

# Seed Optimization with Frozen Generator for Superior Zero-shot Low-light Image Enhancement

Yuxuan Gu, Yi Jin, *Member, IEEE*, Ben Wang, Zhixiang Wei, Xiaoxiao Ma, Haoxuan Wang, Pengyang Ling, Huaian Chen, and Enhong Chen, *Fellow, IEEE*

**Abstract**—In this work, we observe that the generators, which are pre-trained on massive natural images, inherently hold the promising potential for superior low-light image enhancement against varying scenarios. Specifically, for the low-light image enhancement process of a single image, we introduce the pre-trained generators to restore the details and colors degraded by low-light conditions, thereby improving the visual effect. Taking one step further, we introduce a novel optimization strategy, which backpropagates the gradients to the input seeds rather than the parameters of the low-light image enhancement model, thus intactly retaining the generative knowledge learned from natural images and achieving faster convergence speed. Benefiting from the pre-trained knowledge and seed-optimization strategy, the low-light image enhancement model can significantly regularize the visibility and fidelity of the enhanced result, thus rapidly generating high-quality images without training on any low-light dataset. Extensive experiments on various benchmarks demonstrate the effectiveness of the proposed method, showing its potential advantages over numerous state-of-the-art methods both qualitatively and quantitatively.

**Index Terms**—Low-light image enhancement, retinex decomposition model, generative priors.

## I. INTRODUCTION

Capturing high-quality images under poor lighting conditions is an extremely challenging task since the captured image easily suffers severe visual degradation, such as poor illumination and color distortion. Such degradation significantly affects the performance of the downstream tasks [1]–[3]. To address this problem, low-light image enhancement (LIE) is about improving the visual quality of a low-light image to have the one with better visibility. Such a technique not only demonstrates its practical value in digital photography, but also

Manuscript received 7 May 2024, revised 11 August 2024; accepted 28 August 2024. This work was supported in part by National Natural Science Foundation of China under Grant 62401532, in part by Anhui Provincial Key Research and Development Plan under Grant 202304a05020072, in part by Anhui provincial Natural Science Foundation under Grant 23080850F226, in part by the China Postdoctoral Science Foundation under Grant 2022M720137 and in part by the Fundamental Research Funds for the Central Universities under Grant WK2090000065. This article was recommended by Associate Editor R. Ionescu. (*Yuxuan Gu and Yi Jin contributed equally to this work.*) (*Corresponding authors:* Yi Jin; Huaian Chen.)

Yuxuan Gu, Yi Jin, Ben Wang, Zhixiang Wei, Xiaoxiao Ma, Haoxuan Wang, Pengyang Ling, and Huaian Chen are with the School of Engineering Science, University of Science and Technology of China, Hefei, Anhui 230022, China (e-mail: jinyi08@ustc.edu.cn; anchen@ustc.edu.cn).

Enhong Chen is with the Department of Computer Science, University of Science and Technology of China, Hefei, Anhui 230022, China.

Copyright © 2024 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org).

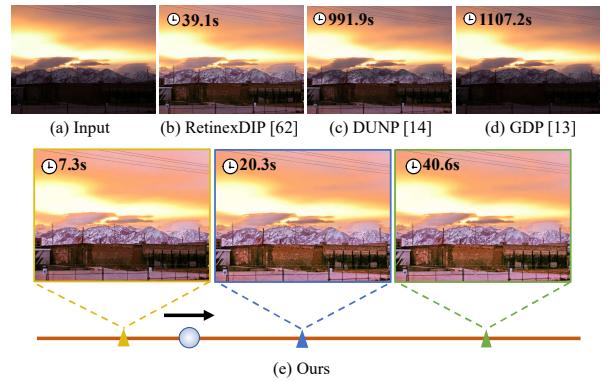


Fig. 1. Visual Comparisons on a real low-light image. As iterative optimization proceeds, the method generates brighter and more visually appealing results within a shorter time compared to two other leading zero-shot low-light image enhancement approaches. The image outlined in blue represents the output using our recommended iteration number.

benefits many downstream computer vision applications such as surveillance and tracking in low-light conditions.

In previous studies, various robust frameworks for low-light image enhancement have been explored, including Histogram equalization [53]–[55], inverse domain operation [56], [57], Retinex decomposition [50]–[52]. Methods based on Histogram equalization (HE) address low-light image enhancement issue by flattening the histogram and extending the dynamic range, which can produce high-contrast images. Inverse domain operation [56], [57] utilizes the inverse connection between the images with insufficient illumination and those in hazy environments. Although these methods may produce promising results in some cases, they still suffer from the limitation in designing well-working constraints for various scenes. The Retinex decomposition model (RDM), which assumes that an image can be deconstructed into the element-wise product of a reflectance layer and an illumination layer, has emerged as a significant framework for the development of advanced low-light image enhancement techniques [50]–[52]. However, the effectiveness of existing RDM-based LIE methods is substantially dependent on manually designed priors. These priors may not provide an accurate representation of the reflectance and illumination layers in real-world images, thereby posing challenges to the fidelity and applicability of these enhancement techniques.

More recently, with the advancement of deep learning technology in the field of computer vision, recent works [4]–[7] have applied supervised learning to achieve low-light image

enhancement, in which the model is trained with an elaborately collected dataset containing enormous low-light/normal-light image pairs. These methods have achieved remarkable performance, but since the generalization ability of deep learning models heavily depends on the amount of data, they have to collect a large number of paired images. Such a data collection operation is tedious and time-consuming.

To eliminate the requirements of low/normal-light image pairs, many unsupervised learning approaches have been proposed [8]–[12], [84]–[89]. These methods leverage the non-reference low-light images or unpaired low/normal-image images for training, and thus greatly reduce the efforts of data collection. As typical examples, EnlightenGAN [8] leverages the unpaired images with a generative adversarial network, and Zero-DCE [10] estimates pixel-wise high-order curves with a set of non-reference losses. These methods can effectively brighten the low-light images. However, the performance of these methods highly depends on the data distribution between the training samples and testing samples. The performance may suffer a dramatic decline when the training and testing samples have a large discrepancy.

Subsequently, some methods [13], [14] have been devoted to exploring a zero-shot LIE approach that can be trained with a low-light image itself. Representative, the pioneering work GDP [13] utilizes the generative diffusion prior for model training, which achieves more competitive visual quality. However, GDP’s performance is hindered by its departure from the robust framework leads to a reliance on a simplistic degradation model, limiting its effectiveness in complex scenarios. In contrast, instead of using pre-trained model, Liang *et al.* [14] applies the deep image prior (DIP) [15] to LIE tasks, abbreviated as DUNP, which successfully brightens the low-light images. However, DUNP utilizes randomly initialized weights, relying solely on information from a single low-light image. Consequently, it struggles to accurately decompose images, especially in extremely low-light cases.

In this work, we observe that the generators, which are pre-trained on massive natural images, inherently hold the promising potential for superior low-light image enhancement against varying scenarios. Based on this observation, we propose a novel zero-shot LIE perspective, *i.e.*, strengthening Retinex decomposition with a well-trained generator, to leverage abundant pre-trained knowledge for low-light image enhancement from a single image. Such a perspective takes advantage of high-quality structure and texture priors learned by a well-trained deep generative model to enhance the realness and fidelity of the enhanced image. Starting from this perspective, we design a zero-shot LIE model based on the Retinex theory, which rapidly decomposes an image into a reflectance map and an illumination map. Instead of fine-tuning the parameters of the generative model, we backpropagate the gradients calculated from the loss functions to the input seeds, thus intactly retaining the deep generative knowledge learned from large-scale training samples. Our method iteratively optimizes on only one low-light image, then eliminating the need for low-light datasets at all. In summary, the contributions of this work are listed as follows:

- We propose a new perspective for LIE tasks, *i.e.*,

strengthening Retinex decomposition with a well-trained generator, which exhaustively leverages pre-trained knowledge to effectively address the challenges of insufficient information and severe feature degradation in single-image low-light conditions.

- We design a zero-shot LIE framework by embedding seed-optimization into a Retinex-based enhancement framework, which achieves faster convergence speed and superior visual results. To the best of our knowledge, this is the first learning-based LIE approach that does not need to optimize the parameters of the enhancement model.
- We demonstrate the superiority of the proposed method through extensive experiments. With only limited time (<10s), the proposed method can even achieve superior performance against state-of-the-art methods trained on large-scale LIE datasets.

## II. RELATED WORK

### A. Conventional Low-light Image Enhancement

In early works, low-light image enhancement techniques have primarily focused on adjusting image intensity to augment contrast. Conventional low-light image enhancement methods [46], [48]–[57] can generally be categorized into Histogram equalization, inverse domain operation and Retinex decomposition model. Histogram equalization [46], [53]–[55], aims to modify pixel values to conform to a specified distribution, further contributing to the suite of intensity and color adjustment methods. Some researchers [56], [57] noticed the similarity between haze images and the inverted low-light images. Thus, the inverse domain-based methods applied dehazing methods to enhance low-light images. However, these contrast enhancement approaches often lead to visuals that appear unnaturally altered.

In recent years, the Retinex decomposition model that assumes an image could be decomposed as the element-wise product of a reflectance layer and an illumination layer has been one prominent choice for developing powerful LIE techniques [48]–[52]. The performance of existing RDM-based LIE methods heavily relies on hand-crafted priors which might be inaccurate for characterizing the reflectance and illumination layers on real-world images. A commonly employed prior for the illumination layer is local smoothness [48], which is often achieved through various techniques including bilateral filtering, the application of an L2-penalty on gradients, and a weighted L1-penalty on gradients. In comparison to illumination, reflectance is more challenging to characterize. A widely-used assumption is that a reflectance layer contains fine textures [52], and thus the piece-wise continuity prior is often used for reflectance layers in existing work, *e.g.*, [58]–[60]. Note that the texture of a reflectance map is easily confused with noise, which makes generating a reflectance map more challenging.

### B. Supervised Low-light Image Enhancement

A series of supervised learning-based methods such as [4], [16], [19], [20], have achieved impressive results. To establish the mapping from low-light image to normal-light image,

most of these methods involve supervised learning strategy, thus posing the serious reliance on the datasets containing numerous low-light and normal-light samples. As an early supervised learning endeavor, RetinexNet [16] explores end-to-end Retinex decomposition. KinD [20] and KinD++ [19] adopt a trainable denoising module for reflectance restoration, further exploring the potential of CNN. To obtain better utilization of training data, advanced supervised methods such as LLflow [64], SNR [5] and Retinexformer [4] explore various model architectures and training paradigms. LLflow [64] introduces the Normalizing flow to effectively learn the local pixel correlation and global image characteristics of the image by mapping the distribution of normally exposed images into a Gaussian distribution. SNR [5] proposes a Signal-to-Noise-Ratio-aware network by combining CNN and Transformer. Retinexformer [4] proposes an attention mechanism based on illumination to fully explore the potential of Transformer. Although the above methods significantly improve the model performance, the acquisition cost of paired images required by these methods is extremely high in real-world scenarios. Such constraints are reflected in the poor generalization capability of supervised methods.

### C. Unsupervised Low-light Image Enhancement

To reduce the requirement for paired datasets, many efforts [8], [10]–[12], [17], [18], [41], [42], [84]–[89] have been directed toward zero-shot learning in recent years. Some research use adversarial training methods to convert low-light images into normal-light images. Particularly, EnLightenGAN [8] employs a global-local discriminator structure to effectively utilize unpaired data as both positive and negative samples with the assistance of self-regularized attention. Following the path of EnlightenGAN, UEGAN [17] uses the unidirectional GAN framework and quality evaluation function to alleviate the Over-enhancement problem. Taking one step further, NeRCo [11], CLIP-LIT [12] introduces multi-modal adversarial learning to low-light image enhancement, resulting in better perceptual results. These methods significantly alleviate the need for paired training samples, but they still need large-scale training samples highly correlated with application scenarios in terms of both low-light and normal-light conditions, which is still costly in practice. To achieve non-reference low-light image enhancement, many methods apply diverse frameworks and training strategies. RUAS [42] build a Retinex-inspired unrolling framework with architecture search. SCI [41] proposes a Self-Calibrated Illumination learning framework for fast, flexible and robust low-light image enhancement. Methods such as Zero-DCE [10] propose learning parameterized S-curves and spatial consistency in the loss function to improve the convergence of the S-curve parameters. Then Zero-DCE++ [18] is proposed, which uses iterative approximations of a gamma function through quadratic polynomials. However, these approaches still necessitate datasets with varied illumination for Optimization, which makes their performance still limited by the dataset distribution.

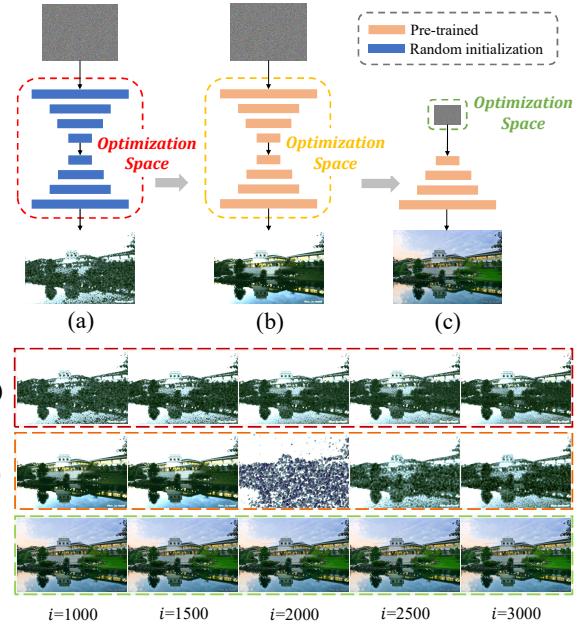


Fig. 2. Retinex decomposition results under different optimization settings. The upper part of the Figure shows the mode of generating reflectance and the lower part shows the results of iterative optimization. (d), (e), (f) are results produced by mode (a), (b), (c), respectively. The white right arrow indicates a strategy shift, and the model used is VQ-VAE-2. The seed-optimization strategy shows better performance in terms of quality.

### D. Low-light Enhancement from Single Image

Subsequently, other methods [13], [14], [62] use only one low-light image for enhancement. GDP [13] leverages a degradation model and pre-trained Denoising Diffusion Probabilistic Model (DDPM) [21] to directly obtain enhanced images. Although GDP attempts to introduce pre-trained models to mitigate optimization difficulties, its performance still heavily relies on the manually defined degradation model and extensive sampling time. Inspired by the concept of leveraging model structure prior, known as Deep Image Prior [15], approaches like RetinexDIP [62], DUNP [14] aim to thoroughly obviate the need for training data. These methods decompose the illumination and reflection components from a single image and subsequently enhance it based on the enhancement results. Nevertheless, DUNP and RetinexDIP employ randomly initialized parameters, which results in a diminished correlation with real-world variance. In summary, while existing zero-shot learning methods alleviate the dependency on training data, their computational efficiency and perception of natural image characteristics remain unsatisfactory.

## III. METHOD

### A. Motivation

Generative models have achieved remarkable success in computer vision tasks. These models, adept at encapsulating the intricate distributions of natural image data due to their extensive training on vast image datasets, show promising versatility for various applications [29], [32].

To mine the knowledge about natural images contained in the pre-trained generator and be able to assist in recovering low-light image features, we tried various ways of embedding

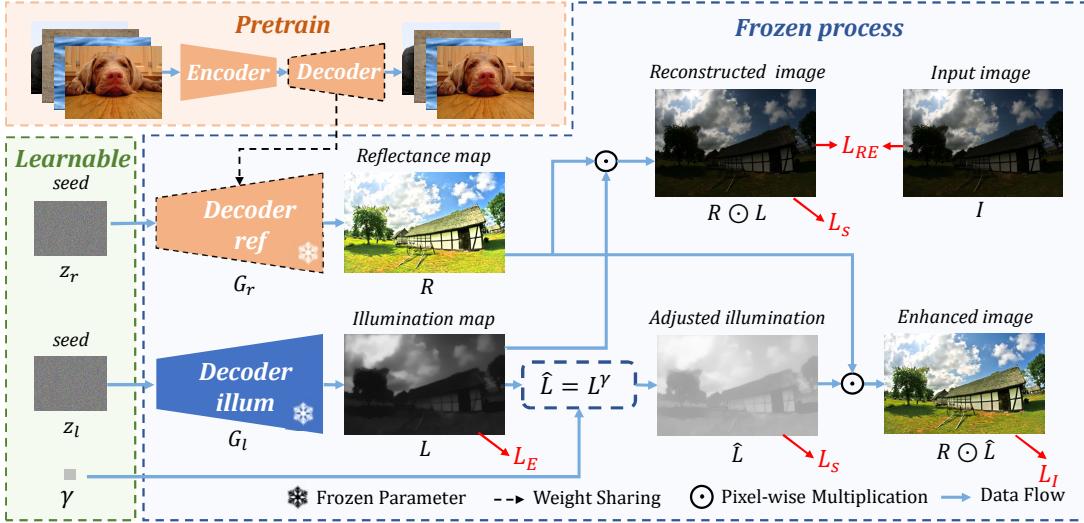


Fig. 3. Overview of the Proposed Low-Light Image Enhancement Method. A conventionally trained image generator is utilized as the reflectance decoder in the preparation phase, while the illumination decoder is initialized randomly. Various image generation models can be applied to the proposed method; in practice, we utilize VQ-VAE-2 as the pre-trained model. For an input low-light image, our design involves three optimizable seeds:  $z_r$ ,  $z_l$ , and  $\gamma$ .  $z_r$  and  $z_l$  are dedicated to generating a reflectance map rich in detail and a comparatively smoother illumination map, respectively. The input image  $I$  is used to supervise the retinex reconstruction results of the reflection map and illumination map. Concurrently,  $\gamma$  is employed for gamma correction of the image. The final enhancement result of the low-light image is achieved solely through iterative optimization of these three inputs.

them in low-light image enhancement tasks. An intuitive way to leverage this generative knowledge is by taking a well-trained generator as the fundamental architecture and then fine-tuning it for image enhancement. However, such fine-tuning can disrupt the intricate generative knowledge cultivated on datasets of high-quality images. Our experiments, as illustrated in Figure 2 (a) and (b), reveal that although initializing with pre-trained weights enhances the decomposition quality in terms of detail and color, it falls short of mirroring the performance attainable in image generation domains.

Recognizing the limitations of conventional fine-tuning, we shift our strategy towards optimizing the input to the model rather than its parameters. As demonstrated in Figure 2 (c) and (f), the proposed methodology, which we refer to as the seed-optimization strategy, significantly outperforms standard fine-tuning techniques in image enhancement tasks. Motivated by this, we introduce a novel Retinex-based Low-light Image Enhancement (LIE) model grounded in the double-DIP framework. Contrary to traditional backpropagation methods that adjust model parameters, our approach optimizes the input seeds, thereby preserving the integrity of the generative knowledge acquired from well-lit image scenarios. This seed-centric optimization, guided by generative knowledge, facilitates the fast and superior enhancement of low-light images.

#### B. Retinex Model with Pre-trained Knowledge

Given an input low-light image  $I$ , we apply Retinex theory to decompose it into a reflectance map  $R$ , and an illumination map  $L$ . This decomposition can be formally expressed as:

$$I = R \odot L, \quad (1)$$

where  $\odot$  represents element-wise multiplication and  $R$  denotes the reflectance map and  $L$  denotes the illumination map. Cor-

respondingly, the ideal high-light image  $\hat{I}$ , can be represented as:

$$\hat{I} = \hat{R} \odot \hat{L}. \quad (2)$$

$\hat{R}$  and  $\hat{L}$  represent the reflection map and illumination map of the normal lighting map. Based on the framework above, we use two generators, denoted as  $G_r$  and  $G_l$ , to generate the components  $L$  and  $R$  respectively, which can be formulated as:

$$I = G_r(z_r, \theta_r) \odot G_l(z_l, \theta_l), \quad (3)$$

where  $z_r$  and  $z_l$  denote the random input seed, while  $\theta_r$  and  $\theta_l$  denote the weights of model. The reflectance map contains abundant structure and texture details of the image, which follows the prior learned by the generative model. We employ a generative model with pre-trained weights  $\tilde{\theta}_r$ , which aims to effectively leverage generative knowledge from high-quality images to reconstruct the reflectance under varied lighting conditions. In contrast, the illumination map, primarily representing ambient light, deviates significantly from these high-quality image priors due to its inherent low-light characteristics, which has been extensively discussed and verified in [77]–[82]. In addition, through motivation and subsequent experiments, we found that different initialization methods have different properties, which we will further discuss in section IV. Therefore, we utilize a generator initialized with random weights, denoted as  $\hat{\theta}_l$ , to generate the illumination map. Now the decomposition process is defined as:

$$I = G_r(z_r, \tilde{\theta}_r) \odot G_l(z_l, \hat{\theta}_l), \quad (4)$$

Having obtained the reflectance and illumination maps, we further apply a gamma transformation to effectively control the brightness of the enhanced result, bringing  $L$  closer to  $\hat{L}$ . Unlike previous works that use specialized modules for adjusting gamma values, we integrate the gamma transformation as a learnable parameter within our enhancement framework.

Denoting the gamma transformation factor as  $\gamma$ . The result of gamma transformation is denoted as  $G_l(z_l, \hat{\theta}_l)^\gamma$ , and the final enhanced result is defined as:

$$\hat{I} = G_r(z_r, \tilde{\theta}_r) \odot G_l(z_l, \hat{\theta}_l)^\gamma. \quad (5)$$

This approach effectively incorporates generative knowledge into the Retinex model. Our next task is to define the optimization process for it.

### C. Prior-Constrained Seed Optimization

According to equation 5, for convenience, we simplify the expression of our model as:

$$\hat{I} = G(z, \theta), \quad (6)$$

where  $G$ ,  $z$ , and  $\theta$  denote all the generators, random seeds, and parameters of the generator respectively. The objective of our model is to minimize the discrepancy between the generated image and the target normal-light image using the loss function. Here,  $G$ , as well as all instances of  $G$  not specifically indicated otherwise, utilize the decoder of VQ-VAE-2 as the generator. In practice, we found that both GAN-based and VAE-based generators can be used in our method. However, GAN models generally have a larger number of parameters, whereas VAE models are smaller and more convenient to use. Among them, VQ-VAE-2 exhibits the best efficiency and performance. We will discuss the characteristics of various pre-trained models in detail in Section IV. For an  $M \times N$  low-light image,  $z$  is of size  $M/2^n \times N/2^n$ , where  $n$  represents the number of upsampling layers in the decoder. To reduce the difficulty of fitting, we use the mean and variance of a normal distribution derived from a large number of latent codes encoded from natural images to initialize the Gaussian distribution of  $z$ . To completely retain the deep generative priors learned from large-scale training samples, instead of fine-tuning the parameters of the generative model, we turn to optimize the inputs, which are represented as:

$$\min_z \mathcal{L}_{all}(G(z, \theta), I). \quad (7)$$

In particular, we backpropagate the gradients calculated from the loss functions to the input seeds as:

$$z_{\text{new}} = z - lr \times \nabla z \mathcal{L}_{all}(G(z, \theta), I), \quad (8)$$

where  $lr$  is the learning rate and  $\nabla z$  is the derivative function. Based on this seed optimization strategy, we can leverage the generative knowledge to optimize the random seeds. By integrating this process with the Retinex framework, we formulate the seed optimization objective as follows:

$$\begin{aligned} \arg \min_{z_r, z_l, \gamma} & \lambda_1 \|G_r(z_r, \tilde{\theta}_r) G_l(z_l, \hat{\theta}_l) - I\| + \\ & \lambda_2 \|G_r(z_r, \tilde{\theta}_r)\| + \\ & \lambda_3 \|G_l(z_l, \hat{\theta}_l)\| + \lambda_4 \|\hat{I}\|. \end{aligned} \quad (9)$$

Next, we will detail the specific implementation of the regularization terms used in this process.

### D. Loss Functions

To optimize the seeds, We use a series of concise and classic regularization terms to optimize the model. The overall loss function can be defined as:

$$\mathcal{L}_{all} = \lambda_{RE} \mathcal{L}_{RE} + \lambda_E \mathcal{L}_E + \lambda_S \mathcal{L}_S + \lambda_I \mathcal{L}_I, \quad (10)$$

Among them,  $\mathcal{L}_{RE}$  measures the reconstruction error,  $\mathcal{L}_E$  is used to regularize the intensity distribution,  $\mathcal{L}_S$  is used to regularize the highly ill-posed Retinex framework, and  $\mathcal{L}_I$  is used to enhance low-light images to a sufficient brightness, while  $\lambda_{RE}$ ,  $\lambda_E$ ,  $\lambda_S$  and  $\lambda_I$  is the balance factor.

**Reconstruction Loss.** To achieve accurate Retinex decomposition, the reconstruction loss is defined as:

$$\mathcal{L}_{RE} = \|I - G_r(z_r, \tilde{\theta}_r) \odot G_l(z_l, \hat{\theta}_l)\|^2. \quad (11)$$

The reconstruction loss facilitates the standard Retinex decomposition, corresponding to the first regularization term in equation 9.

**Illumination-consistency Loss.** According to previous studies [77], [78], Retinex theory necessitates a relatively smooth illumination map. To constrain the general structure of the illumination map, we define the illumination regularization loss as:

$$\mathcal{L}_E = \left\| \max_{c \in \{R, G, B\}} gauss(I)^c - G_l(z_l, \hat{\theta}_l) \right\|, \quad (12)$$

where  $gauss$  donates Gaussian blur with kernel size=25 and  $\sigma = 2.0$  and  $c$  donates the channel of the input image. The setting and configurations of Gaussian blur are designed to obtain the smooth illumination map, following the theory and settings in [77]–[82], corresponding to the third regularization term in equation 9.

**Smoothness Loss.** To regularize the highly ill-posed Retinex framework, we introduce a smoothness loss, which is defined to constrain the smoothness of both the illumination and reflectance maps, thereby avoiding inaccuracies in Retinex decomposition caused by noise and unnecessary textures. Therefore, we employ TV loss and a parameter  $\tau$  to control its influence. This can be expressed as:

$$\mathcal{L}_S^r = \|\nabla G_r(z_r, \tilde{\theta}_r)\|, \quad (13)$$

For the illumination map, we assume that illumination changes gradually across the surface of an object but varies significantly between different objects. In practice, we impose a higher variation at image edges and stricter smoothness constraints elsewhere. After enhancing loss stability using an exponential function, this constraint can be expressed as:

$$\mathcal{L}_S^l = \left\| \frac{\nabla G_l(z_l, \hat{\theta}_l)}{\exp \|\nabla G_r(z_r, \tilde{\theta}_r)\|} \right\|, \quad (14)$$

By combining these constraints, we obtain the final smoothness loss, corresponding to the second and third regularization term in equation 9, where  $\tau$  denotes the balance factor.

$$\mathcal{L}_S = \mathcal{L}_S^l + \tau \mathcal{L}_S^r \quad (15)$$

TABLE I  
PSNR/SSIM/NIQE/MUSIQ/NoRVDNet $\pm$  SCORES ON FIVE DATASETS. THE BEST AND SECOND-BEST ZERO-SHOT METHODS ARE HIGHLIGHTED IN RED AND BLUE RESPECTIVELY. THE OVERALL 'RANK' IS CALCULATED BY AVERAGING THE PER-DATASET AVERAGE RANKINGS AMONG EVERY METHOD, WHILE 'ROR' DENOTES THE RANK OF THE OVERALL RANK.

Method	LOL					MEF			LIME			DICM			NPE			Rank	RoR	
	PSNR	SSIM	NIQE	MUSIQ	N+	NIQE	MUSIQ	N+	NIQE	MUSIQ	N+	NIQE	MUSIQ	N+	NIQE	MUSIQ	N+			
No-Learning	STAR [49]	12.64	0.62	7.40	54.92	55.28	3.63	63.43	65.26	3.73	62.44	62.31	3.73	60.09	67.35	3.82	67.95	61.40	10.17	6
	LR3M [48]	10.22	0.46	7.57	46.19	57.96	5.48	63.19	61.85	5.20	64.71	61.77	4.57	62.70	60.75	4.90	70.69	58.54	14.22	19
	LIME [51]	14.02	0.62	8.09	56.55	50.37	3.42	64.48	59.68	3.85	60.66	57.04	3.76	60.75	59.94	4.27	67.39	56.41	12.89	15
Paired Training	Retinex-Net [16]	16.77	0.42	8.87	57.26	47.36	4.40	66.71	59.49	4.60	62.90	56.64	4.43	63.11	62.17	4.47	65.16	57.37	13.89	18
	KinD++ [19]	21.80	0.83	5.12	70.54	67.29	3.76	66.11	71.50	4.73	64.00	68.98	3.78	62.81	71.04	3.76	65.53	64.17	5.44	2
	SNR [5]	24.61	0.84	5.17	59.79	72.55	4.15	57.49	70.72	5.69	52.52	69.28	4.64	50.19	70.97	5.70	50.67	69.07	10.56	8
	LLFlow-SKF [61]	26.37	0.93	5.55	74.04	63.25	4.80	51.50	65.24	4.93	49.74	63.90	4.21	56.35	65.99	4.54	64.45	70.96	10.33	7
Unpaired Training	Retinexformer [4]	25.15	0.90	3.47	63.15	55.41	3.44	62.16	67.23	3.86	59.45	64.58	3.93	58.32	70.47	3.97	66.22	62.72	7.39	4
	EnGan [8]	17.48	0.65	4.68	56.60	55.69	3.20	63.03	66.49	3.66	59.07	63.26	3.56	57.03	66.32	3.70	59.41	58.00	8.89	5
	SCL-LLE [43]	12.42	0.52	7.63	56.39	53.06	3.28	65.01	66.56	3.77	61.45	61.77	3.57	60.43	60.57	3.59	63.75	57.06	11.00	10
Zero Referenced	NeRCo [11]	19.74	0.80	3.40	69.87	67.74	3.76	61.91	67.84	3.59	61.24	68.92	3.80	64.08	70.46	3.69	65.55	71.46	5.61	3
	RUAS [42]	16.40	0.50	6.34	59.34	56.89	5.47	55.53	60.73	5.37	55.43	59.70	7.20	50.41	60.74	7.16	50.78	55.63	16.28	21
	SCI [41]	14.78	0.52	7.87	57.12	52.42	3.62	62.59	61.41	4.18	59.11	58.80	4.13	53.90	63.08	4.02	56.94	57.37	14.89	20
	SGZ [39]	15.93	0.57	7.82	56.53	52.65	3.34	63.87	65.03	3.91	60.41	61.62	3.56	57.42	61.38	3.52	60.82	58.31	11.39	12
	ZeroDCE++ [10]	14.86	0.56	7.77	55.85	52.88	3.32	65.13	65.76	3.79	60.73	61.21	3.57	59.31	61.16	3.65	63.13	57.81	10.89	9
	ZeroIG [75]	17.69	0.71	4.74	53.33	55.87	4.50	56.66	59.18	4.69	58.36	59.57	4.41	57.61	62.44	4.49	58.05	57.73	13.78	17
	NeurBR [76]	11.38	0.44	7.50	59.07	57.57	3.34	63.59	62.11	3.80	60.75	60.95	3.83	58.03	62.06	3.83	65.37	62.08	12.67	14
	GDP [13]	15.89	0.54	6.13	60.25	54.62	4.08	59.64	60.98	4.40	59.69	59.92	4.11	58.83	64.98	3.67	59.91	57.90	12.56	13
	RetinexDIP [62]	11.66	0.56	7.69	56.97	51.38	3.32	65.06	58.73	3.86	61.19	56.87	3.82	60.16	61.68	4.04	67.96	57.38	12.94	16
	DUNP [14]	13.21	0.46	4.27	67.06	61.27	3.10	65.08	60.42	4.12	56.90	61.22	3.65	56.37	60.60	3.81	68.03	65.49	11.28	11
<b>Ours</b>		<b>18.10</b>	<b>0.75</b>	<b>2.80</b>	<b>58.98</b>	<b>68.02</b>	<b>3.16</b>	<b>65.19</b>	<b>67.67</b>	<b>3.64</b>	<b>63.12</b>	<b>66.57</b>	<b>3.43</b>	<b>65.19</b>	<b>67.18</b>	<b>3.65</b>	<b>68.24</b>	<b>70.95</b>	<b>3.78</b>	<b>1</b>

**Illumination Control Loss.** The illumination loss encourages the overall illumination of the enhanced image closer to the desired illumination. To this end, we define the illumination control loss, corresponding to the last regularization term in equation 9. as:

$$\mathcal{L}_I = \left\| E - G_r(z_r, \tilde{\theta}_r) \odot G_l(z_l, \hat{\theta}_l)^\gamma \right\|, \quad (16)$$

where  $E$  is the well-exposedness level and is set to 0.6 in our experiments, following the setting in [10], [18].

See Algorithm 1 for the pseudo-code of the proposed method.

**Algorithm 1** Low-light image enhancement using seed optimization with frozen generators

**Input:** Low-light image  $I$ ; parameters  $\tau, \lambda_{RE}, \lambda_E, \lambda_S$ , and  $\lambda_I$ ; maximum iterations  $S$  for optimization of random seeds; pretrained generators  $\theta_r$  and randomly initialized  $\theta_l$ ; Seed initialization mean  $\mu$  standard variance  $\sigma$ .

**Output:** Normal-light image  $\hat{I}$

- 1: Sample  $\{z_l^{(s)}, z_r^{(s)}\}$  from  $\mathcal{N}(\mu, \sigma^2)$
- 2: Initialize  $\gamma$  to 0.5
- 3: **for**  $s = 0$  to  $S$  **do**
- 4:      $R^{(s)} \leftarrow G_r(z_r^{(s)}, \tilde{\theta}_r)$
- 5:      $L^{(s)} \leftarrow G_l(z_l^{(s)}, \hat{\theta}_l)$
- 6:      $I^{(s)} \leftarrow R^{(s)} \odot L^{(s)}$
- 7:      $\hat{I}^{(s)} \leftarrow R^{(s)} \odot L^{(s)\gamma}$
- 8:     Compute the gradients w.r.t  $z_l^{(s)}, z_r^{(s)}, \gamma$
- 9:     Update  $z_l^{(s)}, z_r^{(s)}, \gamma$  using the Adam
- 10: **end for**
- 11:  $\hat{I} \leftarrow \hat{I}^{(S)}$
- 12: **return**  $\hat{I}$

## IV. EXPERIMENTS

### A. Implementation Details

The proposed method is based on pytorch and runs with an NVIDIA TITAN RTX GPU and an Intel(R) Xeon(R) Gold 6252 CPU @ 2.10GHz. The Adam optimizer is set to do the train with learning rate of 1e-2. For the non-referenced datasets, each image undergoes an optimization process over 5000 iterations, and  $\tau$  is set to 0.2, while images from the LOL dataset are optimized for 2500 iterations, and  $\tau$  is set to 0.6. The learnable param gamma is initialized as 0.5, and  $\lambda_{RE}$ ,  $\lambda_E$ ,  $\lambda_S$ , and  $\lambda_I$  is set to be 12, 0.05, 0.03, and 0.01 respectively.

For the pre-trained model, we use the VQ-VAE-2 model with the Adam optimizer at learning rate of 3e-4 and train on the FFHQ [24] and ImageNet [44] dataset with batchsize of 128. In practice, we find that different types of generative models can achieve good results under their respective recommended data pipelines, which will be demonstrated in Sec IV-D.

### B. Datasets

We choose 5 referenced or non-reference datasets to comprehensively evaluate various methods in low-light image enhancement tasks: 1) LOL [16] included 15 indoor images in severely underexposed conditions. 2) NPE [34] contains 8 low-light nature images. 3) MEF [35] comprised 17 low-light images in different scenes. 4) LIME [33] featured 10 dark scene images of different resolutions. 5) DICM [36] contains 69 images from low-light and normal-light conditions. We use the most universal NIQE [38], two state-of-the-art metrics MUSIQ [37], NoRVDNet++ [83] to compare the performance of all methods on these datasets. It is noted that LOL is collected under extreme low-light conditions with normal light references provided, while the other four are

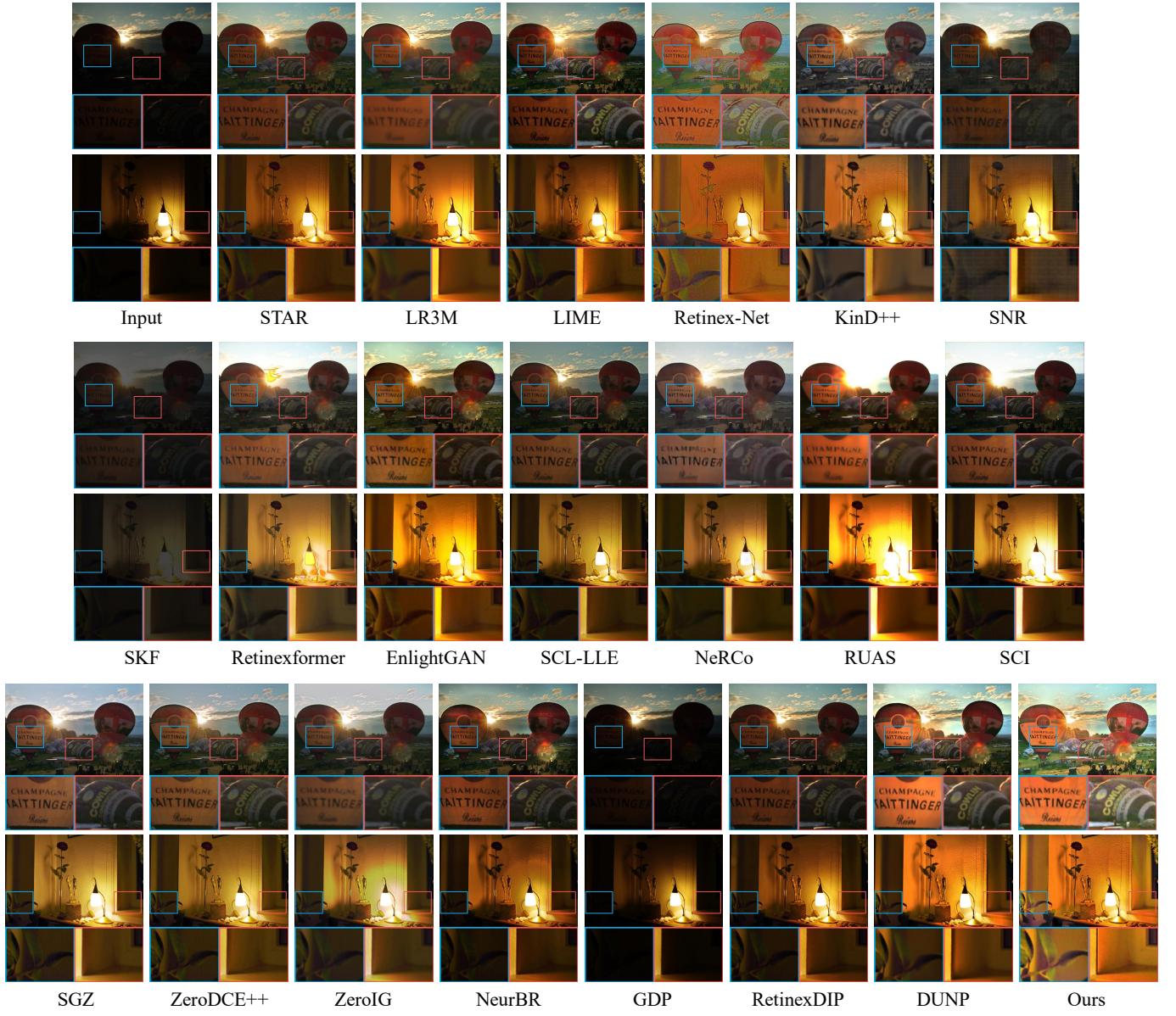


Fig. 4. Visual quality comparisons on MEF dataset with several low-light image enhancement state-of-the-art methods. Our approach demonstrates strong detail and color recovery capabilities.

collected under moderate low-light conditions without ground truths. For LOL datasets with paired images, we additionally use PSNR and SSIM [40] to evaluate the method based on reference images. In alignment with referenced studies [14], we employ non-parametric metrics to facilitate a more holistic and equitable comparison, incorporating both the average ranking (denoted by 'Rank') and the rank of rank (denoted by 'RoR'). Specifically, we first compute the average rankings of different metrics across each dataset and subsequently average these to determine the overall rank. The RoR is then defined as the rank assigned to this overall average rank.

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

For a more comprehensive analysis, we compare our method with four non-learning methods (i.e., LIME [51], STAR [49], and LR3M [48]), five advanced supervised learning methods (i.e., RetinexNet [16], KinD++ [19], SNR [5], SKF [61] and Retinexformer [4]), and nine unsupervised learning methods,

including RetinexDIP [62], DUNP [14], GDP [13], ZeroDCE++ [10], ZeroIG [75], NeurBR [76], SCL-LLE [43], RUAS [42], EnlightGAN [8], SGZ [39], NeRCo [11] and SCI [41]. We use the configuration recommended in the original paper for our experiments. For methods with multiple pre-trained weights, we use the weights that perform best under our ranking rule, so Retinexformer, SKF, and SNR use weights trained on LOL-v1.

1) *Quantitative Analysis.* We evaluated the performance of various methods using their official pre-trained models and publicly available code. As presented in Table I, the proposed method achieves the highest average ranking across all benchmarks on full-reference and no-reference metrics. When compared to recent prominent same-category techniques like DUNP and GDP, our method demonstrates distinct advantages. We incorporate more robust generative priors than DUNP, leveraging additional pre-trained knowledge. Furthermore, thanks to the well-established Retinex model,



Fig. 5. Visual quality comparisons on DICM dataset with several low-light image enhancement state-of-the-art methods. Our approach demonstrates strong detail and color recovery capabilities.

our method consistently outperforms GDP, especially on a series of challenging datasets. For a more intuitive visual comparison, we provide results from all methods, including DUNP, and GDP, in scenarios characterized by significant brightness degradation in the subsequent sections. Compared with other zero-shot methods (where RUAS, Zero-DCE, SCI, and SGZ use low-light datasets), our method achieves SOTA on 82% of the metrics, and even when compared with dataset-based methods achieved first overall ranking.

**2) Qualitative Analysis.** For a more visual understanding, we present the visual outcomes of all methodologies in Figure 4 and Figure 5. It can be seen that recent deep learning-based techniques often struggle to produce enhanced images with natural illumination. Specifically, the outputs from EnlightGAN [8], KinD++ [17], and DUNP [14] exhibit uneven illumination and poor details. There are noticeable over-exposures

in the results from RUAS [42] and SCI [41]. Both GDP [13] and SNR [5] fall short in effectively enhancing extremely dark regions. RetinexNet [16]'s output is characterized by a pronounced animated style with non-authentic delineations. Similarly, Zero-DCE++ [10] and others occasionally introduce unnatural noise and pronounced bright spots. By comparison, our model realizes the best visual quality with prominent contrast and vivid colors, while restoring intricate details.

**3) Computational Efficiency.** In addressing the computational efficiency of zero-shot low-light image enhancement methods, Figure 6 provides a comparative analysis of the runtime for several state-of-the-art techniques that do not require training on low-light datasets. These methods were evaluated under uniform conditions at a resolution of  $600 \times 400$  pixels, following their respective recommended configurations. The evaluation encompasses the performance metrics and

datasets previously detailed. Notably, as shown in Table III our proposed method, even when limited to 7.3 seconds, outperforms others in terms of accuracy and computational efficiency. Additionally, we analyzed the model's runtime speed at different resolutions. To ensure fairness, we removed the resize preprocessing from all methods. The comparative results with similar methods are shown in Table II. We observed that the proposed method outperforms others in terms of speed across various resolutions, further demonstrating the efficiency of the proposed approach.

TABLE II

COMPARISON OF PROCESSING TIME: THE AVERAGE TIME(GPU SECONDS) FOR DIFFERENT IMAGE SIZES. THE BEST RESULTS ARE MARKED IN BOLD. #OOM MEANS THAT THE TEST CANNOT BE PERFORMED BECAUSE THE GPU MEMORY CAPACITY IS EXCEEDED

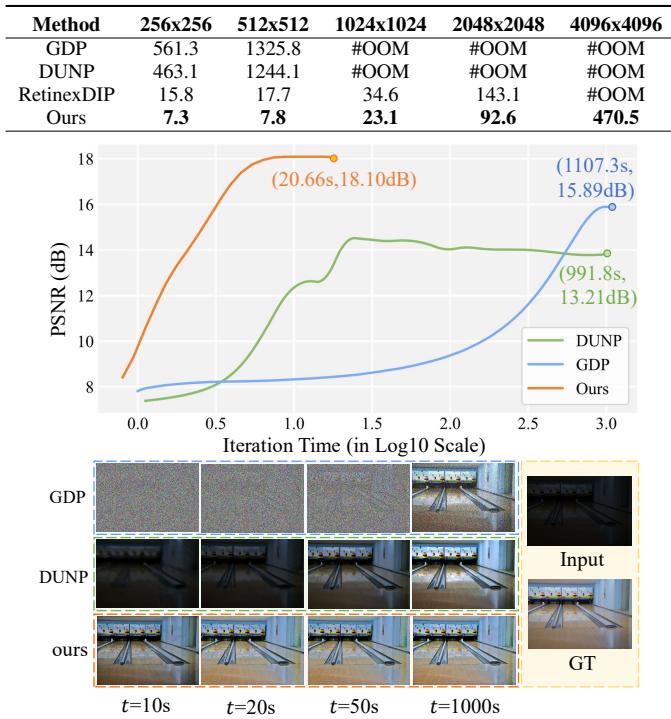


Fig. 6. Comparison of efficiency with other methods that can be trained with a low-light image itself. All results are obtained on the LOL dataset using the same computing resources.

TABLE III

COMPARISON OF PROCESSING TIME: THE AVERAGE TIME(GPU SECONDS) FOR ENHANCING ONE LOW-LIGHT IMAGE WITH THE SIZE OF  $600 \times 400$ . 'RANK' AND 'RoR' FOLLOW THE DEFINITION IN TABLE I. THE BEST RESULTS ARE MARKED IN BOLD.

Method	Total Time	Iteration Time	Rank	RoR
RetinexDIP	39.1	0.13052	12.94	16
GDP	1107.2	1.10702	12.56	13
DUNP	991.9	0.06611	11.28	11
Ours(900 iters)	<b>7.3</b>	<b>0.00813</b>	3.97	<b>1</b>
Ours(2500 iters)	20.3	<b>0.00813</b>	3.78	<b>1</b>
Ours(5000 iters)	40.6	<b>0.00813</b>	3.66	<b>1</b>
Ours(10000 iters)	81.3	<b>0.00813</b>	<b>3.58</b>	<b>1</b>

#### D. Ablation Studies

We conducted extensive ablation Studies to validate the efficacy of our proposed perspective and evaluate the contribution

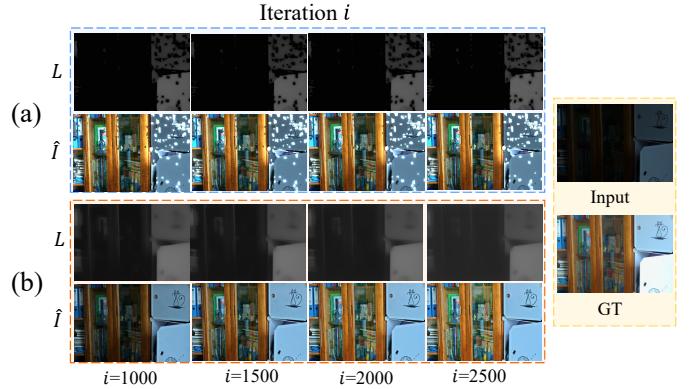


Fig. 7. Ablation Study of Pre-training. (a): Initialization of both the illumination decoder and reflectance decoder with weights from pre-trained models. (b): Random initialization of the illumination decoder.  $i$  is denoted as the number of iterations.

of each model component. All evaluations were carried out on the well-referenced LOL dataset.

1) *Effect of Pre-trained Weights.* We further delve into the impact of using pre-trained weights in our method by conducting experiments under three settings: (a) Load pre-trained models for both reflectance decoder and illumination decoder. (b) Load pre-trained models for reflectance decoder only. (b) is the setting used in the proposed method. Figure 7 graphically illustrates the outcomes for these settings. Setting (a) incorporates pre-trained information into the enhancement process, resulting in significant visual improvements. However, it does not consider the smoothness characteristic of the illumination map as outlined in Retinex theory. Consequently, the illumination map contains unnecessary details and becomes difficult to optimize. Our setting (b) effectively combines pre-trained knowledge with the generation of a more desirable illumination map, leading to the best enhancement results. The quantitative results in Table IV further indicate the consequential role that pre-trained weights play in the enhancement model's performance.

TABLE IV

QUANTITATIVE ANALYSIS FROM ABLATION STUDIES DEMONSTRATING THE IMPACT OF PRETRAINED WEIGHTS ON REFLECTANCE AND ILLUMINATION DECODER. ✓ MEANS LOADING PRE-TRAINED WEIGHT TO THIS GENERATOR. THE BEST RESULTS ARE MARKED IN BOLD.

Setting	$G_r$	$G_l$	PSNR↑	SSIM↑	NIQE↓	MUSIQ↑
(a)	✓	✓	16.83	0.65	3.35	58.42
(b)	✓		<b>18.10</b>	<b>0.75</b>	<b>2.80</b>	<b>58.98</b>

2) *Optimization Mode.* To evaluate the benefits of retaining the deep generative principles learned from normal light images, we conduct experiments with three settings: (a) optimizing only the gamma, reflectance decoder, and illumination decoder; (b) optimizing all the inputs, and decoders; (c) optimizing the parameters of illumination decoder and input of reflectance decoder. (d) is the setting used in the proposed method. Through experiments, as shown in Table VI and Figure 8, it can be seen that setting (a) is similar to traditional DIP methods in three metrics, and directly using pre-trained models does not significantly improve model performance; Compare to setting (a), setting (b) additionally optimizes the

TABLE V  
TRAINING DETAILS AND EFFECTS USED BY DIFFERENT PRETRAINED GENERATORS.

Method	DCGAN	EAGAN	VAE	VQVAE-2	VAE (SDv2.1)	DDPM
Dataset	Imagenet-1K	Imagenet-1K	COCO	Imagenet+FFHQ	LAION-5B	Imagenet-1K
PSNR↑	16.09	16.94	17.19	18.10	19.03	18.22
SSIM↓	0.69	0.74	0.72	0.75	0.78	0.73
Parameters (M)	3.58	11.55	1.63	0.36	49.49	147.01
Pretraining Hours	61.47	120.26	12.32	30.31	—	402.17
Pretraining Steps	$1.3 \times 10^6$	$1.3 \times 10^6$	$1.1 \times 10^5$	$1.4 \times 10^6$	—	$1.3 \times 10^5$
Pretraining lr	$1e - 3$	$1e - 3$	$1e - 4$	$1e - 4$	—	$2e - 5$
Pretraining Optimizer	SGD	Adam	AdamW	Adam	—	Adam
Iteration	4000	3000	5000	2500	10000	10000
Iteration time (s)	0.06259	0.18580	0.01035	0.00813	0.93807	8.54128
Total time (s)	250.8	557.4	51.75	20.3	9380.7	85412.8



Fig. 8. Ablation Study of Optimization Mode. (a): Consistent seed with model fine-tuning. (b): Joint optimization of seed and model parameters. (c): Joint optimization of reflectance seed and illumination model parameters. (d): Consistent model with seed fine-tuning.

input, but this training method actually makes the coordination between the model and input noise more chaotic, leading to further performance degradation; (c) achieves almost similar performance to our method but still does not surpass it. This is because the simple structure of the illumination map makes it difficult to improve by adding more parameters. Our approach directly optimizes the input and achieves the best results, which is consistent with our observation in Sec III-A.

TABLE VI

QUANTITATIVE ANALYSIS FROM ABLATION STUDIES DEMONSTRATING THE IMPACT OF OPTIMIZATION MODE. ✓ MEANS PLACING THIS PARAMETER IN THE OPTIMIZER FOR ITERATIVE TRAINING. \* INDICATES OPTIMIZING  $\theta_l$  AND FREEZING  $\theta_r$ . THE BEST RESULTS ARE MARKED IN BOLD.

Setting	$z$	$\theta$	PSNR↑	SSIM↑	NIQE↓	MUSIQ↑
(a)		✓	15.51	0.68	3.61	56.92
(b)	✓	✓	14.88	0.61	4.08	56.58
(c)	✓	*	17.62	0.69	3.16	57.83
(d)	✓		<b>18.10</b>	<b>0.75</b>	<b>2.80</b>	<b>58.98</b>

3) *Pre-trained Model Selection.* To elucidate the benefits of our proposed training framework, we train our enhancement network utilizing various generative pre-trained models from both GAN [23], [45] and VAE [26], [28] families, and DDPM [21]. As shown in Figure 9, the images produced by different pre-trained models have different styles. GANs are more likely to generate dark and cold-toned images, while VAEs tend to generate high-contrast images. For the VAE models, we adhere to the configurations outlined in Section III-C, specifically employing solely the Decoder component. Regarding the GAN models, we utilize the Generator component, excluding its initial convolutional layer, to enhance control over the content of the generated images. Overall, we believe that both GANs and VAEs can encapsulate extensive

pre-trained knowledge and be effectively applied to low-light enhancement. However, an interesting fact is that VAEs often have fewer parameters and faster runtime compared to GANs. This is likely due to VAEs having a well-defined latent space, whereas GANs rely solely on the generator to produce images. In terms of efficiency, VAEs demonstrate higher potential compared to GANs, which ultimately led us to choose VAE-based models for our work. Remarkably, as shown in Table V, VQ-VAE-2 pre-trained model outperforms the others on the balance of speed and effect, albeit other models demonstrate near-par performance. This prompted us to select the VQ-VAE-2 decoder for final implementation. All these results suggest that our approach is versatile enough to be compatible with diverse image generation techniques, even those not explicitly designed for low-light image enhancement tasks.

4) *Analysis of Loss Contributions.* In our framework, the loss components  $\mathcal{L}_{RE}$  and  $\mathcal{L}_I$  are crucial for image reconstruction and brightness enhancement in low-light image enhancement tasks, while  $\mathcal{L}_E$  and  $\mathcal{L}_S$  contribute to stable Retinex decomposition and improved image quality. Table VII details the impact of omitting  $\mathcal{L}_E$  and  $\mathcal{L}_S$  on the model's performance, and Figure 10 illustrates the consequences of removing each regularization term. Our findings indicate that the absence of any loss component significantly degrades enhancement performance. Specifically, excluding  $\mathcal{L}_{RE}$  hinders the model's ability to accurately represent the original image. Omitting  $\mathcal{L}_I$  and  $\mathcal{L}_E$  disrupts the light map's enhancement capabilities, and without  $\mathcal{L}_S$ , the model struggles with efficient noise reduction in extremely dark areas, adversely affecting the fitting speed.

TABLE VII  
QUANTITATIVE ANALYSIS FROM ABLATION STUDIES DEMONSTRATING THE IMPACT OF LOSSES. THE BEST RESULTS ARE MARKED IN BOLD.

Setting	PSNR↑	SSIM↑	NIQE↓	MUSIQ↑
w/o $\mathcal{L}_{RE}$	11.67	0.55	4.74	40.49
w/o $\mathcal{L}_S$	15.66	0.64	4.22	50.79
w/o $\mathcal{L}_E$	9.93	0.42	5.68	48.46
w/o $\mathcal{L}_I$	8.31	0.26	5.78	41.09
Proposed	<b>18.10</b>	<b>0.75</b>	<b>2.80</b>	<b>58.98</b>

5) *Hyperparameter Settings.* In section IV, we proposed different hyperparameter configurations for supervised and unsupervised datasets. This distinction is necessary because the LOL dataset contains more noise, requiring stronger smoothness constraints, while the non-reference datasets are



Fig. 9. Visual quality comparison of different pre-training models. \* means loading pre-trained weight to this generator, e.g., DCGAN\* for loading pre-trained weights while DCGAN for random initialization.

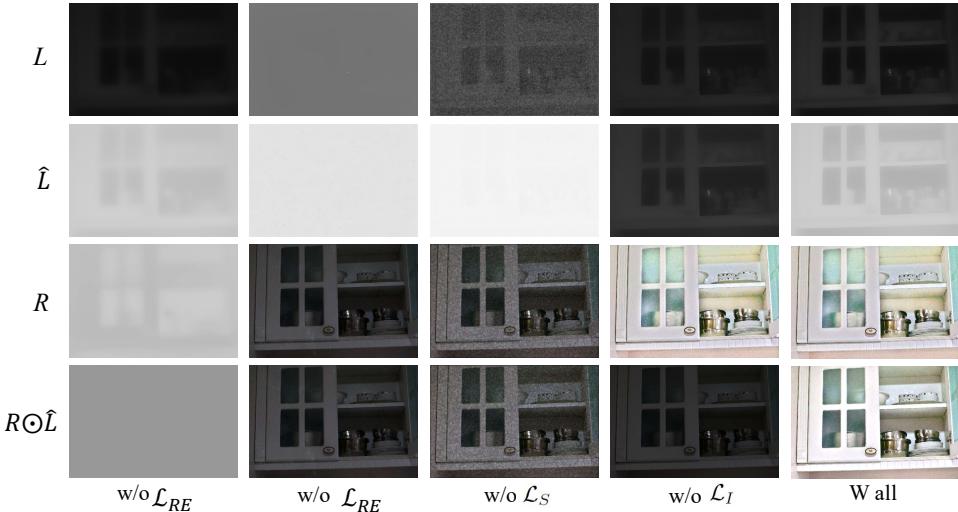


Fig. 10. Visual quality comparison of ablating different losses,  $L$  denotes the illumination map,  $\hat{L}$  denotes the illumination map after gamma transformation and  $R$  denotes the reflectance map.

more complex, necessitating more iterations. Using different parameters allows the model to achieve optimal performance. However, applying the same set of parameters across datasets also enables the proposed method to achieve state-of-the-art performance. As shown in Table VIII, we have listed several combinations of  $\tau$  and iteration settings along with their rankings in Table I.

TABLE VIII  
PERFORMANCE SCORES OF OUR METHODS UNDER DIFFERENT SETTINGS.

Method	Rank	RoR
Ours ( $\tau = 0.2, iters = 3750$ )	1	4.70
Ours ( $\tau = 0.2, iters = 2500$ )	1	4.72
Ours ( $\tau = 0.4, iters = 5000$ )	1	4.33
Ours ( $\tau = 0.4, iters = 3750$ )	1	4.64
Ours ( $\tau = 0.4, iters = 2500$ )	1	4.36
Ours ( $\tau = 0.6, iters = 5000$ )	1	4.43
Ours ( $\tau = 0.6, iters = 3750$ )	1	4.74
Ours ( $\tau = 0.2, iters = 5000$ )	1	4.78
Ours ( $\tau = 0.6, iters = 2500$ )	1	4.89
Ours	1	3.78

### E. User Study

We conduct a user study to more comprehensively evaluate the visual quality of enhanced results obtained by different methods. All methods mentioned previously were included in this user study. The datasets utilized in the user study encompass all those mentioned in Section IV-C for testing purposes, which include LOL, MEF, LIME, DICM, and NPE datasets. For each test image, participants were provided with the input low-light image, alongside the corresponding images enhanced by our method and by others. We designed a dedicated User Interface to present all images and collect the results given by users. All images are randomly arranged on this User Interface to ensure fairness. All users were asked to view all the images on a PHL 272 S9 screen, with Intel(R) UHD Graphics 730. All remaining parameters were configured with their default settings. A total of 21 participants were enlisted to assess the color accuracy, noise removal, clarity, lighting coordination, and overall quality of these images on a scale from 1 (worst) to 5 (best), with the average rating

serving as the result of the user study. Before the user study was officially conducted, all users were asked to complete the pre-training by conducting a test vote that was exactly the same as the formal voting process. Ten random image was used to detect whether the user understood the User Interface. After the user completed the test vote, he reported to the tester what he thought was the "worst image" and "best image" on each indicator. If the user fully understood the meaning of 1 and 5, the voting officially began. These 21 participants had different research backgrounds and personal situations, including whether they had low-level vision, graphic imagery, or photography aesthetics research background, age, and gender information. We reported the basic situation of the participants in Figure 11 with their permission.

The findings of the user study are encapsulated in Table IX. The distribution of votes indicates a pronounced preference for the results generated by our method, showcasing a clear advantage over competing methods. Most methods have good performance in one indicator but are not advantageous in other aspects. For example, SNR and dump have good performance in noise removal. Our approach attained an average score of 3.18 across all images and also ranked first in color accuracy, clarity, and lighting coordination, evidencing that it produces results that are more visually appealing in comparison to other techniques. We also use ANOVA (Analysis of Variance) to analyze the significance of the user study data. We first formulated the hypotheses as follows:

- **Null Hypothesis ( $H_0$ ).** There is no difference in the average ratings of pictures produced by different models.
- **Alternative Hypothesis ( $H_0$ ).** There is a difference in the average ratings of pictures produced by different models.

We set the significance level ( $\alpha$ ) to 0.05. As shown in the last row in Table IX, the p-value of each item is less than 0.05, which indicates that the user study is statistically significant.

TABLE IX  
AVERAGE SCORES FROM USER STUDY. COLOR INDICATES THE SCORE FOR COLOR ACCURACY, NOISE INDICATES THE SCORE FOR NOISE REMOVAL, CLARITY INDICATES TO THE SCORE FOR IMAGE CLARITY, AND QUALITY INDICATES THE OVERALL QUALITY SCORE.  $p$  INDICATES P-VALUES BY ANALYSIS OF VARIANCE. THE HIGHEST SCORES ARE HIGHLIGHTED IN BOLD.

Method	Color	Noise	Clarity	Lighting	Quality
STAR [49]	3.20	3.86	3.41	3.18	2.46
LR3M [48]	2.69	3.58	2.75	3.57	1.93
LIME [51]	3.85	2.53	3.17	3.12	2.39
Retinex-Net [16]	3.27	2.70	1.28	1.35	1.69
KinD++ [19]	1.08	3.33	3.11	3.48	2.32
SNR [5]	2.91	<b>4.70</b>	2.85	1.16	2.04
LLFlow-SKF [61]	1.47	3.56	1.57	2.78	2.40
Retinexformer [4]	3.09	2.66	2.06	3.10	2.88
EnGan [8]	2.82	3.27	2.08	3.47	2.29
RUAS [42]	3.45	3.47	2.77	3.73	1.82
SCI [41]	2.44	2.75	3.79	1.24	2.12
SGZ [39]	3.67	3.68	1.84	1.52	2.65
SCI-LLE [43]	2.26	2.45	3.28	2.75	2.10
ZeroDCE++ [10]	2.95	1.41	1.84	1.65	2.67
GDP [13]	3.61	2.01	2.87	1.98	1.71
NeRCo [11]	1.64	1.44	3.63	1.55	1.99
RetinexDIP [62]	1.79	2.47	4.28	1.39	2.31
DUNP [14]	4.35	4.38	2.91	2.87	2.59
Ours	<b>4.91</b>	3.42	<b>3.96</b>	<b>4.12</b>	<b>3.18</b>
$p$	0.006	0.007	0.011	0.003	0.002

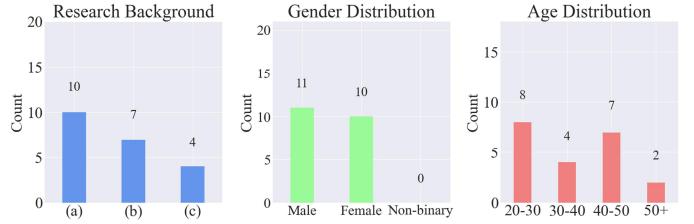


Fig. 11. The user background survey of the users participating in the user study consists of three items: research background, gender and age. (a): who have research experience in low-level vision tasks, (b): who have received education in photographic aesthetics, (c): who have no relevant experience.

#### F. Applications and Generalization Study

1) **Generalization to Various Low-Light Inputs.** To evaluate the model's capabilities in practical applications, we utilized the following datasets to test its performance in medical image enhancement and remote sensing image enhancement:

- **RFMiD:** The medical images are sourced from the Retinal Fundus Multi-Disease Image Dataset (RFMiD) [71]. This dataset includes around 860 retinal fundus images from a research facility in Jalna and Nanded, where patients from across Maharashtra come for preventative and therapeutic eye care. We selected 168 low-light retinal fundus images from this dataset to test the performance of various low-light enhancement methods. Following the protocol in [72], we employ NIQE, BRISQUE, and PIQE metrics to evaluate the performance of each method.
- **NWPU VHR-10:** This dataset [69], [70] contains a total of 800 VHR optical remote sensing images, with 715 color images acquired from Google Earth at a spatial resolution ranging from 0.5 to 2 meters, and 85 pan-sharpened color infrared images acquired from Vaihingen data at a spatial resolution of 0.08 meters. The low-light remote sensing images are generated using the synthesis method in [73]. Following the protocols in [73], we employ PSNR, SSIM, and LPIPS metrics on these remote sensing images.
- **SID.** The SID [68] dataset consists of raw short-exposure night-time images with corresponding long-exposure reference images. The extensive noise present in SID makes it highly challenging. We preprocessed the images using HDR-Toolbox [67], randomly selected 25 pairs of images from it, and enhanced them using the proposed method.
- **DICM-J.** We compressed the DICM dataset using the JPEG compression algorithm from [74] and we got DICM-J, with all parameters set to default. Although the brightness did not change significantly, the introduction of JPEG compression introduced additional noise, making the DICM dataset more challenging.

All experimental settings are consistent with those in Section IV. The experimental results are shown in Table X. As observed, other methods exhibit significant variability in performance across different datasets, whereas the proposed method achieves notable results. The reason for our method's excellent generalization ability is the seed optimization strategy, which allows for individual tuning for each specific input. This approach fully leverages the potential of the pre-trained

TABLE X  
TEST RESULTS ON NWPU VHR-10, RFMID, SID AND DICM DATASETS. THE BEST AND SECOND-BEST ZERO-SHOT METHODS ARE HIGHLIGHTED IN RED AND BLUE RESPECTIVELY.  $\uparrow$  ( $\downarrow$ ) MEANS HIGHER (LOWER) IS BETTER.

Method	NWPU VHR-10				RFMID			SID			DICM		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	NIQE $\downarrow$	BRISQUE $\downarrow$	PIQE $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	NIQE $\downarrow$	MUSIQ $\uparrow$	NIQE $\downarrow$	MUSIQ $\uparrow$	
No-Learning	STAR [49]	18.81	0.68	0.34	6.504	39.59	30.38	12.50	0.59	7.22	52.44	3.14	46.63
	LR3M [48]	19.11	0.46	0.42	5.510	33.19	34.18	10.01	0.44	7.17	44.01	5.01	60.09
	LIME [51]	19.90	0.65	0.41	6.369	41.89	28.05	14.01	0.60	8.35	51.01	3.86	49.96
Paired Training	Retinex-Net [16]	18.00	0.46	0.42	4.894	35.96	24.13	16.60	0.42	9.09	57.43	4.96	47.73
	KinD++ [19]	19.62	0.75	0.38	5.068	45.03	30.01	19.83	0.85	5.22	70.02	4.06	48.34
	SNR [5]	23.80	0.82	0.44	7.603	35.21	26.81	22.25	0.79	5.31	63.14	4.77	49.74
	LLFlow-SKF [61]	21.72	0.88	0.40	5.716	36.02	33.01	26.96	0.89	6.06	70.84	4.59	50.85
	Retinexformer [4]	23.00	0.83	0.49	5.642	29.39	36.42	23.59	0.85	3.27	62.62	4.52	50.82
Unpaired Training	EnGan [8]	21.77	0.65	0.35	5.473	34.30	31.86	16.58	0.60	4.82	53.16	4.04	45.59
	SCL-LLE [43]	22.35	0.52	0.41	6.488	42.84	25.41	11.76	0.52	7.73	55.20	3.06	51.00
	NeRCo [11]	23.73	0.86	0.30	6.549	32.18	36.06	18.44	0.83	3.33	76.55	4.08	57.55
Zero Referenced	RUAS [42]	<b>21.03</b>	0.55	0.39	9.620	39.96	37.39	17.06	0.47	5.90	55.45	6.73	44.56
	SCI [41]	20.98	0.49	0.37	5.971	38.81	<b>24.20</b>	13.86	0.54	8.18	56.32	4.63	46.81
	SGZ [39]	20.25	<b>0.61</b>	0.33	6.093	37.37	29.95	<b>17.47</b>	0.53	8.28	60.92	<b>3.45</b>	48.50
	ZeroDCE++ [10]	20.69	0.57	0.33	5.792	<b>33.68</b>	33.01	14.55	<b>0.56</b>	8.12	53.39	3.87	<b>53.57</b>
	GDP [13]	15.91	0.56	0.34	<b>4.863</b>	34.49	37.27	15.71	0.51	6.44	<b>61.98</b>	3.94	48.62
	RetinexDIP [62]	<b>22.63</b>	0.60	<b>0.30</b>	6.465	37.43	37.24	11.35	0.40	7.26	54.72	4.72	45.18
	DUNP [14]	19.88	0.47	0.36	5.560	34.41	31.73	14.20	0.45	<b>4.45</b>	60.66	3.78	47.48
	<b>Ours</b>	20.87	<b>0.74</b>	<b>0.28</b>	<b>5.188</b>	<b>29.86</b>	<b>23.98</b>	<b>18.06</b>	<b>0.71</b>	<b>3.69</b>	<b>61.14</b>	<b>3.18</b>	<b>54.58</b>

TABLE XI  
QUANTITATIVE ANALYSIS DEMONSTRATING THE IMPACT OF IMAGE DECOMPOSITION FRAMEWORKS. (A) CAREAGA AND AKSOY'S METHOD. (B) RETINEX.

Setting	PSNR $\uparrow$	SSIM $\uparrow$	NIQE $\downarrow$	MUSIQ $\uparrow$
(A)	17.32	0.70	3.26	<b>59.76</b>
(B)	<b>18.10</b>	<b>0.75</b>	<b>2.80</b>	58.98

model, ensuring good adaptability to data from different domains.

2) *Different Image Decomposition Frameworks.* To evaluate the robustness of the proposed method within different image decomposition frameworks, we replaced the illumination decoder and reflectance decoder in the Retinex framework with the method proposed by Careaga and Aksoy [66]. Careaga and Aksoy's method [66] addresses the long-tailed distribution effect observed in shading estimation by transforming the shading into an Inverse Shading Representation:

$$D = \frac{1}{S+1} \quad (17)$$

where  $S$  represents the shading and  $D$  denotes the inverse shading. Under the framework of Careaga and Aksoy's method, the proposed model achieves the enhanced image results using the following formulation:

$$A = \frac{I \times G(z, \tilde{\theta})}{1 - G(z, \tilde{\theta})} \quad (18)$$

where  $I$  is the input low-light image,  $G$  is the shading estimation model proposed in Careaga and Aksoy's method,  $\tilde{\theta}$  represents the pre-trained weights, and  $A$  is the final enhanced output. We maintained all prior constraints and optimization processes unchanged. The results obtained on the LOL dataset are shown in Table XI. Although the performance of our model using Careaga and Aksoy's method did not surpass that of the Retinex-based model, it still yielded competitive results.

This indicates that the proposed method is not exclusively dependent on the Retinex framework and can be generalized to various other applications.

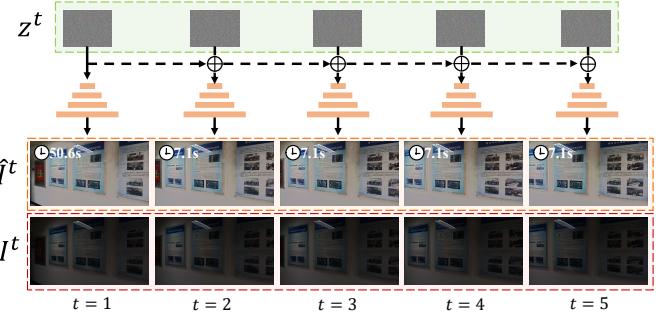


Fig. 12. Results of video low-light enhancement using the adapted proposed method, the seed optimization idea we proposed can also be used to improve efficiency in the process of video processing.

3) *low-light video enhancement.* We also extended the proposed method to the video processing domain by leveraging the similarity between adjacent frames to reduce the number of iterations starting from the second frame. Specifically, for the first frame, we performed the complete seed optimization process. From the second frame onward, we directly updated the seed from the previous frame using residual updates. This approach allowed the residual updates to quickly adapt to the new image with minimal time required. Taking the DID [65] dataset as an example, Figure 12 illustrates the pipeline of this process.

## V. CONCLUSION

In this work, we introduce a zero-shot low-light enhancement approach that integrates a pre-trained generative model into a Retinex-based enhancement framework. We discuss the application of pre-trained generators in the image enhancement process and address how to embed pre-trained knowledge into

the zero-shot image enhancement process. To counter the slow fine-tuning process and potential forgetting issues, we employ a seed optimization strategy for efficient iterative optimization. Throughout our experiments, we thoroughly examined the functionality of different model components and validated the robustness and effectiveness of the proposed framework across various datasets. Our approach significantly reduces the dependence on extensive low-light datasets, demonstrating its effectiveness across a range of lighting conditions and scenes.

## REFERENCES

- [1] C. Zheng, D. Shi, and W. Shi, "Adaptive unfolding total variation network for low-light image enhancement," pp. 4439–4448, October 2021.
- [2] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21–37.
- [3] M. J. Islam, C. Edge, Y. Xiao, P. Luo, M. Mehtaz, C. Morse, S. S. Enan, and J. Sattar, "Semantic segmentation of underwater imagery: Dataset and benchmark," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 1769–1776.
- [4] Y. Cai, H. Bian, J. Lin, H. Wang, R. Timofte, and Y. Zhang, "Retinexformer: One-stage retinex-based transformer for low-light image enhancement," *arXiv preprint arXiv:2303.06705*, 2023.
- [5] 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. 17714–17724.
- [6] H. Jiang, A. Luo, S. Han, H. Fan, and S. Liu, "Low-light image enhancement with wavelet-based diffusion models," *arXiv preprint arXiv:2306.00306*, 2023.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] S. Yang, M. Ding, Y. Wu, Z. Li, and J. Zhang, "Implicit neural representation for cooperative low-light image enhancement," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 12 918–12 927.
- [12] Z. Liang, C. Li, S. Zhou, R. Feng, and C. C. Loy, "Iterative prompt learning for unsupervised backlit image enhancement," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 8094–8103.
- [13] B. Fei, Z. Lyu, L. Pan, J. Zhang, W. Yang, T. Luo, B. Zhang, and B. Dai, "Generative diffusion prior for unified image restoration and enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 9935–9946.
- [14] 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.
- [15] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9446–9454.
- [16] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," *arXiv preprint arXiv:1808.04560*, 2018.
- [17] Z. Ni, W. Yang, S. Wang, L. Ma, and S. Kwong, "Towards unsupervised deep image enhancement with generative adversarial network," *IEEE Transactions on Image Processing*, vol. 29, pp. 9140–9151, 2020.
- [18] C. Li, C. Guo, and C. C. Loy, "Learning to enhance low-light image via zero-reference deep curve estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 8, pp. 4225–4238, 2021.
- [19] Y. Zhang, X. Guo, J. Ma, W. Liu, and J. Zhang, "Beyond brightening low-light images," *International Journal of Computer Vision*, vol. 129, pp. 1013–1037, 2021.
- [20] 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.
- [21] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.
- [22] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.
- [23] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.
- [24] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4401–4410.
- [25] A. Brock, J. Donahue, and K. Simonyan, "Large scale gan training for high fidelity natural image synthesis," *arXiv preprint arXiv:1809.11096*, 2018.
- [26] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
- [27] A. Van Den Oord, O. Vinyals *et al.*, "Neural discrete representation learning," *Advances in neural information processing systems*, vol. 30, 2017.
- [28] A. Razavi, A. Van den Oord, and O. Vinyals, "Generating diverse high-fidelity images with vq-vae-2," *Advances in neural information processing systems*, vol. 32, 2019.
- [29] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10 684–10 695.
- [30] R.-Q. Wu, Z.-P. Duan, C.-L. Guo, Z. Chai, and C. Li, "Ridep: Revitalizing real image dehazing via high-quality codebook priors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 22 282–22 291.
- [31] K. C. Chan, X. Wang, X. Xu, J. Gu, and C. C. Loy, "Glean: Generative latent bank for large-factor image super-resolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 14 245–14 254.
- [32] S. Zhou, K. Chan, C. Li, and C. C. Loy, "Towards robust blind face restoration with codebook lookup transformer," *Advances in Neural Information Processing Systems*, vol. 35, pp. 30 599–30 611, 2022.
- [33] 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.
- [34] 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.
- [35] 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.
- [36] 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.
- [37] J. Ke, Q. Wang, Y. Wang, P. Milanfar, and F. Yang, "Musiq: Multi-scale image quality transformer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5148–5157.
- [38] 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, 2012.
- [39] S. Zheng and G. Gupta, "Semantic-guided zero-shot learning for low-light image/video enhancement," in *Proceedings of the IEEE/CVF Winter conference on applications of computer vision*, 2022, pp. 581–590.
- [40] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [41] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward fast, flexible, and robust low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5637–5646.
- [42] 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.

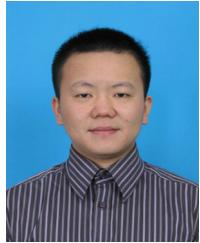
- [43] D. Liang, L. Li, M. Wei, S. Yang, L. Zhang, W. Yang, Y. Du, and H. Zhou, "Semantically contrastive learning for low-light image enhancement," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 2, 2022, pp. 1555–1563.
- [44] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
- [45] G. Ying, X. He, B. Gao, B. Han, and X. Chu, "Eagan: Efficient two-stage evolutionary architecture search for gans," in *European Conference on Computer Vision*. Springer, 2022, pp. 37–53.
- [46] S. K. Dhara and D. Sen, "Exposedness-based noise-suppressing low-light image enhancement," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 6, pp. 3438–3451, 2021.
- [47] Y. Ren, Z. Ying, T. H. Li, and G. Li, "LECARM: Low-light image enhancement using the camera response model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 4, pp. 968–981, 2018.
- [48] 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.
- [49] J. Xu, Y. Hou, D. Ren, L. Liu, F. Zhu, M. Yu, H. Wang, and L. Shao, "Star: A structure and texture aware retinex model," *IEEE Transactions on Image Processing*, vol. 29, pp. 5022–5037, 2020.
- [50] S.-Y. Yu and H. Zhu, "Low-illumination image enhancement algorithm based on a physical lighting model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 1, pp. 28–37, 2017.
- [51] 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.
- [52] 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.
- [53] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 355–368, 1987.
- [54] 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.
- [55] 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.
- [56] L. Li, R. Wang, W. Wang, and W. Gao, "A low-light image enhancement method for both denoising and contrast enlarging," in *2015 IEEE international conference on image processing (ICIP)*, 2015, pp. 3730–3734.
- [57] L. Zhang, P. Shen, X. Peng, G. Zhu, J. Song, W. Wei, and H. Song, "Simultaneous enhancement and noise reduction of a single low-light image," *IET Image Processing*, vol. 10, no. 11, pp. 840–847, 2016.
- [58] W. Ma, J.-M. Morel, S. Osher, and A. Chien, "An L1-based variational model for retinex theory and its application to medical images," in *CVPR 2011*. IEEE, 2011, pp. 153–160.
- [59] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel, "A variational framework for retinex," *International Journal of computer vision*, vol. 52, pp. 7–23, 2003.
- [60] W. Ma, S. Osher *et al.*, "A tv bregman iterative model of retinex theory," *Inverse Probl. Imaging*, vol. 6, no. 4, pp. 697–708, 2012.
- [61] 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.
- [62] 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.
- [63] J. Li, X. Feng, and Z. Hua, "Low-light image enhancement via progressive-recursive network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 11, pp. 4227–4240, 2021.
- [64] 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.
- [65] H. Fu, W. Zheng, X. Wang, J. Wang, H. Zhang, and H. Ma, "Dancing in the dark: A benchmark towards general low-light video enhancement," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 12877–12886.
- [66] C. Careaga and Y. Aksoy, "Intrinsic image decomposition via ordinal shading," *ACM Transactions on Graphics*, vol. 43, no. 1, pp. 1–24, 2023.
- [67] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced High Dynamic Range Imaging (2nd Edition)*. Natick, MA, USA: AK Peters (CRC Press), July 2017.
- [68] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3291–3300.
- [69] H. Su, S. Wei, M. Yan, C. Wang, J. Shi, and X. Zhang, "Object detection and instance segmentation in remote sensing imagery based on precise mask r-cnn," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 1454–1457.
- [70] H. Su, S. Wei, S. Liu, J. Liang, C. Wang, J. Shi, and X. Zhang, "Hqisnet: High-quality instance segmentation for remote sensing imagery," *Remote Sensing*, vol. 12, no. 6, p. 989, 2020.
- [71] S. Panchal, A. Naik, M. Kokare, S. Pachade, R. Naigaonkar, P. Phadnis, and A. Bhange, "Retinal fundus multi-disease image dataset (rfmid) 2.0: a dataset of frequently and rarely identified diseases," *Data*, vol. 8, no. 2, p. 29, 2023.
- [72] Y. Ma, J. Liu, Y. Liu, H. Fu, Y. Hu, J. Cheng, H. Qi, Y. Wu, J. Zhang, and Y. Zhao, "Structure and illumination constrained gan for medical image enhancement," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3955–3967, 2021.
- [73] M. Zhao, R. Yang, M. Hu, and B. Liu, "Deep learning-based technique for remote sensing image enhancement using multiscale feature fusion," *Sensors*, vol. 24, no. 2, 2024.
- [74] J. Jiang, K. Zhang, and R. Timofte, "Towards flexible blind jpeg artifacts removal," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 4997–5006.
- [75] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 3015–3024.
- [76] Z. Zhao, H. Lin, D. Shi, and G. Zhou, "A non-regularization self-supervised retinex approach to low-light image enhancement with parameterized illumination estimation," *Pattern Recognition*, vol. 146, p. 110025, 2024.
- [77] D. J. Jobson, Z.-u. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE transactions on image processing*, vol. 6, no. 3, pp. 451–462, 1997.
- [78] —, "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.
- [79] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman, "Ground truth dataset and baseline evaluations for intrinsic image algorithms," in *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009, pp. 2335–2342.
- [80] C. Rother, M. Kiefel, L. Zhang, B. Schölkopf, and P. Getler, "Recovering intrinsic images with a global sparsity prior on reflectance," *Advances in neural information processing systems*, vol. 24, 2011.
- [81] J. T. Barron and J. Malik, "Color constancy, intrinsic images, and shape estimation," in *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012, Proceedings, Part IV 12*. Springer, 2012, pp. 57–70.
- [82] Y. Li and M. S. Brown, "Single image layer separation using relative smoothness," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 2752–2759.
- [83] F. Banterle, A. Artusi, A. Moreo, F. Carrara, and P. Cignoni, "Norvdppn++: Real-time no-reference image quality metrics," *IEEE Access*, vol. 11, pp. 34 544–34 553, 2023.
- [84] Y. Jin, W. Yang, and R. T. Tan, "Unsupervised night image enhancement: When layer decomposition meets light-effects suppression," in *European Conference on Computer Vision*. Springer, 2022, pp. 404–421.
- [85] W. Xiong, D. Liu, X. Shen, C. Fang, and J. Luo, "Unsupervised low-light image enhancement with decoupled networks," in *2022 26th International Conference on Pattern Recognition (ICPR)*. IEEE, 2022, pp. 457–463.
- [86] N. Zheng, J. Huang, F. Zhao, X. Fu, and F. Wu, "Unsupervised underexposed image enhancement via self-illuminated and perceptual guidance," *IEEE Transactions on Multimedia*, vol. 25, pp. 5469–5484, 2022.
- [87] Y. Li, Y. Niu, R. Xu, and Y. Chen, "Zero-referenced low-light image enhancement with adaptive filter network," *Engineering Applications of Artificial Intelligence*, vol. 124, p. 106611, 2023.
- [88] W. Sun, L. Tian, Q. Wang, R. Cui, J. Lu, X. Yang, and Y. Zhang, "Explore unsupervised exposure correction via illumination component divided guidance," *Knowledge-Based Systems*, vol. 276, p. 110730, 2023.
- [89] 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.



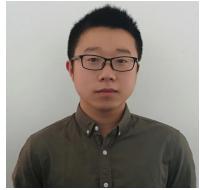
**Yuxuan Gu** received the B.S. degree in Computer Science from Anhui University, Hefei, China, in 2023. He is currently working toward the M.S. degree in Instrument Science and Technology at the School of Engineering Science, University of Science and Technology of China, Hefei, China. His research interests include computer vision and low-level vision.



**Huaiyan Chen** received the Ph.D. degree from the University of Science and Technology of China, Hefei, China, in 2022. He is currently an Associate Researcher with the School of Engineering Science, University of Science and Technology of China. His research interests include deep learning, image/video segmentation, and image/video restoration. He has published more than 20 papers in refereed conferences and journals, including the IEEE Transactions on Image Processing, IEEE Transactions on Multimedia, CVPR, ICCV, and NeurIPS.



**Yi Jin** received the Ph.D. degree from the University of Science and Technology of China, Hefei, China, in 2013. He is currently an Associate Professor at the School of Engineering Science, University of Science and Technology of China. He has authored or co-authored over 50 refereed articles. His current research interests include intellectual detection, image processing, and artificial intelligence. Dr. Jin was a recipient of the Key Innovations Award of the Chinese Academy of Sciences and the First Class Science and Technology Progress Award of Anhui Province in 2016 and 2019.



**Ben Wang** Ben Wang received the B.S. degree in measurement and control technology and instruments from the Anhui University, Hefei, China, in 2019. He is currently working toward the M.S. in mechanical engineering at the School of Engineering Science, University of Science and Technology of China, Hefei, China. He research interests include machine vision, 3-D vision, and computational imaging.



**Zhixiang Wei** received the B.S. degree in Communication Engineering from the Anhui University, Hefei, China, in 2021. He is currently working toward the Ph.D degree in the University of Science and Technology of China. His current research interests include computer vision, and domain adaptation & generalization.



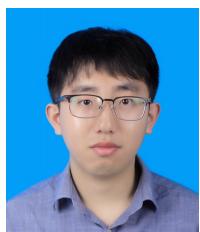
**Enhong Chen** (Fellow, IEEE) received the Ph.D. degree from the University of Science and Technology of China, Hefei, China. He is a professor and vice dean of the School of Computer Science, USTC. His general areas of research include data mining and machine learning, social network analysis, and recommender systems. He has published more than 100 papers in refereed conferences and journals, including IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Industrial Electronics, KDD, ICDM, NIPS, and CIKM. He was on program committees of numerous conferences including KDD, ICDM, and SDM. He received the Best Application Paper Award on KDD-2008, the Best Research Paper Award on ICDM-2011, and the Best of SDM-2015. His research is supported by the US National Science Foundation for Distinguished Young Scholars of China. He is a fellow of the IEEE.



**Xiaoxiao Ma** received the B.S. degree in Computer Science from China Agricultural University, Beijing, China, in 2022. She is currently working toward the M.S. degree in Instrument Science and Technology at the School of Engineering Science, University of Science and Technology of China, Hefei, China. Her research interests include computer vision and FPGA acceleration.



**Haoxuan Wang** received the B.S. degree in measurement and control technology and instruments from the Anhui University, Hefei, China, in 2020. He is currently working toward the Ph.D degree in the University of Science and Technology of China. His current research interests include computer vision, and model compression & acceleration.



**Pengyang Ling** received the B.S degree from Hefei University of Technology, Hefei, China, in 2020. He is currently working toward the Ph.D degree in the University of Science and Technology of China. His current research interests include image restoration, image generation, and image editing.