Article

Contextual Information Aided Generative Adversarial Network for Low-Light Image Enhancement

Shiyong Hu , Jia Yan and Dexiang Deng *

School of Electronic Information, Wuhan University, Wuhan 430072, China; husy@whu.edu.cn (S.H.);
yanjia2011@gmail.com (J.Y.)

* Correspondence: ddx@whu.edu.cn

Abstract: Low-light image enhancement has been gradually becoming a hot research topic in recent years due to its wide usage as an important pre-processing step in computer vision tasks. Although numerous methods have achieved promising results, some of them still generate results with detail loss and local distortion. In this paper, we propose an improved generative adversarial network based on contextual information. Specifically, residual dense blocks are adopted in the generator to promote hierarchical feature interaction across multiple layers and enhance features at multiple depths in the network. Then, an attention module integrating multi-scale contextual information is introduced to refine and highlight discriminative features. A hybrid loss function containing perceptual and color component is utilized in the training phase to ensure the overall visual quality. Qualitative and quantitative experimental results on several benchmark datasets demonstrate that our model achieves relatively good results and has good generalization capacity compared to other state-of-the-art low-light enhancement algorithms.

Keywords: generative adversarial network; attention mechanism; contextual information

Citation: Hu, S.; Yan, J.; Deng, D. Contextual Information Aided Generative Adversarial Network for Low-Light Image Enhancement. *Electronics* **2022**, *11*, 32. <https://doi.org/10.3390/electronics11010032>

Academic Editor: Chiman Kwan

Received: 20 November 2021

Accepted: 20 December 2021

Published: 23 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Nowadays, in a world where multimedia equipment is much more easily accessible, images and videos have become the most ubiquitous ways to convey and record information. However, encountering imperfect capturing conditions, such as insufficient lighting and short exposure time, is inevitable, making the captured images suffer from detail loss and low brightness. Their practical value may be limited. Manual operations such as increasing exposure time and setting high ISO can mitigate the degradation. However, typical users may not quite have the required skills to operate the photograph device. Hence, enhancing contrast and details of low-light images using software algorithms is highly desirable, which can be beneficial to both low-level and high-level real-world image applications such as night imaging and video surveillance.

Low-light image enhancement has gradually become an active research topic in computer vision field in recent years. Current methods can be roughly classified into three categories: histogram-based algorithms, Retinex-based algorithms and data-driven algorithms.

An image histogram counts all pixels in an image and depicts the frequency distribution of pixel values. Histogram-based algorithms directly manipulate image histogram and stretch it to obey uniform distribution, which can widen the dynamic range of the low-light image. Some of them [1–4] adopted global histogram of input image to estimate the pixel transformation function, but they ignored the unevenly distributed darkness and might introduce over-exposure distortion in some areas of the enhanced result. To alleviate this problem, local histogram-based methods were proposed [5,6]. They used the local image histogram to make the entire transformation function more adaptive. Later on, some methods [7,8] adopted different restraints on image histogram to further consider correlation of adjacent regions and reduce overstretching distortion. While these methods

are intuitive and easily implemented, they still amplify noise hidden in dark areas and introduce some distortion in their results.

Retinex-based algorithms decompose the low-light image into illumination and reflectance components based on the Retinex theory [9]. According to this theory, an image is the dot product of these two components, and the reflectance component is always fixed with varied illumination under different scenes. Jobson et al. extended previous single-scale Retinex [10] to multi-scale Retinex [11] for simultaneous dynamic range compression and tonal rendition in enhancing process. Guo et al. [12] developed a structure-aware smoothing model to improve illumination estimation for image enhancement. Wang et al. [13] decomposed an image using a bright-pass filter and used bi-log transformation to improve the illumination and preserve image naturalness. Fu et al. [14] proposed a new weighted variational model to estimate reflectance with more details and suppress noise. All these methods require additional prior information on the decomposed illumination or reflectance component. Either the enhanced reflectance or reflectance combined with enhanced illumination component is treated as the final enhanced output. While Retinex-based algorithms have achieved promising results in many works, decomposing the input image into two components is an ill-posed problem in essence. Designing handcrafted prior information requires extra knowledge, which is cumbersome and cannot be embedded in an automatic system.

With numerous pieces of data and powerful computing resources easily accessible, recent years have seen remarkable progress in deep learning technology in many image processing and computer vision problems [15,16]. The number of works using deep neural networks on low-light image enhancement has been growing. According to whether paired training data are used, deep-learning-based methods can be further divided into three subclasses: supervised learning methods, unsupervised learning methods, zero-shot learning methods. Lore et al. [17] proposed a stacked sparse denoising autoencoder to adaptively brighten natural low-light images, which was also the pioneering work of learning-based methods for low-light image enhancement. Chen et al. [18] introduced a dataset of raw low-light images with corresponding reference images to develop an enhancing pipeline using a fully convolutional neural network (CNN). However, images in raw the data domain are less common in daily applications, reducing their practical value. Wei et al. [19] incorporated Retinex theory into the deep neural network and introduced RetinexNet for simultaneous decomposition and illumination adjustment. They also built a low-light image dataset (LOL) containing low- and normal-light pairs captured in real scenarios for the development of research community. Hua et al. [20] proposed a generative adversarial network (GAN) guided by image quality assessment (IQA) techniques to enhance low-light images. While supervised learning-based methods achieve decent performance, using training data may restrict their generalization capacity to a specific dataset. Hence, unsupervised learning-based methods have emerged and attracted lots of attention in research community. Jiang et al. proposed EnlightenGAN [21] for single image low-light enhancement without paired training data. This is the first work using GAN with an unsupervised learning strategy, which means researchers no longer need to collect paired low- and normal-light images in a real scene. Ni et al. [22] further extended EnlightenGAN with a global attention and modulation module for enhancing the aesthetic quality of images. Chen et al. [23] treated image enhancement as an image-to-image translation task and used a two-way GAN with some modifications and unpaired dataset. Apart from supervised and unsupervised learning-based methods, zero-shot learning based methods have flourished in recent works. Zhu et al. [24] introduced a novel three-branch CNN termed RRDNet to simultaneously perform denoising and restoring underexposed images without any prior image examples or training in advance. Different from the above methods, Guo et al. [25] formulated image enhancement as a task of image-specific curve estimation. This method does not require any paired or unpaired training data and directly estimates pixel-wise curve parameters to adjust the input brightness. Taking advantage of strong feature representation power, CNN can exploit relevant features from input image

or enough volumes of training data and automatically learn a non-linear mapping from a low-light to normal-light image. More powerful feature representation capacity usually requires a more complicated CNN structure, resulting in a large computational burden. While all the above methods have achieved promising performance, some results will still suffer from detail loss and noise amplification when images are taken under unbalanced lightness distribution.

In this paper, we propose an improved GAN for low-light image enhancement. We use U-net [26] as our backbone. U-Net has been applied to many image restoring tasks due to its encoder–decoder structure, which is beneficial to propagating semantic information across layers. To ease the gradient flow and boost feature propagation across different hierarchical layers, we build our encoder–decoder U-Net-like network using densely residual blocks. Previous works adopt attention mechanism [27,28] to make the enhancement model focus on important regions, but they all focus on spatial or channel dimension and ignore multiscale context information. Context information is important for image restoration tasks, where pixel-to-pixel correspondence is learned from input image to output image. In image restoration tasks, removing degraded image content and preserving desired spatial details can be realized by enlarging receptive field, leading to richer context information [29]. Thus, to reduce unwanted distortion in the enhanced results and integrate more context information in multiple scales, we introduce multiscale context attention module (MSCAM) to better refine feature maps and capture more salient features. To mitigate color deviation in the enhanced result, we further use a hybrid loss incorporating a perceptual loss in feature domain and a color-related component computing in HSV color space during the training phase. Compared to traditional techniques, our method can directly learn mappings from low-light images to normal-light images without heavy regularization design and manually selected parameters, and produce robust, visually pleasing results fast. Additionally, like previous works [30,31] using GAN in practical application, our method can be also helpful in a realistic visual monitoring system, such as enhancing visibility for face detection in the dark [25].

The main contributions of this paper are listed as follows:

- (1) We design a GAN framework built upon encoder–decoder structure using residual dense blocks for better gradient backpropagation and feature fusion across different layers, which plays an import role in image enhancement.
- (2) We introduce a novel and lightweight MSCAM module to further refine feature maps using context information in multiple scales. The MSCAM can make our model enlarge receptive fields and highlight important features, which can reduce the enhancement distortion. We train our GAN framework with the joint perceptual and color-related loss to mitigate color deviation and detail loss in the enhanced results.
- (3) We validate our method on several benchmark datasets. Experimental results demonstrate the superiority against many state-of-the-art methods qualitatively and quantitatively.

The remainder of this paper is organized as follows. Some related topics of this work are introduced in Section 2. We describe our proposed method in detail in Section 3. Experimental results and related analysis are presented in Section 4. Section 5 concludes this work.

2. Related Work

2.1. Generative Adversarial Network

GAN, proposed by Goodfellow et al. [32], has attracted lots of attention and become the hottest research topic in the deep learning community. GAN is comprised of two forward networks, i.e., a discriminator and a generator. The generator is trained to produce fake images indistinguishable from real images to fool the discriminator, while the discriminator is trained to differentiate fakes images from real images as much as possible. The two networks have completely opposite training targets, and their relationship can be viewed as

a minimum–maximum game in which they compete. To minimize (maximize) the training loss for the generator (discriminator), the adversarial loss can be formulated as follows:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

where D and G denote the discriminator and generator, and z and x denote random noise vector and real image, respectively. $p_z(z)$ represents sampling from random noise, and $p_{data}(x)$ represents sampling from real world.

Recently, GAN has been widely applied in many low-level image restoration applications. Ledig et al. [33] utilized GAN to recover fine texture details and infer photo-realistic natural images together using a novel perceptual loss during the training process. Jiang et al. [21] constructed an unsupervised GAN framework to perform low-light image enhancement without paired training samples. In addition, they used self-regularized attention mechanism and double discriminators, i.e., local and global discriminators, to handle the unevenly distributed lighting in the input image. All these works proved that GAN has great potential in low level image processing, and encoder–decoder structure in the generator plays a pivotal role for feature representation.

2.2. Attention Mechanism

The attention mechanism is inspired by the fact that human eyes can quickly scan the global image content to obtain target areas that need to be focused on and pay more attention to them [34]. It has been widely applied in various computer vision tasks, such as image classification [35] and image restoration [36]. Wang et al. [37] introduced non-local operation into CNN to capture distant information for spatial attention. Hu et al. [38] proposed a squeeze-and-excitation block to adaptively recalibrate important information along channel axes. In the field of low-light image enhancement, Atoum et al. [27] introduced color-wise attention map to provide auxiliary information for image enhancement. Lv et al. [28] proposed a fully CNN containing four subnets to perform brightness enhancement and denoising tasks with the guidance of two attention maps. While these works have demonstrated relatively good performance, they all capture spatial information within the same scale. In contrast, we sequentially apply the attention mechanism along both spatial and channel axes to refine feature maps, and also incorporate multiscale context information to alleviate distortion.

2.3. Dilated Convolution

Dilated convolution was first proposed for image semantic segmentation [39]. It uses dilated convolutions by plugging holes between adjacent locations in normal convolutional kernels, which can increase the size of receptive field without additional parameters and computing costs. Dilated convolution has been applied widely in the field of object detection [40]. In this paper, we further exploit multi-scale information hidden in feature maps by using multiple dilated convolutions with different dilated rates, which can adaptively aggregate contextual information without losing resolution.

3. Proposed Method

In this section, we elaborate on the detailed architecture and loss function of our proposed GAN for low-light image enhancement.

3.1. Overall Network Architecture

The overall architecture of our proposed generator is shown in Figure 1a. Our generator has adopted the encoder–decoder structure that has been proven effective in many image restoration tasks. The encoder extracts local and global information hidden in feature maps at different levels, such as image contrast and texture, with the increasing sizes of receptive fields. The decoder utilizes relevant information and up-samples the feature maps to produce the final enhanced result. ResNet, proposed by He et al. [15], has achieved great

breakthroughs in many low- and high-level vision problems. Its identity mapping based on short connection can effectively solve the gradient vanishing problem occurred in training deep networks. Later on, Huang et al. [41] proposed DenseNet by connecting current layer with all preceding layers directly using short connections, which is helpful to feature reuse and information flow. Inspired by the structure of ResNet and DenseNet, we insert local residual dense blocks into the generator of our proposed GAN. The detailed structure of local residual dense block is illustrated in Figure 2. It consists of five convolutional layers with batch normalization and Leaky rectified linear unit (LReLU) activation, and each latter layer directly connects with all preceding layers. X_I and X_O denote the input and output of local residual dense block; the intermediate output X_k after k -th convolutional layer can be formulated as:

$$X_k = \Phi([X_0, X_1, \dots, X_{k-1}]), k = 1, \dots, 4 \quad (2)$$

where Φ denotes the composite function of each convolutional layer, and $[X_0, X_1, \dots, X_{k-1}]$ refers to the concatenation of feature maps produced by all preceding layers and the input. To preserve feed-forward and ease the gradient back-propagation, local residual connection is added within each dense block. Hence, the final output of each local residual dense block is as follows:

$$X_O = \Phi([X_4, X_3, \dots, X_0]) + X_I \quad (3)$$

Concretely, in the encoder part, we insert one local residual dense block between the convolutional layer and the max pooling layer. Similarly, in the decoder counterpart, we insert one residual dense block after the up-sampling and convolutional layer. Hence, in our proposed generator, hierarchical features can be fully exploited and fused across multiple convolutional layers.

Figure 1. Overall architecture of our proposed method.

Figure 2. Detailed structure of local residual dense block.

At the same spatial resolution level, we introduce a short-path connection and insert our proposed MSCAM between the encoder and decoder part in order to remedy the information loss in the down-sampling process and further refine feature maps for better representation. Apart from the local residual and dense connections, inspired by DnCNN [42], we also adopt the global residual learning strategy to ease the training process.

The discriminator of our proposed GAN aims to judge whether the generated results can be distinguished from the real normal-light images. We adopt the structure of PatchGAN [43] without batch normalization as our discriminator, as shown in Figure 1b. The discriminator fully adopts a convolutional structure, mapping the input to an $N \times N$ matrix. Each element X_{ij} represents the probability that each patch within a receptive field of input image is true. The final output of the discriminator is the average value of all X_{ij} , representing whether the enhanced result is close to the real image.

3.2. Multi-Scale Context Attention Module

Previous works [21,27,28] adopt the attention mechanism to guide the enhancement process. However, local distortion may appear in their enhancement results. To reduce local distortion, non-local context information can be integrated into the model. Context information has been successfully applied to many works [40,44] due to the auxiliary information from surrounding content. The down-sampling operation in our generator will inevitably causes information loss with the spatial size of feature maps becoming smaller. To fully exploit hidden information and further refine feature maps in the generator, motivated by [40,45], we introduce MSCAM embedded at each spatial resolution level for better feature representation. As depicted in Figure 3, our MSCAM consists of two main parts: multiscale context spatial attention submodule (MCSA) and channel attention submodule.

Figure 3. Detailed structure of multiscale context attention module.

$X_i \in \mathbb{R}^{C \times H \times W}$ denote the feature map after each residual dense block at layer i ($i = 1, 2, 3, 4$), where C is the channel number, and H and W represent the height and

width of the feature map, respectively. Firstly, our MSCAM concatenates the output feature from the MSCAM located at the next layer X_O^{i+1} , and merge the two feature maps using a 1×1 convolutional operation to form a new feature map X . Then, the following two attention modules further exploit and generate a 2D spatially refined feature map and a channel-wise refined feature map separately, which encourage the generator to capture more relevant information on spatial and channel dimensions. The whole process can be expressed as follows:

$$X = \text{Conv}_{1 \times 1}(\text{Cat}(X_i, X_O^{i+1})) \quad (4)$$

$$X_{MC} = F_{MC}(X) \quad (5)$$

$$X_C = F_C(X_{MC}) = F_C(F_{MC}(X)) \quad (6)$$

where F_{MC} denotes the multiscale context spatial attention module, $X_{MC} \in \mathbb{R}^{C \times H \times W}$ represents the spatially refined feature maps, F_C represents the channel attention module, and $X_C \in \mathbb{R}^{C \times H \times W}$ represents the channel-wise refined feature maps. To facilitate convergence during the training process, we add residual connection to connect the input feature maps and the refined feature maps as follows:

$$X_O = X_C + X \quad (7)$$

We also duplicate the output feature maps, denoted as X_O^i , and transmit them to the MSCAM at previous shallow layer for cross-layer feature interaction. Note that for the deepest layer, only feature maps from the current layer are needed.

The detailed structure of the MCSA submodule is inspired by the fact that local details are easily noticed by human eyes over global content, and different distances can make human eyes focus on different ranges. Therefore, we adopt four convolutional layers with different dilated rates to excavate multiscale context information hidden in feature maps, aggregating non-local information of different scales into the generator. As is shown in Figure 4, four dilated convolutional layers are arranged in four branches, inspired by the structure proposed in [40]. In each branch, we use the spatial attention module to highlight important regions. Concretely, we apply both average pooling and max pooling operations along channel dimension on feature maps after dilated convolutional operators, producing two feature descriptors with the size of $H \times W \times 1$. After concatenating them on the channel dimension, we use a convolutional layer with kernel size of 7×7 to squeeze its dimension to $H \times W \times 1$. Finally, the sigmoid function is employed to produce the attention map $M_i (i = 1, 2, \dots, 4)$, which is then multiplied with the input feature map in each branch $X_i (i = 1, 2, \dots, 4)$ element wise. The results from four branches are concatenated to produce the final refined feature maps X_{MC} .

Apart from spatial attention mechanism that captures informative regions, feature channels also contain import cues. Different channels contain key information with different levels of importance. Inspired by previous work [38,45], we also adopt channel attention submodule to model channel interdependencies on channel dimension. We present the channel attention submodule in Figure 5. Feature maps after MCSA submodule are encountered with global and average pooling operations along spatial axes to obtain two channel descriptors with the shape of $C \times 1 \times 1$. Then, both descriptors are, respectively, sent to a multi-layer perceptron consisting of two shared fully connected (FC) layers with ReLU activation function. The length of the features after first FC layer is set to $(C/r) \times 1 \times 1$ to reduce complexity, where r is the reduction ration, while the length of the output after the second FC layer remains unchanged. After the shared network is applied to both descriptors, we adopt element-wise summation operation to merge the two output feature maps. A sigmoid activation function is followed to turn the merged feature into the final channel weights for different channels, denoted as M_C . Finally, the input feature map X_{MC} to channel attention submodule is multiplied with the channel attention map in an element-wise style to obtain the channel-wise refined feature map X_C .

Figure 4. Diagram of multiscale context spatial attention (MCSA) submodule.

Figure 5. Diagram of channel attention submodule.

3.3. Loss Function

Loss function plays a pivotal role in training deep neural networks. An appropriate loss function can stabilize the training process for better convergence and contribute to producing visually pleasing enhanced results. In this paper, we adopt a compound loss function L_{total} consisting of three components: GAN loss L_{gan} , perceptual loss L_p , and color loss L_c , which can be calculated as follows:

$$L_{total} = L_{gan} + \lambda_1 L_p + \lambda_2 L_c \quad (8)$$

where λ_1 and λ_2 represent corresponding weight coefficients for perceptual loss and color loss to balance relative importance of each component. We detail each loss component below.

GAN aims at generating fake images that are similar enough to real high-quality images that can deceive the discriminator. Standard GAN loss tries to guide the generator to produce images with the distribution matching to that of real natural images. Here, we adopt relativistic average discriminator [46] which predicts the relative probability that real images are more realistic than fake images and leads the generator to produce more realistic results. The loss function of relativistic average discriminator can be written as:

$$D_{Ra}(x_r, x_f) = \sigma \left(C(x_r) - E_{x_f \sim P_{fake}} [C(x_f)] \right) \quad (9)$$

$$D_{Ra}(x_f, x_r) = \sigma \left(C(x_f) - E_{x_r \sim P_{real}} [C(x_r)] \right) \quad (10)$$

where C represents the discriminator, x_r and x_f represent samples from real distribution P_{real} and fake distribution P_{fake} , respectively, σ denotes the sigmoid function, and E represents taking average value of all samples. We use least square GAN loss instead of sigmoid function in the discriminator. Therefore, the loss function of our generator and discriminator can be formulated as:

$$L_G = E_{x_f \sim P_{fake}} [(D_{Ra}(x_f, x_r) - 1)^2] + E_{x_r \sim P_{real}} [D_{Ra}(x_r, x_f)^2] \quad (11)$$

$$L_D = E_{x_f \sim P_{fake}} [D_{Ra}(x_f, x_r)^2] + E_{x_r \sim P_{real}} [(D_{Ra}(x_r, x_f) - 1)^2] \quad (12)$$

where L_G and L_D represent the respective loss of generator and discriminator. During the training phase, the generator and the discriminator are optimized alternately.

Perceptual loss proposed in [47] has been successfully applied in many low-level visual applications. It calculates and minimizes the distances between generated images and ground-truth images in feature domain by feeding images into a pretrained VGG-19 network, guiding the results to have visual appearances similar to the target. Unlike some previous works that calculate feature distances after only one layer in VGG-19 model, we choose feature maps after multiple convolutional layers since information within multiscale receptive fields can be involved. Hence, the perceptual loss is defined as the l_2 norm of feature maps:

$$L_p = \frac{1}{H_j W_j} \sum_j \|IN(\phi_j(\hat{y})) - IN(\phi_j(y))\|_2 \quad (13)$$

where $\hat{y}$ and y denote the ground-truth image and the generated result, $\phi_j(\cdot)$ denotes feature extraction after j -th convolutional layer of pretrained VGG-19 model, $IN(\cdot)$ denotes instance normalization [48], H_j and W_j denote spatial height and width of feature map.

Color deviation inevitably occurs during training process when enhancing low-light images. To mitigate color deviation and further improve image naturalness as much as possible, motivated by [49], we utilize color loss computed in HSV color space, which is closer to human perception and has been widely applied in image processing application. Color loss is formulated as:

$$L_c = \|\hat{S} \cdot \hat{V} \cdot \cos(\hat{H}) - S \cdot V \cdot \cos(H)\|_1 \quad (14)$$

where H , S , and V represent the hue, saturation and value components of generated image respectively, and $\hat{H}$, $\hat{S}$, and $\hat{V}$ represent those of the target image.

4. Experimental Evaluation and Analysis

4.1. Dataset Description and Evaluation Metrics

We conduct extensive experiments to evaluate our proposed method systematically. LOL [19] and SICE [50] are two common paired datasets in the low-light image enhancement research community. Like many previous works [25,51] did, we use them as our benchmark datasets to validate our model. Some previous works [52,53] also evaluate their methods on MIT-Adobe FiveK [54] dataset. Nevertheless, this dataset is originally constructed for enhancing aesthetic quality of images, and only a small portion of images were taken under low-light conditions. Thus, we do not use it as our benchmark dataset.

The original LOL dataset contains 500 real-captured low-/normal-light image pairs with the size of 400×600 , which were captured in real scenes by change exposure time and ISO of camera. To further enrich content diversity, 1000 more low-/normal-light images pairs with the size of 384×384 were synthesized as supplements by analyzing the illumination distribution of real captured low-light images. The constructors of LOL dataset divide the 500 real-captured image pairs into the training set with 485 image pairs and the testing set with 15 image pairs. However, their original splitting scheme cannot guarantee the comparative methods' performance to be fully evaluated since the number of testing set is too small and the contents of training and testing sets are repetitive. Hence, we adopt the modified version

of LOL dataset proposed in DRBN [55], which consists of 789 low-/normal-light image pairs. The 689 pairs are served as training samples, while the other 100 pairs are the testing set. We also use the 1000 synthesized image pairs in the LOL dataset as supplementary training data, so our training data contain 1689 pairs including real-captured and synthesized images. In addition, the contents of testing set are unseen during the training phase which is important for fair comparison.

The original SICE dataset contains 589 sequences of different scenes, and each sequence includes multiple images with different exposure levels and a high-quality reference image selected by multi-exposure fusion algorithms. We filter out SICE images containing misalignment caused by moving objects and image distortion during the collecting process. Additionally, we also reduce the spatial size of SICE images due to limited computing resources. The resized images have the same length-width ratio as the original images with the long side containing 800 pixels. Finally, we choose 1300 under-exposure images with their corresponding references as training set, and another 120 images as testing set. The contents of testing set do not overlap with the training set.

In addition, we also evaluate the generalization capacity of our method on several widely-used benchmark low-light image datasets, i.e., LIME [12] (10 images), MEF [56] (17 images), DICM [57] (69 images), MFUSION [58] (10 images), and VV [59] (24 images). These datasets have no corresponding normal-light images as ground-truth images. Their image contents are much more diverse and not included in LOL and SICE dataset. All the datasets were built in natural scenes, where images do not contain null data. In addition, we deleted completely dark images before experiments, rather than changing and re-scaling them [60], since they do not contain any useful information.

We use four common image quality metrics to measure the perceptual quality of enhanced results, i.e., peak signal-to-noise ratio (PSNR), structure similarity [61] (SSIM) index, natural image quality evaluator [62] (NIQE) and integrated local NIQE [63] (IL-NIQE). The PSNR is ratio between the maximum power of normal light image and the power of background noise which degrades image fidelity. SSIM measures image quality from the perspective of image structure since human eyes pay more attention to image edges and details, and image distortion degrades structural information. These two metrics require corresponding high-quality images as references when evaluating perceptual quality. NIQE extracts relevant features from high-quality natural images to build a perfect quality model, and measures image quality by calculating distance from it. IL-NIQE further extends NIQE by taking local patches into consideration. These two metrics can directly obtain perceptual quality scores from input images only without any references.

4.2. Implementation Detail

Our method is implemented using Pytorch framework and trained on a PC with Nvidia 2080Ti GPU and i7-8700k CPU. We adopt the Adam optimization algorithm [64] with the batch size of 4, and the default parameters of Adam ($\beta_1 = 0.9, \beta_2 = 0.999$) are fixed. The initial learning is set to 10^{-5} . During the training process, we crop patches at random location in all training images with the unified size of $320 \times 320 \times 3$. The loss weight λ_1 and λ_2 are all set to 5. For low-/normal-light image pairs, the cropping position in coupled images is kept same to avoid unwanted pixel misalignment. For data augmentation, we randomly flip training images horizontally. The training images are normalized into the range of $[0, 1]$. All image batches are packed into 4-D tensors to serve as the CNN inputs. Batch normalization is adopted after each convolutional layer for better convergence during the training process. All training parameters in convolutional layers are initialized using the method introduced in [65]. We have trained our network for 120 epochs in total, and the initial learning rate is decayed by multiplying with 0.5 per 30 epochs. During the testing stage, the batch size is set to 1. Hence, our network can process images with arbitrary spatial size.

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

We evaluate our method on benchmark datasets and compare it with several state-of-the-art methods. The competitive methods include: (1) Retinex theory-based methods: LIME [12], BIMEF [66], NPE [13], and SRIE [14]; (2) deep-learning-based methods: RetinexNet [19], GLADNet [67], Zero-DCE [25], RRDNet [24], and EnlightenGAN [21]. We choose these comparative methods like many previous pieces of work [27,28,55] did since they are representative in the field of low-light image enhancement. For learning-based methods, we use source codes provided by their corresponding authors. We retrain and test them on the same training and testing dataset as the proposed method for fair comparison. For methods that do not need training process, we directly test them on the same testing set. Note that when evaluating on LOL and SICE, we use their respective training and testing samples.

We first list the quantitative results of all competing methods on the LOL and SICE datasets. PSNR and SSIM are chosen as evaluation metrics since both LOL and SICE have referenced ground-truths. Besides, we also use NIQE score as extra index to measure quality without reference. For PSNR and SSIM, a larger score means better perceptual quality, while for NIQE, a lower score means better perceptual quality. The average scores of all methods on each dataset are listed in Table 1. As is shown in Table 1, our method has outperformed all state-of-the-art methods on LOL dataset in terms of PSNR and SSIM even though we do not directly adopt MSE loss and SSIM loss in the optimization process, demonstrating the superiority of our method. For NIQE score on LOL dataset, our score has made the top 2 and is only inferior to GLADNet. For SICE dataset, our method has achieved good results except for SSIM, which is only slightly lower than GLADNet. Nearly all methods' results degrade on SICE dataset compared to LOL dataset. From Table 1, we can also find that methods without any training data, such as RRDNet and SRIE, perform poorly, indicating the importance of data-driven routine. Overall, learning-based methods perform better than traditional Retinex-based methods, attributable to the strong representation power of CNNs.

Table 1. Quantitative comparison between our method and other state-of-the-art methods on LOL dataset and SICE dataset. The best results are highlighted in boldface.

Methods	LOL (100 Images)			SICE (120 Images)		
	PSNR	SSIM	NIQE	PSNR	SSIM	NIQE
LIME [12]	15.2524	0.4729	9.8031	12.8109	0.5277	3.1757
NPE [13]	17.3451	0.5170	9.6743	16.2730	0.5509	2.6215
SRIE [14]	14.4877	0.5440	7.9710	15.0963	0.5004	2.5757
BIMEF [66]	17.8781	0.6540	8.1987	16.2763	0.5087	2.6501
RRDNet [24]	13.0681	0.4324	7.6332	14.7994	0.4839	2.9285
GLADNet [67]	18.2202	0.7160	3.8418	17.0483	0.5557	2.9591
RetinexNet [19]	17.2003	0.5474	10.5765	15.1248	0.5283	3.8574
EnlighenGAN [21]	18.6588	0.6696	5.1479	16.2803	0.5513	2.7427
Zero-DCE [25]	17.6891	0.5344	9.0115	14.2244	0.4027	3.4510
Ours	20.2265	0.7378	4.2319	17.0500	0.5522	2.3240

Then we show some qualitative results of all competing methods. Figure 6 displays visual comparisons with state-of-the-art methods on the LOL dataset. To make image details more visible, we also zoom in on some details in the bounding boxes and show them in the bottom left corner. Specifically, in Figure 6b–e, traditional Retinex-based methods fail to generate enhanced results with sufficient brightness and their results contain intense noise. Zero-shot learning-based method RRDNet cannot enhance the low-light input, as shown in Figure 6f, indicating the importance of training data. In Figure 6g–i, these three learning-based methods also produce images with severe noise, such as the surrounding of the wheel under ping-pong table shown in red rectangles. On the contrary, in Figure 6j, GLADNet generates relatively good enhanced result compared to Retinex, Zero-DCE, and EnlightenGAN.

However, it still brings some noise and color deviation, making the results not as close to the ground-truth image as our method.

Figure 7 also displays visual comparison against state-of-the-art methods on the SICE dataset. We can see clearly that LIME improves the illumination, but brings color deviation. NPE, SRIE, BIMEF and RRDNet yield results without sufficient brightness. RetinexNet and Zero-DCE yield results with much noise and color distortion. In Figure 7i, the result of EnlightenGAN has obvious over-saturation and under-exposure, such as “sky” and “building”. It seems that the result of GLADNet and our method look very similar, which corresponds to the results in Table 1. However, in Figure 7k, our result contains a little more texture, such as “tree” in the red rectangle.

Figure 6. Visual comparisons of different methods on a real image from LOL dataset.

Figure 7. Visual comparisons of different methods on a real image from SICE dataset.

Apart from the above two benchmark datasets, we also evaluate our method on five real-world datasets widely used in many low-light enhancement algorithms to test the generalization capacity. Since these datasets have no ground-truth reference normal-light images, we adopt NIQE and ILNIQE as the evaluation metrics. A lower quality score means better perceptual quality. Before we test our model on these datasets, we train our model using LOL training samples. The quantitative results are listed in Table 2. We also

list weighted average indexes for all methods, where the weight is calculated according to the number of images in each dataset.

Table 2. Quantitative comparison between our method and other state-of-the-art methods on 5 real-world datasets. The best results are highlighted in boldface, while the second results are underlined.

Methods	MEF (17 Images)		DICM (69 Images)		LIME (10 Images)	
	NIQE	ILNIQE	NIQE	ILNIQE	NIQE	ILNIQE
LIME [12]	3.8269	26.1817	3.4268	28.4170	4.5555	32.5244
NPE [13]	3.5469	24.1136	<u>3.1234</u>	25.8219	4.0250	29.4629
SRIE [14]	3.2041	25.8736	3.1628	25.9804	<u>3.6523</u>	29.2414
BIMEF [66]	3.0651	24.4423	3.1595	26.3395	3.7455	29.1967
RRDNet [24]	3.1307	25.5280	3.3946	26.7397	3.7228	28.8370
GLADNet [67]	3.5555	27.9934	3.2777	26.1807	3.9811	30.5339
RetinexNet [19]	4.6230	23.6211	3.6606	25.2090	5.2911	29.0565
EnlightenGAN [21]	3.0102	25.3575	3.1548	<u>24.9817</u>	3.6473	26.7167
Zero-DCE [25]	3.3075	24.0260	3.7409	25.3581	4.2958	30.3389
Ours	<u>3.0201</u>	<u>23.6694</u>	2.8861	24.7051	3.7494	<u>26.9393</u>
Methods	MFUSION (10 Images)		VV (24 Images)		Weighted Average	
	NIQE	ILNIQE	NIQE	ILNIQE	NIQE	ILNIQE
LIME [12]	3.4202	29.3796	3.5141	25.4747	3.5816	27.9715
NPE [13]	2.9492	24.4639	2.7336	22.7550	3.1628	25.2079
SRIE [14]	2.8223	23.4508	2.7102	23.9456	3.0961	25.6470
BIMEF [66]	3.0228	23.3601	2.8987	24.1321	3.1336	25.6745
RRDNet [24]	3.1714	24.4290	3.2492	24.4691	3.3413	26.1456
GLADNet [67]	3.2519	24.3996	3.1450	24.4801	3.3416	26.3016
RetinexNet [19]	4.1025	24.1310	3.4800	20.8992	3.9125	24.4187
EnlightenGAN [21]	2.9095	<u>22.2765</u>	<u>2.6486</u>	21.9790	<u>3.0615</u>	<u>24.4019</u>
Zero-DCE [25]	3.8680	25.4857	3.5473	23.1790	3.7009	25.1746
Ours	<u>2.8259</u>	21.4988	2.5801	<u>21.1579</u>	2.9089	23.8401

As can be seen from Table 2, our method obtains relatively good performances on all five datasets. Even though our method does not achieve the best results in terms of NIQE and ILNIQE scores on some datasets such as MEF and LIME, it is still in the top three results and the scores are only slightly lower than the best. Our method has achieved the best average indexes, which indicates that our method has relatively good generalization capacity. Some visual results on these datasets are shown in Figures 8–11.

Figures 8 and 9 show representative enhancement results of different methods on two real-world datasets, i.e., MEF and VV. As can be seen from above, our model can effectively enhance real-world low-light images. Concretely, LIME can enhance visual visibility, but it produces result with over-exposure and color shift, such as “balloon” in Figure 8b and “plant” in Figure 9b. Obviously, NPE, SRIE, BIMEF and RRDNet fail to improve visual visibility and weaken image contrast. RetinexNet and Zero-DCE can restore image brightness but amplify tensive noise in some areas, such as top part of “balloon” in Figure 8 and “plant” in Figure 9. EnlightenGAN yields dark results near the characters on the “balloon” in Figure 8 and also contains halo effects near the “woman” contour in Figure 9. GLADNet also amplifies noise and produces over-exposure in “woman face” area. Contrarily, our method can produce results with rich texture, sufficient brightness, and less distortion.

Figure 8. Visual comparisons of different methods on a real image from MEF dataset.

Figure 9. Visual comparisons of different methods on a real image from VV dataset.

Figures 10 and 11 are another two representative enhancement results from two real-world datasets, i.e., MFUSION and LIME. In Figure 10, LIME and GLADNet can enhance image lightness, but they tend to be enhanced in the “tree” region. NPE and RetinexNet can produce results with suitable exposure, but they introduce more color distortion on the pavement. The result of Zero-DCE looks unnatural, and SRIE, BIMEF, RRDNet, and EnlightenGAN can not improve image lightness sufficiently, such as tree texture, as our method. In Figure 11, NPE, SRIE, BIMEF, and RRDNet contain under-exposed region. LIME, RetinexNet, and EnlightenGAN have obvious color deviation on the pavement near the car. Compared to GLADNet, our result has clearer blue traffic sign and less blurred pattern on the pavement.

Figure 10. Visual comparisons of different methods on a real image from MFUSION dataset.

Figure 11. Visual comparisons of different methods on a real image from LIME dataset.

4.4. Ablation Analysis

To evaluate the effectiveness of each component in our model, we conduct several ablation experiments. Specifically, we remove one component of our model each time, and re-train the model using the same parameter settings and the LOL dataset. To verify the role of the proposed MSCAM, we replace them with short connection to directly link

relevant layers at the same spatial level. To evaluate the effectiveness of multiple residual dense blocks (RDBs), we replace them with convolutional layers of same channel numbers, same activation function and batch normalization layers. Moreover, to demonstrate the performance gain introduced by different parts of loss function, we set the weights λ_1 and λ_2 of perceptual loss and color loss to 0, respectively. Our model is GAN-based method; hence, adversarial loss cannot be omitted. Each model variant is evaluated on LOL testing set. All the quantitative results are listed in Table 3.

As can be seen from Table 3, our model with all components achieves the best results, demonstrating all components in our model have their own contributions to the final results. Obviously, model without RDBs or MSCAM degrades the performance by a large margin. Besides, each component in loss function plays a critical role during the training process.

Table 3. Quantitative results of all ablation experiments on LOL testing set. The best results are highlighted in boldface. ‘w/o’ means without.

Model Variants	LOL (100 Images)		
	PSNR	SSIM	NIQE
model w/o MSCAM	20.0449	0.7314	4.3938
model w/o RDBs	18.9750	0.7038	4.2789
model w/o L_p	19.4423	0.6871	4.3879
model w/o L_c	19.8560	0.7331	4.2076
model (proposed)	20.4342	0.7520	4.2042

To demonstrate the results in a more intuitive way, we also display some qualitative results in Figure 12. All model variants can improve the brightness, but they still contain some distortion. Without RDBs or MSCAM, the enhanced results contain color shift, namely, the “back wall area” in Figure 12b, and noise amplification, such as the “door” region in Figure 12c. Without perceptual loss, the enhanced result in Figure 12d yields a severe color shift. The result in Figure 12e contains less color deviation, but without color loss, the result still includes slightly color shift in “back wall area near the chair”, degrading the enhancing result. In contrast, our result looks more visually pleasing, and contains vivid color and less noise, validating the effectiveness of each component.

Figure 12. Visual comparison of ablation studies on LOL testing set.

We also investigate the impacts of the number of convolutional layers in RDBs. We gradually change the number of convolutional layers from two to five. Each time we re-train and evaluate our model using LOL dataset with other configuration fixed. Using more than

five layers will exceed the GPU resources at our disposal. The quantitative results are in Table 4. We can see that using five convolutional layers in each RDB can improve performance.

Table 4. Ablation study on the number of convolutional layers in RDBs. The best results are highlighted in boldface.

Number of Conv Layers	LOL (100 Images)		
	PSNR	SSIM	NIQE
2 conv layers	19.8414	0.7229	4.2620
3 conv layers	19.9364	0.7248	4.5254
4 conv layers	19.5328	0.6873	4.3288
5 conv layers (proposed)	20.4342	0.7520	4.2042

In addition, we study the impacts of two weight coefficients λ_1 and λ_2 in the training loss function. Each time we vary one of them and fix the other to 1, then we re-train and test our model using the LOL dataset. We plot the performance curves with respect to λ_1 and λ_2 in Figure 13. We can see that in general SSIM and NIQE are stable with respect to λ_1 and λ_2 at a wide range. PSNR has larger fluctuations than SSIM and NIQE since it calculates pixel-wise difference between two images. To balance three metrics and intuitively observe loss curve in the training process, we empirically set both λ_1 and λ_2 to 5.

Figure 13. (a) Performance variation with respect to λ_1 . (b) Performance variation with respect to λ_2 .

4.5. User Study

To validate the visual similarity between the enhanced results and the ground-truth normal light images subjectively, we have invited 20 volunteer college students without extra image processing knowledge to judge whether the enhanced results are similar to the ground-truths. We used ten images chosen from the LOL dataset, and requested all the volunteers to give their ratings on similarity level according to their visual perception. The ratings consisted of five discrete scores, where 5 denotes the best visual similarity to the ground-truth image and 1 denotes the worst. All the volunteers were sitting at a viewing distance about 2.5 times of the monitor's height to assign scores to every enhanced image. Each time we presented an enhanced result of one comparative method with its corresponding ground-truth image to one subject. The order of displaying method is random, and all subjects do not know which images are generated by our method. For each competitive method, we averaged the scores of all images given by all subjects as the final similarity score. Here, we only conduct subjective tests on learning-based methods. The results are listed in Table 5. It is observed that our method has obtained superior mean opinion score (MOS) compared to others, meaning our enhanced results are most similar to the ground-truth normal-light images. RetinexNet and RRDNet obtain the lowest scores among all methods, indicating their enhanced results have distinct differences to the ground truths. EnlightenGAN and GLADNet are slightly inferior to our method, which are consistent with previous analyses.

Table 5. The mean opinion scores (MOS) of different methods in user study. The best result is highlighted in boldface.

Methods	RRDNet	RetinexNet	GLADNet	EnlightenGAN	Zero-DCE	Ours
MOS	2.48	2.34	3.18	3.36	2.62	3.42

4.6. Computational Complexity

An ideal enhancement algorithm is expected to produce excellent enhancing results as well as to have low computational complexity. We also test the running speed of each comparative method. We choose an image from the LOL dataset with a fixed spatial resolution of $400 \times 600 \times 3$ for all the methods. All the comparative methods are running on a desktop computer with a 3.7 GHz CPU and a 32 GB internal memory. For learning-based methods, we also use a NVIDIA GTX1080ti GPU to accelerate the testing phase. We list the running time of each method in Table 6. Our method ranks third among all the methods. Clearly, Zero-DCE has achieved the fastest inference speed since it adopts a light-weight architecture to estimate enhancement curve. Although our method has more extra computational burden than EnlightenGAN, it can still provide a relatively good trade-off between efficiency and effectiveness.

Table 6. Runtime comparisons of different methods. The best result is highlighted in boldface.

Methods	Running Time (s)	Platform
LIME [12]	0.1899	MATLAB (CPU)
NPE [13]	6.0977	MATLAB (CPU)
SRIE [14]	5.7245	MATLAB (CPU)
BIMEF [66]	0.1729	MATLAB (CPU)
RRDNet [24]	53.3276	Pytorch (GPU)
GLADNet [67]	0.0323	Tensorflow (GPU)
RetinexNet [19]	0.0358	Tensorflow (GPU)
EnlightenGAN [21]	0.0092	Pytorch (GPU)
Zero-DCE [25]	0.0062	Pytorch (GPU)
Ours	0.0218	Pytorch (GPU)

5. Conclusions

In this paper, we propose an improved GAN for low-light image enhancement. To promote hierarchical feature fusion across multiple layers at different depths and ease information flow, we insert local residual dense blocks into the generator. Then, to alleviate local distortion and reduce noise, we introduce multiscale context attention module integrating more contextual information from multiple scales, and enable feature interaction across layers. Spatial and channel-wise dependencies are also modeled to adaptively refine and dynamically recalibrate feature maps. A hybrid loss function containing perceptual and color component is utilized to train the proposed model. Quantitative and qualitative results demonstrate that our proposed method can generate visually pleasing enhanced results. However, our method still has some limitations. For instance, it cannot process images taken in extremely dark scenes, and parameters in our model can be further reduced. In the future work, we will explore the use of transfer learning to reduce the complexity and further improve the performance.

Author Contributions: Conceptualization, methodology, validation, investigation, and writing—original draft preparation, S.H.; writing—review and editing, S.H. and J.Y.; project administration, and funding acquisition, J.Y. and D.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded in part by the National Natural Science Foundation of China grant number 61701351.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zuiderveld, K. Contrast Limited Adaptive Histogram Equalization. In *Graphics Gems IV*; Academic Press Professional, Inc.: Cambridge, MA, USA, 1994; pp. 474–485.
2. Pisano, E.D.; Zong, S.; Hemminger, B.M.; DeLuca, M.; Johnston, R.E.; Muller, K.; Braeuning, M.P.; Pizer, S.M. Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *J. Digit. Imaging* **1998**, *11*, 193. [[CrossRef](#)] [[PubMed](#)]
3. Lee, C.; Lee, C.; Kim, C.S. Contrast Enhancement Based on Layered Difference Representation of 2D Histograms. *IEEE Trans. Image Process.* **2013**, *22*, 5372–5384. [[CrossRef](#)]
4. Abdullah-Al-Wadud, M.; Kabir, M.H.; Akber Dewan, M.A.; Chae, O. A Dynamic Histogram Equalization for Image Contrast Enhancement. *IEEE Trans. Consum. Electron.* **2007**, *53*, 593–600. [[CrossRef](#)]
5. Stark, J. Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Trans. Image Process.* **2000**, *9*, 889–896. [[CrossRef](#)]
6. Capra, A.; Castorina, A.; Corchs, S.; Gasparini, F.; Schettini, R. Dynamic range optimization by local contrast correction and histogram image analysis. In Proceedings of the 2006 Digest of Technical Papers International Conference on Consumer Electronics, Las Vegas, NV, USA, 7–11 January 2006; pp. 309–310. [[CrossRef](#)]
7. Celik, T.; Tjahjadi, T. Contextual and Variational Contrast Enhancement. *IEEE Trans. Image Process.* **2011**, *20*, 3431–3441. [[CrossRef](#)]
8. Ibrahim, H.; Pik Kong, N.S. Brightness Preserving Dynamic Histogram Equalization for Image Contrast Enhancement. *IEEE Trans. Consum. Electron.* **2007**, *53*, 1752–1758. [[CrossRef](#)]
9. Land, E.H. The retinex theory of color vision. *Sci. Am.* **1977**, *237*, 108–129. [[CrossRef](#)]
10. Jobson, D.; Rahman, Z.; Woodell, G. Properties and performance of a center/surround retinex. *IEEE Trans. Image Process.* **1997**, *6*, 451–462. [[CrossRef](#)]
11. Jobson, D.; Rahman, Z.; Woodell, G. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans. Image Process.* **1997**, *6*, 965–976. [[CrossRef](#)]
12. Guo, X.; Li, Y.; Ling, H. LIME: Low-Light Image Enhancement via Illumination Map Estimation. *IEEE Trans. Image Process.* **2017**, *26*, 982–993. [[CrossRef](#)]
13. Wang, S.; Zheng, J.; Hu, H.M.; Li, B. Naturalness Preserved Enhancement Algorithm for Non-Uniform Illumination Images. *IEEE Trans. Image Process.* **2013**, *22*, 3538–3548. [[CrossRef](#)] [[PubMed](#)]
14. Fu, X.; Zeng, D.; Huang, Y.; Zhang, X.P.; Ding, X. A Weighted Variational Model for Simultaneous Reflectance and Illumination Estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
15. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
16. Ren, W.; Ma, L.; Zhang, J.; Pan, J.; Cao, X.; Liu, W.; Yang, M.H. Gated Fusion Network for Single Image Dehazing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
17. Lore, K.G.; Akintayo, A.; Sarkar, S. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognit.* **2017**, *61*, 650–662. [[CrossRef](#)]
18. Chen, C.; Chen, Q.; Xu, J.; Koltun, V. Learning to See in the Dark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
19. Wei, C.; Wang, W.; Yang, W.; Liu, J. Deep Retinex Decomposition for Low-Light Enhancement. In Proceedings of the British Machine Vision Conference, Newcastle, UK, 3–6 September 2018.
20. Hua, W.; Xia, Y. Low-Light Image Enhancement Based on Joint Generative Adversarial Network and Image Quality Assessment. In Proceedings of the 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Beijing, China, 13–15 October 2018; pp. 1–6. [[CrossRef](#)]
21. Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; Wang, Z. EnlightenGAN: Deep Light Enhancement Without Paired Supervision. *IEEE Trans. Image Process.* **2021**, *30*, 2340–2349. [[CrossRef](#)] [[PubMed](#)]
22. Ni, Z.; Yang, W.; Wang, S.; Ma, L.; Kwong, S. Towards Unsupervised Deep Image Enhancement with Generative Adversarial Network. *IEEE Trans. Image Process.* **2020**, *29*, 9140–9151. [[CrossRef](#)]
23. Chen, Y.S.; Wang, Y.C.; Kao, M.H.; Chuang, Y.Y. Deep Photo Enhancer: Unpaired Learning for Image Enhancement From Photographs with GANs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
24. Zhu, A.; Zhang, L.; Shen, Y.; Ma, Y.; Zhao, S.; Zhou, Y. Zero-Shot Restoration of Underexposed Images via Robust Retinex Decomposition. In Proceedings of the 2020 IEEE International Conference on Multimedia and Expo (ICME), London, UK, 6–10 July 2020; pp. 1–6. [[CrossRef](#)]
25. Guo, C.; Li, C.; Guo, J.; Loy, C.C.; Hou, J.; Kwong, S.; Cong, R. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
26. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.

27. Atoum, Y.; Ye, M.; Ren, L.; Tai, Y.; Liu, X. Color-Wise Attention Network for Low-Light Image Enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Seattle, WA, USA, 14–19 June 2020.
28. Lv, F.; Li, Y.; Lu, F. Attention guided low-light image enhancement with a large scale low-light simulation dataset. *Int. J. Comput. Vis.* **2021**, *129*, 2175–2193. [\[CrossRef\]](#)
29. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H.; Shao, L. Learning enriched features for real image restoration and enhancement. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020, Proceedings, Part XXV 16*; Springer: Cham, Switzerland, 2020; pp. 492–511.
30. Lio, G.E.; Ferraro, A.; Ritacco, T.; Aceti, D.M.; De Luca, A.; Giocondo, M.; Caputo, R. Leveraging on ENZ Metamaterials to Achieve 2D and 3D Hyper-Resolution in Two-Photon Direct Laser Writing. *Adv. Mater.* **2021**, *33*, 2008644. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Lio, G.E.; Ferraro, A. LIDAR and Beam Steering Tailored by Neuromorphic Metasurfaces Dipped in a Tunable Surrounding Medium. *Photonics* **2021**, *8*, 65. [\[CrossRef\]](#)
32. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2014; Volume 27.
33. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
34. Itti, L.; Koch, C.; Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 1254–1259. [\[CrossRef\]](#)
35. Wang, F.; Jiang, M.; Qian, C.; Yang, S.; Li, C.; Zhang, H.; Wang, X.; Tang, X. Residual Attention Network for Image Classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
36. Liu, X.; Ma, Y.; Shi, Z.; Chen, J. GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019.
37. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-Local Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
38. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
39. Yu, F.; Koltun, V. Multi-Scale Context Aggregation by Dilated Convolutions. In Proceedings of the 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, 2–4 May 2016.
40. Li, Y.; Chen, Y.; Wang, N.; Zhang, Z. Scale-Aware Trident Networks for Object Detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019.
41. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
42. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [\[CrossRef\]](#)
43. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
44. Chen, D.; He, M.; Fan, Q.; Liao, J.; Zhang, L.; Hou, D.; Yuan, L.; Hua, G. Gated Context Aggregation Network for Image Dehazing and Deraining. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa Village, HI, USA, 7–11 January 2019; pp. 1375–1383. [\[CrossRef\]](#)
45. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
46. Jolicoeur-Martineau, A. The relativistic discriminator: A key element missing from standard GAN. *arXiv* **2018**, arXiv:1807.00734.
47. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016; pp. 694–711.
48. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Instance normalization: The missing ingredient for fast stylization. *arXiv* **2016**, arXiv:1607.08022.
49. Moran, S.; McDonagh, S.; Slabaugh, G. CURL: Neural Curve Layers for Global Image Enhancement. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 9796–9803. [\[CrossRef\]](#)
50. Cai, J.; Gu, S.; Zhang, L. Learning a Deep Single Image Contrast Enhancer from Multi-Exposure Images. *IEEE Trans. Image Process.* **2018**, *27*, 2049–2062. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Lu, K.; Zhang, L. TBEFN: A Two-Branch Exposure-Fusion Network for Low-Light Image Enhancement. *IEEE Trans. Multimed.* **2021**, *23*, 4093–4105. [\[CrossRef\]](#)
52. Ren, W.; Liu, S.; Ma, L.; Xu, Q.; Xu, X.; Cao, X.; Du, J.; Yang, M.H. Low-Light Image Enhancement via a Deep Hybrid Network. *IEEE Trans. Image Process.* **2019**, *28*, 4364–4375. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Wang, R.; Zhang, Q.; Fu, C.W.; Shen, X.; Zheng, W.S.; Jia, J. Underexposed Photo Enhancement Using Deep Illumination Estimation. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 6842–6850. [\[CrossRef\]](#)

54. Bychkovsky, V.; Paris, S.; Chan, E.; Durand, F. Learning photographic global tonal adjustment with a database of input/output image pairs. In Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 97–104. [\[CrossRef\]](#)
55. Yang, W.; Wang, S.; Fang, Y.; Wang, Y.; Liu, J. From Fidelity to Perceptual Quality: A Semi-Supervised Approach for Low-Light Image Enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
56. Ma, K.; Zeng, K.; Wang, Z. Perceptual Quality Assessment for Multi-Exposure Image Fusion. *IEEE Trans. Image Process.* **2015**, *24*, 3345–3356. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Lee, C.; Lee, C.; Kim, C.S. Contrast enhancement based on layered difference representation. In Proceedings of the 2012 19th IEEE International Conference on Image Processing, Orlando, FL, USA, 30 September–3 October 2012; pp. 965–968. [\[CrossRef\]](#)
58. Fu, X.; Zeng, D.; Huang, Y.; Liao, Y.; Ding, X.; Paisley, J. A fusion-based enhancing method for weakly illuminated images. *Signal Process.* **2016**, *129*, 82–96. [\[CrossRef\]](#)
59. Vonikakis, V. Datasets. Available online: <https://sites.google.com/site/vonikakis/datasets> (accessed on 9 July 2017).
60. Gonzalez-Abril, L.; Angulo, C.; Ortega, J.A.; Lopez-Guerra, J.L. Generative Adversarial Networks for Anonymized Healthcare of Lung Cancer Patients. *Electronics* **2021**, *10*, 2220. [\[CrossRef\]](#)
61. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [\[CrossRef\]](#)
62. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a “Completely Blind” Image Quality Analyzer. *IEEE Signal Process. Lett.* **2013**, *20*, 209–212. [\[CrossRef\]](#)
63. Zhang, L.; Zhang, L.; Bovik, A.C. A Feature-Enriched Completely Blind Image Quality Evaluator. *IEEE Trans. Image Process.* **2015**, *24*, 2579–2591. [\[CrossRef\]](#)
64. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
65. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
66. Ying, Z.; Li, G.; Gao, W. A bio-inspired multi-exposure fusion framework for low-light image enhancement. *arXiv* **2017**, arXiv:1711.00591.
67. Wang, W.; Wei, C.; Yang, W.; Liu, J. GLADNet: Low-Light Enhancement Network with Global Awareness. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018), Xi’an, China, 15–19 May 2018; pp. 751–755. [\[CrossRef\]](#)