


# Generative adversarial network for low-light image enhancement

 Fei Li<sup>1,\*</sup>  | Jiangbin Zheng<sup>1</sup> | Yuan-fang Zhang<sup>2,3,\*</sup>

<sup>1</sup> School of Software, Northwestern Polytechnical University, Xi'an, Shaanxi 710100, P. R. China

<sup>2</sup> School of Computer Science, Northwestern Polytechnical University, Xi'an, Shaanxi 710100, P. R. China

<sup>3</sup> Faculty of Engineering and IT, University of Technology Sydney, Sydney, NSW 2007, Australia

## Correspondence

Jiangbin Zheng, School of Computer Science, Northwestern Polytechnical University, Xi'an, Shaanxi, 710100, P. R. China.

Email: [foreverfei875@gmail.com](mailto:foreverfei875@gmail.com); [zhengjb@nwpu.edu.cn](mailto:zhengjb@nwpu.edu.cn)

\*Fei Li and Yuan-fang Zhang have contributed equally to this work.

## Funding information

Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University, Grant/Award Number: CX201959; Synergy Innovation Foundation of the University and Enterprise for Graduate Students at Northwestern Polytechnical University, Grant/Award Number: XQ201910; National Natural Science Foundation of China, Grant/Award Number: 61972321

## Abstract

Low-light image enhancement is rapidly gaining research attention due to the increasing demands of extreme visual tasks in various applications. Although numerous methods exist to enhance image qualities in low light, it is still undetermined how to trade-off between the human observation and computer vision processing. In this work, an effective generative adversarial network structure is proposed comprising both the densely residual block (DRB) and the enhancing block (EB) for low-light image enhancement. Specifically, the proposed end-to-end image enhancement method, consisting of a generator and a discriminator, is trained using the hyper loss function. The DRB adopts the residual and dense skip connections to connect and enhance the features extracted from different depths in the network while the EB receives unique multi-scale features to ensure feature diversity. Additionally, increasing the feature sizes allows the discriminator to further distinguish between fake and real images from the patch levels. The merits of the loss function are also studied to recover both contextual and local details. Extensive experimental results show that our method is capable of dealing with extremely low-light scenes and the realistic feature generator outperforms several state-of-the-art methods in a number of qualitative and quantitative evaluation tests.

## 1 | INTRODUCTION

Generally, images captured in the low-light environment suffer from various visual quality degradations, including poor visibility [1], low contrast [2], unexpected noise [3] etc. These interference factors degrade the quality of obtained pictures and result in failures in most subsequent computer vision tasks, be it low-level or high-level, such as person re-identification [4] in night video surveillance. On the other hand, low-light images analysis is key to the understanding of scenes under some extreme vision conditions, such as automatic machines [5], monitors and pieces of automatic equipment [6].

Therefore, low-light image enhancement is gradually becoming one of the useful and urgent research problems to be resolved. In short, it aims at restoring the image captured under low-light condition to achieve perceptive details similar to those from a natural light images, with higher contrast, less noise contaminations and superior visibility. Generally

speaking, enhancement algorithms consist of a denoising and a brightness adjustment step. The enhancement should allow pertinent visual interpretations of these images and it is thus key to most computer-vision based intelligent systems [7–9], for example, automated driving and video surveillance.

However, it is non-trivial to enhance low-light images, since noises are easily amplified but hard to be removed in this ill-posed inverse problem. For this task, massive restoration algorithms are proposed in the past decade, including [10–13]. These works mostly attempt to use handcraft features or priors to exploit the hidden information in low-light patches. For example, Cheng et al. [14] was the first to propose the histogram equalisation(HE) approach for image enhancement. The main idea is to stretch the dynamic range in the original low-light image to that of a natural light image. However, it often introduced undesirable illumination distortions as well as increased noises levels.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *IET Image Processing* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology

Recently, numerous enhancement methods [13, 15, 16] based on convolutional neural networks (CNNs) have been proposed to improve enhancement performances. There has been growing research interests in end-to-end deep neural network architectures to model the mapping transfer between low-light image input and desired image as output. Specifically, these learning methods extract the abstract features and learn the non-linear mapping functions between input and output using a considerable volume of training data. The state-of-the-art RetinexNet [17] and the See-in-Dark [15] are the typical examples of these methods. The RetinexNet was a model consisting of a Decompose and a Relight network layer. The former decomposed the input low-light images into illumination and reflectance map while the latter adjusted the entire light distribution. While Chen et al. [15] focused on the enhancement of the RAW images that contain more details collected from the cameras, their methods are less efficient with compressed image datasets.

Furthermore, the above-mentioned images enhancement methods suffer performance losses in several extreme scenarios and shortcomings that yield insufficient enhancement qualities, for example, noises and unbalanced light distributions. Meanwhile, disconnection with high-level applications could also hamper the performances of enhancement. Finally, these studies did not give much attention to the uncertain relationships between spatial features of various size.

To address these problems in an attempt to further the enhancement results, we introduce in the present study a densely residual Generative Adversarial Network (DRGAN) to focusing on the feature extraction and the functional practice, that is, enhancement used in application of high-level vision applications, such as face detection in dark [18].

In particular, we propose a novel feature extraction module for the low-light enhancement by exploiting relationships between the extracted features of various sizes. Specifically, we feed the GAN network with image pairs of synthetic low-light images and their ground truth (GT) counterparts. Additionally, inspired from feature pyramid network and the multi-scaled feature fusion strategy [19], we design the enhancing blocks to extract features of different sizes to be concatenated for intermediate processing in order to improve the model's feature representation capacity. To further improve the feature representation, we modify the standard discriminator by increasing the last feature size and allowing the generator to distinguish between synthetic images and the perspectives of the patch level. Experimental results show that our enhancements are more accurate and realistic due to the proposed module as compared to the outcomes of reference algorithms.

The main contributions of this paper include: (1) A novel low-light enhancement network comprising the DRB and EB module and achieving the start-of-the-art performances on several widely used low-light datasets. (2) A novel loss function designed for image detail preservation. (3) An extensive experimental validation to demonstrate the improvements in both pixel mapping and high-level visual tasks.

The remainder of this paper is organised as follows: In Section 2, we give a brief overview of the background knowledge and related topics. We describe the proposed model in Section 3 and provide the experimental details and result analysis with per-

formance comparisons with previous works in Section 4. And finally, we conclude this work in Section 5.

## 2 | RELATED WORKS

Compared with the natural-light images containing higher contrast and more detailed information, low-light images have low illumination, often resulting in poorer performance in high-level vision tasks. Normally, low-light environment means limited light sources with weak lighting. Only target objects close to the light sources are visible while considerable illumination variations occur in one image. In this section, we briefly review and analyse the following three items: the traditional methods, the Retinex theory and the learning-based methods.

It is generally acknowledged that low-light image enhancement has gradually become a popular research topic in computer vision while a number of methods have been proposed recently. One typical characteristic of low-light image is its lower dynamic range and thus the most common solution consists of raising the contrast by stretching the range. In particular, a series of approaches, such as histogram equalisation (HE) [14, 20], aims at recovering the visibility of dark regions by contrast enhancement. Other well-known contrast enhancement methods based primarily on improving image contrast proposed in the past decades, for example, contrast-limiting adaptive histogram equalisation (CLAHE) [21] and brightness preserving bi-histogram equalisation (BBHE) [22]. However, these global enhancement approaches do not target particular regions for enhancement. For example, dark regions should be treated with priority compared to those with sufficient object details.

Unlike the above-mentioned contrast enhancement methods, the Retinex-based method [17] performs the joint illumination adjustment and noise removal by decomposing the captured image into different reflectance regions and their corresponding illumination components. This study generates high-quality output by processing reflectance and illumination. Other variations include the single-scale Retinex (SSR) [23], the multi-scale Retinex (MSR) [24] and the robust Retinex [25, 26], all having the potential to adjust the illumination and remove the noises. However, these methods may also yield over-enhancement or under-enhancement due to simple and single constraints, resulting in unnatural outputs with intensive noise artifacts.

With the rapid emergence of the computing device and neural network theory, learning-based methods have proven their excellent learning ability in image reconstruction and enhancement. This is primarily due to the more sophisticated loss function than the Euclidean distance, prone to produce blurring results. The LLNet [13] was the first to have introduced one auto-encoder for the low-light image enhancement. Inspired by the Retinex theory, the MSR-net [27] was then proposed to learn an end-to-end mapping between dark and bright images. Motivated by image components' decomposition and illumination, the RetinexNet [17] proposed two networks for decomposition and relight and learn the key constraints between decomposition and illumination maps. To further remove the noises, the RetinexNet added the joint denoising module. More recently, Chen et al. [15] introduced a universal pipeline for low-light

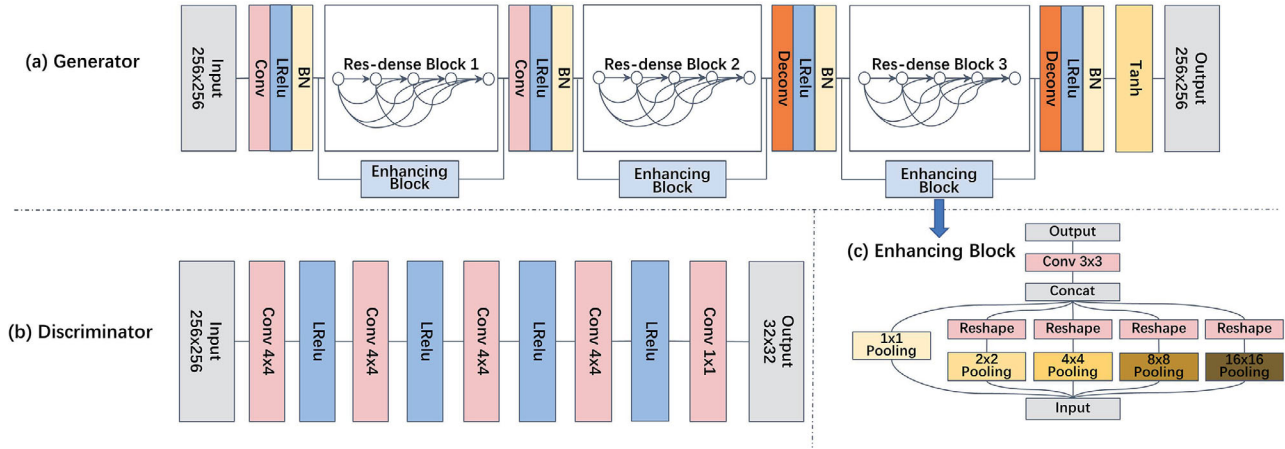


FIGURE 1 Overview of our network architecture

image processing based on the end-to-end training of a fully convolutional network. Despite its effectiveness with the RAW sensor data, this pipeline could not be applied to more generic and publicly available dataset.

### 3 | PROPOSED METHODS

In this section, we firstly discuss the formulation of the low-light image, then the overall architecture of the proposed densely residual generative adversarial network (DRGAN). Finally, we detail the loss function design to resolve the limitation of simple constraint problem in the training process.

#### 3.1 | Low-light image formulation

In order to understand the details and resolve the low-light image enhancing problem, Guo et al. [28] introduced the initial definition of low-light image as follows:

$$L(x) = R(x) \circ T(x), \quad (1)$$

where  $L(x)$  and  $R(x)$  denotes the degraded and the original image, respectively. And  $T(x)$  is the illumination map to encode the light intensity condition with the  $\circ$  a pixel-wise multiplication operator. Hence, we formulated the problem of low-light image enhancement to be the estimation of the non-linear degradation function between the normal- and low-light images. The main goal here is to accurately simulate the mapping function to recover the original image  $R(x)$ .

#### 3.2 | Overview of network architecture

Classical computer vision algorithms, such as image denoising, de-blurring and super-resolution, are all inclined to use conventional CNNs architecture module to achieve image enhancement and reconstruction. However, these existing methods

usually only consider the pixel mapping between low-light image and corresponding ground truth (GT) while ignoring the similarities in the feature level. Based on these researches, we attempt to convert these methods into the GANs model. The primary purpose is to generate high-quality and robust features to reconstruct the degraded images. It is proven that deep learning networks containing such modules have excellent performances for image reconstruction tasks. With insufficient training datasets at hand, we adopt the GANs module to increase the volume and diversity of the training images. The whole architecture is similar to the standard GANs network, one generator and its corresponding discriminator. Nevertheless, we adapt these modifications to improve the performance of this architecture and employ the whole resnet-based architecture for the generator, whose details are shown in Figure 1.

##### 3.2.1 | Densely residual block (DRB)

In light of the huge successes of almost all CNN-based algorithms [19], we adopt the densely connected scheme and residual strategy to design a novel feature generator, to combine the advantages of both standard CNNs and GANs. Specifically, the generator has a modular architecture composed of three DRB, and each block consists of five convolutional layers with densely skip connections, as shown in Figure 1(a). Each convolutional layer has a  $3 \times 3$  kernel.

##### 3.2.2 | Enhancement block (EB)

To further improve the diversity of extracted features, we introduce the enhancement block (EB), illustrated in Figure 1(c), to extract intermediate features with different scales. Specifically, the EB can extract multi-scale features from low-level edge features to high-level semantic features. And the initial motivation of this strategy is to establish a connection between the local patch and the global contents. We expect to improve the feature representation capacity with the effective fusion of multi-scale

features. The block receives five features processed by an average pooling layer with pooling sizes of 1/2, 1/4, 1/8 and 1/16, respectively. Then, we concatenate these features as input of the convolutional layer with a  $3 \times 3$  kernel. Afterwards, we alter the filter size and padding to align the input and output matrices to avoid the overlapping and grid from the de-convolution and up-sampling operations.

### 3.2.3 | Discriminator

Inspired by [29], we propose to remove the batch normalisation to improve computing efficiency, as shown in Figure 1(b). Indeed, WGAN-GP [30] alters the norm of the gradient of the discriminator with respect to each input, invalidating the batch normalisation. Therefore, the proposed discriminator follows the basic structure of PatchGAN [31] without batch normalisation. Furthermore, we introduce one binary scale value (either real or fake), and the discriminator produces a corresponding  $32 \times 32$  feature matrix to represent the result from the perspective of high level. Consequently, the discriminator could differentiate images at the feature patch level.

### 3.2.4 | Loss function

Recovering high-quality images with high contrast and chromatic richness from low-light images is a highly ill-posed problem, in which the design of appropriate loss function is often essential. The better loss function is supposed to constrain the training process to ensure optimal network training. In the following, we will present each component's effect in the joint loss function and illustrate the contributions in producing sharper edges and more detailed textures. In the optimisation process, the proposed joint loss  $\mathcal{L}_{RDGAN}$  consists of the GAN loss, the perceptual loss  $L_{per}$  and the contextual loss  $L_{CX}$  as follows:

$$\mathcal{L}_{RDGAN} = \mathcal{L}_{Gan} + \lambda_p \mathcal{L}_{per} + \lambda_c \mathcal{L}_{CX}. \quad (2)$$

#### Gan loss

Recently, the relativistic discriminator structure has been widely adopted in several researches [32]. This function estimates the probability that real data is more realistic than fake data, and also directs the generator to synthesise a fake image that is more realistic than the real images. The definition writes:

$$D_{Ra}(x_r, x_f) = \sigma \left( C(x_r) - \mathbb{E}_{x_f \sim \mathbb{P}_{fake}} [C(x_f)] \right), \quad (3)$$

and

$$D_{Ra}(x_f, x_r) = \sigma \left( C(x_f) - \mathbb{E}_{x_r \sim \mathbb{P}_{real}} [C(x_r)] \right), \quad (4)$$

where  $C$  indicates the discriminator,  $x_r$  and  $x_f$  are samples selected from the real  $\mathbb{P}_{fake}$  and fake distribution  $\mathbb{P}_{fake}$ . And  $\sigma$  represents activation function.

For the discriminator, we employ the relativistic discriminator and take the least square GAN (LSGAN) [33] as the activation function.

Thus, the  $\mathcal{L}_{Gan}$  is the sum of  $\mathcal{L}_G$  (generator loss) and  $\mathcal{L}_D$  (discriminator loss) as defined by:

$$\begin{aligned} \mathcal{L}_D = & \mathbb{E}_{x_r \sim \mathbb{P}_{real}} \left[ (D_{Ra}(x_r, x_f) - 1)^2 \right] + \\ & \mathbb{E}_{x_f \sim \mathbb{P}_{fake}} \left[ D_{Ra}(x_f, x_r)^2 \right], \end{aligned} \quad (5)$$

and

$$\begin{aligned} \mathcal{L}_G = & \mathbb{E}_{x_f \sim \mathbb{P}_{fake}} \left[ (D_{Ra}(x_f, x_r) - 1)^2 \right] + \\ & \mathbb{E}_{x_r \sim \mathbb{P}_{real}} \left[ D_{Ra}(x_r, x_f)^2 \right], \end{aligned} \quad (6)$$

where  $D$  indicates the discriminator,  $x_r$  and  $x_f$  are samples selected from the real  $\mathbb{P}_{fake}$  and fake distribution  $\mathbb{P}_{fake}$ , respectively.

#### Perceptual loss

To obtain realistic images and preserve the semantic details properly, we introduce the perceptual loss [34] based on the pre-trained VGG features to constrain the brighter region with rich structured features, which is defined as

$$\mathcal{L}_{per} = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W \|\phi_H(x_g)_{i,j} - \phi_H(x_c)_{i,j}\|_2^2, \quad (7)$$

where  $\phi_H(\cdot)$  represents the feature extractor and  $\phi_H(\cdot)_{i,j}$  indicates pixel in the  $i$ th column and  $j$ th row of the network feature, each of which is of size  $H \times W$ . In this study, we adopt the VGG-19 network pre-trained on the ImageNet [35] as a feature extractor. The perceptual loss function  $\mathcal{L}_{per}$  is designed to measure the differences between images in the feature space instead of the pixel space and guide the training process on the semantic level.

#### Contextual loss

Contextual loss [36, 37] has been recently studied to improve the visual quality of generated images in the GANs network. For image style transfer and super-resolution, the main purpose is to establish the similarity between the input and the desired target. The available strategies include both pixel and global content loss, such as mean square error (MSE) and perceptual loss. However, the pixel loss function constrains the model per paired pixel between the input and the ground truth, probably resulting in over-smoothing, while the content loss is unregulated in the local patch and cannot preserve the details in the generated image. Therefore, the contextual loss focuses on the similarity between the features regardless of the spatial positions.

We aim at targeting the darker regions with low illumination, with responses across multiple channels by spatial weighting feature maps. Specifically, minimising the differences between

the weighted low-level feature maps should improve the perceptual quality of brightness region in the enhanced outputs. Therefore, the loss is defined as by:

$$\mathcal{L}_{CX}(x, y, l) = -\log(CX(\Phi^l(x), \Phi^l(y))), \quad (8)$$

where  $x$  denotes the input image and  $y$  the target image, and  $CX$  the similarity measures the features maps  $\Phi^l(x)$  and  $\Phi^l(y)$  from the  $l$ th layer in the perceptual network VGG19  $\Phi(\cdot)$ . Note that the similarity is measured by the sum of regions with the same objects invariant of the corresponding spatial locations. Overall, a pair of images is considered similar when most features of one image can also be found in the other. Hence, the contextual similarity function  $CX$  could be defined as follows:

$$CX(X, Y) = \frac{1}{N} \sum_j \max_i CX_{ij}. \quad (9)$$

Then, we detail the similarity definitions between features. The loss function relies on the cosine distance, noted as  $d_{ij}$ , between the feature  $x_i$  and  $y_j$ :

$$d_{ij} = \left( 1 - \frac{(x_i - \mu_x) \cdot (y_j - \mu_y)}{\|x_i - \mu_x\|_2 \|y_j - \mu_y\|_2} \right) \text{ where } \mu_x = \frac{1}{N} \sum_i x_i, \mu_y = \frac{1}{N} \sum_j y_j. \quad (10)$$

When  $d_{ij} \ll d_{ik}$ , we assume that features  $x_i$  and  $y_j$  have similar contexts. To simplify the calculation, the cosine distance is normalised as follows:

$$\tilde{d}_{ij} = \frac{d_{ij}}{\min_k d_{ik} + \epsilon}, \quad (11)$$

with  $\epsilon = 1e - 5$ . Using an exponential operation, we transformed the distance into similarity:

$$w_{ij} = \exp\left(\frac{1 - \tilde{d}_{ij}}{b}\right), \quad (12)$$

where we set  $b = 0.5$ . Hence, the normalised similarity to define the contextual similarity between features is as follows:

$$CX_{ij} = w_{ij} / \sum_k w_{ik}. \quad (13)$$

The main objective of this loss function is to guide the model to generate images with natural image feature distribution. Hence, the function measures the differences in each spatial location feature map per channel.

## 4 | EXPERIMENTAL VALIDATION

In this section, we discuss the dataset for synthetic low-light image and the detailed setups of the proposed method. Then,

**TABLE 1** Quantitative performance comparison of our method with those state-of-the-arts on LOL dataset [17] by all metrics. (w/o means without and **Red** and **blue** indicate the best and the second best performance, respectively)

Methods	PSNR	SSIM	NIQE
BIMEF [38]	13.7891	0.6386	7.7684
LIME [28]	17.0994	0.5491	8.5237
LECARM [39]	14.2233	0.5789	8.0693
RetinexNet [17]	17.0921	0.4956	9.3127
EnlightenGAN [40]	<b>17.1891</b>	<b>0.6761</b>	<b>4.8344</b>
Zero-DCE [41]	14.5370	0.6067	8.0894
Ours w/o $\mathcal{L}_{per}$	17.5062	0.7729	4.2569
Ours w/o $\mathcal{L}_{CX}$	17.6245	0.7609	3.7228
Ours	<b>18.0224</b>	<b>0.7784</b>	<b>3.7959</b>

we present the performances of our DRGAN in comparison with the reference state-of-the-art methods for several image quality evaluation metrics. Finally, we conclude the ablation of losses in this model and compare the performances in face detection, a high-level visual task.

### 4.1 | Dataset collection and implementation details

#### 4.1.1 | Synthetic low-light image

Our method is conducted by 30K paired images, that is, low-light and bright, synthesised by VOC2007 dataset. Each low-light image is randomly generated from the original image and the non-linear degradation function by the following simulation method:

$$P_{image} = F(X + G(\sigma)), \quad (14)$$

where  $F(\cdot)$  represents the gamma adjustment function and  $G(\cdot)$  the noise component with the given standard deviation  $\sigma$ . Random gamma darkening with controlled noise levels allows to generate a huge variety of synthetic training images to validate the robustness of the whole model. Specifically, we adopt the additive Gaussian noise in the synthetic images to model the noises in the camera shooting process.

However, synthetic images cannot completely replace the role of real-life low-light image data. To fully evaluate the performances of the proposed method, we also include images from various scenes from the LOL [17] and the Exdark [2] datasets in compression experiments. The LOL dataset is used for objective and subjective evaluations since it includes highly degraded images for which most methods cannot achieve promising results. And the ExDark dataset consists of 7363 low-light images with annotation of 12 object classes. Due to the relative small volumes of the datasets, such as NPE [16], MEF



**FIGURE 2** Visual comparison from the loss ablation study. (b)–(d) demonstrates the effectiveness of each component ( $L_{per}$  and  $L_{CX}$ ) in the whole loss function (a) and (e) represent the original input and ground truth (GT)



**FIGURE 3** Visual comparison with state-of-the-art methods on the LOL [17] dataset

**TABLE 2** Quantitative performance comparison of ours with state-of-the-arts on Exdark by NIEQ metric. (Red and blue indicate the best and the second best performance, respectively)

Methods	Bicycle	Boat	Bottle	Bus	Car	Cat	Chair	Cup	Dog	Motorbike	People	Table	AVG
LIME [28]	3.9576	4.0194	4.2548	3.8171	3.9347	4.5830	4.1763	4.2295	4.2985	4.1775	4.2356	3.9695	4.1378
LECARM [39]	3.7934	3.9805	4.2371	3.7233	3.9653	4.5598	4.1312	4.2945	4.2415	4.0095	4.1941	3.9429	4.0894
BIMEF [38]	<b>3.6109</b>	3.9547	4.0674	<b>3.5739</b>	3.9241	4.5547	4.0647	4.2360	4.1619	<b>3.7907</b>	4.0049	3.9082	3.9877
RetinexNet [17]	4.5941	4.5034	4.5482	4.4448	<b>3.3964</b>	4.8918	4.5698	4.3482	4.7851	4.5616	4.7954	4.2984	4.5614
EnlightenGAN [40]	<b>3.6415</b>	<b>3.8128</b>	4.0143	3.6750	3.8546	<b>4.0936</b>	<b>3.8371</b>	4.0704	<b>3.9455</b>	4.0197	3.9267	<b>3.7987</b>	<b>3.8908</b>
Zero-DCE [41]	3.6628	3.9176	<b>3.9600</b>	3.5882	3.7837	4.5695	3.8696	4.0639	4.0681	<b>3.7070</b>	<b>3.8165</b>	<b>3.6962</b>	3.8919
Ours*	3.7201	<b>3.6022</b>	<b>3.9843</b>	<b>3.5550</b>	<b>3.6046</b>	<b>4.2363</b>	<b>3.8071</b>	<b>4.0104</b>	<b>3.9055</b>	3.8197	<b>3.7339</b>	4.1626	<b>3.8451</b>

[42], we need to make sure the robustness and scalability of methods.

## 4.2 | Implementation details

For hyper parameters  $\lambda_p, \lambda_c$  in the loss function, we empirically use 0.5 and 0.5 to weight the component adopted in whole function. All convolutional layer kernels are set to  $3 \times 3$  in size except in the *EB*, where the  $1 \times 1$ ,  $3 \times 3$  and  $5 \times 5$  kernels are used to extract multiple feature, following concentration to rebuild the original dimension by one  $1 \times 1$  convolutional layer. Specifically, we trained all models for 200 epochs with a batch-size 16, and the loss was minimised using the Adam [43] optimiser with a learning rate of  $10^{-4}$ . And we adopt the TensorFlow [44] libraries to implement the proposed network with two NVIDIA GeForce GTX 1080TI GPUs for computing acceleration.

## 4.3 | Performance evaluation

To assess the performance of the proposed model, we adopt three metrics for quantitative comparisons, divided into referenced and non-referenced metrics. Besides, we take the state-of-the-art methods of BIMEF [38], LIME [28], LECARM [39], RetinexNet [17], EnlightenGAN [40] and Zero-DCE [41] as the references.

### 4.3.1 | Referenced metrics

Two standard metrics are adopted to investigate the performances of the enhancement, namely the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM). The PSNR approximates the reconstruction quality of a generated image  $x$  compared to the corresponding GT  $y$  based on the Mean Squared Error(MSE) as follows:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2, \quad (15)$$

**TABLE 3** Quantitative performance comparison of ours with state-of-the-arts on DARK FACE dataset [18]. (Red and blue indicate the best and the second best performance, respectively)

Methods	Raw	LIME [28]	RetinexNet [17]
AP	0.2348	0.3070	<b>0.3087</b>
NIQE	4.5391	3.8960	5.3313
Methods	EnlightenGAN [40]	Zero-DCE [41]	Ours
AP	0.2952	<b>0.3111</b>	0.2958
NIQE	<b>2.9591</b>	3.7444	<b>3.2272</b>

$$PSNR = 10 \cdot \log_{10} \left( \frac{\max(I)^2}{MSE} \right). \quad (16)$$

Here,  $\max(I)$  is the maximum possible pixel value of the image  $I$ . On the other hand, the SSIM measures the image patches based on three properties: luminance, contrast, and structure. The metric is formulated as follows:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (17)$$

where  $\mu_x$  and  $\mu_y$  denotes the mean,  $\mu_x^2$  and  $\mu_y^2$  the variance of  $x$  and  $y$ , respectively. And the  $\sigma_{xy}$  denotes the cross-correlation between  $x$  and  $y$ . We fix  $c_1 = (255 \times 0.01)^2$  and  $c_2 = (255 \times 0.03)^2$  to ensure numeric stability.

### 4.3.2 | Non-referenced metric

To address the limitation of ground truth image in the ExDark dataset, non-referenced evaluation methods are also needed. We adopt the Natural Image Quality Evaluator(NIQE) to examine the performance differences among the compared methods.

Tables 1 and 2 report the numerical results among the competitors on the LOL and the ExDark dataset individually. For the test image in LOL dataset, we investigate both of referenced metric and non-referenced metric, while in the ExDark only the non-referenced metric is compared. Firstly, Table 1 compares the numerical results among the competitors on LOL dataset. In



**FIGURE 4** Visual comparison with state-of-the-art methods on the Exdark [2] dataset

this dataset, each low-light image has its corresponding normal-light image, and we take both referenced and non-referenced metrics. Obviously, we conclude that the proposed model significantly outperforms all the other reference methods in all the metrics. It can be noticed that these traditional methods, including LIME, LECARM, and BIMEF, generate huge random noises in several scenes, while the CNN-based or GAN-based methods, including our DRGAN, RetinexNet, Enlight-

enGAN and Zero-DCE, could effectively overcome this issue. in these scenes.

For the Exdark with a larger number (7363) of low-light images and high diversity of exclusive light conditions, Table 2 illustrates the non-referenced metric NIQE scores. The proposed network demonstrates clear advantages over the others while showing some slight weaknesses in several classes, that is, Bicycle, Car, Cat, Motorbike and Table. Furthermore, the



Zero-DCE, EnlightenGAN and BIMEF are comparable in the total average score and have the best performances in certain categories. Overall, our DRGAN performs significantly better compared with competitors.

Figures 2 and 3 illustrate the visual comparisons on some selected images from the LOL and the Exdark datasets. We can notice that most of the methods brighten the low-quality images. However, severe distortions exist due to inappropriate light adjustment, obstinate noise artifacts and colour alterations. For instance, the results from the RetinexNet induce significant noises while the EnlightenGAN and Zero-DCE could not enhance effectively in several extreme dark regions with low noise levels. By contrast, the proposed method outperforms in these cases and recovers the darker regions more successfully. The edge preservation and noise rejection results both corroborate the superiority of our method.

#### 4.4 | Ablation study

Figure 4 presents the ablation study results to show the effects of each component,  $L_{per}$  and  $L_{CX}$ , as part of the loss function. We can clearly observe that the results without  $L_{per}$  has relatively lower contrasts and model removing the  $L_{CX}$  fails to recover the colour variations, and the contextual details. The results in Figure 2 regulated by all the loss components contain clearer details and higher contrasts, especially in the zoomed regions. By introducing the joint loss function, the network keeps focus on the local patches in order to recover the details, such as edges and smaller objects. Hence, we could conclude that both loss components have played a significant role in the proposed model. In addition, Table 1 presents the loss component ablation results from the image metrics point of view.

#### 4.5 | Analysis: Face detection in the dark

To further analyse the effect brought by low-light enhancement methods, we also investigate the face detection task as an extra experimental task. Firstly, we take the Dark Face dataset [18], with over 10,000 images in low-light conditions, as a testing dataset to measure the performances. Secondly, the Dual Shot Face Detector(DFSD) [45], trained on the Wider Face dataset [46], is used as the baseline model. Finally, to guarantee fair comparison, we select 1000 images from the train set in the Dark Face and feed the enhanced results by the above methods to the baseline. Furthermore, we examine the performances by the average precision (AP), shown in the precision-recall (P-R) curves in Figure 5. We also add the AP curve from the standard toolkit provided in the Dark Face dataset [18].

Overall, the precision of DSFD increases considerably compared to that using only the original low-light images, which means the enhancing methods play critical roles in improving the precision in the high-level task of face detection. The Dce-

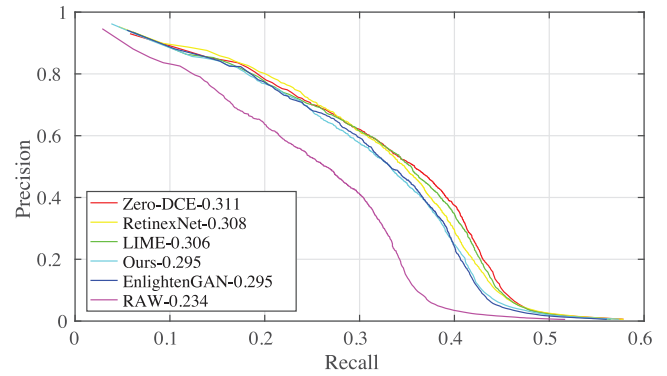


FIGURE 5 The performance of face detection in the DARK FACE [18], contains the P-R curves and AP

zero and the RetinexNet perform the best with the AP metric but neither could achieve high scores in pixel-measured metrics, as computed in Table 3. The major reason is that these enhancing methods introduce noise artifacts during the enhancement and significantly reduce the performances in image quality evaluation metrics.

By contrast, the EnlightenGAN and proposed method achieve the best performances in these quality metrics but are insufficient in the task of face detection. This can be explained by the fact that both methods are based on the GAN that might introduce additive features to distort the original ones and thus interfere with the detectors.

As a general rule, higher performances in pixel-wise metrics cannot guarantee better results in high-level visual tasks.

## 5 | CONCLUSIONS AND FUTURE WORKS

In this work, we proposed a deep network for low-light image enhancement with the objective of information retrieval instead of physical restoration. We make several adaptations in the loss function design and basic architecture to establish a robust connection between the local patches and global contents. Experimental results demonstrate the superiority of the proposed enhancement method and show competitive performances over existing light enhancement methods, both qualitatively and quantitatively. In future work, we intend to exploit the more effective low-light enhancement frameworks via unsupervised learning, to reduce the dependency for paired training data. Besides, limiting the interferences brought by generated feature is an interesting topic, to improve the performance measured by both pixel-wise metrics and high-level visual tasks.

### ACKNOWLEDGEMENTS

This work is sponsored by Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University (CX201959) and Synergy Innovation Foundation of the University and Enterprise for Graduate Students at Northwestern Polytechnical University (XQ201910). This work is also

supported in part by the [National Natural Science Foundation of China](#) under Grant 61972321.

## CONFLICT OF INTEREST

The authors declare that they do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

Fei Li  <https://orcid.org/0000-0002-4184-7203>

## REFERENCES

- Zhang, Y., et al.: Kindling the darkness: A practical low-light image enhancer. In: Proceedings of the 27th ACM International Conference on Multimedia. ACM, New York (2019)
- Loh, Y. P., Chan, C. S.: Getting to know low-light images with the exclusively dark dataset. *Comput. Vis. Image Underst.* 178, 30–42 (2019)
- Jiang, L., et al.: Deep refinement network for natural low-light image enhancement in symmetric pathways. *Symmetry* 10(10), 491 (2018)
- Gala, A., Shah, S.: A survey of approaches and trends in person re-identification. *Image Vis. Comput* 32, 270–286 (2014)
- Rashed, H., et al.: FuseMODNet: Real-time camera and LiDAR based moving object detection for robust low-light autonomous driving. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pp. 2393–2402. IEEE, Piscataway, NJ (2019)
- Lee, U., et al.: Development of a self-driving car that can handle the adverse weather. *Int. J. Automot. Technol.* 19(1), 191–197 (2018)
- Khan, M., et al.: An implementation of optimized framework for action classification using multilayers neural network on selected fused features. *Pattern Anal. Appl.* 22, 1377–1397 (2019)
- Khan, M., et al.: License number plate recognition system using entropy-based features selection approach with SVM. *IET Image Proc.* 12, 200–209 (2018)
- Sharif, M., et al.: Human action recognition: A framework of statistical weighted segmentation and rank correlation-based selection. *Pattern Anal. Appl.* 23, 281–294 (2019)
- Gehler, P. V., et al.: Bayesian color constancy revisited. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE, Piscataway, NJ (2008)
- Kang, S. B., et al.: Personalization of image enhancement. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1799–1806. IEEE, Piscataway, NJ (2010)
- Hwang, S. J., et al.: Context-based automatic local image enhancement. *European Conference on Computer Vision*, pp. 569–582. Springer, Berlin Heidelberg (2012)
- Lore, K. G., et al.: Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognit.* 61, 650–662 (2017)
- Cheng, H. D., Shi, X. J.: A simple and effective histogram equalization approach to image enhancement. *Digital Signal Process.* 14(2), 158–170 (2004)
- Chen, C., et al.: Learning to see in the dark. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3291–3300. IEEE, Piscataway, NJ (2018)
- Wang, S., et al.: Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Trans. Image Process.* 22(9), 3538–3548 (2013)
- Wei, C.: Deep Retinex decomposition for low-light enhancement. *arXiv:1808.04560* (2018)
- Yuan, Y., et al.: UG<sup>2+</sup> Track 2: A collective benchmark effort for evaluating and advancing image understanding in poor visibility environments. *arXiv:1904.04474* (2019)
- Huang, G., et al.: Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269. IEEE, Piscataway, NJ (2017)
- Abdullah-Al-Wadud, M., et al.: A dynamic histogram equalization for image contrast enhancement. *IEEE Trans. Consum. Electron.* 53(2), 593–600 (2007)
- Reza, A.: Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. *J. VLSI Signal Process. Syst. Signal Image Video Technol.* 38(1), 35–44 (2004)
- Kim, Y.-T.: Contrast enhancement using brightness preserving bi-histogram equalization. *IEEE Trans. Consum. Electron.* 43, 1–8 (1997)
- Jobson, D. J., et al.: Properties and performance of a center/surround retinex. *IEEE Trans. Image Process.* 6(3), 451–462 (1997)
- Rahman, Z., et al.: Multi-scale retinex for color image enhancement. In: Proceedings of 3rd IEEE International Conference on Image Processing, vol. 3, pp. 1003–1006. IEEE, Piscataway, NJ (1996)
- Li, M., et al.: Structure-revealing low-light image enhancement via robust retinex model. *IEEE Trans. Image Process.* 27(6), 2828–2841 (2018)
- Ren, X., et al.: Joint enhancement and denoising method via sequential decomposition. In: 2018 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1–5. IEEE, Piscataway, NJ (2018)
- L. Shen, et al.: MSR-net: Low-light image enhancement using deep convolutional network. *arXiv:1711.02488* (2017)
- Guo, X., et al.: Lime: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.* 26(2), 982–993 (2017)
- Radford, A., et al.: Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv:1511.06434* (2016)
- Gulrajani, I., et al.: Improved training of Wasserstein GANs. *arXiv:1704.00028* (2017)
- Isola, P., et al.: Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5967–5976. IEEE, Piscataway, NJ (2017)
- Jolicoeur-Martineau, A.: The relativistic discriminator: A key element missing from standard GAN. *arXiv:1807.00734* (2018)
- Mao, X., et al.: Least squares generative adversarial networks. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2813–2821. IEEE, Piscataway, NJ (2017)
- Johnson, J., et al.: Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision, pp. 694–711. Springer, Cham (2016)
- Krizhevsky, A., et al.: Imagenet classification with deep convolutional neural networks. *Commun. ACM* 60(6), 84–90 (2017)
- Roey, I.R., et al.: The contextual loss for image transformation with non-aligned data. *arXiv:1803.02077* (2018)
- Mechrez, I. R., et al.: Maintaining natural image statistics with the contextual loss. In: Asian Conference on Computer Vision, pp. 427–443. Springer, Cham (2018)
- Ying, Z., et al.: A bio-inspired multi-exposure fusion framework for low-light image enhancement. *arXiv:1711.00591* (2017)
- Ren, Y., et al.: LECARM: Low-light image enhancement using the camera response model. *IEEE Trans. Circuits Syst. Video Technol.* 29(4), 968–981 (2019)
- Jiang, Y., et al.: EnlightenGAN: Deep light enhancement without paired supervision. *arXiv:1906.06972* (2019)
- Guo, C., et al.: Zero-reference deep curve estimation for low-light image enhancement. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1777–1786. IEEE, Piscataway, NJ (2020)
- Lee, C., et al.: Power-constrained contrast enhancement for emissive displays based on histogram equalization. *IEEE Trans. Image Process.* 21, 80–93 (2012)
- Kingma, D.P., Adam, J.B.: Adam: A method for stochastic optimization. *arXiv:1412.6980* (2015)
- Abadi, M., et al.: Tensorflow: Large-scale machine learning 467 on heterogeneous distributed systems. *arXiv:1603.04467* (2016)

45. Li, J., et al.: DSFD: Dual shot face detector. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5055–5064. IEEE, Piscataway, NJ (2019)
46. Yang, S., et al.: Wider face: A face detection benchmark. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5525–5533. IEEE, Piscataway, NJ (2016)

**How to cite this article:** Li F, Zheng J, Zhang Yuan-fang. Generative adversarial network for low-light image enhancement. *IET Image Process.* 2021;15:1542–1552.  
<https://doi.org/10.1049/ipr2.12124>