

ReLLIE: Deep Reinforcement Learning for Customized Low-Light Image Enhancement

Rongkai Zhang*

Lanqing Guo*

rongkai002@e.ntu.edu.sg

lanqing001@e.ntu.edu.sg

Nanyang Technological University

Siyu Huang

Nanyang Technological University

siyu.huang@ntu.edu.sg

Bihan Wen†

Nanyang Technological University

bihan.wen@ntu.edu.sg

ABSTRACT

Low-light image enhancement (LLIE) is a pervasive yet challenging problem, since: 1) low-light measurements may vary due to different imaging conditions in practice; 2) images can be enlightened subjectively according to diverse preference by each individual. To tackle these two challenges, this paper presents a novel deep reinforcement learning based method, dubbed ReLLIE, for customized low-light enhancement. ReLLIE models LLIE as a markov decision process, *i.e.*, estimating the pixel-wise image-specific curves sequentially and recurrently. Given the reward computed from a set of carefully crafted non-reference loss functions, a lightweight network is proposed to estimate the curves for enlightening of a low-light image input. As ReLLIE learns a policy instead of one-one image translation, it can handle various low-light measurements and provide customized enhanced outputs by flexibly applying the policy different times. Furthermore, ReLLIE can enhance real-world images with hybrid corruptions, *e.g.*, noise, by using a plug-and-play denoiser easily. Extensive experiments on various benchmarks demonstrate the advantages of ReLLIE, comparing to the state-of-the-art methods. (Code is available: <https://github.com/GuoLanqing/ReLLIE>.)

CCS CONCEPTS

- Computing methodologies → Image processing; Sequential decision making.

KEYWORDS

low-light image enhancement; deep reinforcement learning

ACM Reference Format:

Rongkai Zhang, Lanqing Guo, Siyu Huang, and Bihan Wen. 2021. ReLLIE: Deep Reinforcement Learning for Customized Low-Light Image Enhancement. In *Proceedings of the 29th ACM International Conference on Multimedia (MM '21), October 20–24, 2021, Virtual Event, China*. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3474085.3475410>

*Both authors contributed equally to this research.

†Bihan Wen is the corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MM '21, October 20–24, 2021, Virtual Event, China

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8651-7/21/10...\$15.00

<https://doi.org/10.1145/3474085.3475410>

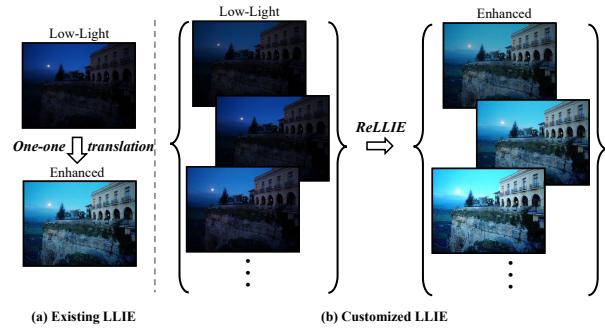


Figure 1: (a) Most of the existing LLIE methods produce one-one image translation; (b) We propose a customized LLIE scheme using the ReLLIE method.

1 INTRODUCTION

Low-light images captured under insufficient lighting conditions are pervasive in real-life scenarios due to inevitable environmental/technical constraints. Suffering from compromised aesthetic quality and unsatisfactory transmission of information, such low-light images are forbidden from many computer vision applications which therefore motivate plenty of low-light image enhancement (LLIE) methods [4, 6, 8, 10, 16, 30, 31]. Based on Retinex theory [12, 24, 32], a low-light image can be modeled by the following degradation process:

$$S_{low} = R \circ I_{low} + n_{add}, \quad (1)$$

where S_{low} is the low-light image, R denotes the underlying reflectance, I_{low} is the insufficient illumination, n_{add} is the additive noise, and \circ denotes the element-wise multiplication. The LLIE task aims to recover the “optimal” illumination I_{opt} from the observation S_{low} with the consistent reflectance R , meanwhile, suppressing the noise n_{add} . Most of the existing methods establish a one-one image translation model under the assumption that there exists only one deterministic output for an input. However, as an intrinsic nature, the LLIE task is complicated in practice, since both S_{low} and I_{opt} may be diverse for different individuals/applications. As shown in Fig. 1, an LLIE method should be more customized that it can 1) handle inputs S_{low} with *varied degrees of degeneration* (which are possibly different from the training data) and 2) provide candidate outputs with *different subjective I_{opt}* so as to meet the preference of different users.

In this paper, we present a deep reinforcement learning (DRL) based method named ReLLIE, to achieve more customized LLIE results. Instead of simply performing one-one paired image translation, ReLLIE reformulates LLIE as a *sequential image-specific curve estimation problem*. Specifically, ReLLIE takes a low-light or intermediate image as input and produces second-order curves as its output at each step following a learned policy. The policy is parameterized by a lightweight fully convolutional network and trained using a set of non-reference loss functions specially designed for LLIE. In a recurrent manner, ReLLIE employs the image-specific curves to deliver a robust and accurate dynamic range adjustment.

To the best of our knowledge, ReLLIE is the first non-reference DRL based method for pixel-wise LLIE. Compared to the existing methods, ReLLIE has the following advantages. Firstly, ReLLIE learns a more flexible stochastic policy other than the deterministic one-one image translation. It can deal with inputs of different low-light degrees and provide customized enhancement outputs. The number of enhancement steps can be flexibly determined by the users (*i.e.* less or more than which used in training). Secondly, while existing deep LLIE methods require large-scale paired images or additional high-quality images for training which are expensive to collect. ReLLIE adopts non-reference loss functions as its reward function such that it does not require any paired or even unpaired data in its training process. Therefore, ReLLIE enables non-reference [6] and zero-shot [6] image enlightening which are more flexible for real-world scenarios. Thirdly, ReLLIE can be flexibly equipped with additional enhancement modules, *e.g.*, denoiser, to tackle the hybrid image degeneration according to personalized preference. In extensive experiments, we show that ReLLIE can perform on par with other existing LLIE methods that require paired or unpaired data for training. ReLLIE also achieves the state-of-the-art performance on zero-shot scenarios.

Our contributions are summarized as follows.

- (1) Recognizing the gap between real-world scenarios and the limitations of existing LLIE methods, we present a DRL based lightweight framework namely ReLLIE, towards a more customized LLIE scheme.
- (2) Accompanied with ReLLIE, we propose a new non-reference LLIE loss namely channel-ratio constancy loss (CRL) and a new channel dependent momentum update (CDMU) module, for training more robust LLIE models. We also propose enhancement-guided refinement (RF) module to handle the additive noise in LLIE scenarios.
- (3) Extensive experiments show that the proposed ReLLIE can be effectively applied to zero-shot and unsupervised LLIE benchmarks.

2 RELATED WORK

2.1 Deep Reinforcement Learning for Image Restoration

Recently, DRL has gathered considerable interest in image processing tasks. For instance, Yu *et al.* [26] proposes RL-Restore to learn a policy for selecting appropriate tools from predefined toolbox to progressively restore the quality of a corrupted image. However, it requires sufficient paired training data to train the agent using \mathcal{L}_2 loss function. More related to this work, Park *et al.* [19] proposes a

DRL based color enhancement method to tackle the need of paired data via a “distort-and-recover” training scheme. Their scheme only requires high-quality reference images for training instead of input and retouched image pairs. In parallel with [19], Hu *et al.* [9] enables a paired image-free photo retouching method with DRL and generative adversarial networks (GANs). While these methods focus on global image restoration, Furuta *et al.* [3] proposes pixelRL to enable pixel-wise image restoration which is more flexible. More recently, Zhang *et al.* [29] proposes R3L, which applies DRL to pixel-wise image denoising via direct residual recovery. However, the aforementioned methods all require the external set of “high-quality” training images, which can be highly limited in practice. Furthermore, no work to date has exploited DRL for LLIE problem.

2.2 Low-Light Image Enhancement

The LLIE task aims to increase the image visibility so as to benefit a series of downstream tasks including classification, detection, and recognition. Histogram equalization (HE) [1] and its follow-ups [14] achieve uniformly contrast improvement by spreading out the most frequent intensity values, providing undesirable amplified noise. Later on, Retinex theory [12], which assumes an image can be decomposed into reflectance and illumination, has been widely used in traditional illumination-based methods [2, 8]. For instance, NPE [23] jointly enhances contrast and illumination, and LIME [8] proposes a structure-aware smoothing model to estimate the illumination map. These hand-craft methods impose priors on the decomposed illumination and reflectance, which achieve impressive results in illumination adjustment but presenting intensive noises and artifacts.

Recently, the deep learning based methods commonly apply high-quality normal-light ground truth as guidance to learn how to improve low-light image [16, 24, 30]. LL-Net [16] proposes a stacked auto-encoder to simultaneously conduct denoising and enhancement using synthesized low/normal-light image pairs. However, the distribution of synthetic data inevitably deviates from real-world images due to the domain gap, leading to severe performance degradation when transferring to real-world cases. Later on, Wei *et al.* [24] collects a real-world dataset with low/normal-light image pairs, based on which the Retinex-Net is proposed to decompose images into illumination and reflectance in a data-driven way. Following that, various other neural networks [25, 30] have been proposed for supervised LLIE. More recent methods [10] focus on unsupervised LLIE which directly enlightens low-light images without any paired training data. The very recent Zero-DCE [6] trains the deep LLIE model using non-reference losses. However, existing deep methods produce one-one image mapping for LLIE, while neglecting different low-light imaging conditions in practice and diverse subjective preference by each individual.

Our proposed ReLLIE is significantly different from other counterparts by achieving a more customized LLIE via learning a stochastic enhancement policy rather than the one-one image translation model. The enhancement operation can be conducted multiple times, which is highly flexible for real-world scenarios. In addition, ReLLIE can be applied to zero-shot and unsupervised LLIE scenarios by employing the non-reference losses as reward function.

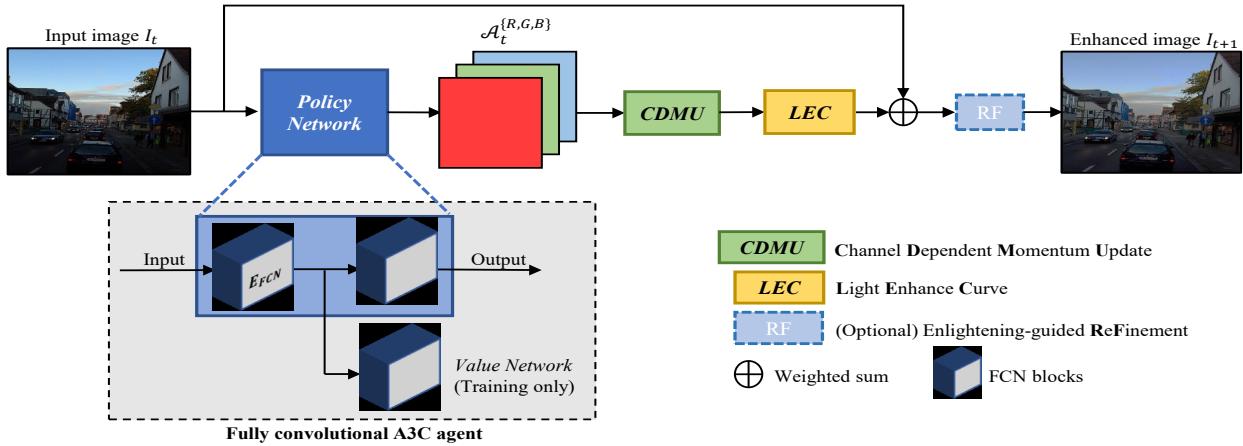


Figure 2: Illustration of the ReLLIE pipeline by taking the t -th enhancement step as an example.

3 PROBLEM DEFINITION

3.1 LLIE via Curve Adjustment

LLIE can be achieved by human experts via applying the curve adjustment in photo editing software, where the self-adaptive curve parameters are solely dependent on the input images. The optimal curves for challenging low-light images are often of very high order. Zero-DCE [6] suggests this procedure can be realized equally by recurrently applying the low-order curves. In this work, we apply a second-order light enhancement curve (LEC) at each step, which can be formulated as:

$$\text{LE}(I(x); \mathbf{A}(x)) = I(x) + \mathbf{A}(x)I(x)(1 - I(x)), \quad (2)$$

where I is the input low-light image and x denotes the pixel coordinates. $\text{LE}(I(x); \mathbf{A}(x))$ outputs the enhanced image at x , using the learned feature parameter $\mathbf{A}(x)$, which has the same size as the image. LE can be applied multiple times to approximate higher-order LEC. At the t -th step ($t \geq 1$), the enhanced output is:

$$\text{LE}_t(x) = \text{LE}_{t-1}(x) + \mathbf{A}_t(x)\text{LE}_{t-1}(x)(1 - \text{LE}_{t-1}(x)), \quad (3)$$

which models the enhancement of a low-light image as a sequential decision making problem by finding the optimal pixel-wise parameter map $\mathbf{A}_t(x)$ at each step t .

3.2 LLIE as Markov Decision Process

Based on (3), we show that LLIE can be formulated as a markov decision process (MDP) [21] consisting of the task-specific *state*, *action* and *reward*.

state: At each step t , the low-light image $I_t \in \mathbb{R}$ is the *state* ($s_t \in \mathcal{S}$), where $t = 0$ denotes the initial state with raw inputs and $t \geq 1$ denotes the intermediate states with partially enhanced images from the previous step. *action*: The *action* at s_t is to select a parameter $\alpha_t(x)$ for the LEC of each pixel, where $\alpha_t(x)$ is constrained in a predefined range \mathcal{A} and all α_t constitute a parameter map $\mathbf{A}_t(x)$. Applying a sequence of parameter maps to the input raw images results in a trajectory T of *states* and *actions*:

$$T = (s_0, \mathbf{A}_0, s_1, \mathbf{A}_1, \dots, s_{N-1}, \mathbf{A}_{N-1}, s_N, \mathbf{A}_N),$$

where N is the number of steps, and s_N is the stopping state. *reward*: The *reward* $r : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ evaluates the actions given a *state*. Our goal is to obtain a policy π that maximizes the accumulated *reward* during the MDP. To this end, we employ a stochastic policy agent parameterized by $\pi_\theta(\mathbf{A}_t | s_t)$ with trainable parameters θ . The policy $\pi_\theta : \mathcal{S} \rightarrow \mathbb{P}(\mathcal{A})$ maps the current *state* $s_t \in \mathcal{S}$ to $\mathbb{P}(\mathcal{A})$ the set of probability density functions over the *actions*, as $P(\mathbf{A}_t | s_t)$. In summary, when an agent enters a *state*, it samples one *action* according to the probability density functions, receives the *reward*, and transits to the next *state*.

More specifically, given a trajectory T , the return r_k^Y is the summation of discounted *rewards* after s_k :

$$r_k^Y = \sum_{k'=0}^{N-k} \gamma^{k'} r(s_{k+k'}, \mathbf{A}_{k+k'}), \quad (4)$$

where $\gamma \in [0, 1]$ is a discount factor, which places greater importance on *rewards* in the nearer future. To evaluate a policy, we have the following objective:

$$J(\pi_\theta) = \mathbb{E}_{s_0 \sim \mathcal{S}_0} [r_0^Y | \pi_\theta], \quad (5)$$

where s_0 is the input image and \mathcal{S}_0 is the input distribution, e.g., a dataset. Intuitively, the objective in Eq. 5 describes the expected return over all possible trajectories induced by the policy π_θ . The goal of the agent is to maximize the objective $J(\pi_\theta)$, which is related to the final image quality defined by *reward* r , since images (*states*) with a higher quality are more greatly rewarded.

4 PROPOSED RELLIE

4.1 Agent

With the MDP formulation of LLIE, we can apply a DRL based agent to conduct such task. Inspired by [3], we employ fully convolutional networks (FCNs) based asynchronous advantage actor-critic (A3C) [18] framework as our stochastic policy agent. The overall framework of ReLLIE is depicted in Fig. 2. In A3C, we use a policy network π_θ and a value network V_{θ_v} to make DRL training more stable and efficient [22]. The FCN-based encoder EFCN extracts the features of the input image I_t then outputs s_t , the representation of state t .

EFCN is shared by both π_θ and V_{θ_v} . Taking s_t , the policy network π_θ outputs the probability $P(\mathbf{A}_t|s_t, \theta_\pi)$, from which a parameter map $\mathbf{A}_t(x)$ is sampled. The value network outputs $V_{\theta_v}(s_t)$ which is an estimation of the long term discounted rewards:

$$V_{\theta_v}(s_t) = \mathbb{E}_{s_0=s_t} [r_0^\gamma]. \quad (6)$$

We also include a skip link in ReLLIE to make the update of the input image I_t a weighted sum of raw input image I_0 and the enhanced one. The update process is

$$I_t = \omega L E_t(x) + (1 - \omega) I_0, \quad (7)$$

where ω is a tunable parameter and empirically set as 0.8. After color enhancement, our framework includes an optional denoising module (which can be arbitrary image enhancing method) for further enhancement.

Without loss of generality, we consider the one-step learning case ($N = 1$) here for convenience. The gradients of the parameters of these two networks θ_π, θ_v are calculated as:

$$\begin{aligned} r_t^Y &= r_t + \gamma V(s_{t+1}), \\ d\theta_v &= \nabla_{\theta_v} (r_t^Y - V_{\theta_v}(s_t))^2, \\ d\theta_\pi &= -\nabla_{\theta_\pi} \log P(\mathbf{A}_t|s_t, \theta_\pi) (r_t^Y - V_{\theta_v}(s_t)). \end{aligned} \quad (8)$$

Action space. As mentioned in Section 3.2, the *action* for state s_t selects a parameter $\alpha_t(x)$ for LEC of a pixel, where $\alpha_t(x)$ is constrained in a predefined range \mathcal{A} and all α_t constitute the parameter map $\mathbf{A}_t(x)$. The range \mathcal{A} is critical for the performance of our agent, since a too narrow range results in a limited enhancement while a too wide one results in an exhaustively large search space. Here, we empirically set the range $\mathcal{A} \in [-0.3, 1]$ with graduation as 0.05. This setting ensures that 1) each pixel is in the normalized range of $[0, 1]$ and 2) LEC is monotonous. Meanwhile, it alleviates the cost of searching suitable LEC for low-light image enhancement. Fig. 3 shows that LEC can effectively cover the pixel value space under the proposed action space setting, with respect to different choices of N .

Reward. Many metrics have been proposed for image quality assessment, e.g., the \mathcal{L}_2 distance between enhanced/groundtruth outputs and the adversarial loss learned from a predefined set of “high-quality” images. In this work, we adopt four non-reference losses to assess an enhanced image and use the negative weighted sum of them as the *reward* to train our agent. On one hand, the using of non-reference losses gets rid of the need of expensively collected *paired* data and even does not require the so-called “high-quality” images. On the other hand, a weighted sum of different non-reference losses introduces more flexibility for user preference.

4.2 Non-Reference Losses

For zero-reference LLIE, spatial consistency loss, exposure control loss, and illumination smoothness loss are exploited in [6]. In addition to these losses, in this work we propose a new non-reference loss, namely channel-ratio constancy loss (CRL), for more robust and effective learning of zero-reference LLIE models. We discuss the details of the four losses in the following.

Spatial consistency loss. The spatial consistency loss L_{spa} encourages the preservation of the difference among neighboring

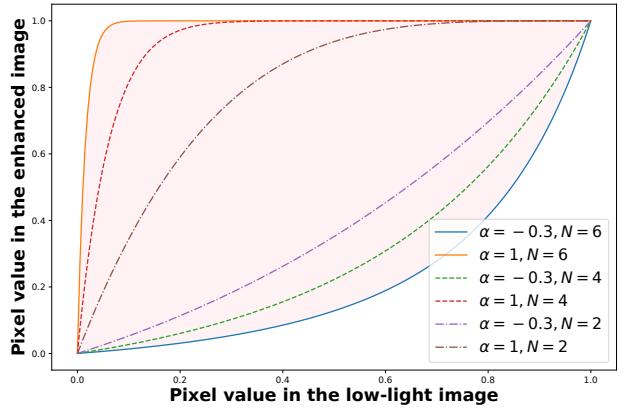


Figure 3: Illustration of how the adjustment ranges with different N and action range $\mathcal{A} \in [-0.3, 1]$.

regions during the enhancement:

$$L_{spa} = \frac{1}{K} \sum_{i=1}^K \sum_{j \in \Omega(i)} (|Y_i - Y_j| - |I_i - I_j|)^2, \quad (9)$$

where K is the number of local region and $\Omega(i)$ is the four neighboring regions (top, down, left, right) centered at the region i . Y and I denote the average intensity value of the local region in the enhanced version and input image, respectively. Here, the local region is set to 4×4 empirically.

Exposure control loss. The exposure control loss L_{exp} measures the distance between the average intensity value of a local region to a predefined well-exposedness level E , i.e., the gray level in the RGB color space [17]. It is written as:

$$L_{exp} = \frac{1}{M} \sum_{k=1}^M |Y_m - E|, \quad (10)$$

where M represents the number of non-overlapping local regions of size 16×16 , Y_m is the average intensity value of a local region m in the enhanced image. According to [6], E is set to 0.6.

Illumination smoothness loss. To avoid aggressive and sharp changes between neighboring pixels, we employ illumination smoothness loss L_{tvA} to control the curve parameter map \mathbf{A} at every state, as:

$$L_{tvA} = \frac{1}{N} \sum_{t=1}^N \sum_{c \in \epsilon} (|\nabla_x A_t^c| + |\nabla_y A_t^c|)^2, \epsilon = R, G, B, \quad (11)$$

where N is the number of iteration and ∇_x and ∇_y denote the horizontal and vertical gradient operations, respectively.

Channel-ratio constancy loss. In addition to the above three losses, we propose a channel-ratio constancy loss L_{crl} to constrain the ratio among three channels to prevent potential color deviations in the enhanced image. CRL L_{crl} is formulated as:

$$L_{crl} = \sum (|\frac{I_R}{I_G} - \frac{Y_R}{Y_G}| + |\frac{I_R}{I_B} - \frac{Y_R}{Y_B}| + |\frac{I_G}{I_B} - \frac{Y_G}{Y_B}|)^2, \quad (12)$$

where $\frac{I_R}{I_G}$ denotes the pixel-wise ratio between R channel and G channel of input image I , $\frac{Y_R}{Y_G}$ denotes the pixel-wise ratio between

R channel and G channel of enhanced one Y , and Σ denotes the summation of all the ratios. L_{crl} constrains the intrinsic ratio among channels of the input images and thereby avoiding color casts.

Agent reward. The total learning objective is

$$L_{total} = W_{spa}L_{spa} + W_{exp}L_{exp} + W_{tvA}L_{tvA} + W_{crl}L_{crl}, \quad (13)$$

where W_{spa} , W_{exp} , W_{tvA} and W_{crl} are tunable parameters which can be set according to user preference. Hence, for a given enhanced image, the *reward* r at a certain state s_t is

$$r(s_t, \mathbf{A}_t) = -L_{total}(s_{t+1}). \quad (14)$$

4.3 Channel Dependent Momentum Update

We further propose a channel dependent momentum update (CDMU) for color images with RGB channels. At each state, the agent outputs $\mathbf{A}_R(x)$, $\mathbf{A}_G(x)$, $\mathbf{A}_B(x)$ for the pixel in different channels respectively. The real parameter maps $\mathbf{A}_R^*(x)$, $\mathbf{A}_G^*(x)$, and $\mathbf{A}_B^*(x)$ applied to each channel is computed as:

$$\begin{aligned} \mathbf{A}_R^*(x) &= \mathbf{A}_R(x), \\ \mathbf{A}_G^*(x) &= \omega_{CD}\mathbf{A}_G(x) + (1 - \omega_{CD})\mathbf{A}_R(x), \\ \mathbf{A}_B^*(x) &= \omega_{CD}\mathbf{A}_B(x) + (1 - \omega_{CD})\mathbf{A}_R(x). \end{aligned} \quad (15)$$

where ω_{CD} is a tunable parameter which controls the dependence among channels. It is reasonable to perform CDMU among different channels, since in natural images the RGB channels are usually related to each other. Such update avoids aggressive modifications on an individual channel which may result in unbalanced tone performance. Note that any a single channel can be used as the reference channel, *i.e.*, \mathbf{A}_R . The ablation study in Section 5.2 reveals that a totally independent update leads to tone failure and unstable training.

4.4 Enlightening-guided Recursive Refinement

For low-light images, the degeneration model can be hybrid in practice. For instance, the image noise in shadows may become more pronounced after operations of brightness boosting. However, rare existing methods consider explicit denoising during enlightening process. To this end, this work introduces an optional denoising block to perform enlightening-guided recursive refinement (RF). In general, many existing denoisers can be good candidates for the denoising block. In light of the competitive performance of pretrained FFDNet [27], we adopt FFDNet as the denoiser block, as well as an additional noise level map as a guidance to handle the spatially variant noise. Here the noise level map refers to the ratio that each pixel enlightened, inspired by the empirical evidence the noise level map can indicate the degree of involved noise [32].

We note that the denoising blocks are *totally optional* in our framework, as they are not involved in the training process. Our agent learns the policy in a “denoising-free” setting, and users can use FFDNet to denoise the enhanced images optionally at each step of the testing phase. Such regime not only makes training more stable, but also allows a larger flexibility of using other denoisers [5, 7, 15] in testing phase. Moreover, compared with the supervised one-one image translation methods, our method allows to address various types of degeneration rather than the noise by simply employing the restoration methods accordingly.

Table 1: Quantitative results on LOL dataset [24]. +FFDNet denotes employing an external FFDNet [27] denoiser for post-processing the enhanced results.

Methods		Metrics		
		LPIPS ↓	SSIM ↑	PSNR ↑
Supervised	Retinex-Net [24]	0.4739	0.5336	16.77
	KinD [30]	0.1593	0.8784	20.38
Unsupervised	EnlightenGAN [6]	0.3661	0.6601	17.02
	EnlightenGAN+FFDNet	0.2219	0.8130	17.63
	Zero-DCE [6]	0.3352	0.6632	14.86
	Zero-DCE+FFDNet	0.2179	0.7674	15.03
Zero-shot	ReLLIE+FFDNet	0.1974	0.8268	19.52
	LIME [8]	0.3724	0.6216	14.02
	LIME+FFDNet	0.2819	0.7419	14.20
	Kar <i>et al.</i> [11] ¹	-	0.6950	17.50
Zero-shot	ReLLIE	0.3976	0.6413	18.37
	ReLLIE+FFDNet (ZS)	0.2618	0.7733	18.99

5 EXPERIMENTS AND RESULTS

In this section, we show how the proposed method ReLLIE achieves a more customized LLIE in real-world scenarios and thereby boosts the LLIE performance.

5.1 Experiments Setting

Datasets and baselines. We conduct experiments on two types of LLIE datasets, the standard dataset with paired data (LOL [24]), and the datasets without ground truth images (LIME [8], NPE [23], and DICM [13]). We compare our methods against several state-of-the-art LLIE baselines. The baselines can be classified into three categories, the supervised methods (Retinex-Net [24] and KinD [30]), the unsupervised methods (EnlightenGAN [10] and Zero-DCE [6]), and the zero-shot methods (LIME [8] and Kar *et al.*[11]). All the baselines are implemented using the publicly available codes as well as recommended parameters.

We note that the definition of “zero-shot” in this paper is different from the conventional “zero-shot learning” which often refers to using the learned models to handle images of unseen categories. In this paper, the “zero-shot” setting indicates the model can only observe *a single image* in training process. This setting is very challenging, since most of the learning-based models have much more parameters which require sufficient data in training.

Implementation details. We implement the proposed method using PyTorch framework [20]. We implement two versions of ReLLIE for both unsupervised and zero-shot settings. For unsupervised learning with sufficient training data, we adopt a seven-layer neural network as the policy agent. For zero-shot learning, we adopt a four-layer neural network as the policy agent. Except the number of layers, all the hyperparameters are identical for both of them. The coefficients in the loss is set as $W_{spa} = 1$, $W_{exp} = 100$, $W_{crl} = 20$, and $W_{tvA} = 200$. In CDMU, $\omega_{CD} = 0.2$ is set as the default. For agent learning, the discount factor γ is 0.95, the learning rate is 0.001, and the number of training iterations is 20,000 and 1,000 for unsupervised and zero-shot setting, respectively. All the experiments are conducted on a GTX 1080Ti GPU.



Figure 4: Examples of enhancement results on LOL evaluation dataset. For each two rows: Input image, LIME [8], RetinexNet [24], EnlightenGAN [10], Zero-DCE [6], KinD [30], ReLLIE (ours), ground truth. Zoom in to better see the details.

5.2 Quantitative Comparison

For quantitative comparison with existing methods, we employ three metrics including Peak Signal-to-Noise Ratio (PSNR, dB), Structural Similarity (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) [28]. Table 1 summarizes the performances of ReLLIE and baselines on the test images of LOL dataset. Guided by the paired data (*i.e.*, supervised learning), KinD [30] achieves the

best performance. Except KinD [30], our ReLLIE outperforms all the other baselines under both unsupervised and zero-shot settings. It demonstrates the efficacy of DRL for LLIE tasks.

In Fig. 5, we show the results of zero-shot LLIE. The upper row shows that ReLLIE preserves more contextual information with a better contrast. The lower row shows that ReLLIE avoids artifacts which exist in all the other baselines. More details are zoomed in with red boxes for further comparison.

¹Since the authors [11] have not released the code, we only report the SSIM and PSNR referring to their paper.



Figure 5: Examples of enhancement results on LIME evaluation dataset. We show the estimated results of (b) LIME [8], (c) EnlightenGAN [10], (d) Zero-DCE [6], and (e) Our ReLLIE.

5.3 Visual Quality Comparison

Figs. 4 and 5 compare subjective visual quality on low-light images. Fig. 4 shows the unsupervised LLIE setting that the ground truth is available. The enhanced images provided by our ReLLIE is more visually pleasing without obvious noise and color casts. Moreover, the results of ReLLIE are more sharp with more details remained and therefore preserving more visual information. It should be noticed that we adopt $N = 6$ for all the images, even though N can be changed according to users' preference for better performance (see Fig. 8). Results on the third sample reveal that for some very dark low-light images, $N = 6$ may result in under-enhancement. However, ReLLIE can still enhance the image with a relatively good contrast and yields visually pleasure results.

5.4 Visualization of Customized LLIE

The favorite illumination strengths of different persons may be pretty diverse. Therefore, a practical approach needs to consider the user-orientated goals by providing various enhancement options. Fig. 8 shows the customized enhanced images provided by the proposed ReLLIE in zero-shot scenarios and Fig. 6 demonstrates the different SSIM and PNSR achieved with different N . Given a single low-light image, we train a randomly initialized agent with a fixed amount of steps, *i.e.*, $N = 8$, for 1,000 iterations until it converges. ReLLIE is customized, because: 1) Even though the policy network is trained with $N = 8$, in testing phase the images can be enhanced for arbitrary steps; 2) Although no refinement module is involved in training, it can be employed at arbitrary steps to improve the testing performance, as shown in Fig. 7. In one word, our ReLLIE provides more candidate enhanced images to users. Therefore, it is more customized and suitable for real-world applications.

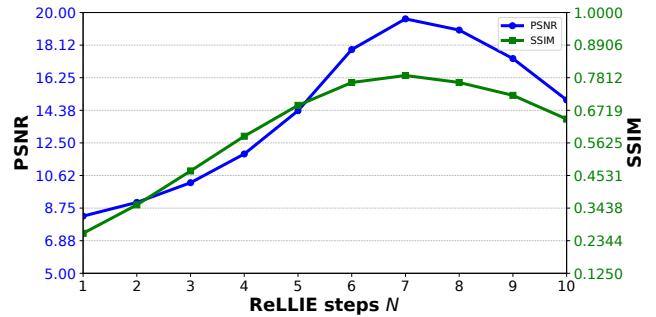


Figure 6: Quantitative performance of ReLLIE on LOL evaluation set with different number of enhancement steps N .



Figure 7: Enhancement examples of ReLLIE (zero-shot) without and with RF, respectively.



Figure 8: Examples of customized LLIE with different enhancement steps.

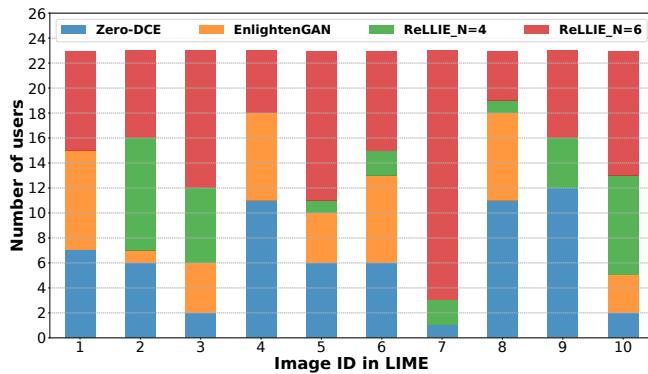


Figure 9: User study results on LIME [8] evaluation set.

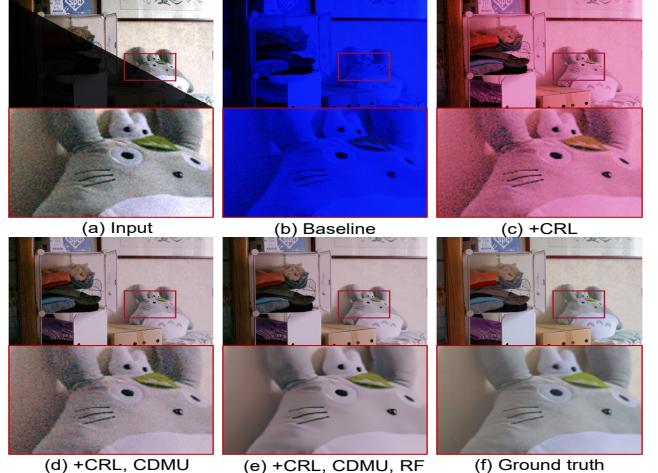


Figure 10: The effect of components in ReLLIE. (b) is the baseline without using all the components.

Table 2: Ablation study on the components of ReLLIE.

CRL	CDMU	RF	ReLLIE		
			LPIPS ↓	SSIM ↑	PSNR ↑
			0.6746	0.3798	7.76
✓			0.4824	0.6653	15.40
✓	✓		0.3450	0.6730	18.74
✓	✓	✓	0.1974	0.8268	19.52

6 CONCLUSION

In this paper, we have proposed a non-reference DRL based framework, ReLLIE, for efficient, robust, and customized low-light image enhancement. By learning a stochastic image translation policy instead of a one-one translation model, ReLLIE provides diverse image enhancement candidates to meet different individuals' preference. In addition, we have proposed a series of learning modules including CRL, CDMU and RF to enhance the robustness of LLIE methods. Extensive qualitative and quantitative experiments and user study have validated the superiority of ReLLIE against existing methods on unsupervised/zero-shot LLIE scenarios.

REFERENCES

- [1] Mohammad Abdullah-Al-Wadud, Md Hasanul Kabir, M Ali Akber Dewan, and Oksam Chae. 2007. A dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics* 53, 2 (2007), 593–600.
- [2] Xueyang Fu, Delu Zeng, Yue Huang, Yinghao Liao, Xinghao Ding, and John Paisley. 2016. A fusion-based enhancing method for weakly illuminated images. *Signal Processing* 129 (2016), 82–96.
- [3] Ryosuke Furuta, Naoto Inoue, and Toshihiko Yamasaki. 2019. Fully convolutional network with multi-step reinforcement learning for image processing. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 3598–3605.
- [4] Michaël Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédéric Durand. 2017. Deep bilateral learning for real-time image enhancement. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 118.
- [5] Shuhang Gu, Lei Zhang, W. Zuo, and Xiangchu Feng. 2014. Weighted Nuclear Norm Minimization with Application to Image Denoising. *2014 IEEE Conference on Computer Vision and Pattern Recognition* (2014), 2862–2869.
- [6] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. 2020. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1780–1789.
- [7] Lanqing Guo, Zhiyuan Zha, Saiprasad Ravishankar, and Bihai Wen. 2021. Self-convolution: A highly-efficient operator for non-local image restoration. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1860–1864.
- [8] Xiaoje Guo, Yu Li, and Haibin Ling. 2016. LIME: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing* 26, 2 (2016), 982–993.
- [9] Yuanning Hu, Hao He, Chenxi Xu, Baoyuan Wang, and Stephen Lin. 2018. Exposure: A white-box photo post-processing framework. *ACM Transactions on Graphics (TOG)* 37, 2 (2018), 1–17.
- [10] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. 2019. Enlightengan: Deep light enhancement without paired supervision. *arXiv preprint arXiv:1906.06972* (2019).
- [11] Aupendu Kar, Sobhan Kanti Dhara, Debashis Sen, and Prabir Kumar Biswas. 2021. Zero-shot Single Image Restoration through Controlled Perturbation of Koschmieder's Model. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [12] Edwin H Land. 1977. The retinex theory of color vision. *Scientific american* 237, 6 (1977), 108–129.
- [13] Chulwoo Lee, Chul Lee, and Chang-Su Kim. 2012. Contrast enhancement based on layered difference representation. In *2012 19th IEEE International Conference on Image Processing*. IEEE, 965–968.
- [14] Chulwoo Lee, Chul Lee, and Chang-Su Kim. 2013. Contrast enhancement based on layered difference representation of 2D histograms. *IEEE transactions on image processing* 22, 12 (2013), 5372–5384.
- [15] Ding Liu, Bihai Wen, Yuchen Fan, Chen Change Loy, and Thomas S Huang. 2018. Non-Local Recurrent Network for Image Restoration. In *Advances in Neural Information Processing Systems*, Vol. 31.
- [16] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. 2017. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition* 61 (2017), 650–662.
- [17] Tom Mertens, Jan Kautz, and Frank Van Reeth. 2009. Exposure fusion: A simple and practical alternative to high dynamic range photography. In *Computer graphics forum*. Vol. 28. Wiley Online Library, 161–171.
- [18] Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. 2016. Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*. 1928–1937.
- [19] Jongchan Park, Joon-Young Lee, Donggeun Yoo, and In So Kweon. 2018. Distort-and-recover: Color enhancement using deep reinforcement learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5928–5936.
- [20] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in pytorch. (2017).
- [21] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.
- [22] Richard S Sutton, David McAllester, Satinder Singh, and Yishay Mansour. [n.d.]. Policy Gradient Methods for Reinforcement Learning with Function Approximation. In *Advances in Neural Information Processing Systems*. 1057–1063.
- [23] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. 2013. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing* 22, 9 (2013), 3538–3548.
- [24] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. 2018. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560* (2018).
- [25] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. 2020. From Fidelity to Perceptual Quality: A Semi-Supervised Approach for Low-Light Image Enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [26] Ke Yu, Chao Dong, Liang Lin, and Chen Change Loy. 2018. Crafting a toolchain for image restoration by deep reinforcement learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2443–2452.
- [27] Kai Zhang, Wangmeng Zuo, and Lei Zhang. 2018. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Transactions on Image Processing* 27, 9 (2018), 4608–4622.
- [28] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *CVPR*.
- [29] Rongkai Zhang, Jiang Zhu, Zhiyuan Zha, Justin Dauwels, and Bihai Wen. 2021. R3L: Connecting Deep Reinforcement Learning to Recurrent Neural Networks for Image Denoising via Residual Recovery. *arXiv:2107.05318*
- [30] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. 2019. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM International Conference on Multimedia*. 1632–1640.
- [31] Ziqiang Zheng, Yang Wu, Xinran Han, and Jianbo Shi. 2020. ForkGAN: Seeing into the Rainy Night. In *The IEEE European Conference on Computer Vision (ECCV)*.
- [32] Anqi Zhu, Lin Zhang, Ying Shen, Yong Ma, Shengjie Zhao, and Yicong Zhou. 2020. Zero-shot restoration of underexposed images via robust retinex decomposition. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 1–6.