Computers & Graphics 104 (2022) 95–105

Contents lists available at ScienceDirect

Computers & Graphics

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

Technical Section

Combining attention mechanism and Retinex model to enhance low-light images[☆,☆☆]

Yong Wang[a,1], Jin Chen[a,*,1], Yujuan Han[b], Duoqian Miao[c]

[a] School of Artificial Intelligence, Chongqing University of Technology, Chongqing, 400054, China
[b] School of Electronic Engineering, Xidian University, Xi'an, 710075, China
[c] Department of Computer Science and Technology, Tongji University, Shanghai, 201804, China

ARTICLE INFO

Article history:
Received 18 January 2022
Received in revised form 16 March 2022
Accepted 6 April 2022
Available online 14 April 2022

Keywords:
Low-light image enhancement
Retinex model
Attention map
Image decomposition
Image restoration

ABSTRACT

Low-light image enhancement is challenging due to intractable problems such as color distortion and noise, which hide in the dark. Simply enhancing the brightness of dark areas will inevitably amplify hidden artifacts. We have observed more noise in the underexposed areas of images than in the normally exposed areas. Attention mechanism can be used to emphasize the vital information of the processed object and suppress some irrelevant information. Inspired by these observations, we propose a deep network that Combines Attention mechanism and Retinex (CA&R Net) model to enhance low-light images. Firstly, we develop an attention map to evaluate the degree of image underexposure and guide enhancement in a region-adaptive manner. This way, it can enhance underexposed areas and avoid over-enhancing normally exposed areas. Secondly, we use the reconstructed reflectance and low illumination to predict the illumination layers of the image jointly. This joint prediction utilizes the attention mechanism, making illumination adjustment achieve better results. The quantitative experimental results show that the CA&R Net can successfully handle noise, color distortion, and multiple types of degradation with the power of attention information. Moreover, both SSIM and PSNR are better than other advanced methods.

1. Introduction

As digital imaging devices are increasingly widely deployed, people can take photos anytime and anywhere. However, the shooting scenes of many photos are underexposed. Photos taken under underexposure conditions are challenging to show the details of the scenery and people, which cannot meet people's ideal visual effects and needs. At the same time, these low-visibility photos also bring significant challenges to traditional computer vision tasks such as image segmentation, target detection [1], and tracking [2]. Therefore, designing a practical algorithm to enhance low-light images is necessary.

Although some existing technologies can enhance low-light images, such as setting long exposure, high ISO, and flash, there are some deficiencies with these methods. For example, long exposure has limitations when shooting static scenes, high sensitivity will increase noise and blur the images. High ISO increases the sensitivity of the image sensor to light while also amplifying noise. The utilization of flash can illuminate the environment to a certain extent. Still, it will introduce unexpected highlights and unbalanced light in the photo, making the photo visually unpleasant (see Fig. 1). .

Many researchers have conducted massive research and proposed many solutions to these problems. Early research [3–6] mainly focused on contrast enhancement. These methods have specific deficiencies in restoring image details and colors. Research in recent years [7–10] takes deep learning methods to adjust images. These methods simultaneously learn and adjust color, brightness, contrast, and saturation to achieve better results. However, these existing methods still have limitations for enhancing low-light images that are seriously underexposed.

In this paper, we propose the CA&R Net that combines attention mechanism and Retinex model to solve the above mentioned problems. Specifically, we suggest a novel information extraction network that learns to acquire the reflectance (R), illumination (I), and attention map (A_{map}) of the image. Then, A_{map} is used as a guide for the Restore-Net stage to restore the reflection component. Finally, the recovered reflectivity and low illumination

☆ This work was supported by the National Natural Science Foundation of China (Grant nos. 61502065, and 61976158).
☆☆ This article was recommended for publication by J. Zheng.

* Corresponding author.
E-mail addresses: ywang@cqut.edu.cn (Y. Wang), chenjin@2020.cqut.edu.cn (J. Chen), yujuanhan@stu.xidian.edu.cn (Y. Han), dqmiao@tongji.edu.cn (D. Miao).
1 Yong Wang and Jin Chen contributed equally to this work, and they are co-first authors.

https://doi.org/10.1016/j.cag.2022.04.002
0097-8493/

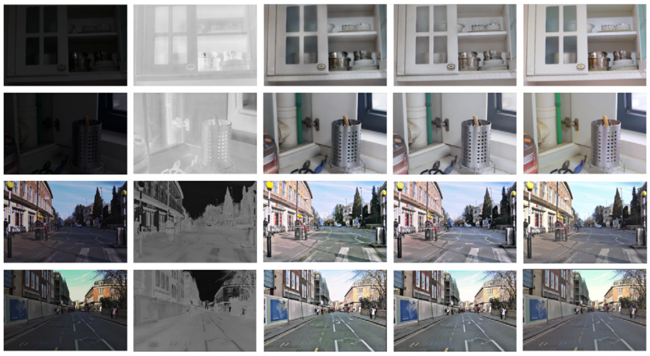

Fig. 1. Our low-light enhancement method can reconstruct the visual quality of underexposure images. Our result is a more natural color and lighting distribution guided by attentional information. Where (b) is the attention map of the image, which is used to evaluate the degree of image underexposure. For the source low-light image, the higher the illumination, the lower the attention map value, and its value range is [0, 1]. By inputting the attention map into the Attention Residual in Residual (ARIR) module, the network can determine which areas are underexposed and which sites are exposed according to their value and process them respectively to achieve a better enhancement effect.

are used to adjust the illumination component. The motivation of our design is that the noise in the underexposed areas of images is more severe than that in the normally exposed areas. Simple enhancement of underexposed areas will amplify the noise hidden in the dark. Therefore, we introduce an attention mechanism to emphasize the vital information of the processed object and suppress some irrelevant details. We extract the attention map to evaluate the degree of underexposure and guide the enhancement in a region-adaptive manner. By doing this, more attention will be paid to the underexposed areas during the enhancement process, which can enhance the underexposed areas and avoid over-enhancing the normal exposure areas. Besides, instead of just using the estimated low illumination for prediction, we use the reconstructed reflectance and low illumination for jointly predicting to illumination layer of the image. This joint prediction utilizes an attention mechanism to make illumination adjustment achieve better results. We conduct extensive experiments on the LOL dataset. Experimental results show that CA&R Net can successfully handle noise, color distortion, and multiple types of degradation with the power of attention information. Furthermore, both SSIM and PSNR outperform Retinex-Net and KinD. Compared with Retinex-Net, CA&R Net dramatically improves the overall performance, with PSNR by 5.38 and SSIM by 26.86%. Compared to KinD, CA&R Net improves PSNR by 1.95 and SSIM by 5.25%.

The main contributions of this paper are summarized as follows:

- We propose the CA&R Net that combines attention mechanism and Retinex model to enhance low-light images. With the power of attention information, the CA&R Net can successfully handle noise, color distortion, and multiple types of degradations.
- We develop an attention map to guide the reflectance restoration in a region-adaptive manner so that it can pay more attention to underexposed areas during the enhancement process and avoid over-enhancing the normal exposure areas.
- Instead of just using the estimated low illumination for prediction, we use the reconstructed reflectance and low illumination to jointly predict the image's illumination layer. This joint prediction utilizes an attention mechanism to make illumination adjustment achieve better results.
- Extensive experiments have been conducted to evaluate our method, and the superiority of our approach has been proved qualitatively and quantitatively.

2. Related work

Traditional Methods. The earliest low-light image enhancement algorithm is to adjust the light distribution of low-light images uniformly. Histogram equalization (HE) achieves illumination enhancement by expanding the dynamic range of images [11] so that the details hidden in the dark area are redisplayed. Later, some other optimization algorithms were proposed, such as adaptive HE algorithm [12], average intensity retention [13], black and white stretching [14], and a novel LHE algorithm [15]. These methods can improve the visibility of images. Still, they mainly improve the contrast of images without considering the effect of illumination, which leads to the over and under illumination in enhancement results. Some scholars also use frequency-domain methods to enhance images, such as the homomorphic filtering algorithm based on spatial filters [16], the two-channel high-frequency color image enhancement method based on HSV color space [17]. These methods can effectively highlight the details of the image by enhancing the transformation parameters. The disadvantage is that these methods amplify the noise hidden in images and require a lot of calculations. The selection of transformation parameters usually requires manual intervention and cannot be automatically selected. Li et al. proposed a method based on the degradation model to enhance low-light images [18]. Li et al. used image fusion methods to synthesize high-quality images [19,20]. The enhancement effects of these fusion methods are relatively sound. However, they rely on multiple different images in the same scene. When faced with scenes that require real-time monitoring, it is not easy to achieve enhancement through image fusion in a short time. There are also some researchers keen on defogging models. Dong et al. proposed a defogging model to enhance low-light images [21]. The defogging algorithm is applied to the inverted low-light vision to enhance the visibility of images. This type of algorithm requires low computational complexity and can achieve good performance. However, the defogging algorithm lacks a physical explanation of the basic model and requires some additional noise reduction processing to eliminate noise. Therefore, the direct application of defogging algorithms to enhance low-light images still has its shortcomings.

Retinex-Based Methods. Another method for low-light image enhancement is based on the Retinex theory [22]. This theory first decomposes the observed image into illumination and reflectance and then enhances these two components separately. In order to suppress the noise in the reflectance and make the enhanced illumination more natural, it is necessary to introduce various prior conditions into the model to guide the enhancement of these two components, such as weighted variation [4], structure perception prior [23]. At the same time, in order to achieve better low light enhancement results, a variety of Retinex variants are proposed to balance the layer separation and manipulation, such as single-scale Retinex (SSR) [24], multi-scale Retinex (MSR) [25], and multi-scale Retinex with color restoration (MSRCR) [26]. Elad et al. proposed a non-iterative Retinex algorithm that can suppress dark area noise and process image edge parts [27]. Fu et al. proposed a weighted variational model to estimate the illumination and reflectance of the image [4]. This model can accurately estimate reflectivity while suppressing noise. Wang et al. used local nonlinear transformation to enhance the separated illumination, making the image brighter and more natural [28]. Xiao et al. introduced the enhancement adjustment factor to adjust the enhancement degree of different exposure areas to avoid problems such as color distortion and amplified noise [29]. In order to solve the problems of the halo effect, loss of detail, and color distortion encountered by MSRCR when enhancing color, Zhao et al. proposed a Markov random field model [30]. These methods achieve not only good results in enhancing the brightness and contrast of the image but also have obvious advantages in processing color images. However, Retinex-based algorithms are built under hand-crafted filters. Therefore, they are insufficient in processing the complex signal characteristics of various images.

Learning-Based Methods. In recent years, deep learning has been widely used in underlying image processing and has achieved great success. While bringing significant changes to the underlying image processing tasks, it has also brought significant improvements in the performance of low-light image enhancement. Dai et al. proposed a novel enhancer for low light image enhancement, which can flexibly adjust the brightness [31]. Zhang et al. proposed a learning-based decomposition-enhancement method to restore low-light images. [32]. Lore et al. created a deep autoencoder to enhance low-light images to achieve adaptive denoising and brightness adjustment [33]. Ren et al. introduced joint denoising and low-light enhancement strategy to eliminate inherent noise in the process of enhancing low-light images [34]. Lv et al. proposed a novel network consisting of a feature extraction module, an enhancement module, and a fusion module [35]. Cai et al. used CNN to achieve single image contrast enhancement [36]. Wang et al. proposed a GLobal illumination-Aware and Detail-preserving Network(GLAD) [37]. Chen et al. introduced the short-exposure low-light image dataset SID in Raw format and proposed a method based on FCN to process these images [38]. This method improves the traditional process of processing low-light images and can successfully suppress noise and correctly implement the color conversion, but there are certain limitations. For example, HDR tone mapping cannot be resolved, and magnification cannot be learned in the input. Wei et al. constructed a paired dataset containing low/normal light images and proposed a deep network called Retinex-Net [39]. The network combines the Retinex model with deep learning for the first time, breaking the limitation that the traditional Retinex model requires manual constraints. However, the subsequent enhancement network only designs a Relight-Net to re-estimate the illumination. It does not consider the influence of the reflectance component and illumination components on the image, respectively. Therefore, the desired enhancement cannot be achieved. Zhang et al. built a simple yet effective network for Kindling the Darkness(KinD) [40]. They designed an excellent image decomposition network and created a superb loss function that achieves desired results. However, in the subsequent enhancement network, the reflectivity recovery network only uses the U-Net structure, and the design of the illumination adjustment network is also straightforward. No more information is introduced to recover these two components, resulting in the enhanced image still existing the problem of low brightness. A multi-branch convolutional neural network based on two attention maps is proposed. The two attention maps extracted by this method are used to guide image denoising and enhancement tasks, respectively [41]. Wang et al. proposed a network that estimates the image-to-illumination mapping to enhance low-light images, which breaks the traditional image-to-image mapping methods [42]. But this method cannot handle texture details and clean noise well. Xu et al. proposed a frequency-based decomposition and enhancement model. This model first uses the attention context coding module to restore the low-frequency layer image content and denoise and then uses the converted image to enhance high-frequency details [43]. Fan et al. introduced semantic information into low-light image enhancement and used it to guide subsequent enhancement [44]. The introduction of semantic information can provide more details for enhancement, leading to better image restoration. However, in this method, the semantics of the images are divided into three parts: sky, ground, and background, limiting the dataset's structure. In addition, datasets with semantic information in the low-light image domain are difficult to obtain, so this method can only be applied to specific datasets. Yang et al. designed an end-to-end signal prior-guided decomposition and data-driven mapping network, which utilizes the constraints of decomposition to enhance a single low-light image [45]. Jiang et al. applied Generative Adversarial Networks to enhance low-light images. This model adopts unpaired data to train the network, which solves the problem of difficulty in obtaining paired datasets [46].

3. Methodology

3.1. Retinex model for low-light image enhancement

The classic Retinex theory models human color perception. It assumes that the observed image can be decomposed into two components, reflectance and illumination:

$$S = I \circ R, \tag{1}$$

where S denotes the input image, R denotes reflectance, I denotes illumination and $\circ$ denotes element-wise multiplication. Reflectance describes the inherent property of the captured object, which is consistent under any brightness conditions. Illumination means various brightness on the object.

For the low-light image enhancement based on the Retinex model, the low-light image S_{low} is first decomposed into illumination I_{low} and reflection R_{low}:

$$[I_{low}, R_{low}] = f_{decom}\left(S_{low}\right), \tag{2}$$

$$\left[I_{high}, R_{high}\right] = f_{decom}\left(S_{high}\right), \tag{3}$$

where $f_{decom}(\cdot)$ represents the learning process of decomposing the input image into reflectance and illumination. Then, according to the decomposed reflectance R_{low} and illumination I_{low}, the enhanced reflectance $\hat{R}_{high}$ and illumination $\hat{I}_{high}$ are inferred:

$$\hat{R}_{high} = f_{restore}\left(R_{low}\right), \tag{4}$$

$$\hat{I}_{high} = f_{adjust}\left(I_{low}\right), \tag{5}$$

The final enhancement result is reconstructed by the following formula:

$$\hat{S}_{high} = \hat{R}_{high} \circ \hat{I}_{high}. \tag{6}$$

where $\hat{}$ represents the predicted result. In our method, $f_{decom}(\cdot)$ is not a manual constraint but a learnable process, which is jointly modeled with $f_{restore}(\cdot)$ and $f_{adjust}(\cdot)$ to achieve enhancement in a data-driven manner. In addition, to allow $f_{restore}(\cdot)$ to pay more attention to the underexposed areas, we extract A_{map} to evaluate the degree of underexposure. Then, A_{map} together with R_{low} is input to the Restore-Net to guide the enhancement in a region-adaptive manner. Making it pay more attention to the underexposed areas during the enhancement process can enhance the underexposed areas and avoid over-enhancing the normally exposed areas. Therefore, in our scheme, we modify the formula (4) as:

$$\hat{R}_{high} = f_{restore}\left(R_{low}, A_{map}\right), \tag{7}$$

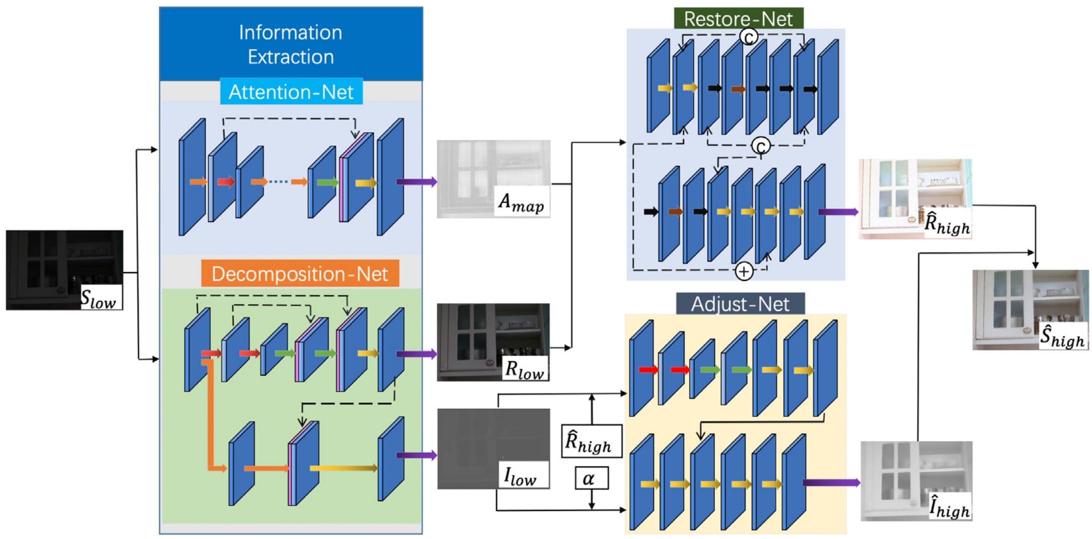

Fig. 2. The architecture of our CA&R Net consists of three components: information extraction (including attention network and image decomposition network), reflectance restoration, and illumination adjustment. Firstly, we use the information extraction network to acquire the reflectance (R), illumination (I), and attention map (A_{map}) from the input image (S). Secondly, use Restore-Net to restore the reflectance component under the guidance of A_{map}. Finally, use the reconstructed reflectance and low illumination for jointly predicting to illumination layer of the image. The final enhancement result is reconstructed by $\hat{S}_{high} = \hat{R}_{high} \circ \hat{I}_{high}$.

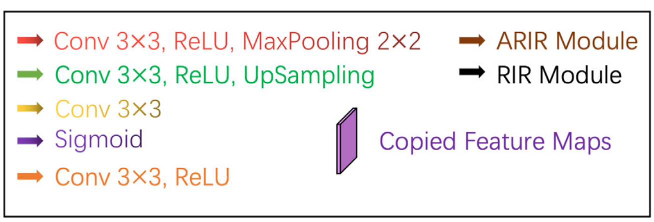

Fig. 3. The description of network architecture.

modify the formula (5) as:

$$\hat{I}_{high} = f_{adjust}\left(I_{low}, \hat{R}_{high}\right). \tag{8}$$

formula (5) utilizes the reconstructed reflectance $\hat{R}_{high}$ and low illumination I_{low} for jointly predicting to illumination layer of the image, instead of just utilizing the estimated low illumination to predict. This joint prediction adopts an attention mechanism so that more attention can be paid to underexpose areas and learn the details of the image better. And thus, the illumination adjustment can achieve better results.

3.2. Overall network architecture

In this paper, we design the CA&R Net that combines attention mechanism and Retinex model to enhance low-light images. The whole network architecture is shown in Fig. 2. It consists of three stages: information extraction (including attention network and image decomposition network), reflectance restoration, and illumination adjustment. For these three stages, three sub-networks are built to model $f_{decom}(\cdot)$, $f_{restore}(\cdot)$, $f_{adjust}(\cdot)$, respectively.

At the decomposition stage, the network utilizes Decomposition-Net to decompose the paired low/normal light images into reflectance R_{low}/R_{high} and illumination I_{low}/I_{high} (denoted as $f_{decom}(\cdot)$). In this stage, under the constraints of low/normal light images sharing the same reflectance and illumination smoothness, autonomous learning is carried out in a data-driven manner. At the restoration stage, the noise in the reflectance is suppressed by a Restore-Net. Under the guidance of A_{map}, the reflectance restore is realized in a regional adaptive manner (denoted as $f_{restore}(\cdot)$). At the adjustment stage, we use the reconstructed reflectance and low illumination for jointly predicting to illumination layer of the image. This joint prediction utilizes an attention mechanism, which can make illumination adjustment achieve better results (denoted as $f_{adjust}(\cdot)$). Finally, we use the reconstructed reflectance and adjusted illumination to reconstruct the image according to element-wise multiplication to obtain the final enhancement results. Table 1 summarizes the functions of each sub-networks in the network.

Next, we will introduce in detail the three sub-networks of information extraction (Section 3.3), reflectivity restoration (Section 3.4), and illumination adjustment (Section 3.5).

3.3. Information extraction net

Image decomposition network. Recovering two components from images is a challenging task. The ground-truth images require professionals with excellent photography skills and massive processing. Normally, such image data is difficult to obtain, so it is crucial to design a proper loss function to constrain network training without ground-truth guidance. In our network, we use low/normal light images [S_{low}, S_{high}] with different exposure levels for training. According to the constraint that low/normal light images share the same reflectance and the smooth and consistent nature of the illumination map, a loss function for the constraint is designed. Specifically, it includes four parts: reflectance similarity loss, illumination smoothness loss, mutual consistency loss, and reconstruction loss. Zhang et al. created an excellent loss function for image decomposition that can achieve an ideal result [40]. We here use this loss to constrain the training of the decomposition network.

Loss of reflectance similarity:

$$L_{rs} = \left\| R_{low} - R_{high} \right\|_2, \tag{9}$$

We simply utilize L_{rs} to regularize the reflectance similarity, where $\|\cdot\|_2$ denotes the ℓ^2 norm (MSE).

Loss of illumination smoothness:

$$L_{is} = \left\| \frac{\nabla I_{low}}{\max\left(\left|\nabla S_{low}\right|, \varepsilon\right)} \right\|_1 + \left\| \frac{\nabla I_{high}}{\max\left(\left|\nabla S_{high}\right|, \varepsilon\right)} \right\|_1, \tag{10}$$

Table 1
The functionality of each sub-network in our network architecture.

Sub-network	Decom-net	Restore-net	Adjust-net
Input	S_{high} (S_{low})	R_{low}, A_{map}	I_{low}, $\hat{R}_{high}$
Output	R_{high}, I_{high} (R_{low}, I_{low})	$\hat{R}_{high}$	$\hat{I}_{high}$
Process	$f_{decom}(\cdot)$	$f_{restore}(\cdot)$	$f_{adjust}(\cdot)$
Functionality	Decomposition	Reflectance restoration	Illumination adjustment

The illumination smoothness is constrained by L_{is}, where ∇ stands for the first-order derivative operator, $\|\cdot\|_1$ denotes the ℓ^1 norm, and $|\cdot|$ denotes the absolute value operator. In addition, we added a small positive constant ε for avoiding zero denominators (0.01 in this work) (see Fig. 3).

Loss of mutual consistency:

$$L_{mc} = \|X \circ \exp(-c \cdot X)\|_1, \tag{11}$$

$$X = |\nabla I_{low}| + \left|\nabla I_{high}\right|, \tag{12}$$

Mutual consistency is represented by L_{mc}, where c is the parameter (set to 10 in this paper).

Loss of reconstruction:

$$L_{rec} = \|S_{low} - R_{low} \circ I_{low}\|_1 + \left\|S_{high} - R_{high} \circ I_{high}\right\|_1, \tag{13}$$

Decomposition-Net decomposes the input image into illumination and reflectance. When reconstructing these two components, the input image should also be obtained. We use L_{rec} to constrain it.

Therefore, the total loss function of this stage is expressed as:

$$L_{decom} = L_{rec} + \lambda_{rs}L_{rs} + \lambda_{is}L_{is} + \lambda_{mc}L_{mc}. \tag{14}$$

where λ_{rs}, λ_{is}, and λ_{mc} are the weight coefficients of reflectance similarity loss, illumination smoothness loss and mutual consistency loss.

Attention-Net. We use the Attention-Net to extract attention map A_{map} of images and guide the reflectance recovery in Restore-Net. We have observed that there is more noise in the underexposed areas of images than that in the normally exposed areas. To effectively enhance the contrast of the image and suppress noise, the key is to solve the problem in an adaptive way of area perception. Therefore, we designed an A_{map} to evaluate the degree of underexposure and guide the enhancement in a region-adaptive manner, to enhance the underexposed areas and avoid over-enhancing the normally exposed areas. In our method, we directly use U-Net to achieve this goal. The purpose is to provide guidance for the subsequent enhancement of underexposed areas and avoid over-enhancing normally exposed areas. The output is an A_{map}, which is used to indicate the degree of underexposed areas. It can be expressed by the following formula:

$$A = \frac{\left|\max_c\left(S_{high}\right) - \max_c\left(S_{low}\right)\right|}{\max_c\left(S_{high}\right)}, \tag{15}$$

where $\max_c(\cdot)$ denotes returning the maximum value of the three color channels, S_{high} denotes high-light image, S_{low} denotes low-light images. In order to obtain A_{map}, we use this formula to constrain it:

$$L_{attention} = \|f_a(S) - A\|_2. \tag{16}$$

where S denotes low-light images, $f_a(\cdot)$ and A denote predicted and expected results, respectively.

3.4. Reflectance restoration net

As the solid noise and color deviation in the reflectance R_{low} from Decomposition-Net, we use the Restore-Net to restore R_{low} under the guidance of attention information.

In the reflectance recovery phase, inspired by Fan et al. [44], we introduced the Residual in Residual (RIR) module. The RIR module utilizes deep residual learning to provide powerful pixel adjustment capabilities for the reflectance recovery network, which can effectively suppress noise and perform color correction. To introduce attention information, we designed an Attention Residual in Residual (ARIR) module. This module first adopts two two-layer convolutions to extract features F_1 and F_2 from A_{map}. Then, several convolution layers are used to extract the feature F_3 from the reflectance R_{low} through the RIR module. Finally, the three features of F_1, F_2, and F_3 are merged by multiplying F_3 by F_2 and then adding F_1. After the fusion, the output of the ARIR module is obtained through convolution operation. At this stage, we first use the convolution operation to extract features. Then the features are fused through a series of RIR and ARIR modules. Finally, the output result $\hat{R}_{high}$ is obtained through convolution and the Sigmoid layer.

Restore-Net utilizes the reflectance of normal-light images as a constraint for training. Specifically, the final loss at this stage includes three parts: Mean Squared Error (MSE), Structure SIMilarity index ($SSIM$), and Gradient loss ($Grad$). The loss function of the reflectance recovery stage can be expressed as:

$$L_{restore} = MSE\left(\hat{R}_{high}, R_{high}\right) + \lambda_1 SSIM\left(\hat{R}_{high}, R_{high}\right)$$

$$+\lambda_2 Grad\left(\hat{R}_{high}, R_{high}\right). \tag{17}$$

where λ_i is the weight coefficient, $\hat{R}_{high}$ is denoted as the reconstruction result by Restore-Net.

$$Grad\left(\hat{R}_{high}, R_{high}\right) = \sqrt{\left|\nabla\hat{R}_{high} - \nabla R_{high}\right|}. \tag{18}$$

where ∇ stands for the first-order derivative operator, $\sqrt{\cdot}$ means arithmetic square root, and $|\cdot|$ means the absolute value operator, which ensures that arithmetic square root operations can be performed normally.

3.5. Illumination adjustment net

The illumination I_{low} directly decomposed by Decomposition-Net is usually in a low-visibility. In order to adjust I_{low}, we designed an Adjust-Net and used the reconstructed reflectance $\hat{R}_{high}$ and low illumination I_{low} for jointly predicting to illumination layer of the image.

Firstly, we utilize U-Net to extract features from $\hat{R}_{high}$. In this way, the information of A_{map} will be introduced into the Adjust-Net. Secondly, a two-layer convolution network is adopted to extract features from enhancement ratio α and I_{low}. This feature is fused with the features extracted by U-Net and then obtained the adjusted illumination through a convolutional operation. The result $\hat{I}_{high}$ of the illumination adjustment network is limited to [0,1] through the Sigmoid layer. Finally, $\hat{R}_{high}$ and $\hat{I}_{high}$ are rebuilt according to the element-wise multiplication. The CA&R Net obtains the final prediction $\hat{S}_{high} = \widehat{R}_{high} \circ \hat{I}_{high}$.

Since there is no actual light level of the image, we need to design a light condition conversion mechanism to meet the needs

Table 2
Quantitative measurement results on LOL dataset.

Metrics	BIMEF [47]	CRM [48]	Dong [21]	LIME [23]	MF [49]	RRM [50]
PSNR	13.86	17.20	16.72	16.76	18.79	13.88
SSIM	0.58	0.64	0.58	0.56	0.64	0.66
Metrics	MSR [26]	CLAHE [51]	DHECI [52]	AGLLIE [41]	DeepUPE [42]	EFF [53]
PSNR	13.17	13.13	14.64	19.48	13.27	17.85
SSIM	0.48	0.37	0.45	0.81	0.45	0.65
Metrics	SICE [36]	GLAD [37]	JED [34]	LLNet [33]	Retinex-Net-TIP [45]	KinD [40]
PSNR	19.40	19.72	17.33	17.56	20.06	20.87
SSIM	0.69	0.70	0.67	0.55	0.82	0.80
Metrics	NPE [54]	Retinex-Net [39]	BPDHE [13]	MBLLEN [35]	SRIE [4]	**ours**
PSNR	16.97	16.77	13.84	18.56	17.34	**22.15**
SSIM	0.59	0.56	0.43	0.75	0.69	**0.83**

of different scenes. In our dataset, we adopt paired low/normal light images. That is, we have paired illumination images. Even if we do not know the exact relationship between I_{low} and I_{high}, we can use $\alpha = I_{high}/I_{low}$ to calculate their intensity ratio. This ratio parameter can be used to guide the illumination adjustment process. In the testing phase, α can be given by users. We utilize the illumination I_{high} of the normal-light image to guide the adjustment network, introducing more context information for model training. The loss functions at this stage can be expressed as:

$$L_{adjust} = MSE\left(\hat{I}_{high}, I_{high}\right) + \lambda_1 SSIM\left(\hat{I}_{high}, I_{high}\right)$$
$$+\lambda_2 Grad\left(\hat{I}_{high}, I_{high}\right). \qquad (19)$$

where λ_i is the weight coefficient.

4. Experimental validation

4.1. Implementation details

In this paper, our network is trained using the public dataset LOL, which includes 500 low/normal light image pairs. Among them, 485 image pairs were used for training, 15 image pairs were used for testing, and the synthetic dataset proposed by Fan et al. [44] includes 2458 low/normal light image pairs, of which 2118 pairs are used for training the other 340 pairs for evaluation. Our network consists of three parts: information extraction (including attention and decomposition network), reflectance recovery, and illumination adjustment. The training process is also split into four stages, corresponding to Attention-Net, Decomposition-Net, Restore-Net, and Adjust-Net. During the training process, the batch size is set to be 10 and the patch size to be 48. The initial learning rate is set to 0.0001. In the Restore-Net, when the epoch reaches 300, 500, 1500, the learning rate is updated to 1/2, 1/4, 1/8 of the initial values. We use stochastic gradient descent (SGD) technology to optimize our network. The entire network is trained on Tesla V100 GPU using the TensorFlow framework.

4.2. Quantitative comparisons

In this section, we evaluate the performance of our method through quantitative experiments. Since assessing the quality of enhanced images is not a simple task, in this paper, we adopt both Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) that are commonly used in the field of image enhancement to evaluate the results. PSNR is the ratio between the maximum possible power of the normal-light image and the power of the enhanced image. This indicator can be used to measure the fidelity between the normal-light image and the enhanced image. SSIM judges the similarity of two images from three aspects: structure, brightness, and contrast. The higher the values of these two indicators are, the better the image quality is.

In order to evaluate the superiority of the network from a quantitative perspective, we conducted massive experiments on the public dataset LOL. The state-of-the-art methods of BIMEF [47], CRM [48], Dong [21], LIME [23], MF [49], RRM [50], NPE [54], Retinex-Net [39], MSR [26], CLAHE [51], DHECI [52], AGLLIE [41], DeepUPE [42], EFF [53], BPDHE [13], MBLLEN [35], SICE [36], GLAD [37], JED [34], LLNet [33], Retinex-Net-TIP [45], KinD [40], SRIE [4] are involved as the competitors. It can be seen from Table 2 that our method is superior to all the most advanced methods in these two indicators.

Among them, Retinex-Net [39], Retinex-Net-TIP [45], and KinD [40] participating in the comparison are all based on the Retinex model. Without considering the effects of the two components on the image, Retinex-Net [39] only designed an enhancement network to re-estimate the image's brightness in the subsequent enhancement network and thus cannot achieve ideal enhancement. The image decomposition-net of KinD [40] is well-designed, but in the subsequent enhancement network, the reflectance recovery network only uses the U-Net structure. The design of the illumination adjustment network is also relatively simple. Our CA&R Net extracts the attention map as a guide for the reflex recovery stage. The attention map can evaluate the degree of underexposure, and more attention can be paid to the underexposed area in the reflection recovery stage to enhance this component well. In addition, since the noise in the dark area of a low-light image is more serious than the normal bright area, if the entire part is directly enhanced, the noise hidden in the dark area will be amplified. Therefore, using the attention map as a guide can effectively suppress the noise during the enhancement process. In the illumination adjustment stage, in addition to the enhancement rate, the restored reflectance is also used as input to introduce attention information so that the illumination adjustment can focus on more context information, so as achieving better results. The quantitative results of the experiment also proved the effectiveness and superiority of CA&R Net.

4.3. Qualitative evaluation

Figs. 4–6 show the visual comparison between our method and other methods on the LOL dataset. Experiment results show that the four ways of BIMEF [47], SRIE [4], CRM [48], and Dong [21] are insufficient in enhancing image brightness. Even after improving the image, the problem of dark light still exists. Although the four methods GLAD [37], LIME [23], MF [49], and NPE [54] have improved the brightness of the image to a certain extent, the enhanced image has apparent noise and severe color distortion. The improved image of Retinex-Net [39] is blurry. The enhancement effect of KinD [40] is relatively good. The brightness of images

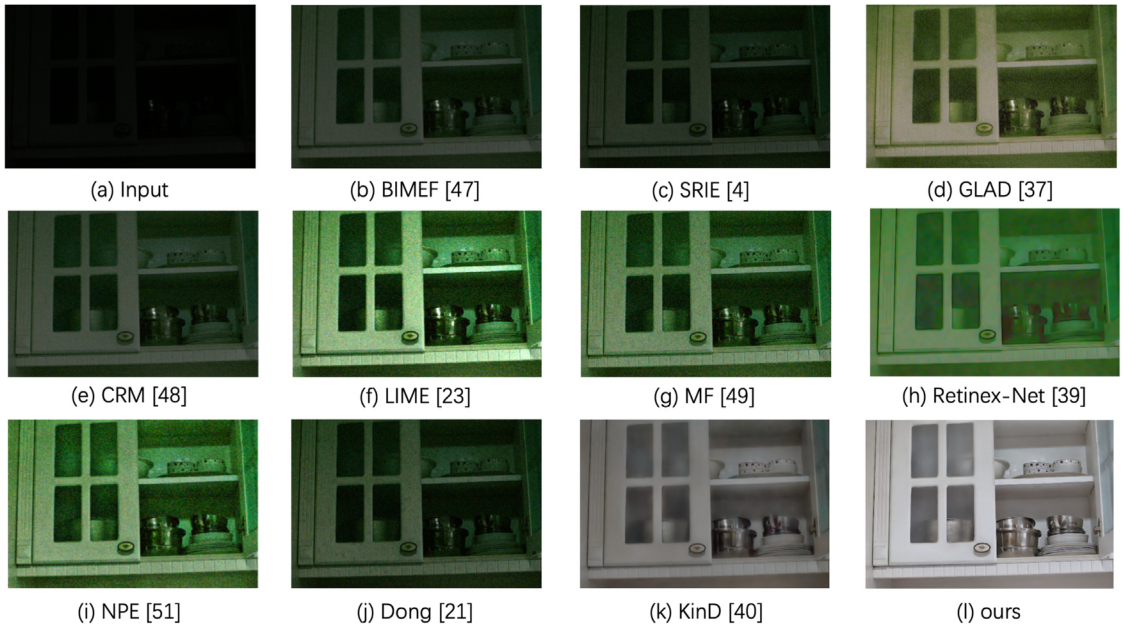

Fig. 4. Visual comparison with other low-light image enhancement methods.

Fig. 5. Visual comparison with other low-light image enhancement methods.

Table 3
Quantitative measurement results on synthetic datasets.

Metrics	BIMEF [47]	LIME [23]	EnlightenGAN [46]	MF [49]	JED [34]	DeepUPE [42]
PSNR	22.642	12.304	16.953	20.115	21.604	22.503
SSIM	0.762	0.508	0.731	0.651	0.798	0.710
Metrics	SICE [36]	MBLLEN [35]	Retinex-Net [39]	KinD [40]	**ours**	
PSNR	16.227	17.228	23.166	21.630	**23.806**	
SSIM	0.783	0.744	0.676	0.903	**0.923**	

is improved, and the noise is also handled well. Moreover, it can enhance images with considerable fidelity, but the improved image still has low brightness (see Table 3).

Figs. 7 and 8 show the visual comparison of our method with Retinex-Net and KinD on synthetic datasets. The experimental results show that the enhanced images of Retinex-Net have intense noise and color distortion. Compared with Retinex-Net, KinD has significantly less overall noise and distortion, but the above problems still exist in some areas. For example, (c1), (c2), (c3), the recovery of some areas is poor. The noise and distortion

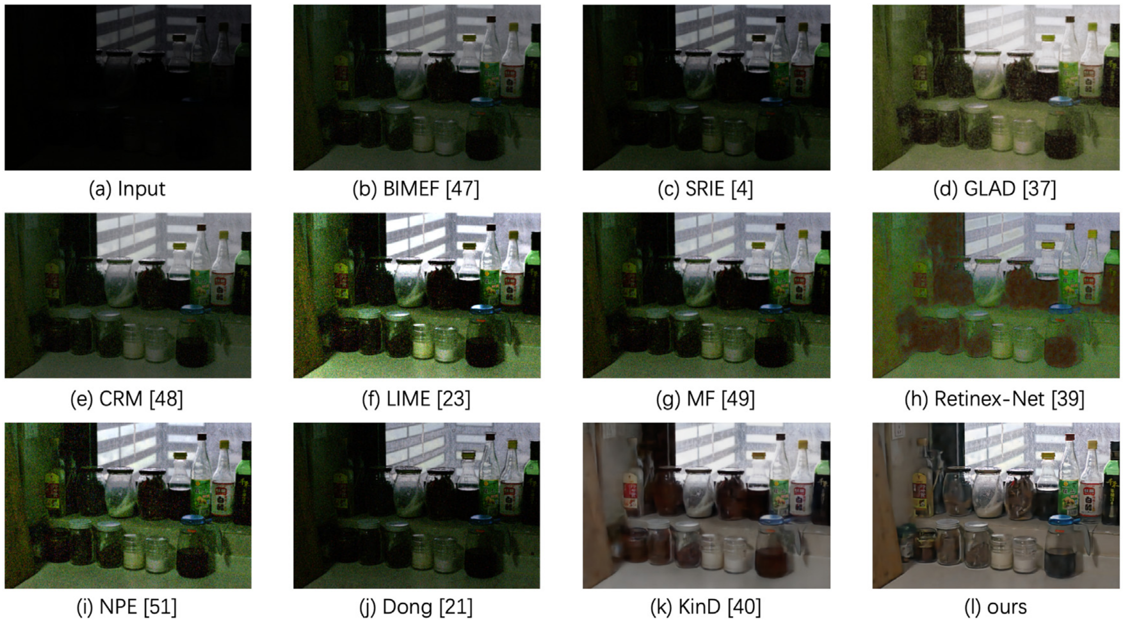

Fig. 6. Visual comparison with other low-light image enhancement methods.

Fig. 7. Visual comparison with other low-light image enhancement methods.

of (g1) and (g2) are more negligible, but the enhancement result still has the problem of low brightness. The color distortion of the (g3) red area is severe, and the white site is over-enhanced. Overall, although Retinex-Net improves the image's brightness, the restoration result is poor. This is because it only designs a Relight-Net to re-estimate the illumination of the image and does not separately consider the effects of the reflection component and the illumination component on the image. Reflectance is an inherent property of an object, including features such as the color of an image. The reflectivity obtained by the decomposition of low-light images is very rough. If it is not processed, the enhancement results will inevitably have noise and color distortion problems. KinD's reflectivity recovery net is only implemented with U-Net, and no more information is provided for it as a guide. This simple uniform enhancement can lead to over-enhancement and noise in some areas. Because for low-light images, the noise in dark places is more severe than in normally exposed areas, enhancements that do not distinguish the entire region are bound to amplify the noise hidden in dark areas. KinD's light adjustment network also only uses a few simple convolutions without providing it with more contextual information. Therefore, when KinD processes the overall dark image, there will be a problem that the illumination of the enhanced result is still insufficient. The restoration result of our network is more natural, which solves

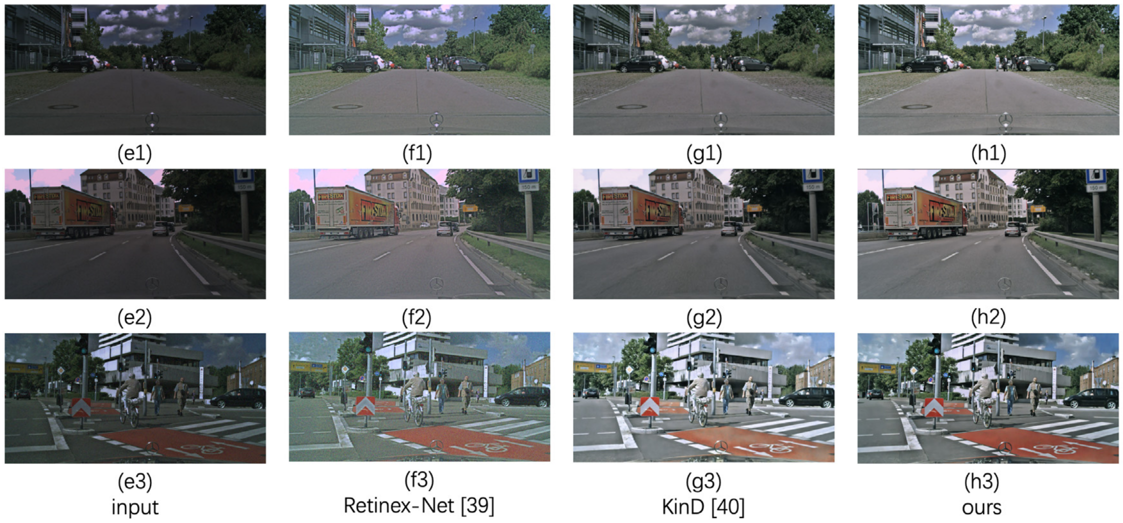

Fig. 8. Visual comparison with other low-light image enhancement methods.

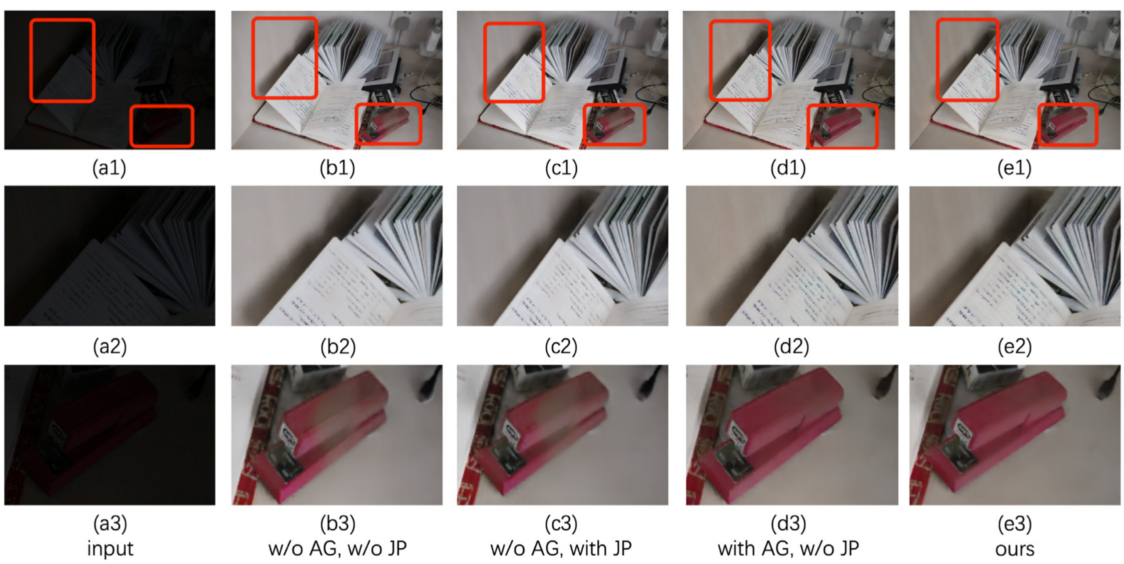

Fig. 9. Ablation studies of the proposed method.

Table 4
The setting of our ablation study.

Methods	Attention guide	Joint prediction
w/o AG, w/o JP	×	×
w/o AG, with JP	×	✓
with AG, w/o JP	✓	×
ours	✓	✓

Table 5
Ablation study of our proposed network.

Methods	PSNR	SSIM
Retinex-Net	16.7740	0.5594
w/o AG, w/o JP	20.2009	0.7755
w/o AG, with JP	21.0028	0.7917
with AG, w/o JP	21.1708	0.8170
ours	22.1506	0.8280

the problem of harsh noise and color distortion in Retinex-Net and the problem of poor restoration of some regions of KinD. We design the attention map to evaluate the degree of underexposure and use it as a guide for subsequent enhancements to enhance the underexposed areas better and avoid over-enhancement of the customarily exposed areas, such as (h3). In addition, the joint prediction can provide more information for the lighting adjustment so that the lighting adjustment can achieve better results, such as (h1) and (h2).

4.4. Ablation study

In this section, we conduct ablation studies to evaluate the effectiveness of each component in CA&R Net. The setting can be found in Table 4. Quantitative ablation study results of PSNR and SSIM can be found in Table 5. The CA&R Net is based on Retinex-Net [39], and KinD [40] is the baseline, w/o AG and w/o JP is the result of retraining KinD under our conditions. Quantitative results show that compared with the Retinex-Net, CA&R Net largely improves the overall performance, with PSNR improving 5.38 and SSIM improving 26.86%. Based on the baseline, the injection of joint prediction (w/o AG, with JP) improves the PSNR by 0.80 and SSIM by 1.62%. The injection of attention information (with AG, w/o JP) improves the PSNR by 0.97 and SSIM by 4.15%. The injection of attention information and joint prediction (ours) improves the PSNR by 1.95 and SSIM by 5.25%.

Qualitative results are shown in Fig. 9. For (a2), there is obvious noise in the shadow area of the enhanced photo (b2). The injection of joint prediction (w/o AG, with JP) eliminates the noise in the shadow area, but the background has visible color distortion (c2). This is because joint prediction can provide more contextual information for Adjust-Net. The injection of attention information (with AG, w/o JP) solves the problem of background color distortion, but the noise in the shadow part still exists (d2). This is because attention information can distinguish areas with different exposures and successfully handle these areas. The enhancement results of CA&R Net (ours) are considerable, which successfully solves the existing noise and color distortion problems (e2). For (a3), the enhanced result suffers from severe artifacts and degradation (b3). Even the injection of joint prediction cannot deal with the problems of artifacts and degradation in severely underexposed areas (c3). However, after the injection of attention information, these defects are eliminated (d3). It proves once again that the attention information can distinguish the areas of different exposures, and pay more attention to these areas during the enhancement process, to enhance results better.

5. Conclusion

In this paper, we designed a CA&R Net to solve the problem of low-light image enhancement. We have observed more noise in the underexposed areas of images than in the normally exposed areas. Attention mechanism can be used to emphasize the vital information of the processed object and suppress some irrelevant information. Based on these observations, we propose the CA&R Net that combines attention mechanism and Retinex model to enhance low-light images. The CA&R Net includes three stages of information extraction (including attention network and image decomposition network), reflectance recovery, and illumination adjustment. Firstly, we use Attention-Net to obtain the attention map of the image and then decouple the original space into two smaller subspaces through the Decomposition-Net. Secondly, the attention map is adopted to guide the reflectance recovery in a region-adaptive manner. Finally, the recovered reflectance and low illumination are used to predict the illumination layer of the image jointly. Extensive experiments show that our CA&R Net is superior to other advanced methods in low-light image enhancement. It can successfully handle noise, color distortion, and multiple types of degradations.

CRediT authorship contribution statement

Yong Wang: Validation, Formal analysis, Resources, Writing – review & editing, Supervision, Funding acquisition. **Jin Chen:** Conceptualization, Methodology, Software, Validation, Data curation, Writing – original draft. **Yujuan Han:** Validation, Formal analysis, Investigation, Data curation, Writing – review & editing, Visualization. **Duoqian Miao:** Formal analysis, Investigation, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

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.

References

[1] Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, p. 2117–25.
[2] Chu Q, Ouyang W, Li H, Wang X, Liu B, Yu N. Online multi-object tracking using CNN-based single object tracker with spatial-temporal attention mechanism. In: Proceedings of the IEEE international conference on computer vision. 2017, p. 4836–45.
[3] Yuan L, Sun J. Automatic exposure correction of consumer photographs. In: European conference on computer vision. Springer; 2012, p. 771–85.
[4] 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. 2016, p. 2782–90.
[5] Tian Q-C, Cohen LD. A variational-based fusion model for non-uniform illumination image enhancement via contrast optimization and color correction. Signal Process 2018;153:210–20.
[6] Cai B, Xu X, Guo K, Jia K, Hu B, Tao D. A joint intrinsic-extrinsic prior model for retinex. In: Proceedings of the IEEE international conference on computer vision. 2017, p. 4000–9.
[7] Ignatov A, Kobyshev N, Timofte R, Vanhoey K, Van Gool L. Dslr-quality photos on mobile devices with deep convolutional networks. In: Proceedings of the IEEE international conference on computer vision. 2017, p. 3277–85.
[8] 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. 2018, p. 6306–14.
[9] Hu Y, He H, Xu C, Wang B, Lin S. Exposure: A white-box photo post-processing framework. ACM Trans Graph 2018;37(2):1–17.
[10] Park J, Lee J-Y, Yoo D, Kweon IS. Distort-and-recover: Color enhancement using deep reinforcement learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, p. 5928–36.
[11] Pizer SM, Johnston RE, Ericksen JP, Yankaskas BC, Muller KE. Contrast-limited adaptive histogram equalization: speed and effectiveness. In: [1990] proceedings of the first conference on visualization in biomedical computing. IEEE Computer Society; 1990, p. 337–8.
[12] Kim TK, Paik JK, Kang BS. Contrast enhancement system using spatially adaptive histogram equalization with temporal filtering. IEEE Trans Consum Electron 1998;44(1):82–7.
[13] Ibrahim H, Kong NSP. Brightness preserving dynamic histogram equalization for image contrast enhancement. IEEE Trans Consum Electron 2007;53(4):1752–8.
[14] Arici T, Dikbas S, Altunbasak Y. A histogram modification framework and its application for image contrast enhancement. IEEE Trans Image Process 2009;18(9):1921–35.
[15] Huang S-C, Yeh C-H. Image contrast enhancement for preserving mean brightness without losing image features. Eng Appl Artif Intell 2013;26(5–6):1487–92.
[16] Xiao J, Song S-p, Ding L-j. Research on the fast algorithm of spatial homomorphic filtering. J Image Grap 2008;13(12):2302–6.
[17] Han L-n, Xiong J, Geng G-h, Zhou M. Using HSV space real-color image enhanced by homomorphic filter in two channels. Comput Eng Appl 2009;45(27):18–20.
[18] Li P, Liang J, Zhang M. A degradation model for simultaneous brightness and sharpness enhancement of low-light image. Signal Process 2021;189:108298.
[19] Li ZG, Zheng JH, Rahardja S. Detail-enhanced exposure fusion. IEEE Trans Image Process 2012;21(11):4672–6.
[20] Yun B-J, Hong H-D, Choi H-H. A contrast enhancement method for HDR image using a modified image formation model. IEICE Trans Inf Syst 2012;95(4):1112–9.
[21] Dong X, Wang G, Pang Y, Li W, Wen J, Meng W, et al. Fast efficient algorithm for enhancement of low lighting video. In: 2011 IEEE international conference on multimedia and expo. IEEE; 2011, p. 1–6.
[22] Land EH. The retinex theory of color vision. Sci Am 1977;237(6):108–29.
[23] Guo X, Li Y, Ling H. LIME: Low-light image enhancement via illumination map estimation. IEEE Trans Image Process 2016;26(2):982–93.
[24] Jobson DJ, Rahman Z-u, Woodell GA. Properties and performance of a center/surround retinex. IEEE Trans Image Process 1997;6(3):451–62.
[25] Rahman Z-u, Jobson DJ, Woodell GA. Multi-scale retinex for color image enhancement. In: Proceedings of 3rd IEEE international conference on image processing, Vol.3. IEEE; 1996, p. 1003–6.
[26] Jobson DJ, Rahman Z-u, Woodell GA. A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE Trans Image Process 1997;6(7):965–76.
[27] Elad M. Retinex by two bilateral filters. In: International conference on scale-space theories in computer vision. Springer; 2005, p. 217–29.

[28] Wang D, Niu X, Dou Y. A piecewise-based contrast enhancement framework for low lighting video. In: Proceedings 2014 IEEE international conference on security, pattern analysis, and cybernetics. IEEE; 2014, p. 235–40.
[29] Xiao J, Shan D, Duan P, Chao-Ping T, Ben-Shun Y. A fast image enhancement algorithm based on fusion of different color spaces. Acta Automat Sinica 2014;40(4):697–705.
[30] Zhao H, Xiao C, Yu J, Bai L. A Retinex algorithm for night color image enhancement by MRF. Opt Precis Eng 2014;22(4):1048–55.
[31] Dai C, Guan Z, Lin M. Single low-light image enhancer using Taylor expansion and fully dynamic convolution. Signal Process 2021;189:108280.
[32] Zhang S, Lam EY. An effective decomposition-enhancement method to restore light field images captured in the dark. Signal Process 2021;189:108279.
[33] Lore KG, Akintayo A, Sarkar S. LLNet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognit 2017;61:650–62.
[34] Ren X, Li M, Cheng W-H, Liu J. Joint enhancement and denoising method via sequential decomposition. In: 2018 IEEE international symposium on circuits and systems. IEEE; 2018, p. 1–5.
[35] Lv F, Lu F, Wu J, Lim C. MBLLEN: Low-light image/video enhancement using CNNs. In: BMVC. 2018, p. 220.
[36] Cai J, Gu S, Zhang L. Learning a deep single image contrast enhancer from multi-exposure images. IEEE Trans Image Process 2018;27(4):2049–62.
[37] Wang W, Wei C, Yang W, Liu J. GLADNet: Low-light enhancement network with global awareness. In: 2018 13th IEEE international conference on automatic face & gesture recognition. IEEE; 2018, p. 751–5.
[38] 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. 2018, p. 3291–300.
[39] Wei C, Wang W, Yang W, Liu J. Deep retinex decomposition for low-light enhancement. In: British machine vision conference 2018. BMVA Press; 2018, p. 155, URL http://bmvc2018.org/contents/papers/0451.pdf.
[40] Zhang Y, Zhang J, Guo X. Kindling the darkness: A practical low-light image enhancer. In: Proceedings of the 27th ACM international conference on multimedia. 2019, p. 1632–40.
[41] 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(7):2175–93.
[42] 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 IEEE/CVF conference on computer vision and pattern recognition. 2019, p. 6849–57.
[43] Xu K, Yang X, Yin B, Lau RW. Learning to restore low-light images via decomposition-and-enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020, p. 2281–90.
[44] Fan M, Wang W, Yang W, Liu J. Integrating semantic segmentation and retinex model for low-light image enhancement. In: Proceedings of the 28th ACM international conference on multimedia. 2020, p. 2317–25.
[45] Yang W, Wang W, Huang H, Wang S, Liu J. Sparse gradient regularized deep retinex network for robust low-light image enhancement. IEEE Trans Image Process 2021;30:2072–86.
[46] Jiang Y, Gong X, Liu D, Cheng Y, Fang C, Shen X, et al. Enlightengan: Deep light enhancement without paired supervision. IEEE Trans Image Process 2021;30:2340–9.
[47] Ying Z, Li G, Gao W. A bio-inspired multi-exposure fusion framework for low-light image enhancement. 2017, arXiv preprint arXiv:1711.00591.
[48] Ying Z, Li G, Ren Y, Wang R, Wang W. A new low-light image enhancement algorithm using camera response model. In: Proceedings of the IEEE international conference on computer vision workshops. 2017, p. 3015–22.
[49] 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.
[50] Li M, Liu J, Yang W, Sun X, Guo Z. Structure-revealing low-light image enhancement via robust retinex model. IEEE Trans Image Process 2018;27(6):2828–41.
[51] Reza AM. Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. J VLSI Signal Process Syst Signal Imag Video Technol 2004;38(1):35–44.
[52] Nakai K, Hoshi Y, Taguchi A. Color image contrast enhacement method based on differential intensity/saturation gray-levels histograms. In: 2013 international symposium on intelligent signal processing and communication systems. IEEE; 2013, p. 445–9.
[53] Ying Z, Li G, Ren Y, Wang R, Wang W. A new image contrast enhancement algorithm using exposure fusion framework. In: International conference on computer analysis of images and patterns. Springer; 2017, p. 36–46.
[54] Wang S, Zheng J, Hu H-M, Li B. Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Trans Image Process 2013;22(9):3538–48.