Self-Supervised Unfolding Network with Shared Reflectance Learning for Low-Light Image Enhancement

Jia Liu, Yu Luo, *Member, IEEE*, Guanghui Yue, *Member, IEEE*, Jie Ling, Liang Liao, *Senior Member, IEEE*, Chia-Wen Lin, *Fellow, IEEE*, Guangtao Zhai, *Fellow, IEEE*, Wei Zhou, *Senior Member, IEEE*

***Abstract*—Recently, incorporating Retinex theory with unfolding networks has attracted increasing attention in the low-light image enhancement field. However, existing methods have two limitations, i.e., ignoring the modeling of the physical prior of Retinex theory and relying on a large amount of paired data. To advance this field, we propose a novel self-supervised unfolding network, named S^{2}UNet, for the LIE task. Specifically, we formulate a novel optimization model based on the principle that content-consistent images under different illumination should share the same reflectance. The model simultaneously decomposes two illumination-different images into a shared reflectance component and two independent illumination components. Due to the absence of the normal-light image, we process the low-light image with gamma correction to create the illumination-different image pair. Then, we translate this model into a multi-stage unfolding network, in which each stage alternately optimizes the shared reflectance component and the respective illumination components of the two images. During progressive multi-stage optimization, the network inherently encodes the reflectance consistency prior by jointly estimating an optimal reflectance across varying illumination conditions. Finally, considering the presence of noise in low-light images and to suppress noise amplification, we propose a self-supervised denoising mechanism. Extensive experiments on nine benchmark datasets demonstrate that our proposed S^{2}UNet outperforms state-of-the-art unsupervised methods in terms of both quantitative metrics and visual quality, while achieving competitive performance compared to supervised methods. The source code will be available at https://github.com/J-Liu-DL/S2UNet.**

***Index Terms*—Low-light image enhancement, Retinex theory, unfolding network, self-supervised, shared reflectance learning.**

This work was supported in part by the National Natural Science Foundation of China under Grant 62371305, in part by the Project of Department of Education of Guangdong Province under Grant 2025KTSCX111, in part by the Natural Science Foundation of Shenzhen under Grant JCYJ20230808105906013, and in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515030025. (Corresponding author: Guanghui Yue)

J. Liu, Y. Luo, and J. Ling are with the School of Computer Science, Guangdong University of Technology, Guangzhou 510006, China (e-mail: 2112405108@mail2.gdut.edu.cn; yuluo@gdut.edu.cn; jling@gdut.edu.cn).

G. Yue is with the School of Biomedical Engineering, Shenzhen University Medical School, Shenzhen 518054, China (email: yueguanghui@szu.edu.cn).

L. Liao is with Hangzhou Institute of Technology, Xidian University, Hangzhou 311231, China (e-mail: liaoliangwhu@whu.edu.cn).

C.-W. Lin is with the Department of Electrical Engineering and the Institute of Communications Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan (e-mail: cwlin@ee.nthu.edu.tw).

G. Zhai is with the Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: zhaiguangtao@sjtu.edu.cn).

W. Zhou is with the School of Computer Science and Informatics, Cardiff University, CF10 3AT Cardiff, U.K. (e-mail: zhouw26@cardiff.ac.uk).

I. INTRODUCTION

IMAGE capture under low-light conditions is a common and unavoidable scenario. However, images captured in such conditions often suffer from various unpredictable forms of image degradation, such as lower contrast, increased noise, detail loss, and color distortion [1]–[4]. Restoring quality and enhancing visual effects for such low-light images typically requires manual post-processing with professional editing software, which is not only time-consuming and laborious but also demands professional skills [5], [6]. Consequently, low-light image enhancement (LIE) holds significant research importance and practical value.

Over the past few decades, numerous attempts have been made in LIE, including histogram equalization [7], gamma correction [8], and Retinex-based methods [9]. Among these traditional methods, Retinex-based methods have received particular attention due to their theoretical foundations and practical effectiveness. Based on Retinex theory [10], an observed image can be modeled as the product of reflectance and illumination. The reflectance encodes the inherent characteristics of the scene, which are assumed to be invariant to changes in illumination conditions. Inspired by this, some studies enhance low-light images by directly estimating and refining the reflectance [11]. Others instead focus on modifying the illumination, which is then fused with the reflectance to generate the enhanced output [12]. To address this ill-posed decomposition problem, various handcrafted priors have been introduced to constrain the solution space [13], [14]. However, these handcrafted priors prove difficult to adapt to the diverse and complex scenarios encountered in practical applications, which restricts the practical deployment of such methods.

The rise of deep learning has spurred the integration of Retinex-based concepts into neural networks to tackle the LIE problem [15]–[18]. Representative methods directly decompose the input into reflectance and illumination via encoder-decoder networks [15], [16] or cascaded decomposition-adjustment networks [17], [18]. These deep learning-based methods achieve superior performance and adaptability compared to traditional methods. However, as data-driven models, their black-box characteristics are highly controversial. Recently, more and more studies have attempted to incorporate deep unfolding networks with Retinex theory to improve both the enhancement effect and model interpretability [19]–[21]. These methods translate the iterative process of the tradi-

tional Retinex optimization model into a multi-stage unfolding network, where each stage corresponds to a single iteration step towards estimating the reflectance and illumination components. Generally, these methods usually estimate the reflectance component and minimize the reflectance difference between low-light and reference images.

Although effective in certain scenarios, existing unfolding-based LIE methods exhibit several limitations. On the one hand, decomposing low-light and reference images separately ignores the modeling of the reflectance consistency prior inherent in Retinex theory. This paradigm also prevents effective information interaction across different illumination conditions. In this case, the network may simply memorize patterns from previously seen images, instead of acquiring a generalizable capability to uncover latent reflectance. On the other hand, most existing methods rely on a large number of paired data. However, in real-world scenarios, acquiring a high-quality reference image that precisely corresponds to a given low-light input is often infeasible.

To advance this field, this paper formulates a new optimization model for the LIE task, which enforces a shared reflectance component while simultaneously decomposing illumination-different images. Based on this model, we propose a self-supervised unfolding network, named S^{2}UNet. Specifically, the network takes a low-light image and its illumination-different auxiliary image as input for decomposition and joint optimization. To construct the required image pair without external data, we generate the auxiliary image from the low-light image by applying gamma correction. In S^{2}UNet, an Initial Decomposition (ID) module first decomposes these two images into their respective reflectance and illumination components, providing initial estimates for the subsequent UnFolding (UF) module, which includes multiple unfolding stages. At each stage, the UF module iteratively optimizes the shared reflectance component and the independent illumination components. Through progressive multi-stage optimization, the network leverages complementary information from both images to collaboratively estimate the reflectance. This process results in an optimal and consistent reflectance across varying illumination. Finally, the output of the UF module, i.e., the reflectance component, is considered the enhanced image. Considering the presence of noise in low-light images and to suppress noise amplification, we propose a self-supervised denoising mechanism.

Our study differs fundamentally from existing unfolding-based LIE methods in several key aspects, as shown in Fig. 1. First, in terms of the constraint for reflectance consistency, existing methods [19], [22] generally follow a learning-driven paradigm and typically keep reflectance consistency via the loss function in the last stage of the network. Under this paradigm, the network fails to fully ensure the decomposition results satisfying reflectance consistency, which may lead to suboptimal enhancement performance. In contrast, by enforcing an image and its auxiliary version to share the same reflectance component during decomposition through model optimization at each stage of the network, our method directly embeds the prior of reflectance consistency into the network architecture, resulting in better enhancement performance.

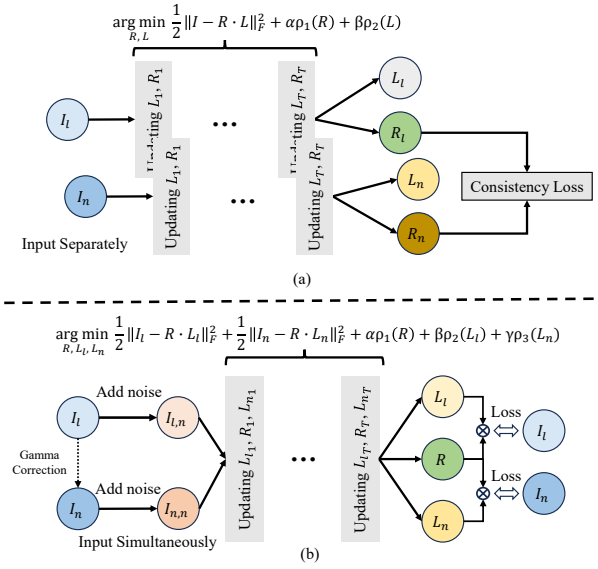


Fig. 1. Comparison between existing unfolding-based LIE methods and our proposed paradigm. (a) Existing methods separately decompose low-light and reference images, and learn the reflectance mapping relationship between the two images through a loss function; and (b) our method introduces a joint optimization model that directly embeds the reflectance-consistent physical prior into the network without reference image.

Second, existing unfolding-based methods [20], [21] usually require aligned low-light and normal-light image pairs during optimization. In contrast, we construct a novel joint optimization model that simultaneously takes the low-light image and its gamma-corrected auxiliary image as decomposition targets. This allows the network to leverage illumination variation to enhance optimization, without requiring any normal-light images as reference. Finally, contrary to methods [23], [24] that handle denoising and enhancement sequentially, our method integrates a self-supervised denoising mechanism into the Retinex decomposition process, achieving end-to-end joint enhancement and denoising.

The core contributions of this study can be outlined as follows:

- By incorporating the Retinex theory and unfolding networks, we formulate a new optimization model for the LIE task. This model shares an identical reflectance between illumination-different images during the optimization process, effectively encoding the physical prior of Retinex theory into the network.
- To address the challenge of the absence of normal-light references, we propose to utilize gamma correction to create an illumination-different auxiliary image for the low-light image and propose a self-supervised denoising mechanism to achieve simultaneous enhancement and denoising in a completely self-supervised manner.
- Extensive experiments on nine benchmark datasets show that the proposed LIE method outperforms three conventional methods and eight state-of-the-art unsupervised methods while achieving competitive performance compared to three supervised methods. Also, it increases the performance of the downstream face detection task by providing high-quality inputs.

II. RELATED WORK

A. Traditional Methods

Traditional methods for LIE are often classified into three representative types: Histogram Equalization (HE), Gamma Correction (GC), and Retinex-based methods. Among them, HE expands the dynamic range by redistributing pixel intensity values [25], [26], while GC employs power-law transformations to adjust pixel values [27], [28]. Although these two types of methods effectively enhance overall image brightness and contrast, they commonly suffer from issues such as over-enhancement, loss of detail, and noise amplification in complex scenarios. The Retinex theory models an image as a combination of reflectance and illumination components [10], where the former preserves scene details and object colors, while the latter accounts for brightness variations. In recent years, the use of the Gaussian filter to decompose the reflectance and illumination components of an image has become popular [29], [30]. While filter-based Retinex methods are computationally efficient, they often suffer from halo artifacts, detail loss, and reliance on manual parameter tuning.

Alternatively, some methods formulate energy functions with handcrafted priors under a variational framework and solve the Retinex decomposition problem through iterative optimization [31]–[33]. Ng et al. [9] introduced total variation regularization terms to enforce spatial smoothness in the illumination and piecewise continuity in the reflectance. Guo et al. [31] initialized the illumination using the maximum value of the pixel channel, and refined the illumination estimation using an augmented Lagrangian multiplier method along with a weighting strategy. Fu et al. [32] proposed a weighted variational model that introduces exponential weighting in the regularization term to mitigate the unbalanced effects of logarithmic gradients in bright and dark regions. Li et al. [33] incorporated an explicit noise map into their framework and proposed a novel objective function to simultaneously enhance low-light images and suppress noise.

Although Retinex-based methods help improve image brightness and contrast, they heavily rely on handcrafted priors. The design of such priors is not only challenging but also typically based on ideal assumptions, making them difficult to adapt to complex real-world scenarios.

B. Deep Learning-Based Methods

Recently, deep learning has achieved remarkable success in low-level computer vision tasks [34]–[40], especially in the LIE field [41], [42]. Deep learning-based LIE methods can be classified as supervised and unsupervised methods. Supervised methods require strictly aligned image pairs to learn mappings from low-light to normal-light domains for enhancement. In the work of Wei et al. [43], the image is first decomposed into reflectance and illumination components, after which illumination refinement is conducted using an enhancement network. Building upon [43], Zhang et al. [17] implemented degradation removal and flexible brightness adjustment through dedicated reflectance restoration and illumination adjustment modules after decomposition. Wang et al. [44] introduced an intermediate illumination representation to establish correlations between the low-light input and the expected enhanced results. Liu et al. [45] introduced a three-stage framework, where residual-quantized codebooks learn normal illumination priors. While these methods typically yield good enhancement results, their dependence on paired data limits their applicability.

Unsupervised methods eliminate the dependency on paired training data by leveraging physical priors or self-supervised strategies for model training. While such methods still have room for performance improvement, they offer greater practical applicability in real-world scenarios. Kandula et al. [46] developed a two-stage unsupervised LIE framework, generating coarse estimates in the first stage, followed by a context-adaptive illumination-guided norm for refinement. Guo et al. [47] employed a lightweight neural network to estimate an image-specific set of curves for pixel-wise dynamic range adjustment, achieving zero-reference training. Luo et al. [48] generated a pseudo high-quality reference for the low-light image using quadratic curves, and fed the image pairs into parallel homogeneous branches for mutual learning, resulting in a good enhancement effect. Wu et al. [49] designed a deep neural network guided by Retinex theory to concurrently address illumination adjustment and denoising without supervised labels.

While achieving performance improvements, these deep learning-based methods primarily focus on designing neural network architectures that lack interpretability, making their working mechanisms unclear.

C. Unfolding-Based Methods

Generally, model-based methods use mathematical constraints through physical mechanisms, yielding outputs with clear physical interpretability. Learning-based methods, on the other hand, demonstrate unique advantages in feature representation through mapping relationship learning using large datasets. Deep unfolding methods, combining the strengths of both paradigms, have emerged as a novel direction for addressing LIE problems. Liu et al. [50] pioneered an unsupervised unfolding framework for illumination estimation and denoising, employing neural architecture search to automatically discover low-light priors. However, this method does not explicitly formulate the optimization objective of the Retinex model, neglecting the physical coupling between illumination and reflectance. Subsequently, Wu et al. [19] unfolded the Retinex decomposition process into a trainable network architecture, allowing the model to learn implicit priors from data adaptively. Zheng et al. [22] integrated a pre-trained masked autoencoder into the proximal operator network to customize learnable priors for the deep unfolding paradigm. Liu et al. [20] designed an optimization model for algorithm unrolling, which integrates explicit structure-revealing priors with network-learned implicit priors, to facilitate better decomposition. Wang et al. [21] unfolded a multi-scale Retinex optimization model to enforce consistency in multi-scale reflectance maps.

Although these methods attempt to incorporate unfolding frameworks with Retinex theory, they independently decompose differently illuminated images and lack essential

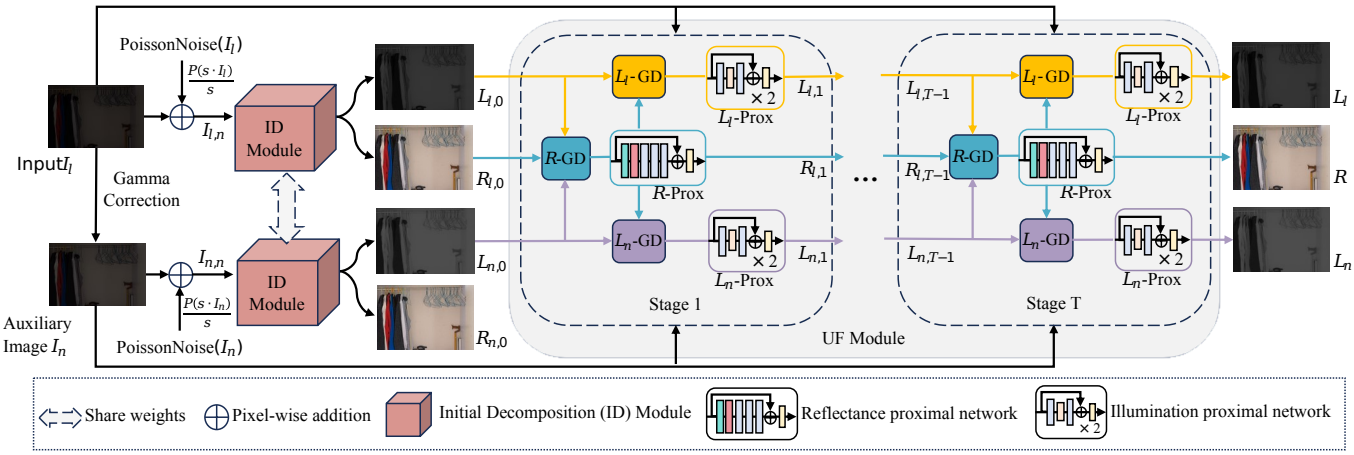


Fig. 2. The overall architecture of our proposed S^{2}UNet. It includes an ID module and a UF module. The former is used to obtain the initial estimates of the reflectance and illumination, and the latter iteratively optimizes the shared reflectance $R_{l,0}$ and independent illumination $L_{l,0}$ and $L_{n,0}$ of the two images. The UF module includes T unfolding stages, where each stage alternately updates three variables ($R_{l,k}$, $L_{l,k}$ and $L_{n,k}$) according to Eq. (7). The process first updates the shared reflectance $R_{l,k}$, then updates both illuminations $L_{l,k}$ and $L_{n,k}$ based on the updated $R_{l,k}$. Here, X-GD denotes gradient descent computations, and X-Prox means the proximal network.

information interaction and co-optimization, resulting in a poor enhancement effect under complex scenarios. Moreover, these methods all employ fully supervised training, ignoring the challenges of acquiring precisely aligned low-light and normal-light image pairs for real-world applications.

III. PROPOSED METHOD

A. Motivation

According to Retinex theory, the observed image I is typically modeled as the element-wise product of the reflectance R and the illumination L:

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

These two components can be estimated by minimizing the following energy function as an optimization paradigm:

$$E(R, L) = \frac{1}{2}\|I - R \cdot L\|_F^2 + \alpha\rho_1(R) + \beta\rho_2(L), \tag{2}$$

where $\|\cdot\|_F$ indicates the Frobenius norm, ρ_1 and ρ_2 are regularization terms that encode the priors imposed on R and L, and α and β represent trade-off parameters.

Existing unfolding-based methods typically follow a similar optimization model in Eq. (2) and unfold it into deep neural networks. However, these methods rely on supervised training with aligned low- and normal-light image pairs, and their optimization framework fails to incorporate the essential prior knowledge that reflectance should remain invariant under varying illumination conditions. To address these limitations, we propose a joint optimization model that simultaneously decomposes two illumination-different images into a shared reflectance component and two independent illumination components. To construct the required image pairs, we employ gamma correction to generate an illumination-different auxiliary image for the low-light image. In this way, the two images have different illumination yet the same reflectance. Considering the absence of the normal-light reference, we leverage self-supervised training using these constructed image pairs and estimate the optimal reflectance under varying illumination conditions through shared reflectance learning.

B. Proposed Optimization Model

In contrast to Eq. (2), we derive an auxiliary image I_n from the low-light input I_l via gamma correction and estimate the reflectance and illumination through energy minimization with regularization:

$$E(R, L_l, L_n) = \frac{1}{2}\|I_l - R \cdot L_l\|_F^2 + \frac{1}{2}\|I_n - R \cdot L_n\|_F^2 + \alpha\rho_1(R) + \beta\rho_2(L_l) + \gamma\rho_3(L_n), \tag{3}$$

where R denotes the shared reflectance component between the two images, L_l and L_n represent the illumination components of I_l and I_n, respectively. ρ_1, ρ_2, and ρ_3 are regularization terms encoding the priors imposed on R, I_l, and I_n, while α, β, and γ are trade-off parameters.

By fixing other variables and alternately updating individual variables through iterative optimization, we can decompose the problem of minimizing Eq. (3) into the following three univariate subproblems:

$$R_k = \arg\min_R \frac{1}{2}\|I_l - R \cdot L_{l,k-1}\|_F^2 + \frac{1}{2}\|I_n - R \cdot L_{n,k-1}\|_F^2 + \alpha\rho_1(R), \tag{4}$$

$$L_{l,k} = \arg\min_{L_l} \frac{1}{2}\|I_l - R_k \cdot L_l\|_F^2 + \beta\rho_2(L_l), \tag{5}$$

$$L_{n,k} = \arg\min_{L_n} \frac{1}{2}\|I_n - R_k \cdot L_n\|_F^2 + \gamma\rho_3(L_n), \tag{6}$$

where k denotes the iteration index. In this study, we apply the proximal gradient descent algorithm [51] to solve the aforementioned three subproblems.

Update rule: From the proximal gradient descent algorithm, we derive the following update steps:

$$\begin{cases} X_k' = X_{k-1} - \lambda_k \nabla f(X_{k-1}), \\ X_k = \text{prox}(X_k'), \end{cases} \tag{7}$$

where X denotes the variables to be updated, $\nabla f(\cdot)$ is the gradient computation, λ is the learnable step size, and $\text{prox}(\cdot)$

represents the proximal network. We iteratively update R, L_l, and L_n via Eq. (7), where $\nabla f(R_{k-1}) = L_{l,k-1} \cdot (R_{k-1} \cdot L_{l,k-1} - I_l) + L_{n,k-1} \cdot (R_{k-1} \cdot L_{n,k-1} - I_n)$, $\nabla f(L_{l,k-1}) = R_k \cdot (R_k \cdot L_{l,k-1} - I_l)$, and $\nabla f(L_{n,k-1}) = R_k \cdot (R_k \cdot L_{n,k-1} - I_n)$.

Compared to traditional optimization paradigm shown in Eq. (2), our optimization paradigm (see Eq. (3)) has the following theoretical differences. On the one hand, traditional paradigm fail to express the reflectance consistency prior in the decomposition stage, whereas our paradigm enforces this prior as a prerequisite by making the two data fidelity terms share the same reflectance variable. This design ensures that the decomposition results inherently follow the physical prior, enhancing both interpretability and theoretical self-consistency of the model. On the other hand, traditional paradigm models and solves for a single image, which usually leads to ill-posedness. In contrast, our paradigm simultaneously treats two illumination-different images as observations, providing additional information and constraints from the same scene. This can reduce the ambiguity of the solution space, leading to more stable and reliable results.

C. Overall Network Architecture

Fig. 2 presents the overall architecture of our proposed method, named S^{2}UNet. Given a low-light image I_l, we first generate an auxiliary image I_n by processing it with gamma correction. We then add Poisson noise to both I_l and I_n to form training inputs $I_{l,n}$ and $I_{n,n}$. Specifically, for each pixel of the input, we extract its value and scale it to the range of [0, s] using a scale factor of s. This scaled value is fed into a Poisson distribution to obtain a sampled Poisson random value, which is then re-scaled back to the original range (i.e., [0, 1]). After processing all pixels, we can produce the noisy image of the input. Following this, an Initial Decomposition (ID) module is used to produce preliminary reflectance components ($R_{l,0}$, $R_{n,0}$) and illumination components ($L_{l,0}$, $L_{n,0}$) for both images, respectively. Next, $R_{l,0}$, $L_{l,0}$, and $L_{n,0}$ are fed into a well-designed UnFolding (UF) module for iterative optimization. This module translates the iterative update steps described in Section III-B into an unfolding module consisting of T unfolding stages, where each stage corresponds to one optimization iteration. The optimization process enforces both images to maintain identical reflectance while alternately optimizing this shared reflectance component and their respective illumination components. This compels the network to fully leverage information from both images to jointly estimate optimal reflectance under varying illumination conditions, thereby encoding the reflectance consistency physical prior of Retinex theory into the network architecture. The final stage outputs the optimized results R, L_l, and L_n. Following previous works [16], [31], [44], we use the refined reflectance R as our final enhancement result.

For an observed low-light image, it is difficult to obtain an image with different illumination yet the same reflectance in practice. In this study, we propose to operate a controllable transformation to generate an auxiliary image that has the same reflectance but different illumination for the observed low-light image. In this way, we are able to build the required

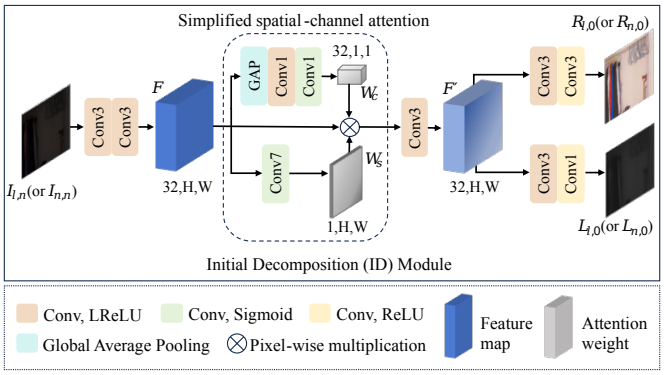


Fig. 3. Structure of the ID module.

image pairs for joint optimization. Following [52], for low-light image $I_l = R \cdot L_l$, we apply gamma correction to generate its auxiliary image I_n:

$$I_n = I_l^\gamma = R^\gamma \cdot L_l^\gamma = R \cdot R^{\gamma-1} \cdot L_l^\gamma \approx R \cdot L_l^\gamma, \qquad (8)$$

where γ denotes the correction factor. $R^{\gamma-1} \approx 1$ when γ approaches 1. According to the experiment in Section IV-B5, we set γ to 0.8.

D. Initial Decomposition Module

The ID module aims to provide high-quality initialization for subsequent unfolding optimization. Fig. 3 illustrates its structure. The input noised image, i.e., $I_{l,n}$ or $I_{n,n}$, is first passed through two 3×3 convolution layers with the LReLU activation function to extract a preliminary feature F. F is then processed by a simplified spatial-channel attention and a 3×3 convolution layer with the LReLU activation function, resulting in a refined feature F' that highlights important feature regions. In this study, the spatial attention uses a 7×7 convolution layer with the Sigmoid activation function to produce spatial attention weights W_s, while the channel attention applies global average pooling followed by two 1×1 convolution layers with LReLU and Sigmoid activation functions to generate channel attention weights W_c. It is worth noting that our focus is to utilize the attention mechanism for effective feature extraction, and selecting an optimal attention block is beyond the scope of the current study. We leave the attention selection task as future work. Finally, two separate convolution branches are used to generate initial estimates of reflectance $R_{l,0}$ (or $R_{n,0}$) and illumination $L_{l,0}$ (or $L_{n,0}$), respectively. The reflectance branch contains two 3×3 convolution layers, each followed by an LReLU activation function and a ReLU activation function, respectively. The illumination branch consists of a 3×3 convolution layer followed by an LReLU activation function, and a 1×1 convolution layer followed by a ReLU activation function.

E. Unfolding Module

The unfolding module takes the decomposition results of the ID module as initial values and alternately optimizes the reflectance R shared by the two images and their respective illuminations L_l and L_n. It translates the iterative optimization

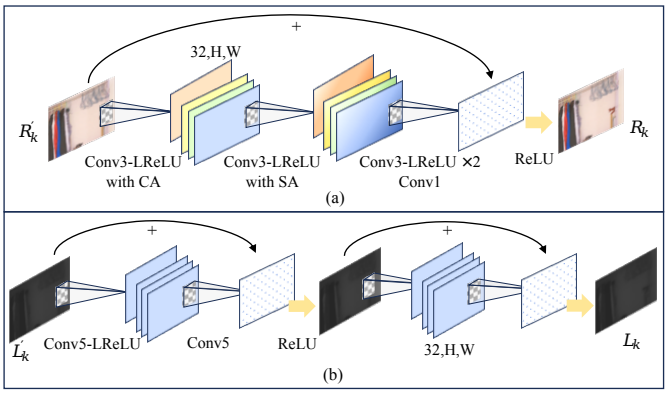


Fig. 4. Structure of the proximal networks at the k-th unfolding stage: (a) Reflectance proximal network (R-Prox), and (b) Illumination proximal network (L_l-Prox and L_n-Prox, which share identical network structure).

process into T unfolding stages, each containing the same network structure, as illustrated in Fig. 2. The notations R-GD, L_l-GD, and L_n-GD represent gradient descent computations based on the formulas in Eq. (7), while R-Prox, L_l-Prox, and L_n-Prox denote the corresponding proximal networks. Taking the k-th unfolding stage as an example: First, R-GD performs gradient descent using R_{k-1}, $L_{l,k-1}$, and $L_{n,k-1}$ from the previous stage. Then, the reflectance proximal network R-Prox generates the updated R_k for the current stage. Similarly, $L_{l,k-1}$ and $L_{n,k-1}$ are updated based on R_k to obtain $L_{l,k}$ and $L_{n,k}$, where illumination proximal networks L_l-Prox and L_n-Prox use the same structure.

1) Reflectance Proximal Network: The specific structure of the reflectance proximal network is shown in Fig. 4(a), in which the learning process of the adaptive reflectance prior is formulated in a residual manner. Specifically, this network consists of five convolution layers (each of the first four layers includes a 3×3 convolution and a LReLU activation function, while the last layer consists of a 1×1 convolution and a residual connection operation to the input, followed by a ReLU activation function). Channel and spatial attention modules [53] are used after the first and second convolution layers to enhance the expression of key structural and textural features in the reflectance.

2) Illumination Proximal Network: As detailed in Fig. 4(b), the illumination proximal network employs two cascaded residual blocks to learn adaptive illumination priors in a residual manner. Each residual block consists of two 5×5 convolution layers, considering that larger receptive fields help to better model the smooth characteristics of illumination. An LReLU/ReLU activation function is used after the first/second convolution layer. Additionally, a residual connection operation to the input is added to the second convolution layer.

F. Self-Supervised Denoising Mechanism

Considering the presence of noise in low-light images and to suppress noise amplification, inspired by N2N [54], we implement a self-supervised denoising mechanism during network training. Specifically, we inject Poisson noise into I_l and I_n to generate noisy image pairs $I_{l,n}$ and $I_{n,n}$ as input to the network :

$$\begin{cases} I_{l,n} = \frac{\mathcal{P}(s \cdot I_l)}{s} = R \cdot L_l + \epsilon_l, \\ I_{n,n} = \frac{\mathcal{P}(s \cdot I_n)}{s} = R \cdot L_n + \epsilon_n, \end{cases} \tag{9}$$

where $\mathcal{P}(\cdot)$ denotes Poisson sampling, and ϵ indicates zero-mean Poisson noise correlated with image intensity. Motivated by the observation that photon arrival at sensors follows a Poisson process [55], we simulate the stochasticity of actual photon counts received at each pixel location by performing Poisson sampling after scaling pixel values in Eq. (9), where s is the scaling factor. Due to the smoothness of illumination, we assume that noise is primarily introduced to the reflectance component. Based on this, we introduce a consistency loss L_{con}, as detailed later in Section III-G. L_{con} guides the network to ignore noise and extract the identical underlying reflectance during the decomposition process. Thus, our method is robust to the perturbations from Poisson sampling, effectively preserving the property of reflectance consistency. Accordingly, we update Eq. (9) as follows:

$$\begin{cases} I_{l,n} = (R + N_l) \cdot L_l = R_{l,n} \cdot L_l, \\ I_{n,n} = (R + N_n) \cdot L_n = R_{n,n} \cdot L_n, \end{cases} \tag{10}$$

where N_l and N_n denote ϵ_l / L_l and ϵ_n / L_n, respectively. After generating these noisy images, i.e., $I_{l,n}$ and $I_{n,n}$, we feed them into the ID module to generate initial reflectance ($R_{l,0}$ and $R_{n,0}$) and illumination ($L_{l,0}$ and $L_{n,0}$). To endow the ID module with the ability to denoise, the obtained reflectance and illumination should reconstruct the original images, i.e., $R_{l,0} \cdot L_{l,0} \approx I_l$ and $R_{n,0} \cdot L_{n,0} \approx I_n$, which is constrained by an initial reconstruction loss L_{ir}, introduced later in Section III-G. Subsequently, $R_{l,0}$, $L_{l,0}$, and $L_{n,0}$ are fed into the UF module, which performs multi-stage iterative optimization on these variables through shared reflectance learning. At each stage, we also require the shared reflectance R_k and respective illuminations $L_{l,k}$ and $L_{n,k}$ to reconstruct the original images, i.e., $R_k \cdot L_{l,k} \approx I_l$ and $R_k \cdot L_{n,k} \approx I_n$, through a reconstruction loss L_{rec}, introduced later in Section III-G.

The extremely low signal-to-noise ratio makes accurately separating the noise within the image a challenging ill-posed problem. Rather than separating noise directly, we guide the network via decomposition-reconstruction constraints to recover a clean reflectance, i.e., the enhanced result. This effectively recasts denoising as a self-supervised prior learning task intrinsically aligned with Retinex decomposition. Unlike methods that treat denoising as an independent pre-/post-processing step or rely on explicit noise modeling, our perspective leverages the need for accurate image reconstruction to force the network to internally learn a noise-robust representation of the reflectance. This reconstruction-oriented learning idea aligns the denoising process with the optimization objective, thereby achieving joint enhancement and denoising. To the best of our knowledge, we conduct the pioneering attempt to explore this research perspective, offering a more natural and integrated solution to the LIE challenge. Notably, our focus is not on the technical improvement of the N2N method, but on its cohesive integration to address the denoising challenge within the LIE task, which differs from existing cascaded solutions.

G. Loss Function

In summary, our network is optimized end-to-end, guided by a combined loss defined as the sum of the losses from both the ID and unfolding modules:

$$L_{total} = L_{id} + L_{uf}. \tag{11}$$

1) Loss Function for the ID Module: The ID provides initial estimates that simultaneously comply with Retinex physical principles and exhibit favorable numerical properties for subsequent unfolding optimization. To reduce optimization complexity and enhance convergence efficiency, it includes three key loss constraints:

$$L_{id} = L_{ir} + L_{con} + L_{iie}. \tag{12}$$

where $L_{ir} = \|I_l - R_{l,0} \cdot L_{l,0}\|_1 + \|I_n - R_{n,0} \cdot L_{n,0}\|_1$ is the initial reconstruction loss that enforces adherence to the Retinex model, $L_{con} = \|R_{l,0} - R_{n,0}\|_1$ is the consistency loss that ensures reflectance invariance under different illumination conditions, and $L_{iie} = \|L_{l,0} - \max_{c\in\{R,G,B\}} I_l^{(c)}\|_1 + \|L_{n,0} - \max_{c\in\{R,G,B\}} I_n^{(c)}\|_1$ is the initial illumination estimation loss that implements the initial illumination approximation using maximum RGB channel values per pixel [31]. $\|\cdot\|_1$ denotes the l_1 norm.

2) Loss Function for the UF Module: The unfolding module unfolds the iteration process of our proposed joint optimization model into T stages, where the network progressively updates the shared reflectance R and respective illuminations L_l and L_n of two images. The loss L_{uf} for the UF module is defined as follows:

$$L_{uf} = L_{rec} + L_R + L_s + \mu L_{il}, \tag{13}$$

where μ is the trade-off parameter. L_{rec}, L_R, L_s, and L_{il} are the reconstruction loss, the reflectance loss, the structure-aware illumination smoothness loss, and the illumination estimation loss, respectively.

The complex prior modeling of the optimization objective is decomposed into the proximal networks at each stage k ($k \in \{1, 2, \cdots, T\}$). Based on this, we impose constraints of L_{rec}, L_R, and L_s not only on the final estimation results but also on each intermediate stage. L_{rec} ensures that the optimization results at each stage strictly satisfy the Retinex model:

$$L_{rec} = \sum\nolimits_{k=1}^{T} (||I_l - R_k \cdot L_{l,k}||_1 + ||I_n - R_k \cdot L_{n,k}||_1). \tag{14}$$

L_R is used to guide the enhancement direction and suppress the noise of reflectance. Formally, it combines the maximum entropy [16] and gradient regularization:

$$L_R = \sum\nolimits_{k=1}^{T} \omega_k \left(\lambda_1 || \max_{c\in\{R,G,B\}} R_k^{(c)} - \mathcal{H}(\max_{c\in\{R,G,B\}} I_l^{(c)})||_1 + \lambda_2 ||\nabla R_k||_1\right), \tag{15}$$

where λ_1 and λ_2 are trade-off parameters for the maximum entropy term L_{Rm} and the gradient regularization term L_{Rg}, respectively. $\mathcal{H}(\cdot)$ denotes the histogram equalization operation, and ∇ represents the gradient computation in the horizontal and vertical directions. ω_k is the weight parameter used to balance the contributions of each unfolding stage. In our experiments, ω_k is set to 0.3 when $k < T$ while to 1 when $k = T$, enabling progressive constraint to prevent local optima. L_s aims to maintain image structure while smoothing illumination:

$$L_s = \sum\nolimits_{k=1}^{T} \lambda_3 (\|\nabla L_{l,k} \cdot \exp(-\theta \nabla R_k)\|_1 + \|\nabla L_{n,k} \cdot \exp(-\theta \nabla R_k)\|_1), \tag{16}$$

where λ_3 and θ are trade-off parameters. The illumination estimation loss L_{il} is used to constrain the final optimization result of illumination:

$$L_{il} = \|L_l - \max_{c\in\{R,G,B\}} I_l^{(c)}\|_1 + \|L_n - \max_{c\in\{R,G,B\}} I_n^{(c)}\|_1. \tag{17}$$

Algorithm 1 The training procedure of our method

Input: Low-light image set I_l.
Output: Normal-light image R.
Initialize: The parameters in the ID module $\mathcal{G}_{ID}$ and the UF module $\mathcal{G}_{UF}$, and the number of unfolding stage T.
repeat
1: $I_n = I_l^{\gamma}$;
2: $I_{l,n} = \frac{\mathcal{P}(s \cdot I_l)}{s}$, $I_{n,n} = \frac{\mathcal{P}(s \cdot I_n)}{s}$;
3: $[R_{l,0}, L_{l,0}] = \mathcal{G}_{ID}(I_{l,n})$, $[R_{n,0}, L_{n,0}] = \mathcal{G}_{ID}(I_{n,n})$;
4: $R, L_l, L_n = \mathcal{G}_{UF}(R_{l,0}, L_{l,0}, L_{n,0})$:
while $1 \le k \le T$ **do**
Update R_k, $L_{l,k}$ and $L_{n,k}$ based on Eq. (7),
$k \leftarrow k + 1$
end while
5: Update the parameters of $\mathcal{G}_{ID}$ and $\mathcal{G}_{UF}$ by minimizing the loss function Eq. (11).
until convergence

IV. EXPERIMENTS

A. Experiments Setup

1) Implementation Details: Our proposed method is implemented with PyTorch on an NVIDIA GeForce RTX 3090Ti GPU. The network parameters are initialized using the Kaiming normal method. The AdamW optimizer is used to update the network parameters with a fixed learning rate of 0.001, and default β values (0.9, 0.999). The network is trained for 120 epochs with a batch size of 16. Considering the limited computational resource, all images are resized to 256×256 and no data augmentation strategy is used. The hyperparameters λ_1, λ_2, λ_3, θ, μ, and s in Eq. (15), Eq. (16), Eq. (13), and Eq. (9) are set to 0.15, 0.05, 0.1, 0.1, 0.01, and 1000, respectively. Algorithm 1 depicts the training procedure for our method.

2) Datasets: To comprehensively evaluate the performance of LIE methods, we use 485 training samples from the LOL dataset [43] as the training set and select 393 images from nine datasets as the test set. Specifically, supervised methods are trained on 485 low/normal-light image pairs, while unsupervised methods (including our method) use only the 485 low-light images from these pairs. The test set comprises four paired datasets, including LOL (15 images) [43], LSRW (50 images) [56], LOL-syn (100 images) [18], and LOL-real

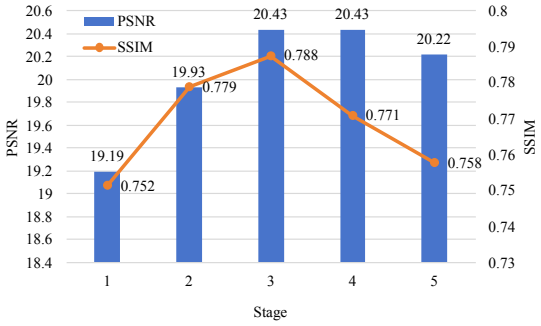


Fig. 5. Ablation study on stage numbers of the UF module.

(100 images) [18], along with five unpaired datasets, including LIME (10 images) [31], MEF (17 images) [57], NPE (8 images) [12], DICM (69 images) [58], and VV[1] (24 images). For clarity, we present the number of images within the test set of each dataset in parentheses in the above description.

3) Evaluation Metrics: To compare different LIE methods on paired datasets, we employ two full-reference metrics: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) [59]. To compare different LIE methods on unpaired datasets, we adopt four no-reference metrics: Natural Image Quality Evaluator (NIQE) [60], Perception Index (PI) [61], Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [62], and Lightness Order Error (LOE) [12]. Higher values of PSNR and SSIM, whereas lower values of NIQE, PI, BRISQUE, and LOE indicate superior performance.

4) Compared Methods: We select 14 state-of-the-art LIE methods for comparison, including three traditional methods (NPE [12], SRIE [63], LIME [31]), three supervised learning-based methods (RetinexNet [43], KinD [17], URetinexNet [19]), and eight unsupervised learning-based methods (ZeroDCE [47], RUAS [50], SCI [64], MLNet [48], ZeroIG [23], the method proposed by Li et al. [52], SPNet [24], and CLODE [65]). To avoid performance bias, these methods were trained and tested using their official source code, with the same data division settings as our method.

B. Ablation Studies

Before the main experiment, a series of ablation studies are carried out on the LOL dataset to evaluate the contribution of key configurations in our proposed method.

1) Number of Unfolding Stages: Our method uses unfolding modules to progressively refine decomposition results through multi-stage optimization. Here, we conduct experiments to investigate the impact of the number T of unfolding stages on performance. As shown in Fig. 5, the enhancement performance shows steady improvement when increasing T from 1 to 3, demonstrating the effectiveness of multi-stage unfolding for progressive decomposition optimization. Further unfolding beyond 3 stages causes performance degradation, likely due to optimization challenges in overly deep architectures. Thus, we set T to 3 in our final model.

[1]https://sites.google.com/site/vonikakis/datasets

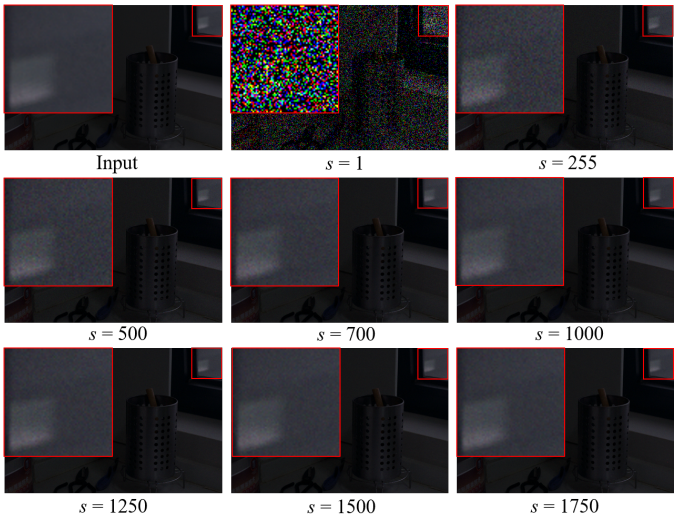


Fig. 6. Visual results after Poisson sampling at different scaling factors.

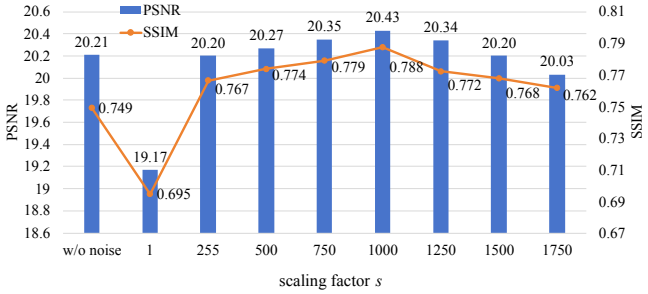


Fig. 7. Ablation study on added noise and scaling factor s of the self-supervised denoising mechanism.

2) Self-Supervised Denoising Mechanism: The proposed self-supervised denoising mechanism employs Eq. (9) to generate noisy images through controllable Poisson noise injection. The scaling factor s serves as a key parameter regulating noise intensity to simulate realistic noise distribution characteristics. Here, we investigate the effectiveness of the noise-adding approach and the impact of the scaling factor. Fig. 6 demonstrates the visual effects of Poisson sampling on the same image at different scaling factors. As shown in Fig. 7, when scaling is not performed (that is, s = 1), the performance decreases greatly. One possible reason for this is that the inherently low pixel values in low-light images lead to excessive noise intensity during direct Poisson sampling, causing severe distortion and substantially compromising training stability. In contrast, properly scaled noise injection improves SSIM compared to both unscaled and noise-free conditions, indicating the efficacy of the proposed self-supervised denoising mechanism. PSNR and SSIM achieve an optimal balance when $s = 1000$. Further increasing s diminishes performance, likely due to insufficient noise intensity hindering the network's ability to distinguish between added noise and image content. Thus, we select $s = 1000$ as our default configuration.

3) Analysis of Loss Contributions: In this study, we apply multiple loss functions during network training. Here, We analyze their contributions through ablation experiments. Table I shows the experimental results. By comparing the results in the first and last rows, it is clear that the removal of L_{con} decreases the performance of the network, indicating its

TABLE I
ABLATION RESULTS FOR DIFFERENT LOSS FUNCTION.

Configuration	PSNR ↑	SSIM ↑
w/o $\mathcal{L}_{con}$	20.14	0.761
w/o $\mathcal{L}_{iie}$	19.87	0.768
w/o $\mathcal{L}_{Rg}$	20.20	0.767
w/o $\mathcal{L}_{s}$	20.29	0.757
w/o $\mathcal{L}_{il}$	19.91	0.776
Ours	**20.43**	**0.788**

TABLE II
ABLATION RESULTS FOR DIFFERENT ATTENTION BLOCKS.

Attention Block	PSNR ↑	SSIM ↑	Time (s) ↓
w/o attention	19.22	0.761	**0.0544**
SE	19.86	0.776	0.0547
CBAM	19.98	0.773	0.0774
Ours	**20.43**	**0.788**	0.0577

positive role. Likewise, the absence of regularization L_{iie} and L_{il} for illumination adversely affects reflectance refinement, leading to PSNR reductions of 0.56dB and 0.52dB, respectively. This demonstrates the strong coupling relationship between illumination and reflectance, where proper illumination estimation guides the network to produce high-quality reflectance. As shown in the third and fourth rows, SSIM is decreased by 0.021 and 0.031 due to the lack of gradient regularization loss L_{Rg} for reflectance and smoothness loss L_s for illumination, respectively, indicating their positive effect on structural recovery. In summary, all loss functions guide the network in enhancing low-light images, and their complementary integration improves network performance.

4) Effectiveness of the Attention Block: Our ID module incorporates a simplified spatial-channel attention block for feature enhancement. Here, we conduct an ablation study to compare it with two mainstream attention blocks: SE [66] and CBAM [53]. As shown in Table II, the usage of attention blocks boosts performance. Compared to SE and CBAM, our proposed attention block brings more performance improvement. In addition, it has a comparable inference speed to SE and a faster inference speed than CBAM.

5) Impact of the Gamma Correction Factor: As formulated in Eq. (8), our method applies gamma correction to generate an auxiliary image for the low-light input. Here, we investigate the impact of the gamma correction factor γ on the enhancement performance. As shown in Fig. 8, the PSNR value increases as γ increases from 0.6 to 0.8. However, further increasing γ beyond 0.8 results in performance degradation. A possible reason for this is that, an excessively small γ violates the condition $R^{\gamma-1} \approx 1$, thereby compromising reflectance consistency. Conversely, an overly large γ fails to produce an auxiliary image with sufficient illumination difference, potentially resulting in a suboptimal reflectance estimate. Therefore, we select $\gamma = 0.8$ as our default configuration.

6) Impact of Auxiliary Image Generation Methods: In the literature, many techniques have been proposed for image enhancement. Here, we compare the effectiveness between gamma correction and three common image enhancement techniques (histogram equalization, Laplacian sharpening, and logarithmic transformation) in our LIE task. As shown in

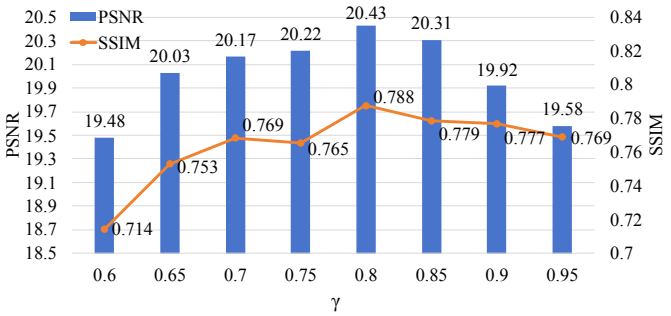


Fig. 8. Ablation study on gamma correction factor γ for generating auxiliary images.

TABLE III
QUANTITATIVE COMPARISON OF DIFFERENT AUXILIARY IMAGE GENERATION METHODS.

Method	PSNR ↑	SSIM ↑
Histogram Equalization	17.56	0.395
Laplacian Sharpening	19.46	0.760
Logarithmic Transformation	19.91	0.771
Gamma Correction (Ours)	**20.43**	**0.788**

Table III, gamma correction achieves the best performance, followed by logarithmic transformation, Laplacian sharpening, and histogram equalization. Possible reasons for this are as follows. Histogram equalization potentially disrupts the image's inherent reflectance properties by redistributing pixel intensities. Laplacian sharpening amplifies both edges and noise, when applied to low-light images with low signal-to-noise ratios. Logarithmic transformation $I_n = c \cdot \log(1+I_l)$, while also a nonlinear enhancement method, differs fundamentally from gamma correction in its mathematical form. Unlike the result of gamma correction in Eq. (8), where I_n shares the identical reflectance R with I_l, logarithmic transformation does not guarantee such a factorizable product structure with a shared R. Since the logarithmic transformation cannot strictly satisfy the shared reflectance prior, it forces the network to resolve the inherent discrepancy during optimization. This introduces an additional burden that limits its effectiveness within our framework. In contrast, gamma correction provides a clear physical guidance by ensuring the shared reflectance prior, which contributes to its superior performance and confirms its suitability as an auxiliary image generation method.

7) Analysis of Loss Weight Hyperparameters: In our study, the initial estimates for the hyperparameters in Eq. (13), Eq. (15), and Eq. (16) are empirically set by considering the relative magnitudes of the different loss terms to ensure stable optimization. Subsequently, we conduct a series of ablation studies to determine these values. Specifically, we begin with univariate ablation, varying only a single weight at a time while keeping all others fixed. The interaction effects between different weights are then examined through a joint ablation study, implemented as a local grid search in the vicinity of the baseline. Finally, the combination that achieves the best PSNR/SSIM values is selected. The ablation results are shown in Fig. 9 and Fig. 10. Generally, a more extensive search over the hyperparameter space might lead to further performance gains. However, given the substantial computational cost of such a search and the fact that our method already achieves

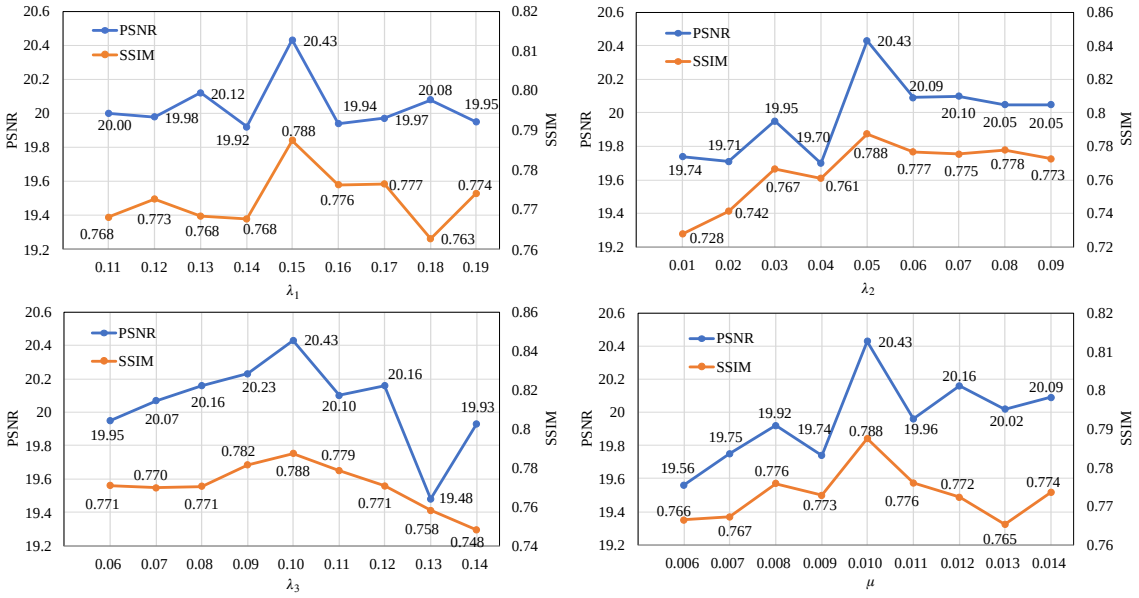


Fig. 9. Results of the univariate ablation study on loss weight hyperparameters.

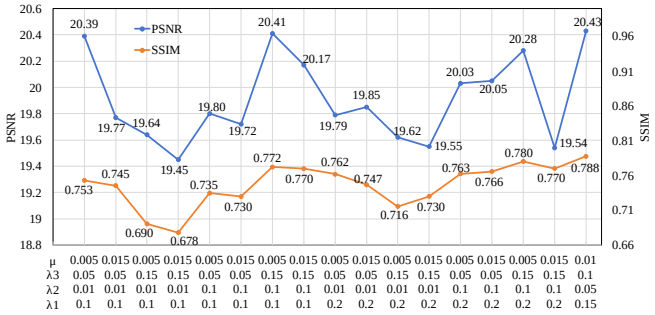


Fig. 10. Results of the joint ablation study on loss weight hyperparameters.

TABLE IV
ABLATION RESULTS FOR DIFFERENT NOISE TYPES.

Noise	PSNR ↑	SSIM ↑
Gaussian noise	20.22	0.771
Poisson noise (Ours)	**20.43**	**0.788**

competitive performance under current settings, we consider a more exhaustive hyperparameter optimization to a valuable direction for future work.

8) Analysis of Noise Types: Our self-supervised denoising mechanism constructs training pairs by introducing Poisson noise to the original input. To justify this design, we compare it with the additive white Gaussian noise (AWGN). As shown in Table IV, the model trained with Poisson noise achieves superior performance. This advantage stems from the closer alignment of Poisson noise with the physical imaging process under low-light conditions. In camera imaging, the stochastic arrival of photons causes the number of photons accumulated at each pixel to follow a Poisson distribution [67], where the noise variance is proportional to the signal intensity. As such, learning to remove Poisson noise compels the network to distinguish image content from noise. This design endows the network with strong generalization and adaptability when dealing with real-world noise.

C. Quantitative Evaluations

Table V tabulates the results of our method and compared methods on nine datasets. Specifically, among unsupervised methods, our method achieves an average improvement of 0.245dB in PSNR over the second-best method MLNet on two paired datasets (LOL and LOL-syn), as well as a 0.38dB advantage over the second-best method CLODE on LSRW (17.59dB vs. 17.21dB). On the LOL-real dataset, our method achieves a 0.52dB improvement over the second-best Li's method. In terms of the SSIM metric, our method also outperforms the compared method MLNet on three datasets: LOL (0.788 vs. 0.758), LSRW (0.528 vs. 0.519), and LOL-real (0.776 vs. 0.760). The performance on paired datasets demonstrates our method's capability to effectively ensure enhancement fidelity and restore image structures. On LIME, NPE, DICM, and VV, four unpaired datasets, our method also achieves superior performance over compared methods in no-reference metrics NIQE and PI. Specifically, it achieves an average improvement of 9.1% in NIQE compared to Li's method, and 13.7% in PI compared to SPNet, indicating superior capabilities in both preserving image naturalness and enhancing perceptual quality. In addition, our method maintains top-two performance across all three metrics (NIQE, PI and BRISQUE) on both the LIME and VV datasets, achieving a BRISQUE value of 21.57 on the VV dataset. This shows that our method has a lower degree of distortion in the enhancement results.

In addition to these traditional evaluation metrics, we also apply the evaluation metrics "Rank" and "RoR" to ensure more comprehensive comparisons. The "Rank" metric first ranks all methods in terms of each metric on all datasets, then averages all ranking values in all metrics to obtain an overall ranking. A value of "Rank" closer to 1 signifies better performance. "RoR" ranks the "Rank" results, converting them into a more intuitive sequential ranking, where 1 indicates the best performance. Among unsupervised methods, our

TABLE V
QUANTITATIVE RESULTS OF VARIOUS LIE METHODS ACROSS NINE BENCHMARK DATASETS. FOR UNSUPERVISED MODELS, THE BEST PERFORMANCES ARE MARKED IN RED. THE "RANK" METRIC IS COMPUTED BY AVERAGING THE RANKINGS OF EACH METHOD ACROSS ALL METRICS. THE "ROR" FURTHER TRANSFORMS THE "RANK" RESULTS INTO A MORE INTUITIVE SEQUENTIAL RANKING.

Dataset	Metric	Traditional			Supervised			Unsupervised								
		NPE	SRIE	LIME	RetinexNet	KinD	URetinexNet	ZeroDCE	RUAS	SCI	MLNet	ZeroIG	Li et al.	SPNet	CLODE	**Ours**
LOL	PSNR	16.97	11.86	16.76	16.77	17.65	19.95	16.26	16.34	16.02	20.23	19.36	19.82	20.12	19.16	20.43
	SSIM	0.472	0.493	0.434	0.417	0.773	0.805	0.516	0.498	0.534	0.758	0.750	0.744	0.739	0.620	0.788
LSRW	PSNR	16.19	13.36	15.61	15.49	16.41	17.09	14.28	14.27	15.77	16.88	16.74	14.90	16.64	17.21	17.59
	SSIM	0.390	0.422	0.365	0.354	0.484	0.539	0.403	0.465	0.437	0.519	0.516	0.490	0.511	0.479	0.528
LOL-syn	PSNR	16.60	14.50	16.88	17.14	17.28	17.38	14.66	13.44	15.76	18.46	17.24	17.34	17.00	18.37	18.75
	SSIM	0.778	0.663	0.758	0.758	0.758	0.787	0.773	0.644	0.764	0.806	0.744	0.790	0.778	0.851	0.792
LOL-real	PSNR	17.33	14.45	15.24	16.10	20.59	18.74	14.21	15.40	17.30	17.88	16.95	18.13	17.98	17.63	18.65
	SSIM	0.453	0.520	0.409	0.399	0.818	0.825	0.466	0.489	0.536	0.760	0.722	0.759	0.758	0.583	0.776
LIME	NIQE	3.90	3.79	4.15	4.60	4.76	4.79	4.48	5.36	4.19	4.62	4.94	4.32	4.34	4.01	3.94
	PI	2.87	2.76	3.08	3.08	3.72	3.51	3.21	3.91	3.06	3.62	4.49	3.45	3.42	3.03	3.03
	BRISQUE	18.09	16.93	20.32	26.34	25.21	26.34	26.26	29.58	22.19	16.57	36.10	18.37	16.12	14.50	19.08
	LOE	1471	824	1324	1882	830	601	451	1188	215	702	1088	1110	764	722	679
NPE	NIQE	3.95	3.99	4.27	4.59	4.17	4.59	4.64	7.08	4.61	3.85	4.96	4.12	4.55	3.97	3.62
	PI	2.92	2.94	3.27	3.13	3.11	3.24	3.52	5.22	3.56	3.39	4.45	3.28	3.52	3.16	2.90
	BRISQUE	14.60	17.00	18.40	22.58	18.09	26.12	32.93	49.38	34.51	20.81	37.32	28.77	24.83	21.08	18.23
	LOE	646	533	1120	1217	460	858	1077	2147	1058	1309	1292	905	1289	471	729
MEF	NIQE	3.52	3.48	3.70	4.42	3.87	4.31	3.65	5.38	3.65	4.28	5.06	3.84	4.13	3.64	3.88
	PI	2.51	2.61	2.92	2.87	3.03	3.48	2.59	4.05	2.63	3.66	4.80	3.42	3.34	2.80	3.11
	BRISQUE	15.74	17.92	17.95	20.07	27.51	25.42	18.50	34.10	15.77	23.52	41.82	23.91	20.04	12.07	23.85
	LOE	1158	754	1079	1777	690	621	406	1032	178	667	857	797	709	379	603
DICM	NIQE	3.76	3.90	3.86	4.43	4.14	4.44	4.00	7.14	4.00	4.13	4.75	3.90	4.02	3.68	3.68
	PI	3.12	3.34	3.58	3.24	3.58	3.55	3.59	5.74	3.65	3.76	4.53	3.51	3.73	3.31	3.17
	BRISQUE	22.98	24.59	24.78	29.60	28.83	24.94	36.30	47.38	31.90	28.62	39.77	31.93	32.72	23.06	24.55
	LOE	662	623	1261	1542	781	875	953	2497	1093	1362	1727	988	1456	711	861
VV	NIQE	2.52	2.85	2.48	2.70	3.03	3.52	3.48	5.34	3.58	3.90	4.87	3.70	3.80	3.51	3.33
	PI	3.00	3.24	2.85	2.95	3.48	3.79	3.50	4.98	3.66	4.15	4.83	3.99	4.00	3.71	3.57
	BRISQUE	23.25	24.61	27.88	27.37	23.44	26.43	35.15	51.53	30.67	31.80	46.47	34.04	32.99	29.71	21.57
	LOE	821	551	1275	1388	625	533	532	1447	508	748	902	759	884	420	688
Rank		5.6	6.5	8.5	9.9	6.8	6.9	9.0	13.9	8.1	7.6	11.8	8.1	8.6	4.4	3.9
RoR		3	4	10	13	5	6	12	15	8	7	14	8	11	2	1

TABLE VI
A COMPARISON OF MODEL PARAMETERS, FLOPS, AND INFERENCE TIME (GPU) BETWEEN LEARNING-BASED METHODS AND OUR METHOD. THE SOURCE OF "ROR" IS REFERRED TO TABLE V.

Method	Params (M) ↓	FLOPs (G) ↓	Time (s) ↓	RoR ↓
Retinex-Net	0.4446	235.64	0.1891	13
KinD	8.0160	127.72	0.2025	5
URetinexNet	0.3607	208.50	0.0283	6
ZeroDCE	0.0794	19.01	0.0033	12
RUAS	0.0034	0.78	0.0115	15
SCI	**0.0003**	**0.13**	**0.0019**	8
MLNet	0.4718	58.22	0.0426	7
ZeroIG	0.0866	118.73	0.0069	14
Li et al.	0.3453	74.71	0.0574	8
SPNet	1.2667	264.64	0.0307	11
CLODE	0.2865	1703.12	3.9033	2
Ours	0.1487	44.59	0.0577	**1**

method ("Rank" = 3.9) outperforms CLODE ("Rank" = 4.4) and MLNet ("Rank" = 7.6), which rank second and third, respectively. According to the "RoR" metric, our method is ranked first ("RoR" = 1) among all compared methods (including traditional, supervised, and unsupervised methods), fully demonstrating its comprehensive performance advantage.

We also analyze the model complexity of the proposed method and compare it with competing methods in terms of model parameters (Params), floating point operations (FLOPs), and inference time. All metrics are measured on a PC equipped with an Intel i7-12700K CPU and an NVIDIA GeForce RTX 3090 Ti GPU, using images of size 600×400×3 as input. As shown in Table VI, while our S^{2}UNet incurs higher computational complexity than three lightweight methods (ZeroDCE, RUAS, and SCI), it is at a comparable or better level compared to the state-of-the-art methods. Considering that an LIE model with higher-quality outputs is highly desired for practical applications, our main focus in the current study is to improve the performance. In the future, we will optimize the network structure of S^{2}UNet for a better trade-off between efficiency and enhancement quality, which is crucial for deployment on resource-constrained platforms.

D. Qualitative Evaluations

Figs. 11 and 12 present some representative enhancement results of twelve LIE methods on the LOL and LOLv2-real datasets, respectively. Here, we do not present the results of conventional methods, as we are more focused on comparing deep learning-based methods. These images with indoor scenes were captured using adjusted exposure times and camera parameters. As shown in the first image of Fig. 11, the enhancement results of KinD, URetinexNet, ZeroDCE, Li's method, and CLODE include noticeable color shifts,

Fig. 11. Visual comparison of state-of-the-art LIE methods on low-light images in LOL dataset. Please zoom in to view details.

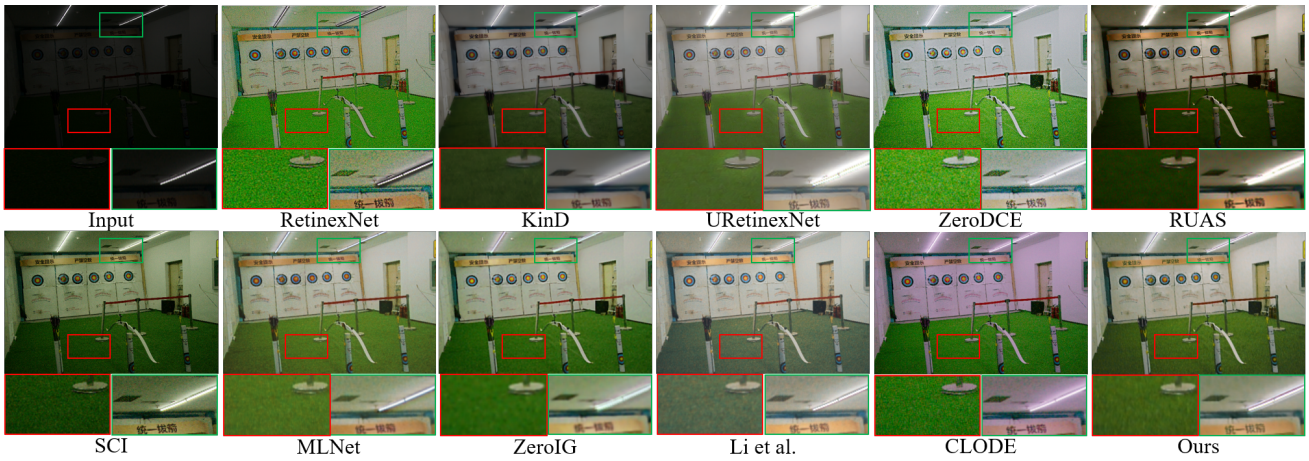


Fig. 12. Visual comparison of state-of-the-art LIE methods on low-light images in the LOLv2-real dataset. Please zoom in to view details.

while ZeroIG oversmooths wooden textures when enhancing an image with rich color and texture details. When processing the second image, RetinexNet amplifies noise during enhancement, SCI produces an underexposed appearance, and RUAS shows unnatural local contrast. In contrast, our method achieves brightness recovery while maintaining a more natural color rendition and contrast balance. The results in Fig. 12 also demonstrate that the compared methods usually yield unsatisfactory enhancement effects. For instance, KinD, RUAS, and SCI produce underexposed effects, RetinexNet, ZeroDCE, and CLODE fail to suppress noise, and both URetinexNet and ZeroIG generate blurred outputs. In contrast, our method effectively enhances the overall brightness of the image while smoothing noise without overly blurring details.

Figs. 13, 14, and 15 present the enhancement results of twelve LIE methods when processing real-world low-light images captured outdoors. As shown in Fig. 13, when enhancing a challenging low-light image from the VV dataset, URetinexNet, ZeroDCE, RUAS, Li's method, SPNet, and CLODE exhibit varying degrees of overexposure in important objects, e.g., the man in the foreground and the book in the lower right corner. Although KinD achieves proper exposure for the foreground, it fails to reveal the background details of the trees and leaves within the red box. ZeroIG produces unnatural artifacts in the book region within the green box and distorts the skin tone of the man. RetinexNet and MLNet demonstrate poor noise suppression in dark regions. In contrast, our method successfully reveals details in dark areas while performing denoising and avoids over-enhancing well-exposed regions. On the DICM dataset, which includes natural scene images with non-uniform illumination conditions, the compared methods produce poor enhancement

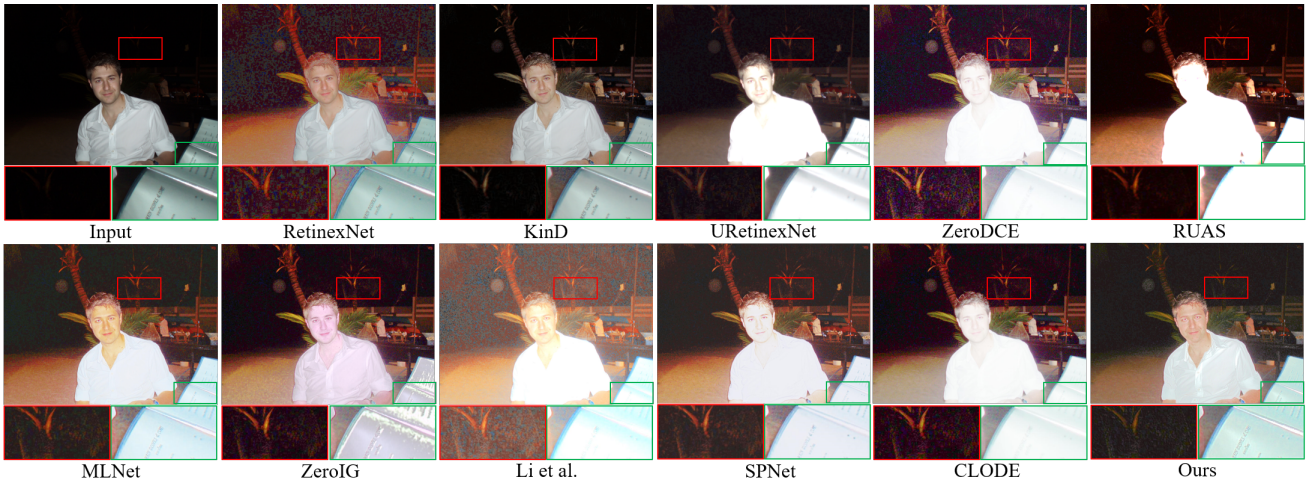


Fig. 13. Visual comparison of state-of-the-art LIE methods on a challenging example in the VV dataset. Please zoom in to view details.

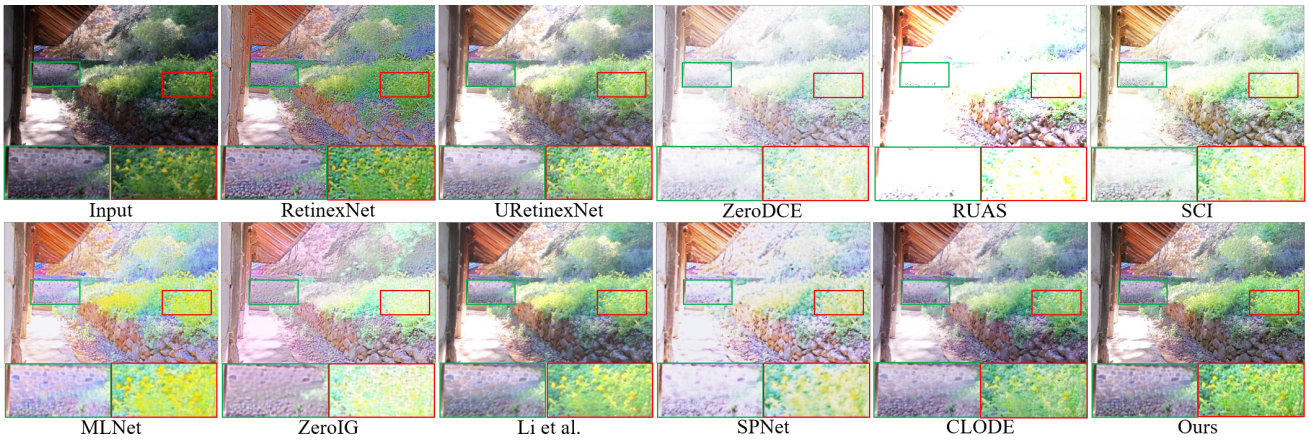


Fig. 14. Visual comparison of state-of-the-art LIE methods on low-light images in the DICM dataset. Please zoom in to view details.

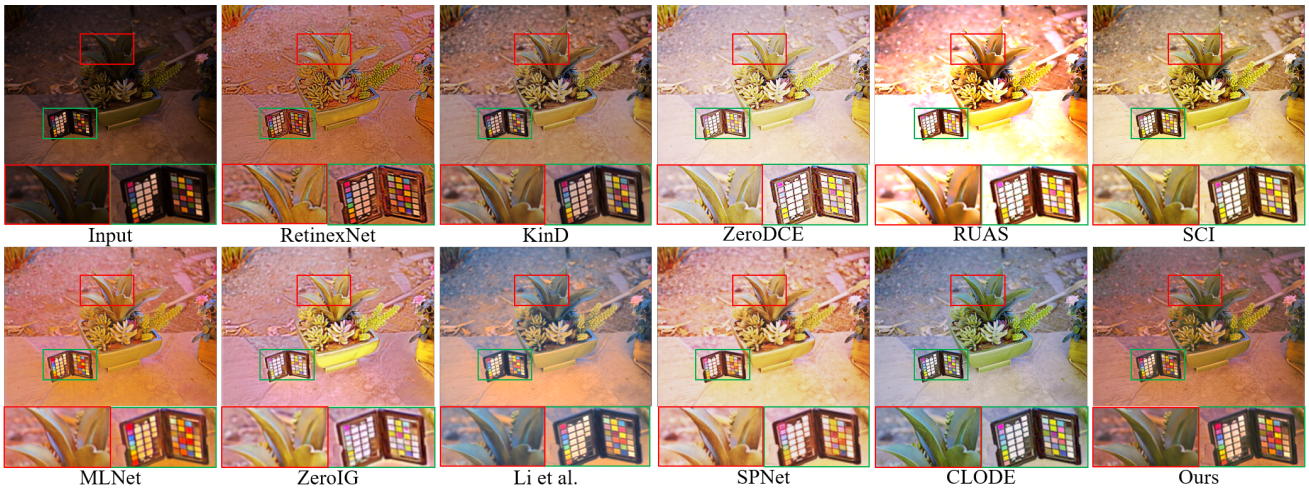


Fig. 15. Visual comparison of state-of-the-art LIE methods on low-light images in the LIME dataset. Please zoom in to view details.

results, as shown in Fig. 14. For example, Retinex results in unnatural colors, MLNet presents unnatural saturation in the flower areas, Li's method and CLODE blur texture details, and URetinexNet, ZeroDCE, RUAS, SCI, ZeroIG, and SPNet lead to overexposure. In contrast, our method can enhance locally underexposed areas while maintaining the natural colors and textures of the flowers and plants. As shown in Fig. 15, our method also demonstrates advantages in color fidelity and visual naturalness compared to other methods on the LIME dataset. On the contrary, RetinexNet exhibits unnatural artifacts at the edges, and ZeroIG loses edge details due to over-smoothing. RUAS again exhibits overexposure issues, Li's method and CLODE bring an unrealistic blue color bias, while the other compared methods generally result in low color saturation in plant regions, losing natural appearance.

To better understand our method, we present the visu-

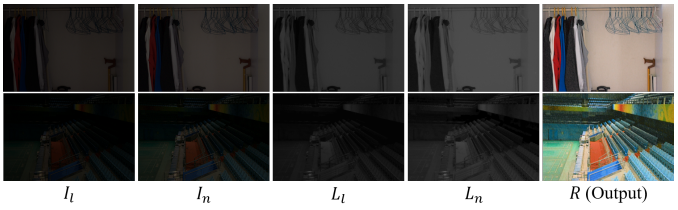


Fig. 16. Visualization results of the decomposed components.

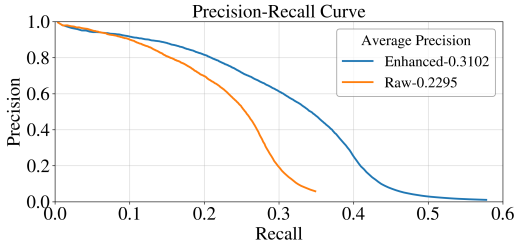


Fig. 17. The Precision-Recall curves on the DARK FACE dataset, where Raw denotes detection results from original images and Enhanced represents results from images enhanced by our method.

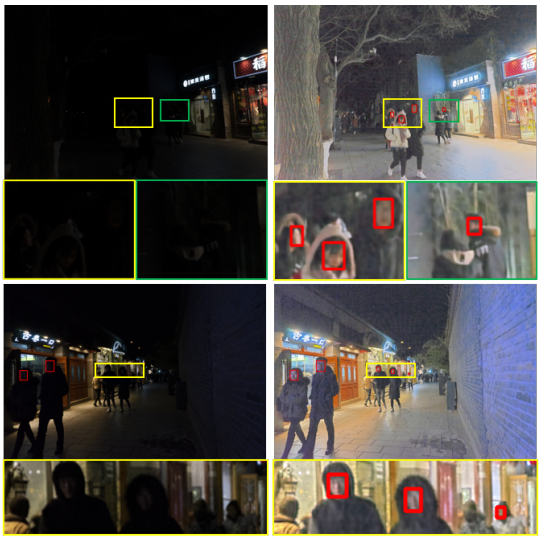


Fig. 18. Visual comparison of face detection results in raw low-light images and its enhanced versions.

alizations of the decomposition results in Fig. 16, which include the input image I_l, the auxiliary image I_n, and their corresponding illumination components L_l, L_n, along with the shared reflectance component R. The auxiliary image is generated by applying gamma correction to the input image, yielding a different brightness level. The illumination components capture global lighting variations, and their differences reflect the distinct illumination conditions between I_l and I_n.

E. Applications in Low-Light Face Detection

In this section, we investigate the impact of S^{2}UNet as a preprocessing step to improve the performance of a downstream task, i.e., low-light face detection. Specifically, we conduct experiments on the DARK FACE dataset [68], which contains 6,000 real-world nighttime low-light images with annotated face bounding boxes. The enhanced images are fed into the pre-trained DSFD face detector [69], and evaluated using the DARK FACE evaluation tool[2] at an IoU threshold of 0.5 to plot the precision-recall (P-R) curves and compute the average precision (AP). As shown in Fig. 17, compared to raw low-light images, the AP of DSFD improves from 0.2295 to 0.3102, demonstrating the positive role of our method in increasing the performance of face detection in low-light conditions. To illustrate this more intuitively, we present an example in Fig. 18, where the small targets in low-light regions are accurately detected after enhancing the low-light image with our S^{2}UNet.

V. Conclusion

In this paper, we propose a novel self-supervised unfolding network, named S^{2}UNet, for low-light image enhancement. The key motivation of our S^{2}UNet lies in modeling the mapping from the low-light to the high-light domain without strictly paired images through shared reflectance learning. Compared to existing works, our S^{2}UNet has the following advantages. First, we construct a new optimization model that enforces the decomposed reflectance from illumination-different images to be the same, effectively encoding the physical prior of Retinex theory into the network. Second, we generate an illumination-different auxiliary image for the low-light image and train the network in a self-supervised manner, thereby eliminating dependence on strictly paired images. Third, we propose a new denoising mechanism that does not require a high-quality image as a reference, making it more suitable for practical applications. Extensive experiments on nine benchmark datasets demonstrate that our method outperforms three conventional methods and eight state-of-the-art unsupervised methods, while achieving competitive performance compared to three supervised methods. In addition, our S^{2}UNet also increases the performance of face detection in low light conditions, a typical downstream task.

References

[1] W. Ren, S. Liu, L. Ma, Q. Xu, X. Xu, X. Cao, J. Du, and M.-H. Yang, "Low-light image enhancement via a deep hybrid network," *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4364–4375, 2019.

[2] H. Zhou, W. Dong, X. Liu, Y. Zhang, G. Zhai, and J. Chen, "Low-light image enhancement via generative perceptual priors," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 10, 2025, pp. 10 752–10 760.

[3] Q. Jiang, Y. Kang, Z. Wang, W. Ren, and C. Li, "Perception-driven deep underwater image enhancement without paired supervision," *IEEE Transactions on Multimedia*, vol. 26, pp. 4884–4897, 2024.

[4] G. Yue, J. Gao, R. Cong, T. Zhou, L. Li, and T. Wang, "Deep pyramid network for low-light endoscopic image enhancement," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 5, pp. 3834–3845, 2024.

[5] Z. Zeng, Z. Wang, Z. Wang, Y. Zheng, Y.-Y. Chuang, and S. Satoh, "Illumination-adaptive person re-identification," *IEEE Transactions on Multimedia*, vol. 22, no. 12, pp. 3064–3074, 2020.

[6] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3291–3300.

[7] S. Pizer, R. Johnston, J. Ericksen, B. Yankaskas, and K. Muller, "Contrast-limited adaptive histogram equalization: speed and effectiveness," in *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*, 1990, pp. 337–345.

[8] H. Farid, "Blind inverse gamma correction," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1428–1433, 2001.

[2]https://github.com/Ir1d/DARKFACE_eval_tools

[9] M. K. Ng and W. Wang, "A total variation model for retinex," *SIAM Journal on Imaging Sciences*, vol. 4, no. 1, pp. 345–365, 2011.

[10] E. H. Land and J. J. McCann, "Lightness and retinex theory," *Journal of the Optical society of America*, vol. 61, no. 1, pp. 1–11, 1971.

[11] D. Jobson, Z. Rahman, and G. Woodell, "Properties and performance of a center/surround retinex," *IEEE Transactions on Image Processing*, vol. 6, no. 3, pp. 451–462, 1997.

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

[13] X. Ren, W. Yang, W.-H. Cheng, and J. Liu, "Lr3m: Robust low-light enhancement via low-rank regularized retinex model," *IEEE Transactions on Image Processing*, vol. 29, pp. 5862–5876, 2020.

[14] L. Wang, L. Xiao, H. Liu, and Z. Wei, "Variational bayesian method for retinex," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3381–3396, 2014.

[15] C. Li, J. Guo, F. Porikli, and Y. Pang, "Lightennet: A convolutional neural network for weakly illuminated image enhancement," *Pattern Recognition Letters*, vol. 104, pp. 15–22, 2018.

[16] Y. Zhang, X. Di, B. Zhang, and C. Wang, "Self-supervised image enhancement network: Training with low light images only," *arXiv preprint arXiv:2002.11300*, 2020.

[17] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 1632–1640.

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

[19] W. Wu, J. Weng, P. Zhang, X. Wang, W. Yang, and J. Jiang, "Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 5891–5900.

[20] X. Liu, Q. Xie, Q. Zhao, H. Wang, and D. Meng, "Low-light image enhancement by retinex-based algorithm unrolling and adjustment," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 11, pp. 15 758–15 771, 2024.

[21] H. Wang, X. Hou, J. Li, Y. Yan, W. Sun, X. Zeng, K. Zhang, and X. Cao, "Multi-scale retinex unfolding network for low-light image enhancement," *IEEE Transactions on Multimedia*, pp. 1–13, 2025.

[22] N. Zheng, M. Zhou, Y. Dong, X. Rui, J. Huang, C. Li, and F. Zhao, "Empowering low-light image enhancer through customized learnable priors," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 12 525–12 535.

[23] 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 *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 3015–3024.

[24] Y. Luo, X. Chen, J. Ling, C. Huang, W. Zhou, and G. Yue, "Unsupervised low-light image enhancement with self-paced learning," *IEEE Transactions on Multimedia*, vol. 27, pp. 1808–1820, 2025.

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

[26] H. Ibrahim and N. S. P. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 4, pp. 1752–1758, 2007.

[27] S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," *IEEE Transactions on Image Processing*, vol. 22, no. 3, pp. 1032–1041, 2013.

[28] W. Wang, N. Sun, and M. K. Ng, "A variational gamma correction model for image contrast enhancement," *Inverse Problems and Imaging*, vol. 13, no. 3, pp. 461–478, 2019.

[29] D. Jobson, Z. Rahman, and G. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing*, vol. 6, no. 7, pp. 965–976, 1997.

[30] Z. Rahman, D. Jobson, and G. Woodell, "Multi-scale retinex for color image enhancement," in *Proceedings of 3rd IEEE International Conference on Image Processing*, vol. 3, 1996, pp. 1003–1006 vol.3.

[31] X. Guo, Y. Li, and H. Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 982–993, 2017.

[32] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2782–2790.

[33] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust retinex model," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2828–2841, 2018.

[34] G. Yue, S. Wu, R. Tian, H. Lin, J. Li, T. Yuan, H. Lv, Z. Yu, N. Mao, and X. Song, "Benchmarking laryngeal neoplasm segmentation: A multicenter dataset and an effective method," *IEEE Transactions on Image Processing*, vol. 34, pp. 7362–7377, 2025.

[35] Y. Liu, Z. Qi, J. Cheng, and X. Chen, "Rethinking the effectiveness of objective evaluation metrics in multi-focus image fusion: A statistic-based approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 8, pp. 5806–5819, 2024.

[36] Z. Wan, X. Yan, Z. Li, X. Fan, W. Zuo, and D. Zhao, "No-reference stereoscopic omnidirectional image quality assessment via a binocular viewport hypergraph convolutional network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 35, no. 7, pp. 7196–7209, 2025.

[37] G. Yue, L. Zhang, J. Du, T. Zhou, W. Zhou, and W. Lin, "Subjective and objective quality assessment of colonoscopy videos," *IEEE Transactions on Medical Imaging*, vol. 44, no. 2, pp. 841–854, 2025.

[38] Z. Wan, H. Qin, R. Xiong, Z. Li, X. Fan, and D. Zhao, "Predicting 360 video saliency: a convlstm encoder-decoder network with spatio-temporal consistency," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 14, no. 2, pp. 311–322, 2024.

[39] G. Yue, W. Li, C. Zhao, Z. Wu, T. Zhou, Q. Jiang, and R. Cong, "Text-guided semantic alignment network with spatial-frequency interaction for infrared-visible image fusion under extreme illumination," *IEEE Transactions on Image Processing*, vol. 34, pp. 7943–7958, 2025.

[40] L. Zheng, Y. Luo, Z. Zhou, J. Ling, and G. Yue, "Cdinet: Content distortion interaction network for blind image quality assessment," *IEEE Transactions on Multimedia*, vol. 26, pp. 7089–7100, 2024.

[41] C. Li, C. Guo, L. Han, J. Jiang, M.-M. Cheng, J. Gu, and C. C. Loy, "Low-light image and video enhancement using deep learning: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 12, pp. 9396–9416, 2021.

[42] G. Yue, L. Zhang, W. Liu, J. Du, T. Zhou, H. Lin, Q. Jiang, and W. Ren, "Sgnet: Style-guided network with temporal compensation for unpaired low-light colonoscopy video enhancement," *IEEE Transactions on Image Processing*, accepted, in press, DOI: 10.1109/TIP.2025.3644172, 2025.

[43] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," *arXiv preprint arXiv:1808.04560*, 2018.

[44] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Under-exposed photo enhancement using deep illumination estimation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 6842–6850.

[45] 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 *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 12 106–12 115.

[46] P. Kandula, M. Suin, and A. N. Rajagopalan, "Illumination-adaptive unpaired low-light enhancement," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 8, pp. 3726–3736, 2023.

[47] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 1777–1786.

[48] Y. Luo, B. You, G. Yue, and J. Ling, "Pseudo-supervised low-light image enhancement with mutual learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 1, pp. 85–96, 2024.

[49] K. Wu, J. Huang, Y. Ma, F. Fan, and J. Ma, "Cycle-retinex: Unpaired low-light image enhancement via retinex-inline cyclegan," *IEEE Transactions on Multimedia*, vol. 26, pp. 1213–1228, 2024.

[50] 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 *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 10 556–10 565.

[51] N. Parikh, S. Boyd *et al.*, "Proximal algorithms," *Foundations and trends® in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

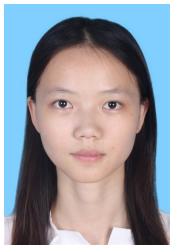
[52] L. Huaqiu, H. Wang *et al.*, "Interpretable unsupervised joint denoising and enhancement for real-world low-light scenarios," in *The Thirteenth International Conference on Learning Representations*, 2025.

[53] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.

[54] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2noise: Learning image restoration without clean data," *arXiv preprint arXiv:1803.04189*, 2018.

[55] Q. Yang, C. Jung, Q. Fu, and H. Song, "Low light image denoising based on poisson noise model and weighted tv regularization," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 3199–3203.
[56] J. Hai, Z. Xuan, R. Yang, Y. Hao, F. Zou, F. Lin, and S. Han, "R2rnet: Low-light image enhancement via real-low to real-normal network," *Journal of Visual Communication and Image Representation*, vol. 90, p. 103712, 2023.
[57] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3345–3356, 2015.
[58] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation of 2d histograms," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5372–5384, 2013.
[59] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
[60] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.
[61] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6228–6237.
[62] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.
[63] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2782–2790.
[64] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward fast, flexible, and robust low-light image enhancement," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 5627–5636.
[65] D. Jung, D. Kim, and T. H. Kim, "Continuous exposure learning for low-light image enhancement using neural odes," in *The Thirteenth International Conference on Learning Representations*, 2025.
[66] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.
[67] X.-Y. Kong, L. Liu, and Y.-S. Qian, "Low-light image enhancement via poisson noise aware retinex model," *IEEE Signal Processing Letters*, vol. 28, pp. 1540–1544, 2021.
[68] W. Yang, Y. Yuan, W. Ren, J. Liu, W. J. Scheirer, Z. Wang, and et al., "Advancing image understanding in poor visibility environments: A collective benchmark study," *IEEE Transactions on Image Processing*, vol. 29, pp. 5737–5752, 2020.
[69] J. Li, Y. Wang, C. Wang, Y. Tai, J. Qian, J. Yang, C. Wang, J. Li, and F. Huang, "Dsfd: Dual shot face detector," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 5055–5064.

Jia Liu received the B.S. degree in software engineering from Guangdong University of Technology in 2024. He is currently pursuing an M.S. degree at the Guangdong University of Technology. His research interests primarily lie in the field of computer vision and image processing.

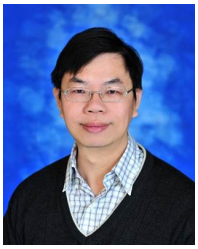


Yu Luo (Member, IEEE) received her Ph.D. degree in School of Computer Science and Technology from South China University of Technology, China, in 2016. She is currently an associate professor at Guangdong University of Technology, China. Her research interests include image recovery, medical imaging and deep learning.

Guanghui Yue (Member, IEEE) received the B.S. degree in communication engineering and the Ph.D. degree in information and communication engineering from Tianjin University, Tianjin, China, in 2014 and 2019, respectively. He was a joint Ph.D. student with the School of Computer Science and Engineering, Nanyang Technological University, Singapore, from September 2017 to January 2019. He is currently an Associate Professor with the School of Biomedical Engineering, Shenzhen University Medical School, Shenzhen University. His research interests include medical image analysis, bioelectrical signal processing, image quality assessment, 3D image visual discomfort prediction, pattern recognition, and machine learning.

Jie Ling received the Ph.D. degree in computer science from Sun Yat-sen University, China, in 1998. He is currently a Professor with the School of Computer Science, Guangdong University of Technology. His main research interests include computer applications, intelligent video processing technology.

Liang Liao (Senior Member, IEEE) received the B.S. degree from the International School of Software, Wuhan University, Wuhan, China, in 2013, and the Ph.D. degree from the National Engineering Research Center for Multimedia Software, School of Computer Science, Wuhan University, in 2019. He was a Research Fellow with the School of Computer Science and Engineering, Nanyang Technological University, Singapore, from 2022 to 2024, and a Project Researcher with the National Institute of Informatics, Japan, from 2019 to 2022. He is currently a Professor with Hangzhou Institute of Technology, Xidian University. His research interests include image/video processing and quality assessment.



Chia-Wen Lin (Fellow, IEEE) received the Ph.D. degree in electrical engineering from National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2000. He was with the Department of Computer Science and Information Engineering, National Chung Cheng University, Chiayi, Taiwan, from 2000 to 2007. He is currently a Distinguished Professor with the Department of Electrical Engineering and the Institute of Communications Engineering, NTHU. He is also the Deputy Director of the AI Research Center, NTHU. His research interests include image and video processing, computer vision, and video networking. Dr. Lin is currently serving as an Associate Editor-in-Chief of IEEE Transactions on Circuits and Systems for Video Technology.

Guangtao Zhai (Fellow, IEEE) received the B.E. and M.E. degrees from Shandong University, Shandong, China, in 2001 and 2004, respectively, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2009. He is currently a Research Professor with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University. His research interests include multimedia signal processing and perceptual signal processing.

Wei Zhou (Senior Member, IEEE) received the Ph.D. degree jointly from the University of Science and Technology of China, Hefei, China, in 2021, and University of Waterloo, Waterloo, ON, Canada. Wei Zhou was a Visiting Scholar with the National Institute of Informatics, Chiyoda City, Japan, Research Assistant with Intel, Research Intern with Microsoft Research and Alibaba Cloud, and Postdoctoral Fellow with the University of Waterloo. Wei Zhou is currently an Assistant Professor with Cardiff University, Cardiff, U.K. Wei Zhou's research interests include multimedia computing, perceptual image processing, and computational vision. Wei Zhou is also an Associate Editor for IEEE Transactions on Neural Networks and Learning Systems, Pattern Recognition, and Neurocomputing.