Deep Reinforcement Learning for Lunar Polar Low-Light Enhancement

Kaichen Chi †, Qiang Li †, *Member, IEEE*, Jun Chu, Junjie Li, and Qi Wang, *Senior Member, IEEE*

***Abstract*—As a bridge between the moon and human perception, the lunar optical image reflects lunar topography, geology, and evolution. Unfortunately, the permanent shadow regions (PSRs) near the lunar poles suffer from information contamination due to insufficient illumination. Low-light enhancement is a subjective process whose target is tied to human visual perception. However, existing low-light enhancement methods often operate as opaque "black box", lacking transparency and failing to accommodate diverse perceptual preferences. To this end, we explore a PSRs Low-Light Enhancer (PSRs-LLE) that treats low-light enhancement as a Markov decision process, thereby dynamically fitting perceptual preferences. Specifically, a deep Q network as an agent integrates multiple user-friendly attributes (*e.g.*, brightness, contrast, chroma, and detail) through actions recursion (*i.e.*, a candidate set of image enhancement operations). Such transparent and specific action sequences satisfy customization preferences of users while providing convincing interpretability, compared with the "black box" paradigm of deep learning. More importantly, a well-designed non-reference loss function liberates PSRs-LLE from the dilemma of virtual assumptions and paired data, which further enhances usability. Extensive experiments demonstrate that PSRs-LLE outperforms state-of-the-art methods in both qualitative and quantitative comparisons.**

***Index Terms*—Low-light enhancement, remote sensing, reinforcement learning, Markov decision, personalization.**

I. Introduction

As the only natural satellite of the earth, the exploration and exploitation of the moon is a promising direction. Unfortunately, due to the axial tilt and undulating topography, the permanent shadow regions near the poles are never illuminated by the sun [1]–[3]. Although PSRs contain precious water-ice and polar volatiles, image formation drawbacks such as low visibility, low contrast, and low signal-to-noise ratio lead to information contamination [4]. Therefore, designing a low-light enhancement method to repair the visual quality of PSRs is of significance.

Traditional low-light enhancement methods focus on statistics of illumination degradation through hand-crafted prior knowledge, *e.g.*, histogram equalization [5] and Retinex theory [6], thus relighting dark pixels. However, traditional methods are highly restrictive in practice because prior assumptions typically fail to satisfy PSRs. In contrast, benefiting from large-scale supervision signals, deep learning-based methods learn the mapping from the low-light sample to the normal exposed one in an end-to-end manner, which have become mainstream [7]. Pioneering ideas have been elaborated from different perspectives, *e.g.*, multi-view collaboration [2], image decomposition [3], prompt learning [8], and Fourier transform [9]. Nevertheless, such end-to-end schemes implicitly assume only one deterministic normal-light output for a low-light sample, which fails to satisfy diverse perceptual demands [10].

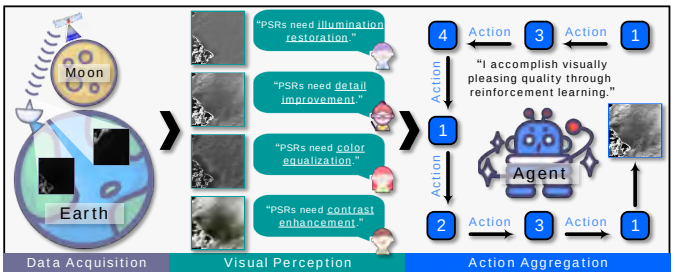

Fig. 1. Schematic of our basic idea. PSRs-LLE employs a series of transparent actions (actions $1 \mapsto 4$ represent the illumination restoration operator, the detail improvement operator, the color equalization operator, and the contrast enhancement operator, respectively) driven by rewards to restore the illumination.

For the PSRs sample, enhancing illumination is not sufficient because it suffers from multiple concurrent quality degradations. How to satisfy comprehensive visual quality restoration based on visual perception is a challenge. As well known, deep learning, a "black box" scheme, does not support preference-based guidance [11]. Reinforcement learning thus emerges as a highly promising solution to this challenge. In this work, we employ reinforcement learning to organize a series of actions to progressively relight dark pixels, as shown in Fig. 1. Driven by perceptual rewards, stepwise actions align the appearance of PSRs more in line with the comprehensive visual attractiveness. In particular, our perceptual rewards are non-reference, suggesting that PSRs-LLE gets rid of the laborious and tedious collection of data pairs. Moreover, our action set is designed to fit the built-in functions in commercial photo retouching software [12]. Therefore, a reasonable combination of multiple conventional image processing techniques can gracefully cope with tricky low-light degradation. In summary, the main contributions are as follows.

- **Perspective contribution.** We rethink low-light enhancement from the perspective of reinforcement learning, employing perceptual scores to emulate subjective user evaluations, and subsequently as a reward organizing a series of basic image enhancement operators into a

This work was supported in part by the National Natural Science Foundation of China under Grant 62301385, 62471394, and U21B2041, and in part by the Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University under Grant CX2024107. Kaichen Chi and Qiang Li have equal contribution. Corresponding author: Qi Wang.

The authors are with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, China (e-mail: chikaichen@mail.nwpu.edu.cn, liqmges@gmail.com, junchu@mail.nwpu.edu.cn, junjieli@mail.nwpu.edu.cn crabwq@gmail.com).

comprehensive illumination restoration scheme.

- **Technical contribution.** We propose PSRs-LLE, a deep Q network that simulates quality evaluation through reward feedback and subsequently selects commercial photo retouching actions to dynamically tune the visual quality of PSRs. Such a manner provides a transparent processing stream for low-light enhancement.
- **Practical contribution.** Without fancy virtual assumptions, empirical manual processing, and tedious data acquisition, our method achieves nontrivial visual quality in a non-reference form.

II. Related work

In this section, we first review reinforcement learning-based low-level methods, and subsequently discuss existing low-light enhancement schemes.

A. Reinforcement Learning

Benefit from the principled decision pattern and highly explainable action flows, reinforcement learning's emergence as an upstart in low-level task. Yin *et al.* [13] employed Gaussian and Laplacian pyramids to fuse overexposed and underexposed images to generate an intermediate version. Later on, a deep Q network progressively adjusted the contrast, brightness, and saturation of the intermediate version to produce visually pleasing high dynamic range images. To generate realistic biological details, Chen *et al.* [14] proposed a patch level Markov decision process for pathology image super-resolution. On the one hand, a spatial manager was responsible for identifying corrupted pathology patches. On the other hand, a temporal manager was responsible for evaluating when patch recovery should be terminated. Wang *et al.* [15] designed a meta Atlantis for underwater image enhancement, consisting of meta submergence, meta relief, and meta ebb. Specifically, meta submergence embedded the Akkaynak-Treibitz image formation model into a generative adversarial network to synthesize physically plausible virtual underwater samples. Meta relief was responsible for generating depth maps for underwater samples. Meta ebb employed reinforcement learning to mine the optimal underwater imaging parameter configuration, and then inversely solved the image formation model to repair visual quality. To save the computational overhead of omnidirectional image super-resolution, Deng *et al.* [16] used reinforcement learning to discard unnecessary high latitude bands. Similarly, Yu *et al.* [17] proposed a reinforcement learning scheme with a difficulty feedback reward to cue the difficulty of distortion region repair. Subsequently, an appropriate denoising route was selected to achieve a balance between performance and complexity. Obviously, reinforcement learning presents quite promising potential in the context of low-light enhancement.

B. Low-Light Enhancement

Traditional-based methods usually tend to explore various illumination properties. Arici *et al.* [18] regarded the low-light enhancement as an optimization for minimizing the cost function. Therefore, they introduced penalty terms in the histogram equalization, thus integrating white/black stretching and luminance preservation into the image restoration process. Wang *et al.* [19] proposed a bright-pass filter to decouple the low-light sample into reflectance and illumination. Subsequently, a bi-log transformation was used for illumination to cope with non-uniform luminance. Based on maximum color channel values, Guo *et al.* [20] estimated the illumination map, and then maintained its fidelity and smoothness through Frobenious and ℓ_1 norms. In addition, they designed a sped-up solver to reduce the computational load while recovering the luminance. Considering inevitable noise in low-light conditions, Li *et al.* [21] proposed an alternating direction minimization strategy to optimize the parameter estimation of the Retinex model, thus accomplishing low-light enhancement in the noisy environment. Hao *et al.* [22] proposed a two-stage Retinex model to repair visibility. Specifically, they designed a Gaussian total variation driven edge filter to smooth the illumination layer. Then, under the Retinex constraint, they applied a regularization term to the reflectance layer to suppress the noise. Nevertheless, traditional-based methods fail to generalize to real-world low-light enhancement because of ill-posed prior assumptions.

Recently, deep learning-based methods have displayed superior performance. Saini *et al.* [23] designed an additive factorization scheme to decompose the low-light sample into multiple specular components through pixel sparsity. Subsequently, they employed the fusion network to perform fusion, enhancement, and denoising on specular components to restore illumination. Li *et al.* [24] proposed a diffusion paradigm to integrate multi-modal information, such as depth maps, visible samples, and text captions, to cope with the low-light issue through semantic consistency. Chi *et al.* [25] embedded Retinex theory into Mamba as well as devised reflectance gradient and illumination contour losses to mimic the biologically plausible low-light enhancement in the human visual system. Wang *et al.* [26] decoupled low-light enhancement into illumination recovery and color transfer phases, and then propagated context between low-light and normal-light nodes to restore luminance. Lv *et al.* [27] embedded the Fourier transform into a diffusion model, employing amplitude to align luminance with normal-light distribution. Yang *et al.* [28] leveraged the neural representation normalization to normalize low-light samples, thus avoiding the regeneration of extreme degradation. Besides, they introduced semantic supervision and high-frequency supervision in the discriminator to recover luminance. Liu *et al.* [29] proposed a query module to capture normal-light features, and then fuse low-light features and normal-light features through similarity. Such a manner is able to dynamically update the luminance and texture. Cai *et al.* [30] proposed an illumination-guided transformer to explore the interaction between regions with different exposure levels. In addition, a corruption restorer is proposed to reduce noise, artifacts, and color deviation. Nevertheless, the above methods tend to compromise on a deterministic answer, *i.e.*, ground truth, and not applicable to PSRs. Besides, the supervised learning scheme on limited scenarios also weaken their generalization ability.

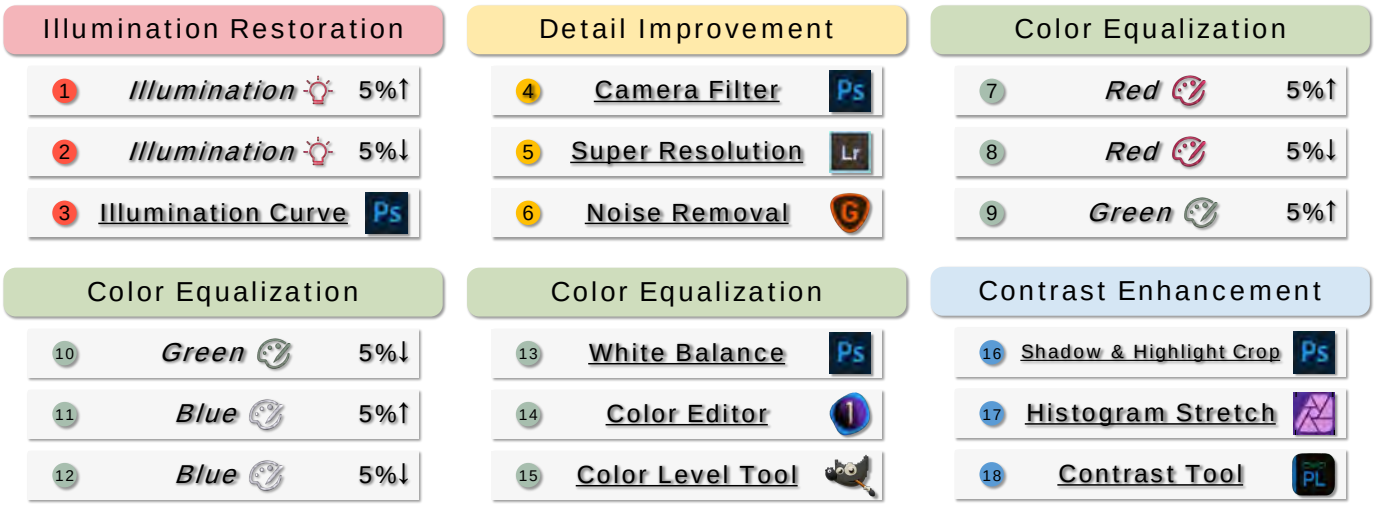

Fig. 2. Schematic of the action candidate. The candidates balance *basic image enhancement operations* and commercial photo retouching operations because of their powerful generalization and convincing interpretability. Notably, the icons of commercial software have been converted to the cartoon clay style, thus there are no copyright issues.

III. Methodology

In this section, we first describe an overview of reinforcement learning. Subsequently, the pipeline for low-light enhancement is illustrated.

A. Preliminaries

Reinforcement learning is defined as a Markov decision process, providing the best strategy for organizing image enhancement methods to achieve the low-light enhancement goal. Specifically, reinforcement learning consists of two elements, *i.e.*, agent and environment. The agent accumulates experience through environmental interactions, emulating human learning to master skills. A four-tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R})$ is responsible for describing such collaboration process. $\mathcal{S}$ is the state, representing the perception of the agent towards the environment. $\mathcal{A}$ is the action, representing the description of agent behaviors. $\mathcal{P}$ is the state transition function, representing the change probability of the environment. $\mathcal{R}$ is the reward, representing the feedback from the post-action environment to the agent. Regardless of the reward, PSRs-LLE updates towards the high reward, thus maximizing the overall reward. In what follows, we describe this Markov decision process.

Agent. We leverage a deep Q network [31] as the agent, which consists of four fully connected layers. For the deep Q network, the input term is the current state, *i.e.*, the image representation, while the output term is the Q value, *i.e.*, the action-wise value function.

Environment. We treat the PSRs sample as the environment. This environment comprises both the original image and the step-wise processing version of the Markov decision process.

State. We consider PSRs features as states, which can be expressed as a mixture of illumination and perception:

$$\mathcal{S} = \mathrm{Cat}[\mathcal{F}_{\mathcal{I}}, \mathcal{F}_{\mathcal{P}}], \tag{1}$$

where $\mathcal{F}_{\mathcal{I}}$ represents the illumination feature, $\mathcal{F}_{\mathcal{P}}$ represents the perception feature, and $\mathrm{Cat}[\cdot]$ represents the dimension concatenation operator. Both $\mathcal{F}_{\mathcal{I}}$ and $\mathcal{F}_{\mathcal{P}}$ are normalized between [0,1]. The illumination feature $\mathcal{F}_{\mathcal{I}}$ is derived from Retinex theory, which reflects the intensity and distribution of luminance. In contrast, the perception feature $\mathcal{F}_{\mathcal{P}}$ aligns with human visual perception and focuses on measuring the image quality [32]. Specifically, Zhang *et al.* [32] employed deep networks such as VGG for similarity judgment and quality evaluation on a perceptual judgment benchmark. They found that the judgment of the deep network was surprisingly consistent with humans and was able to perceive visual distortions. Inspired by this, Sun *et al.* [11] used VGG-19 to extract perceptual cues, thus providing color, contrast, and saturation distributions. In summary, our state design accommodates both task-relevant luminance cues and perceptual cues, which achieves the best of both worlds.

Action. We employ basic image enhancement operators and popular commercial photo retouchers to establish an action set. For these actions, we either use open source code, or we understand the motivation and implement it ourselves. According to the mainstream enhancement preferences, the action set is divided into four subsets, *i.e.*, illumination restoration, detail improvement, color equalization, and contrast enhancement. The action set is depicted in Fig. 2, listing a total of 18 actions.

In the illumination restoration subset, action ❶ represents a 5% increase in luminance, and action ❷ is the opposite. The main reason we set the magnitude to 5% is to avoid irreversibility due to over- or under-enhancement [11]. Action ❸ is the illumination curve. It adjusts exposure through anchor point manipulation on the illumination curve, typically used for the Photoshop software. In the detail improvement subset, action ❹ is the camera filter. Photoshop combines texture, clarity, mist removal, and granularity to accomplish it. Action ❺ is the super resolution, and Lightroom employs this technique to render clear details and textures. Action ❻ is the noise removal. Topaz Gigapixel AI introduces the denoising operation to boost lossless enlargement. In the color equalization subset, action ❼ ↦ action ⓬ represent the increase or decrease of

RGB color channels, respectively. Action ⓭ is white balance. Photoshop uses it to equalize color channels. Action ⓮ from Capture One combines hue, saturation, and brightness revisions to correct color distortion. Action ⓯ is the color level tool, which formulates a color level distribution to change the hue. In the contrast enhancement subset, action ⓰ is a crop operation for shadow and highlight. As the name suggests, it truncates the endpoints of the histogram to enhance contrast. Action ⓱ represents the histogram stretch from Affinity Photo. In DxO PhotoLab, action ⓲ collaborates contrast, micro contrast, and fine contrast to sharpen a photo. We choose the above simple yet effective methods because they have stood the test of time and market. More importantly, our actions are extensible and replaceable. Nearly all available methods are plug and play.

State Transition Function. The state transition function $\mathcal{P}(s_{t+1}|s_t, a_t)$ depicts the transition from state s_t to state s_{t+1}. In particular, $s_t, s_{t+1} \in \mathcal{S}$ and $a_t \in \mathcal{A}$.

Reward. We employ non-reference perceptual loss functions as a reward. In the Markov decision process, these loss functions progressively induce PSRs-LLE to converge to the desirable quality.

To suppress over- or under-exposure, the exposure control loss $\mathcal{L}_\mathcal{E}$ [12] measures the deviation between average intensity of image patches and a satisfactory luminance level $\mathcal{E}$:

$$\mathcal{L}_\mathcal{E} = \frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} |\mathcal{A}_m - \mathcal{E}|, \tag{2}$$

where $\mathcal{M}$ represents the number of non-overlapping image patches of size 16×16, $\mathcal{A}_m$ represents the average intensity of an image patch in the enhanced version $\mathcal{I}_\mathcal{E}$, and $\mathcal{E}$ represents the gray level in the RGB color space, set to 0.6 [12]. The illumination reward is the change of $\mathcal{L}_\mathcal{E}$, expressed as follows:

$$\mathcal{R}_\mathcal{I} = -[\mathcal{L}_\mathcal{E}(t) - \mathcal{L}_\mathcal{E}(t-1)], \tag{3}$$

where t represents the step t in the Markov decision process.

The smooth loss $\mathcal{L}_\mathcal{S}$ [23] measures the difference between neighboring pixels to polish details:

$$\mathcal{L}_\mathcal{S} = \frac{1}{|\Omega|} \sum_{\Omega} [(\nabla_x \mathcal{I}_\mathcal{E})^2 + (\nabla_y \mathcal{I}_\mathcal{E})^2], \quad \Omega \in \{\mathcal{C} \times \mathcal{H} \times \mathcal{W}\} \tag{4}$$

where ∇_x and ∇_y represent horizontal and vertical gradient operations, respectively. The detail reward is the change of $\mathcal{L}_\mathcal{S}$, expressed as follows:

$$\mathcal{R}_\mathcal{D} = -[\mathcal{L}_\mathcal{S}(t) - \mathcal{L}_\mathcal{S}(t-1)], \tag{5}$$

The color constancy loss $\mathcal{L}_\mathcal{C}$ [12] follows the Gray World assumption to eliminate potential color cast:

$$\mathcal{L}_\mathcal{C} = \sum_{\forall (p,q) \in \xi} (\mathcal{J}^p - \mathcal{J}^q)^2, \ \xi = \{(\mathcal{R}, \mathcal{G}), (\mathcal{R}, \mathcal{B}), (\mathcal{G}, \mathcal{B})\} \tag{6}$$

where $\mathcal{J}$ represents the average intensity of the color channel. The color reward is the change of $\mathcal{L}_\mathcal{C}$, expressed as follows:

$$\mathcal{R}_\mathcal{C} = -[\mathcal{L}_\mathcal{C}(t) - \mathcal{L}_\mathcal{C}(t-1)], \tag{7}$$

The histogram loss $\mathcal{L}_\mathcal{H}$ [33] promotes histogram stretching, driving the distribution of PSRs to span a wider range of gray scales:

$$\mathcal{L}_\mathcal{H} = \sum_c \mathrm{MSE}[H^c(\mathcal{I}_\mathcal{E}) - \sum H^c(\mathcal{I}_\mathcal{E})/255], \tag{8}$$

where $\mathrm{MSE}[\cdot]$ represents the mean squared error operation. Besides, $H^c = \{i, H_i^c\}_{i=0}^{255}$ represents the differentiable histogram, and H_i^c represents the pixel number of intensity value i. The contrast reward is the change of $\mathcal{L}_\mathcal{H}$, expressed as follows:

$$\mathcal{R}_\mathcal{T} = -[\mathcal{L}_\mathcal{H}(t) - \mathcal{L}_\mathcal{H}(t-1)], \tag{9}$$

In summary, the reward is a weighted sum of the above perceptual rewards:

$$\mathcal{R}(s_t, a_t) = \alpha \mathcal{R}_\mathcal{I} + \beta \mathcal{R}_\mathcal{D} + \eta \mathcal{R}_\mathcal{C} + \zeta \mathcal{R}_\mathcal{T}. \tag{10}$$

We organize 30 volunteers to provide the importance of four enhancement factors for visual perception. Based on the quantitative scores from volunteers, the enhancement attribute index $\{\alpha, \beta, \eta, \zeta\}$ was set to $\{0.537, 0.196, 0.098, 0.169\}$.

B. Pipeline

Fig. 3 illustrates an overview of PSRs-LLE. The PSRs-LLE consists of four parts. The first part is a bi-dimensional encoder that extracts both luminance and perception features. We employ Retinex-Net [49] as the illumination encoder, which is trained on LOL-Real [69] and LOL-Synthetic [69] to obtain the pre-trained weights. Referring to [11], we employ VGG-19 trained on ImageNet classification as the perception encoder. Our perception features are derived from the specific 4096-D activation of the fully connected layer in the VGG-19 model. The concatenation of luminance features and perception features represent the state from the environment. The second part is the evaluation network, which receives luminance and perception features and selects the optimal action. Taking time step t as an example, the agent selects action ❸ as the optimal action under the guidance of the ε-greedy algorithm. This is because restoring luminance is the most pressing need in the current step. The third part is the target network, which provides a value $\mathcal{Q}(s, a)$ for each action a. If the $\mathcal{Q}$ value is negative, the inference stops, otherwise the action selection-action value calculation process is repeat. The last part is the reward computation towards the visual quality restoration, which only operates in the training phase but not in the inference phase.

Specifically, the evaluation network selects the optimal action with a probability of 1-ε through the ε-greedy strategy:

$$a_t = \underset{a_t(i)}{\mathrm{argmax}}\, \mathcal{Q}(s_t, a_t(i)), \tag{11}$$

where $a_t(i)$ represents the action set. Therefore, i is set to 18. After action a_t, the visual quality of PSRs is improved. This likewise implies that the state transitions from s_t to s_{t+1}.

Subsequently, the target network provides the value of the action a_t through a single step approximation [68]:

$$\mathcal{Q}(s_t, a_t) = \begin{cases} \mathcal{Q}_{\mathrm{output}}, & \text{if END} \\ \mathcal{R}(s_t, a_t) + \gamma \mathcal{Q}(s_{t+1}, a_{t+1}), & \text{ELSE} \end{cases} \tag{12}$$

where "if END" indicates the value of the action a_t is negative and s_t is the terminal state. Besides, $\mathcal{Q}_{\mathrm{output}}$ represents the value of the PSRs-LLE output term (low-light enhancement result). "ELSE" indicates the updated state s_{t+1} repeats again the above process (Perform action $\mapsto$ Update state).

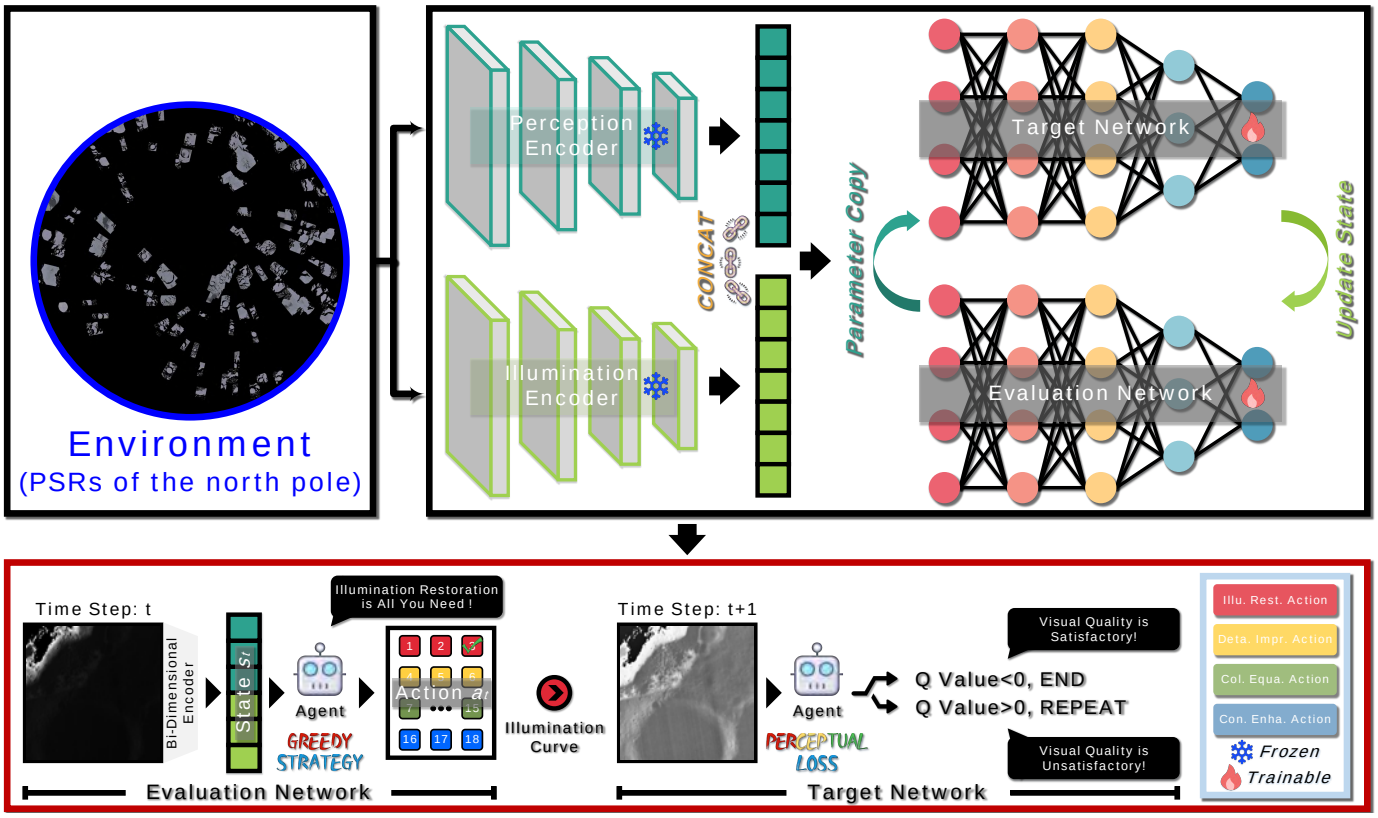

Fig. 3. Schematic diagram of PSRs-LLE. The black box is the network framework for PSRs-LLE. PSRs-LLE consists of an illumination encoder, a perception encoder, an evaluation network, and a target network. Besides, the red box is the computational process for PSRs-LLE from time step t to t+1.

As is well-known, the reinforcement learning pipeline iterates the above time steps with the goal of maximizing the cumulative reward. Extending from a single time step to a dynamic flow, the cumulative reward unfolds in a recursive manner:

$$\begin{aligned}\mathcal{Q}(s_t,a_t) &= \mathcal{R}(s_t,a_t)+\gamma\mathcal{R}(s_{t+1},a_{t+1})+\gamma^2\mathcal{R}(s_{t+2},a_{t+2})+\ldots \\ &= \mathcal{R}(s_t,a_t)+\gamma[\mathcal{R}(s_{t+1},a_{t+1})+\gamma\mathcal{R}(s_{t+2},a_{t+2})+\ldots] \\ &= \mathcal{R}(s_t,a_t)+\gamma\mathcal{Q}(s_{t+1},a_{t+1}),\end{aligned} \tag{13}$$

where $\gamma \in [0,1]$ represents the discount factor. The advantage of such recursive paradigm is to drive the agent to continuously explore new actions to better perceive and understand the environment, thus acquiring potentially high rewards. This contributes to enhance generalization performance [68].

The evaluation network and the target network are updated leveraging the following loss function:

$$\mathcal{L}(\theta) = [\hat{\mathcal{Q}}(s_t,a_t;\hat{\theta}) - \mathcal{Q}(s_t,a_t;\theta)]^2. \tag{14}$$

where θ and $\hat{\theta}$ represent the parameters of the target network and the evaluation network.

IV. Experiment

A. Experimental Setting

Implementation. We train the agent with a batch size 4 for 50,000 epochs on an NVIDIA RTX 3090 GPU. During training, the perception and illumination encoders are frozen, while the target and evaluation networks are update. In the training process, we employ the ε-greedy policy [34], which randomly samples actions with a probability of ε while selecting an action with the maximum $\mathcal{Q}$-value with a probability of $1-\varepsilon$. ε annealed linearly from 1.0 to 0.1. The discount rate for reward is set to 0.95. The initial learning rate is set to 10^{-5}, and decays by a factor of 0.96 every 5,000 episodes. Following [68], parameter copying is performed every 1,000 episodes. In addition, the Adam optimizer serves network optimization. In the inference stage, the illumination and perception features of a low-light sample are the initial state s_0. Subsequently, the agent selects the optimal action a_0 by maximizing $\mathcal{Q}(s_0,a)$. PSRs-LLE applies the action a_0 to the low-light sample to obtain the enhanced version. Based on the enhanced version, PSRs-LLE repeats the above steps until all action values are negative. As shown in Figs. 4-7, multiple actions must be executed to achieve optimal quality restoration, as a single action is insufficient. In addition, the number of actions executed varies across samples, depending on their degree of quality degradation.

Benchmark. We perform qualitative and quantitative experiments on the LROC-NAC dataset [4]. LROC-NAC provides full-color images acquired by the narrow angle camera aboard the lunar reconnaissance orbiter, with a resolution of 0.5-2 m and a coverage boundary of 10 km. These images cover polar regions extending from 80°S to the South Pole and from 80°N to the North Pole. Overall, the training set contains 12 north pole and 10 south pole images with very large size. Therefore, the training samples are available for user cropping. Besides, the testing set consisted of 40 PSRs images of size 400×400,

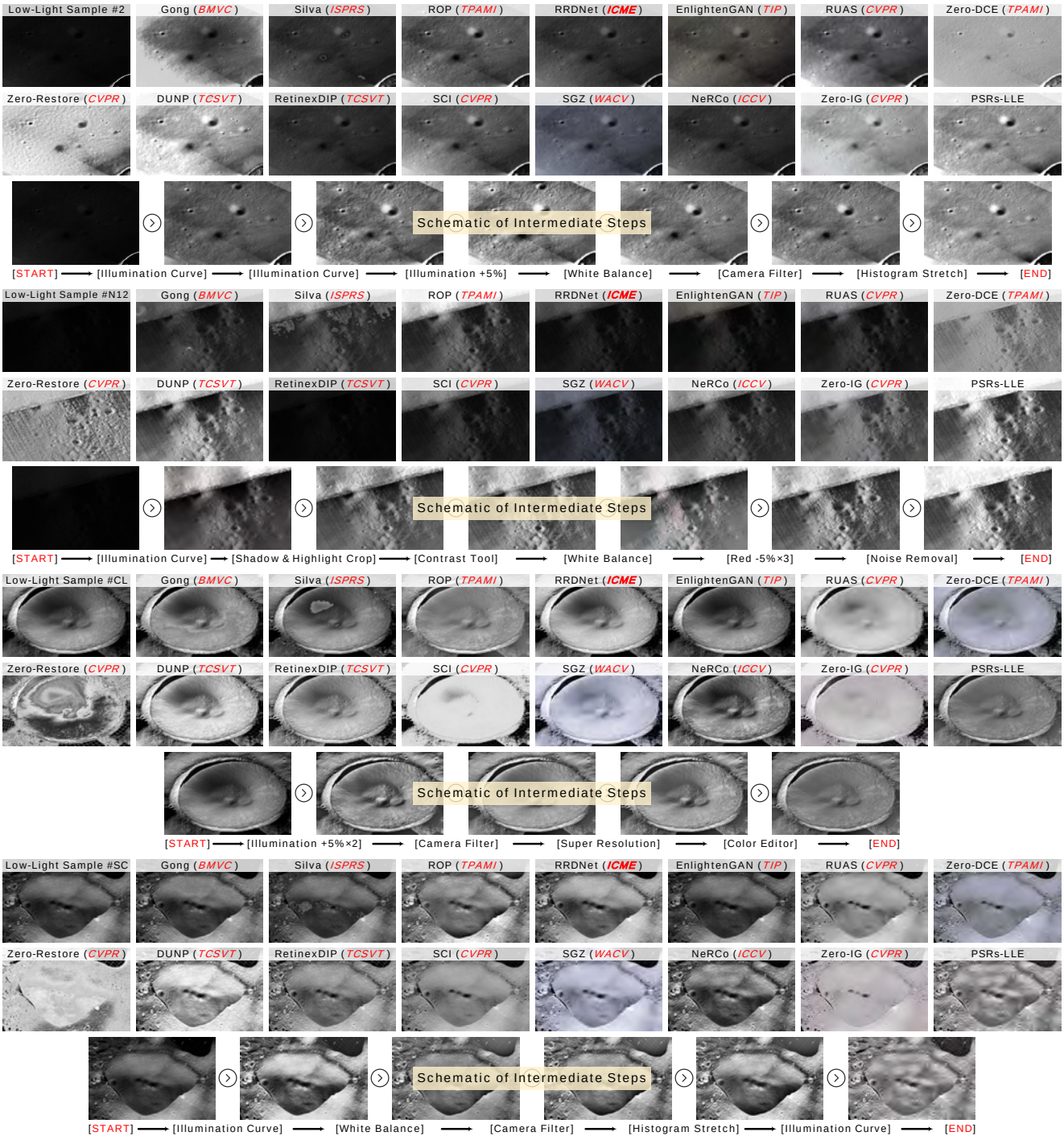

Fig. 4. Qualitative comparison on LROC-NAC [4]. Notably, we provide a series of actions chosen by the agent and a corresponding image sequence.

with 20 images from each pole.

Competitor. We compare PSRs-LLE with the following competitors: Gong [7], Silva [35], ROP [36], RRDNet [37], EnlightenGAN [38], RUAS [39], Zero-DCE [12], Zero-Restore [40], DUNP [41], RetinexDIP [42], SCI [43], SGZ [44], NeRCo [28], and Zero-IG [45]. Following the experimental setting provided by the authors, all competitors are retrained on LROC-NAC [4] until they achieve the best quantitative scores.

Metric. Since permanent shadow regions lack corresponding ground truths, we employ the natural image quality evaluator (NIQE) [46], the contrast-distorted image quality evaluator (CEIQ) [47], the entropy enhancement evaluator (EME) [48], and the lightness-order-error (LOE) [19] to perform the non-reference evaluation. Specifically, NIQE evaluates the distortion degree based on a quality-aware feature collection from a natural scenario statistical model. CEIQ mimics the sensitivity of the human visual system to spatial intensity, spatial distribution, and orientation of structures, thus calculating contrast degradation in terms of both luminance and chrominance. EME introduces an alpha coefficient into the Weber-Fechner law to

TABLE I
QUANTITATIVE COMPARISON ON LROC-NAC [4] AND RSLLE-4K. † REPRESENTS TRADITIONAL METHODS AND ‡ REPRESENTS ZERO-SHOT OR UNSUPERVISED METHODS. THE BEST SCORE IS IN RED, THE SECOND-BEST SCORE IS IN BLUE, AND THE THIRD-BEST SCORE IS IN GREEN.

Method	LROC-NAC				RSLLE-4K			
	NIQE ↓	CEIQ ↑	EME ↑	LOE ↓	NIQE ↓	CEIQ ↑	EME ↑	LOE ↓
Gong [7] †	9.7885	2.4484	5.7138	267.1895	4.5725	2.5239	6.3074	356.8747
Silva [35] †	7.7170	2.7270	6.2409	394.6249	5.0749	2.2115	5.9172	127.5841
ROP [36] †	10.8806	2.5486	4.7879	215.4844	4.5091	3.1173	7.2707	297.5018
RRDNet [37] ‡	7.0837	2.1050	5.2877	133.3124	4.6758	2.6849	6.5826	47.2457
EnlightenGAN [38] ‡	7.7027	2.1993	5.4055	183.5800	5.2229	2.9170	6.8135	412.0616
RUAS [39] ‡	10.1035	1.9574	3.6693	150.9587	6.1901	2.7682	5.1110	312.7937
Zero-DCE [12] ‡	9.6586	2.3879	5.7195	189.5603	4.5562	2.3247	5.7550	328.7847
Zero-Restore [40] ‡	11.3308	3.2026	7.0451	557.3289	5.1726	2.9220	7.4320	653.2968
DUNP [41] ‡	7.3601	3.2349	7.0956	290.3495	4.6000	2.8270	7.2101	554.6531
RetinexDIP [42] ‡	10.0608	1.6020	3.6180	269.9995	5.3621	2.4963	5.8206	907.1642
SCI [43] ‡	10.0281	1.9860	3.5672	150.1192	4.7807	2.5791	6.2491	62.7923
SGZ [44] ‡	9.7920	2.1529	5.5225	196.6376	4.7722	2.5211	6.7114	334.6987
NeRCo [28] ‡	10.3841	2.6165	6.0465	194.2642	4.5028	2.9610	6.8891	135.6728
Zero-IG [45] ‡	10.8718	2.7361	6.2612	199.3751	4.4871	2.5509	6.6437	303.3698
PSRs-LLE ‡	7.3134	3.4138	7.3653	120.9972	4.4270	3.1358	7.4409	42.3827

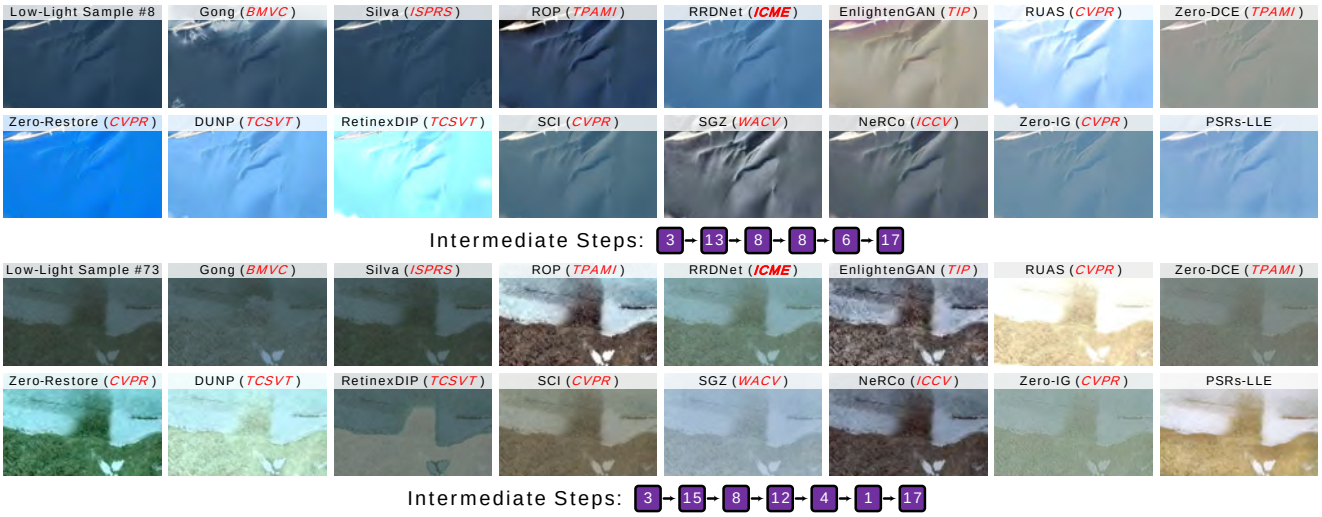

Fig. 5. Qualitative comparison on RSLLE-4K.

describe human perception of detail stimuli. LOE assesses the illumination recovery effect based on the luminance order error between the quality degradation image and the corresponding enhancement version.

B. Qualitative Comparison

We display a visual comparison on PSRs in Fig. 4. Gong [7], RRDNet [37], EnlightenGAN [38], RUAS [39], RetinexDIP [42], SCI [43], SGZ [44], NeRCo [28], and Zero-IG [45] fail to provide satisfactory luminance for challenging PSRs. For Gong [7], the main reason for failing to provide satisfactory luminance is that the user needs to manually label the normal-light and low-light regions. Such a manner is stereotypical and prohibitive. For RRDNet [37], RUAS [39], RetinexDIP [42], and Zero-IG [45], these methods rely on Retinex theory to optimize illumination. However, this artificial assumption fails to adequately describe the low-light issue. EnlightenGAN [38], SGZ [44], and NeRCo [28] use either attention maps or enhancement factors to localize low-light regions, yet they often miss dim pixels. In addition, RUAS [39], Zero-DCE [12], SCI [43], SGZ [44], and Zero-IG [45] obviously change the intrinsic hue of sample #CL. This is because their core insight is luminance transfer but ignores color consistency. Silva [35] introduces artificial artifacts in all samples. This is because Silva [35] localizes low-light regions through thresholding, which fails to apply to the complex PSRs. Subsequently, Silva [35] employs low-light pixels multiplied by the illumination ratio to restore luminance, while this violent multiplication operation leads to severe content distortion. Zero-Restore [40] fails to apply to samples #CL and #SC, leading to image content contamination. Although ROP [36] and DUNP [41] improve luminance, they are not stable. For example, ROP [36] fails to restore the luminance of sample #N12, while DUNP [41] overexposes sample #2. In terms of detail, instead

Fig. 6. Qualitative comparison on LOL [49].

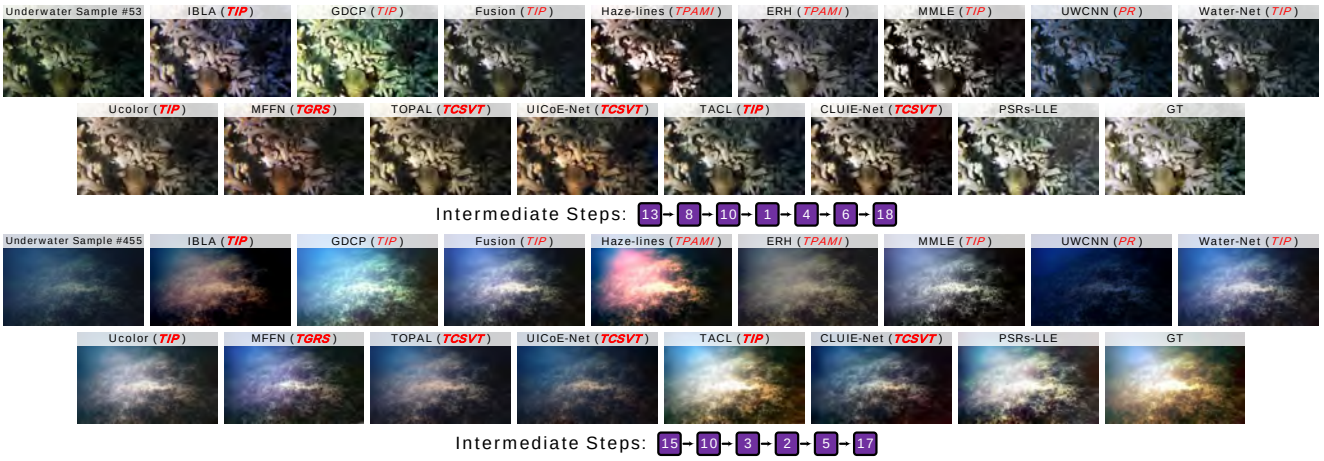

Fig. 7. Qualitative comparison on UIEB [52].

of enhancing detail, most compared methods lead to detail loss. For example, RUAS [39], Zero-Restore [40], DUNP [41], SCI [43], SGZ [44], and Zero-IG [45] toward samples #CL and #SC. This is because reflectance, reflecting texture and detail, is not optimized. In contrast, our PSRs-LLE not only restores luminance, but enriches detail and contrast. More importantly, PSRs-LLE maintains color consistency before and after enhancement. This is credited to our action set covering luminance, detail, color, and contrast enhancement operations. PSRs-LLE is able to select actions to address the corresponding quality degradation based on the feedback of perceptual loss.

C. Quantitative Comparison

We report the NIQE, CEIQ, EME, and LOE scores of PSRs-LLE and competitors in Table I. In Table I, our method achieves three best scores and one second-best score. Compared with the top-performing competitor, PSRs-LLE accomplishes the percentage gain of 5.2% / 3.7% / 10.2% in terms of CEIQ / EME / LOE. There is an interesting finding from the quantitative comparison. Although both Zero-DCE [12] and our method refer to commercial photo retouchers, their performance differs significantly. This is because Zero-DCE [12] only employs a neural network to fit illumination curves, ignoring other enhancement preferences. However, detail, color, and contrast are likewise quite promising for PSRs. This likewise suggests that employing reinforcement learning to organize a series of actions to accomplish illumination restoration is reasonable.

D. Scenario Adaptability

To validate the scenario adaptability of PSRs-LLE, we generalize it to the remote sensing scene, the natural scene, and the underwater scene. Similarly, all competitors are retrained on the corresponding scenario datasets and achieve the best quantitative scores.

Remote Sensing Scene. For qualitative and quantitative comparisons, we establish a remote sensing low-light enhancement dataset (RSLLE-4K)[1]. RSLLE-4K comprises 100 low-light samples with 4K resolution, captured over the Sierra Nevada, Rocky, and Andes mountains. Moreover, we provide 100 normal-light samples with similar but inconsistent content and luminance, again at 4K resolution. Thus, RSLLE-4K is able to support zero-shot and unsupervised methods. For training, we divide 80 sample pairs into a training set, while the remaining

[1]https://github.com/chi-kaichen/RSLLE-4K

TABLE II
QUANTITATIVE COMPARISON ON LOL [49] AND UIEB [52]. ┼ REPRESENTS FULLY SUPERVISED METHODS.

Method	LOL		Method	UIEB			
	PSNR ↑	SSIM ↑		PSNR ↑	SSIM ↑	UCIQE ↑	UIQM ↑
RRDNet [37] ‡	11.1600	0.4450	IBLA [53] †	13.8129	0.3712	0.6277	1.5450
EnlightenGAN [38] ‡	18.9005	0.7627	GDCP [54] †	14.0055	0.3625	0.6253	1.5351
RUAS [39] ‡	16.4792	0.6083	Fusion [55] †	15.1764	0.4382	0.5930	1.4982
Zero-DCE [12] ‡	15.2586	0.5367	Haze-lines [56] †	14.5129	0.4366	0.6566	1.6313
Zero-Restore [40] ‡	12.6676	0.3371	ERH [57] †	14.3570	0.4021	0.5402	1.4652
DUNP [41] ‡	13.8080	0.4772	MMLE [58] †	13.5866	0.3655	0.6153	1.8579
RetinexDIP [42] ‡	9.2708	0.2940	UWCNN [59] ┼	11.1229	0.2184	0.4919	1.5570
SCI [43] ‡	14.7617	0.5525	Water-Net [52] ┼	15.8373	0.4302	0.5943	1.4814
SGZ [44] ‡	14.9161	0.4264	Ucolor [60] ┼	16.4070	0.4637	0.5673	1.3558
NeRCo [28] ‡	15.0944	0.5376	MFFN [61] ┼	14.4192	0.4056	0.5981	1.5198
LLFormer [50] ┼	23.9799	0.8305	TOPAL [62] ┼	14.5744	0.4244	0.5701	1.4660
PairLIE [51] ┼	18.4691	0.7567	UICoE-Net [63] ┼	15.3769	0.6023	0.5718	1.4790
RQ-LLIE [29] ┼	25.5871	0.8521	TACL [64] ┼	17.2567	0.5154	0.6279	1.5981
Zero-IG [45] ‡	18.1603	0.6198	CLUIE-Net [65] ┼	14.2944	0.4061	0.5869	1.5363
PSRs-LLE ‡	19.3566	0.8081	PSRs-LLE ‡	18.0450	0.7142	0.6354	1.6054

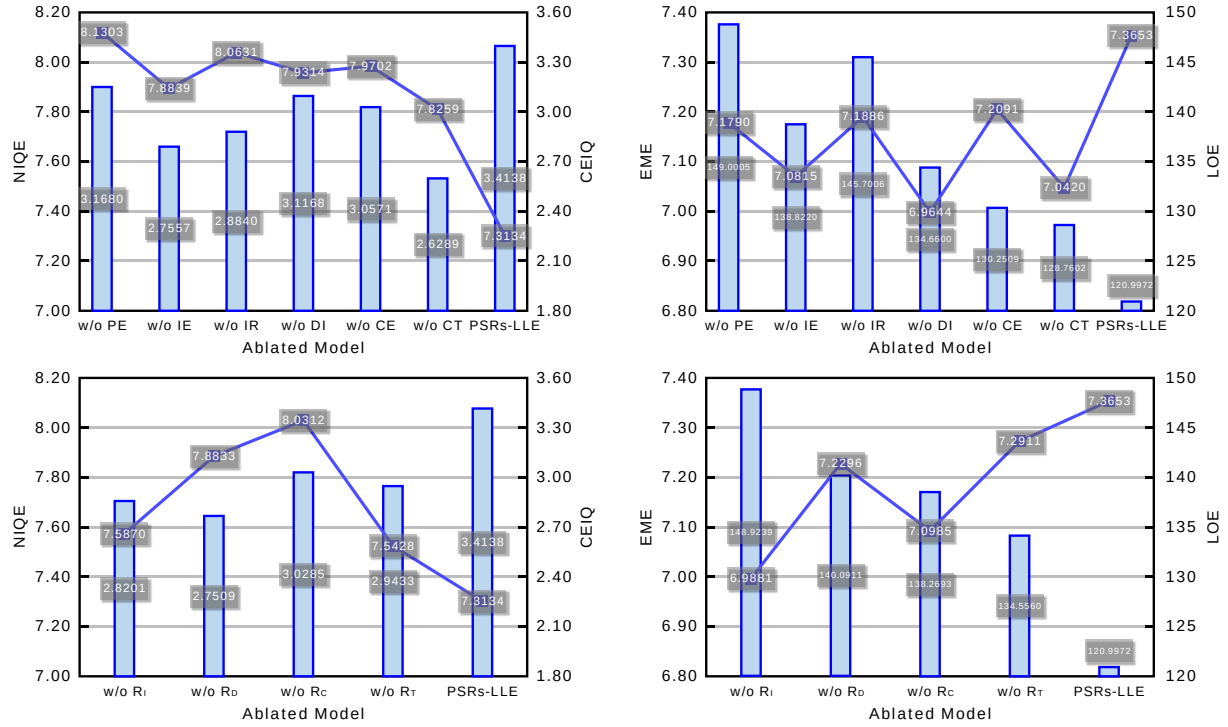

Fig. 8. Quantitative score of the ablation study on LROC-NAC [4].

20 samples are divided into a testing set. In addition, training samples are further augmented through cropping. A qualitative comparison between PSRs-LLE and the above competitors is shown in Fig. 5. RUAS [39] introduces over-exposure in sample #73 and RetinexDIP [42] introduces over-exposure in sample #8, so they are unstable. This is due to over-enhancement of the illumination component. EnlightenGAN [38], Zero-DCE [12], and Zero-Restore [40] produce yellowish, grayish, and greenish color deviations, respectively. For sample #73, ROP [36] suffers from noise. This is because ROP [36] restores illumination through spectrum estimation. However, ROP [36] employs the whole low-light sample to acquire a unified spectrum, and the noise is amplified if the distance between nearby and distant objects is too large [36]. Gong [7], Silva [35], SCI [43], SGZ [44], and NeRCo [28] underperform against tricky sample #8. Although RRDNet [37], DUNP [41], and Zero-IG [45] achieve luminance restoration, they produce unrealistic color representations. In contrast, our method restores luminance while retaining realistic colors. More importantly, PSRs-LLE outperforms all compared methods in terms of quantitative comparison, as shown in Table I. This likewise demonstrates that PSRs-LLE conquers the remote sensing scene.

Natural Scene. Qualitative and quantitative comparisons of natural scenes are performed on the LOL dataset [49]. Notably,

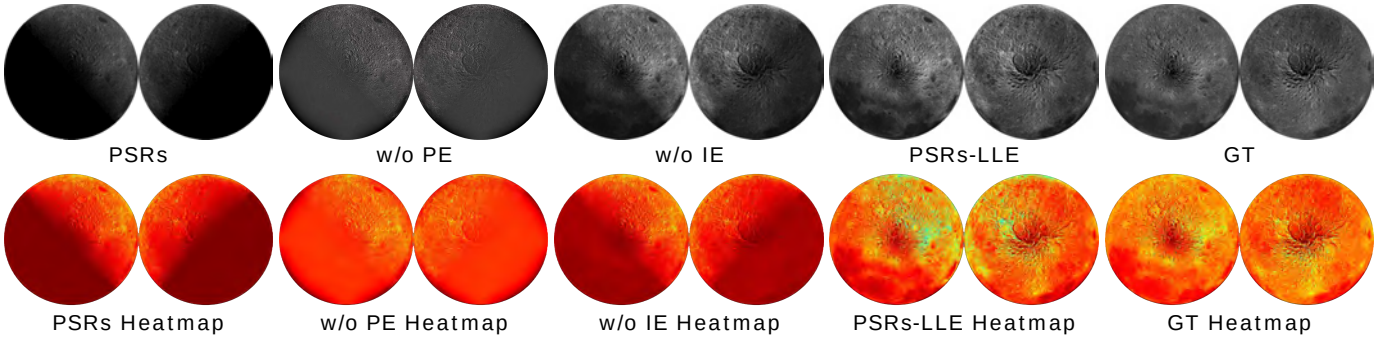

Fig. 9. Ablation study toward the feature selector. According to Retinex theory, we employ illumination maps to reflect luminance intensity. Notably, the illumination maps are rendered as heatmaps, thus highlighting the visualization differences.

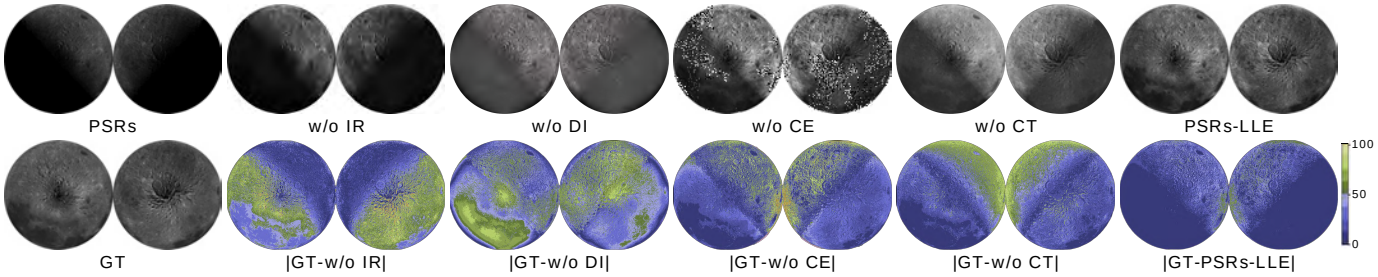

Fig. 10. Ablation study toward the action space. We employ error maps to display the visualization of pixel differences. Notably, the error map represents the absolute difference between the ground truth and the ablated/full model.

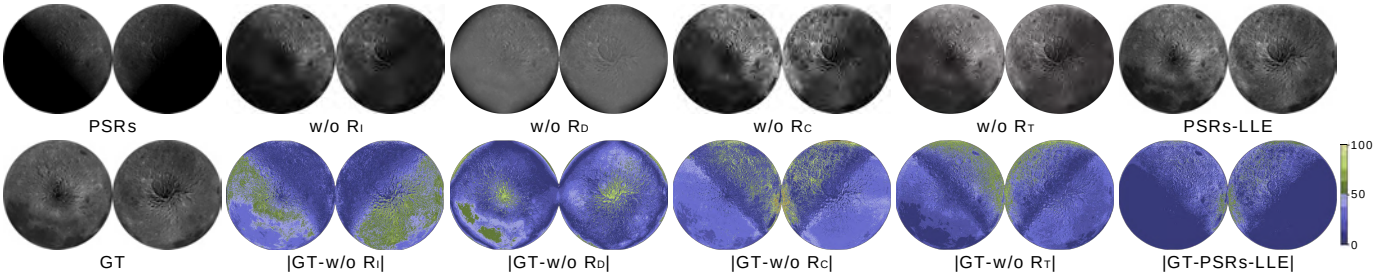

Fig. 11. Ablation study toward the reward function.

competitors add LLFormer [50], PairLIE [51], and RQ-LLIE [29]. In Fig. 6, RRDNet [37], RetinexDIP [42], SCI [43], SGZ [44], and NeRCo [28] perform poorly under low-light conditions. In addition, some compared methods introduce greenish color deviations in sample #55 and sample #665, such as EnlightenGAN [38], RUAS [39], Zero-DCE [12], Zero-Restore [40], DUNP [41], PairLIE [51], and Zero-IG [45]. In contrast, our method, LLFormer [50], and RQ-LLIE [29] are closer to ground truths in terms of color and luminance. For LLFormer [50] and RQ-LLIE [29], the main reason for accomplishing color consistency is that they introduce color-aware strategies. Therefore, we supplement color equalization operations to the action set is reasonable. For the quantitative comparison, our method achieves the third-best PSNR score and SSIM score, as shown in Table II. As is well-known, the fully supervised method has an overwhelming advantage in full-reference evaluation. Nevertheless, our method still outperforms PairLIE [51], which is nontrivial. More importantly, compared with fully supervised methods, we do not rely on paired data, thereby providing superior usability.

Underwater Scene. We perform qualitative and quantitative experiments on the UIEB dataset [52]. Notably, we select low-light samples from UIEB for both training and testing. Besides, we compare PSRs-LLE with IBLA [53], GDCP [54], Fusion [55], Haze-lines [56], ERH [57], MMLE [58], UWCNN [59], Water-Net [52], Ucolor [60], MFFN [61], TOPAL [62], UICoE-Net [63], TACL [64], and CLUIE-Net [65]. In Fig. 7, Haze-lines [56], IBLA [53], and MFFN [61] produce reddish color deviations due to overcompensation in the red channel. Fusion [55], MMLE [58], UWCNN [59], and Water-Net [52] fail to conquer low-light. This is because Fusion [55], MMLE [58], and Water-Net [52] all employ a fusion strategy to integrate color-balanced and sharpened versions while ignoring the low-light issue. Moreover, UWCNN [59] uses the neural network to simulate Jerlov water types (10 types). However, all but Type-9 are not applicable to the low-light environment. ERH [57] veils sample #455, resulting in blurred textures and structures. This is because ERH [57] employs a fusion strategy that ignores detail enhancement. GDCP [54], Ucolor [60], TOPAL [62], UICoE-Net [63], TACL

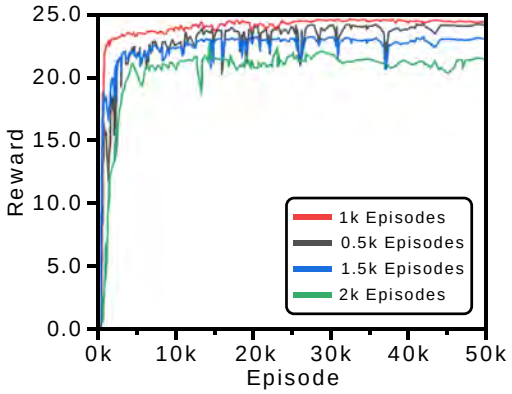

Fig. 12. Ablation study toward the parameter update interval.

[64], and CLUIE-Net [65] convert the coral in sample #53 from the whitish tone to greenish and yellowish tones. In comparison, our method restores satisfactory luminance and realistic colors. More importantly, our full-reference scores even surpass all fully supervised methods, as shown in Table II. This is because ground truths for UIEB are provided by multiple underwater image enhancement methods, and thus pseudo ground truths are the upper limit for end-to-end networks. Conversely, an exploration and exploitation training manner enables our method to take trails beyond the reference. The second-best UCIQE [66] score and the third-best UIQM [67] score demonstrate that our method achieves an excellent balance among color, luminance, saturation, sharpness, and contrast. In summary, our action set accounts for nearly all common degradation factors. With the reasonable combination, the action flow is applicable to multiple scenarios. This excellent scene adaptation competitors do not possess.

E. Ablation Study

In this subsection, we explore the contribution of action spaces and reward functions. We study the effectiveness of the feature selector. Besides, we analyze the effect of the parameter update interval. More specifically,

- w/o PE and w/o IE refer to the feature selector without the perception encoder and the illumination encoder, respectively.
- w/o IR, w/o DI, w/o CE, and w/o CT refer to the action space without the illumination restoration operation, the detail improvement operation, the color equalization operation, and the contrast enhancement operation, respectively.
- w/o $\mathcal{R}_{\mathcal{I}}$, w/o $\mathcal{R}_{\mathcal{D}}$, w/o $\mathcal{R}_{\mathcal{C}}$, and w/o $\mathcal{R}_{\mathcal{T}}$ refer to the reward function without the illumination reward, the detail reward, the color reward, and the contrast reward, respectively.

The quantitative scores of ablated models are depicted in Fig. 8. The effects of the feature selector, the action space, and the reward function are depicted in Figs. 9-11, respectively. In addition, the average rewards for different update intervals during training is shown in Fig. 12. The conclusions drawn from the ablation study are as follows.

- As depicted in Fig. 8, the full PSRs-LLE accomplishes the best NIQE, CIEQ, EME, and LOE scores compared with ablated models, suggesting that the integration of four enhancement operations and the design of reward functions are reasonable.
- In Fig. 9, the heatmap of the ablated model w/o PE is smooth and fails to reflect the bumpy topography. The ablated model w/o IE fails to restore luminance. In contrast, the full model balances illumination and perception representations, accomplishing a win-win situation for both luminance and detail.
- In Fig. 10 and Fig. 11, the ablated models w/o IR and w/o $\mathcal{R}_{\mathcal{I}}$ fail to restore luminance. The ablated models w/o DI and w/o $\mathcal{R}_{\mathcal{D}}$ lose fine structural and textural information. The gray-scale distributions of the ablated models w/o CE and w/o $\mathcal{R}_{\mathcal{C}}$ are not balanced. Besides, PSRs-LLE shows high contrast compared with the ablated models w/o CT and w/o $\mathcal{R}_{\mathcal{T}}$, which contributes to highlighting the topography of PSRs. Therefore, the mutual benefits of four enhancement operations allow PSRs-LLE to cope with PSRs.
- In Fig. 12, the average reward increases rapidly when the update interval is set to 1,000 episodes and stabilizes after approximately 9,000 episodes. In contrast, the average rewards of other update interval models either fluctuate more widely or fail to maximize. We therefore set the update interval to 1,000 episodes [68], as this configuration enables the agent to stably and efficiently learn an effective low-light enhancement strategy.

V. Conclusion

In this paper, we propose a reinforcement learning-driven lunar polar low-light enhancement method. Under user recommendation, our method autonomously selects a set of photo retouchers and organizes them into an optimal sequence. Such a manner breaks through the limitations of each photo retoucher and creates a complementary collaboration. More importantly, our method is more transparent than the end-to-end deep learning paradigm, where each step in the low-light enhancement process is optional, replaceable, and visible to the user. Extensive experiments have demonstrated the superiority of PSRs-LLE. In addition, PSRs-LLE is able to be generalized to multiple scenarios, which further enhances its utility.

References

[1] W. Cai, T. Xu, M. Shu, and Y. Wu, "Using the moon for on-orbit absolute radiometric calibration of GaoFen-4 PMS," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–13, May. 2024.
[2] H. Luo, B. Chen, L. Zhu, P. Chen, and S. Wang, "RCNet: Deep recurrent collaborative network for multi-view low-light image enhancement," *IEEE Trans. Multimedia*, vol. 27, pp. 2001–2014, Jan. 2025.
[3] Y. Luo, X. Chen, J. Ling, C. Huang, W. Zhou, and G. Yue, "Unsupervised low-light image enhancement with self-paced learning," *IEEE Trans. Multimedia*, vol. 27, pp. 1808–1820, Apr. 2025.
[4] F. Zhang, Z. Tu, W. Hao, Y. Chen, F. Li, and M. Ye, "Zero-shot parameter learning network for low-light image enhancement in permanently shadowed regions," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–16, Jul. 2024.
[5] T. Celik and T. Tjahjadi, "Contextual and variational contrast enhancement," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3431–3441, Dec. 2011.
[6] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2782–2790.

[7] H. Gong and D. Cosker, "Interactive shadow removal and ground truth for variable scene categories," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2014, pp. 1–11.

[8] Z. Liang, C. Li, S. Zhou, R. Feng, and C. C. Loy, "Iterative prompt learning for unsupervised backlit image enhancement," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 8060–8069.

[9] K. Chi, J. Li, W. Jing, Q. Li, and Q. Wang, "Neural implicit fourier transform for remote sensing shadow removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–10, Jun. 2024.

[10] L. Li, D. Liang, Y. Gao, S.-J. Huang, and S. Chen, "ALL-E: Aesthetics-guided low-light image enhancement," in *Proc. IJCAI Int. Jt. Conf. Artif. Intell.*, Aug. 2023, pp. 1062–1070.

[11] S. Sun *et al.*, "Underwater image enhancement with reinforcement learning," *IEEE J. Ocean. Eng.*, vol. 49, no. 1, pp. 249–261, Jan. 2024.

[12] C. Li, C. Guo, and C. C. Loy, "Learning to enhance low-light image via zero-reference deep curve estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 8, pp. 4225–4238, Aug. 2022.

[13] J.-L. Yin, B.-H. Chen, Y.-T. Peng, and H. Hwang, "Automatic intermediate generation with deep reinforcement learning for robust two-exposure image fusion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 7853–7862, Dec. 2022.

[14] W. Chen, J. Liu, T. W. S. Chow, and Y. Yuan, "STAR-RL: Spatial-temporal hierarchical reinforcement learning for interpretable pathology image super-resolution," *IEEE Trans. Med. Imaging*, vol. 43, no. 12, pp. 4368–4379, Dec. 2024.

[15] H. Wang, W. Zhang, L. Bai, and P. Ren, "Metalantis: A comprehensive underwater image enhancement framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–19, Apr. 2024.

[16] X. Deng, H. Wang, M. Xu, L. Li, and Z. Wang, "Omnidirectional image super-resolution via latitude adaptive network," *IEEE Trans. Multimedia*, vol. 25, pp. 4108–4120, Apr. 2022.

[17] K. Yu, X. Wang, C. Dong, X. Tang, and C. C. Loy, "Path-Restore: Learning network path selection for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 7078–7092, Oct. 2022.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] S. Hao, X. Han, Y. Guo, X. Xu, and M. Wang, "Low-light image enhancement with semi-decoupled decomposition," *IEEE Trans. Multimedia*, vol. 22, no. 12, pp. 3025–3038, Dec. 2020.

[23] S. Saini and P. J. Narayanan, "Specularity factorization for low-light enhancement," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 1–12.

[24] J. Li *et al.*, "Light the night: A multi-condition diffusion framework for unpaired low-light enhancement in autonomous driving," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 15205–15215.

[25] K. Chi, S. Guo, J. Chu, Q. Li, and Q. Wang, "RSMamba: Biologically plausible retinex-based Mamba for remote sensing shadow removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, pp. 1–10, Jan. 2025.

[26] Q. Wang, K. Chi, W. Jing, and Y. Yuan, "Recreating brightness from remote sensing shadow appearance," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–11, May. 2024.

[27] X. Lv *et al.*, "Fourier priors-guided diffusion for zero-shot joint low-light enhancement and deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 25378–25388.

[28] S. Yang, M. Ding, Y. Wu, Z. Li, and J. Zhang, "Implicit neural representation for cooperative low-light image enhancement," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 12872–12881.

[29] Y. Liu, T. Huang, W. Dong, F. Wu, X. Li, and G. Shi, "Low-light image enhancement with multi-stage residue quantization and brightness-aware attention," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 12106–12115.

[30] Y. Cai, H. Bian, J. Lin, H. Wang, R. Timofte, and Y. Zhang, "Retinexformer: One-stage retinex-based transformer for low-light image enhancement," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 12470–12479.

[31] V. Mnih *et al.*, "Playing atari with deep reinforcement learning," 2013, *arXiv:1312.5602*.

[32] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 586–595.

[33] Y. Wu *et al.*, "Towards a flexible semantic guided model for single image enhancement and restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 9921–9939, Dec. 2024.

[34] K. M. Lee, H. Myeong, and G. Song, "SeedNet: Automatic seed generation with deep reinforcement learning for robust interactive segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 1760–1768.

[35] G. F. Silva, G. B. Carneiro, R. Doth, L. A. Amaral, and D. F. G. de Azevedo, "Near real-time shadow detection and removal in aerial motion imagery application," *ISPRS J. Photogramm. Remote Sens.*, vol. 140, pp. 104–121, Jun. 2018.

[36] J. Liu, R. W. Liu, J. Sun, and T. Zeng, "Rank-one prior: Real-time scene recovery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 7, pp. 8845–8860, Jul. 2023.

[37] A. Zhu, L. Zhang, Y. Shen, Y. Ma, S. Zhao, and Y. Zhou, "Zero-shot restoration of underexposed images via robust retinex decomposition," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2020, pp. 1–6.

[38] Y. Jiang *et al.*, "EnlightenGAN: Deep light enhancement without paired supervision," *IEEE Trans. Image Process.*, vol. 30, pp. 2340–2349, Jan. 2021.

[39] 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 Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10556–10565.

[40] A. Kar, S. K. Dhara, D. Sen, and P. K. Biswas, "Zero-shot single image restoration through controlled perturbation of koschmieder's model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 16200–16210.

[41] J. Liang, Y. Xu, Y. Quan, B. Shi, and H. Ji, "Self-supervised low-light image enhancement using discrepant untrained network priors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 11, pp. 7332–7345, Nov. 2022.

[42] 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.

[43] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward fast, flexible, and robust low-light image enhancement," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5627–5636.

[44] S. Zheng and G. Gupta, "Semantic-guided zero-shot learning for low-light image/video enhancement," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops (WACVW)*, Jan. 2022, pp. 581–590.

[45] Y. Shi, D. Liu, L. Zhang, Y. Tian, X. Xia, and X. Fu, "ZERO-IG: Zero-shot illumination-guided joint denoising and adaptive enhancement for low-light images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 3015–3024.

[46] 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.

[47] Y. Zhou, L. Li, H. Zhu, H. Liu, S. Wang, and Y. Zhao, "No-reference quality assessment for contrast-distorted images based on multifaceted statistical representation of structure," *J. Vis. Commun. Image Represent.*, vol. 60, pp. 158–169, Apr. 2019.

[48] S. S. Agaian, B. Silver, and K. A. Panetta, "Transform coefficient histogram-based image enhancement algorithms using contrast entropy," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 741–758, Mar. 2007.

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

[50] T. Wang, K. Zhang, T. Shen, W. Luo, B. Stenger, and T. Lu, "Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method," in *Proc. AAAI Conf. Artif. Intell.*, Feb. 2023, pp. 2654–2662.

[51] Z. Fu, Y. Yang, X. Tu, Y. Huang, X. Ding, and K.-K. Ma, "Learning a simple low-light image enhancer from paired low-light instances," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 22252–22261.

[52] C. Li *et al.*, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, Feb. 2020.

[53] Y.-T. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1579–1594, Apr. 2017.

[54] Y.-T. Peng, K. Cao, and P. C. Cosman, "Generalization of the dark channel prior for single image restoration," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2856–2868, Jun. 2018.
[55] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 379–393, Jan. 2018.
[56] D. Berman, D. Levy, S. Avidan, and T. Treibitz, "Underwater single image color restoration using haze-lines and a new quantitative dataset," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 8, pp. 2822–2837, Aug. 2021.
[57] H. Song, L. Chang, Z. Chen, and P. Ren, "Enhancement-registration-homogenization (ERH): A comprehensive underwater visual reconstruction paradigm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6953–6967, Oct. 2022.
[58] W. Zhang, P. Zhuang, H.-H. Sun, G. Li, S. Kwong, and C. Li, "Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement," *IEEE Trans. Image Process.*, vol. 31, pp. 3997–4010, Jun. 2022.
[59] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognit.*, vol. 98, pp. 107038, Feb. 2020.
[60] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, "Underwater image enhancement via medium transmission-guided multi-color space embedding," *IEEE Trans. Image Process.*, vol. 30, pp. 4985–5000, May. 2021.
[61] R. Chen, Z. Cai, and W. Cao, "MFFN: An underwater sensing scene image enhancement method based on multiscale feature fusion network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, Mar. 2022.
[62] Z. Jiang, Z. Li, S. Yang, X. Fan, and R. Liu, "Target oriented perceptual adversarial fusion network for underwater image enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 10, pp. 6584–6598, Oct. 2022.
[63] Q. Qi *et al.*, "Underwater image co-enhancement with correlation feature matching and joint learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1133–1147, Mar. 2022.
[64] R. Liu, Z. Jiang, S. Yang, and X. Fan, "Twin adversarial contrastive learning for underwater image enhancement and beyond," *IEEE Trans. Image Process.*, vol. 31, pp. 4922–4936, Jul. 2022.
[65] K. Li *et al.*, "Beyond single reference for training: Underwater image enhancement via comparative learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 6, pp. 2561–2576, Jun. 2023.
[66] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6062–6071, Dec. 2015.
[67] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE J. Ocean. Eng.*, vol. 41, no. 3, pp. 541–551, Jul. 2016.
[68] R. Peng, S. Tan, X. Mo, B. Li, and J. Huang, "Employing reinforcement learning to construct a decision-making environment for image forgery localization," *IEEE Trans. Inf. Forensics Secur.*, vol. 19, pp. 4820–4834, Mar. 2024.
[69] 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, Jan. 2021.

Kaichen Chi received the B.E. degree in electronic and information engineering and the M.E. degree in communication and information system from Liaoning Technical University, Huludao, China, in 2019 and 2022 respectively. He is currently working toward the Ph.D. degree in the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. His research interests include image processing and deep learning.

Qiang Li (Member, IEEE) is currently with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University. His research interests include remote sensing image processing, particularly for image quality enhancement, object/change detection.

Jun Chu received the B.E. degree in automation from Northwestern Polytechnical University, Xi'an, China, in 2024. He is currently pursuing the M.S. degree with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. His research interests include deep learning and computer vision.

Junjie Li received the B.E. degree in software engineering from Zhengzhou University, Zhengzhou, China, in 2024. He is currently working toward the M.S. degree in the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision, pattern recognition and remote sensing.

Qi Wang (Senior Member, IEEE) received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively. He is currently a Professor with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision, pattern recognition and remote sensing. For more information, visit the link (https://crabwq.github.io/).