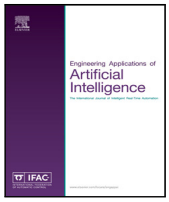
Engineering Applications of Artificial Intelligence 117 (2023) 105632

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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



LACN: A lightweight attention-guided ConvNeXt network for low-light image enhancement

Saijie Fan [a], Wei Liang [a], Derui Ding [a,*], Hui Yu [b]

[a] Department of Control Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China
[b] School of Creative Technologies, University of Portsmouth, PO1 2DJ, UK

ARTICLE INFO

Keywords:
Low-light image enhancement
ConvNeXt networks
Selective kernel attention modules
Feature fusion

ABSTRACT

Images captured under low-light conditions usually have poor visual quality, and hence greatly reduce the accuracy of subsequent tasks such as image segmentation and detection. In the low-light image enhancement task, noises in the dark areas are generally amplified while the images' brightness is enhanced. It should be pointed out that many deep learning methods cannot effectively suppress the noise at this stage and capture important feature information. To address the above problem, this paper proposes a Lightweight Attention-guided ConvNeXt Network (LACN) for low-light image enhancement. A novel Attention ConvNeXt Module (ACM) is first proposed by introducing a parameter-free attention module (i.e. SimAM) into the ConvNeXt backbone network. Then, a nontrivial lightweight network LACN based on a multi-attention mechanism is established through stacking two ACMs and fusing their features. In what follows, an improved hybrid attention mechanism, Selective Kernel Attention Module (SKAM), is adopted to effectively extract both global and local information. Such a module realizes the evaluation of lighting conditions for the whole image and the adaptive adjustment of the receptive field. Finally, through the feature fusion module, the features of different stages are aggregated to improve the ability of network to retain color information. Numerous experiments on low-light image enhancement are implemented via comparison with other state-of-the-art methods. Experiments show that the proposed method significantly improves the brightness and contrast of low-illumination images, preserves color information, and suppresses the generation of noises after image brightening.

1. Introduction

It should be pointed out that images captured in poor lighting conditions usually exhibit low brightness, low contrast, as well as poor color information due mainly to the influence of both the photography environment and the limitation of the equipment, Various shortcomings such as uneven shadows and overexposure still occur even under the scene with the device's flash turned on while shooting. These images not only have poor visual perception but also have a serious impact on subsequent advanced computer vision tasks, such as image segmentation, object detection, automatic driving, and so on. An effective approach to dealing with the above problems is a low-light image enhancement and hence receives an ever-increasing research topic in computer vision tasks.

The traditional low-light image enhancement methods mainly include two major categories, that is, the methods dependent on histogram equalization and ones based on the Retinx theory. Main idea of histogram equalization methods (Pizer et al., 1987; Pisano et al., 1998) is to increase the dynamic range of the gray value difference between pixels by adjusting the gray distribution of the image, thereby the overall contrast of the image is enhanced. The last one (Land, 1977; Ju et al., 2021; Jobson et al., 1997b; Hu et al., 2021; Jobson et al., 1997a) is based on the picture decomposition of reflection components, and illumination components, where the first component reflects the inherent attribute of the scene, including textures, edge details, and color information, while the second component roughly contains the distribution of contour information and lighting. The representative methods include, but are not limited to, SSR (Jobson et al., 1997b), and MSRCR (Jobson et al., 1997a) algorithms.

Low-light image enhancement is a sub-field of image restoration. Nowadays, some advanced image enhancement algorithms have gradually emerged benefiting from the rapid development of deep learning in the field of computer vision (Liang et al., 2022; Gao et al., 2019; Liu et al., 2021b; Zhang et al., 2021b). In light of the learning characteristics, there are the end-to-end learning methods (i.e. HWMNet Fan et al., 2022 and SARN Wei et al., 2021), the method based on learning of illumination and reflection components (i.e. Retinex-Net Wei et al., 2018, KinD Zhang et al., 2019 and KinD++ Zhang et al., 2021a), the unsupervised learning method (i.e. EnlightenGAN Jiang et al., 2021) as well as the semi-supervised learning method (i.e. DRBN Yang

* Corresponding author.
E-mail address: deruiding2010@usst.edu.cn (D. Ding).

https://doi.org/10.1016/j.engappai.2022.105632
Received 9 September 2022; Received in revised form 27 October 2022; Accepted 9 November 2022
Available online 21 November 2022

et al., 2020). It is worth mentioning that denoising is rarely considered in these methods, and the resulting huge amount of parameters is time-consuming and thus greatly limits their practicality.

Through the survey of traditional and deep learning methods, it is not difficult to find that there are still three considerable challenges in low-light image enhancement, which are disclosed as follows. (1) Serious noise effects are still ubiquitous after image brightening. Some deep-learning-based models developed in the past few years have not achieved ideal results. Generally, denoising before enhancement will cause information loss and make reconstruction difficult, while the output image could be blurred if denoising is performed after enhancement. (2) The ability of adaptive enhancement is relatively weak. For example, there may be overexposed after enhancement for the areas that are originally bright on the image and hence overall visibility is unsatisfactory. (3) The number of network parameters is huge. Specifically, the network's structure of many developed methods with good visual performance is commonly large and with a large number of parameters, which will result in time-cost in practical engineering. In this paper, useful information and global illumination information can be extracted by adding an attention mechanism to suppress noises and overexposures, and a parameter-free attention mechanism should be taken into consideration due mainly to the number of parameters of the overall network.

To sum up, this paper proposes a novel lightweight attention-guided ConvNeXt network (denoted as LACN), which can not only improve the image contrast and suppress noises but also enhance the ability of both color information retention and detail recovery. First, by introducing the attention mechanism into the ConvNeXt (Liu et al., 2022b) backbone network, an Attention ConvNeX Module (ACM) is constructed, and then two ACM modules are stacked. Second, inspired by SKNet (Li et al., 2019) (Selective Kernel Networks) and CBAM (Woo et al., 2018) (Convolutional Block Attention Module), an improved hybrid attention mechanism SKAM (Selective Kernel Attention Module) is adopted. Such a module is configured between two stacked ACMs to achieve adaptive adjustment of the receptive field. It has the capability of improving the lighting evaluation and the detail recovery while effectively suppressing noises. Third, the features of different layers, that is, shallow and deep features, are adequately fused to retain the color information and restore texture details. In conclusion, the model in this paper is lightweight and has a faster inference speed compared with reported approaches while brightening low-light images. The main contributions of this paper are as follows:

- A novel ACM is proposed by introducing the attention mechanism into the ConvNeXt backbone network. In light of such a module, a novel lightweight network LACN based on a multi-attention mechanism is established through both stacking two ACMs and fusing their features;
- The designed SKAM is nontrivial and effectively extracts both global and local information. Hence, it is beneficial to evaluate the lighting conditions of the whole image while realizing the adaptive adjustment of the receptive field;
- Two skip connections are adopted to make that shallow features can be backpropagation. Furthermore, a Feature Fusion Module (FFM) is employed to fuse features from different layers to retain abundant color information, that is, reducing the loss of color information;
- On the public data set, the superiority of the LACN model proposed in this paper is verified by comparative experiments. Furthermore, the effectiveness of the designed modules is also fully revealed via the ablation experiments.

2. Related work

2.1. Low-light image enhancement methods based on Retinex

Inspired by Retinex, some representative algorithms such as Single Scale Retinex (SSR), Multi-Scale Retinex (MSR), and Multi-Scale Retinex with Color Restoration (MSRCR) have been developed one by one in the past few years. The SSR algorithm estimates the illumination component through the Gaussian function and then subtracts the original image from the logarithmic domain to obtain the reflection component. MSR based on the SSR algorithm implements Gaussian filtering on the image at different scales and then obtains the illumination component by a weighted average of the results. The MSRCR algorithm solves the problem of color distortion caused by the enhancement of image brightness based on the MSR algorithm by adding a color recovery factor. Although achieved excellent results in the enhancement of brightness and contrast, the above methods still have certain limitations. For instance, the processing of high-noise pictures is still unsatisfactory, and the received image could occur insufficient enhancement of local brightness and loss of details.

2.2. Low-light image enhancement methods based on deep learning

In recent years, benefiting from the development of artificial intelligence, various deep learning models continuously emerge, and great breakthroughs have been made in the fields of image recognition (Li et al., 2022b; Lu et al., 2021; Xu et al., 2021), object detection (Dong et al., 2022; Zeng et al., 2022; Wu et al., 2022a), image super resolution (Yin et al., 2022; Kwasniewska et al., 2020; Liu et al., 2022a), image segmentation (Ronneberger et al., 2015; Yamanakkanavar and Lee, 2022), and so on. At the same time, the research of deep learning in the field of low-light image enhancement has also made some important development. As the first deep learning-based image enhancement method, LLNet (Lore et al., 2017) brightens and denoises low-light images through stacked sparse denoising autoencoders. GLADNet (Wang et al., 2018) consists of two sub-networks, a global illumination estimation network and a detail reconstruction network, which respectively realize the prediction of image illumination and the restoration of details. MBLLEN (Lv et al., 2018) involves three parts: a feature extraction module, an enhancement module, and a fusion module, and uses a new loss function to improve the quality of low-light image enhancement. By resorting to the advantages of convolutional neural networks, Retinex-Net (Wei et al., 2018) exploits a decomposition network that decomposes images into light maps and reflection maps and an augmentation network that brightens the lightness. Similarly, KinD (Zhang et al., 2019) is a deep learning network designed based on Retinex, where a reflectivity recovery network is added for denoising. In light of the KinD framework, KinD++ (Zhang et al., 2021a) has been developed to further improve the effect of image enhancement via a multi-scale illumination attention module. DLN (Wang et al., 2020a) regards the low-light image enhancement task as a residual learning problem and designs a lighting back-projection module using back-projection theory (Haris et al., 2018) in image super-score. DRBN (Yang et al., 2020) includes two steps, recursive paired supervised learning and unpaired GAN network learning, and hence integrates the advantages of supervised learning and unsupervised learning. EnlightenGAN (Jiang et al., 2021) is the first low-light image enhancement method based on unsupervised learning and does not require paired data for training. Zero-DCE (Guo et al., 2020) achieves brightness enhancement by designing a deep network to estimate both pixels and curves that can be dynamically adjusted. RUAS (Liu et al., 2021c) uses neural architecture search methods (Li et al., 2022a, 2021; Lyu et al., 2021a,b) to obtain an efficient network structure for both lighting estimation and denoising tasks. Very recently, the structure of the proposed URetinex-Net (Wu et al., 2022b) is composed of three modules for initialization, optimization, and illumination adjustment, and hence achieves noise suppression and detail preservation. Compared with traditional learning, deep learning methods have greater capability of reducing noise. However, a large-scale network is introduced for image enhancement, and the huge number of parameters makes it time-consuming. When applied to real-time work, huge delays could be caused.

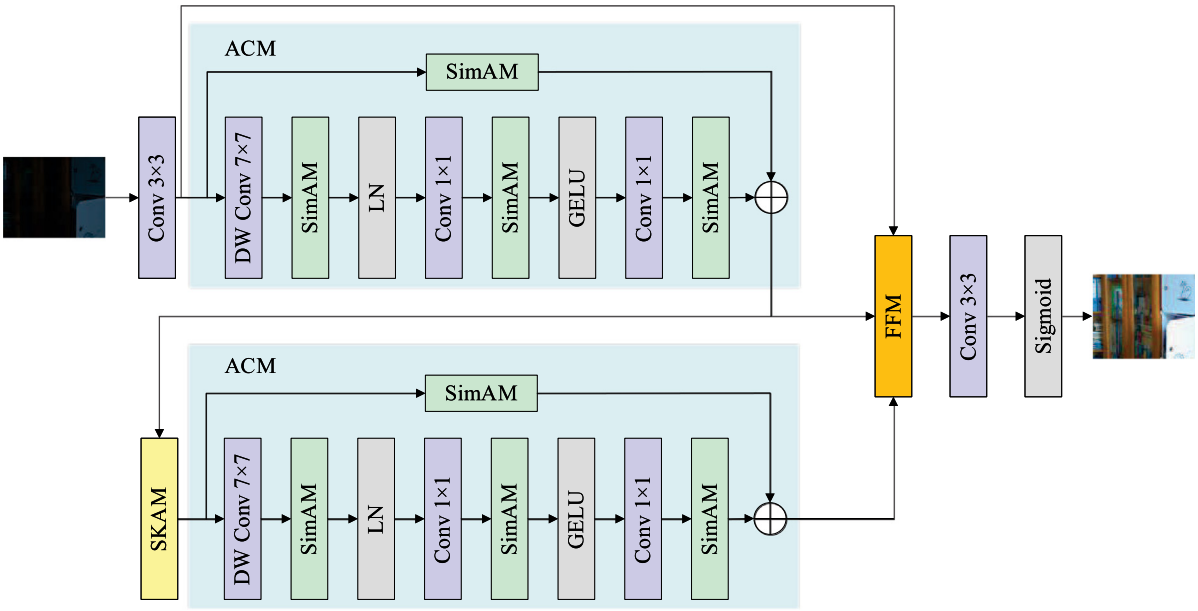


Fig. 1. Model structure of LACN. The brief process is to first perform shallow feature extraction on the image, then send it to ACM to extract features, enter ACM again after passing SKAM, moreover send the features extracted twice by ACM and shallow features into FFM, and finally carry out feature reconstruction to get an enhanced image.

2.3. Attention mechanisms

The purpose of attention mechanisms is to enhance the expressive ability of the network by emphasizing useful information and suppressing useless information and hence they have been widely used in the field of computer vision. SENet (Hu et al., 2018) (Squeeze-and-Excitation Networks) focuses on the relationship between channels and assigns weights to each channel after learning the weight information of the channels. SKNet (Li et al., 2019), an improved version of SENet, can adaptively adjust the size of the convolution kernel to obtain feature map information about different receptive fields with the help of multiple scales of input information. ECA (Wang et al., 2020b) (Efficient Channel Attention) performs the cross-channel interaction while maintaining the SENet performance and reducing the number of parameters. The hybrid attention CBAM (Woo et al., 2018), considering both the channel and the spatial dimension, not only assigns different importance degrees to different channels, but also assigns different importance degrees to different location features in the same channel. It is worth mentioning that the above attention modules often use additional sub-networks to generate attention weights. Recently, a new attention mechanism SimAM (Yang et al., 2021) (Simple Attention Module) was born. It calculates the attention weights by designing an energy function so that no additional sub-networks are needed to generate the attention weights and no additional parameters are needed to be introduced. In summary, the attention mechanism should be able to effectively suppress unnecessary noise when applied to the task of low-light image enhancement.

3. The proposed method

3.1. The overall structure of the new LACN

In recent years, Transformer (Vaswani et al., 2017) has provided deep insights into the field of computer vision and has also achieved some interesting results in image restoration (Qu et al., 2022). Subsequently, an improved version, Swin-transformer (Liu et al., 2021a), has been developed and further promoted its applications. Inspired by the Swin-transformer, the emerged ConvNeXt is comparable to Transformers in terms of accuracy, scalability, and robustness by improving key components in ResNet. In consideration of its advantages, this paper proposes a lightweight network LACN by adopting ConvNeXt as the backbone network to achieve better enhancement effects and faster inference speeds. The network is mainly composed of three modules: ACM, SKAM, and FFM, and its overall structure is roughly shown in Fig. 1. Particularly, ACM pays more attention to the learning of brightness information to preliminarily realize noise suppression. The purpose of SKAM added in the middle of the stacked ACM modules is to extract both global information and local information to improve the illumination evaluation ability. While guaranteeing higher processing power, it has better denoising ability than traditional ConvNeXt modules for real-world low-light images with noise levels. Finally, the function of the FFM is to adequately fuse shallow features and high-level features to keep richer color information. In the following subsections, the functions and novelties of each module as well as the adopted loss functions are presented in detail.

3.2. Design of attention ConvNeXt modules

The task of low-light image enhancement is not only to improve the overall brightness of the image, but also to restore the texture details and boundary information of objects while suppressing their surrounding noise information. As such, it is necessary to focus on extracting the boundary and texture information in shallow features and the brightness information in high-level features. To this end, this paper inserts some attention (i.e. SimAM attention) modules in the identity mapping path and after each convolutional layer in ConvNeXt. It can be seen from Fig. 1 that four attention modules are inserted into ConvNeXt. This kind of operation constructs a novel ACM based on ConvNeXt frameworks. It should be pointed out that the number of parameters of the network will inevitably increase if traditional SENet, ECANet, and other attention are used to replace the adopted SimAM. In other words, the adopted attention module is to assign higher weights to important features without increasing the number of parameters.

Specifically, the extracted shallow features x is inputted into the ACM module and a 7×7 depth-wise convolution is applied to make each convolution kernel process single channel information. Through the SimAM attention module and normalization layers LN, the feature x' with rich texture and boundary information will be extracted

$$x' = LN\left(S_A\left(f_{DW}^{7\times7}(x)\right)\right) \tag{1}$$

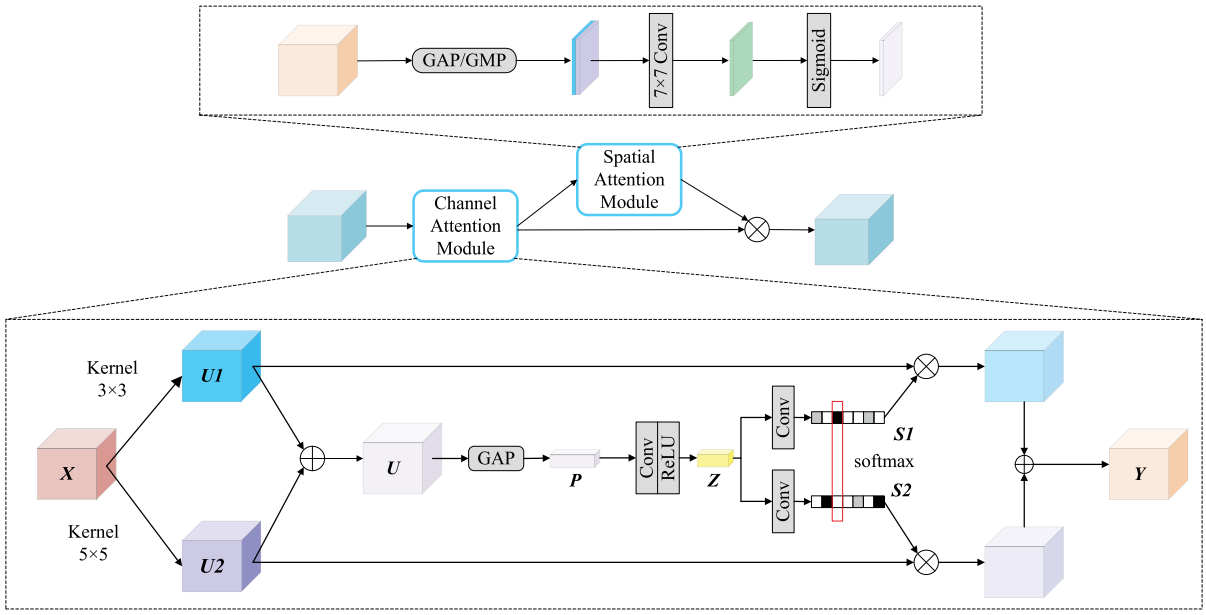


Fig. 2. Structure of SKAM.

where $f_{DW}(\cdot)$ stands for the depth-wise convolution operation, $LN(\cdot)$ means the layer normalization operator, and $S_A(\cdot)$ represents the SimAM attention operator. In what follows, the feature x'' is generated after performing 1×1 convolution, SimAM, and activation operations

$$x'' = \vartheta\left(S_A\left(f^{1\times1}\left(x'\right)\right)\right) \tag{2}$$

Among them, $f(\cdot)$ is the convolution operation, and ϑ means the GELU activation function. Then the received feature x'' goes through a layer of convolution and attention modules to fully extract the brightness information. Finally, the feature output y of the ACM module is generated via the outputs of the SimAM attention module combined with the outputs of the identity mapping path

$$y = S_A\left(f^{1\times1}\left(x''\right)\right) + S_A(x) \tag{3}$$

3.3. Design of selective kernel attention modules

The processing mode of observing a certain area for the first time is essentially an application of the attention mechanism. In this process, the brain will focus on this area and devote more attention, while ignoring other irrelevant information. In short, the attention mechanism can flexibly capture the connection between global information and local information. Its purpose is to allow the module to obtain the target area that needs to be focused on, that is, highlighting the salient useful features and suppressing and ignoring irrelevant features via putting big weights on this part.

The hybrid attention mechanism combines the channel attention mechanism and the spatial attention mechanism organically to make up for their respective shortcomings. As in the human brain, neurons in the visual cortex can adjust the receptive field in response to stimuli. Inspired by such a mechanism, the selective kernel attention module SKAM can be constructed by adaptively adjusting the weight of receptive fields of the channel attention in CBAM. In low-light image enhancement tasks, a small receptive field is helpful for detail recovery, and a large receptive field contributes to the global illumination evaluation. The serial form of channel attention and spatial attention is adopted in SKAM in this paper, whose structure is shown in Fig. 2. The channels of the image contain rich color features, but not all color features contribute greatly to the enhancement of low-light images. Therefore, the role of channel attention is to configure higher weights to useful color feature channels to achieve the higher attention, equivalently, suppress useless color information. There are three steps in the channel attention part.

The first step is the calculation of the multi-scale features. The feature maps of $H \times W \times C$ are performed the convolution operation with the convolution kernels of size 3×3 and 5×5 respectively, and the corresponding outputs are denoted as U_1 and U_2. In order to further improve the efficiency, the dilated convolution with an expansion rate of 2 is used to replace the 5×5 convolution kernel. Finally, the multi-scale features are combined into U, denoted as:

$$U = U_1 + U_2 = \delta\left(f^{3\times3}(X)\right) + \delta\left(f^{5\times5}(X)\right) \tag{4}$$

where δ represents the ReLU activation function.

The second step is global average pooling for U. The vector P of $1 \times 1 \times C$ is obtained by calculating the average value of each channel. Then, a compact feature representation Z of $1 \times 1 \times r$ is generated after the channel reduction of the fully connected layer. The process can be expressed as:

$$P = \frac{1}{H \times W} \sum_{i}^{H} \sum_{j}^{W} U_C(i,j) \tag{5}$$

$$Z = \delta\left(W_0 P\right) \tag{6}$$

where U_C is the feature of the Cth channel of U, W and H are respectively the width and height of the feature, and W_0 is the weight of the fully connected layer. In the experiments, r is set to be 32.

The third step is the calibrated feature map. For the received Z, two parallel convolutional layers are employed to generate Z_1 and Z_2, which are further utilized to generate attention activation S_1 and S_2 through Softmax. Such two activation matrices are adopted to adaptively recalibrate the feature maps U_1 and U_2. In summary, the calibrated feature map Y is calculated by

$$Y = S_1 \cdot U_1 + S_2 \cdot U_2 \tag{7}$$

where $S_1 = \frac{e^{Z_1}}{e^{Z_1}+e^{Z_2}}$ and $S_2 = \frac{e^{Z_2}}{e^{Z_1}+e^{Z_2}}$.

In comparison with channel attention providing large weights for useful color feature channels, the purpose of spatial attention is to generate large weights for useful spatial location information. In the task of low-light image enhancement, the influence of noises is more serious than that in traditional image tasks. The spatial attention mechanism has the capability of capturing the position information of the noises and processes the information unrelated to the task to restrain the noise

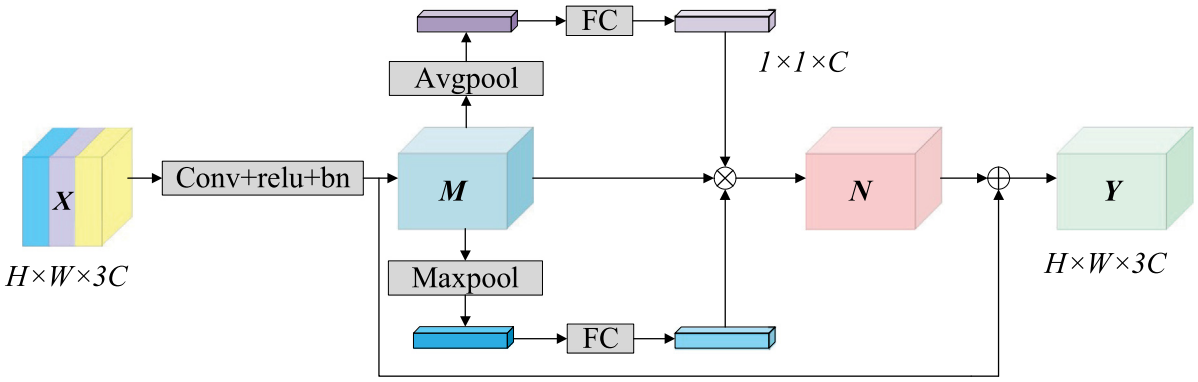


Fig. 3. Structure of feature fusion module.

effect. At the same time, through the captured brightness information, a larger weight is assigned to the dark area of the image, and a smaller weight is assigned to the highlighted area to avoid overexposure.

Two feature vectors of $H \times W \times 1$ are obtained in SKAM when Global Average Pooling (GAP) and Global Max Pooling (GMP) are simultaneously performed on the input features in the channel dimension. Jointing these two vectors as an input, the spatial attention vector Q is obtained after a 7×7 convolution and activation operation

$$Q = \sigma\left(f^{7\times7}\left(Z_{GAP}, Z_{GMP}\right)\right) \tag{8}$$

The above formula is a simple representation of the spatial attention mechanism, where σ represents the Sigmoid activation function, and Z_{GAP} and Z_{GMP} are the feature values obtained by global average pooling and global max pooling in the channel dimension.

3.4. Design of feature fusion modules

By resorting to ACM and SKAM, the brightness of the low-light image can be improved and the noise is well suppressed too. However, the experiment found that the output result is still seriously lost in color information compared with GroundTruth (GT) and hence is quite different from the real-world image. In the convolutional neural network, the shallow features contain rich texture information and color information, and the high-level features obtained by multi-layer convolution include global information. For low-light image enhancement, it is necessary to retain the color information in the output result as much as possible. To this end, an image channel is introduced in this paper to retain the information with specific color characteristics so that the information in dark areas can be better recovered, and the one in bright areas can prevent overexposure. Specifically, two skip connections are designed into the network to send shallow features to the back to solve the problem of loss of color information. At the same time, an FFM module as shown in Fig. 3 is proposed to fuse shallow features and high-level features more effectively. Its input consists of three parts, namely the features originally obtained by 3×3 convolution and the features of the two ACM outputs. This module contains two steps.

In the first step, the information from three feature maps of size $256 \times 256 \times 64$ is connected to form a feature map X of size $256 \times 256 \times 192$. Taking such a feature map as an input, a feature map M with a size of $256 \times 256 \times 192$ is further obtained via a convolution kernel of size 3×3 and step-size 1. This process can effectively capture color information and can be expressed as follows

$$M = \delta\left(f^{3\times3}(X)\right) \tag{9}$$

In the second step, the feature map M is recalibrated by assigning appropriate weights to different channels. For this purpose, a module consisting of two weighted branches and a short connection is constructed, as shown in Fig. 3. First, the feature maps are compressed into $1 \times 1 \times 192$ vectors by global average pooling. At the same time, such a feature map is also subjected to global max pooling to preserve the texture details as much as possible. Each element in the obtained $1 \times 1 \times 192$ vector represents channel information. Their formulas are described by

$$Z_{\mathrm{GAP}} = \frac{1}{H \times W} \sum_{i}^{H} \sum_{j}^{W} M_C(i, j) \tag{10}$$

$$Z_{\mathrm{GMP}} = \max\left(M_C\right) \tag{11}$$

where Z_{GAP} and Z_{GMP} are the feature values obtained by global average pooling and global max pooling respectively, and M_C is the feature of the Cth channel of M. In what follows, these two compressed vectors are weighted through two fully connected layers to obtain their respective weight vectors. Finally, a recalibrated feature map Y with a size of $256 \times 256 \times 192$ is received via the feature vectors multiplied by the corresponding weights

$$Y = \sigma\left(W_2\delta\left(W_1 Z_{\mathrm{GAP}}\right)\right) \cdot M \cdot \sigma\left(W_2\delta\left(W_1 Z_{\mathrm{GMP}}\right)\right) + M \tag{12}$$

where W_1 and W_2 are the weights of the two fully connected layers.

3.5. Loss function

To obtain an excellent image enhancement effect, the loss function used in this paper consists of four parts: Charbonnier L_1 loss function L_{char}, perceptual loss (Johnson et al., 2016) L_{perc}, smoothness loss L_{TV}, and SSIM structural similarity loss L_{ssim}. As such, the mixed loss function is expressed as follows:

$$Loss = L_{char} + L_{perc} + L_{TV} + L_{ssim} \tag{13}$$

(1) Charbonnier L_1 loss function L_{char}: The L_{char} loss, a variant of L_1, is used to preserve the color and brightness of the image in image enhancement. The Charbonnier loss used in this paper is compared with the L_1 loss, it can improve the model performance while dealing with model outliers, and can be expressed by

$$L_{char} = \sqrt{\|X_{output} - X_{high}\|^2 + \epsilon^2} \tag{14}$$

where X_{output} represents the enhanced image, X_{high} means the real image, ϵ is a constant and is set to be 10^{-3}.

(2) Perceptual loss L_{perc}: Such a loss is employed to constrain the similarity between the real image and the predicted image and hence can be utilized to improve the visual quality of the result. Its formula is denoted as follows

$$L_{perc} = \frac{1}{CHW}\|F\left(X_{output}\right) - F\left(X_{high}\right)\|_2^2 \tag{15}$$

where $F(\cdot)$ is an operator that extracts features from the pre-trained model VGG16 (Simonyan and Zisserman, 2014).

Smoothness loss L_{TV}: The loss is to minimize the gradient of the whole image to remove noises in the enhancement result, improve the

visual quality, and maintain the relationship between adjacent pixels. Its value can be calculated by

$$L_{TV}(Y) = \sum_{i=0}^{H}\sum_{j=0}^{W}\sqrt{\left(y_{i,j} - y_{i+1,j}\right)\left(y_{i,j} - y_{i,j+1}\right)} \tag{16}$$

where Y is the measured image, y stands for the pixel value on the image, and i and j are the indices of the pixels.

(4) SSIM structural similarity loss L_{ssim}: This loss function can be utilized to preserve the structure and texture information of the image by considering three aspects of contrast, brightness, and structure. The value range of SSIM is between 0 and 1, and the higher the value, the better the similarity. The loss function is defined as

$$L_{ssim}\left(X_{output}, X_{high}\right) = 1 - SSIM\left(X_{output}, X_{high}\right) \tag{17}$$

where $SSIM(x,y) = \frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1} \cdot \frac{2\sigma_{xy}+C_2}{\sigma_x^2+\sigma_y^2+C_2}$.

In the above formula, x and y represent the two measured images, μ and σ are the mean and variance of the pixel values of the measured images, and C_1 and C_2 are two set constants.

4. Experimental results and analysis

4.1. Experimental datasets

In this section, the developed LACN network will be trained on the LOL (Wei et al., 2018) dataset and its synthetic training dataset and then evaluated on its corresponding validation set. The details of the training and test sets are as follows.

(1) The LOL dataset contains 485 training images and 15 testing images, where the size of each image is 600 × 400, and each low-light image matches a well-exposed image as its reference image. Furthermore, to increase the number of training samples, this paper randomly crops each training image into 10 small images of size 256 × 256.

(2) The LOL synthetic dataset contains 1000 pairs of low-light/normal images, all of which are 384 × 384 in size. In this experiment, 30 pairs of images are randomly selected as the test set, and the remaining 970 pairs of images are used as the training set. Each image is randomly cropped into 4 small images of size 256 × 256 during training.

(3) The LSRW (Hai et al., 2021) dataset contains 5650 paired images, of which 3170 paired images were shot by Nikon D7500, of which 3150 were used as the training set, and 20 were used as test set; 2480 paired images were taken by Huawei P40 Pro, of which 2450 were used as the test set. The training set, 30 images are used as the test set, and the size of all images is 960 × 720. This experiment only evaluates its results on its test set.

4.2. Experimental setup

The experiments are implemented in the equipment with the Ubuntu18.04 operating system, an NVIDIA GeForce RTX 3090 graphics card, and the configured Pytorch deep learning framework. During the training process, the input image size is 256 × 256, the batch size is set to 2, the epoch is selected as 300, the gradient descent optimization algorithm used is Adam, and the initial learning rate is set to 5×10^{-5}. To make the network converge more stably and quickly, the cosine annealing strategy is employed to reduce the learning rate to 1×10^{-5}. The formula for decay is $\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min})\left(1 + \cos\left(\frac{T_{\text{cur}}}{T}\pi\right)\right)$, where $\eta_{\max}$ and $\eta_{\min}$ respectively represent the maximum and minimum learning rates, T_{cur} stands for how many epochs are currently executed, and T means the total number of epochs.

To quantitatively evaluate the low-light image enhancement performance of different methods, this paper adopts Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) (Zhang et al., 2018) as objective evaluation indicators. Among them, PSNR, which measures the fidelity of the image reconstruction, is based on the mean square error between the pixel points of the original image and the processed image; SSIM is to evaluate the similarity of the two images. The higher the PSNR and SSIM, the better the enhanced images. LPIPS with lower values representing closer restoration performance to human perception has the capability of evaluating the perceptual distance between images. The PSNR and LPIPS formulas are as follows

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right) = 20\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \tag{18}$$

$$LPIPS = d\left(X_{output}, X_{high}\right) = \sum_l \frac{1}{H_l W_l}\sum_{h,w}\left\|\omega_l \odot \left(\hat{X}^l_{high} - \hat{X}^l_{output}\right)\right\|_2^2 \tag{19}$$

where MSE means the mean square error between pixel points, MAX_I represents the maximum value of the image color point, d is the perceptual distance between the real image X_{high} and the predicted image X_{output}, l stands for the feature map of the lth layer, and $\hat{X}^l_{high}$ and $\hat{X}^l_{output}$ are the average of the real image features and the predicted image features of the l layers, respectively.

Table 1
Results of three quantitative indicators for different methods on the LOL dataset (Red: the best results, and Blue: the second-best ones).

Network models	PSNR↑	SSIM↑	LPIPS↓
Zero-DCE (2020)	14.56	0.553	0.313
RetinexNet (2018)	16.77	0.563	0.473
EnlightenGAN (2021)	17.48	0.578	0.322
MBLLEN (2018)	17.86	0.701	0.234
RUAS (2021)	18.23	0.717	0.354
DRBN (2020)	20.13	0.832	0.164
KinD (2019)	20.87	0.803	0.172
KinD++ (2021)	21.30	0.818	0.158
URetinexNet (2022)	21.33	0.833	0.121
LACN(ours)	22.54	0.843	0.113

4.3. Experimental results

To verify the superiority of the proposed network structure, lots of careful experiments are carried out in comparison with the state-of-the-art methods such as RetinexNet, MBLLEN, KinD, DRBN, Zero-DCE, KinD++, EnlightenGAN, RUAS, and URetinexNet.

4.3.1. Objective evaluation results

First, combined with various advanced methods, the PSNR, SSIM, and LPIPS results based on the LOL (Wei et al., 2018) dataset are shown in Table 1. It is not difficult to find from this table that the LACN model established in this paper achieves the best results in all three indicators. Specifically, the PSNR and SSIM indicators are 1.21 and 0.01 higher than the second-ranked URetinexNet network, respectively, and the LPIPS indicator is reduced by 0.008 and attains the smallest in all methods. Therefore, it can be verified that the LACN network proposed in this paper has great advantages in brightness enhancement, noise suppression, detail recovery, and color information retention.

The average scores of PSNR and SSIM indicators are shown in Tables 2 and 3 based on the LOL synthetic dataset as well as the LSRW dataset. Specifically, the SSIM metric achieves the highest value on both datasets, and the PSNR metric performs best on the LOL synthetic dataset, ahead of the second-ranked URetinexNet by 5.28, but slightly lower than the URetinexNet on the LSRW dataset.

4.3.2. Subjective evaluation results

In addition to the objective evaluation analysis, this subsection selects the visualization results of representative images on the LOL test set to further test the effectiveness.

It can be seen from Fig. 4 that the visualization effect of Zero-DCE and RUAS is unsatisfactory obviously. Zero-DCE cannot improve the overall brightness and does not adequately suppress noises; RUAS

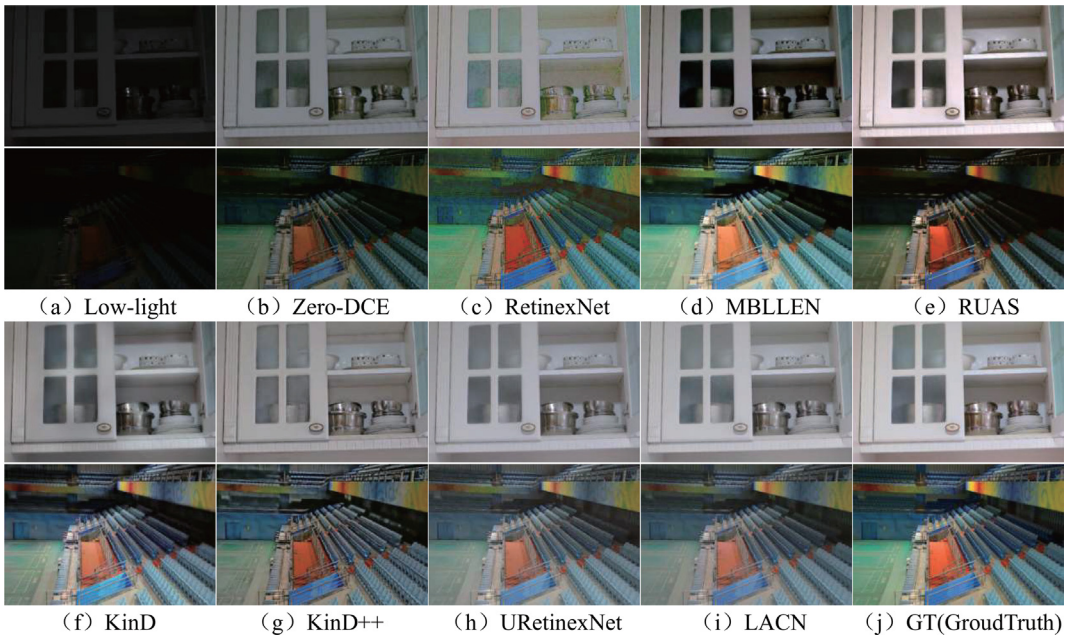


Fig. 4. Visualization results of different methods on the LOL dataset.

Table 2
Results of quantitative indicators for different methods on the LOL synthetic dataset (Red: the best results, and Blue: the second-best ones).

Network models	PSNR↑	SSIM↑
Zero-DCE(2020)	16.52	0.817
RetinexNet (2018)	17.20	0.764
EnlightenGAN (2021)	15.27	0.752
MBLLEN (2018)	14.26	0.655
RUAS (2021)	13.20	0.652
DRBN (2020)	16.24	0.732
KinD (2019)	16.65	0.759
KinD++ (2021)	16.96	0.794
URetinexNet (2022)	18.70	0.842
LACN(ours)	24.59	0.933

Table 3
Results of quantitative indicators for different methods on the LSRW dataset (Red: the best results, and Blue: the second-best ones).

Network models	PSNR↑	SSIM↑
Zero-DCE(2020)	15.83	0.467
RetinexNet(2018)	15.91	0.373
EnlightenGAN(2021)	16.31	0.470
MBLLEN(2018)	17.77	0.518
RUAS(2021)	14.44	0.428
DRBN(2020)	16.15	0.542
KinD(2019)	17.03	0.499
KinD++(2021)	16.97	0.412
URetinexNet(2022)	19.14	0.557
LACN(ours)	18.01	0.564

exposes the shortcoming of insufficient local area enhancement; RetinextNet, compared with the previous two methods, shows the advantage in the overall brightness of the image but, unfortunately, seriously amplifies the noises at the same time, which greatly affects the image quality. When MBLLEN is a concern, the local area enhancement is insufficient, the noises are not well suppressed, and the color is also over-saturated. Particularly, the color exists inconsistent in comparison with the GT. KinD and KinD++, developed by the same author, suppress noises very well and achieve a good performance in the overall brightness of the image, but the first makes the final result blurred and loses details, and the second generates obvious artifacts. URetinexNet is slightly imperfect in color retention. For example, the enhanced result is whiter in the image of the cupboard while the GT is slightly yellowish overall. Focusing on the method developed in this paper, the enhanced images have good preservation of details, excellent global and local brightness, almost no noises, artifacts, and exposure area, as well as the perfect color information as the same as that of GT.

The developed LACN network in this paper is also subjectively evaluated on the LOL synthetic dataset and the LSRW dataset. It can be found from Fig. 5 that images of bicycles and mountains enhanced by RetinexNet exist severe noises and artifacts although color information preservation is very well. The best visual effect in Fig. 6 is URetinex-Net and LACN in this article, but from the keyboard picture, it can be seen that after the URetinex-Net method is enhanced, the color of the keyboard has become much whiter, and the color information is seriously lost. These shortages are effectively solved by our network. In summary, the method in this paper can achieve good visual effects (including brightness, noises, color, textures, and so forth) on LOL dataset, the synthetic dataset and LSRW dataset.

4.4. Ablation experiments

This subsection performs ablation experiments to verify the effectiveness of each module in the proposed network, and the results are listed in Table 4. On one hand, by adding SimAM and SKAM in order, it can be seen from the first, second, and fifth rows that all indicators of PSNR, SSIM, and LPIPS have significantly improved. For instance, compared with the first and second rows, the contribution of SKAM reaches 1.39, 0.015, and 0.018 in these three indicators. When CBAM is replaced by SKAM, these indicators are also doing very well, which have been checked by the scores in the fourth and fifth rows. On the other hand, it can be seen from Fig. 7 that there are slight noises in the first and second images, the whole images are slightly blurred, and the color of the pillow is not well preserved. Furthermore, enhancement in the third picture is insufficient in some areas when CBAM attention was applied. In summary, the ablation experiment discloses that the contribution of the designed attention module is prominent and has an important improvement in image quality.

In what follows, the effectiveness of the added FFM module can be easily verified by the comparison between the third and fifth rows of Table 4. Particularly, the scores of PSNR and SSIM indicators are respectively increased by 1.14 and 0.016, and the LPIPS score is reduced by 0.011. Furthermore, it can be seen from Fig. 8 that the color

Fig. 5. Visualization results of different methods on the LOL synthetic dataset.

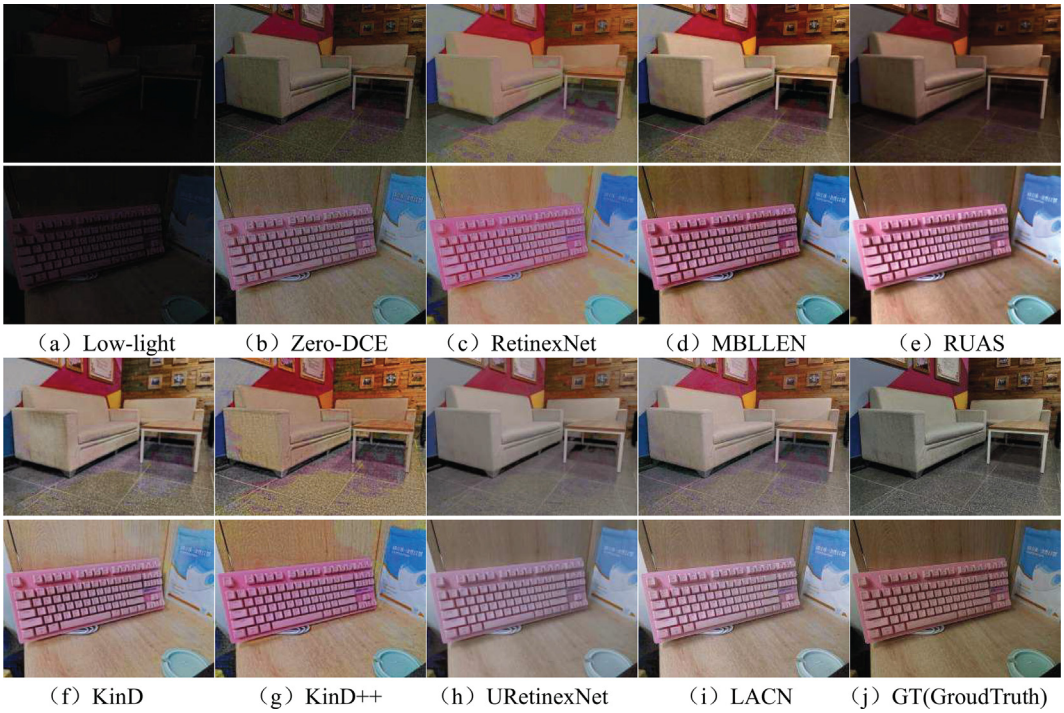


Fig. 6. Visualization results of different methods on the LSRW dataset.

Fig. 7. Visualization of attentional modules.

Fig. 8. Visualization of feature fusion modules.

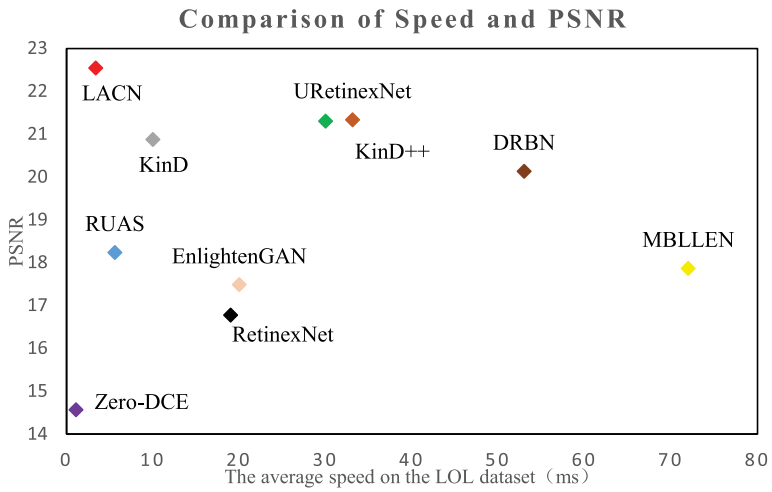


Fig. 9. Processing speeds of different networks.

Table 4
Evaluation of different modules on the LOL dataset (Red: the best results, and Blue: the second-best ones).

Models	PSNR↑	SSIM↑	LPIPS↓
w/o SimAM and SKAM	20.57	0.814	0.153
w/o SimAM	21.96	0.828	0.135
w/o FFM	21.40	0.827	0.124
SKAM→CBAM	22.19	0.832	0.125
LACN	22.54	0.843	0.113

of the bookcase door frame is significantly different from that of GT when there is no FFM, that is, the original brown door frame is much whiter than that after enhancement. In conclusion, from the objective and subjective evaluation, each module in the developed network plays an important role to increase the overall brightness and suppress noise, and also retain the color of the image well.

4.5. Supplementary experiments

This paper also compares the average speed of the model processing images on the LOL dataset. As shown in Fig. 9, the network proposed in this paper only takes 3.38 ms. Although the processing speed of Zero-DCE is only 1.12 ms, as shown in Fig. 4, Fig. 5, and Fig. 6, it has poor visualization and no denoising. The network designed in this paper is also lightweight and has the functions of brightness enhancement and denoising at the same time.

5. Conclusions

In this paper, a lightweight low-light image enhancement network, LACN, has been proposed, where the attention mechanism SimAM has been embedded in ConvNeXt to form ACM. It can effectively realize the extraction of texture details and brightness information, and suppress noise information. Furthermore, an improved SKAM has been adopted to effectively extract both global information and local information. Such a module realized the evaluation of the lighting conditions of the whole image and the adaptive adjustment of the receptive field. Furthermore, a special fusion module has been designed to better preserve the color information in the enhanced results. Extensive experiments demonstrated the superiority of our method in enhancing brightness, denoising ability, and color information retention, as well as the effectiveness of each designed module. The method in this paper solves the problems of underexposure, noise interferences, and loss of color information in low-illumination images to a certain extent. In the follow-up work, how to combine deep learning to propose a new network model to better retain color information is an important research

direction. Furthermore, the methods for low-light video enhancement can be further studied.

CRediT authorship contribution statement

Saijie Fan: Methodology, Software, Writing – original draft. **Wei Liang:** Formal analysis, Software, Investigation. **Derui Ding:** Conceptualization, Writing – review & editing, Supervision, Project administration. **Hui Yu:** Writing – review & 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.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61973219.

References

Dong, X., Yan, S., Duan, C., 2022. A lightweight vehicles detection network model based on YOLOv5. Eng. Appl. Artif. Intell. 113, http://dx.doi.org/10.1016/j.engappai.2022.104914, Art no. 104914.

Fan, C.-M., Liu, T.-J., Liu, K.-H., 2022. Half wavelet attention on M-net+ for low-light image enhancement. http://dx.doi.org/10.48550/arXiv.2203.01296, arXiv preprint arXiv:2203.01296.

Gao, S.-H., Cheng, M.-M., Zhao, K., Zhang, X.-Y., Yang, M.-H., Torr, P., 2019. Res2net: A new multi-scale backbone architecture. IEEE Trans. Pattern Anal. Mach. Intell. 43 (2), 652–662. http://dx.doi.org/10.1109/TPAMI.2019.2938758.

Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R., 2020. Zero-reference deep curve estimation for low-light image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1780–1789. http://dx.doi.org/10.1109/cvpr42600.2020.00185.

Hai, J., Xuan, Z., Yang, R., Hao, Y., Zou, F., Lin, F., Han, S., 2021. R2rnet: Low-light image enhancement via real-low to real-normal network. http://dx.doi.org/10.48550/arXiv.2106.14501, arXiv preprint arXiv:2106.14501.

Haris, M., Shakhnarovich, G., Ukita, N., 2018. Deep back-projection networks for super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1664–1673. http://dx.doi.org/10.1109/cvpr.2018.00179.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7132–7141. http://dx.doi.org/10.1109/TPAMI.2019.2913372.

Hu, J., Zhang, H., Liu, H., Yu, X., 2021. A survey on sliding mode control for networked control systems. Internat. J. Systems Sci. 52 (6), 1129–1147. http://dx.doi.org/10.1080/00207721.2021.1885082.

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

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

Jobson, D.J., Rahman, Z.-u., Woodell, G.A., 1997b. Properties and performance of a center/surround retinex. IEEE Trans. Image Process. 6 (3), 451–462. http://dx.doi.org/10.1109/83.557356.

Johnson, J., Alahi, A., Fei-Fei, L., 2016. Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision. Springer, pp. 694–711. http://dx.doi.org/10.1007/978-3-319-46475-6_43.

Ju, Y., Tian, X., Liu, H., Ma, L., 2021. Fault detection of networked dynamical systems: A survey of trends and techniques. Internat. J. Systems Sci. 52 (16), 3390–3409. http://dx.doi.org/10.1080/00207721.2021.1998722.

Kwasniewska, A., Ruminski, J., Szankin, M., Kaczmarek, M., 2020. Super-resolved thermal imagery for high-accuracy facial areas detection and analysis. Eng. Appl. Artif. Intell. 87, http://dx.doi.org/10.1016/j.engappai.2019.103263, Art no. 103263.

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

Li, S., Li, W., Wen, S., Shi, K., Yang, Y., Zhou, P., Huang, T., 2021. Auto-fernet: A facial expression recognition network with architecture search. IEEE Trans. Netw. Sci. Eng. 8 (3), 2213–2222. http://dx.doi.org/10.1109/TNSE.2021.3083739.

Li, X., Wang, W., Hu, X., Yang, J., 2019. Selective kernel networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 510–519. http://dx.doi.org/10.1109/CVPR.2019.00060.

Li, W., Wen, S., Shi, K., Yang, Y., Huang, T., 2022a. Neural architecture search with a lightweight transformer for text-to-image synthesis. IEEE Trans. Netw. Sci. Eng. 9 (3), 1567–1576. http://dx.doi.org/10.1109/TNSE.2022.3147787.

Li, H., Zeng, N., Wu, P., Clawson, K., 2022b. Cov-net: A computer-aided diagnosis method for recognizing COVID-19 from chest X-ray images via machine vision. Expert Syst. Appl. 207, http://dx.doi.org/10.1016/j.eswa.2022.118029, Art no. 118029.

Liang, W., Ding, D., Wei, G., 2022. Siamese visual tracking combining granular level multi-scale features and global information. Knowl.-Based Syst. 252, http://dx.doi.org/10.1016/j.knosys.2022.109435, Art no. 109435.

Liu, L., Jiang, Q., Jin, X., Feng, J., Wang, R., Liao, H., Lee, S.-J., Yao, S., 2022a. CASR-net: A color-aware super-resolution network for panchromatic image. Eng. Appl. Artif. Intell. 114, http://dx.doi.org/10.1016/j.engappai.2022.105084, Art no. 105084.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021a. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10012–10022. http://dx.doi.org/10.1109/ICCV48922.2021.00986.

Liu, L., Ma, L., Zhang, J., Bo, Y., 2021b. Distributed non-fragile set-membership filtering for nonlinear systems under fading channels and bias injection attacks. Internat. J. Systems Sci. 52 (6), 1192–1205. http://dx.doi.org/10.1080/00207721.2021.1872118.

Liu, R., Ma, L., Zhang, J., Fan, X., Luo, Z., 2021c. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10561–10570. http://dx.doi.org/10.1109/cvpr46437.2021.01042.

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., Xie, S., 2022b. A convnet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11976–11986. http://dx.doi.org/10.1109/CVPR52688.2022.01167.

Lore, K.G., Akintayo, A., Sarkar, S., 2017. Llnet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognit. 61, 650–662. http://dx.doi.org/10.1016/j.patcog.2016.06.008.

Lu, P., Song, B., Xu, L., 2021. Human face recognition based on convolutional neural network and augmented dataset. Syst. Sci. Control Eng. 9 (sup2), 29–37. http://dx.doi.org/10.1080/21642583.2020.1836526.

Lv, F., Lu, F., Wu, J., Lim, C., 2018. MBLLEN: Low-light image/video enhancement using CNNs. In: BMVC. Vol. 220, (1), p. 4.

Lyu, B., Wen, S., Shi, K., Huang, T., 2021a. Multiobjective reinforcement learning-based neural architecture search for efficient portrait parsing. IEEE Trans. Cybern. http://dx.doi.org/10.1109/TCYB.2021.3104866.

Lyu, B., Yang, Y., Wen, S., Huang, T., Li, K., 2021b. Neural architecture search for portrait parsing. IEEE Trans. Neural Netw. Learn. Syst. http://dx.doi.org/10.1109/TNNLS.2021.3104872.

Pisano, E.D., Zong, S., Hemminger, B.M., DeLuca, M., Johnston, R.E., Muller, K., Braeuning, M.P., Pizer, S.M., 1998. Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. J. Digital Imaging 11 (4), 193–200. http://dx.doi.org/10.1007/bf03178082.

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

Qu, L., Liu, S., Wang, M., Song, Z., 2022. Transmef: A transformer-based multi-exposure image fusion framework using self-supervised multi-task learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 36, (2), pp. 2126–2134. http://dx.doi.org/10.1609/aaai.v36i2.20109.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241. http://dx.doi.org/10.1007/978-3-319-24574-4_28.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. http://dx.doi.org/10.48550/arXiv.1409.1556, arXiv preprint arXiv:1409.1556.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.

Wang, L.-W., Liu, Z.-S., Siu, W.-C., Lun, D.P., 2020a. Lightening network for low-light image enhancement. IEEE Trans. Image Process. 29, 7984–7996. http://dx.doi.org/10.1109/TIP.2020.3008396.

Wang, W., Wei, C., Yang, W., Liu, J., 2018. Gladnet: Low-light enhancement network with global awareness. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). IEEE, pp. 751–755. http://dx.doi.org/10.1109/FG.2018.00118.

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q., 2020b. ECA-net: Efficient channel attention for deep convolutional neural networks. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, Seattle, WA, USA. pp. 11531–11539. http://dx.doi.org/10.1109/cvpr42600.2020.01155.

Wei, C., Wang, W., Yang, W., Liu, J., 2018. Deep retinex decomposition for low-light enhancement. http://dx.doi.org/10.48550/arXiv.1808.04560, arXiv preprint arXiv:1808.04560.

Wei, X., Zhang, X., Li, Y., 2021. SARN: A lightweight stacked attention residual network for low-light image enhancement. In: 2021 6th International Conference on Robotics and Automation Engineering. ICRAE, IEEE, pp. 275–279. http://dx.doi.org/10.1109/ICRAE53653.2021.9657795.

Woo, S., Park, J., Lee, J.-Y., Kweon, I.S., 2018. Cbam: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 3–19. http://dx.doi.org/10.1007/978-3-030-01234-2_1.

Wu, P., Li, H., Zeng, N., Li, F., 2022a. FMD-yolo: An efficient face mask detection method for COVID-19 prevention and control in public. Image Vis. Comput. 117, http://dx.doi.org/10.1016/j.imavis.2021.104341, Art no. 104341.

Wu, W., Weng, J., Zhang, P., Wang, X., Yang, W., Jiang, J., 2022b. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5901–5910. http://dx.doi.org/10.1109/CVPR52688.2022.00581.

Xu, L., Song, B., Cao, M., 2021. A new approach to optimal smooth path planning of mobile robots with continuous-curvature constraint. Syst. Sci. Control Eng. 9 (1), 138–149. http://dx.doi.org/10.1080/21642583.2021.1880985.

Yamanakkanavar, N., Lee, B., 2022. MF2-net: A multipath feature fusion network for medical image segmentation. Eng. Appl. Artif. Intell. 114, http://dx.doi.org/10.1016/j.engappai.2022.105004, Art no. 105004.

Yang, W., Wang, S., Fang, Y., Wang, Y., Liu, J., 2020. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 3063–3072. http://dx.doi.org/10.1109/CVPR42600.2020.00313.

Yang, L., Zhang, R.-Y., Li, L., Xie, X., 2021. Simam: A simple, parameter-free attention module for convolutional neural networks. In: International Conference on Machine Learning. PMLR, pp. 11863–11874.

Yin, P., Liu, Z., Wu, D., Huo, H., Wang, H., Zhang, K., 2022. Unsupervised simple siamese representation learning for blind super-resolution. Eng. Appl. Artif. Intell. 114, http://dx.doi.org/10.1016/j.engappai.2022.105092, Art no. 105092.

Zeng, N., Wu, P., Wang, Z., Li, H., Liu, W., Liu, X., 2022. A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection. IEEE Trans. Instrum. Meas. 71, 1–14. http://dx.doi.org/10.1109/TIM.2022.3153997.

Zhang, Y., Guo, X., Ma, J., Liu, W., Zhang, J., 2021a. Beyond brightening low-light images. Int. J. Comput. Vis. 129 (4), 1013–1037. http://dx.doi.org/10.1007/s11263-020-01407-x.

Zhang, Q., He, Y., Luo, N., Patel, S.J., Han, Y., Gao, R., Modak, M., Carotta, S., Haslinger, C., Kind, D., et al., 2019. Landscape and dynamics of single immune cells in hepatocellular carcinoma. Cell 179 (4), 829–845. http://dx.doi.org/10.1016/j.cell.2019.10.003.

Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O., 2018. The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 586–595. http://dx.doi.org/10.1109/cvpr.2018.00068.

Zhang, J., Song, J., Li, J., Han, F., Zhang, H., 2021b. Observer-based non-fragile H_∞-consensus control for multi-agent systems under deception attacks. Internat. J. Systems Sci. 52 (6), 1223–1236. http://dx.doi.org/10.1080/00207721.2021.1884917.