URetinex-Net: Retinex-based Deep Unfolding Network for Low-light Image Enhancement

Wenhui Wu[1] Jian Weng[2] Pingping Zhang[3] Xu Wang[2*] Wenhan Yang[4] Jianmin Jiang[2]
[1]College of Electronics and Information Engineering, Shenzhen University
[2]College of Computer Science and Software Engineering, Shenzhen University
[3]Department of Computer Science, City University of Hong Kong
[4]School of Electrical and Electronic Engineering, Nanyang Technological University

Abstract

Retinex model-based methods have shown to be effective in layer-wise manipulation with well-designed priors for low-light image enhancement. However, the commonly used hand-crafted priors and optimization-driven solutions lead to the absence of adaptivity and efficiency. To address these issues, in this paper, we propose a Retinex-based deep unfolding network (URetinex-Net), which unfolds an optimization problem into a learnable network to decompose a low-light image into reflectance and illumination layers. By formulating the decomposition problem as an implicit priors regularized model, three learning-based modules are carefully designed, responsible for data-dependent initialization, high-efficient unfolding optimization, and user-specified illumination enhancement, respectively. Particularly, the proposed unfolding optimization module, introducing two networks to adaptively fit implicit priors in data-driven manner, can realize noise suppression and details preservation for the final decomposition results. Extensive experiments on real-world low-light images qualitatively and quantitatively demonstrate the effectiveness and superiority of the proposed method over state-of-the-art methods. The code is available at `https://github.com/AndersonYong/URetinex-Net`.

1. Introduction

Images captured in a poor light environment always suffer from low contrast and low visibility, which pose challenges for both human visualization and numerous high-level vision tasks such as object detection [21, 23, 35]. To uncover the buried details in the low-light image and avoid degenerated performance of the subsequent vision tasks, researchers have made great efforts on contrast enhancement, texture recovering and noise removal for the low-light image. Especially for the low-light image enhancement (LLIE), many methods have been proposed, including histogram equalization [29], unsharp masking algorithms [7], Retinex-based methods [12, 27], multiple exposure fusion [5], and deep learning-based methods [11, 34].

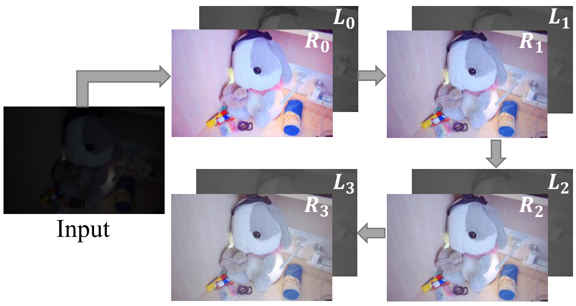

Figure 1. Decomposition results in each unfolding stage, where decomposed components step-wise get rid of degradation. Gamma correction is applied for the decomposed illumination layer for better visual effect.

Since Retinex theory well models color perception of human vision, LLIE methods based on Retinex theory have attracted much attention. As stated in Retinex theory, an image can be decomposed into two components, i.e., reflectance and illumination. Mathematically, the observed image I can be expressed by

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

where R, L and $\cdot$ denote reflectance, illumination and element-wise multiplication, respectively. In some early Retinex-based methods [16, 17], illumination is first estimated, and then reflectance is treated as the final enhanced results. Although details can be largely recovered from the input, it often leads to an unnatural and over-exposed look. Afterward, a number of model-based methods which have

*Corresponding author. wangxu@szu.edu.cn

good interpretability are proposed to solve the ill-posed decomposition problem in Eq. (1), where various hand-crafted priors are designed as the regularization terms introduced into models [12, 13, 22, 27, 30]. Then, to reveal the low-light image, illumination is further brightened up by Gamma correction. Designing explicit prior to fit data is the key to making models well-work, but it is challenging for model-based methods to be adaptive enough in various scenes. Furthermore, most model-based methods adopting conventional iteration optimization schemes are costly for a single image adjustment, which will hinder their development in real-time applications.

Due to these limitations existing in model-based methods, researchers take advantage of deep networks [3, 11, 15, 25, 34, 42] to restore low-light images in a data-driven manner. Among these learning-based methods, Retinex-based ones [34, 39, 41] use deep networks to estimate reflectance and illumination, and brighten up illumination. However, most of these methods perform denoising operations on reflectance after decomposition, resulting in the loss of details. Furthermore, learning-based methods suffer from a lack of interpretability and flexibility, which brings difficulties in analyzing the potential limitations of the designed networks.

To this end, we propose a Retinex-based deep unfolding network (URetinex-Net) to reveal low-light images in RGB color space. To integrate the strengths from model-based and learning-based methods, we formulate Retinex-based decomposition problem as an implicit priors regularized model, where robust regularization terms are inferred by learnable networks instead of using hand-crafted priors. Specifically, the energy function of the formulated model is split into four univariate subproblems via alternating half-quadratic splitting algorithm, and this optimization problem can be solved by iteratively minimizing four subproblems. Then, we unfold the optimization scheme into a deep network. For subproblems regarding to prior terms, two networks are introduced to adaptively fit implicit priors, while the others regarding to the fidelity term are solved by corresponding close-form solutions. During unfolding optimization, the decomposed reflectance and illumination step-wise get rid of degradation (see Fig. 1). Meanwhile, the formulated model avoids designing explicit prior terms. Furthermore, considering the important effect of initialization on optimization, we propose an initialization module to favor the optimization. Finally, we design an illumination adjustment module to flexible brighten up illumination map according to use-specified light level.

In summary, the contributions of this paper lie threefold:

- Based on traditional model-based methods, we propose a novel deep unfolding network for LLIE (URetinex-Net), consisting of three functionally clear modules corresponding to initialization, optimization, and illumination adjustment, respectively, which inherits the flexibility and interpretability from model-based methods.
- The optimization module in our proposed URetinex-Net unfolds optimization procedure into a deep network, which leverages the powerful model ability of learning-based methods to adaptively fit data-dependent priors.
- Extensive experiments on real-world datasets are conducted to demonstrate high efficiency and superiority of our URetinex-Net, which can realize noise suppression and details preservation for the final enhanced results.

2. Related Work

2.1. Retinex Based LLIE Methods

Model-Based Methods: Classic Retinex theory models the Human Visual System (HVS), which assumes that the observed color depends on the intrinsic components of the object itself and the extrinsic non-uniform light source falls onto the object. Naturally, the image can be decomposed into reflectance and illumination as denoted in Eq. (1).

Several Retinex decomposition models [9, 18, 27] have been proposed under variational frameworks. Then adjusting the estimated illumination, the target low-light image is restored. Afterward, several model-based methods whose energy functions are under Maximum a posteriori (MAP) framework are proposed [12, 13, 22, 30]. Guo *et al.* [12] proposed a structure-aware regularization model to refine the illumination map based on the initial one. In order to model degradation caused by noise, Li *et al.* [22] further introduced a noise term into the objective function to help remove noise while amplifying the details. Hao *et al.* [13] use Gaussian total variation as the regularization term to build decomposition model. In general, conventional model-based methods most rely on carefully designed hand-craft priors or certain statistical models. However, such priors are limited by model capacity when applied in various scenes.

Deep Learning-Based Methods: In the past decade, deep learning-based methods have provided promising results for LLIE problems [20]. Inspired by Retinex theory, Wei *et al.* [34] proposed an end-to-end trainable network named Retinex-Net, which includes a decomposition module and an illumination adjustment module. To realize denoising operation in Retinex-Net, BM3D [6] is used as a post-processing layer for reflectance, which implies that Retinex-Net can not handle heavy noise lying in extremely low-light images. Following [34], KinD [41] adopts a trainable denoising module for reflectance restoration. Moreover, a learnable mapping function is designed in the illumination adjustment module, in which images can be flexibly

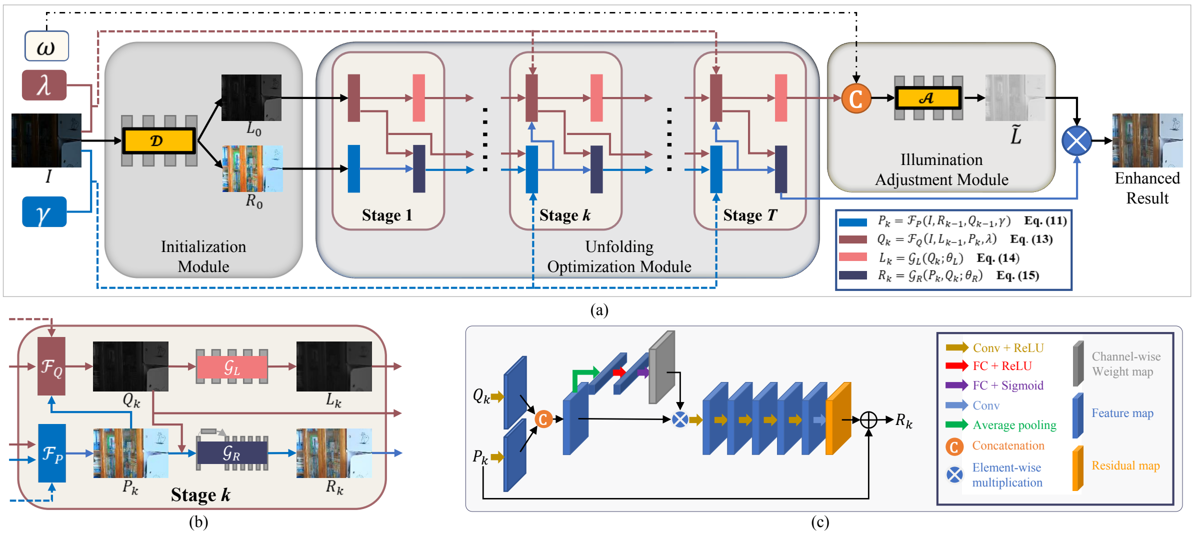

Figure 2. Illustration of our proposed URetinex-Net. (a) Overall framework of URetinex-Net, (b) details about each stage in URetinex-Net, (c) the specific network structure of $\mathcal{G}_{\mathcal{R}}$ applied in each stage. Particularly, URetinex-Net for the LLIE problem includes three learnable modules. By passing a target low-light image into initialization module, the initial reflectance and illumination are generated. After that, the unfolding optimization module refines the reflectance and illumination layers iteratively. Finally, the illumination adjustment module outputs the enhanced normal-light version according to the user-defined ratio.

restored under a user-specific light level. More recently, inspired by Retinex theory combined with maximum entropy, Zhang *et al.* [39] proposed a self-supervised framework utilizing histogram equalization operator to impose the constraint on reflectance.

Although these methods have shown remarkable performance on LLIE, they lack interpretability which will hinder their development on LLIE. Besides, based on the theory that reflectance depicting intrinsic components should be consistent under different light environments, most Retinex-based deep learning methods restore reflectance after decomposition, which will result in the loss of details [22].

2.2. Deep Unfolding Methods

Model-based LLIE methods are highly interpretable and flexible, while learning-based LLIE methods show superiority in learning complicated mapping in a data-driven manner. In addition, deep neural networks perform fast during inference, which is particularly computationally efficient. The unfolding (or unrolling) algorithm leveraging the strengths lying in model-based and learning-based methods has attracted much attention in the past decade.

Gregor and Lecun [10] first designed a time-unfolded algorithm to solve the iterative shrinkage and thresholding algorithm in the optimization of sparse coding, such that the proposed algorithm produces competitive performance within fewer iterations. Inspired by such optimization scheme, deep unfolding algorithms have made great impact on many significant image processing problems, such as super-resolution [33] [36], image denoising [4] [38], clutter suppression [31], and rain removal [8].

Recently, Liu *et al.* [24] proposed an unfolding framework for both illumination estimation and noise removal in an unsupervised manner, while the mutual connection between reflectance and illumination will be ignored in this way. Our method is different from it in two main aspects: (1) we tend to simultaneously estimate reflectance and illumination of the input in a unified framework; (2) our network can flexibly enhance illumination via a user-defined ratio.

3. The Proposed Unfolding Network

In this section, we first introduce the formulation of our proposed method and then present the details of the framework.

3.1. Problem Formulation

Classic Retinex-based model assumes that image can be decomposed into reflectance and illumination via Eq. (1), and various hand-crafted priors are developed to solve this ill-posed decomposition problem under MAP framework. Therefore, the reflectance and illumination can be obtained by minimizing the following regularized energy function:

$$E(R, L) = \|I - R \cdot L\|_F^2 + \alpha\Phi(R) + \beta\Psi(L), \quad (2)$$

where $\|\cdot\|_F$ denotes Frobenius norm, $\|I - R \cdot L\|_F^2$ is the fidelity term derived from Eq. (1), $\Phi(R)$ and $\Psi(L)$ are regularization terms denoting imposed priors over R and L, respectively, and α and β are trade-off parameters.

Generally, to facilitate optimization, the fidelity term and the regularization terms are handled separately, such that we introduce two auxiliary variables P and Q to approximate R and L, respectively. Accordingly, this leads to the following minimization problem:

$$\begin{aligned} \min_{P,Q,R,L} \; & \|I - P \cdot Q\|_F^2 + \alpha\Phi(R) + \beta\Psi(L) \\ \text{s.t.} \quad & P = R, \; Q = L. \end{aligned} \tag{3}$$

To deal with the equality constraints, two quadratic penalty terms are introduced, and problem is rewritten as

$$\begin{aligned} \min_{P,Q,R,L} \; & \|I - P \cdot Q\|_F^2 + \alpha\Phi(R) + \beta\Psi(L) \\ & + \gamma\|P - R\|_F^2 + \lambda\|Q - L\|_F^2, \end{aligned} \tag{4}$$

where γ and λ are penalty parameters.

To solve the problem in Eq. (4), the value of one variable is alternatively updated with those of the others fixed. Therefore, we partition the problem into four univariate subproblems, which can be optimized by the following alternating scheme:

$$P_k = \arg\min_P \; \|I - P \cdot Q_{k-1}\|_F^2 + \gamma\|P - R_{k-1}\|_F^2, \tag{5}$$

$$R_k = \arg\min_R \; \alpha\Phi(R) + \gamma\|P_k - R\|_F^2, \tag{6}$$

$$Q_k = \arg\min_Q \; \|I - P_k \cdot Q\|_F^2 + \lambda\|Q - L_{k-1}\|_F^2, \tag{7}$$

$$L_k = \arg\min_L \; \beta\Psi(L) + \lambda\|Q_k - L\|_F^2, \tag{8}$$

where k denotes the iteration index.

3.2. URetinex-Net Framework

Since it is difficult to design specific regularization terms $\Phi(R)$ and $\Psi(L)$, we take advantage of deep networks to adaptively fit physical priors of R and L. Therefore, based on the above-mentioned optimization scheme, we map the update steps to a deep unfolding network architecture, and propose a novel framework for LLIE. As shown in Fig. 2, the proposed URetinex-Net includes three modules, *i.e.* initialization module, unfolding optimization module, and illumination adjustment module.

3.2.1 Initialization Module

Initialization plays an important role during optimization. Random or all-zero initializations are widely used in the conventional optimization scheme, such as ADMM [1]. Considering that a reliable initialization is beneficial for performance, we hope to obtain an initialized illumination and reflectance with richer information instead of random values or all zeros.

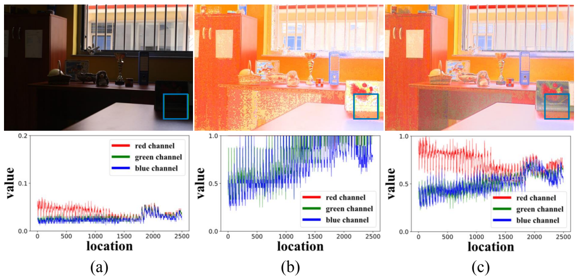

Figure 3. The statistic characteristics of a patch in a low-light image. (a) A low-light image, (b) reflectance initialized by $R_0 = \frac{I}{\max_{c\in\{R,G,B\}} I^{(c)}}$, and (c) reflectance initialized by our proposed module. Obviously, the rigid initialization changes statistical characteristics of the original low-light image, which is well preserved in our initialization modules.

Intuitively, to preserve the overall structure of I, initial illumination L_0 can be initialized by seeking the maximum value of three color channels [12], and initial reflectance R_0 can be accordingly derived by $R_0 = \frac{I}{L_0}$, where $\frac{(\cdot)}{(\cdot)}$ denotes element-wise division. However, initialization in such a rigid manner will cause color distortion. As illustrated in Fig. 3(b), statistical characteristics of intensity for three channels (*e.g.*, {R, G, B}) are changed.

Therefore, in order to reveal coarse details but avoid raising distortion, we propose a data-dependent initialization module named $\mathcal{D}$, which uses a fully convolutional (Conv) network to adaptively and simultaneously learn R_0 and L_0. The initialization module consists of three Conv+LeakyReLU layers, followed by a convolutional layer and ReLU layer. Kernel size is set to be 3×3 across the whole convolutional layers. For initializing two components of low-light images, the loss function is designed as follows:

$$\min_{R_0,L_0} \|I - R_0 \cdot L_0\|_1 + \mu\|L_0 - \max_{c\in\{R,G,B\}} I^{(c)}\|_F^2, \tag{9}$$

where $\|\cdot\|_1$ denotes l_1 norm, μ is the hyper-parameter, and $c \in \{R, G, B\}$ denotes the RGB channels. The first term is the reconstruction loss, and the second term aims to encourage the initialized illumination to preserve the overall structure of I.

Besides, due to the absence of ground-truth reflectance, normal-light images are utilized to generate clear reflectance, which should be close to that of low-light images. Therefore, the reflectance of the normal-light image is used as a reference in the following unfolding optimization module. Based on the network architecture of the initialization module, we further impose structure-aware smooth constraint [34] on the illumination of the normal-light image, and then loss function for decomposing a normal-light image is as follows:

$$\min_{\hat{R},\hat{L}} \|\hat{I}-\hat{R}\cdot\hat{L}\|_1+\hat{\mu}(\|\hat{L}-\max_{c\in\{R,G,B\}} \hat{I}^{(c)}\|_F^2+\|e^{-\epsilon\nabla\hat{I}}\cdot\nabla\hat{L}\|_1), \tag{10}$$

where $\hat{I}$, $\hat{R}$, and $\hat{L}$ represent the normal-light image, reflectance of $\hat{I}$ and illumination of $\hat{I}$, respectively, ϵ and $\hat{\mu}$ are hyper-parameters, and $\nabla(\cdot)$ denotes gradient operation. In the third term, the total variation of the illumination map is weighted by the gradient of the image, such that illumination can be spatially smoothed in a structure-aware manner.

3.2.2 Unfolding Optimization Module

The unfolding optimization module aims to iteratively solve four univariate subproblems to update corresponding variables within T iterations. Through mapping the updating steps to a deep neural network architecture, the inference is unfolded into T stages, each of which corresponds to one iteration where P, Q, L, and R are updated in an alternative manner. In the following, we sequentially present updating rules in the proposed module.

Updating rules for P and Q: Apparently, P-subproblem in Eq. (5) is a classic least square problem, whose closed-form solution can be obtained by differentiating Eq. (5) with respect to P and setting the derivative to 0. Hence, given the initialized reflectance and the closed-form solution to P-subproblem, the updating formula concerning P is as follows:

$$\begin{aligned} P_k &= \mathcal{F}_{\mathcal{P}}(I, R_{k-1}, Q_{k-1}, \gamma) \\ &= \begin{cases} R_0 & ,k=1, \\ \dfrac{\gamma R_{k-1} + I\cdot Q_{k-1}}{Q_{k-1}\cdot Q_{k-1}+\gamma\mathbf{1}}, & else, \end{cases} \end{aligned} \tag{11}$$

where $\mathbf{1}$ denotes all-ones matrix.

Similarly, updating Q can be done via solving Q-subproblem in Eq. (7). As restoring low-light images in RGB space, reflectance layers in three channels share the same illumination layer, such that illumination is assumed to be grayscale. Therefore, Eq. (7) is rewritten as

$$Q_k = \arg\min_{Q} \sum_{c\in\{R,G,B\}} \|I^{(c)}-P_k^{(c)}\cdot Q\|_F^2+\lambda\|Q-L_{k-1}\|_F^2, \tag{12}$$

whose closed-form solution can be found easily. Considering the initialized illumination, the updating formula for Q is obtained as

$$\begin{aligned} Q_k &= \mathcal{F}_{\mathcal{Q}}(I, L_{k-1}, P_k, \lambda) \\ &= \begin{cases} L_0 & ,k=1, \\ \dfrac{\lambda L_{k-1} + \sum_{c\in\{R,G,B\}} I^{(c)}\cdot P_k^{(c)}}{\sum_{c\in\{R,G,B\}} P_k^{(c)}\cdot P_k^{(c)}+\lambda\mathbf{1}}, & else. \end{cases} \end{aligned} \tag{13}$$

Updating rules for L and R: For L- and R-subproblems in Eqs. (8) and (6), instead of introducing hand-crafted priors to manually design specific loss functions, we develop learning-based methods to explore implicit priors from real-world data. In other words, two networks denoted as $\mathcal{G}_{\mathcal{L}}$ and $\mathcal{G}_{\mathcal{R}}$ are introduced to conduct updating for L and R, respectively.

Specifically, the network which is employed to fit the physical prior over L is expressed as

$$L_k = \mathcal{G}_{\mathcal{L}}(Q_k;\theta_L), \tag{14}$$

where Q_k is taken as the input of $\mathcal{G}_{\mathcal{L}}$, and θ_L denotes learnable parameters. We adopt a simple fully convolutional network with five Conv layers followed by ReLU activation to learn implicit priors over L, thereby prior can be learned from training data while avoiding to design a complicated regularization term.

Then, by passing the degraded P_k through a learnable denoising network $\mathcal{G}_{\mathcal{R}}$ in a similar way, a cleaner reflectance can be obtained. However, the level of distortion that appears in reflectance is highly correlative to the luminance of an illumination layer, where darker regions accompany by heavier degradation. Therefore, Q_k is aggregated with P_k as input fed to $\mathcal{G}_{\mathcal{R}}$ for the purpose of guiding reflectance restoration. Thus, the network which used for performing updating of R is expressed as

$$R_k = \mathcal{G}_{\mathcal{R}}(P_k, Q_k;\theta_R), \tag{15}$$

where θ_R denotes learnable parameters in $\mathcal{G}_{\mathcal{R}}$. In order to fuse P_k and Q_k to update R_k, the squeeze-and-excitation (SE) [14] block is employed. Details of $\mathcal{G}_{\mathcal{R}}$ are demonstrated in Fig. 2(c).

The unfolding optimization module is trained in an end-to-end manner, where parameters and network architectures of $\mathcal{G}_{\mathcal{R}}$ and $\mathcal{G}_{\mathcal{L}}$ are shared across different stages. Normal-light reflectance $\hat{R}$ generated by our initialization module is used as the reference during the optimization of unfolding networks. With regarding to the loss function, we adopt the summation of loss functions for reflectance and illumination, which includes the mean squared error (MSE) loss between P_k and R_k in each stage, the MSE loss, the structural similarity loss, and the perceptual loss between $\hat{R}$ and the final restored reflectance R_T, the MSE loss between Q_k and L_k in each stage, and the total variation loss for L_k in each stage. The loss function for the unfolding optimization module is as follows:

$$\begin{aligned} \min_{R_T,L_T} \; & \sum_{k=1}^{T}(\gamma_k\|P_k-R_k\|_F^2+\lambda_k\|Q_k-L_k\|_F^2) \\ & +\beta\|\nabla L_T\|_1+\alpha(\|\phi(\hat{R})-\phi(R_T))\|_1 \\ & +\|\hat{R}-R_T\|_F^2+(1-\text{SSIM}(\hat{R},R_T))), \end{aligned} \tag{16}$$

where T denotes the total number of stages, $\phi(\cdot)$ denotes the high-level feature extractor which is pre-trained on ImageNet by VGG19 network, SSIM$(\cdot)$ represents the structural similarity loss, and γ_k, λ_k, α, and β are trade-off parameters, respectively.

Apparently, even in the deep neural network architecture, the proposed unfolding optimization module has good interpretability, where $\mathcal{F}_{\mathcal{P}}$, $\mathcal{F}_{\mathcal{Q}}$, $\mathcal{G}_{\mathcal{L}}$, and $\mathcal{G}_{\mathcal{R}}$ all have clear meanings. What's more, it avoids explicit regularization designing and adaptively restores illumination and reflectance in the deep-learning manner.

3.2.3 Illumination Adjustment Module

In practice, there is no ground-truth light level for image enhancement, and flexibly adjusting illumination is necessary for fitting different practical requirements. For LLIE, Gamma correction is widely used to brighten up the illumination map [9, 22, 30], i.e., $\tilde{L} = L^{\frac{1}{\gamma}}$, where $\tilde{L}$ represents adjusted map and the changeable factor γ is empirically set to $1/2.2$. However, as suggested in [41], compared with gamma correction, illumination adjustment in the learning manner is more corroborative with actual situations. To this end, we propose an illumination adjustment module, which takes low-light illumination L and user-specific enhancement ratio ω as input, expressed as

$$\tilde{L} = \mathcal{A}(L, \omega; \theta_A), \tag{17}$$

where θ_A denotes learnable parameters of $\mathcal{A}$. Note that $\omega > 1$ denotes the brighten-up case.

As the distribution of the illumination map is non-uniform, we expand ω to a map, whose size is the same as that of L. Thus, the adjusted illumination $\tilde{L}$ is then obtained by passing the concatenation of ω and L through $\mathcal{A}$. In the concern of computational efficiency, we adopt the same lightweight network structure as the initialization module. To maintain the consistency and adjust illumination smoothly, the convolutional kernel size in this module is enlarged to 5×5.

The loss function for illumination adjustment module is as follows:

$$\min_{\tilde{L}} \|\nabla \tilde{L} - \nabla L\|_1 + \|R \cdot \tilde{L} - \hat{I}\|_F^2 + (1 - \mathrm{SSIM}(R \cdot \tilde{L}, \hat{I})), \tag{18}$$

where $\|\nabla \tilde{L} - \nabla L\|_1$ aims to enforce the refined illumination map $\tilde{L}$ to maintain the consistency with L, and the other two terms are both reconstruction losses to guarantee the fidelity constraints for learning $\tilde{L}$. During training, the value of ω is defined by $\frac{mean(\hat{L})}{mean(L)}$. While for inference, the enhancement ratio ω is specified by user.

Although KinD [41] also develops a flexible illumination adjustment net, it depends on the decomposed illumination map of the normal-light image, which may destroy the local structure of the input L. Instead, the proposed module aims to maintain structural consistency with low-light illumination. Moreover, the last two terms in loss function integrate reflectance component, which aims to constrain the refined illumination can naturally reconstruct the normal-light images.

4. Experiments

4.1. Implementation Details and Data sets

To evaluate the performance of our proposed method, we train and test our model on LOL dataset [34], which contains 500 low/normal-light image pairs, and is captured at various exposure times from the real world. We follow the setting of training data as [41]. For a more convincing comparison, we extend our model to SICE dataset [2], which contains 589 natural scenes with multi-level exposure, and randomly select 108 under/normal exposure image pairs from it. Furthermore, we adopt MEF dataset [19] for visual comparison to demonstrate the efficiency of our proposed method.

URetinex-Net is trained separately for each module, where the batch size is set to be 4. We use a small patch size (48×48) for training our unfolding module in the concern of efficiency. Through the ablation experiment, we find that $T = 3$ has already achieved promising results, such that T is empirically set to 3 in the following experiments. Each module is trained using Adam optimizer with a learning rate of 0.0001 and decaying by 10 after 30K iterations. Heuristically, μ, $\hat{\mu}$, α, β and ϵ are set to 0.1, 0.1, 1, 20 and 10 respectively. According to [36], penalty parameters λ and γ are expected to iteratively increase for stable convergence. Here, we initially set λ and γ as 0.5 and 0.1 respectively, both of which are increased by 0.05 in each stage. All experiments are conducted on an NVIDIA Tesla V100 GPU under PyTorch [28] framework.

4.2. Comparison with the state-of-the-art

We qualitatively and quantitatively compare with five traditional Retinex-based methods, including NPE [32], SRIE [9], LIME [12], RRM [22], and LR3M [30]. Furthermore, to verify the efficiency of our model in learning implicit prior from data, we compare with state-of-the-art (SOTA) learning-based methods including Retinex-Net [34], KinD [41], Zero-DCE [11], RUAS [24], AGLL-Net [26] and KinD++ [40]. Retinex-Net, KinD, KinD++ and RUAS have been trained on LOL dataset, such that we utilize their provided models for evaluation. Otherwise, for Zero-DCE, we reorganize the training data with the cropped images in LOL training dataset and fine-tuned it via the expanded dataset following their default settings. To measure the differences in color, structure, and high-level feature

(a) Input (b) LIME (c) RRM (d) Zero-DCE (e) KinD++ (f) RUAS (g) Ours (h) Ground-truth

Figure 4. Visual comparison on SICE dataset with the SOTA LLIE methods. The red and green boxes indicate obvious differences.

Table 1. Quantitative comparison on LOL and SICE datasets. The best and the second best results are boldfaced and underlined.

Dataset	Method	MAE↓	PSNR↑	SSIM↑	LPIPS↓
LOL	NPE [32]	0.1290	16.9697	0.4818	2.0607
	SRIE [9]	0.2571	11.8552	0.4937	1.8616
	LIME [12]	0.1200	16.7586	0.4440	2.0601
	RRM [22]	0.2080	13.8765	0.6696	1.7132
	LR3M [30]	0.3086	10.2228	0.4343	2.3669
	Retinex-Net [34]	0.1256	16.7740	0.4285	2.3346
	KinD [41]	<u>0.0804</u>	20.8665	0.8022	<u>1.4091</u>
	Zero-DCE [11]	0.1370	16.7955	0.5573	2.0038
	RUAS [24]	0.1534	18.2260	0.7170	1.9254
	KinD++ [40]	**0.0679**	<u>21.3003</u>	<u>0.8226</u>	1.4899
	AGLLNet [26]	0.1268	20.2400	0.7900	1.6383
	URetinex-Net	0.0832	**21.3282**	**0.8348**	**1.2234**
SICE	NPE [32]	0.1486	16.1973	0.7177	1.2848
	SRIE [9]	0.2058	14.0580	0.6420	**1.2492**
	LIME [12]	0.1472	15.4572	0.7001	1.3935
	RRM [22]	0.1601	15.9489	0.6892	1.5084
	LR3M [30]	0.1940	14.2086	0.6073	1.9482
	Retinex-Net [34]	0.1382	15.9080	0.7052	1.5377
	KinD [41]	0.1151	17.7330	0.7588	1.6684
	Zero-DCE [11]	0.1330	16.9090	0.7554	<u>1.2613</u>
	RUAS [24]	0.2628	10.9878	0.5451	2.5565
	KinD++ [40]	**0.1060**	<u>18.5942</u>	<u>0.7762</u>	1.4827
	AGLLNet [26]	0.1463	16.0006	0.7336	1.5196
	URetinex-Net	<u>0.1068</u>	**18.9467**	**0.7808**	1.2744

similarity, we use MAE, PSNR, SSIM, and LPIPS [37] as metrics. The higher the PNSR and SSIM values, the better the image quality. On the contrary, the worse the MAE and LPIPS values, the better the image quality.

Quantitative results on LOL dataset are shown in Table 1. It is obvious to learn that our method outperforms others on most of the metrics except slightly worse on MAE than KinD and KinD++. We perform much better results than all other methods in terms of PSNR, SSIM and LPIPS, indicating the effectiveness of the proposed method. Visual comparisons are shown in the supplementary materials.

To validate the generalisation performance of the proposed URetinex-Net, we further evaluate on SICE dataset without retraining or fine-tuning. Results are reported in Table 1 and Fig. 4. Compared to other methods, URetinex-Net generalises better to unseen scenes. It can be observed from Fig. 4 that some models, *e.g.*, NPE, SRIE, LIME, Retinex-Net and Zero-DCE introduce and even amplify noise after enhancement, such that they suffer from severe noise distortion and color degradation when brightening dark regions. RRM introduce a noise term in its objective function with a carefully designed regularization term, but it fails to preserve small details. KinD++ can remove noise but introduce unnatural looking due to over-sharpening (see the wall behind the flower in Fig. 4(e)), which leads to an unnatural look. RUAS results in over-exposure (see the leaves in Fig. 4(f)) due to the ignorance of reflectance in their framework. AGLLNet is a multi-branch CNN based architecture without Retinex-based setting, while apparently our proposed model outperforms it in all the metrics, which demonstrates the effectiveness of our Retinex-based unfolding strategy.

Moreover, we give an extensive visual comparison on MEF dataset. We report the top-5 best methods in terms of performance and speed in Fig. 5. Although traditional model-based methods can somehow restore low-light images, they are time-consuming in the iterative optimization procedure. In contrast, our method saves more time during inference. Furthermore, Zero-DCE as a learning-based LLIE method has a fast processing speed, but it has a limited capacity to achieve noise suppression and reach a satisfactory effect. KinD++ and RUAS remove noise in a post-processing manner which may bring other problems such as losing details, blurring or even worse image quality. In comparison to all of these approaches, our model is capable of noise suppression and detail preservation while sufficiently revealing low-light images. Demonstrating that, compared to meticulously hand-crafted planned priors, our unfolding module can impose a more robust implicit prior. In particular, our model shows the unique advantage of details recovering, which illustrates the superiority of our unfolding optimization. More comparison results and analysis can be found in the supplementary materials.

4.3. Ablation Study

The results of the ablation study are reported in Table 2. We first analyze the effectiveness of our initialization module by comparing with a rigid initialization manner (*e.g.*, $L_0 = \max_{c\in\{R,G,B\}} \hat{I}^{(c)}$). Then, to investigate the effectiveness of fusing illumination layer for the guidance of learning regularization term on reflectance, we remove Q_k from the input of network $\mathcal{G}_{\mathcal{R}}$, while the rest of the setups are as same as URetinex-Net. In order to illustrate the effec-

(a) Input (b) LIME (0.062) (c) RRM (8.550) (d) Zero-DCE (0.002) (e) KinD++ (0.163) (f) RUAS (0.005) (g) Ours (0.027)

Figure 5. Visual Comparison on MEF dataset. Average time (s) of processing these 3 inputs are reported in brackets. Zoom in for best view.

Table 2. Quantitative results of ablation study in terms of PSNR, SSIM, LPIPS and inference time on LOL dataset, where IM, UOM and IG are abbreviations for the proposed initialization module, unfolding optimization module and illumination guiding, respectively. Noted that time is represented in millisecond (ms).

Method	IM	UOM	IG	T	PSNR↑	SSIM↑	LPIPS↓	Time↓
w/o IM	×	✓	✓	3	20.0992	0.7304	1.9847	34.6
w/o UOM	✓	×	✓	-	20.7572	0.8252	1.2675	36.6
w/o IG	✓	✓	×	3	19.6109	0.8173	1.3261	35.8
Ours	✓	✓	✓	1	20.1222	0.8273	1.3165	**14.4**
Ours*	✓	✓	✓	3	21.3282	**0.8348**	**1.2234**	36.7
Ours	✓	✓	✓	5	**21.5767**	0.8334	1.2412	58.7
Ours	✓	✓	✓	7	21.4345	0.8332	1.2332	80.8

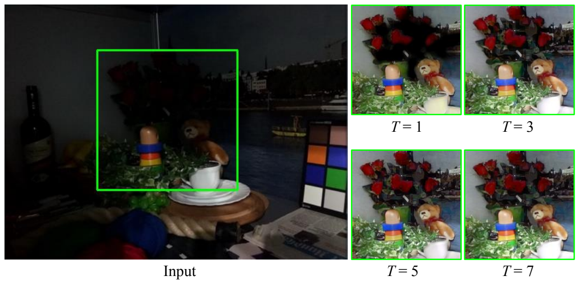

Figure 6. Ablation study of the importance of our unfolding module with different T.

tiveness of our unfolding optimization module, we further simply stack the network of $\mathcal{G}_{\mathcal{R}}$ and $\mathcal{G}_{\mathcal{L}}$ for T times and discard the unfolding optimization, which keeps the same network capacity as URetinex-Net. Finally, we investigate the performance of our URetinex-Net under different choices of stage T. The fifth row in Table 2 shows our default setting as the anchor. The first row shows that removing our initialization module leads to performance significantly drops. This is because rigid initialization may destroy the statistical characteristics of intensity for three channels of the input image, further impacting the learning capacity of our unfolding module. The second row indicates that the performance drops when discarding the unfolding optimization module, which further demonstrates the importance of our unfolding optimization module. In the third row, we observe that the performance degrades when removing the the guidance of the illumination layer. The results of the unfolding stage selection are shown in the final four rows of Table 2. Obviously, the model without unfolding mechanism (i.e., $T = 1$) performs worst in PSNR, SSIM and LPIPS. Through visual comparison in Fig. 6, we can easily observe that poor details preserving and color distortion remains on the enhanced results with $T = 1$, while a cleaner one can be obtained by using unfolding optimization. Based on the trade-off between the image quality and inference time, we select $T = 3$ as our default setting.

5. Conclusion

In this paper, we proposed a URetinex-Net for real-world low-light image enhancement. Based on the optimization procedure of conventional model-based methods, we first formulated Retinex decomposition problem as an implicit priors regularized model, and then unfolded the update steps in the optimization into a deep neural network. By taking advantage of learning-based methods, the estimated reflectance and illumiation can adaptively fit the data-dependent priors. Extensive experiments results verify that the proposed URetinex-Net can high-efficiently enhance low-light images with successfully noise suppressing and details preserving.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China (Grants 61871270 and 62006158), and in part by the Shenzhen Natural Science Foundation under Grants JCYJ20200109110410133, 20200812110350001 and 20200810150732001.

References

[1] Stephen Boyd, Neal Parikh, and Eric Chu. *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011. 4

[2] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018. 6

[3] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018. 2

[4] Yunjin Chen and Thomas Pock. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE transactions on pattern analysis and machine intelligence*, 39(6):1256–1272, 2016. 3

[5] Chunyang Cheng, Xiao-Jun Wu, Tianyang Xu, and Guoyang Chen. Unifusion: A lightweight unified image fusion network. *IEEE Transactions on Instrumentation and Measurement*, 70:1–14, 2021. 1

[6] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising with block-matching and 3d filtering. In *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*, volume 6064, page 606414. International Society for Optics and Photonics, 2006. 2

[7] Guang Deng. A generalized unsharp masking algorithm. *IEEE transactions on Image Processing*, 20(5):1249–1261, 2010. 1

[8] Ying Ding, Xinwei Xue, Zizhong Wang, Zhiying Jiang, Xin Fan, and Zhongxuan Luo. Domain knowledge driven deep unrolling for rain removal from single image. In *2018 7th International Conference on Digital Home (ICDH)*, pages 14–19. IEEE, 2018. 3

[9] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2782–2790, 2016. 2, 6, 7

[10] Karol Gregor and Yann LeCun. Learning fast approximations of sparse coding. In *Proceedings of the 27th international conference on international conference on machine learning*, pages 399–406, 2010. 3

[11] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1780–1789, 2020. 1, 2, 6, 7

[12] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016. 1, 2, 4, 6, 7

[13] Shijie Hao, Xu Han, Yanrong Guo, Xin Xu, and Meng Wang. Low-light image enhancement with semi-decoupled decomposition. *IEEE transactions on multimedia*, 22(12):3025–3038, 2020. 2

[14] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018. 5

[15] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, 30:2340–2349, 2021. 2

[16] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing*, 6(7):965–976, 1997. 1

[17] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. Properties and performance of a center/surround retinex. *IEEE transactions on image processing*, 6(3):451–462, 1997. 1

[18] Ron Kimmel, Michael Elad, Doron Shaked, Renato Keshet, and Irwin Sobel. A variational framework for retinex. *International Journal of computer vision*, 52(1):7–23, 2003. 2

[19] Chulwoo Lee, Chul Lee, Young-Yoon Lee, and Chang-Su Kim. Power-constrained contrast enhancement for emissive displays based on histogram equalization. *IEEE Transactions on Image Processing*, 21(1):80–93, 2012. 6

[20] Chongyi Li, Chunle Guo, Ling-Hao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. Low-light image and video enhancement using deep learning: a survey. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (01):1–1, 2021. 2

[21] Chengxi Li, Xiangyu Qu, Abhiram Gnanasambandam, Omar A. Elgendy, Jiaju Ma, and Stanley H. Chan. Photon-limited object detection using non-local feature matching and knowledge distillation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 3976–3987, October 2021. 1

[22] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing*, 27(6):2828–2841, 2018. 2, 3, 6, 7

[23] Jiaying Liu, Dejia Xu, Wenhan Yang, Minhao Fan, and Haofeng Huang. Benchmarking low-light image enhancement and beyond. *International Journal of Computer Vision*, 129(4):1153–1184, 2021. 1

[24] Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10561–10570, 2021. 3, 6, 7

[25] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017. 2

[26] Feifan Lv, Yu Li, and Feng Lu. Attention guided low-light image enhancement with a large scale low-light simulation dataset. *International Journal of Computer Vision*, 129(7):2175–2193, 2021. 6, 7

[27] Michael K Ng and Wei Wang. A total variation model for retinex. *SIAM Journal on Imaging Sciences*, 4(1):345–365, 2011. 1, 2

[28] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32:8026–8037, 2019. 6

[29] Stephen M Pizer. Contrast-limited adaptive histogram equalization: Speed and effectiveness. In *Proceedings of the First Conference on Visualization in Biomedical Computing, Atlanta, Georgia*, volume 337, 1990. 1

[30] Xutong Ren, Wenhan Yang, Wen-Huang Cheng, and Jiaying Liu. Lr3m: Robust low-light enhancement via low-rank regularized retinex model. *IEEE Transactions on Image Processing*, 29:5862–5876, 2020. 2, 6, 7

[31] Oren Solomon, Regev Cohen, Yi Zhang, Yi Yang, Qiong He, Jianwen Luo, Ruud JG van Sloun, and Yonina C Eldar. Deep unfolded robust pca with application to clutter suppression in ultrasound. *IEEE transactions on medical imaging*, 39(4):1051–1063, 2019. 3

[32] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9):3538–3548, 2013. 6, 7

[33] Zhaowen Wang, Ding Liu, Jianchao Yang, Wei Han, and Thomas Huang. Deep networks for image super-resolution with sparse prior. In *Proceedings of the IEEE international conference on computer vision*, pages 370–378, 2015. 3

[34] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 1, 2, 5, 6, 7

[35] Xin Xu, Shiqin Wang, Zheng Wang, Xiaolong Zhang, and Ruimin Hu. Exploring image enhancement for salient object detection in low light images. 17(1s), 2021. 1

[36] Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding network for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3217–3226, 2020. 3, 6

[37] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018. 7

[38] Xiaoshuai Zhang, Yiping Lu, Jiaying Liu, and Bin Dong. Dynamically unfolding recurrent restorer: A moving endpoint control method for image restoration. *arXiv preprint arXiv:1805.07709*, 2018. 3

[39] Yu Zhang, Xiaoguang Di, Bin Zhang, and Chunhui Wang. Self-supervised image enhancement network: Training with low light images only. *arXiv preprint arXiv:2002.11300*, 2020. 2, 3

[40] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129(4):1013–1037, 2021. 6, 7

[41] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia*, pages 1632–1640, 2019. 2, 6, 7

[42] Zhaoyang Zhang, Yitong Jiang, Jun Jiang, Xiaogang Wang, Ping Luo, and Jinwei Gu. Star: A structure-aware lightweight transformer for real-time image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4106–4115, October 2021. 2