal is smoother than the noisy signal. Proposition 1 offers a theoretical guarantee for the denoising effect of Algorithm 2. Considering that the soft-AE performs a network-based wavelet transform, it inherits such a theoretical guarantee from the wavelet shrinkage theory.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

WU et al.: RETINEX IMAGE ENHANCEMENT BASED ON SEQUENTIAL DECOMPOSITION                                                                 7

## C. Gamma Correction

To further improve the enhancement effect, the gamma correction can be used for illumination adjustment after obtaining the final reflectance $R$ and illumination $L$. Then, the final result can be obtained by

$$S' = R \cdot L' \tag{21}$$

where $\cdot$ denotes the elementwise multiplication and $L' = L^{(1/\gamma)}$. Usually, the value of $\gamma$ is empirically set to 2.2.

This gamma correction method assumes that the illumination map from the Retinex model is close to the real illumination. However, this assumption may not hold for the extremely low illumination. Thus, we adopt the following gamma correction strategy for the extremely low-light images:

$$S' = R' \cdot L' \tag{22}$$

where $R' = R^{(1/\gamma_1)}$ and $L' = L^{(1/\gamma_2)}$. We can manually adjust the values of $\gamma_1$ and $\gamma_2$ according to the characteristics of the images. Finally, the whole procedure of the proposed framework is summarized in Algorithm 1.

## IV. INTERPRETABILITY OF THE PROPOSED FRAMEWORK

The employment of the plug-and-play framework features modularization, which sequentially solves $R$ and $L$ and avoids the crosstalk between them. Our highlight is that the plug-and-play framework can enhance interpretability. Generally, interpretability means to what extent a human can understand and reason a model. In [20], interpretability is divided into post hoc interpretation and ad hoc interpretable modeling. The former is conducted after a model is well learned. The main advantage of post hoc methods is that one does not need to compromise interpretability with the predictive performance since prediction and interpretation are separate. The latter is to prototype an interpretable model. The merit of ad hoc interpretable modeling is that it can avoid the bias of post hoc interpretation. We argue that the modularization in the plug-and-play framework can enhance both post hoc and ad hoc interpretability. On the one hand, we can use a powerful denoising network, conjugated by a dedicated post hoc analysis, to simultaneously enjoy satisfactory denoising performance and interpretability. On the other hand, we can straightly utilize an interpretable denoiser. In the following, we illustrate them in detail.

### A. Post Hoc Analysis

Since our framework is essentially iterative, we develop an explanatory directed graph to illustrate how $R$ is solved as the iteration goes. For example, is it local? When and where the information is lost severely? Specifically, inspired by the study in [58], we set a pixel $R_i^t$ of interest obtained in the $t$th iteration to zero (mask it) and examine how $R$ in the subsequent iteration is altered. Suppose that $R_j^{t+1}$, where $j \neq i$, has a dramatic change; it concludes that $\hat{R}_i^t$ has a major impact on $R_j^{t+1}$. Then, we build a link between them. Repeating this procedure for different pixels and iterative steps, we expect to track the information evolution in the iteration.

### B. Ad Hoc Modeling

The aforementioned Soft-AE is essentially unrolling the cascade wavelet shrinkage algorithm into a network, which is naturally more interpretable than the conventional autoencoders. Let us mathematically illustrate the connection between the Soft-AE and the wavelet shrinkage system.

Without loss of generality, we consider a four-convolutional-layer Soft-AE. Suppose that this four-convolutional-layer Soft-AE contains $N$ filters in the first encoding layer and $M \times N$ filters in the second encoding layer, which are denoted as $\psi_i, i \in [N]$ and $\psi_{ij}, i \in [M]; j \in [N]$, respectively. In symmetry, the two decoding layers of this four-convolutional-layer Soft-AE consists of $N \times M$ and $N$ filters, which are denoted as $\phi_{ij}, i \in [N]; j \in [M]$ and $\phi_i, i \in [N]$, respectively. For simplicity, we use $(\cdot)^+$ to denote the soft-thresholding function $\eta_b(\cdot)$. The final output of this four-convolutional-layer Soft-AE is

$$\sum_k^N \phi_k * \left[ \sum_j^M \phi_{kj} * \left[ \sum_i^N \psi_{ji} * (\psi_i * x)^+ \right]^+ \right] \tag{23}$$

where $*$ represents convolution. We can apply the approximate property of the soft thresholding

$$(h + g)^+ \sim h^+ + g^+. \tag{24}$$

When the threshold is zero, a soft-thresholding activation degenerates into a linear activation. Thus, (24) approximately holds when the threshold is small. As a result, (23) turns into

$$\sum_k^N \phi_k * \sum_j^M \left[ \phi_{kj} * \sum_i^N \psi_{ji} * (\psi_i * x)^+ \right]^+. \tag{25}$$

Let $\Psi$ be a matrix of the size of $M \times N$ whose $(j, i)$-entry is $\psi_{ji}$ and $\Phi$ be the $N \times M$ matrix whose $(k, j)$-entry is $\phi_{kj}$. Since convolution operations conform to the associative laws, (25) is further simplified into the matrix form

$$[\phi_1, \ldots, \phi_N] \otimes \Phi \otimes \Psi \otimes \left[ (\psi_1 * x)^+, \ldots, (\psi_N * x)^+ \right]^T \tag{26}$$

where $(A \otimes B)_{ij} = \sum_k A_{ik} * B_{kj}$, which is analogous to the matrix product, but elements are convolutional filters, and the operation between the elements is convolution. The Soft-AE is to do wavelet shrinkage and can recover the clean signal when the following conditions are satisfied:

$$\begin{cases} \Phi \otimes \Psi = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N)\delta \\ \phi_k = \dfrac{\psi_k^{-1}}{\left| \sum_k^N \lambda_k \right|} \text{ or } \psi_k = \dfrac{\phi_k^{-1}}{\left| \sum_k^N \lambda_k \right|}, \quad k = 1, 2, \ldots, N \end{cases} \tag{27}$$

where $\delta$ is the Dirac function, and the selection of $\Phi \otimes \Psi$ should make $\sum_k^N \lambda_k$ nonzero. Equation (27) can be trivially fulfilled by setting diagonal elements of $\Phi$ and $\Psi$ to be mutually inverse to each other and the rest elements to zero.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

8                                                                                                    IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS

## V. MAIN EXPERIMENTS

Extensive experiments are conducted to evaluate the effectiveness of our framework. First, we introduce the experimental protocols. Second, we analyze via qualitative visualization and quantitative comparisons the image enhancement performance of our proposed framework, with an emphasis on the effects of the initial illumination map and the proposed sequential decomposition.

### A. Experimental Protocols

*1) Datasets:* We conduct extensive experiments on three datasets. The first dataset is Set12, which contains 12 underexposed images with real noise collected in public datasets [10], [61], [62]. The second dataset is LOL [7] (https://daooshee.github.io/BMVC2018website/), which contains 500 low-light images and their normal-light ground truth. We use the fivefold cross validation to test the performance of all supervised learning-based methods on the LOL dataset. Fig. 5 shows the splits of the fivefold cross validation. The third dataset is the Berkeley segmentation dataset (BSD) (https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/) that contains 30 images with normal light and is employed for testing. We use a simple synthetic strategy to transform those normal-light images into low-light versions with noise distortion. First, we generate the low-light image by reducing the brightness of the V-channel of the normal images in HSV color space. Then, the white Gaussian noise with $\sigma = 5$ is added to the low-light images to obtain the desired images.

*2) Compared Methods:* To demonstrate the superiority of our method, we compare it with nine off-the-shelf advanced image enhancement approaches, including three traditional methods FOTV [12], LIME [10] and LR3M [25], and six machine learning-based methods RetinexNet [7], KinD [8], Zero-DCE [59], RetinexDIP [37], RUAS [40], and URetinex [60]. All these methods are either classical benchmarks or state-of-the-arts published in recent two years. We configure the parameters of all compared methods based on the recommendation of the original papers.

*3) Evaluation Metric:* Two types of image quality assessment indexes are employed. The first type is full-reference image quality assessment metrics (FR-IQAs): PSNR, SSIM, and mse that are used when both low-light images and normal-light ground truth are available. The second type is no-reference image quality assessment metrics (NR-IQAs) that are suitable when only low-light images are available. The widely used NR-IQAs are NIQE [63], BTMQI [64], and ARISMC [65]. The lower NIQE, BTMQI, and ARISMC, the better perceptual quality.

### B. Performance Analysis

*1) Qualitative Comparison:* Figs. 6 and 7 show the visual comparison between the proposed method and competitors on LOL179 and Img7.

The highlights of Fig. 6 are given as follows. First, the RetinxDIP method only shows the moderate enhancement effects. Most regions of this enhanced image are still

TABLE I
AVERAGE AND STD NUMERICAL RESULTS ON SET12. THE BEST PERFORMER IS BOLD-FACED. THE RUNNER-UP IS UNDERLINED

| Method | NIQE | BTMQI | ARISMC | Rank |
|---|---|---|---|---|
| LIME [10] | 3.1546±0.8511 | 4.6255±0.8585 | 3.4193±0.9463 | 10 |
| FOTV [12] | 3.0282±0.7282 | 3.6458±0.8939 | 3.4218±0.8790 | 6 |
| LR3M [25] | 3.4828±0.6605 | 3.8042±1.2977 | 2.6589±0.3321 | 7 |
| RetinexNet [7] | 4.6671±0.9783 | 3.4078±0.8609 | 3.0486±0.5056 | 7 |
| KinD [8] | 3.0295±0.6271 | 3.7100±1.4252 | **2.5341±0.1184** | 3 |
| Zero-DCE [59] | 2.8006±0.8096 | 3.5091±1.1919 | 2.7711±0.2568 | 4 |
| RUAS [40] | 4.0258±1.3475 | 5.2329±1.7881 | 2.9608±0.4262 | 11 |
| RetinexDIP [37] | 3.4750±0.7543 | 4.2005±1.2592 | 3.1618±0.5048 | 9 |
| URetinex [60] | 3.4508±0.8024 | 3.4446±1.3577 | 2.6506±0.1011 | 5 |
| Ous-IRCNN | 2.7776±0.7051 | 3.3741±1.4002 | 2.6000±0.1222 | 2 |
| Ours | **2.7198±0.7934** | **3.3436±1.0732** | 2.5780±0.1072 | **1** |
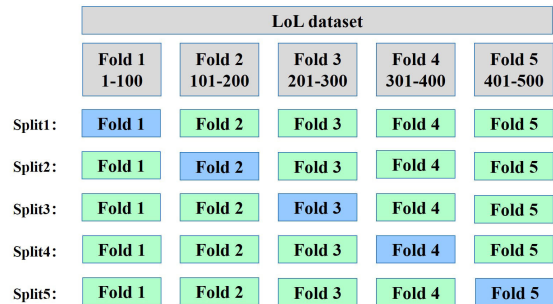


Fig. 5. Split of the fivefold cross validation. In particular, the LOL dataset contains 500 images, and these images are randomly coded from 1 to 500.

extremely dark and insufficiently contrastive. Second, although LIME, RetinexNet, Zero-DCE, and FOTV obtain the desirable enhancement, they fail to eliminate the noise in the dark regions. Third, as shown in the zoomed-in regions, the visual results generated by KinD and LR3M are oversmooth, resulting in unclear and unsharp structural expression, while the URetinex and KinD produce unrealistic results with color distortion. Finally, the proposed model is visually superior, with a better enhancement and a lower noise level. The abovementioned issues, such as heavy noise appearing in the dark regions and unclear structures, are markedly alleviated.

From Fig. 7, we draw four observations. First, most methods perform poorly in denoising, e.g., significant noise remains on the road. Although LR3M and RUAS effectively suppress noise, the texture in the green box is missing because of the overdenoising effect. Second, the results of LIME, RetinexNet, and RUAS are overenhanced, e.g., extra color distortion is present. Third, URetinex and RUAS unsatisfactorily overexpose the sky. Finally, the visual effect created by our proposed model is pleasing. For example, the structures are sharper, and the surface is clearer.

*2) Quantitative Comparison:* Now, we quantitatively compare our framework with its competitors on two test datasets. Specifically, we use the NR-IQAs to measure the performance of all the methods on Set12, and the NR-IQAs and FR-IQAs on the LOL dataset.

Tables I and II summarize the experimental results of all methods on LOL and Set12, respectively. Note that the Ave. Rank is the comprehensive ranking of all numerical results. First, the regularization-based algorithms show suboptimal performance on the Set12 and LOL because those methods ignore the noise distortion in enhancing images. Second,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

WU et al.: RETINEX IMAGE ENHANCEMENT BASED ON SEQUENTIAL DECOMPOSITION

9

TABLE II

AVERAGE AND STD NUMERICAL RESULTS ON THE LOL DATASET. THE BEST PERFORMER IS BOLD-FACED. THE RUNNER-UP IS UNDERLINED

| Average±std | FOTV [12] | LIME [10] | LR3M [25] | RetinexNet [7] | Zero-DCE [59] | KinD [8] | RetinexDIP [37] | RUAS [40] | Ours |
|---|---|---|---|---|---|---|---|---|---|
| PSNR | 13.70±4.14 | 15.24±2.59 | 14.67±4.24 | 15.39±3.22 | 14.16±4.71 | 15.98±4.81 | 8.94±3.20 | 14.24±4.29 | **17.12±3.14** |
| SSIM | 0.5155±0.1703 | 0.4460±0.1556 | 0.6135±0.1775 | 0.5461±0.1594 | 0.5318±0.1893 | 0.6578±0.1803 | 0.2949±0.2022 | 0.4786±0.1670 | **0.6651±0.1667** |
| MSE | 4019.4±3247.6 | 2347.1±1634.8 | 3301.4±2824.4 | 2474.7±2137.7 | 3962.5±3627.4 | 2828.1±3044.1 | 10286.0±5957.3 | 3982.5±4381.8 | **1667.3±1465.6** |
| NIQE | 8.9681±1.7472 | 9.3741±1.9435 | 4.6112±0.7123 | 6.8250±1.3304 | 8.5394±1.8042 | 4.3496±1.1772 | 6.9270±1.3364 | 6.1581±1.3799 | **3.7051±0.8058** |
| BTMQI | 4.9480±1.1103 | 3.7655±0.8957 | 5.5673±1.2495 | **3.1050±1.2812** | 4.6587±1.3103 | 4.5545±1.1164 | 6.2707±1.0134 | 4.6972±1.5844 | 4.4144±1.6022 |
| ARISMC | 3.3168±0.3947 | 3.4268±0.4454 | **2.4946±0.2249** | 3.1503±0.1411 | 3.3750±0.3344 | 2.9158±0.2324 | 3.2709±0.2415 | 3.0668±0.1517 | 2.7255±0.2121 |
| Ave. Rank | 8 | 6 | 4 | 3 | 7 | 2 | 9 | 5 | **1** |
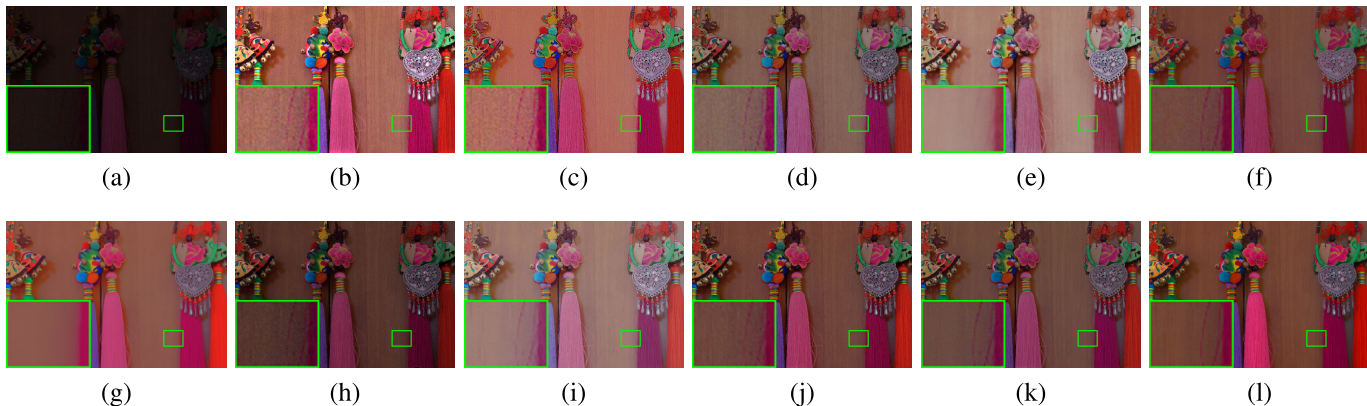




Fig. 6. Enhanced results of existing methods on LOL179. (a) Input. (b) LIME [10]. (c) RetinexNet [7]. (d) Zero-DCE [59]. (e) KinD [8]. (f) FOTV [12]. (g) LR3M [25]. (h) RetinexDIP [37]. (i) URetinex [60]. (j) Ours-IRCNN. (k) Ours. (l) Ground truth.
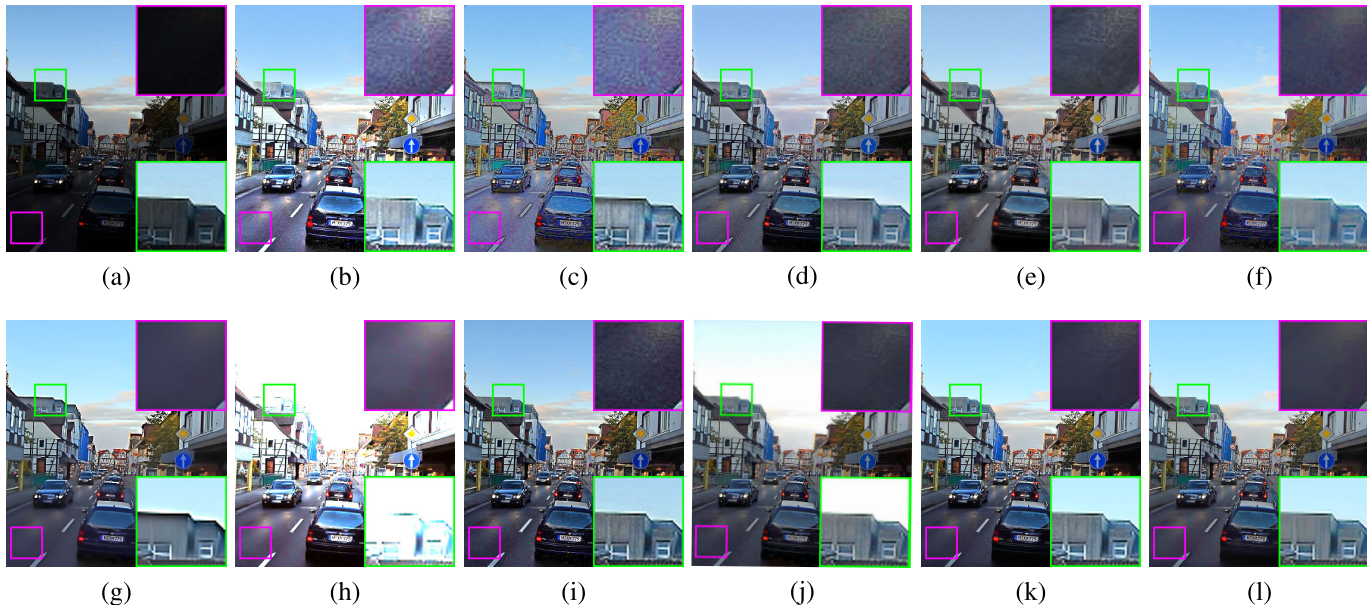




Fig. 7. Image enhancement results of existing methods on Img7. (a) Input. (b) LIME [10]. (c) RetinexNet [7]. (d) Zero-DCE [59]. (e) KinD [8]. (f) FOTV [12]. (g) LR3M [25]. (h) RUAS [40]. (i) RetinexDIP [37]. (j) URetinex [60]. (k) Ours-IRCNN. (l) Ours.
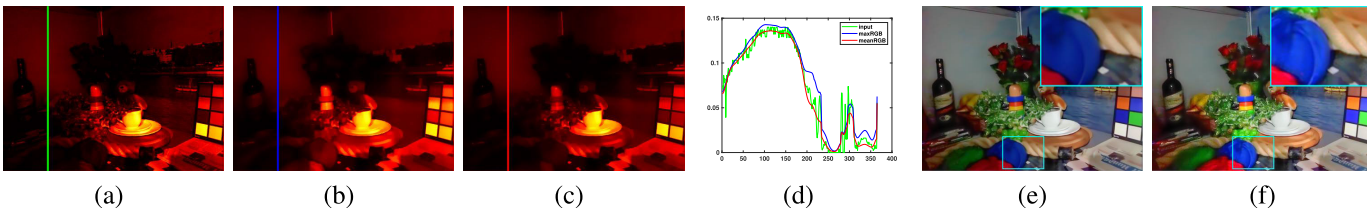


Fig. 8. Comparison between the meanRGB operator and maxRGB operator. Heatmap of (a) input, (b) maxRGB, and (c) meanRGB. (d) 1-D plots of the estimated $L$ map. Enhanced results of (e) maxRGB and (f) meanRGB.

compared to the regularization-based methods, learning-based models achieve better results. Since the performance of pure learning-based methods relies heavily on the number of training samples and this dataset only has limited data, their results are not completely satisfactory. Finally, among all the image enhancement models, our method obtains the best

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

10

IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS

TABLE III
AVERAGE NUMERICAL RESULTS ON 30 SYNTHETIC
LOW-LIGHT IMAGES WITH NOISE DISTORTION

| Method | PSNR | SSIM | MSE |
|---|---|---|---|
| RetinexNet [7] | 14.03±2.14 | 0.7059±0.1242 | 2870.4±1368.3 |
| KinD [8] | 15.33±2.39 | 0.7581±0.1005 | 2118.3±1132.8 |
| Zero-DCE [59] | 17.39±1.79 | 0.7576±0.0850 | 1278.6±493.9 |
| RUAS [40] | 15.09±1.69 | 0.7267±0.0870 | 2155.2±788.5 |
| URetinex [60] | 16.07±3.18 | 0.7605±0.1096 | 2025.6±1415.4 |
| Ours | **19.89±2.18** | **0.8328±0.0678** | **741.7±327.2** |

numerical scores in light of all metrics. In addition, our model has a relatively low std in the LOL dataset, suggesting that our model is more robust and consistent.

To further test the effectiveness of the proposed method in noise suppression, here we specifically compare it with five competitors that also consider noise suppression in image enhancement. The images are from the BSD dataset. Table III reports the quantitative results on the synthetic 30 low-light and noisy images. In Table III, our method achieves the highest PSNR and SSIM values and the lowest mse error.

*3) Initial Illumination Map:* Different from the methods based on alternating iteration, our model estimates $R$ and $L$ separately. The key to the success of the proposed framework is to accurately estimate $L$. In addition to the meanRGB operator mentioned in this article, the maxRGB operator is often used to estimate the initial illumination map, which is mathematically defined as

$$\hat{L}(x) = \max_{c \in \{R,G,B\}} L^c(x). \tag{28}$$

We compare the performance of two initialization operators in Fig. 8, where Fig. 8(a)–(c) shows the heatmaps of the input image and the estimated illumination obtained by maxRGB and meanRGB. As can be seen, both methods are noise-free, but the estimated illumination obtained by meanRGB is smoother than that obtained by maxRGB. The 1-D profile in Fig. 8(d) also implicates that the pixel profile of the illumination map obtained by meanRGB is more faithful to the input image. Furthermore, we can see in Fig. 8(e)–(f) that the meanRGB initial illumination yields a better enhancement result.

*4) Retinex Decomposition:* To validate the effectiveness of our decomposition strategy, we offer the decomposition results of our framework in Fig. 9, where 9(a)–(d) represents the input $S$, the reflectance layer $R^{(k)}$, illumination layer $L$, and the restored result, respectively. In our framework, the noise existing in the illumination component is transferred to the reflectance layer. Thus, the illumination layer should be spatially smooth and only contains simple structures. We only need to denoise the reflectance map. Fig. 9(b) spotlights that our method can effectively remove the noise in the reflectance map and preserve the edge and structure information. In addition, the estimated illumination map [Fig. 9(c)] is smooth and noise-free. Those observations conclude the effectiveness of the decomposition strategy and justify the utility of our plug-and-play framework for image enhancement.

*5) Runtime:* Table IV shows the runtime of our method and other state-of-the-arts for a single image. Note that the
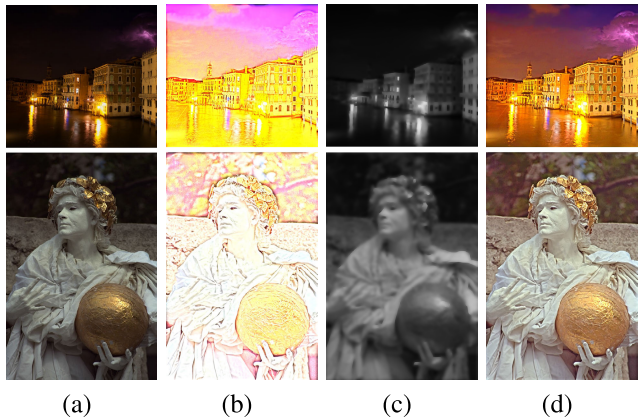


Fig. 9. Decomposition results generated by our framework. (a)–(d) Observed image, estimated reflectance $R^{(k)}$, illumination $L$, and enhanced version, respectively.

TABLE IV
RUNTIME OF DIFFERENT MODELS FOR A SINGLE IMAGE (IMG1)

| Method | Runtime |
|---|---|
| LIME [10] | 4.2861s |
| FOTV [12] | 4.7405s |
| LR3M [25] | 106.6291s |
| RetinexNet [7] | 1.8547s |
| Zero-DCE [59] | 2.3340s |
| KinD [8] | 1.6160s |
| RetinexDIP [37] | 24.4460s |
| RUAS [40] | 2.3995s |
| URetinex [60] | 1.1920s |
| Ours | 16.1763s |

proposed model is faster than LR3M and RetinexDIP but slower than other models. Although our model is not the fastest, it has moderate runtime. We argue that a higher time cost is worthwhile in order to acquire excellent enhancement performance. The most time-consuming part of our framework lies in deriving $L$. The denoising time of our denoiser only takes 0.2601 s.

## C. Interpretability of the Proposed Framework

Earlier experiments show that our designed denoiser can assist the framework to deliver advanced image enhancement performance. Now, we use the aforementioned post hoc dynamics analysis to track the denoising process to address problems such as what kind of information is used in denoising in our framework.

The directed graph is built based on the algorithm in Section IV-A. We demonstrate the information flow of denoising in Fig. 10. Each arrow connects the most related two pixels between images generated by two iterations. From Fig. 10, we draw one interesting observation. It can be seen that the information flow of denoising is highly local, ti.e., the restoration of some pixels is most influenced by its very surrounding pixels. This is surprising because the deeper layers of a CNN usually can extract abstract information, which can cover a large receptive field. Thus, the interaction between two pixels is expected to be local but not highly local. Since various CNN-based denoisers with interpretability can be inserted at will, we conjecture that if we insert a
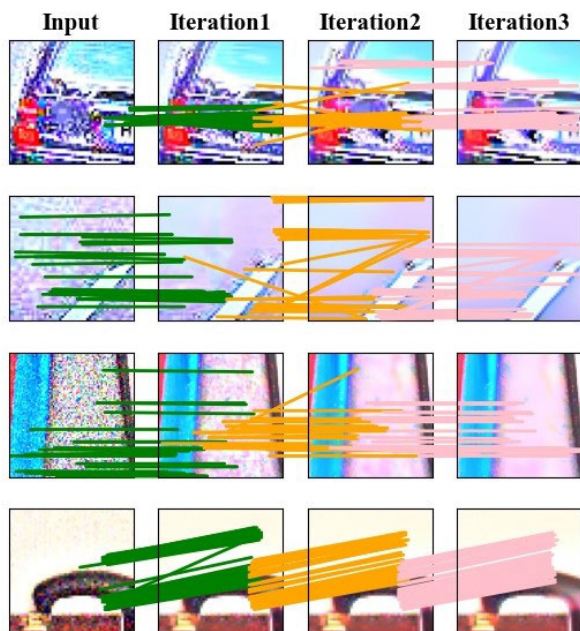
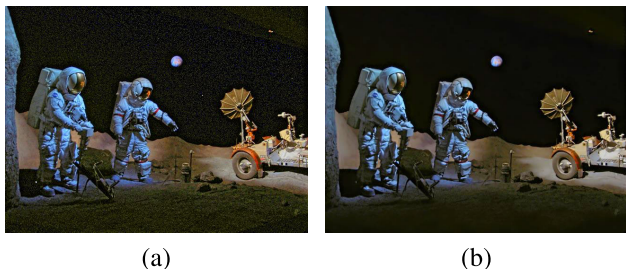Fig. 10. Information flow of the proposed framework.



Fig. 11. Results of our proposed framework with and without the denoiser. (a) W/O denoiser (NIQE: 3.6909). (b) With denoiser (NIQE: 2.8494).

transformer [66] into the denoising module, we might see the interaction in a much longer range.

## VI. ABLATION STUDY AND PARAMETRIC ANALYSIS

### A. Ablation Study

*1) Effectiveness of Denoiser:* To show the necessity of the inserted denoiser, we compare the results of our model with and without denoiser. Fig. 11 reports the comparable results. From it, we can find that the result generated by our proposed model without denoiser has heavy noise, color distortion, and poor numerical result. In contrast, the results created by the model with denoiser have a clear and sharper structure. Such results have demonstrated the effectiveness of the proposed plug-and-play framework with the inserted denoisers, which can alleviate color bias and structural detail loss. By increasing the coefficient of denoising, we can increase the degree of noise suppression in the reflectance map. According to the degree of image degradation, different coefficients of denoising can be employed to achieve the desirable performance.

*2) Effects of Different CNN-Based Denoisers:* We compare the enhancement effects of different denoisers to prove the robustness of our plug-and-play framework. Fig. 12 shows the enhancement results of four different denoisers, including IRCNN [46], FFDNet [67], CBDNet [68], and our CNN-based
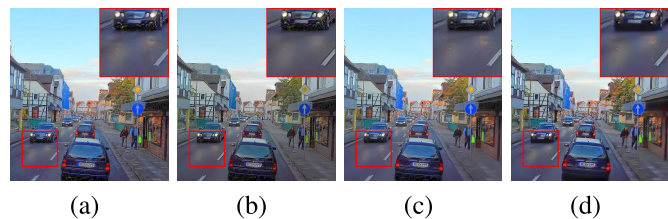


Fig. 12. Performance of the proposed framework with different denoisers. (a) CBDNet (NIQE: 5.6149). (b) FFDNet (NIQE: 6.1877). (c) IRCNN (NIQE: 4.8271). (d) Our CNN-based denoiser (NIQE: 4.7822).
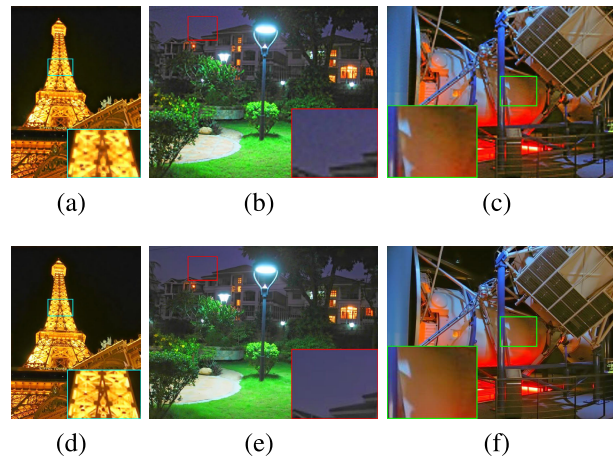




Fig. 13. Visual comparison on (a)–(c) wavelet denoiser and (d)–(f) ours.

denoiser. From Fig. 12, among all the denoisers, we can find that our denoiser achieves the best performance on both the visual effect and numerical result. Obviously, the performance of denoisers has a direct effect on the result of our proposed framework. However, the proposed method is a universal, feasible and superior framework, which allows the insertion of different denoisers based on the intrinsic characteristics of different images.

*3) Effect of an Explainable Denoiser:* In particular, we compare the effect of wavelet shrinkage denoiser (top row) and the Soft-AE (bottom row) with details in Fig. 13. The denoised results of wavelet shrinkage denoiser still suffer obvious punctuate noise (the enlarged area in the red and green boxes) and lose structural details (the enlarged area in the blue box). On the contrary, the Soft-AE satisfactorily suppresses the noise and simultaneously keeps structural details. Thus, we conclude that the Soft-AE has a better performance.

### B. Parametric Analysis

More favorably, for important hyperparameters such as $\alpha, \beta, \lambda, \gamma_1$, and $\gamma_2$, which greatly affect regularization and correction, we have discussed their sensitivity to the performance of the proposed framework in detail.

*1) Regularization Parameters—$\alpha$ and $\beta$:* The regularization parameters $\alpha$ and $\beta$ in (2) have direct effects on the performance of our model. Fig. 14 shows the visual effects under different settings of $\alpha$ and $\beta$. Fig. 14(b) shows the result with $\alpha = 0$, which means that the term $\|\nabla L\|_1$ does not work. Such a setting is not conducive to generating a smooth illumination map and leads to unsatisfactory enhancement results. Fig. 14(c) gives the result with $\beta = 0$, which means that the third term of (2) is inactive. Fig. 14(d) shows the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

12                                                                          IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS



Fig. 14.   Regularization parametric analysis. (a) Input. (b) Enhanced image with $\alpha = 0$. (c) Enhanced image with $\beta = 0$. (d) Enhanced image by our framework.
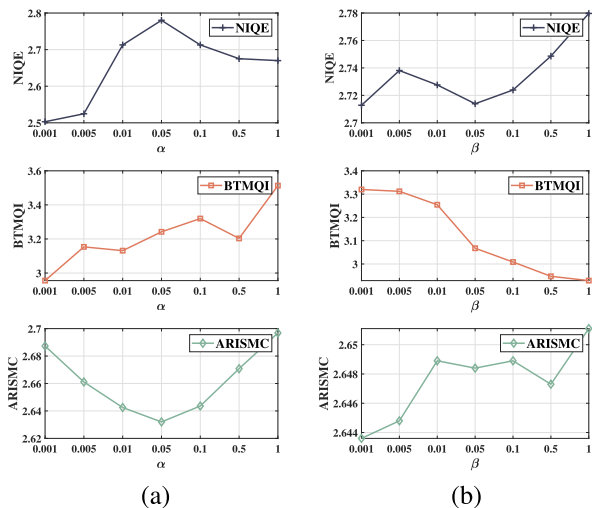


Fig. 15.   Effect of different regularization parameters on NIQE, BTQMI, and ARISMC. We fix (a) $\beta = 0.001$ and (b) $\alpha = 0.1$.
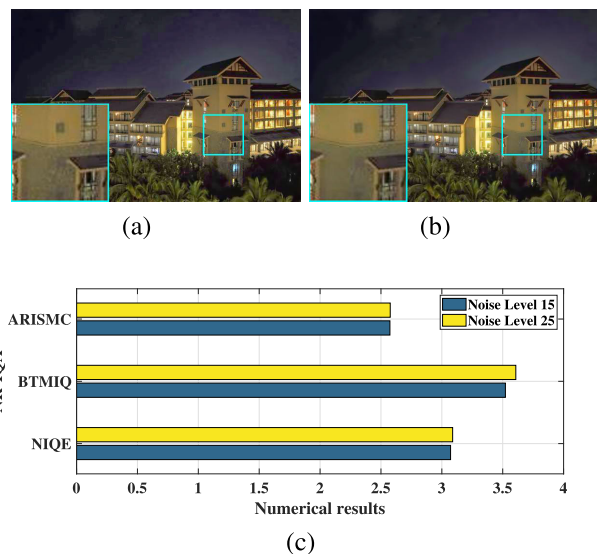


Fig. 16.   Quantitative and qualitative analysis of different $\omega$ values. (a) Noise level 15. (b) Noise level 25. (c) Numerical results for different noise levels.
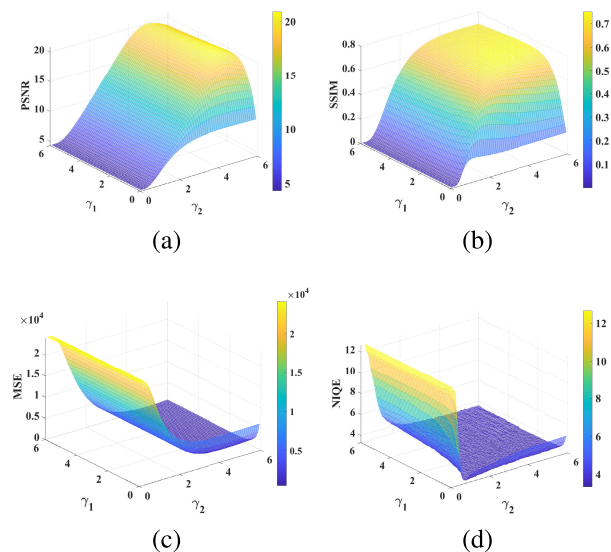


Fig. 17.   Impact of gamma correction parameters on (a) PSNR, (b) SSIM, (c) mse, and (d) NIQE.

result generated by our model with $\alpha = 0.02$ and $\beta = 0.001$. Such results demonstrate the usefulness of the second and third terms in (2). To further test the effects of $\alpha$ and $\beta$, we conduct the experiments by only changing one parameter at a time. Fig. 15 gives the experimental results (NIQE, BTMQI, and ARISMC) on Set12. From it, we can easily observe that different parameters have a direct impact on the performance of our method. Please note that lower NIQE, BTMQI, and ARISMC values represent better visual quality. As can be observed, NIQE always prefers smaller parameters but is not sensitive to $\beta$. BTMQI prefers smaller $\alpha$ and larger $\beta$. ARISMC prefers an intermediate $\alpha$ and is not sensitive to $\beta$.

*2) Noise Level—$\omega$:* The noise level of the denoiser usually determines the effect of denoising. Therefore, we studied the effect of different denoising levels on the enhanced results. Fig. 16 shows the quantitative and qualitative analyses of different $\omega$'s. It is obvious that the CNN-based denoiser with noise level 25 achieves a better visual result. However, the enhanced results obtained by the denoisers with different noise levels have little difference in numerical results. In order to pursue a better denoising performance, we tend to use a CNN-based denoiser with a noise level of 25.

*3) Gamma Correction—$\gamma_1$ and $\gamma_2$:* Gamma correction is a nonnegligible link in the Retinex method. Fig. 17 shows the numerical results with different $\gamma_1$ and $\gamma_2$ values on the test set of LOL. The result indicates that the numerical results are highly affected by $\gamma_2$. A lower $\gamma_2$ value corresponds to a higher NIQE value and a higher mse value. Moreover, PSNR

and SSIM are more inclined to a higher $\gamma_2$. On the contrary, the performance is not sensitive to the variation of $\gamma_1$. Therefore, to obtain stable and superior performance in low-light image enhancement, we adopt different settings of $\gamma_1$ and $\gamma_2$ on different datasets. For the LOL dataset, $\gamma_1$ and $\gamma_2$ are set to 1.5 and 4, respectively, while $\gamma_1 = 1$ and $\gamma_2 = 2.2$ are set on the Set12 dataset.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

WU et al.: RETINEX IMAGE ENHANCEMENT BASED ON SEQUENTIAL DECOMPOSITION 13

## VII. Conclusion

In this article, we have proposed a framework that uses a sequential decomposition strategy to solve $R$ and $L$ sequentially, thereby avoiding the alternating iteration and canceling the mutual interference between solving $R$ and $L$. This framework not only circumvents the reliance on large paired low-/normal-light data, a key problem encountered in image enhancement but also promotes interpretability. Quantitative and qualitative experiments have demonstrated the superiority of our method. However, we find that the visual effect of images is largely affected by the gamma correction parameters that depend on the degree of underexposure. Future work will be investigating the effect of gamma correction and exploring efficient gamma correction methods.

## References

[1] F.-L. Fan et al., "On a sparse shortcut topology of artificial neural networks," *IEEE Trans. Artif. Intell.*, vol. 3, no. 4, pp. 595–608, Aug. 2022.

[2] Y. Lei, H. Zhu, J. Zhang, and H. Shan, "Meta ordinal regression forest for medical image classification with ordinal labels," 2022, *arXiv:2203.07725.*

[3] Z. Huang, J. Chen, J. Zhang, and H. Shan, "Learning representation for clustering via prototype scattering and positive sampling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 6, pp. 7509–7524, Jun. 2023.

[4] S. M. Pizer et al., "Adaptive histogram equalization and its variations," *Comput. Vis., Graph., Image Process.*, vol. 39, no. 3, pp. 355–368, Sep. 1987.

[5] E. D. Pisano et al., "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms," *J. Digit. Imag.*, vol. 11, no. 4, pp. 193–200, Nov. 1998.

[6] M. K. Ng and W. Wang, "A total variation model for Retinex," *SIAM J. Imag. Sci.*, vol. 4, no. 1, pp. 345–365, Jan. 2011.

[7] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep Retinex decomposition for low-light enhancement," in *Proc. Brit. Mach. Vis. Conf.*, 2018, pp. 1–12.

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

[9] E. H. Land and J. J. McCann, "Lightness and Retinex theory," *J. Opt. Soc. Amer.*, vol. 61, no. 1, pp. 1–11, 1971.

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

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

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

[13] H. Chang, M. K. Ng, W. Wang, and T. Zeng, "Retinex image enhancement via a learned dictionary," *Opt. Eng.*, vol. 54, no. 1, Jan. 2015, Art. no. 013107.

[14] Z. Gu, F. Li, and X.-G. Lv, "A detail preserving variational model for image Retinex," *Appl. Math. Model.*, vol. 68, pp. 643–661, Apr. 2019.

[15] F. Lv, Y. Li, and F. Lu, "Attention guided low-light image enhancement with a large scale low-light simulation dataset," *Int. J. Comput. Vis.*, vol. 129, no. 7, pp. 2175–2193, Jul. 2021.

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

[17] M. Zhu, P. Pan, W. Chen, and Y. Yang, "EEMEFN: Low-light image enhancement via edge-enhanced multi-exposure fusion network," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 13106–13113.

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

[19] Y. Wang et al., "Progressive retinex: Mutually reinforced illumination-noise perception network for low-light image enhancement," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 2015–2023.

[20] F. Fan, J. Xiong, M. Li, and G. Wang, "On interpretability of artificial neural networks: A survey," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 5, no. 6, pp. 741–760, Nov. 2021.

[21] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Dec. 2013, pp. 945–948.

[22] F. Fan, M. Li, Y. Teng, and G. Wang, "Soft autoencoder and its wavelet adaptation interpretation," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1245–1257, 2020.

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

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

[25] X. Ren, W. Yang, W. Cheng, and J. Liu, "LR3M: Robust low-light enhancement via low-rank regularized retinex model," *IEEE Trans. Image Process.*, vol. 29, pp. 5862–5876, 2020.

[26] Y. Wang, Z. F. Pang, Y. Duan, and K. Chen, "Image Retinex based on the nonconvex TV-type regularization," *Inverse Problems Imag.*, vol. 15, no. 6, pp. 1381–1407, 2021.

[27] L. Liu, Z.-F. Pang, and Y. Duan, "Retinex based on exponent-type total variation scheme," *Inverse Problems Imag.*, vol. 12, no. 5, pp. 1199–1217, 2018.

[28] J. Liang and X. Zhang, "Retinex by higher order total variation $l_1$ decomposition," *J. Math. Imag. Vis.*, vol. 52, no. 3, pp. 345–355, 2015.

[29] T. J. Cooper and F. A. Baqai, "Analysis and extensions of the Frankle–McCann Retinex algorithm," *J. Electron. Imag.*, vol. 13, no. 1, pp. 85–92, 2004.

[30] E. Provenzi, L. De Carli, A. Rizzi, and D. Marini, "Mathematical definition and analysis of the Retinex algorithm," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 22, no. 12, pp. 2613–2621, 2005.

[31] E. H. Land, "An alternative technique for the computation of the designator in the Retinex theory of color vision," *Proc. Nat. Acad. Sci. USA*, vol. 83, no. 10, pp. 3078–3080, May 1986.

[32] J. McCann, "Lessons learned from Mondrians applied to real images and color gamuts," in *Proc. IS&T/SID Color Imag. Conf.*, 1999, pp. 1–8.

[33] B. K. P. Horn, "Determining lightness from an image," *Comput. Graph. Image Process.*, vol. 3, no. 4, pp. 277–299, Dec. 1974.

[34] J. Michel Morel, A. Belén Petro, and C. Sbert, "A PDE formalization of Retinex theory," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2825–2837, Nov. 2010.

[35] R. Liu, L. Ma, Y. Zhang, X. Fan, and Z. Luo, "Underexposed image correction via hybrid priors navigated deep propagation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 8, pp. 3425–3436, Aug. 2022.

[36] L. Ma, R. Liu, J. Zhang, X. Fan, and Z. Luo, "Learning deep context-sensitive decomposition for low-light image enhancement," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 10, pp. 5666–5680, Oct. 2022, doi: 10.1109/TNNLS.2021.3071245.

[37] 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.*, vol. 32, no. 3, pp. 1076–1088, Mar. 2022.

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

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

[40] 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. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10561–10570.

[41] W. Yang, W. Wang, H. Huang, S. Wang, and J. Liu, "Sparse gradient regularized deep Retinex network for robust low-light image enhancement," *IEEE Trans. Image Process.*, vol. 30, pp. 2072–2086, 2021.

[42] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2862–2869.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

14
IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS

[43] C. Huang, M. K. Ng, T. Wu, and T. Zeng, "Quaternion-based dictionary learning and saturation-value total variation regularization for color image restoration," *IEEE Trans. Multimedia*, vol. 24, pp. 3769–3781, 2022.

[44] F. Fang, J. Li, Y. Yuan, T. Zeng, and G. Zhang, "Multilevel edge features guided network for image denoising," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 9, pp. 3956–3970, Sep. 2021.

[45] K. Zhang, W. Zuo, and L. Zhang, "Deep plug-and-play super-resolution for arbitrary blur kernels," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1671–1681.

[46] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2808–2817.

[47] Y. Sun, J. Liu, and U. Kamilov, "Block coordinate regularization by denoising," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 380–390.

[48] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, Jul. 1995.

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

[50] E. H. Land, "The Retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108–129, Dec. 1977.

[51] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends® Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[52] Z. Ma et al., "Dirichlet process mixture of generalized inverted Dirichlet distributions for positive vector data with extended variational inference," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6089–6102, Nov. 2022, doi: 10.1109/TNNLS.2021.3072209.

[53] R. Du, J. Xie, Z. Ma, D. Chang, Y. Song, and J. Guo, "Progressive learning of category-consistent multi-granularity features for fine-grained visual classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 9521–9535, Dec. 2022, doi: 10.1109/TPAMI.2021.3126668.

[54] J. Kim, S. Kim, and M. Lee, "Convolutional neural network with biologically inspired ON/OFF ReLU," in *Proc. Int. Conf. Neural Inf. Process.*, 2015, pp. 316–323.

[55] W. Shang, K. Sohn, D. Almeida, and H. Lee, "Understanding and improving convolutional neural networks via concatenated rectified linear units," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 2217–2225.

[56] A. L. Maas et al., "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, 2013, vol. 30, no. 1, p. 3.

[57] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[58] D. Wang, F. Fan, Z. Wu, R. Liu, F. Wang, and H. Yu, "CTformer: Convolution-free Token2Token dilated vision transformer for low-dose CT denoising," 2022, *arXiv:2202.13517*.

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

[60] W. Wu, J. Weng, P. Zhang, X. Wang, W. Yang, and J. Jiang, "URetinex-Net: Retinex-based deep unfolding network for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5891–5900.

[61] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation," in *Proc. 19th IEEE Int. Conf. Image Process.*, Oct. 2012, pp. 965–968.

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

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

[64] K. Gu et al., "Blind quality assessment of tone-mapped images via analysis of information, naturalness, and structure," *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 432–443, Mar. 2016.

[65] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "No-reference image sharpness assessment in autoregressive parameter space," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3218–3231, Oct. 2015.

[66] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–15.

[67] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[68] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1712–1722.

**Tingting Wu** received the B.S. and Ph.D. degrees in mathematics from Hunan University, Changsha, China, in 2006 and 2011, respectively.

From 2015 to 2018, she was a Post-Doctoral Researcher with the School of Mathematical Sciences, Nanjing Normal University, Nanjing, China. From 2016 to 2017, she was a Research Fellow with Nanyang Technological University, Singapore. She is currently an Associate Professor with the School of Science, Nanjing University of Posts and Telecommunications, Nanjing. Her research interests include variational methods for image processing and computer vision, optimization methods and their applications in sparse recovery, and regularized inverse problems.

**Wenna Wu** received the B.S. degree in mathematics and applied mathematics from the Taiyuan Institute of Technology, Taiyuan, China, in 2019. She is currently pursuing the M.S. degree with the School of Science, Nanjing University of Posts and Telecommunications, Nanjing, China.

Her research interests include image processing and computer vision, machine learning, and inverse problems.
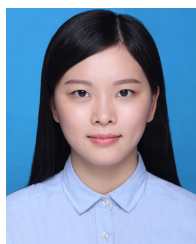
**Ying Yang** received the B.Sc. and Ph.D. degrees in computer science from Chongqing University, Chongqing, China, in 2016 and 2021, respectively.

She is currently a Post-Doctoral Research Fellow with the Department of Mathematics, The Chinese University of Hong Kong, Hong Kong. Her main research interests include image processing, quality evaluation, and multimedia security.

**Feng-Lei Fan** (Member, IEEE) is currently a Research Assistant Professor with the Department of Mathematics, The Chinese University of Hong Kong, Hong Kong. His research interests lie in deep learning theory and methodology.

Prof. Fan was a recipient of the IBM AI Horizon Fellowship and the 2021 International Neural Network Society Doctoral Dissertation Award.

**Tieyong Zeng** received the B.S. degree from Peking University, Beijing, China, in 2000, the M.S. degree from École Polytechnique, Palaiseau, France, in 2004, and the Ph.D. degree from the University of Paris XIII, Paris, France, in 2007.

He worked as a Post-Doctoral Researcher with ENS de Cachan, Cachan, France, from 2007 to 2008, and an Assistant/Associate Professor with Hong Kong Baptist University, Hong Kong, from 2008 to 2018. He is currently a Professor with the Department of Mathematics, The Chinese University of Hong Kong, Hong Kong. His research interests include image processing, machine learning, and scientific computing.