Cycle-Interactive Generative Adversarial Network for Robust Unsupervised Low-Light Enhancement

Zhangkai Ni¹, Wenhan Yang², Hanli Wang¹∗, Shiqi Wang³, Lin Ma³, Sam Kwong⁴∗
¹Department of Computer Science and Technology, Tongji University, Shanghai, China
²School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
³Meituan, Beijing, China, ⁴Department of Computer Science, City University of Hong Kong, Hong Kong
{zkni,hanliwang}@tongji.edu.cn, wenhan.yang@ntu.edu.sg, forest.linma@gmail.com, {shiqwang,cssamk}@cityu.edu.hk

ABSTRACT

Getting rid of the fundamental limitations in fitting to the paired training data, recent unsupervised low-light enhancement methods excel in adjusting illumination and contrast of images. However, for unsupervised low-light enhancement, the remaining noise suppression issue due to the lacking of supervision of detailed signal largely impedes the wide deployment of these methods in real-world applications. Herein, we propose a novel Cycle-Interactive Generative Adversarial Network (CIGAN) for unsupervised low-light image enhancement, which is capable of not only better transferring illumination distributions between low/normal-light images but also manipulating detailed signals between two domains, e.g., suppressing/synthesizing realistic noise in the cyclic enhancement/degradation process. In particular, the proposed low-light guided transformation feed-forwards the features of low-light images from the generator of enhancement GAN (eGAN) into the generator of degradation GAN (dGAN). With the learned information of real low-light images, dGAN can synthesize more realistic diverse illumination and contrast in low-light images. Moreover, the feature randomized perturbation module in dGAN learns to increase the feature randomness to produce diverse feature distributions, persuading the synthesized low-light images to contain realistic noise. Extensive experiments demonstrate both the superiority of the proposed method and the effectiveness of each module in CIGAN.

CCS CONCEPTS

• Computing methodologies → Image processing; Computer vision; Unpaired image enhancement.

KEYWORDS

Low-light image enhancement, generative adversarial network (GAN); quality attention module

∗Corresponding authors: Hanli Wang and Sam Kwong.

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

ACM Reference Format:
Zhangkai Ni¹, Wenhan Yang², Hanli Wang¹∗, Shiqi Wang³, Lin Ma³, Sam Kwong⁴∗. 2022. Cycle-Interactive Generative Adversarial Network for Robust Unsupervised Low-Light Enhancement. In Proceedings of the 30th ACM International Conference on Multimedia (MM ’22), October 10–14, 2022, Lisbon, Portugal. ACM, New York, NY, USA, 9 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Recent years have witnessed an accelerated growth of capturing devices, enabling the ubiquitous image acquisition in various illumination conditions. Typically, images acquired under low-light conditions inevitably degraded by various visual quality impairments, such as undesirable visibility, low contrast, and intensive noise. Low-light image enhancement aims to restore the latent normal-light image from the observed low-light one to simultaneously obtain desirable visibility, appropriate contrast, and suppressed noise [34, 39]. It greatly improves the quality of images to benefit human vision and can also assist in high-level computer vision tasks, such as image classification [22], face recognition [17], and objection detection [22], etc. Pioneering low-light image enhancement methods stretch the dynamic range of low-light images, i.e. Histogram equalization (HE) [1, 2, 8, 33], or adjust the decomposed illumination and reflectance layers adaptively, i.e. Retinex-based approaches [11, 13, 18, 35].

Recently, learning-based approaches have achieved remarkable successes [4, 5, 27, 34]. Most of these methods follow the paradigm of supervised learning and heavily rely on the well-prepared paired normal/low-light images to train and evaluate models. However, the commonly seen paired training datasets suffer from their respective limitations. First, synthesized data via a simplified simulated imaging pipeline [23] might fail to capture intrinsic properties of real low-light images. Second, it is quite labor-intensive and time-consuming to create manual retouching data [3, 34] by expert retouchers. It also takes the risk of personal quality bias of the retouchers to adopt such kinds of data as the training data. Third, real captured data [37] might capture real degradation but fail to cover diverse scenes and objects in the wild. Besides, the ground truths captured with a pre-defined setting, i.e. the exposure time and ISO, might not be optimal. Therefore, the reliance of supervised methods on the paired data inevitably leads to the domain shift between the training data and testing data in the real world, further bringing challenges to the generalization on real low-light images.

Recently, a series of unsupervised low-light enhancement methods are proposed. These methods have no reliance on the paired...
training data and only require two unpaired collections of low/normal-light images. They are built based on the uni-directional generative adversarial network (GAN) [17] or learnable curve adjustment [12]. These methods achieved promising performances in illumination/contrast adjustment. However, due to the absence of supervision of detailed signal, the quality for some challenging real low-light images with intensive noise are not satisfactory. As a very similar topic, image aesthetic quality enhancement benefits from CycleGAN [6, 28, 44] to deliver state-of-the-art performance. We argue that, these CycleGANs neither handle low-light image enhancement problem effectively. First, low-light degradation introduces information lost, which makes the enhancement problem ambiguous. In other words, the mapping between low/normal-light images is one-to-many mapping. However, CycleGAN can only lead to one-to-one discriminative mapping [7]. Second, the intrinsic dimensions of low/normal-light domains are imbalanced as low-light images with intensive noise reflect more complicated properties. The imbalance might disturb the training of CycleGANs, namely that the degradation generator fails to synthesize realistic noise and subsequently the enhancement generator cannot handle the realistic degradation.

In this paper, we propose a novel Cycle Interactive GAN (CIGAN) for unsupervised low-light image enhancement to simultaneously adjust illumination, enhance contrast and suppress noise. The more comprehensive consideration of image degradation leads to more effective degradation and enhancement processes in cycle modeling. In other words, the more realistic and diverse the low-light images generated in image degradation, the better and more robust the results in image enhancement. To address the above-mentioned issues of CycleGANs, efforts have been made in three aspects. First, we make the degradation and enhancement generators in our CIGAN interact with each other. More specifically, we propose a novel low-light guided transformation to transfer the features of real low-light images from the enhancement generator to the degradation generator. With the information of different real low-light images as the reference during the whole training process, more diverse low-light images are synthesized, which is beneficial for modeling multiple mappings relationship between low/normal-low images. Second, to handle the domain imbalance issue, we incorporate a novel feature randomized perturbation in the degradation generator. The perturbation applies a learnable randomized affine transform to the intermediate features, which balances the intrinsic dimensions of the features in two domains and is beneficial for synthesizing realistic noise. Last but not least, we design a series of advanced modules to improve the modeling capacities of our CIGAN, such as a dual attention module at the generator side, a multi-scale feature pyramid at the discriminator side, a logarithmic image processing model as the fusion operation of enhancement generator. Extensive experimental results show that our method is superior to existing unsupervised methods and even the state-of-the-art supervised methods on real low-light images. To summarize, the main contributions of our paper are three-fold:

- We propose a novel CIGAN for unsupervised low-light image enhancement to simultaneously adjust illumination, enhance contrast, and suppress noise, which excellent in image enhancement and significantly surpasses most previous works in image degradation modeling.
- We propose a low-light guided transformation (LGT) that allows generators of degradation/enhancement to interact, which helps to generate low-light images with more diverse and realistic illumination and contrast.
- We propose a learnable feature randomized perturbation (FRP) to produce diverse feature distributions, which makes the generated low-light images with more realistic noise and benefits the low-light image enhancement process.

2 RELATED WORK

2.1 Traditional Image Enhancement

Histogram equalization (HE). HE focuses on fitting the illumination histogram to a specific distribution according to local or global statistical characteristics [1, 2, 8, 33]. For example, Arici et al. [2] cast HE as an optimization problem to improve image contrast while suppressing unnatural effects. Abdullah et al. [1] proposed a dynamic HE technique using partition operation. Stark et al. [33] presented an adaptive contrast enhancement based on generalizations of HE. The main problem of HE is that it easily causes over-enhancement and noise amplification.

Retinex-based approaches. The Retinex-based method decomposes low-light images into an illumination layer and a reflectance layer to adaptively perform joint illumination adjustment and noise suppression [11, 13, 18, 35]. Wang et al. [35] proposed a naturalness Retinex for non-uniform illumination image enhancement. Fu et al. [11] introduced a weighted variation Retinex that simultaneously estimates the illumination and reflectance layer. These methods have shown satisfying performance in illumination adjustment, however, hand-crafted constraints are difficult to accurately decompose the low-light image into the illumination and reflection layers, resulting in unnatural visual effects.
2.2 Learning-based Image Enhancement

Low-light image enhancement has achieved great successes with the booming of deep learning [12, 17, 23, 34, 39]. Broadly speaking, learning-based image enhancement methods can be roughly divided into three categories according to training data: supervised [4, 23, 29, 34], semi-supervised [39], and unsupervised [12, 17, 27, 28, 38]. LLNet [23] is the first attempt to introduce deep learning into the problem of low-light image enhancement. For enhanced performance, various supervised methods are proposed by designing sophisticated network architectures and optimization objects, such as MSR-net [31], DRD [37], SICE [4], DHN [29], and UPE [34]. However, these supervised methods are subject to the common restriction of highly dependent on paired data, which limits the performance of these methods on real testing data. Most recently, Yang et al. [39] proposed a semi-supervised low-light enhancement method. Jiang et al. [17] proposed the first unsupervised model based on GAN. Guo et al. [12] adopted the no-reference optimization without paired or unpaired data. These unsupervised methods achieve promising performance in illumination adjustment, however, noise suppression has not been considered.

3 METHOD

As shown in Fig. 2, our proposed CIGAN aims to improve the perceptual quality of low-light images by simultaneously adjusting illumination, enhancing contrast and suppressing noise under the supervision of unpaired data. It consists of complementary degradation GAN (dGAN) and enhancement GAN (eGAN).

1) dGAN: It aims to synthesize a realistic low-light image \( \hat{I}_l \in \mathbb{L} \) (low-light image domain) from the input normal-light image \( I_n \in \mathbb{N} \) (normal-light image domain) with the help of a reference low-light image \( I_l \in \mathbb{L} \). We design two modules to synthesize more realistic low-light images with low-light illumination and contrast as well as intensive noise. As denoted by the red dotted line in Fig. 2, \( L_{GT} \) (see Sec. 3.2-1) helps dGAN synthesize \( I_l \) with the feature information of \( I_l \), which preserves the content of \( I_n \) while the introduced low-light attributes of \( I_l \) makes \( I_l \) have more realistic and diverse low-light illumination and contrast. As denoted by the blue dotted line in Fig. 2, \( FRP \) (see Sec. 3.2-2) learns to inject random noise into features to make the feature distributions more diverse and synthesize more realistic image noise. Furthermore, an exposure assessment loss \( L_{exp} \) (see Sec. 3.3-1) is adopted to keep the local average illumination of synthesized low-light images close to a low value.

2) eGAN: Conversely, it focuses on learning to recover the latent normal-light image \( I_n \) from the synthesized low-light image \( I_l \). To make the generator of eGAN yield high-quality normal-light images, we design a flexible logarithmic image processing (LIP) fusion model and a dual attention module (DAM) (see Sec. 3.2-3).
Figure 3: The detailed structure of proposed (a) LGT, (b) FRP, and (c) DAM. The Conv and LReLU are convolution and LeakyReLU operations, respectively. The FRP used in dGAN helps synthesize realistic noise. The DAM is used in eGAN and dGAN to effectively model contextual information.

3.1 Model Architecture

1) Generator of dGAN. Given an input normal-light image \( I_n \) and a reference low-light image \( I_l \), we adopt the pre-trained VGG-19 network [32] \( E(\cdot) \) to extract their multi-scale feature representations as \( E^i(I_n) \) and \( E^i(I_l) \), respectively. The LGT uses the features \( E^i(I_l) \) extracted from \( I_l \) to adaptively modulates the features \( E^i(I_n) \) of \( I_n \), which helps to synthesize more diverse illumination and contrast under the guidance of various unpaired reference low-light images. The DAM is designed to capture context information from spatial and channel dimensions. The FRP learning randomly perturbs the features of decoder \( G_i \) to help synthesize low-light images with realistic noise. Therefore, the synthesized low-light image \( \tilde{l}_i \) can be expressed as:

\[
\tilde{l}_i = G_l(E^i(I_n), E^i(I_l), T_i, A_i, P_i),
\]

where \( T_i, A_i, \) and \( P_i \) are LGT, DAM, and FRP at the \( i \)-th scale, respectively. Basically, the multi-scale features \( E^i(\cdot) \) is relu_1(i.e., relu1_1, relu2_1, relu3_1, relu4_1, and relu5_1, respectively), where the parameters in \( E(\cdot) \) are fixed during the training phase.

2) Generator of eGAN. The generator of eGAN is dedicated to recovering the normal-light image \( \tilde{l}_n \) from the synthesized low-light image \( \tilde{l}_i \):

\[
\tilde{l}_n = F(G_N(E^i(\tilde{l}_i), A_i), \tilde{l}_i).
\]

where \( G_N \) is the decoder of the generator of eGAN.

Different from most previous methods that subtract the output of the network from the input low-light image to obtain the final enhanced image. We propose a flexible LIP model \( F(\cdot) \) to fuse the input \( \tilde{l}_i \) and output \( \tilde{l}_n \) into one image to combine information from two sources, which are formulated as follows,

\[
\tilde{l}_n = \frac{\tilde{l}_i + \tilde{l}_n}{\lambda + \tilde{l}_n},
\]

where \( \lambda \) is a scalar controlling the enhancement process, which is set to 1 in our work. The proposed LIP model effectively improves the stability and performance of model training (see Sec. 4.4).

3) Multi-scale Feature Pyramid Discriminator. A critical issue associated with GAN is to design a discriminator that can distinguish real/fake images based on local details and global consistency. Our solution is to design a discriminator network that can simultaneously focus on low-level texture and high-level semantic information. Therefore, we propose a multi-scale feature pyramid discriminator (MFPD) as shown in Fig. 4. The intermediate layer of the discriminator has a smaller receptive field to make the generator pay more attention to texture and local details, while the last layer has a larger receptive field to encourage the generator to ensure global consistency [27]. In short, the proposed MFPD uses multi-scale intermediate features and a pyramid scheme to guide the generators to generate images with finer local details and appreciable global consistency.

3.2 Module Design

1) Low-light Guided Transformation. We propose a novel low-light guided transformation (LGT) module to transfer the low illumination and contrast attributes of low-light images from the enhancement generator to the degradation generator, which adaptively modulate the features of normal-light images to generate low-light images with more diverse and realistic illumination and contrast. As shown in Fig. 3 (a), our LGT has two inputs at the \( i \)-th scale: the intermediate features \( E^i(I_n) \in \mathbb{R}^{b\times c\times h\times w} \) of the normal-light image and the intermediate features \( E^i(I_l) \in \mathbb{R}^{b\times c\times h\times w} \) from the reference low-light image, where \( b \) represents the batch size, \( c \) is the number of feature channels, \( h \) and \( w \) are height and width of feature, respectively. The transformation parameters \( w(E^i(I_l)) \) and \( b(E^i(I_l)) \) are learned from the reference features \( E^i(I_l) \) by two convolution layers, where the first convolution is shared. The modulated intermediate features \( \tilde{E}^i(I_n) \) can be produced via affine transformation as follows:

\[
\tilde{E}^i(I_n) = E^i(I_n) \odot w(E^i(I_l)) + b(E^i(I_l)),
\]

where \( \odot \) and \( + \) are Hadamard element-wise product and element-wise addition, respectively.

2) Feature Randomized Perturbation. A critical issue associated with GAN is to design a discriminator that can distinguish real/fake images based on local details and global consistency. Our solution is to design a discriminator network that can simultaneously focus on low-level texture and high-level semantic information. Therefore, we propose a multi-scale feature pyramid discriminator (MFPD) as shown in Fig. 4. The intermediate layer of the discriminator has a smaller receptive field to make the generator pay more attention to texture and local details, while the last layer has a larger receptive field to encourage the generator to ensure global consistency [27]. In short, the proposed MFPD uses multi-scale intermediate features and a pyramid scheme to guide the generators to generate images with finer local details and appreciable global consistency.
Compared with AdaIN [15] using statistical information to perform denormalization on channel-wise, our LGT processes at the element level and provides a flexible way to spatially modulate normal-light image features \(E^i(l_n)\). In this way, the proposed LGT incorporates the low illumination and contrast attributes of the reference low-light image into the synthesized low-light image through element-wise affine parameters \(w \in \mathbb{R}^{b \times c \times 1 \times 1}\) and \(b \in \mathbb{R}^{b \times 1 \times h \times w}\) are sampled from the standard Gaussian distributions, then fused as:

\[
\hat{x} = (1 + \theta_1 \cdot \alpha)x + \theta_2 \cdot \beta, \tag{5}
\]

where \(\{\theta_1, \theta_2\} \in \mathbb{R}^{1 \times c \times 1 \times 1}\) are two learnable scalar weights, which are learned together with all other parameters of the network by back-propagation. As shown in Fig. 2, we embed the proposed FRP module into generator of dGAN at multiple scales to make the noise of synthesized low-light images close to real-light images.

**3.3 Training Objectives**

1) **Exposure Assessment Loss.** We propose the exposure assessment loss to control the exposure consistency between the synthesized low-light images and the real ones. Our insight is to keep the average intensity of local regions of the synthesized low-light images close to a low value. Inspired by [24], we formulate \(L_{\text{exp}}\) as:

\[
L_{\text{exp}} = 1 - \exp\left(-\frac{(i - e)^2}{2\sigma^2}\right). \tag{6}
\]

where \(i\) is the average intensity of a local region, \(e\) is the desired intensity, which should be close to a low value, and \(\sigma\) controls the smoothness of the Gaussian curve. In our work, \(\sigma, e\) and the local region size are set to 0.1, 0.1, and \(7 \times 7\), respectively.

2) **Adversarial Loss.** We adopt the relativistic average Hinge loss GAN (RaHingeGAN) [19, 27] to guide dGAN to synthesize realistic low-light images. The RaHingeGAN loss of dGAN can be formulated as:

\[
\begin{align*}
L^G_G &= \mathbb{E}_{l_r \sim L} \left[ \max\left(0, 1 - \left( \mathbb{E}_{l_i \sim L} D_L(l_i) - \mathbb{E}_{l_r \sim L} D_L(l_r) \right) \right) \right] \\
&\quad + \mathbb{E}_{l_i \sim L} \left[ \max\left(0, 1 + \left( \mathbb{E}_{l_R \sim L} D_L(l_R) - \mathbb{E}_{l_i \sim L} D_L(l_i) \right) \right) \right], \\
L^D_G &= \mathbb{E}_{l_i \sim L} \left[ \max\left(0, 1 + \left( \mathbb{E}_{l_R \sim L} D_L(l_R) - \mathbb{E}_{l_i \sim L} D_L(l_i) \right) \right) \right] \\
&\quad + \mathbb{E}_{l_R \sim L} \left[ \max\left(0, 1 - \left( \mathbb{E}_{l_i \sim L} D_L(l_i) - \mathbb{E}_{l_R \sim L} D_L(l_R) \right) \right) \right],
\end{align*}
\]

where \(l_i\) is the real low-light image from the domain of low-light images \(L\), and \(l_R\) is the synthesized data from the domain of synthesized low-light images \(L\). Similarly, the RaHingeGAN loss of eGAN is expressed as:

\[
\begin{align*}
L^N_G &= \mathbb{E}_{\tilde{l}_n \sim \tilde{N}} \left[ \max\left(0, 1 - \left( \mathbb{E}_{l_n \sim N} D_N(l_n) - \mathbb{E}_{\tilde{l}_n \sim \tilde{N}} D_N(l_n) \right) \right) \right] \\
&\quad + \mathbb{E}_{l_n \sim N} \left[ \max\left(0, 1 + \left( \mathbb{E}_{\tilde{l}_R \sim \tilde{N}} D_N(l_R) - \mathbb{E}_{l_n \sim N} D_N(l_n) \right) \right) \right], \\
L^N_D &= \mathbb{E}_{\tilde{l}_n \sim \tilde{N}} \left[ \max\left(0, 1 + \left( \mathbb{E}_{l_R \sim N} D_N(l_R) - \mathbb{E}_{\tilde{l}_n \sim \tilde{N}} D_N(l_n) \right) \right) \right] \\
&\quad + \mathbb{E}_{l_R \sim N} \left[ \max\left(0, 1 - \left( \mathbb{E}_{\tilde{l}_n \sim \tilde{N}} D_N(l_n) - \mathbb{E}_{l_R \sim N} D_N(l_R) \right) \right) \right],
\end{align*}
\]

where \(N\) and \(\tilde{N}\) are the real normal-light image domain and synthesized normal-light image domain, respectively.

3) **Cycle-Consistency Loss.** It consists of two terms: (1) \(L_{\text{con}}\) calculates the L1 distance between the input images \(l_n\) and \(l_R\) and the cycled images \(\tilde{l}_n / \tilde{l}_R\), (2) \(L_{\text{per}}\) is formulated as the L2 norm between the feature maps of the input images and those of the cycled images, as follows:

\[
\begin{align*}
L_{\text{con}} &= ||l_n - \tilde{l}_n||_1 + ||l_R - \tilde{l}_R||_1, \\
L_{\text{per}} &= ||\phi^j(l_n) - \phi^j(\tilde{l}_n)||_2 + ||\phi^j(l_R) - \phi^j(\tilde{l}_R)||_2.
\end{align*}
\]

where \(\phi^j(\cdot)\) is the feature map of the \(j\)-th layer of the VGG-19 network [32], and relu4_1 is used in our work.

**Total Loss.** The proposed CIGAN is optimized with the following objective:

\[
L_G = L^G_G + L^N_G + \lambda_{\text{exp}} L_{\text{exp}} + \lambda_{\text{con}} L_{\text{con}} + \lambda_{\text{per}} L_{\text{per}}, \tag{10}
\]

\[
L_D = L^G_D + L^N_D, \tag{11}
\]

where \(\lambda_{\text{exp}}, \lambda_{\text{con}},\) and \(\lambda_{\text{per}}\) are positive constants to control the relative importance of \(L_{\text{exp}}, L_{\text{con}},\) and \(L_{\text{per}},\) respectively.

**4 EXPERIMENTS**

In this section, the performance of the proposed method is validated through quantitative and qualitative comparisons as well as user study.

**Dataset.** We follow [39] to comprehensively evaluate our proposed method on LOL dataset [37] with diverse scenes and much variability. It consists of 689 training image pairs and 100 test image pairs, all of which are captured in real-world scenarios. To meet the requirement of unpaired learning, the training set is divided into two

---

**Figure 4: The detailed structure of proposed MFPD.**
Table 1: Quantitative comparisons of different methods on real low-light test images in LOL-Real dataset [37]. EG denotes EnlightenGAN.

| Metric    | BIME   | BPDHE  | CRM    | DHECE  | Dong   | EFF    | CLAHE  | LIME   | MF     | CycleGAN | QAGAN |
|-----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|----------|-------|
| PSNR      | 17.85  | 13.84  | 19.64  | 14.64  | 17.26  | 17.85  | 13.13  | 15.24  | 18.73  | 18.80    | 18.97 |
| PSNR-GC   | 24.72  | 19.55  | 24.92  | 16.31  | 20.57  | 24.72  | 16.60  | 17.19  | 20.98  | 23.48    | 24.43 |
| SSIM      | 0.6526 | 0.4254 | 0.6623 | 0.4450 | 0.5270 | 0.6526 | 0.3709 | 0.4702 | 0.5590 | 0.6316   | 0.6081 |
| SSIM-GC   | 0.7231 | 0.5936 | 0.6968 | 0.4521 | 0.5715 | 0.7231 | 0.3947 | 0.4905 | 0.5765 | 0.6648   | 0.6513 |
| PSNR-GC   | 19.89  | 17.33  | 17.34  | 14.45  | 15.48  | 17.35  | 17.60  | 19.60  | 18.23  | 18.07    | 19.89 |
| SSIM      | 0.4269 | 0.6654 | 0.6859 | 0.5421 | 0.5672 | 0.4521 | 0.6906 | 0.6575 | 0.6165 | 0.6030   | 0.7817 |
| SSIM-GC   | 0.5158 | 0.7236 | 0.7459 | 0.7075 | 0.7476 | 0.7051 | 0.7250 | 0.6727 | 0.6452 | 0.6739   | 0.8189 |

partitions: 344 low-light images and another 345 normal-light images with no intersection with each other. Furthermore, we collect more norm/low-light images from the publicly accessible datasets to expand the training images to 1000 unpaired normal/low-light images.

**Baselines.** To carry out an overall comparison and evaluation, the proposed CIGAN is compared with twenty-one classical and state-of-the-art methods, including BIMEF [40], BPDHE [16], CRM [42], DHECE [26], Dong [9], EFF [41], CLAHE [45], LIME [13], MF [10], MR [18], JED [30], RRM [21], SICE [11], DRD [37], UPE [34], SICE [4], CycleGAN [44], EnlightenGAN [17], QAGAN [28], UEGAN [27], and ZeroDCE [12], where SICE and UPE, EnlightenGAN and ZeroDCE are the leading supervised and unsupervised methods for low-light image enhancement, respectively.

**Evaluation Metrics.** We follow [4, 34, 39] and adopt the most widely-used full-reference image quality assessment (FR-IQA) metrics: PSNR and SSIM [36]. And calculated the PSNR and SSIM of the Gamma correction results. To study how users prefer the enhanced results of each method, we perform a user study with 24 participants and 30 images of seven methods using pairwise comparisons. Each time the participants are randomly present with the enhanced results of two different methods of the same test image, they are then asked to select their favorite result from the two presented images. Table 2 tabulates the results of the pairwise comparison, from which we can observe

### 4.1 Quantitative Comparison

Table 1 compares the proposed CIGAN with the classical and state-of-the-art methods on LOL dataset [37]. It can be observed that the proposed method outperforms all previous methods in the comparison because it consistently achieves the highest scores in terms of PSNR, PSNR-GC, SSIM, and SSIM-GC. This reveals that the proposed CIGAN is much more effective in illumination enhancement, structure restoration, and noise suppression. From Table 1, we can see that the proposed method is significantly superior to other state-of-the-art unsupervised methods (i.e., CycleGAN, EnlightenGAN, and ZeroDCE). This is because, on one hand, dGAN makes the attributes of synthesized low-light images consistent with those of real ones, and on the other hand, eGAN is able to restore high-quality normal-light images. Another interesting observation is that the proposed CIGAN even achieves better performance than leading supervised methods (i.e., DRD, UPE, and SICE) trained on a large number of paired images. It is worth noting that the larger PSNR gap between with and without Gamma correction shows that our method can effectively remove intensive noise and restore vivid details.
that the enhanced results of the proposed CIGAN are more favorite with users because CIGAN is selected more frequently than the comparison methods. This is consistent with the quantitative and qualitative results, and further consolidates the conclusion that the proposed CIGAN is superior to the state-of-the-art methods.

### 4.4 Ablation Study

We conduct extensive ablation studies to quantitatively evaluate the effectiveness of each component in our proposed CIGAN. The variant of CIGAN w/o $L_{\text{exp}}$ replaces the proposed LIP-based fusion by subtracting the network output from the input low-light image. We perform an ablation analysis on the real low-light image in Fig. 7. It can be observed that the result produced by CIGAN is obviously better than its variants. Table 3 lists the performance of different variants of our proposed CIGAN on 100 testing images in the LOL dataset in terms of average PSNR, PSNR-GC, SSIM, and SSIM-GC. From Table 3, we want to emphasize three key components. First, it is critical for LGT to adaptively modulate the normal-light image features with low-light image features. Without it, generator $G_L$ fails to learn domain-specific properties directly.

![Figure 5: Visual quality comparisons of state-of-the-art enhancement methods. Upper left: original results. Lower right: the corresponding results after Gamma transformation correction for better comparison.](image-url)
Figure 6: The visual quality comparison for a close-up region of state-of-the-art enhancement methods.

Figure 7: Ablation study of the effectiveness of three key components (i.e., FRP, LGT, and $L_{\text{exp}}$) in our proposed CIGAN. The (e) w/o all means the CIGAN without FRP, LGT, and $L_{\text{exp}}$ that very similar to vanilla CycleGAN.

5 CONCLUSIONS

This paper aims to improve the perceptual quality of real low-light images using unpaired data only in an unsupervised manner. To this end, we propose a novel unsupervised CIGAN, which contains three elaborately designed components: (1) LGT module adaptively modulates normal-light image features with low-light image features to synthesize more diverse low-light images; (2) FRP module encourages the synthesis of low-light images with realistic noise; (3) MFPD improves image quality from coarse-to-fine. Finally, a novel exposure assessment loss is formulated to control the exposure of synthesized low-light images and attention mechanisms are adopted to further improve the image quality. Extensive experiments on real-world low-light images show that our method achieves the superior performance in both quantitative and qualitative evaluations.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous referees for their insightful comments and suggestions. This work was supported in part by National Natural Science Foundation of China under Grant 61976159, Shanghai Innovation Action Project of Science and Technology under Grant 20511100700.
