Knowledge-Based Systems 240 (2022) 108010

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

LE-GAN: Unsupervised low-light image enhancement network using attention module and identity invariant loss

Ying Fu [a,*], Yang Hong [a], Linwei Chen [a], Shaodi You [b]

[a] School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China
[b] Institute of Informatics, University of Amsterdam, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 25 July 2021
Received in revised form 16 December 2021
Accepted 17 December 2021
Available online 23 December 2021

Keywords:
Low-light image enhancement
Illumination-aware attention module
Identity invariant loss
Paired normal/low-light images dataset

ABSTRACT

Low-light image enhancement aims to recover normal-light images from the images captured under very dim environments. Existing methods cannot well handle the noise, color bias and over-exposure problem, and fail to ensure visual quality when lacking paired training data. To address these problems, we propose a novel unsupervised low-light image enhancement network named LE-GAN, which is based on generative adversarial networks and is trained with unpaired low/normal-light images. Specifically, we design an illumination-aware attention module that enhances the feature extraction of the network to address the problems of noise and color bias, as well as improve the visual quality. We further propose a novel identity invariant loss to address the over-exposure problem to make the network learn to enhance low-light images adaptively. Extensive experiments show that the proposed method can achieve promising results. Furthermore, we collect a large-scale low-light dataset named Paired Normal/Lowlight Images (PNLI). It consists of 2,000 pairs of low/normal-light images captured in various real-world scenes, which can provide the research community with a high-quality dataset to advance the development of this field.

1. Introduction

Compared with normal-light images, quality degradation of low-light images captured under terrible lighting conditions is serious due to inevitable environmental or technical constraints, leading to unpleasant visual perception including details degradation, color distortion, and severe noise. These phenomena have a significant impact on the performance of advanced downstream visual tasks, such as image classification, object detection, semantic segmentation [1–4], *etc*. To mitigate the degradation of image quality, low-light image enhancement has become an important topic in the low-level image processing community to effectively improve visual quality and restore image details.

In the early research on low-light image enhancement, traditional methods [5,6] generally use hand-crafted features as input and utilize optimization strategy and rules to improve image quality, which are greatly dependent on the precision of their assumption of hand-crafted priors. Recently, deep learning methods have become more and more popular in the computer vision community, and have achieved unprecedented improvements in low-light enhancement. Many researchers [7,8] train deep models in a supervised manner with low/normal-light image pairs. However, these methods usually do not generalize well on real-world images because their performance is highly affected by the paired data of training sets. Besides, it is also very difficult to simultaneously capture low-light and ground truth images of the same visual scenes. Another line of methods [9,10] attempts to address the low-light enhancement task in an unsupervised manner. Among these methods, EnlightenGAN [10] is the first unsupervised approach to solve the low-light enhancement problem using a global–local discriminator structure with well-designed losses. Zero-DCE [9] proposed a deep curve estimation network to treat this problem as a task of image-specific curve estimation. Nevertheless, they inevitably have severe color deviation and artifacts in some cases.

To address the above issues, in this paper, we present a low-light enhancement generative adversarial network (LE-GAN) using a cyclic architecture to transform low-light images into the corresponding normal-light ones in an unsupervised way without relying on exactly paired images. To improve the visual quality of recovered images, we propose an illumination-aware module to utilize contextual and global information. Specifically, it consists of a spatial-illumination attention branch and a global-illumination attention branch. The spatial-illumination attention module encodes a wider range of contextual information to dig more discriminative spatial features, which helps the network generate images with better visual quality and avoid the negative effects of noise. The global-illumination attention module is

* Corresponding author.
E-mail address: fuying@bit.edu.cn (Y. Fu).

https://doi.org/10.1016/j.knosys.2021.108010
0950-7051/

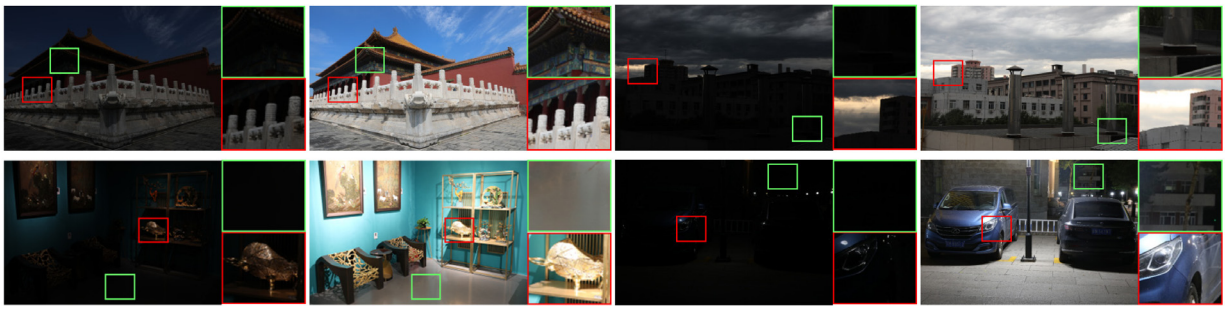

Fig. 1. Representative visual examples by enhancing typical low-light images from our PNLI dataset using LE-GAN. The proposed LE-GAN achieves visually pleasing results in both dark and regions where the brightness changes drastically.

designed to learn the global correlation along feature channels, which improves the correction of colors and eliminates color bias in low-light images. Besides, as shown in Fig. 1, it is a challenging issue of enhancing the regions with relatively high brightness. Previous works [11–13] tend to improve the overall brightness of the image, which will easily cause over-exposure in these regions and result in unsatisfactory visual feeling. To address the above issues, we propose the identity invariant loss, which imposes an identity constraint on the output to focus on learning to enhance low-light images adaptively.

In addition, to better improve and validate low-light enhancement methods, we collect a large-scale low-light dataset under real-world scenarios named **P**aired **N**ormal/**L**ow-light **I**mages (**PNLI**). It consists of 2,000 high-quality image pairs with 6720 $\times$ 4480 resolution, which are taken in various indoor and outdoor scenes. Our PNLI has larger scale of the dataset, better scene diversity, and higher image resolution over existing datasets. We believe that the dataset can greatly advance the development of this field. Extensive experiments on our PNLI dataset as well as existing datasets [8,14] show that our method outperforms state-of-the-arts and has better generalization ability than other methods.

In summary, the main contributions of our paper are summarized as follows:

- We present a novel unsupervised low-light enhancement method named LE-GAN, which outperforms state-of-the-art competitors.
- We propose an illumination-aware attention module and an identity invariant loss to enhance network feature extraction ability and solve the over-exposure problem, aiming to further improve the visual quality of enhanced results.
- We build a new large-scale dataset consisting of high-quality low/normal-light image pairs captured in complex and diverse scenes from the real world. To the best of our knowledge, the PNLI dataset is the currently largest real-world paired images dataset for low-light image enhancement.

2. Related work

In this section, we briefly review the related research in low-light image enhancement, which includes traditional methods and deep learning methods.

2.1. Traditional methods

Histogram equalization (HE) [15] and Retinex theory [16,17] are the most primarily and widely used methods in low-light image enhancement. The primary HE method often causes loss of contextual details, poor color restoration, and noise disturbance. Thus, many algorithms have been proposed to solve these issues, *e.g.*, adaptive histogram equalization [18], contrast-limited adaptive histogram equalization [19], dualistic sub-image histogram equalization method [20], and brightness bi-histogram equalization method [21]. However, these methods fail to solve the color bias in low-light enhancement completely. Retinex-based methods [5,22] decompose images into reflectance and illumination and then enhance images by manipulating illumination. Celik et al. [23] and Lee et al. [24] utilize the relationship between adjacent pixels and large gray-level differences to adjust brightness at local levels. Land and Jobson propose Retinex [16] and multi-scale Retinex model [25] to formulate the light enhancement as an illumination estimation problem. More recently, Wang et al. [26] propose an enhancement algorithm for non-uniform illumination images, utilizing a bi-log transformation to make a balance between details and naturalness. Based on the previous investigation of the logarithmic transformation, Fu et al. [22] propose a weighted variational model to estimate both the reflectance and the illumination from an observed image with imposed regularization terms. Nevertheless, these methods cannot always achieve satisfying performance and even generate additional artifacts.

2.2. Deep learning methods

Recently, deep learning has achieved great success in image restoration/enhancement tasks [27–31]. In this section, we introduce some deep learning low-light image enhancement methods, which can be divided into fully supervised methods and unsupervised methods.

Fully supervised methods. Most existing works for low-light image enhancement rely on paired images for the image enhancement task [32–35]. Chen et al. [36] use an end-to-end network to obtain enhanced images from extremely low-light raw images. Kin et al. [37] introduce a stacked auto-encoder to learn low-light enhancement with denoising. In addition, some fully supervised methods incorporate the traditional Retinex theory-based methods with CNNs to obtain enhanced images. Retinex-Net [8] uses a decomposition network to decompose the input images into reflectance and illumination, and then an encoder–decoder network is used to adjust the illumination. Pineda et al. [38] extend Retinex-Net by incorporating another reflectance restoration network to improve the reflectance. All of these methods are highly dependent on the dataset and cannot work in real-world scenarios where no paired data exist.

Unsupervised methods. Benefit from the development of unsupervised learning methods [11,12] in image processing, many unsupervised methods have been proposed to address the task of low-light enhancement. Jiang et al. [10] propose EnlightenGAN, which is a one-way GAN using global–local discriminator structure with a well-designed self feature preserving loss. Guo et al. [9] present a method named Zero-Reference Deep Curve

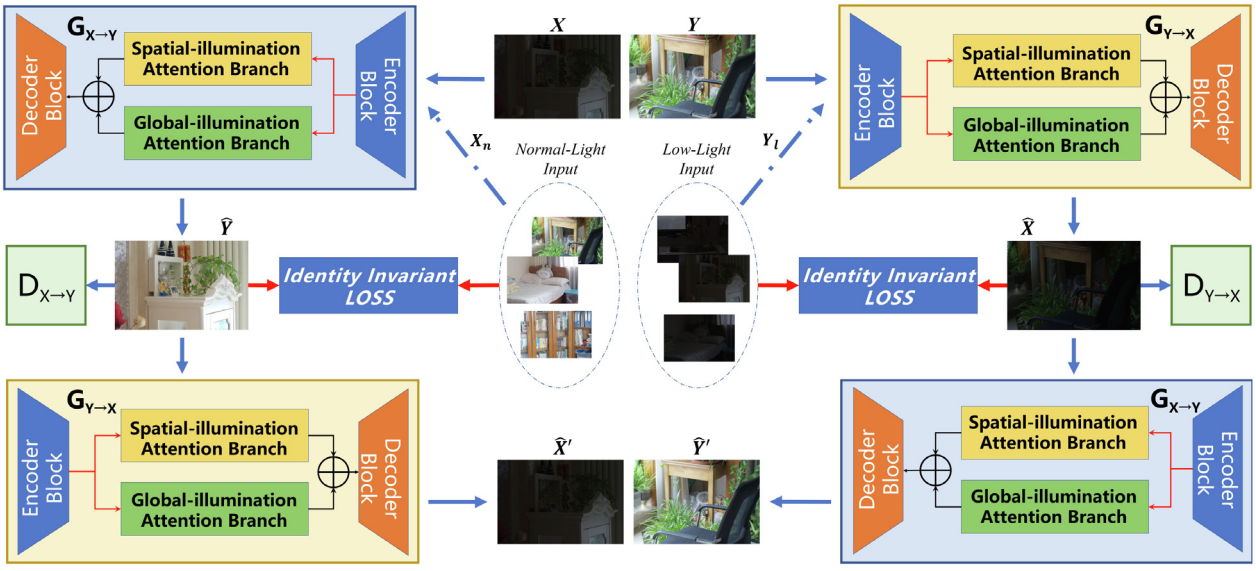

Fig. 2. The framework of LE-GAN. The model consists of two generative adversarial networks, $G_{X\to Y}$ and $G_{Y\to X}$ that form a cyclic network.

Estimation (Zero-DCE), which formulates low-light enhancement as a task of image-specific curve estimation with a deep network. Although these methods solve the problem of unpaired data, the quality of the enhanced images is limited, *e.g.*, these methods enhance the holistic brightness of the input but cause over-exposure in the regions with relatively high brightness. Besides, these methods cannot deal with the noise well simultaneously, which has a significant impact on the visual quality of the enhanced images.

3. Method

As illustrated in Fig. 2, our LE-GAN is designed to resolve the low-light image enhancement issue in an unsupervised way. In this section, we first introduce the network architectures of our LE-GAN. Then, we introduce our illumination-aware attention module, Finally, the identify loss function as well as other loss functions are described.

3.1. Network architectures

Our LE-GAN consists of two generative adversarial networks forming a cycle network [13]. $G_{X\to Y}$ takes the low-light image $\boldsymbol{X}$ as input and then outputs a normal-light image $\hat{\boldsymbol{Y}}$. $G_{Y\to X}$ generates the inverse low-light image $\hat{\boldsymbol{X}}'$ from the corresponding normal-light image $\hat{\boldsymbol{Y}}$. $D_{Y\to X}$ is used to distinguish between input image $\boldsymbol{X}$ and the generated low-light images $\hat{\boldsymbol{X}}$ and $D_{X\to Y}$ aims to discriminate between the normal-light image $\boldsymbol{Y}$ and the enhanced result $\hat{\boldsymbol{Y}}$. Besides, X_n and Y_l are selected from *Normal-Light Input* and *Low-Light Input*, respectively, and they are served as the input of the generator for the calculation of the identity invariant loss. More details can be found in Section 3.3.

We adopt a U-Net [39] style network with a spatial-illumination attention branch and a global-illumination attention branch as the generators. Firstly, the encoder consisting of several Resnet-D blocks extracts features and maps the input low-light images to high-dimensional representations. Then, these intermediate features are fed into two attention branches respectively. After that, the output feature maps are elementally added to obtain the representation of the enhanced result. At last, the decoder is composed of several bilinear interpolation operations and convolution(BIC) layers, aiming to reconstruct the enhanced normal-light image from the mapped features. Besides, skip connections are utilized between the encoder layers and the corresponding decoder layers to make full use of hierarchical multi-scale information from both low-level features and high-level features. For the discriminators, we use the commonly used VGG architecture to identify real and fake images.

3.2. Illumination-aware attention module

To handle the noise, low brightness, and color bias simultaneously, we propose the illumination-aware attention module consisting of a spatial-illumination attention branch and a global illumination attention branch is shown in Fig. 3.

Spatial-illumination attention branch. In a normal-light image, the illumination of pixels may change dramatically in a small neighborhood, *e.g.*, two pixels are next to each other and one is in the shadow while the other one is in the light. The features of regions to which the two pixels belong should be quite different in the convolutional neural network. However, due to the nature of the convolution kernel, its receptive field is usually small and limited to the local area. It often ignores the contextual information and makes features of adjacent pixels similar to each other. This leads to inaccurate brightness estimation and noise suppression, which makes low-light enhancement results unsatisfying. Therefore, we add spatial-illumination attention branch to our generator, which uses non-local attention, and can encode a wider range of contextual information into local features, *i.e.*, the features of all pixels in the shadow together and making them more discriminative and different from features of neighbor pixels in the light.

As shown in Fig. 3, the spatial-illumination attention branch first takes the encoded representation as input and generates three intermediate features, *i.e.*, $\mathbf{f}(\mathbf{x})$, $\mathbf{g}(\mathbf{x})$, and $\mathbf{h}(\mathbf{x})$. Then, We calculate the dot products of $\mathbf{f}(\mathbf{x})$ with $\mathbf{g}(\mathbf{x})$ and apply the softmax function to obtain the weights on $\mathbf{h}(\mathbf{x})$. Finally, we further calculate the dot products of weights with all $\mathbf{h}(\mathbf{x})$ to attain the spatial-illumination feature map.

Global-illumination attention branch. In high-level features, each channel can be regarded as a kind of semantic response that is relevant to illumination estimation and final normal-light image generation. Considering different semantic responses contribute to the final estimation differently, the less relevant or irrelevant semantic responses may introduce noise and lead to

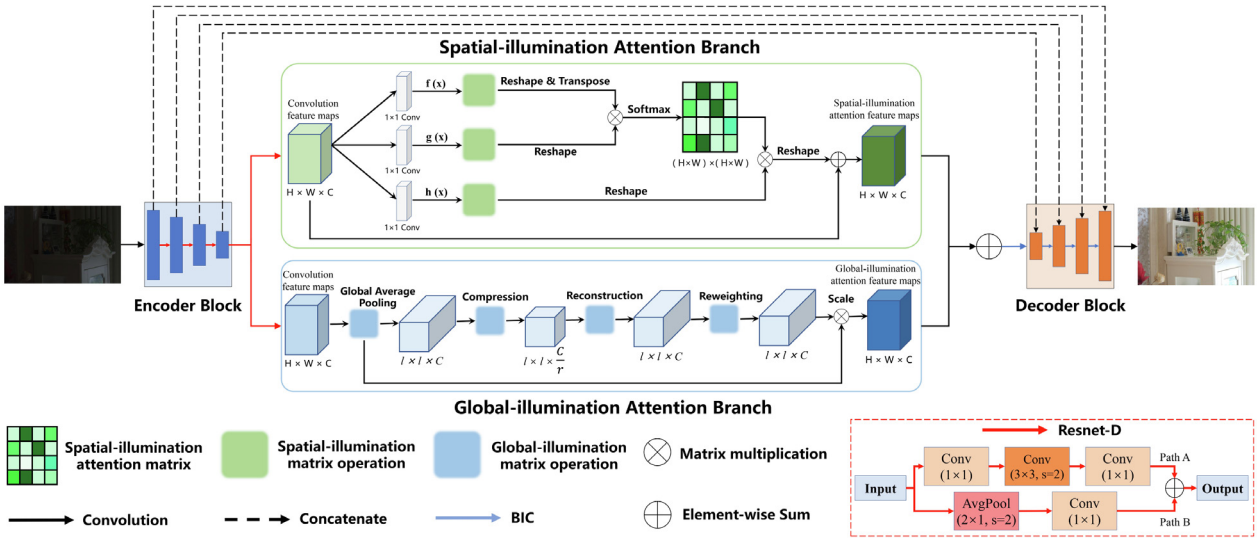

Fig. 3. The architecture of the proposed generator, which consists of an encoder–decoder module and an illumination-aware attention module. In the encoder block, the Resnet-D (red arrows) blocks are used to downsample feature maps; while in the decoder block, the **B**ilinear **I**nterpolation operations and **C**onvolution (BIC, blue arrows) layers are used to upsample the feature maps.

color bias. From the view of low-level vision, different channels may represent different global information, thus we believe that channel attention is helpful to utilize color information adaptively. Therefore, we propose a global-illumination attention branch that allows the network to learn to use global information to selectively emphasize informative semantic responses and suppress less useful ones. Such a global-illumination attention branch can help the network achieve better color estimation and noise suppression, and make the enhanced images realistic.

As illustrated in Fig. 3, the global-illumination attention branch is built upon a transformation, which maps the encoded features into the global-illumination attention feature maps. The input is first processed by a global average pooling in spatial dimensions. Then features encoded in channel dimension are passed into two fully connected layers for compression and reconstruction. After that, the reweighting attention map is multiplied by each channel of the original input of the branch. At last, feature maps generated by the spatial-illumination and the global-illumination attention branch are elementally added as the intermediate feature that passes into the decoder.

3.3. Loss functions

In this subsection, we introduce the loss functions of our method. We first propose a novel identify invariant loss to solve the over-exposure problem in the low-light enhancement task. Besides, we describe the adversarial loss and the cycle consistency loss for the training.

Identity invariant loss. Over-exposure problem leads to loss of image details, and it usually occurs in the relatively bright region of low-light images where enhancement methods tend to improve the overall brightness. These bright regions in low-light images account for a small proportion of pixels and have different brightness distributions from most of the pixels in the low-light image, which makes them hard samples. To solve this problem, we randomly input bright normal-light images to give the model more samples of the bright region and propose an identity invariant loss, imposing an identity constraint on the output, avoiding over-enhancement for the bright region. In this way, the model is able to learn to identify the bright regions and enhance them adaptively to avoid over-exposure. The identity invariant loss can be defined as:

$$\mathcal{L}_{Identity} = \mathcal{L}_{Identity_l} + \mathcal{L}_{Identity_n}, \tag{1}$$

where

$$\mathcal{L}_{Identity_l} = |Y_l - \mathrm{G}_{\mathrm{Y}\to\mathrm{X}}(Y)|, \tag{2}$$

and

$$\mathcal{L}_{Identity_n} = |X_n - \mathrm{G}_{\mathrm{X}\to\mathrm{Y}}(X)|. \tag{3}$$

Adversarial loss. The adversarial loss is used to encourage the distribution of the enhanced image to be close to the normal-light image, and can be described as:

$$\mathcal{L}_{\mathrm{G}_{\mathrm{X}\to\mathrm{Y}}} = -\log\left(1 - \mathrm{D}_{\mathrm{X}\to\mathrm{Y}}\left(\mathrm{G}_{\mathrm{X}\to\mathrm{Y}}(X)\right)\right). \tag{4}$$

And the discriminative $\mathrm{D}_{\mathrm{X}\to\mathrm{Y}}$ is defined as:

$$\mathcal{L}_{\mathrm{D}_{\mathrm{X}\to\mathrm{Y}}} = \log\left(\mathrm{D}_{\mathrm{X}\to\mathrm{Y}}(X_n)\right) + \log\left(1 - \mathrm{D}_{\mathrm{X}\to\mathrm{Y}}\left(\mathrm{G}_{\mathrm{X}\to\mathrm{Y}}(X)\right)\right). \tag{5}$$

Cycle consistency loss. Inspired by CycleGAN [13], to make different regions of the images generated by the two generators correspond to each other, we define our consistency loss as:

$$\mathcal{L}_{cyc_x} = |X - \hat{X}'|, \quad \text{where } \hat{X}' = \mathrm{G}_{\mathrm{Y}\to\mathrm{X}}\left(\mathrm{G}_{\mathrm{X}\to\mathrm{Y}}(X)\right). \tag{6}$$

Thus, we can obtain the final cyclic consistency loss as:

$$\mathcal{L}_{cycle} = \mathcal{L}_{cyc_x} + \mathcal{L}_{cyc_y}. \tag{7}$$

The total loss for $\mathrm{G}_{\mathrm{X}\to\mathrm{Y}}$ and $\mathrm{G}_{\mathrm{Y}\to\mathrm{X}}$ is a combination of all these losses and can be expressed as:

$$\mathcal{L}_{sum_G} = \mathcal{L}_{\mathrm{G}_{\mathrm{X}\to\mathrm{Y}}} + \mathcal{L}_{\mathrm{G}_{\mathrm{Y}\to\mathrm{X}}} + \lambda_1\mathcal{L}_{Identity} + \lambda_2\mathcal{L}_{cycle}, \tag{8}$$

where λ_1 and λ_2 are the loss weights and we empirically set them to 5 and 10 respectively. The overall loss of the discriminators is:

$$\mathcal{L}_{sum_D} = \mathcal{L}_{\mathrm{D}_{\mathrm{X}\to\mathrm{Y}}} + \mathcal{L}_{\mathrm{G}_{\mathrm{Y}\to\mathrm{X}}}. \tag{9}$$

4. Paired normal/low-light images dataset

Deep learning methods for low-light image enhancement based on the synthetic dataset [7,8,37,40] have been studied

Fig. 4. Several representative examples for low/normal-light images in PNLI dataset, LOL dataset, SYN dataset and EnlightenGAN dataset. Objects and scenes captured in our PNLI dataset are more diverse, abundant and superior.

Table 1
Comparison of attributes between our PNLI dataset and the other representative datasets. SYN denotes SYNthesized dataset, EnlightenGAN denotes the dataset used in EnlightenGAN, LOL denotes LOw-Light dataset.

	Dataset	Scale	Paired/Unpaired	Size
Synthetic	SYN	1,000 image pairs	Paired	384 × 384
	EnlightenGAN	914 low-light images1,016 normal-light images	Unpaired	600 × 400
Real-scenes	LOL	500 image pairs	Paired	600 × 400
	Ours (PNLI)	**2,000 image pairs**	Paired	**6720 × 4480**

for several years. However, the synthesized images of existing datasets are usually not photo-realistic, so that models designed and trained on these datasets perform poorly in real-world low-light images. Several real-world datasets are proposed recently [8,10]. Nonetheless, the data is unpaired in [10], and the LOL dataset [8] has a small scale with limited scenes. Considering the defects of existing datasets, we propose a new dataset named **P**aired **N**ormal/**L**ow-light **I**mages (**PNLI**), which contains 2,000 low/normal-light image pairs. Compared to previous datasets, our dataset has larger scale and richer scenes. Besides, the images in our dataset have a higher resolution.

We use Canon EOS 5D Mark IV to capture the data. The resolution of captured images is set to 6720 × 4480. To capture low/normal-light image pairs, the camera was mounted on a sturdy tripod and controlled remotely via a mobile APP. The camera was not touched between the capture process of normal-light and low-light images to avoid vibration. For each pair, we first take the normal-light image. Then the low-light image is captured by changing the shutter, exposure time, and ISO to simulate low-light conditions.

We capture the images in a variety of scenes, *e.g.*, museums, parks, streets, landscapes, vehicles, plants, buildings, symbols, and furniture. Among these images, the quantity of outdoor images is almost three times bigger than that of indoor images. It is noteworthy that all the scenes in our dataset are static to ensure that the content of the low-light image and its ground-truth are identical. Some representative visual examples of PNLI (Paired Normal/Low-light Images) dataset, LOL (LOw-Light) dataset [8], SYN (SYNthesized) dataset [8], and unpaired dataset used in EnlightenGAN [10] are shown in Fig. 4. To the best of our knowledge, PNLI is the largest real-world paired low/normal-light image dataset for low-light enhancement and will be publicly available. Table 1 shows the comparison of the important attribute (*i.e.*, scale, paired or not, image size) between our PNLI and other representative datasets. Compared to other paired normal/low-light image datasets, PNLI is far more diverse, comprehensive, and challenging. It exhibits the following distinctive characteristics and superiority:

- It contains 2,000 image pairs, which is four times the size of the LOL dataset.
- Different from the existing real scenes dataset, *i.e.*, LOL, there are no repeated scenes in our PNLI dataset, which is more abundant and superior than LOL. (There are many very similar scenes with little difference in the LOL dataset, as shown in Fig. 4)
- All images in PNLI are collected from considerably more real scenes, which contain both indoor and outdoor scenes. In addition, the object categories in images are rich and common.
- Excellent visual quality and clarity, which might help in learning pixel-level contextual information.
- The darkness levels of low-light images in PNLI are rich, and it can truly restore various situations where the actual image brightness is missing due to insufficient ambient light or human operation mistakes. Therefore, it can effectively verify the stability and robustness of our proposed method.

5. Experimental results

In this section, we first introduce the implementation details of our model. Then, we describe the datasets and their usage for training and testing of all methods, and the metrics for quantitative evaluation. Next, we compare our method with several state-of-the-art methods of various types. In addition, the effectiveness of our method is evaluated on two real-world

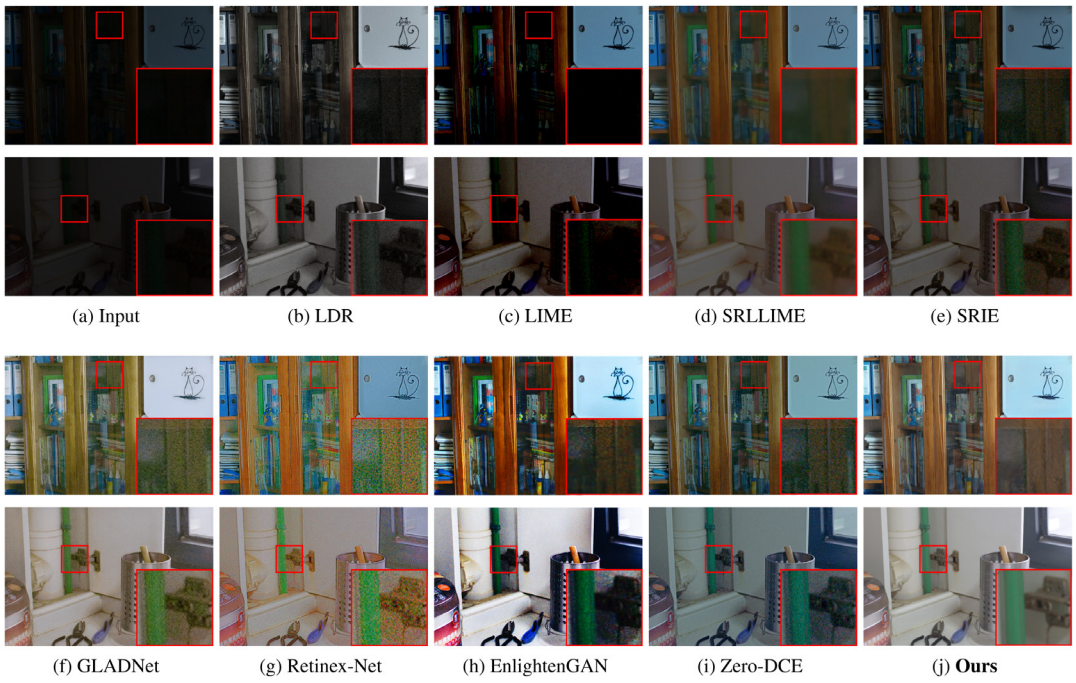

Fig. 5. Visual comparison with other different methods on the LOL dataset [8]. The red boxes represent the saliency regions of the results.

datasets, *i.e.*, PNLI and LOL; a synthesized dataset [8], *i.e.*, SYN; and ExDark [14] dataset containing only low-light images. Finally, we further investigate the effect of our proposed network module and loss function through an ablation study.

5.1. Implementation details

We perform the center crop of size 256 $\times$ 256 to train the model up to 4,000 epochs with Adam optimizer. The learning rate is set to 1e-4, the hyper-parameters β_1, β_2, ε, λ_1, and λ_2 are set to 0.9, 0.999, 1e-8, 5, and 10, respectively. GFLOPs and parameters of the model are 62.66Mb and 7.91Gb. The training time and converging time of our method are 25.6 h and 10.9 h (on the PNLI dataset) on 3 NVIDIA 3090ti GPUs, respectively. And its testing time on a 600 $\times$ 400 image is about 8.43 ms.

5.2. Datasets and metrics

LOL dataset has 500 image pairs, and we randomly select 485 pairs for training and 15 pairs for testing. SYN dataset has 1,000 image pairs, and we randomly select 950 pairs for training and 50 pairs for testing. As for the PNLI dataset with 2,000 image pairs, we randomly select 1,700 pairs for training and 300 pairs for testing. We compare supervised and unsupervised methods on the same dataset. For the supervised methods, we directly use the paired data to train the model. For the unsupervised methods, to ensure that the model cannot see the paired low-light/normal-light images in one iteration, we independently shuffle the low-light images (*e.g.*, 1700 images in the training set of PNLI) and the normal-light images (*e.g.*, 1700 images in the training set of PNLI). The paired dataset is treated as an unpaired dataset for unsupervised methods for a fair comparison. Two image quality metrics are used including Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM).

5.3. Comparison with state-of-the-art methods

We compare our methods with several traditional methods, *i.e.*, LDR [24], LIME [5], SRIE [22] and SRLLIME [6], and several state-of-the-art deep learning methods, *i.e.*, GLADNet [7], Retinex-Net [8], EnlightenGAN [10], and Zero-DCE [9]. All these methods are retrained using the official codes on LOL, SYN, and our PNLI dataset. The experimental results are reported both quantitatively and qualitatively.

5.3.1. Qualitative comparisons

We first compare the visual quality of our LE-GAN with other methods. The results are shown in Figs. 5 and 6. It can be seen that the traditional methods get the worst results, especially on the LOL dataset. Compared to the deep learning methods, these methods can only increase the brightness of the image, but the color saturation of the results is still very low. The deep learning methods can perform low-light enhancement better, but they still suffer from noise and color bias. Besides, we can easily observe that there are some regions of the results generated by the compared methods with relative brightness that are over-exposed after enhancement. In contrast, our method performs the best on all conditions with nearly no artifacts and generates the most realistic normal-light images.

5.3.2. Quantitative comparisons

We also provide quantitative comparisons of our methods with state-of-the-art methods. We report the PSNR and SSIM results of each approach on the PNLI, LOL, and SYN datasets. As shown in Table 2, our method outperforms the other state-of-the-art methods significantly. It is worth noting that our method is unsupervised while it still outperforms current state-of-the-art fully supervised methods, including GLADNet and Retinex-Net. The results strongly prove the effectiveness of our method.

To prove that the improvement of our method does not benefit from more complex models, we do the complexity analysis. Our method has similar GFLOPs to RetinexNet and Enlightten-GAN, and the GFLOPs of GLADNet is larger than that of ours. When it comes to model complexity, though the amount of model parameters of our methods is larger than that of GLADNet, RetinexNet, and Zero-DCE. The PSNR of our method outperforms other methods by a large margin of 3-—5 dB on the PNLI dataset.

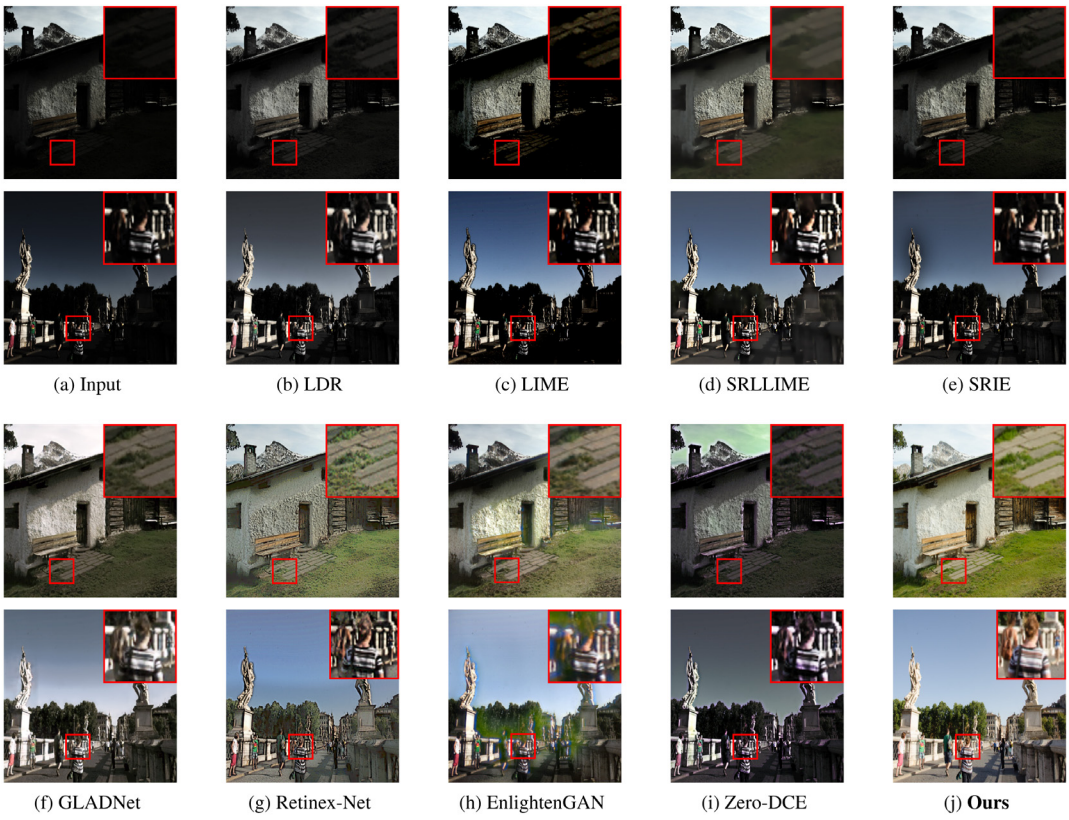

Fig. 6. Visual comparison with other different methods on the SYN dataset [8]. The red boxes represent the saliency regions of the results.

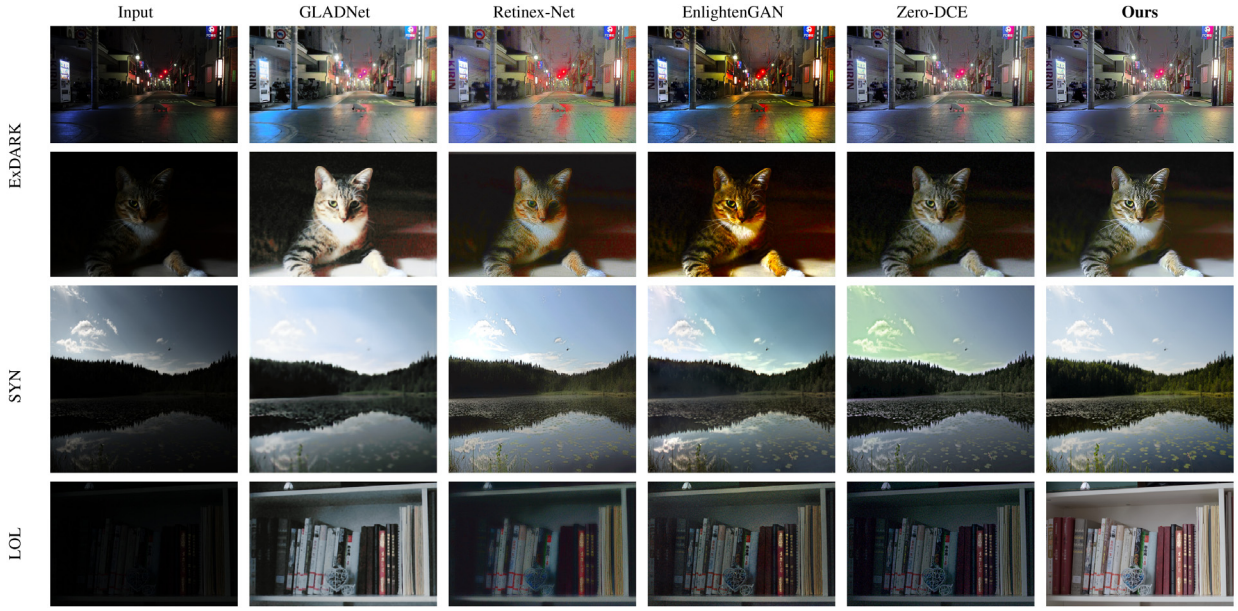

Fig. 7. Visual comparison with other different methods on ExDARK, SYN, and LOL in a zero-shot manner respectively. Our method and other representative deep learning methods are only trained on our PNLI dataset. Our method generates the most visually pleasing results across the three datasets. Please zoom in to see the details.

5.3.3. Generalization ability comparison

Generalization ability is of vital importance in evaluating deep learning algorithms. In this section, we conduct extensive experiments to compare the generalization ability between our methods and state-of-the-art methods. As shown in Table 3, we train models on our PNLI dataset and then test them on LOL and SYN datasets (the left part) and train models on LOL and SYN datasets and then test them on PNLI (the right part). Our method obtains the best performance among all the methods, which illustrates that our trained model is robust and can generalize to more data.

We also provide qualitative results on LOL, SYN, and ExDARK datasets to show the visual quality of these methods trained on PNLI. Since the ExDARK dataset does not have ground-truth,

Fig. 8. Visual comparison with other representative unsupervised low-light image enhancement methods (*i.e.*, EnlightenGAN [10] and Zero-DCE [9]) on PNLI testing set. The red dotted boxes represent the saliency regions of the results.

we only show the visual comparisons without the reference. As shown in Fig. 7, the results generated by other deep learning methods suffer from low color saturation and color bias, especially the predictions of Retinex-Net and EnlightenGAN. However, our method can generate visually satisfying results.

In addition, we conduct experiments of domain adaptation to further show the generalization ability of our method. For this experiment, we compare our method with the other two unsupervised methods, *i.e.*, EnlightenGAN and Zero-DCE, using normal-light data in PNLI and low-light data in LOL as training set and testing the trained models on the PNLI's testing set. In this case, the algorithm needs not only to perform low-light enhancement using unpaired data but also to deal with the domain gap of content between the source domain and the target domain. The visual comparisons are shown in Fig. 8, and the results generated by our method achieve the best visual quality with less color bias compared to EnlightenGAN and Zero-DCE. It can be observed that our method can not only solve the unpaired low-light enhancement but also handle the domain adaptation well. Overall, the above experiments can strongly prove the effectiveness and generalization ability of our method.

5.4. Ablation study

In this section, we conduct an ablation study to further investigate the proposed LE-GAN, including the spatial-illumination attention branch, global-illumination attention branch, and the identity invariant loss.

The importance of the illumination attention branches and identity invariant loss. The quantitative results are demonstrated in Table 4. Compared to the complete method, the PSNR and SSIM significantly drop without the spatial-illumination attention branch. The second important module is the global-illumination attention branch. The identity invariant loss has the least impact on the quantitative metrics, but it still brings about a 2 dB PSNR drop without this loss. The quantitative results validate the effectiveness of the spatial-illumination attention branch, global-illumination attention branch and identity invariant loss.

Table 2
Quantitative results on the PNLI, LOL, and SYN datasets.

Methods	PNLI		LOL		SYN	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
LDR [24]	15.480	0.632	15.484	0.634	13.187	0.591
LIME [5]	13.625	0.412	15.414	0.433	14.318	0.554
SRLLIME [6]	15.953	0.703	13.629	0.706	16.873	0.714
SRIE [22]	16.416	0.617	17.440	0.649	14.478	0.639
GLADNet [7]	21.127	0.773	20.314	0.739	16.761	0.797
Retinex-Net [8]	18.857	0.743	17.780	0.425	16.286	0.779
EnlightenGAN [10]	22.066	0.830	18.850	0.736	16.073	0.827
Zero-DCE [9]	19.083	0.772	16.818	0.741	15.600	0.796
Ours	**24.176**	**0.876**	**22.449**	**0.886**	**24.014**	**0.899**

The superiority of our attention-based methods. In Table 5, we replace our spatial illumination and global illumination attention branches with residual blocks in $G_{x\to y}$ and $G_{y\to x}$ and we denote this model as ours w/o attention. And we also replace the spatial-illumination and global-illumination attention branches with the attention components of DuATM [41], Bilinear CNN [42], and SE-Block [43,44], respectively (denoted as Ours-DuATM, Ours-Bilinear CNN and Ours-SE). As shown in Table 5, the PSNR and SSIM of ours are the highest of all methods, which means that our illumination-aware attention method has more advantages in obtaining the global information of the low-light images. SE-Blocks focuses on the correlation between channels of feature maps and has weaker ability to obtain global information. Please note that all the comparison methods have similar settings (*e.g.*, model size) to ensure fairness.

To further illustrate the impact of the two attention branches and the identity invariant loss on the visual quality, we also provide qualitative results in Fig. 9. We can observe that the two attention branches can significantly improve the visual quality by reducing noise and color bias. Additionally, benefiting from the identify invariant loss, the over-exposure problem can be solved well obviously. The visual results further prove the effectiveness of the proposed modules.

Fig. 9. Visual comparison from the ablation study of our method. Not only the indoor scenes but the outdoor scenes are illustrated. Red and blue boxes represent the key details of the images, and our method performs well in these areas. From left to right, (a) Input. (b) Normal. (c) Ours. (d) w/o Spatial-illumination attention. (e) w/o Global-illumination attention. (f) w/o Identity Invariant Loss.

Table 3
Experimental results about generalization ability.

Methods	Training Set	Testing Set	PSNR	SSIM	Training Set	Testing Set	PSNR	SSIM
GLADNet [7]	PNLI	LOL SYN	18.637 17.035	0.792 0.716	LOL SYN	PNLI	16.912 17.125	0.757 0.730
Retinex-Net [8]	PNLI	LOL SYN	16.035 18.293	0.690 0.800	LOL SYN	PNLI	16.757 17.884	0.764 0.767
EnlightenGAN [10]	PNLI	LOL SYN	18.220 18.707	0.677 0.802	LOL SYN	PNLI	18.117 17.022	0.692 0.646
Zero-DCE [9]	PNLI	LOL SYN	14.213 17.350	0.611 0.847	LOL SYN	PNLI	14.332 17.249	0.733 0.784
Ours	PNLI	LOL SYN	**21.523** **20.011**	**0.812** **0.850**	LOL SYN	PNLI	**20.410** **17.929**	**0.815** **0.798**

Table 4
Ablation study on the PNLI dataset.

Spatial-illumination Attention branch	Global-illumination Attention branch	Identity Invariant Loss	PSNR	SSIM
×	√	√	19.677	0.630
√	×	√	21.825	0.735
√	√	×	22.385	0.724
√	√	√	**24.176**	**0.876**

Table 5
The comparison of attention components.

Method	PSNR	SSIM
Ours w/o attention	19.291	0.627
Ours-DuATM	19.828	0.817
Ours-Bilinear CNN	21.459	0.820
Ours-SE	22.348	0.839
Ours	**24.176**	**0.876**

6. Conclusion

We design a novel unsupervised low-light image enhancement network named LE-GAN. In our model, the illumination-aware attention module consisting of a spatial-illumination attention branch and a global-illumination attention branch is proposed to solve the low brightness with the reduction of noise and color bias. Meanwhile, a novel identity invariant loss is introduced to address the over-exposure problem. Besides, we also build the currently largest real-world paired low-light/normal-light image benchmark dataset for the low-light image enhancement, which consists of large amounts of high-quality images collected from different real-world scenes under different light conditions. The qualitative and quantitative experimental results on various low-light datasets show that our approach outperforms the state-of-the-art approaches. Furthermore, we demonstrate that LE-GAN can obtain preferable generalization ability on several datasets and yield more visually pleasing enhanced images.

CRediT authorship contribution statement

Ying Fu: Methodology, Supervision. **Yang Hong:** Data, Experiment, Writing. **Linwei Chen:** Software, Reviewing and editing. **Shaodi You:** Reviewing and editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research was funded by the National Natural Science Foundation of China under Grants No. 62171038, No. 61827901, and No. 62088101.

References

[1] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. 39 (6) (2016) 1137–1149, http://dx.doi.org/10.1109/TPAMI.2016.2577031.

[2] F. Pérez-Hernández, S. Tabik, A. Lamas, R. Olmos, H. Fujita, F. Herrera, Object detection binary classifiers methodology based on deep learning to identify small objects handled similarly: Application in video surveillance, Knowl.-Based Syst. 194 (2020) 105590, http://dx.doi.org/10.1016/j.knosys.2020.105590.

[3] J. Yuan, X. Hou, Y. Xiao, D. Cao, W. Guan, L. Nie, Multi-criteria active deep learning for image classification, Knowl.-Based Syst. 172 (2019) 86–94, http://dx.doi.org/10.1016/j.knosys.2019.02.013.

[4] Z. Ding, T. Wang, Q. Sun, Q. Cui, F. Chen, A dual-stream framework guided by adaptive gaussian maps for interactive image segmentation, Knowl.-Based Syst. 223 (2021) 107033, http://dx.doi.org/10.1016/j.knosys.2021.107033.

[5] X. Guo, Y. Li, H. Ling, Lime: Low-light image enhancement via illumination map estimation, IEEE Trans. Image Process. 26 (2) (2016) 982–993, http://dx.doi.org/10.1109/TIP.2016.2639450.

[6] M. Li, J. Liu, W. Yang, X. Sun, Z. Guo, Structure-revealing low-light image enhancement via robust retinex model, IEEE Trans. Image Process. 27 (6) (2018) 2828–2841, http://dx.doi.org/10.1109/TIP.2018.2810539.

[7] W. Wang, C. Wei, W. Yang, J. Liu, Gladnet: Low-light enhancement network with global awareness, in: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition, FG 2018, IEEE, 2018, pp. 751–755.

[8] C. Wei, W. Wang, W. Yang, J. Liu, Deep retinex decomposition for low-light enhancement, arXiv preprint arXiv:1808.04560.

[9] C. Guo, C. Li, J. Guo, C.C. Loy, J. Hou, S. Kwong, R. Cong, Zero-reference deep curve estimation for low-light image enhancement, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 1780–1789.

[10] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, Z. Wang, Enlightengan: Deep light enhancement without paired supervision, IEEE Trans. Image Process. 30 (2021) 2340–2349, http://dx.doi.org/10.1109/TIP.2021.3051462.

[11] Y.-S. Chen, Y.-C. Wang, M.-H. Kao, Y.-Y. Chuang, Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6306–6314.

[12] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1125–1134.

[13] Y.P. Loh, C.S. Chan, Getting to know low-light images with the exclusively dark dataset, Comput. Vis. Image Underst. 178 (2019) 30–42, http://dx.doi.org/10.1016/j.cviu.2018.10.010.

[14] L. Tao, C. Zhu, G. Xiang, Y. Li, H. Jia, X. Xie, Llcnn: A convolutional neural network for low-light image enhancement, in: 2017 IEEE Visual Communications and Image Processing, VCIP, IEEE, 2017, pp. 1–4.

[15] M. Abdullah-Al-Wadud, M.H. Kabir, M.A.A. Dewan, O. Chae, A dynamic histogram equalization for image contrast enhancement, IEEE Trans. Consum. Electron. 53 (2) (2007) 593–600, http://dx.doi.org/10.1109/ICCE.2007.341567.

[16] E.H. Land, The retinex theory of color vision, Sci. Am. 237 (6) (1977) 108–129, http://dx.doi.org/10.1038/scientificamerican1277-108.

[17] Y. Wang, Y. Cao, Z.-J. Zha, J. Zhang, Z. Xiong, W. Zhang, F. Wu, Progressive retinex: Mutually reinforced illumination-noise perception network for low-light image enhancement, in: Proceedings of the 27th ACM International Conference on Multimedia, 2019, pp. 2015–2023.

[18] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 2223–2232.

[19] S.M. Pizer, Contrast-limited adaptive histogram equalization: speed and effectiveness, in: Stephen M. Pize, R. Eugene Johnston, James P. Ericksen, Bonnie C. Yankaskas, Keith E. muller (Eds.), in: Proceedings of the First Conference on Visualization in Biomedical Computing, Atlanta Georgia, vol. 337, Medical image display research group, 1990.

[20] Y. Wang, Q. Chen, B. Zhang, Image enhancement based on equal area dualistic sub-image histogram equalization method, IEEE Trans. Consum. Electron. 45 (1) (1999) 68–75, http://dx.doi.org/10.1109/30.754419.

[21] Z. Pan, M. Yu, G. Jiang, H. Xu, Z. Peng, F. Chen, Multi-exposure high dynamic range imaging with informative content enhanced network, Neurocomputing 386 (2020) 147–164, http://dx.doi.org/10.1016/j.neucom.2019.12.093.

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

[23] T. Celik, T. Tjahjadi, Contextual and variational contrast enhancement, IEEE Trans. Image Process. 20 (12) (2011) 3431–3441, http://dx.doi.org/10.1109/TIP.2011.2157513.

[24] C. Lee, C. Lee, C.-S. Kim, Contrast enhancement based on layered difference representation of 2d histograms, IEEE Trans. Image Process. 22 (12) (2013) 5372–5384, http://dx.doi.org/10.1109/TIP.2013.2284059.

[25] D.J. Jobson, Z.-u. Rahman, G.A. Woodell, A multiscale retinex for bridging the gap between color images and the human observation of scenes, IEEE Trans. Image Process. 6 (7) (1997) 965–976, http://dx.doi.org/10.1109/83.597272.

[26] S. Wang, J. Zheng, H.-M. Hu, B. Li, Naturalness preserved enhancement algorithm for non-uniform illumination images, IEEE Trans. Image Process. 22 (9) (2013) 3538–3548, http://dx.doi.org/10.1109/TIP.2013.2261309.

[27] C. Tian, R. Zhuge, Z. Wu, Y. Xu, W. Zuo, C. Chen, C.-W. Lin, Lightweight image super-resolution with enhanced cnn, Knowl.-Based Syst. 205 (2020) 106235, http://dx.doi.org/10.1016/j.knosys.2020.106235.

[28] C. Ren, X. He, L. Qing, Y. Wu, Y. Pu, Remote sensing image recovery via enhanced residual learning and dual-luminance scheme, Knowl.-Based Syst. 222 (2021) 107013, http://dx.doi.org/10.1016/j.knosys.2021.107013.

[29] C. Wang, H.-Z. Shen, F. Fan, M.-W. Shao, C.-S. Yang, J.-C. Luo, L.-J. Deng, Eaa-net: A novel edge assisted attention network for single image dehazing, Knowl.-Based Syst. 228 (2021) 107279, http://dx.doi.org/10.1016/j.knosys.2021.107279.

[30] M.A. Hedjazi, Y. Genc, Efficient texture-aware multi-gan for image inpainting, Knowl.-Based Syst. 217 (2021) 106789, http://dx.doi.org/10.1016/j.knosys.2021.106789.

[31] Y. Dun, Z. Da, S. Yang, Y. Xue, X. Qian, Kernel-attended residual network for single image super-resolution, Knowl.-Based Syst. 213 (2021) 106663, http://dx.doi.org/10.1016/j.knosys.2020.106663.

[32] C. Tian, Y. Xu, W. Zuo, B. Du, C.-W. Lin, D. Zhang, Designing and training of a dual cnn for image denoising, Knowl.-Based Syst. 226 (2021) 106949, http://dx.doi.org/10.1016/j.knosys.2021.106949.

[33] A. Luque-Chang, E. Cuevas, M. Pérez-Cisneros, F. Fausto, A. Valdivia-Gonzalez, R. Sarkar, Moth swarm algorithm for image contrast enhancement, Knowl.-Based Syst. 212 (2021) 106607, http://dx.doi.org/10.1016/j.knosys.2020.106607.

[34] G. Li, Y. Yang, X. Qu, D. Cao, K. Li, A deep learning based image enhancement approach for autonomous driving at night, Knowl.-Based Syst. 213 (2021) 106617, http://dx.doi.org/10.1016/j.knosys.2020.106617.

[35] S.M. Pizer, E.P. Amburn, J.D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J.B. Zimmerman, K. Zuiderveld, Adaptive histogram equalization and its variations, Comput. Vis. Graph. Image Process. 39 (3) (1987) 355–368, http://dx.doi.org/10.1016/S0734-189X(87)80186-X.

[36] C. Chen, Q. Chen, J. Xu, V. Koltun, Learning to see in the dark, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3291–3300.

[37] F. Lv, F. Lu, J. Wu, C. Lim, Mbllen: Low-light image/video enhancement using cnns, in: BMVC, 2018, pp. 220.

[38] X. Alameda-Pineda, S. Arias, Y. Ban, G. Delorme, L. Girin, R. Horaud, X. Li, B. Morgue, G. Sarrazin, Audio-visual variational fusion for multi-person tracking with robots, in: Proceedings of the 27th ACM International Conference on Multimedia, 2019, pp. 1059–1061.

[39] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[40] P.E. Trahanias, A.N. Venetsanopoulos, Color image enhancement through 3-d histogram equalization, in: 11th IAPR International Conference on Pattern Recognition. III. Conference C: Image, Speech and Signal Analysis, Vol. 1, IEEE Computer Society, 1992, pp. 545–548.

[41] J. Si, H. Zhang, C.-G. Li, J. Kuen, X. Kong, A.C. Kot, G. Wang, Dual attention matching network for context-aware feature sequence based person re-identification, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 5363–5372.

[42] T.-Y. Lin, A. RoyChowdhury, S. Maji, Bilinear cnn models for fine-grained visual recognition, in: Proceedings of the IEEE international conference on computer vision, 2015, pp. 1449–1457.

[43] H. Jie, S. Li, S. Gang, et al., Squeeze-and-excitation networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, Vol. 5, 2018.

[44] Z. Zheng, L. Zheng, Y. Yang, Pedestrian alignment network for large-scale person re-identification, IEEE Trans. Circuits Syst. Video Technol. 29 (10) (2018) 3037–3045, http://dx.doi.org/10.1109/TCSVT.2018.2873599.