Unsupervised Low-Light Image Enhancement with Self-Paced Learning

Yu Luo, *Member, IEEE*, Xuanrong Chen, Jie Ling, Chao Huang, Wei Zhou, *Senior Member, IEEE*, and Guanghui Yue, *Member, IEEE*

***Abstract*—Low-light image enhancement (LIE) aims to restore images taken under poor lighting conditions, thereby extracting more information and details to robustly support subsequent visual tasks. While past deep learning (DL)-based techniques have achieved certain restoration effects, these existing methods treat all samples equally, ignoring the fact that difficult samples may be detrimental to the network's convergence at the initial training stages of network training. In this paper, we introduce a self-paced learning (SPL)-based LIE method named SPNet, which consists of three key components: the feature extraction module (FEM), the low-light image decomposition module (LIDM), and a pre-trained denoise module. Specifically, for a given low-light image, we first input the image, its pseudo-reference image, and its histogram-equalized version into the FEM to obtain preliminary features. Second, to avoid ambiguities during the early stages of training, these features are then adaptively fused via an SPL strategy and processed for retinex decomposition via LIDM. Third, we enhance the network performance by constraining the gradient prior relationship between the illumination components of the images. Finally, a pre-trained denoise module reduces noise inherent in LIE. Extensive experiments on nine public datasets reveal that the proposed SPNet outperforms eight state-of-the-art DL-based methods in both qualitative and quantitative evaluations and outperforms three conventional methods in quantitative assessments.**

***Index Terms*—Low-light image enhancement, self-paced learning, histogram equalization, Retinex decomposition.**

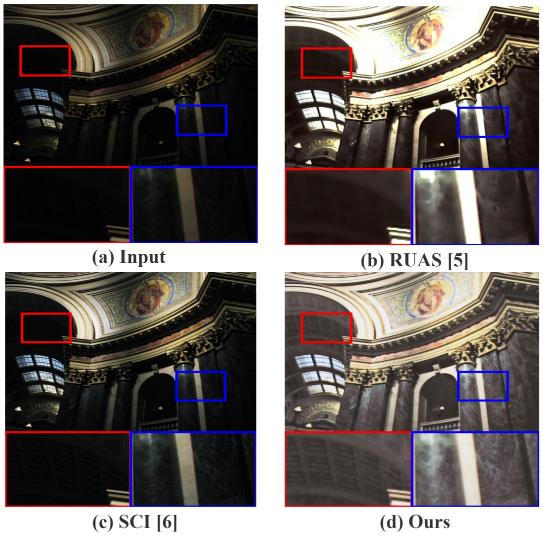


Fig. 1. Samples of naturally low-light images, improved with state-of-the-art unsupervised learning techniques and our SPNet, are presented. Our method can recover more content and details in images, such as the framework of skylights and the patterns on marble surfaces, as highlighted by the color rectangles.

I. Introduction

IMAGES captured under low-light conditions often suffer from a variety of quality issues, such as a significant increase in noise levels, a notable decrease in contrast, and the loss of detailed information. These defects directly impact the usability of images and pose serious challenges to subsequent computer vision tasks, such as object recognition [1]–[3] and image classification [4]. In light of this, the development of effective LIE techniques has become particularly important. These techniques not only improve image aesthetics but also significantly increase the effectiveness of subsequent image analysis tasks.

Histogram equalization is a classic technique for enhancing low-light images by modifying the image's overall brightness contrast. However, as a global processing strategy, it fails to consider the local features of the image, which may lead to an imbalance in local contrast and neglect of important spatial features. Recently, the rise of DL has brought new possibilities to LIE [5]–[9]. However, existing DL-based supervised methods often rely on a certain assumption, i.e., each low-light image has a corresponding version captured under normal lighting, which is often unrealistic in practical applications.

In recent research, various self-supervised and unsupervised strategies for LIE methods have been reported [10], [11]. Nguyen et al. [12] construct pseudo-ground-truth images synthesized from multiple source images that simulate all potential exposure scenarios to train the enhancement network. Wang et al. [13] proposed an adaptive enhancement framework

This work was supported in part by National Natural Science Foundation of China (Nos. 62371305), in part by Guangdong Basic and Applied Basic Research Foundation (No. 2024A1515030025), in part by Natural Science Foundation of Shenzhen (No. JCYJ20230808105906013), in part by the Medicine Plus Program of Shenzhen University (No. 2024YG007), in part by Tencent "Rhinoceros Birds" - Scientific Research Foundation for Young Teachers of Shenzhen University, and in part by International Scientific and Technological Cooperation Project of Huangpu and Development District in Guangzhou (No. 2022GH01). (Corresponding author: Guanghui Yue)

Y. Luo, X. Chen, and J. Ling are with the School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, 510006, China (e-mail: yuluo@gdut.edu.cn; cxr171015@163.com; jling@gdut.edu.cn).

Chao Huang is with the Department of Computer Science, University of Hong Kong, Hong Kong (e-mail: chaohuang75@gmail.com).

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

G. Yue is with the National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Shenzhen University Medical School, Shenzhen 518054, China, and also with the Marshall Laboratory of Biomedical Engineering, Shenzhen University, Shenzhen 518060, China (email: yueguanghui@szu.edu.cn).

for a single low-light image that is based on the strategy of virtual exposure. These methods primarily depend on carefully constructed prior knowledge to guide the training of neural networks, improving image quality without explicit pairwise supervision. Although these works have made some progress, they still face challenges when dealing with individual low-light images, as the limited information contained in images is insufficient to guide the network to learn all necessary features.

More recently, the integration of histogram equalization methods with convolutional neural network (CNN)-based techniques, particularly as a guiding approach, has demonstrated significant effectiveness in the realm of self-supervised and unsupervised image enhancement [4], [14], [15]. Generally speaking, these methods typically function by maintaining consistency between the features of the enhanced images and those extracted via histogram equalization.

While methods based on histogram equalization and CNN have had some effect, two problems remain. First, these methods treat all samples equally, neglecting the negative impact of difficult samples in the early stages of network training, which can lead the network toward suboptimal or erroneous optimizations. Second, in self-supervised or unsupervised scenarios, the lack of normal-light images to constrain Retinex decomposition may result in unsatisfactory outcomes from the enhancement network. Fig. 1 shows examples of restoration results on real low-light images that compare our proposed method with other state-of-the-art unsupervised methods. Our method effectively recovers the skylight textures within the red rectangle and the marble textures within the blue rectangle in the image.

In this paper, we propose a novel unsupervised LIE method named SPNet with an SPL strategy. The main differences between SPNet and the previous LIE method are shown in Fig. 2. Specifically, our network includes three main components: the FEM, the LIDM, and the pre-trained denoise module. FEM extracts simple features from the original low-light image, its pseudo-reference image, and its histogram-equalized versions. We subsequently fuse the features extracted via FEM via an SPL strategy and input the resulting feature into the LIDM for retinex decomposition. By using the SPL strategy, our method can reduce the negative impact of difficult samples to prevent the network from converging toward suboptimal or erroneous solutions during the initial stages of network training and guide the network's training progression from simple to complex tasks. We further enhance network training by constraining the illumination gradient priors between the original and pseudo-reference images. Finally, to eliminate noise during the enhancement process, we integrate a pre-trained denoise module, thus improving the quality of the final images.

The difference between our proposed method and existing methods [16]–[19] is: First, previous methods treat all samples equally during training, which may lead to overly difficult samples pushing the network to a suboptimal solution or even optimizing the network to a suboptimal solution in the early stages of network training. Our method introduces SPL into the LIE task to avoid the negative impact of difficult samples in the early stages of network training, thereby improving the quality of images restored by the network. Second, in the absence of reference images, we further improve the performance of our method by constraining the illumination gradient relationship between the pseudo-reference image and the original image. In summary, the main contributions of this study are as follows:

To prevent the network from optimizing toward suboptimal solutions or erroneous directions in the early training stages, we use an SPL strategy that adaptively adjusts guidance feature weights on the basis of sample difficulty determined through the results of histogram equalization.

To improve the performance of the LIE recovery without reference images, we introduce a novel constraint on the relationship between the gradients of the illumination components derived from the original input image and its pseudo-reference image.

The experimental results on nine datasets show that the proposed SPNet outperforms six state-of-the-art unsupervised LIE methods, three conventional LIE methods, and two supervised LIE methods in terms of performance. The source code of our SPNet will be available at https://github.com/X-Chen-DL/SPNet.

II. Related work

A. Conventional LIE Methods

In the literature of LIE, histogram equalization and its families are popular. Traditional histogram equalization often results in over-enhancement and detail loss [20]. Guo et al. [20] proposed an innovative method to effectively reduce noise amplification. Huang et al. [21] proposed a method to mitigate the problem of excessive enhancement and smoothing. Brightness preserving dynamic histogram equalization [22] extends dynamic histogram equalization techniques, generating output images with average brightness closely matching the input, thus preserving the image's overall luminance. Optimizing histogram equalization parameters enhances contrast and maintains image naturalness [23]. Additionally, detail-weighted histogram equalization has been developed to resolve over-enhancement in peak histograms [24].

Another legends of LIE are Retinex-based methods. The foundational theory of Retinex, proposed by Land and McCann, emphasizes the significance of simulating human visual perception in enhancing images [25]. Building upon this, various researchers have explored different adaptations of Retinex theory to enhance the visibility of images. Wang et al. [26] introduced a LIE technique by improving the realism and detail in photos affected by uneven lighting. Furthermore, Wang et al. [27] developed a technique for adjusting the colors of images through the application of the Retinex model and multi-scale analysis, employing a transformation that is nonlinear in nature as dictated by the Weber-Fechner law. Fu et al. [28] proposed a new probabilistic image enhancement method based on simultaneous estimation of illumination and reflectance in the linear domain. Apart from the aforementioned methods, one notable method is the Single-Scale Retinex method, which improves image brightness and preserves natural color balance simultaneously [29]. However, SSR often leads to over-enhanced results, a limitation that

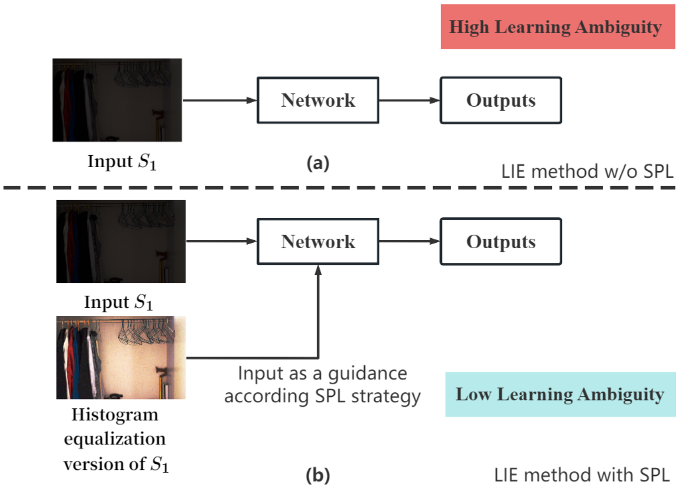


Fig. 2. A comparison between previous LIE methods (a) and our proposed SPNet (b). The key idea of our method is to utilize an SPL strategy, employing histogram equalization results as guidance to reduce the negative impact of difficult samples in the early stages and to diminish learning ambiguities during the initial phase of network training.

Multi-Scale Retinex seeks to address by integrating different scales of Retinex processing [30].

In recent years, the integration of Retinex theory with other computational techniques like optimization algorithms and machine learning has also been explored [31]–[35], indicating a trend towards more sophisticated and adaptive enhancement methods. Nonetheless, in the processing of complex images from real-world scenarios, these techniques usually result in local color distortions [36].

B. Learning-based Methods

In the past decade, a wide range of DL-based LIE algorithms have been proposed, which can be broadly categorized into supervised and unsupervised methods.

Commonly, the supervised methods require paired images, i.e., both low-light image and its content-consistent normal-light image. For example, Wei et al. [16] used deep learning for Retinex decomposition to enhance low-light images by improving brightness and contrast while preserving natural colors and details. Later, Zhang et al. [17] divided the image into reflection and illumination parts and restored them separately while removing noise and chromatic aberration. This method improved the image quality and adapted to different lighting conditions through a sensitivity adjustment network. Wu et al. [18] not only decomposed images into reflection and illumination but also meticulously developed three learning-based modules for data-related initialization, efficient optimization unfolding, and user-specified illumination enhancement. Xu et al. [19] enhance low-light images dynamically through spatially variant operations by integrating signal-to-noise-ratio-aware transformers and Convolution models. Guo et al. [37] enhanced the preliminary illumination representation by incorporating a predetermined framework, resulting in a sophisticated illumination profile facilitating related improvements. Wang et al. [38] proposed an invertible network that learns to map the distribution of normally exposed images into a Gaussian distribution. Wan et al. [39] proposed to enhance the visibility and suppress artifacts by purifying low-light images under the guidance of the NIR enlightened image captured by using the near-infrared light as compensation. Guo et al. [40] proposed a Cross-Image Disentanglement Network with weakly supervised learning, which can simultaneously correct brightness and suppress image artifacts in the feature domain, improving the robustness of pixel shifts between training pairs. Li et al. [41] proposed a knowledge distillation method for LIE task. The proposed method uses a teacher and student framework and knowledge transfer between the teacher and student network is accomplished by distillation loss based on attention maps.

Due to the difficulty in obtaining paired low/normal light images in real-world scenarios, researchers have begun exploring unsupervised methods for LIE. Jiang et al. [42] proposed a paradigm that eschews the use of paired datasets, instead leveraging information extracted directly from the input to constrain network training. Guo et al. [43] proposed a method to enhance image brightness by using DL networks to estimate curves specific to each image. This method uses a set of well-designed loss functions to eliminate the need for data sets in the training process and greatly simplifies the network training process. Nguyen et al. [12] generated synthetic baseline images by amalgamating various original images, thereby replicating every conceivable lighting condition, for the purpose of educating the improvement network. Zhao et al. [10] and Liang et al. [11] utilized the priors of untrained neural networks for unsupervised LIE, requiring no training samples other than the input images themselves. Luo et al. [44] introduced a mutual learning strategy that learns the corresponding normal-light images of low-light images through knowledge distillation via mutual learning across two branches. Fu et al. [45] utilized a straightforward self-supervised method to filter out unsuitable characteristics from the initial images, in order to acquire the hidden attributes common among paired low-light photographs.

C. Self-paced Learning

In recent research, SPL has gained significant traction in the field of computer vision, emphasizing the role of adaptive learning processes in enhancing model performance and data comprehension [46], [47]. SPL's core idea is to allow learning models or algorithms to gradually incorporate complexity, thereby aligning with human learning patterns. Jiang et al. [48] explored SPL in image classification tasks, introducing diversity regularization to improve generalization in neural networks. This approach highlights the effectiveness of SPL in managing varied and complex visual data. Kumar et al. [49] provided a foundational perspective on SPL in latent variable models, offering insights beneficial for optimizing computer vision algorithms. Guo et al. [50] used an SPL strategy to improve the performance of the network by using the Ground Truth (GT) of the attention map as a guidance during difficult samples. Zhang et al. [51], [52] pointed out that in few-shot learning, tasks are often randomly selected for contextual training without considering their difficulty and

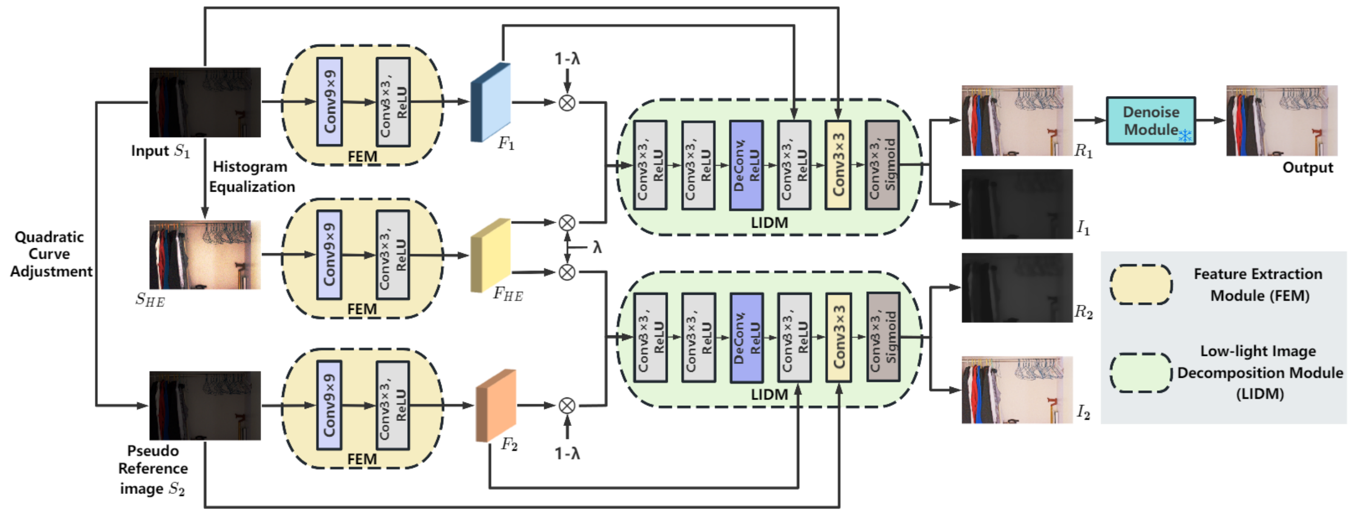


Fig. 3. The overall architecture of our proposed method. S_1, S_{HE}, and S_2 respectively represent the original image, the histogram equalization result of the original image, and the pseudo-reference image. All FEMs and LIDMs share parameters. The FEMs extract the basic features of the input image and support subsequent SPL strategy and image enhancement. The LIDMs use the features extracted from the FEMs to enhance the clarity of low-light images. R_1, I_1, and R_2, I_2 represent the reflectance and illumination components derived from the decomposition of S_1 and S_2, respectively. After obtaining the preliminary enhancement result R_1, it is input into a pre-trained denoise module to obtain the final outcome.

quality, which may hinder the meta-learner from gradually improving its generalization ability. To address this problem, they adopted the SPL strategy to train the meta-learner with tasks of gradually increasing difficulty, thereby improving its generalization ability and performance. Dai et al. [53] proposed a training strategy termed DMH-CL for complex pose estimation, brought from SPL which mainly addresses easy examples in the early training stage and hard ones in the later stage. This approach also significantly enhances the performance of the existing model. These studies collectively showcase the breadth of SPL applications in computer vision, from basic image classification to complex pattern recognition, highlighting its potential to enhance model adaptability and efficiency. Although SPL has shown promising results in the above tasks, for LIE, how to select difficult samples is a critical challenge. In our research, we determine the sample difficulty through the result of histogram equalization of the input image and adaptively adjust the feature guidance weights through the SPL strategy.

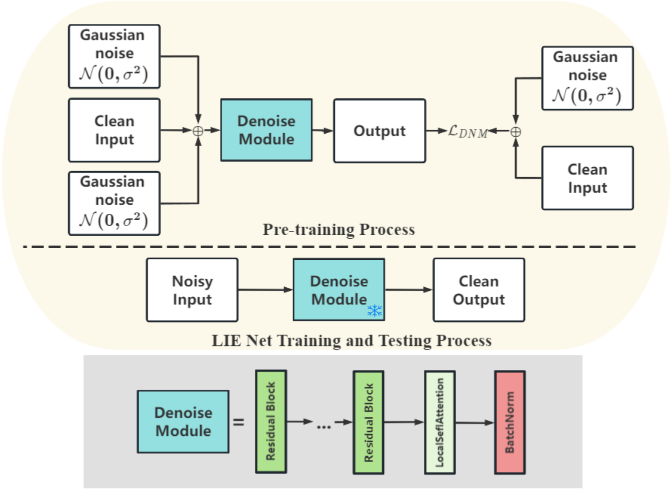


Fig. 4. The upper image is a schematic representation of the denoise module. The initial half illustrates the pre-training process of the denoise module. The subsequent half indicates that the parameters are frozen during both the training and testing phases of LIE. The lower image depicts the specific structure of the denoise module.

III. Proposed Method

A. Motivation

Following Retinex theory [54], a low-light image S can be represented as the product of a reflectance image R and an illumination component I:

$$S = R * I, \tag{1}$$

where $*$ signifies multiplication on an individual element basis. The illumination component I characterizes the brightness level across object surfaces, which are expected to be sectionally smooth and devoid of texture, whereas the reflectance R represents the physical properties of the objects, including the textures and detail information observed in the image.

Although existing methods have achieved certain effects in enhancing low-light images, they usually treat all samples equally [16]–[19]. Owing to the specificity of the LIE task, extremely low light conditions cause the images to contain very little useful information. These are difficult samples in the early stages of network training, which can lead to suboptimal or even erroneous solutions. So we need to train the network from simple samples to difficult samples. Therefore, SPL is highly suitable for the LIE task. To be specific, we use the features of the histogram-equalized version of the input image to select difficult samples and adaptively adjust the weights of the guidance features according to the SPL strategy.

B. Overall Architecture

Fig. 3 shows the overall structure of SPNet, which comprises three primary components: the FEM, the LIDM, and the denoise module. The FEMs aim to extract basic features effectively from the input images, laying the foundation for the subsequent feature fusion of the SPL strategy and image enhancement processing. The LIDMs use the features extracted from the FEMs to enhance the clarity of low-light images. The denoise module processes the preliminary enhancement results to improve the final image quality. We apply histogram equalization and quadratic curve transformation to the original input S_1, generating S_{HE} and S_2. They, along with S_1, are input into the FEM, resulting in F_1, F_{HE}, and F_2. F_1 and F_2 are fused with F_{HE} for feature integration, using the guidance weight λ, which is dynamically adjusted on the basis of the network's restoration quality. Following this, the combined results are fed into the LIDM for retinex decomposition, resulting in R_1, I_1, R_2, and I_2. Following [36], [37], [55], we take the output reflectance part R_1 as the targeted normal-light image. Finally, to reduce the potential noise in the dark area, which is amplified during the LIE process, we input R_1 into the pre-trained denoise module [56] to obtain the final output.

1) Feature Extraction Module (FEM): The FEM primarily comprises two convolutional layers: conv1 and conv2. The conv1 layer employs a 9×9 convolutional kernel, whereas the conv2 layer uses a 3×3 convolutional kernel. The FEMs are designed to share parameters.

2) Low-light Image Decomposition Module (LIDM): As shown in Fig. 3, the LIDM consists of five convolutional layers and a deconvolution layer, gradually increasing the depth of the network to capture more complex features. It ultimately produces the reflectance map (*R*) and the illumination map (*I*). These outputs are processed by a sigmoid function to ensure that the results fall within an appropriate numerical range.

C. Loss Function

Without the normal light image as a reference, the Retinex decomposition will be unstable. Therefore, we incorporate several fundamental constraints intrinsic to Retinex decomposition

$$\begin{aligned}\mathcal{L}_R = \omega_0\|S_1 - R_1 \cdot I_1\|_1 + \omega_1\mathcal{L}_{hep} + \omega_2\|\nabla R_1\|_1 \\ + \omega_3\|\nabla S_1 - \nabla I_1\|_1,\end{aligned} \tag{2}$$

The symbol ∇ signifies horizontal and vertical gradients, and ω_0, ω_1, ω_2, and ω_3 denote the weights. $\|S_1 - R_1 * I_1\|_1$ is the reconstruction loss, which constrains the network to perform a reasonable Retinex decomposition. $\mathcal{L}_{hep}$ refers to the histogram equalization loss. $\|\nabla R_1\|_1$ is used to constrain the smoothness of the generated image. Its primary function is to maintain the naturalness and continuity of images in enhancement tasks, avoiding excessive noise generation within the image. By minimizing $\|\nabla S_1 - \nabla I_1\|_1$, the network is encouraged to generate an illumination component that is gradient-wise similar to the original low-light image, which contributes to preserving the structure and details and improving the brightness and contrast of the image.

In the absence of GT data, imposing effective constraints on the training of unsupervised networks becomes crucial. Our study employs a quadratic curve transformation to generate a sequence of images depicting the same scene under varying brightness conditions. Following [44], we refer to these images as pseudo-reference images. The formula for generating pseudo-reference images S_2 from the input low-light images S_1 is as follows:

$$S_2 = S_1 + \alpha * S_1 * (1 - S_1), \tag{3}$$

where $\alpha \in [0, 1]$ represents the parameter controlling brightness. According to the derivation from [44], we can obtain the representation of the illumination component I_2 of the pseudo reference image from Equation (3) as:

$$I_2 \approx (1 + \alpha) * I_1, \tag{4}$$

where I_1 is the illumination component of the original image. Thus, the gradient relationship of the illumination between the original input image and the pseudo-reference image can be derived:

$$\nabla I_2 \approx (1 + \alpha) * \nabla I_1, \tag{5}$$

We observe that both Equation (4) and Equation (5) can impose constraints on the training of the network. As shown in Table IV, the training results of the network are obviously better when Equation (5) is used than when Equation (4) or neither is used. This is because, in image processing, gradients are usually closely related to the edges and texture information of images. By constraining the gradients, the network can pay more attention to the structural changes in images rather than just the direct changes in pixel values. Moreover, because the human visual system (HVS) is highly sensitive to edges and textures, and gradients can capture this information. So using gradient constraints can align the learned features more closely with the HVS and improve our task's performance. The specific form of the gradient similarity of the illumination constraint is as follows:

$$\mathcal{L}_{GSI} = \|(1 + \alpha) * \nabla I_1 - \nabla I_2\|_1, \tag{6}$$

Our network's overall loss is a linear combination of $\mathcal{L}_{GSI}$ and $\mathcal{L}_R$.

$$\mathcal{L}_{all} = \omega_4 * \mathcal{L}_R + \mathcal{L}_{GSI} \tag{7}$$

where ω_4 is the weighting parameter to balance each loss.

D. Denoise Module

Although the aforementioned LIE module effectively increases the brightness of dark areas, this process also amplifies the potential noise in dark regions. In light of this, our study further uses a denoise module to effectively eliminate the noise hidden in low-light areas. Our denoise module is inspired by [56], and is pre-trained on ten arbitrary low-light images from the training set of the LOL dataset, and its parameters are frozen during the training and testing phases of the FEM and LIDM.

As shown in Fig. 4, during the pre-training phase of the denoise module, we add Gaussian noise $\mathcal{N}(0, \sigma^2)$ to arbitrary images $\mathbf{X}$:

$$\mathbf{Y} = \mathbf{X} + \mathcal{N}(0, \sigma^2), \tag{8}$$

Subsequently, we add the same intensity of Gaussian noise $\mathcal{N}(0, \sigma^2)$ to $\mathbf{Y}$:

$$\mathbf{Z} = \mathbf{Y} + \mathcal{N}(0, \sigma^2), \tag{9}$$

During training denoise module, we use $\mathbf{Y}$ as the GT to constrain the output $\mathcal{F}_\theta(\mathbf{Z})$ obtained from the denoising network. The loss function is as follows:

$$\mathcal{L}_{DNM} = \|\mathcal{F}_\theta(\mathbf{Z}) - \mathbf{Y}\|_2^2, \tag{10}$$

$\mathcal{F}_\theta()$ represents the mapping of the denoise module. As shown in Fig. 4, our denoise module consists of multiple residual blocks and a self-attention block. Each residual block consists of two convolutional layers and a subsequent batch normalization layer and uses the LeakyReLU activation function. After multiple residual blocks, the network employs a local self-attention module [57]. This module calculates attention weights and then applies these weights to the input feature maps, enabling the network to better focus on important areas of the image. The self-attention module is another convolutional layer and batch normalization layer, which further processes the feature maps.

E. Self-Paced Learning

In the absence of normal-light reference images, we utilize an SPL strategy here to allow the network training process to progress from easy to difficult. During training, we apply histogram equalization to the input image S_1 to obtain the S_{HE}. S_{HE} enhances image contrast and improves the quality and clarity of the details in low-light images, providing good guidance for unsupervised network training. According to Zhang's study [15], the feature mapping of histogram-equalized enhanced images on the VGG network is similar to that of the GT. So, we employ the difference between the restored image and histogram-equalized enhanced images in feature space to assess the state of the image restoration and use it as a basis for determining the difficulty level of the samples. The specific representation of the histogram equalization prior loss is as follows:

$$\mathcal{L}_{hep} = \|\phi(R_1) - \phi(S_{HE})\|_2^2, \tag{11}$$

where $\phi()$ represents the feature mappings extracted by the VGG-19 model pre-trained on ImageNet. S_1, S_{HE}, and the pseudo-reference image S_2 are separately fed into the FEM, generating the features F_1, F_{HE}, and F_2, respectively. Following an SPL strategy, F_1 and F_2 are fused with F_{HE} and then input into the LIDM. The expression for the fusion of F_1 with F_{HE} is as follows:

$$F = \lambda \cdot F_{HE} + (1 - \lambda) \cdot F_1 \tag{12}$$

λ is the guidance weight of SPL. The fusion of F_2 with F_{HE} is similar to the fusion of F_1 with F_{HE}. λ dynamically adjusts according to $\mathcal{L}_{hep}$, specifically as follows:

$$\lambda = \begin{cases} 1, & \text{if } \mathcal{L}_{hep} \geq 1, \\ \frac{\mathcal{L}_{hep} - 0.3}{1 - 0.3}, & \text{if } 1 > \mathcal{L}_{hep} > 0.3 \\ 0, & \text{if } \mathcal{L}_{hep} \leq 0.3 \end{cases} \tag{13}$$

Equation (13) is employed for adjusting the weights of F_1 and F_{HE}. At the initial stage of network training, a higher value of $\mathcal{L}_{hep}$ indicates poorer restoration performance, classifying the samples as difficult. In the initial stages of network training, difficult samples may lead the training toward suboptimal solutions or even incorrect optimization directions. To filter out difficult samples in the early stages and guide the network in the correct optimization direction, we set the value of λ to 1. This makes F equivalent to F_{HE}, allowing the network to initially focus on learning from simpler samples. As the training progresses and restoration quality improves, λ decreased to increase the weight of F_1. λ is set to 0 when $\mathcal{L}_{hep}$ is less than 0.3, and F_{HE}'s influence is stopped to avoid over-enhancement and color distortion from the histogram equalization.

Algorithm 1 SPNet for LIE

```
Input: Low-light image set $S_1$, pseudo-reference image set $S_2$, and $S_1$'s histogram-equalized version $S_{HE}$
Output: Normal-light image
Initialize: Initialize FEM $\mathcal{G}_{FEM}$ and LIDM $\mathcal{G}_{LIDM}$; $\lambda$ = 1
repeat
    1: $F_1 = \mathcal{G}_{FEM}(S_1)$, $F_{HE} = \mathcal{G}_{FEM}(S_{HE})$, $F_2 = \mathcal{G}_{FEM}(S_2)$;
    2: According to the SPL strategy, by merging features $F_1$ and $F_{HE}$ as well as $F_2$ and $F_{HE}$ through Equation (12), $F_{fusion1}$ and $F_{fusion2}$ are obtained;
    3: $R_1$, $I_1 = \mathcal{G}_{LIDM}(F_{fusion1})$, $R_2$, $I_2 = \mathcal{G}_{LIDM}(F_{fusion2})$;
    4: Minimize the objective function Equation (7) and update $\mathcal{G}_{FEM}$, $\mathcal{G}_{LIDM}$, $\lambda$ simultaneously;
    5: Input $R_1$ into the pre-trained denoise module to obtain the final output.
until convergence
```

IV. Experiments

A. Implementation Details

Training and testing procedures were executed via PyTorch 1.10.0 on a PC equipped with an Intel(R) Core(TM) i7-12700K and an NVIDIA GeForce RTX 3090 Ti GPU. We iteratively optimized the FEM and LIDM utilizing the Adam optimizer. The pre-training phase of the denoise module was also optimized by the Adam optimizer. The batch size and the patch size are set to 16 and 128. The model's learning rate was initiated at 0.0001. The weights ω_0, ω_1, ω_2, ω_3, and ω_4 are set to 1, 0.2, 0.02, 0.2, and 1, respectively. The number of epochs for the model was 150.

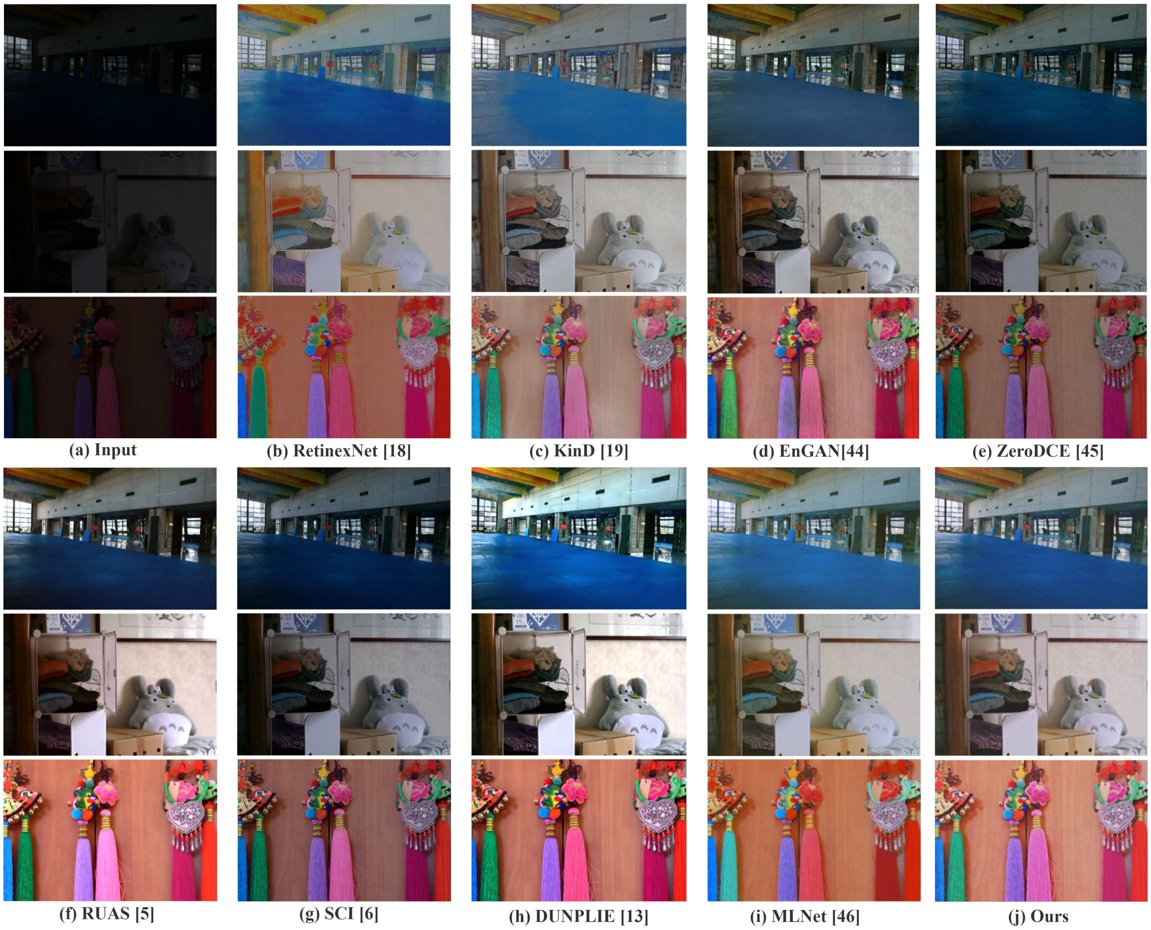
B. Datasets and Criteria

We utilized 324 low-light images gathered from SICE [61] and LOL [16] to train the network. To verify the effectiveness of our proposed method, we conducted tests on the test sets of nine datasets: LOL [16](15 paired low/normal light images for testing), LSRW [6](50 paired low/normal light images for testing), LOL-syn [62](100 paired low/normal light

TABLE I

ANALYSIS COMPARISON OF VARIOUS APPROACHES ACROSS NINE STANDARDIZED DATASETS. FOR DL-BASED UNSUPERVISED METHODS, THE TOP AND SECOND-BEST OUTCOMES ARE DISTINGUISHED BY RED AND BLUE COLORS. THE OVERALL 'RANK' IS CALCULATED BY AVERAGING THE RANKS OBTAINED USING DIFFERENT METHODS FOR EACH DATASET, AND 'ROR' REPRESENTS THE RANK OF THE OVERALL 'RANK'.

Dataset	Metrics	Conventional Methods			Supervised Methods		Unsupervised Methods						
		NPE [26]	SRIE [58]	LIME [37]	RetinexNet [16]	KinD [17]	EnGAN [42]	ZeroDCE [43]	RUAS [59]	SCI [60]	DUNPLIE [11]	MLNet [44]	SPNet
LOL	PSNR	16.94	11.87	16.76	16.77	17.65	17.37	14.83	16.39	16.06	15.49	**20.07**	**20.12**
	SSIM	0.48	0.50	0.56	0.56	0.76	0.66	0.58	0.58	0.49	0.65	**0.76**	**0.79**
LSRW	PSNR	16.12	15.96	16.97	15.90	16.47	16.31	15.83	14.44	15.02	15.04	**16.50**	**16.63**
	SSIM	0.40	0.54	0.40	0.37	0.49	0.47	0.47	0.43	0.48	**0.51**	0.49	**0.52**
LOL-syn	PSNR	16.72	15.56	16.58	17.14	18.32	16.57	12.23	13.40	15.43	16.80	**18.46**	**17.00**
	SSIM	0.77	0.66	0.74	0.76	0.78	0.73	0.73	0.64	0.74	0.75	**0.81**	**0.83**
LOL-real	PSNR	17.33	17.34	15.24	15.47	23.78	**18.23**	**18.06**	15.33	17.30	12.97	17.88	17.98
	SSIM	0.46	0.69	0.47	0.41	0.87	0.61	0.58	0.52	0.54	0.39	**0.76**	**0.79**
LIME	NIQE	3.93	4.53	4.15	4.61	4.73	**3.77**	4.60	4.15	**4.10**	5.40	4.40	4.27
	LOE	431.31	305.02	793.9	539.64	255.80	618.77	298.72	802.25	**84.97**	403.25	246.42	**198.30**
MEF	NIQE	3.52	3.47	3.70	4.41	3.88	**3.04**	4.53	3.77	3.72	3.73	3.78	**3.59**
	LOE	422.32	270.26	939.11	708.25	275.47	589.32	334.18	784.17	**94.97**	527.80	268.22	**222.48**
NPE	NIQE	4.18	4.20	4.27	4.57	4.19	**3.49**	4.18	5.41	3.58	3.85	4.11	**3.54**
	LOE	163.04	164.37	1174.92	653.83	180.91	559.92	440.77	1065.7	**288.36**	457.63	463.6	**438.82**
DICM	NIQE	3.83	4.56	3.84	4.43	4.13	**3.32**	3.41	4.82	3.65	3.95	4.05	**2.97**
	LOE	303.99	547.39	531.51	621.72	250.41	699.61	**336.73**	621.72	**284.12**	422.47	600.86	524.72
VV	NIQE	2.47	3.13	2.47	2.70	3.03	3.62	2.89	3.42	**2.36**	4.37	3.20	**2.76**
	LOE	225.61	117.57	309.56	382.56	133.01	453.87	**150.21**	579.83	**99.24**	401.77	227.66	280.58
Rank		5.7	6.7	7.6	8.7	4.5	6.2	6.9	9.9	5.3	7.9	**4.9**	**3.2**
RoR		5	7	8	11	2	6	8	12	4	10	**3**	**1**


(a) Input (b) RetinexNet [18] (c) KinD [19] (d) EnGAN[44] (e) ZeroDCE [45]

(f) RUAS [5] (g) SCI [6] (h) DUNPLIE [13] (i) MLNet [46] (j) Ours

Fig. 5. Visual comparison of state-of-the-art LIE methods on dimly lit images from the LOL dataset.

images for testing), and LOL-real [62](100 paired low/normal light images for testing), LIME [37](10 low-light images for testing), MEF [63](17 low-light images for testing), NPE [26](8 low/normal light images for testing), DICM [64](69 low-light images for testing), and VV(24 low-light images for testing). Since LOL, LSRW, LOL-syn, and LOL-real include reference images, we quantitatively assessed the enhancement results via two classic full-reference image quality assessment metrics: PSNR and SSIM [65]. For datasets without reference to normal-light images, such as LIME, MEF, NPE, DICM, and VV, we applied NIQE [66] and LOE [26] as the image quality assessment metrics. Higher values of PSNR and SSIM, and lower scores of NIQE and LOE indicate better image quality.

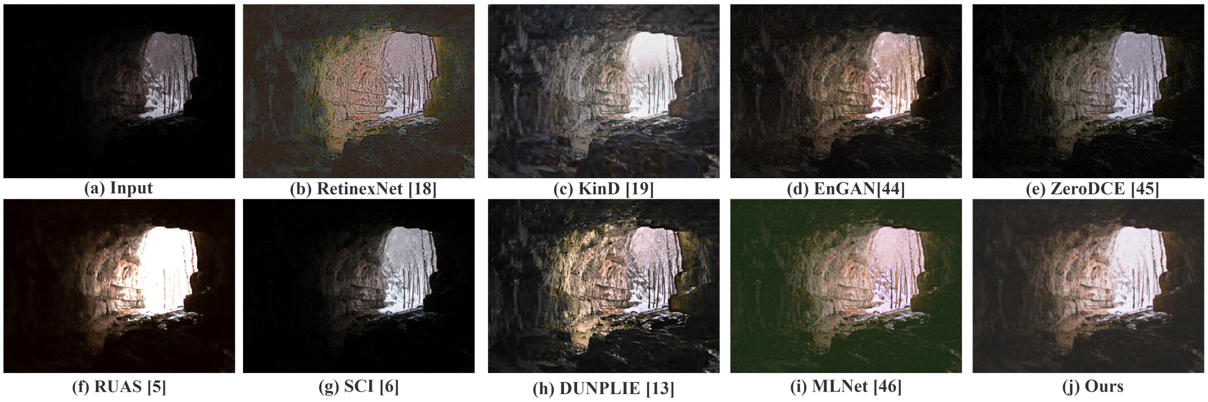


Fig. 6. Visual comparison of real-world images from the MEF dataset using state-of-the-art LIE methods. It is evident that under extreme lighting conditions, our method can recover more content and details in areas compared to other unsupervised approaches.

TABLE II
THE MODEL PARAMETERS AND RUNNING TIMES OF CNN-BASED METHODS AND OUR METHOD.

Method	PARAMETERS(M)	TIME(S)
RetinexNet	0.4448	0.1140
KinD	8.0164	0.1935
EnGAN	8.6373	0.0593
ZeroDCE	0.0764	0.0012
RUAS	0.0034	0.0046
SCI	0.0003	0.0005
MLNet	0.4710	0.0745
SPNet w/o denoise module	0.4856	0.0209
SPNet	1.2226	0.0431

TABLE III
THE ABLATION STUDY OF SPL. THE TOP RESULTS ARE EMPHASIZED IN BOLD.

Configurations	PSNR	SSIM
Setting ($\mathcal{A}$)	19.95	0.76
Setting ($\mathcal{B}$)	19.91	0.75
Setting ($\mathcal{C}$)	19.98	0.77
Setting ($\mathcal{D}$)	**20.12**	**0.79**

C. Comparison with State-of-the-Art Methods

We compare SPNet with three conventional methods, NPE [26], LIME [37], and SRIE [58]; two supervised learning methods, KinD [17] and RetinexNet [16]; and six unsupervised learning methods, EnlightGAN [42], RUAS [59], SCI [60], ZeroDCE [43], DUNPLIE [11], and MLNet [44]. The results of these methods are reproduced by using their official source.

In Table I, we present the result of the proposed SPNet and compare LIE methods on nine datasets. The 'Rank' and 'RoR' show that our SPNet achieves the highest ranking on nine datasets compared with eight DL-based LIE methods and three traditional LIE methods. This demonstrates the effectiveness of our method and its robust ability in a variety of real-world low-light environments. Specifically, it achieves the highest PSNR values on the LOL and LSRW datasets, and the highest SSIM values on the LOL, LSRW, LOL-syn, and LOL-real datasets. Although our method's performance metrics on non-reference datasets are lower than those of SCI, the restored results of our method still have excellent visual quality, as shown in Fig. 6 and Fig. 7.

TABLE IV
THE ABLATION STUDY OF THE ILLUMINATION GRADIENT SIMILARITY LOSS. THE TOP RESULTS ARE EMPHASIZED IN BOLD.

Loss	PSNR	SSIM
$\mathcal{L}_R$	19.91	0.77
$\mathcal{L}_R + \mathcal{L}_{ISL}$	19.99	0.77
$\mathcal{L}_R + \mathcal{L}_{RSL}$	19.81	0.74
$\mathcal{L}_R + \mathcal{L}_{GSI}$	**20.12**	**0.79**

D. Qualitative Evaluations

Fig. 5 shows the enhancement results of nine LIE methods on the LOL dataset. Although RetinexNet enhances the brightness of the picture, it produces some unnatural colors, causing the edges of the objects to become less sharp. The other six unsupervised methods still showed poor restoration effects on the left side of the second image. ZeroDCE, RUAS, SCI, and DUNPLIE lost the texture details on the leftmost side of the image. MLNet restored the color of the toy in the middle of the image to grayish green, indicating color deviation. SPNet restores the brightness of the picture better while keeping the color and contrast natural. In the second image, it restores the extremely dark area on the left side of the image while keeping the other areas from being overexposed.

In Fig. 7, we showcase the results of multiple LIE methods on the LIME dataset. RetinexNet continues to exhibit overly saturated colors, KinD shows ghosting effects, and EnGAN and ZeroDCE demonstrate insufficient restoration. The RUAS and SCI methods result in overexposure. Moreover, our SPNet enhances low-light areas without overexposure. Fig. 6 displays the outcomes of different LIE methods on the MEF dataset. The chosen image is from an environment with an extreme lack of illumination, where it is evident that, compared with ZeroDCE, RUAS and SCI, our method can recover more image information and details under extreme low-light conditions. This phenomenon is particularly noticeable in the depths of the caves in the image that the other LIE methods fail to effectively restore, and our method still manages to reconstruct the contours of the rocks. But there is a slight overexposure at the cave entrance, which may be because SPNet considers the image to be enhanced as a whole and globally enhances it. To address this issue, in future work, we may try to use the attention mechanism to adaptively

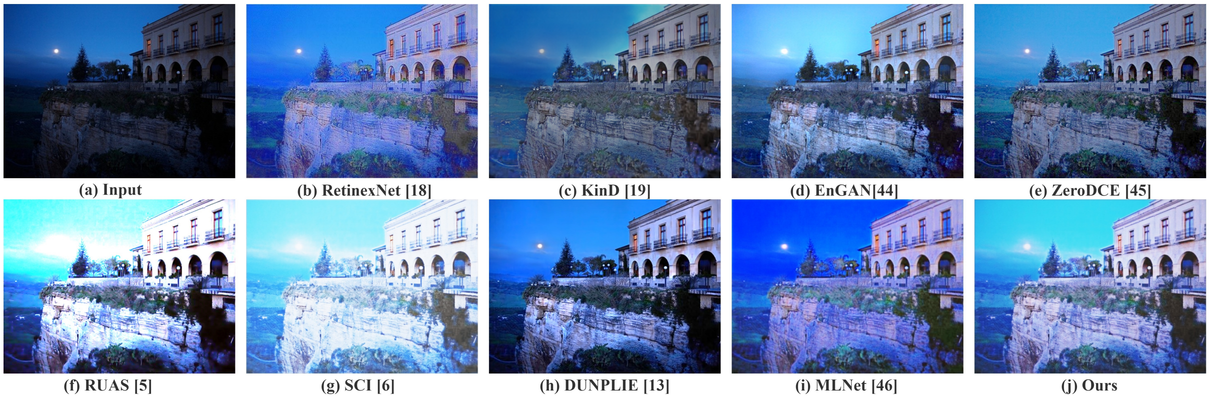


Fig. 7. Visual comparison of real-world images from the LIME dataset using state-of-the-art LIE methods. It can be observed that, compared to other methods, our approach enhances illumination adequately without issues such as overly vivid colors.

TABLE V
THE ABLATION STUDY OF DENOISE MODULE. THE TOP RESULTS ARE EMPHASIZED IN BOLD.

	PSNR	SSIM
w/o denoise module	19.97	0.75
with denoise module	**20.12**	**0.79**

adjust the exposure to avoid local overexposure. Fig. 8 shows the comparison result of our method and other LIE methods on the LSRW dataset. Compared with other methods, our approach can restore extremely dimly lit corridor areas without introducing distortion. RetinexNet, KinD, and MLNet show various levels of distortion. ZeroDCE, RUAS, and SCI are unable to adequately restore image details. Fig. 9 provides the visual comparison results of LIE methods on the LOL-syn dataset. Compared with the competing methods, our method can restore the color better, without distortions. At the same time, our method can restore details such as the windows of the house at the bottom of the picture.

Computational Efficiency We also investigate the computational efficiency of our method and compare it with competing methods in terms of the parameter and the inference time. All experiments are conducted via PyTorch 1.10.0 on a PC equipped with an Intel(R) Core(TM) i7-12700K and an NVIDIA GeForce RTX 3090 Ti GPU. The input image size was 600×400×3. As shown in Table II, our method does not have obvious advantages in computational efficiency. In the current study, the design of SPNet is more focused on improving the image quality from the perspectives of brightness and contrast, and SPNet outperforms other competing methods on average across nine datasets. Additionally, we found that the denoise module accounts for more than 60% of the parameters in our method, which shows that the parameter amount of our LIE module is competitive. Our next research goal is to achieve denoising effects with less computational complexity.

E. Ablation Studies

This section conducts ablation studies on the LOL dataset to investigate the effectiveness of some key components in our proposed network.

1) To validate the effectiveness of the SPL approach, a range of ablation tests were conducted, encompassing four distinct configurations: ($\mathcal{A}$) self-paced learning applied neither in S_1 nor in S_2 ($\mathcal{B}$) self-paced learning applied only in S_1 ($\mathcal{C}$) self-paced learning applied only in S_2 ($\mathcal{D}$) self-paced learning applied in both S_1 and S_2 (as proposed in this study). As shown in Table III, the configuration used in our proposed method outperforms the other settings, which also validates the effectiveness of SPL. In configuration ($\mathcal{A}$), the lack of SPL makes certain samples excessively challenging for the current network, thereby impacting the overall learning effectiveness. In configurations ($\mathcal{B}$) and ($\mathcal{C}$), since either S_1 or S_2 did not employ SPL, the restored reflection layer and illumination layer were less than ideal. This further affects the constraint effectiveness of the illumination gradient similarity loss, and might even have negative implications.

2) To ascertain the effectiveness of the gradient similarity illumination loss, two common loss functions are compared: (a) the training of the network is constrained by the linear relationship between the illumination components I_1 and I_2, as shown in Equation (14); (b) according to Luo et al. [44], the reflection components R_1 and R_2 should be approximately equal, as shown in Equation (15)

$$\mathcal{L}_{ISL} = \|(1+\alpha) * I_1 - I_2\|_1 \tag{14}$$

$$\mathcal{L}_{RSL} = \|R_1 - R_2\|_1 \tag{15}$$

As shown in Table IV, our proposed Illumination Gradient Similarity Loss provides the most significant enhancement in recovery effects for the LIE task.

3) To confirm the ability of the denoise module to eliminate noise that is amplified during the enhancement of low-light images input to the LIE network, we carry out ablation experiment (3). The results, as presented in Table V, demonstrate that the inclusion of the denoise module results in a 0.15dB increase in the PSNR.

4) To verify the appropriateness of the selected lower threshold values in the SPL strategy, we tested lower threshold values ranging from 0.1 to 0.9 in increments of 0.1. Fig. 10 shows the experimental results. Increasing the lower threshold values, which indicates a growing influence of the SPL strategy, improves the network's PSNR and SSIM on the test set. This suggests that enhancing SPL helps our network converge in a more accurate direction. However, when the lower threshold

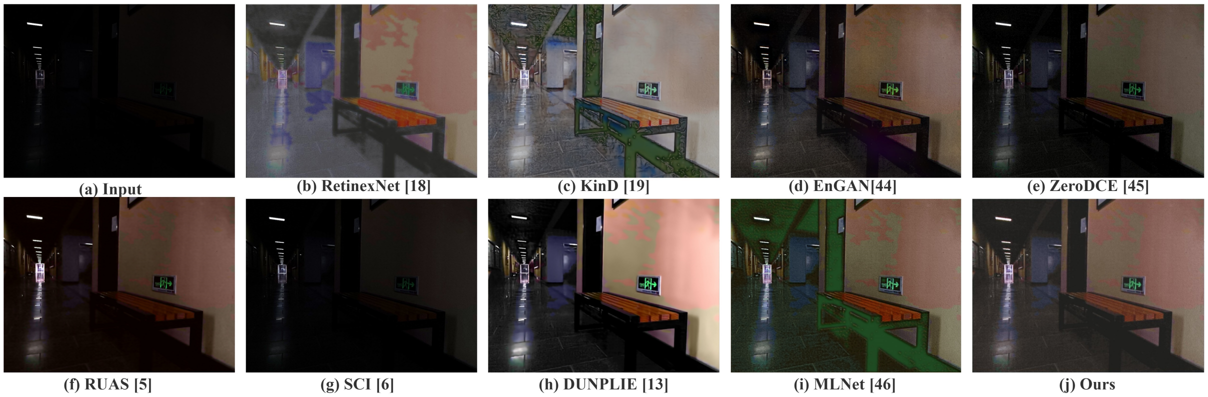


Fig. 8. Visual comparison of real-world images from the LSRW dataset using state-of-the-art LIE methods.

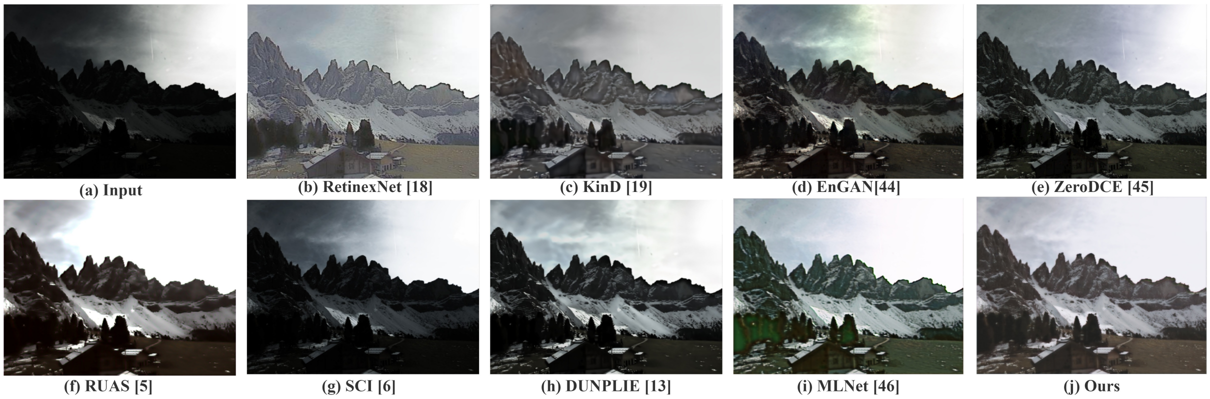


Fig. 9. Visual comparison of real-world images from the LOL-syn dataset using state-of-the-art LIE methods.

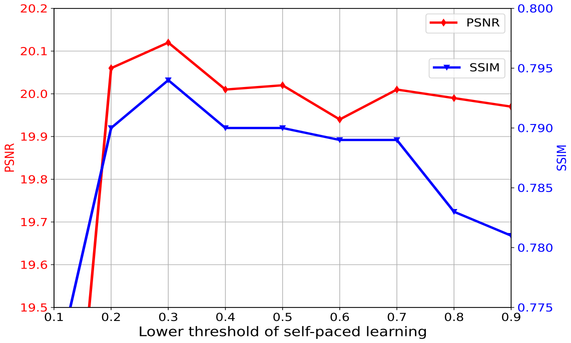


Fig. 10. Ablation study of the lower threshold for the SPL strategy.

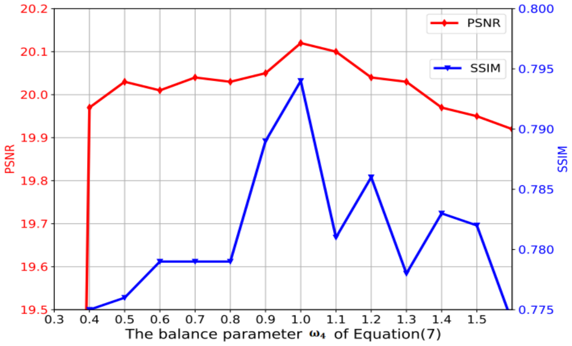


Fig. 11. Ablation study of the balance parameter ω_4 for Equation (7).

values are set too low (below 0.3), the PSNR and SSIM of the network on the test set will decrease. This occurs because setting the lower threshold values too low prevents the network from stopping SPL at the right time. The network does not learn from more challenging samples, leading to poor training results. By comparing the results in Fig. 10, we finally set the threshold to 0.3 due to the superior performance achieved.

5) We conducted an ablation experiment on the key parameter ω_4 in Equation (7). As shown in Fig. 11, the experimental results show that SPNet achieves better results when setting ω_4 to 1.0.

6) In order to verify the rationality of the data setting during the pre-training process of the denoise module, we conducted ablation experiment (6). As shown in Table VI, the noise-free image and singly noise-added image represent the **X** and **Y** in Equation (8), respectively, and the doubly noise-added image represents the **Z** in Equation (9). The experimental results show the effectiveness of our data setting.

TABLE VI
THE ABLATION STUDY OF THE DATA SETTING DURING THE PRE-TRAINING PROCESS OF THE DENOISE MODULE. THE TOP RESULTS ARE EMPHASIZED IN BOLD.

Pre-training data	PSNR	SSIM
singly noise-added image (**Y**), noise-free image (**X**)	19.43	0.763
doubly noise-added image (**Z**), singly noise-added image (**Y**)	**20.12**	**0.794**

V. Conclusion

In this study, we introduce a LIE method featuring an SPL strategy guiding the network's training from simpler to more complex tasks. Additionally, our method constrains network training by leveraging the prior relationship between the illumination components of both low-light images and pseudo-reference images. The preliminary results are then fed into a pre-trained denoise module to achieve the final restoration outcome. Comprehensive experiments across nine standard benchmarks reveal that our method outperforms eight DL-based LIE methods in terms of quality and metrics and achieves performance comparable to three conventional methods. In future work, we will update the denoise module's structure to improve the network's overall computational efficiency. Additionally, we plan to extend our method to other image processing tasks, such as image deraining and MRI reconstruction [67], [68].

References

[1] K. Lu and L. Zhang, "Tbefn: A two-branch exposure-fusion network for low-light image enhancement," *IEEE Transactions on Multimedia*, vol. 23, pp. 4093–4105, 2021.

[2] Q. Ma, Y. Wang, and T. Zeng, "Retinex-based variational framework for low-light image enhancement and denoising," *IEEE Transactions on Multimedia*, vol. 25, pp. 5580–5588, 2023.

[3] J. Xu, M. Yuan, D.-M. Yan, and T. Wu, "Illumination guided attentive wavelet network for low-light image enhancement," *IEEE Transactions on Multimedia*, vol. 25, pp. 6258–6271, 2023.

[4] Q. Jiang, Y. Mao, R. Cong, W. Ren, C. Huang, and F. Shao, "Unsupervised decomposition and correction network for low-light image enhancement," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 19 440–19 455, 2022.

[5] 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*, pp. 1–14, 2023.

[6] 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.

[7] Z. Cui, K. Li, L. Gu, S. Su, P. Gao, Z. Jiang, Y. Qiao, and T. Harada, "You only need 90k parameters to adapt light: a light weight transformer for image enhancement and exposure correction." in *BMVC*, 2022, p. 238.

[8] X. Xu, R. Wang, and J. Lu, "Low-light image enhancement via structure modeling and guidance," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 9893–9903.

[9] Y. Wu, C. Pan, G. Wang, Y. Yang, J. Wei, C. Li, and H. T. Shen, "Learning semantic-aware knowledge guidance for low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 1662–1671.

[10] Z. Zhao, B. Xiong, L. Wang, Q. Ou, L. Yu, and F. Kuang, "Retinexdip: A unified deep framework for low-light image enhancement," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 3, pp. 1076–1088, 2021.

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

[12] H. Nguyen, D. Tran, K. Nguyen, and R. Nguyen, "Psenet: Progressive self-enhancement network for unsupervised extreme-light image enhancement," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 1756–1765.

[13] W. Wang, D. Yan, X. Wu, W. He, Z. Chen, X. Yuan, and L. Li, "Low-light image enhancement based on virtual exposure," *Signal Processing: Image Communication*, vol. 118, p. 117016, 2023.

[14] 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.

[15] F. Zhang, Y. Shao, Y. Sun, C. Gao, and N. Sang, "Self-supervised low-light image enhancement via histogram equalization prior," in *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*. Springer, 2023, pp. 63–75.

[16] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," *British Machine Vision Conference*, 2018.

[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. 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5901–5910.

[19] X. Xu, R. Wang, C.-W. Fu, and J. Jia, "Snr-aware low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 17 714–17 724.

[20] J. Guo, J. Ma, Á. F. García-Fernández, Y. Zhang, and H. Liang, "A survey on image enhancement for low-light images," *Heliyon*, vol. 9, no. 4, 2023.

[21] Z. Huang, Z. Wang, J. Zhang, Q. Li, and Y. Shi, "Image enhancement with the preservation of brightness and structures by employing contrast limited dynamic quadri-histogram equalization," *Optik*, vol. 226, p. 165877, 2021.

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

[23] P. Shanmugavadivu and K. Balasubramanian, "Thresholded and optimized histogram equalization for contrast enhancement of images," *Computers & Electrical Engineering*, vol. 40, no. 3, pp. 757–768, 2014.

[24] Y. Li, Z. Yuan, K. Zheng, L. Jia, H. Guo, H. Pan, J. Guo, and L. Huang, "A novel detail weighted histogram equalization method for brightness preserving image enhancement based on partial statistic and global mapping model," *IET Image Processing*, vol. 16, no. 12, pp. 3325–3341, 2022.

[25] E. H. Land and J. J. McCann, "Lightness and retinex theory," *Josa*, vol. 61, no. 1, pp. 1–11, 1971.

[26] 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.

[27] W. Wang, Z. Chen, X. Yuan, and X. Wu, "Adaptive image enhancement method for correcting low-illumination images," *Information Sciences*, vol. 496, pp. 25–41, 2019.

[28] X. Fu, Y. Liao, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4965–4977, 2015.

[29] 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.

[30] D. J. Jobson, Z.-u. Rahman, and G. A. 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.

[31] X. Fu, Y. Liao, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4965–4977, 2015.

[32] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Processing*, vol. 129, pp. 82–96, 2016.

[33] B. Cai, X. Xu, K. Guo, K. Jia, B. Hu, and D. Tao, "A joint intrinsic-extrinsic prior model for retinex," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4000–4009.

[34] 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.

[35] Z. Rahman, M. Aamir, Y.-F. Pu, F. Ullah, and Q. Dai, "A smart system for low-light image enhancement with color constancy and detail manipulation in complex light environments," *Symmetry*, vol. 10, no. 12, p. 718, 2018.

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

[37] 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, 2016.

[38] Y. Wang, R. Wan, W. Yang, H. Li, L.-P. Chau, and A. Kot, "Low-light image enhancement with normalizing flow," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, no. 3, 2022, pp. 2604–2612.

[39] R. Wan, B. Shi, W. Yang, B. Wen, L.-Y. Duan, and A. C. Kot, "Purifying low-light images via near-infrared enlightened image," *IEEE Transactions on Multimedia*, vol. 25, pp. 8006–8019, 2022.

[40] L. Guo, R. Wan, W. Yang, A. C. Kot, and B. Wen, "Cross-image disentanglement for low-light enhancement in real world," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 4, pp. 2550–2563, 2024.

[41] Z. Li, Y. Wang, and J. Zhang, "Low-light image enhancement with knowledge distillation," *Neurocomputing*, vol. 518, pp. 332–343, 2023.

[42] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "Enlightengan: Deep light enhancement without paired supervision," *IEEE transactions on image processing*, vol. 30, pp. 2340–2349, 2021.

[43] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 1780–1789.

[44] 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, 2023.

[45] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 22 252–22 261.

[46] D. Meng, Q. Zhao, and L. Jiang, "A theoretical understanding of self-paced learning," *Information Sciences*, vol. 414, pp. 319–328, 2017.

[47] Y. Du, J. Deng, Y. Zheng, J. Dong, and S. He, "Dsdnet: Toward single image deraining with self-paced curricular dual stimulations," *Computer Vision and Image Understanding*, vol. 230, p. 103657, 2023.

[48] L. Jiang, D. Meng, S.-I. Yu, Z. Lan, S. Shan, and A. Hauptmann, "Self-paced learning with diversity," *Advances in neural information processing systems*, vol. 27, 2014.

[49] M. Kumar, B. Packer, and D. Koller, "Self-paced learning for latent variable models," *Advances in neural information processing systems*, vol. 23, 2010.

[50] Y. Guo, Y. Gao, W. Liu, Y. Lu, J. Qu, S. He, and W. Ren, "Scanet: Self-paced semi-curricular attention network for non-homogeneous image dehazing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 1884–1893.

[51] J. Zhang, J. Song, L. Gao, Y. Liu, and H. T. Shen, "Progressive meta-learning with curriculum," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 9, pp. 5916–5930, 2022.

[52] J. Zhang, J. Song, Y. Yao, and L. Gao, "Curriculum-based meta-learning," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 1838–1846.

[53] Y. Dai, B. Chen, L. Gao, J. Song, and H. T. Shen, "Dmh-cl: Dynamic model hardness based curriculum learning for complex pose estimation," *IEEE Transactions on Multimedia*, vol. 26, pp. 3180–3193, 2024.

[54] E. H. Land, "The retinex," in *Ciba Foundation Symposium-Colour Vision: Physiology and Experimental Psychology*. Wiley Online Library, 1965, pp. 217–227.

[55] 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.

[56] J. Xu, Y. Huang, M.-M. Cheng, L. Liu, F. Zhu, Z. Xu, and L. Shao, "Noisy-as-clean: Learning self-supervised denoising from corrupted image," *IEEE Transactions on Image Processing*, vol. 29, pp. 9316–9329, 2020.

[57] A. Vaswani, P. Ramachandran, A. Srinivas, N. Parmar, B. Hechtman, and J. Shlens, "Scaling local self-attention for parameter efficient visual backbones," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 12 894–12 904.

[58] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2782–2790.

[59] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10 561–10 570.

[60] 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.

[61] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 2049–2062, 2018.

[62] 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.

[63] 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.

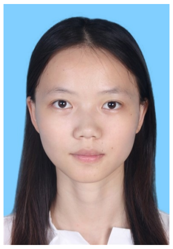
[64] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation of 2d histograms," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5372–5384, 2013.

[65] 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.

[66] 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.

[67] M. Chaverot, M. Carré, M. Jourlin, A. Bensrhair, and R. Grisel, "Improvement of small objects detection in thermal images," *Integrated Computer-Aided Engineering*, no. Preprint, pp. 1–15, 2023.

[68] W. Tang, F. He, and Y. Liu, "Ydtr: Infrared and visible image fusion via y-shape dynamic transformer," *IEEE Transactions on Multimedia*, vol. 25, pp. 5413–5428, 2022.



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.

Xuanrong Chen received the B.S. degree in software engineering from Guangdong University of Technology in 2021. 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.

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.

Chao Huang received the Ph.D. degree from the University of Notre Dame, Notre Dame, IN, USA, in 2019.

He is a tenure-track Assistant Professor with the Department of Computer Science and the Musketeers Foundation Institute of Data Science, The University of Hong Kong, Hong Kong. His research focuses on applied machine learning, graph neural networks, recommendation, and spatialtemporal data mining.

Wei Zhou (Senior Member, IEEE) is an Assistant Professor at Cardiff University, United Kingdom. Dr Zhou was a Postdoctoral Fellow at University of Waterloo, Canada. Wei received the Ph.D. degree from the University of Science and Technology of China in 2021, joint with the University of Waterloo from 2019 to 2021. Wei was a visiting professor at Dalian University of Technology, visiting scholar at National Institute of Informatics, Japan, a research assistant with Intel, and a research intern at Microsoft Research and Alibaba Cloud. Weis research interests span multimedia computing, perceptual image processing, and computational vision.

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.

From September 2017 to January 2019, he was a Joint Ph.D. Student with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. 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.