Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Filter-cluster attention based recursive network for low-light enhancement*

Zhixiong HUANG[†1,3], Jinjiang LI[†‡2,3], Zhen HUA[1,3], Linwei FAN[4]

[1]*School of Information and Electronic Engineering, Shandong Technology and Business University, Yantai 264005, China*

[2]*School of Computer Science and Technology, Shandong Technology and Business University, Yantai 264005, China*

[3]*Co-innovation Center of Shandong Colleges and Universities: Future Intelligent Computing, Shandong Technology and Business University, Yantai 264005, China*

[4]*School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan 250014, China*

[†]E-mail: hzxcyanwind@163.com; lijinjiang@gmail.com

Received Aug. 11, 2022; Revision accepted Nov. 28, 2022; Crosschecked July 1, 2023

Abstract: The poor quality of images recorded in low-light environments affects their further applications. To improve the visibility of low-light images, we propose a recurrent network based on filter-cluster attention (FCA), the main body of which consists of three units: difference concern, gate recurrent, and iterative residual. The network performs multi-stage recursive learning on low-light images, and then extracts deeper feature information. To compute more accurate dependence, we design a novel FCA that focuses on the saliency of feature channels. FCA and self-attention are used to highlight the low-light regions and important channels of the feature. We also design a dense connection pyramid (DenCP) to extract the color features of the low-light inversion image, to compensate for the loss of the image's color information. Experimental results on six public datasets show that our method has outstanding performance in subjective and quantitative comparisons.

Key words: Low-light enhancement; Filter-cluster attention; Dense connection pyramid; Recursive network
https://doi.org/10.1631/FITEE.2200344 **CLC number:** TP391

1 Introduction

The continuous updating and upgrading of current information technology can be seen everywhere in efforts to record real-time information by taking images. However, the camera results are not always satisfactory. In some adverse environments, especially low-light scenes with uneven lighting, the light information received by the camera equipment is insufficient, and the collected images often show low brightness, poor clarity, and weak contrast. This situation causes texture information in the image itself to be hidden in the dark background. This not only violates the photographer's desire to record information, but also affects subsequent collection of information from the image. Although such problems can be alleviated by extending the exposure time, new problems of overexposure and scene blur may arise when using the deep retinex network (Yang WH et al., 2021a). Therefore, it is of vast significance to construct a practical low-light enhancement framework to improve the quality of low-light images. This effort will benefit some vision tasks, e.g., target detection (Liu L et al., 2020; Xie et al., 2021), real-time tracking (Yang B et al., 2021), and autonomous driving (Peng et al., 2020; Aradi, 2022), performed in low-light environments, such as during the night.

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 61772319, 62002200, and 62202268), the Shandong Natural Science Foundation of China (Nos. ZR2021QF134 and ZR2021MF107), the Shandong Provincial Science and Technology Support Program for Youth Innovation Team in Colleges (Nos. 2021KJ069 and 2019KJN042), and the Yantai Science and Technology Innovation Development Plan (No. 2022JCYJ031)

ORCID: Jinjiang LI, https://orcid.org/0000-0002-2080-8678

© Zhejiang University Press 2023

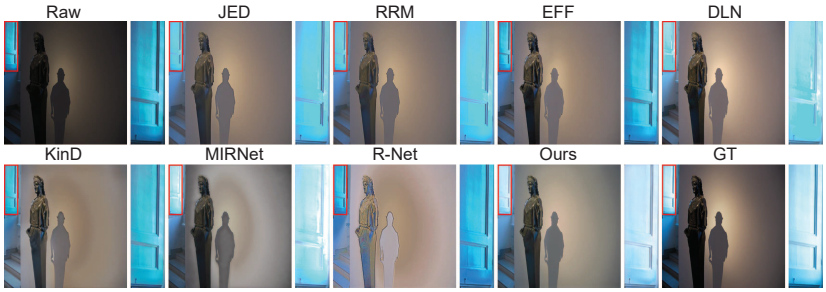

Fig. 1 Detailed enlargement comparison on the MIT-Adobe FiveK dataset

Fig. 1 shows a detailed magnification comparison on an MIT-Adobe FiveK dataset (Bychkovsky et al., 2011). The low-light problems can be seen in the raw image. Compared with other results, our result is most consistent in color with the ground truth, and the texture details on the door have good clarity.

In recent years, low-light image enhancement has become a research hotspot of computer vision. A large number of technical studies have emerged in related fields. They can be roughly classified into three categories: global adjustment based methods, physical model based methods, and deep learning based methods.

1.1 Global adjustment based methods

Low-light images usually show dense black pixels, and histogram equalization (HE) and gamma correction (GC) have served as inspiration for such methods. HE adjusts the probability density function of gray scale to make the distribution of image pixels more uniform, which improves the contrast of the image. Since there is no constraint on the enhancement of local information, early HE-based methods (Cheng and Shi, 2004; Abdullah-Al-Wadud et al., 2007) generated overexposure, chaotic color, and numerous artifacts, so some optimized HE methods were proposed. Celik and Tjahjadi (2011) and Lee et al. (2013) obtained smoother two-dimensional histograms by amplifying the gray-scale difference between adjacent pixels. Wu et al. (2017) adaptively adjusted the contrast gain from HE based on spatial saliency and dark channel prior. Li CL et al. (2020) used the quasi-symmetric correction function to provide pre-processing for the HE input. Xu CR et al. (2020) designed a visibility evaluation model to set the threshold in the histogram to avoid the noise amplification problem.

GC uses the characteristics of illumination distribution to adjust the gamma value in the power exponential function, where all pixels of low-light images are mapped by the function to improve the contrast. Shiau et al. (2015) used bidimensional empirical mode decomposition to assist the contrast adjustment of GC. Singh H et al. (2017) used the optimally weighted interim intensity channel to derive a more accurate gamma value-set. Wang W et al. (2019) proposed a variational model incorporating an energy function to determine the appropriate local gamma value. Li PL et al. (2021) preprocessed images under the iterative back-projection framework, and then adjusted the image brightness by adaptive GC. Xu YD et al. (2021) fused images of different frequency bands by constructing exposure, global contrast, and local contrast weight maps.

These methods have led to great improvement in the contrast of low-light images. However, due to insufficient attention to the natural authenticity of the images, the results generated are prone to uneven exposure or unnatural colors, which leads to a certain deviation between the enhanced image and the real scene.

1.2 Physical model based methods

Based on human visual perceptual characteristics, retinex theory decomposes the image into illumination and reflection layers, and enhances the image by adjusting and combining these two layers. Guo XJ et al. (2017) constructed the illumination layer using R, G, and B channel maximum pixels, and then refined the illumination layer using a prior. Li MD et al. (2018) proposed a robust retinex model to estimate the noise map to deal with strong noise hidden in low-light scenes. Ren XT et al. (2018) enhanced the low-light image in the illumination layer

while suppressing noise in the reflection layer. Hao et al. (2020) estimated the illumination layer by the Gaussian total variation filter and jointly estimated the reflection layer through the retinex constraint. Liu RS et al. (2021) provided a low-light optimization process for images based on retinex rules, and established the low-light prior by the reference-free learning strategy. Yang WH et al. (2021a) obtained the illumination map and reflection map of the enhanced image through the enhancement network and the restoration network, respectively.

Many studies have noticed the connection between other physical models and low-light images as a way to perform low-light image enhancement. Li L et al. (2015) first denoised the image based on the block-matching and three-dimensional filtering (BM3D) and the structural filtering, and then adaptively enhanced the contrast with the dark channel priori model. Yu and Zhu (2019) proposed a physical lighting model with the environmental light and the light-scattering attenuation rate to restore low-light images. Ying et al. (2017) adjusted the color and brightness of low-light images through the camera response model and illumination estimation techniques. Chen et al. (2019) designed an entropy-preserving mapping to enhance image contrast. Wang SH and Luo (2018) enhanced image contrast and naturalness based on a naturalness-preserving enhancement algorithm and multi-layer brightness prior. Wang YF et al. (2019) constructed an absorption light-scattering model to adjust low-light images by reducing the influence of atmospheric light and transmittance. Singh N and Bhandari (2021) improved the image quality based on a reflection model and principal component analysis.

The model-based methods add physical interpretation to low-light image enhancement, enabling significant results in light adjustment of images. However, the parameter adjustment of the model generally originates from manual design, which is not robust enough for low-light image enhancement in a variety of scenes, and the results are prone to overexposure or underexposure. In addition, the noise of the image itself is susceptible to amplification by such methods, resulting in local detail blur.

1.3 Deep learning based methods

Recently, deep learning, which has achieved great success in many fields (Lim et al., 2020; Yan et al., 2020a, 2020b; Li JJ et al., 2021; Huang ZX et al., 2022; Zhang TL et al., 2022), has been widely used in the field of low-light enhancement. Lore et al. (2017) proposed a new stacked sparse denoising autoencoder for image enhancement and noise reduction. Zhang YH et al. (2019) performed light adjustment and degradation removal on images in the network, and trained the network under various lighting conditions. Ren WQ et al. (2019) combined an autoencoder and a spatially variant recurrent neural network to construct more edge details. Xu K et al. (2020) proposed a network for different frequency layers. The network removes noise and enhances details in low-frequency layers and high-frequency layers, respectively. Wang LW et al. (2020) proposed lightening back-projection to assist the network in iterative enhancement. Lv et al. (2021) directed attention to brightness enhancement and noise removal on a synthetic low-light dataset. Ma et al. (2022) designed two estimation networks of brightness and reflectivity, and used spatial correlation to connect these two networks.

Some deep learning methods avoid the requirement for paired training data by generative adversarial networks (GANs) or other methods. Jung et al. (2020) used cycle adversarial networks to transform low-light image styles, and evaluated the image quality through the proposed frame consistency. Guo CL et al. (2020) proposed zero-reference deep curve estimation (Zero-DCE), which converts low-light image enhancement into specific curve estimation, and enhances the visibility of low-light images by adjusting the specific curve. Liu YJ et al. (2021) used Zero-DCE to preprocess images and added fractional order differential gradient masks to the discriminator. Yang WH et al. (2021b) designed a deep recursive band representation to integrate fully supervised and semi-supervised networks. Jiang et al. (2021) used self-regularization in unsupervised networks to guide the low-light image enhancement process.

Deep learning methods have led to impressive improvements in low-light image visibility, but still fall short when there are excessive color changes in low-light scenes, and the color information loss problem is often seen in the deep network architecture. In this work, we propose a recursive network based on filter-cluster attention (FCA), and the main contributions of our method are as follows:

1. We propose FCA to measure the feature space

and channel saliency, which makes the network pay more attention to the defect regions and important channels.

2. We add a dense connection pyramid (DenCP) to extract the inverted image features with high brightness, so the dark regions in each iterative learning can be enhanced.

3. We design a multi-stage recursive network, the network input of which consists of both low-light images and FCA results. After learning three units, i.e., difference concern, gate recurrent, and iterative residual, the learning results are further combined with DenCP to obtain the final output.

2 Methodology

The detailed framework of the proposed network is shown in Fig. 2. In this section, we will introduce the details of the network. First, we will introduce the proposed FCA, then the designed DenCP, and finally the detailed network structure and the loss function.

2.1 Network structure

As shown in Fig. 2, we design a multi-stage recurrent network for low-light image enhancement, which consists of three units: difference concern, gate recurrent, and iterative residual. The difference concern unit focuses on the difference between the learning results of each iteration and the output of FCA, the gate recurrent unit (GRU) selectively remembers the multi-stage global information, and the iterative residual unit performs intensive feature extraction for this iteration. Finally, the enhanced image is obtained by two-step mapping between the brightening result of DenCP and the residual learning result. Next, we describe the three units of the network in detail.

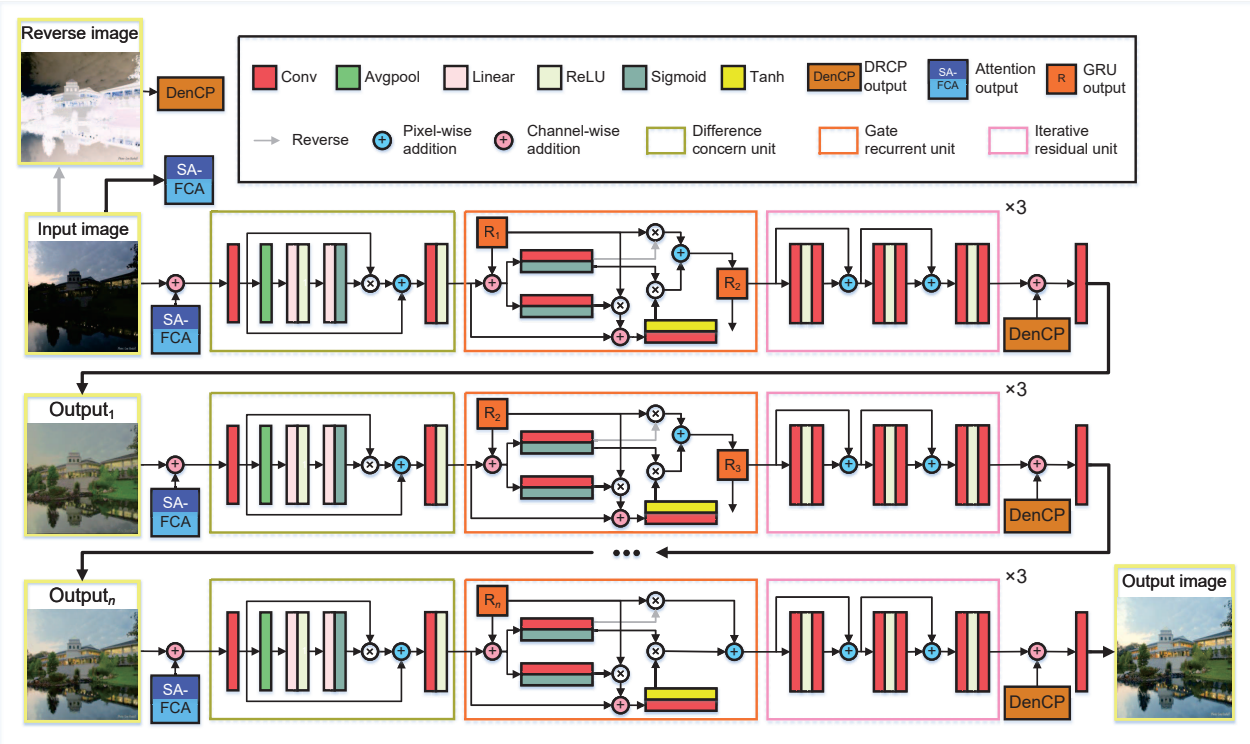

Fig. 2 Architecture of the proposed network. The network transforms low-light images into enhanced images after multiple iterations, and the input of each iteration consists of the result of the previous iteration and the two-stage attention output. The two-stage attention is composed of self-attention and the proposed filter-cluster attention (FCA). Each iteration is divided into three units, difference concern, gate recurrent, and iterative residual, and the results of the three units are superimposed on the dense connection pyramid (DenCP) output to obtain the enhanced image. The detailed structures of FCA and DenCP are shown in Figs. 3 and 4, respectively

2.1.1 Difference concern unit

By two-step calculation of the input image space and channel dependency, FCA highlights the low-light regions and important channels, and the difference concern unit focuses on the difference between each iteration result and the FCA output. This enhances the network's ability to enhance the low-light regions and improve brightness and contrast performance.

This unit first superimposes the three-channel iterative output and the FCA output to obtain a feature input of shape $H \times W \times 6$. To learn the difference between each iteration result and the FCA output, the feature input is first expanded to 32 channels by convolution, and then the feature is compressed into a real sequence $s_{\mathrm{r}} \in \mathbb{R}^{32\times1\times1}$ by the avgpool layer. The sequence not only represents each channel of the previous iteration result, but also includes the salient channels that are labeled in FCA. The difference in the sequence is obtained by crossing the linear layer with the activation function. The process is shown in Eq. (1):

$$d = F_{\mathrm{Sig}}\left(w_2 F_{\mathrm{R}}\left(w_1 s_{\mathrm{r}}\right)\right), \tag{1}$$

where $d \in \mathbb{R}^{32\times1\times1}$ represents the calculated difference, F_{Sig} and F_{R} represent the nonlinear sigmoid function and the linear ReLU function respectively, and w_1 and w_2 represent the weights obtained from the learning of two linear layers. We multiply d by the previous 32 channel features to label the difference of each channel. The labeled results are connected with the input features to prevent the vanishing gradient. Finally, the input for the next stage is obtained by convolution and the ReLU function.

2.1.2 Gate recurrent unit

In the long-term memory and back-propagation of multiple iterations, the network is prone to problems such as vanishing gradient and information loss. Adding recurrent units in network iterations can provide a good solution to such problems. Consistent with the idea of long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997), GRU (Cho et al., 2014) can improve network memory for important past information, and can independently filter the remaining information. Different from LSTM, GRU uses only two reset gates and updates to achieve good performance; that is to say, the cost of GRU calculation is lower, which is undoubtedly more profitable for multi-iteration networks.

The input x of GRU is composed of the previous unit output and the previous GRU output. The sigmoid function is used to convert x to the gate signal in [0, 1]. The two signals are calculated as

$$s = \sigma\left(w\left(x, r_{i-1}\right)\right), \tag{2}$$

where s represents the gate signal, σ represents the sigmoid function, w represents the content selection weight for convolutional learning, x is the output of the previous unit, and r_{i-1} is the GRU output from the previous iteration. The gate signal is then multiplied by r_{i-1} to reset the data, and the reset data are added to the previous unit input and transferred to the tanh function. The current unit information is obtained as

$$r' = \tanh\left(x, s_i r_{i-1}\right). \tag{3}$$

Finally, we update the gate signal s_i to forget and accept information. When s is closer to 1, more information needs to be accepted; when s is closer to 0, more information needs to be forgotten. The update expression is shown as

$$r_i = (1-s) r_{i-1} + s r', \tag{4}$$

where r_i represents the GRU output from this iteration, $(1-s)r_{i-1}$ represents the selective forgetting of the output from the pervious iteration, and sr' represents the selective undertaking of the current unit information.

2.1.3 Iterative residual unit

The superposition of residual blocks brings deep feature extraction, but also tends to bring complex parameters that affect the network performance. Therefore, we choose the iterative approach to share the parameters for each round of residual computation. We design an iterative residual unit containing three rounds of the residual learning process. Each round of residual learning passes through three learning regions consisting of two convolutional layers crossed with the ReLU activation function, in which the convolutional kernel size and step size of all convolutional layers are 3×3 and 1, respectively. Each region is skip connected with the output of the front region after learning.

The result of residual learning is superimposed with the DenCP brightening result, and undergoes a two-step mapping process consisting of convolution and ReLU crossover: $f^{H\times W\times 2\times C} \rightarrow f^{H\times W\times C} \rightarrow f^{H\times W\times 3}$. The first step of mapping learns the weights between the brightening result and the residual result channels, and the channels are fused to compensate for the under-brightened regions in this iteration learning. The second step of mapping restores the features to three channels, thus outputting the enhanced image of this iteration and transferring the image to the next iteration for further enhancement.

2.2 Filter-cluster attention

In this network, the input image first goes through a two-stage attention process, and the first stage is the self-attention process that focuses on region correlation. Because there are defective regions with insufficient brightness in low-light images, by self-attention the dependence between image regions can be calculated to focus on dark pixel concentration regions. Each channel of the feature contains different information and different saliency. Therefore, in the second stage, we design FCA that focuses on channel saliency and dependency. FCA first filters channels with low saliency, then calculates the dependence and redistributes the channel weights, and finally combines each channel group to obtain the final output. Compared with similar channel attention, FCA also undergoes the compression-expansion process, but introduces sparsity and adaptivity in the attention computation process, which enables features to obtain more tightly connected channels after dependency filtering, enabling correlation computation to be performed accurately.

Fig. 3 shows the detailed structure of self-attention (SA) and FCA. At the end of the self-attention, we set a deconvolution layer to expand the feature channel to facilitate further FCA calculation. The self-attention results first pass through the normalization and ReLU activation functions. To calculate the channels, the avgpool operation is used to compress the features f of shape $C \times H \times W$ to the channel real number $\mathbb{R}_c$ of shape $C \times 1 \times 1$, where c is the number of channels. To motivate the network to adopt different sub-channels to express the rich features, $\mathbb{R}_c$ is inputted to the saliency learning region of each group, which consists of two regions where the linear layer and the ReLU function intersect, $\mathbb{R}_c$ undergoes a mapping process ($\mathbb{R}_c \rightarrow \mathbb{R}_{c(1-s)} \rightarrow \mathbb{R}_c$) in this region, and s is the compression rate of the first linear layer. The saliency learning about the ith group is as given in Eqs. (5) and (6):

$$\mathbb{R}_{c(1-s)} = F_{\text{s}}\left(\mathbb{R}_c\right), \tag{5}$$

$$p^i = F_{\text{R}}\left(w^i\left(\mathbb{R}_{c(1-s)}\right) + b^i\right), \tag{6}$$

where p^i represents the saliency vector of the ith group, w^i and b^i represent the weight and bias learned at the two cross regions respectively, $F_{\text{s}}(x)$ represents the channel compression of $\mathbb{R}_c$ at the first

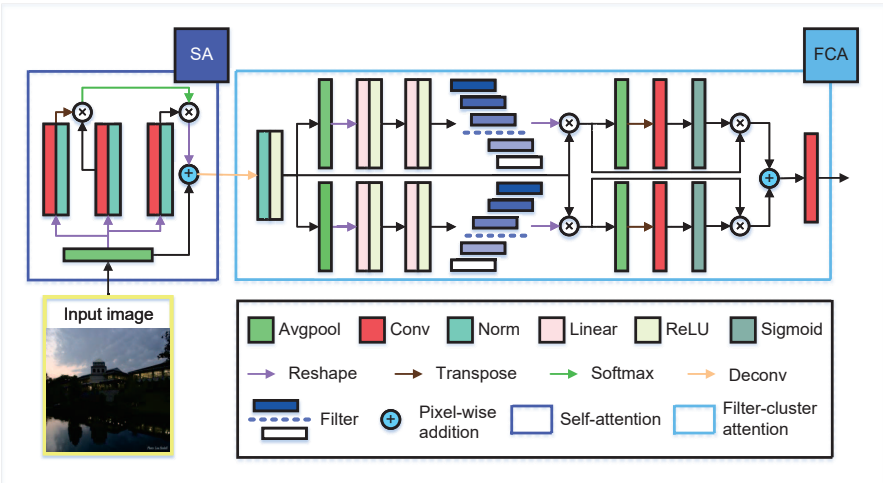

Fig. 3 The two-stage attention involved in the network. The first stage is self-attention (SA) and the second stage is filter-cluster attention (FCA)

cross region, and F_{R} represents the ReLU function. According to the saliency vector, we set the filtering rate for each group to determine the threshold θ^i, and filter each channel with the saliency p^i. The filtering process for p^i is

$$p^i = \begin{cases} p^i, & ||p^i|| > \theta^i, \\ 0, & ||p^i|| \leq \theta^i. \end{cases} \tag{7}$$

The filtered sequence is multiplied by the preprocessed features to remove the less salient channels in the features, and the other channels are amplified to varying degrees to obtain a group of saliency annotation results. Because each channel in the group has strong saliency, we will calculate the dependency in the group more accurately. We refer to the design of efficient channel attention (ECA) (Wang QL et al., 2020) to calculate the within-group dependency. The dependency learning process of the ith group is

$$y = F_{\mathrm{avg}}\left(p^i f\right), \tag{8}$$

$$d^i = F_{\mathrm{Sig}}\left(\sum_{j=1}^{n} w_{j,i} y_{j,i}\right), \tag{9}$$

where y represents the sequence of annotation results compressed by the avgpool operation. Different from saliency calculation, the real number sequence is not further compressed here. After the convolutional layer and the sigmoid function F_{Sig} with nonlinear mechanism, each channel has its corresponding dependency d^i, which is multiplied by the annotation results and combined with other group results to obtain the final attention output.

2.3 Dense connection pyramid

Most of the low-light images have low pixel values, leading to the overall darkness of the image. After network enhancement, the color may appear less visible and less vibrant. To avoid excessive noise, we choose to use simple inversion to obtain the inversion image with high brightness. However, the inverted image is still uneven in color, and the direct linear superposition will destroy the detailed information of the enhanced image. Therefore, we try to extract features from the inverse image to compensate for the less vividly colored regions in the enhancement results. Based on the idea of dense connection and UNet, we design DenCP, which exploits the similarity and difference of multi-scale features to extract the auxiliary brightness information needed by the network. After each iteration, we combine the features of the DenCP output to use the rich color information provided by the inversion image.

Fig. 4 shows the specific structure of DenCP designed in this paper. We design four basic layers in DenCP, which perform feature fusion, up-sampling, down-sampling, and module aggregation, separately. Through these four basic layers and convolution layers, we build the coding module, aggregation module, and decoding module. Considering the parameter complexity in traditional dense connection networks (Huang G et al., 2017), DenCP densely aggregates only the previous layer features at the last layer, while the input of the middle layer comes from the sum of the features of the upper two layers, which

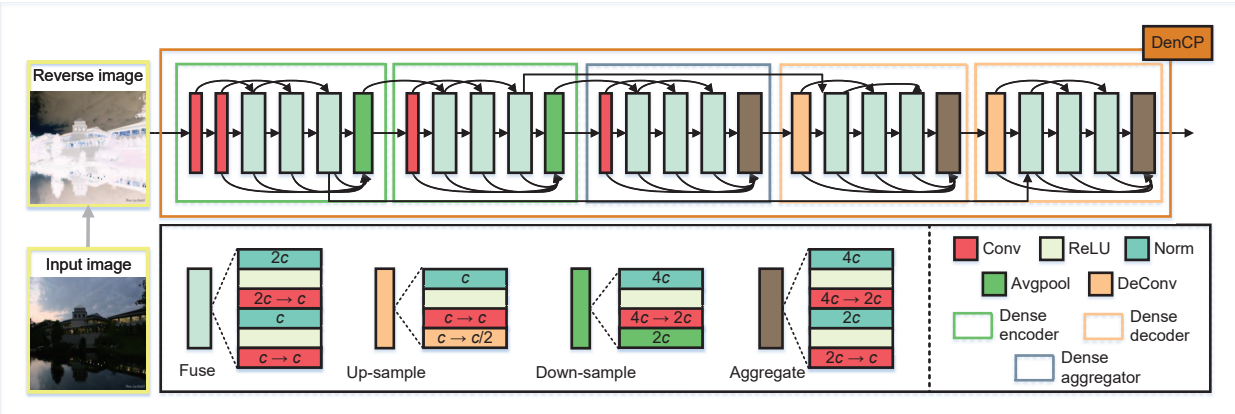

Fig. 4 The specific structure of DenCP. The inverse color image of the original image is used as the input to the pyramid. Above is the DenCP outline framework, and below is the detailed structure of each component and the change of the feature channels in the four basic layers

are expressed as

$$f_i = \begin{cases} F_{\mathrm{f}}(f_{i-2}, f_{i-1}), & i \neq n, \\ F_{\mathrm{a}}(f_0, f_1, \ldots, f_{i-1}), & i = n, \end{cases} \tag{10}$$

where n represents the total number of layers, $f_0, f_1, \ldots, f_{i-1}$ represent the features in each layer of the module, F_{f} represents the feature fusion operation in the middle layer, and F_{a} represents the module aggregation operation for all layer features at the last layer. In the encoding module and decoding module, we design the down-sampling layer and up-sampling layer at the final and initial parts of the module, respectively. The calculation of features in two layers is shown as

$$f_{\mathrm{sc}} = F_{\mathrm{sc}}\left(wF_{\mathrm{R}}\left(f_i\right)_{\mathrm{N}} + b\right), \tag{11}$$

where f_{sc} represents the feature after scale change, F_{sc} represents the scale change function, which is the avgpool in the down-sampling layer and deconvolution in the up-sampling layer. w and b are the weight and bias learned in the convolution layer respectively, and $(\cdot)_{\mathrm{N}}$ represents the normalization operation. DenCP scales down and enlarges the features twice, and performs two skip connections to combine the global information of the network to supplement the information loss in the scale change.

2.4 Loss function

Due to the huge structural differences between low-light and normal images, the proposed network first adopts a supervised learning method for structural similarity. In addition, the enhanced images may have under-enhanced dark or over-enhanced bright spots. The network also focuses on the overall image continuity during the training process. The total loss L_{sum} of the network consists of two kinds of losses: structural similarity (SSIM) loss L_{ssim} and total adjacent variation (TAV) loss L_{tav}. The expression of the total loss is shown in Eq. (12) (see the bottom of this page). I^{e} represents the enhanced image, I^{gt} represents the ground truth, and μ is the hyperparameter used to balance the two losses, which is set to 0.000 01 in this experiment.

2.4.1 SSIM loss

SSIM can measure the structural similarity between the network enhancement results and ground truth during the training process, to motivate the network to generate enhancement results that are more similar to the target. The value range of SSIM is from 0 to 1. The higher the value, the higher the similarity between the two images.

The expression of SSIM is shown in Eq. (13) (see the bottom of this page). $\mu(\cdot)$ is the average function representing image brightness, $\sigma(\cdot)$ is the variance function representing image contrast, and c_1 and c_2 are constants that suppress the denominator to 0. Following the recommended parameters, we set c_1 and c_2 to 0.0001 and 0.0009, respectively. In this study, we define the structural similarity loss as $L_{\mathrm{ssim}}\left(I^{\mathrm{e}}, I^{\mathrm{gt}}\right) = 1 - \mathrm{SSIM}\left(I^{\mathrm{e}}, I^{\mathrm{gt}}\right)$.

2.4.2 TAV loss

The low-light image may have regions with large color-change differences, which leads the network to produce enhanced images with insufficient continuity. To make the generated image smoother, the network needs to pay attention to the color difference in the low-light scene during the training process. Inspired by total variation (TV), we design TAV with a larger calculating domain, paying more attention to the variation between image points and adjacent pixels. The calculating domain of TAV is expanded from the right point and the lower point to the four neighborhoods, and the gradient of the adjacent region is calculated more comprehensively, thereby motivating the network to generate more continuous images in the training process. The TAV loss is shown in Eq. (14) (see the bottom of this page).

$$L_{\mathrm{sum}}\left(I^{\mathrm{e}}, I^{\mathrm{gt}}\right) = L_{\mathrm{ssim}}\left(I^{\mathrm{e}}, I^{\mathrm{gt}}\right) + \mu L_{\mathrm{tav}}\left(I^{\mathrm{e}}\right). \tag{12}$$

$$\mathrm{SSIM}\left(I^{\mathrm{e}}, I^{\mathrm{gt}}\right) = \frac{2\mu\left(I^{\mathrm{e}}\right)\mu\left(I^{\mathrm{gt}}\right) + c_1}{\mu^2\left(I^{\mathrm{e}}\right) + \mu^2\left(I^{\mathrm{gt}}\right) + c_1} \frac{2\sigma\left(I^{\mathrm{e}}\right)\sigma\left(I^{\mathrm{gt}}\right) + c_2}{\sigma^2\left(I^{\mathrm{e}}\right) + \sigma^2\left(I^{\mathrm{gt}}\right) + c_2}. \tag{13}$$

$$L_{\mathrm{tav}}\left(I^{\mathrm{e}}\right) = \frac{1}{3}\sum_{c=1}^{3}\sum_{i=1}^{H}\sum_{j=1}^{W}\sqrt[4]{\left|I^{\mathrm{e}}_{i,j} - I^{\mathrm{e}}_{i-1,j}\right|\left|I^{\mathrm{e}}_{i,j} - I^{\mathrm{e}}_{i,j-1}\right|\left|I^{\mathrm{e}}_{i,j} - I^{\mathrm{e}}_{i+1,j}\right|\left|I^{\mathrm{e}}_{i,j} - I^{\mathrm{e}}_{i,j+1}\right|}. \tag{14}$$

3 Experiments

In this section, we present the implementation details, perform visual and quantitative comparison with other advanced methods on reference and non-reference datasets, and perform ablation studies to demonstrate the optimal number of iterations and the effectiveness of each component. Finally, we provide a specific analysis of the underperforming indexes.

3.1 Implementation details

Training details: the experiments are conducted on a server with the Ubuntu 18.04 operating system, TITAN RTX graphics card with 24 GB graphics memory, Intel Xeo® Silver 4210 CPU, and 376 GB memory. The training datasets are obtained from the LOL dataset (Wei et al., 2018) and the brightening dataset (Wei et al., 2018). The LOL dataset contains 500 pairs of images and the brightening dataset contains 1000 pairs of images, for a total of 1500 pairs of images, of which 1350 pairs are used for training and 150 pairs for reference testing. We build the network model in the pytorch framework with a total training batch size of 100. The learning rate is set to 0.0001, which is automatically adjusted with the Adam optimizer (Kingma and Ba, 2014) during the training process.

We perform subjective and quantitative comparisons with seven advanced methods, including three traditional methods JED (Ren XT et al., 2018), RRM (Li MD et al., 2018), and EFF (Ying et al., 2017), and four deep learning methods DLN (Wang LW et al., 2020), KinD (Zhang YH et al., 2019), MIRNet (Zamir et al., 2020), and R-Net (Wei et al., 2018). We compare the proposed method with advanced methods on the LOL (Wei et al., 2018), MIT-Adobe FiveK (Bychkovsky et al., 2011), NPE (Wang SH et al., 2013), DICM (Lee et al., 2012), ExDark (Loh and Chan, 2019), and LIME (Xu YD et al., 2021) datasets. For the reference datasets LOL and MIT-Adobe FiveK, we use the peak signal-to-noise ratio (PSNR), structural similarity (SSIM), feature similarity (FSIM) (Zhang L et al., 2011), lightness order error (LOE) (Wang SH et al., 2013), and mean-squared error (MSE) indexes for evaluation. For the non-reference datasets NPE, DICM, ExDark, and LIME, we use the BRISQUE (Mittal et al., 2013) and ENTROPY indexes for evaluation.

3.2 Evaluation on reference datasets

Fig. 5 shows the performance of the different methods on the LOL and MIT-Adobe FiveK reference datasets. The results of JED and RRM still have more dark regions, and the image details are blurred in the regions where black pixels are concentrated. EFF shows the best enhancement effect among the traditional methods, but still suffers from under-enhancement. The enhancement effect of DLN and KinD is outstanding and the detailed information is well preserved, but the examples (d) of DLN and MIRNet show exposure at the sky background, the examples (b) of MIRNet have a large amount of patchy noise in the background, and the portrait sculpture also suffers from detail loss due to

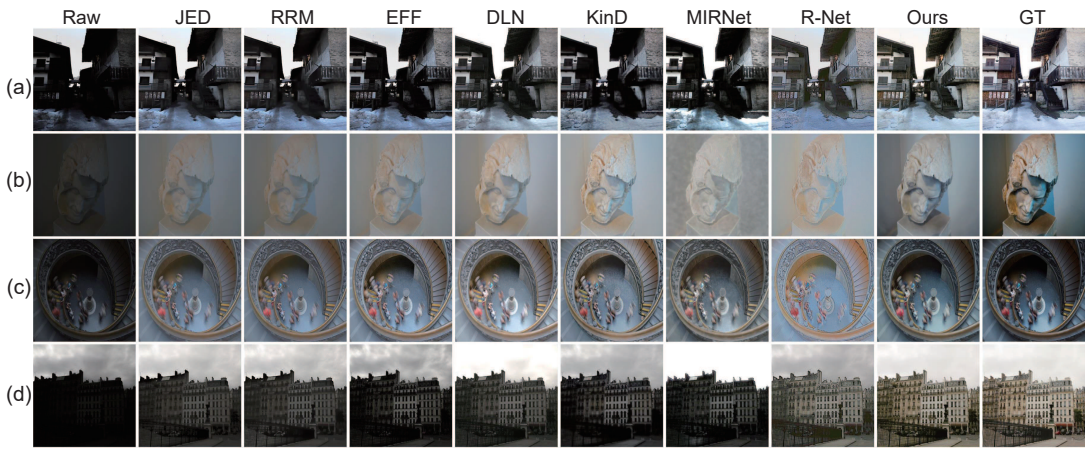

Fig. 5 Comparison on the reference datasets LOL and MIT-Adobe FiveK, with (a) and (d) from the LOL dataset and (b) and (c) from the MIT-Adobe FiveK dataset

over-smoothing. R-Net produces the most vivid colors among all methods, but the image is not clear, and most regions show a foggy hazy effect. Compared with the other methods, our method generates a good light continuity result, effectively improving the brightness and clarity of the image, and no undesirable results of overexposure and noise. Specifically, in the (b) and (c) examples, our method has the closest light, shadow, and color to the ground truth.

In Fig. 6, we show a comparison of the detailed enlargement on the LOL dataset. DLN and R-Net restore some contour to the sculpture, but the images generated are still blurred. The other methods fail to restore the sculpture to its original appearance. Although our method has certain color differences compared to the ground truth, the sculpture itself has the clearest contour among all the results. Tables 1 and 2 compare the indexes of all methods on the LOL and MIT-Adobe FiveK datasets, respectively. The higher the values of PSNR, SSIM, and FSIM, the higher the similarity between the color, structure, and features of the enhancement results and the ground truth. The lower the values of LOE and MSE, the closer the brightness and pixels of the enhancement result to the ground truth. As can be seen in Table 1, RRM achieves suboptimal LOE, and DLN achieves suboptimal performance in terms of other indexes. DLN has exposure in the brightness concentration region in the original images, thus affecting its LOE index. Our method achieves optimal performance in all indexes, which indicates that our method has a higher similarity to the ground truth in LOL.

Table 2 shows the index performance regarding the MIT-Adobe FiveK dataset, where optimal and suboptimal index performances are achieved by EFF, DLN, and our method. The color, structure, and pixels of our method are more similar to the ground truth, while our method is not as good as DLN in terms of feature and brightness approximations.

Table 1 The mean values of PSNR, SSIM, FSIM, LOE, and MSE on the LOL dataset

Method	PSNR	SSIM	FSIM	LOE	MSE
JED	17.662	0.7918	0.8935	70.340	0.006 87
RRM	17.756	0.7872	0.8852	<u>65.590</u>	0.006 66
EFF	17.800	0.7934	0.9096	78.032	0.007 26
DLN	<u>18.492</u>	<u>0.8729</u>	<u>0.9236</u>	78.283	<u>0.005 43</u>
KinD	17.703	0.7961	0.9029	94.753	0.006 81
MIRNet	16.096	0.7618	0.8767	110.860	0.008 95
R-Net	17.321	0.7895	0.8619	223.740	0.005 45
Ours	**23.414**	**0.9337**	**0.9674**	**49.136**	**0.002 46**

The optimal values are in bold and the suboptimal values are underlined

Table 2 The mean values of PSNR, SSIM, FSIM, LOE, and MSE on the MIT-Adobe FiveK dataset

Method	PSNR	SSIM	FSIM	LOE	MSE
JED	17.990	0.7879	0.8936	108.130	0.004 69
RRM	17.999	0.7989	0.8941	108.080	0.004 68
EFF	<u>18.322</u>	<u>0.8463</u>	0.9233	108.580	<u>0.004 60</u>
DLN	17.819	0.8357	**0.9556**	**45.731**	0.005 97
KinD	16.294	0.8148	0.9195	193.400	0.007 09
MIRNet	17.073	0.7924	0.9190	126.600	0.007 09
R-Net	12.596	0.7126	0.8069	571.960	0.015 38
Ours	**19.166**	**0.8523**	<u>0.9428</u>	<u>58.752</u>	**0.004 56**

The optimal values are in bold and the suboptimal values are underlined

3.3 Evaluation on no-reference datasets

We also conduct experiments on NPE, DICM, ExDark, and LIME. The results of all methods are shown in Fig. 7. JED, RRM, and R-Net have improved image brightness and contrast to a certain extent, but make examples (b) and (d) produce severe

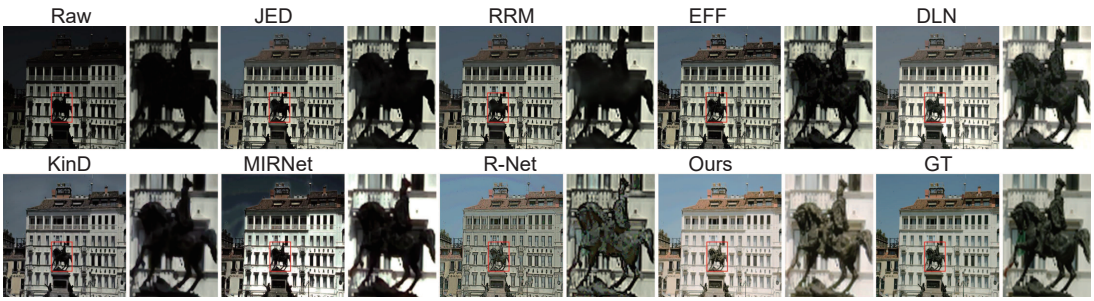

Fig. 6 Detail enlargement comparison on the LOL dataset

color deviation, which affects image visibility. In addition, R-Net spreads the local color in examples (c) to the full image, dramatically deepening the blurring of the image. EFF, DLN, and KinD enhance the visibility of the low-light image better, but still produce color deviations in examples (b) and (d), and DLN over-enhances regions where brightness is concentrated, which also produces the exposure effect. For MIRNet and our method concerning examples (b) and (d), image brightness and contrast are improved without introducing other colors, but our method has better brightness and sharpness. In addition, our method removes the green color deviation in underwater images (c), and makes examples (a) more vibrant, showing that it can produce visually pleasing results in most cases.

Fig. 8 shows the enlarged details of all the results on the DICM dataset. The results of JED and MIRNet are so dark that the figure outline is almost unrecognizable. Other methods improve the clarity of the figure somewhat, but overall is still dark. R-Net and our method improve the brightness of the figure a lot, but the results of R-Net have many blurred patches, affecting the figure clarity. Fig. 9 shows the enlarged details on the ExDark dataset. Different from other methods, EFF, MIRNet, and our method are not affected by the green color, while the brightness of the results of EFF and our method are more prominent. Fig. 10 shows the enlarged details on the NPE dataset. The results of DLN and MIRNet are exposed in light, while in the resulting image of other methods, the periphery of the lampshade is affected

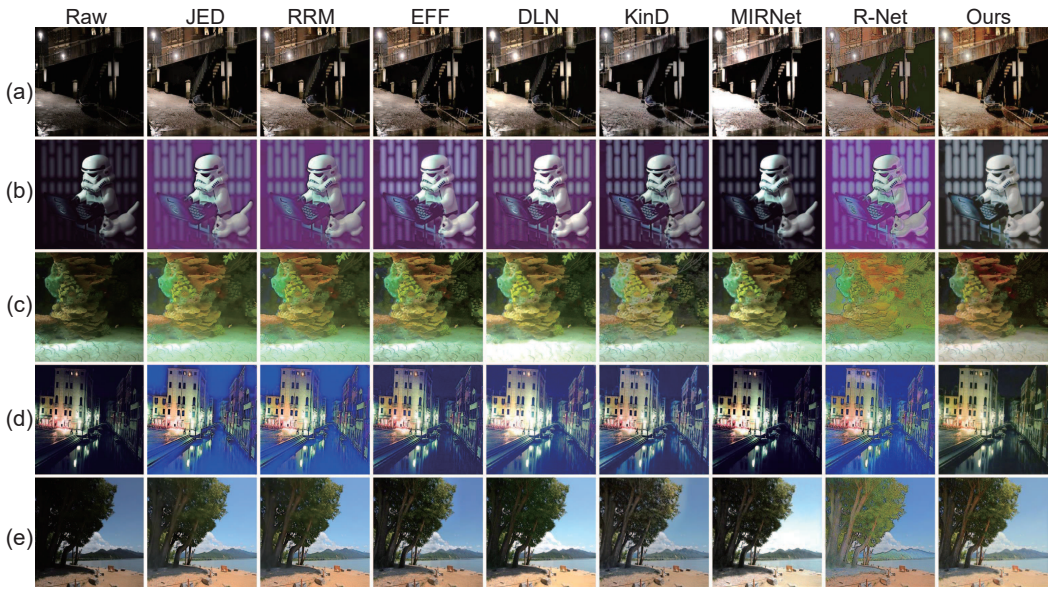

Fig. 7 Comparison on no-reference datasets DICM, ExDark, and LIME, with (a) and (d) from the ExDark dataset, (b) from the LIME dataset, and (c) and (e) from the DICM dataset

Fig. 8 Detailed enlargement comparison on the DICM dataset

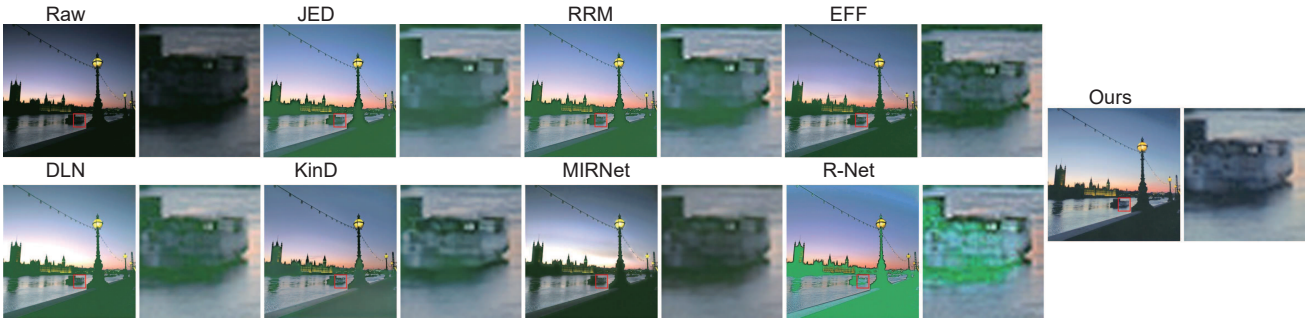

Fig. 9 Detailed enlargement comparison on the ExDark dataset

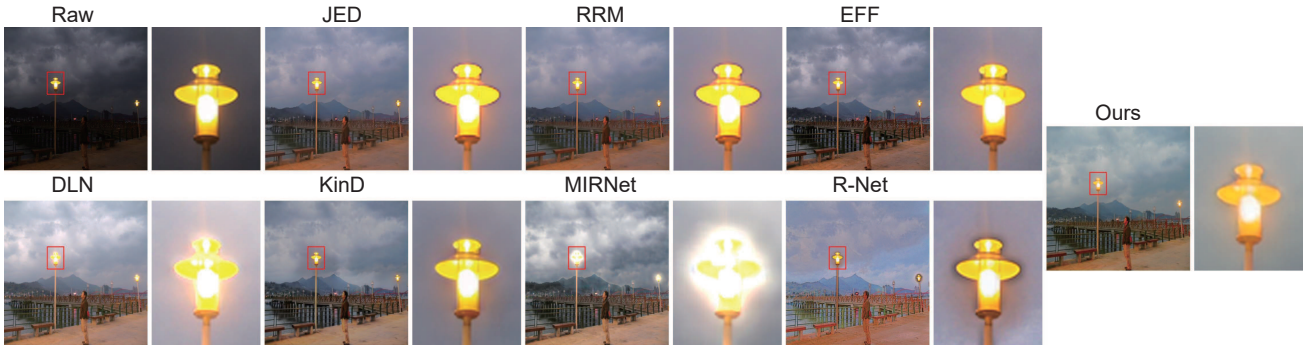

Fig. 10 Detailed enlargement comparison on the NPE dataset

by certain light source outward expansion. Although this phenomenon is not obvious in our method, the color inside the lampshade is not vivid enough.

Table 3 shows the indexes on the four no-reference datasets. The lower the value of BRISQUE, the better the spatial quality of the image. The higher the value of ENTROPY, the better the performance in content richness. Regarding the BRISQUE index, our method achieves optimal performance on NPE, DICM, and ExDark, and suboptimal performance on LIME, and DLN achieves suboptimal and optimal performance on the first three datasets and the LIME dataset, respectively. This indicates that DLN and our method have good visual performance on all four datasets. However, on the LIME dataset, our method is not as prominent as DLN in terms of color vividness, which consequently affects the BRISQUE index. Regarding the ENTROPY index, our method achieves the best performance on the NPE, ExDark, and LIME datasets. This indicates that our method produces content-rich results on all three datasets, while it does not obtain a good ENTROPY index on the DICM dataset. We will analyze this in Section 3.5.

3.4 Ablation studies

We conduct ablation studies on the iteration number and the components of the network, to demonstrate the optimal iteration number and component effectiveness. We train different networks with 1–6 iterations. Fig. 11 shows the comparison of the enhancement effect with different iteration numbers. At iteration 1, the enhancement result is dull and less visible than other results, while at iterations 2, 4, and 5, the network generates other colors, affecting the original color of the image to some extent, and the result at iteration 2 produces exposure in the sky background. The network generates similar results of brightness and color at iterations 3 and 6, achieving a better visual performance. Fig. 12 shows the changes in the different indexes with the increase of the iteration number. It can be seen that the network achieves the best values of PSNR, SSIM, FSIM, and MSE at iteration 3, while at iteration 1 it achieves the best values of the LOE index. In general, we decide to use the network with iteration 3.

We also perform an ablation study to analyze the various components in the network. The

Table 3 The mean values of BRISQUE and ENTROPY on the NPE, DICM, ExDark, and LIME datasets

Method	BRISQUE				ENTROPY			
	NPE	DICM	ExDark	LIME	NPE	DICM	ExDark	LIME
JED	29.908	21.710	29.923	28.883	6.9446	6.8285	7.0567	6.0133
RRM	27.177	28.044	26.948	30.114	7.1836	7.0032	7.1836	6.5648
EFF	16.210	20.034	20.801	18.250	7.1629	6.7907	7.1038	6.0380
DLN	18.988	16.205	17.010	**15.199**	6.5529	6.8133	7.0411	6.1022
KinD	21.364	23.816	22.809	23.988	7.1237	**7.1751**	7.2066	6.1590
MIRNet	19.988	16.945	20.628	21.956	6.7444	6.9382	6.8717	5.9754
R-Net	25.742	32.819	23.955	27.837	7.0555	7.0288	7.3379	6.2415
Ours	**18.680**	**13.854**	**15.543**	17.485	**7.3419**	6.9630	**7.3710**	**6.5861**

The optimal values are in bold and the suboptimal values are underlined

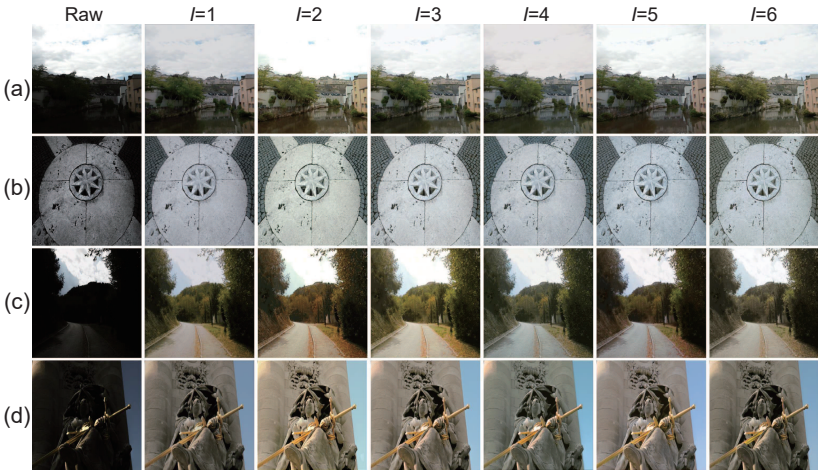

Fig. 11 Comparison of networks with different numbers of iterations, with (a)–(d) from the LOL dataset

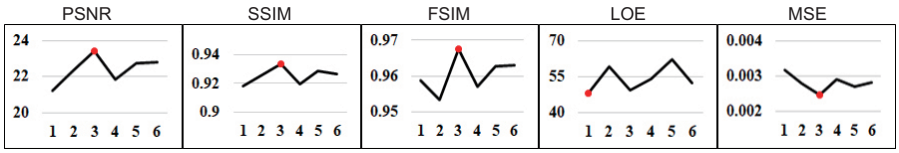

Fig. 12 The index changes of the network under different iteration numbers. The best index is marked with red dots (References to color refer to the online version of this figure)

configurations of each experiment are as follows: (1) -w/o TAV: without TAV loss, (2) -w/o FCA: without FCA, (3) -w/o DCU: without difference concern unit, (4) -w/o GRU: without GRU, (5) -w/o IRU: without iterative residual unit, (6) -w/o DenCP: without DenCP, and (7) ours: complete component network.

In Fig. 13, -w/o TAV is not smooth due to failing to focus adjacent pixels, and the overall color continuity of the tree and tower in examples (a) is not strong. The results of -w/o FCA and -w/o DenCP both show that the color is not bright enough. As can be seen in examples (a), -w/o DenCP even causes the tower spire to be affected by other colors. Due to lack of attention to differences in dark regions, the details of -w/o DCU results are not bright enough and colors are not vivid. -w/o GRU causes the trees in examples (a) to be affected by the red color, and the plants in examples (c) are also duller.

-w/o IRU shows the phenomenon of over-enhancement, and multiple exposures can be seen in the image. The network with complete components has the best performance in brightness and contrast. The face in examples (b) is more ruddy and closer to the ground truth, indicating that the network with complete components produces more visible results.

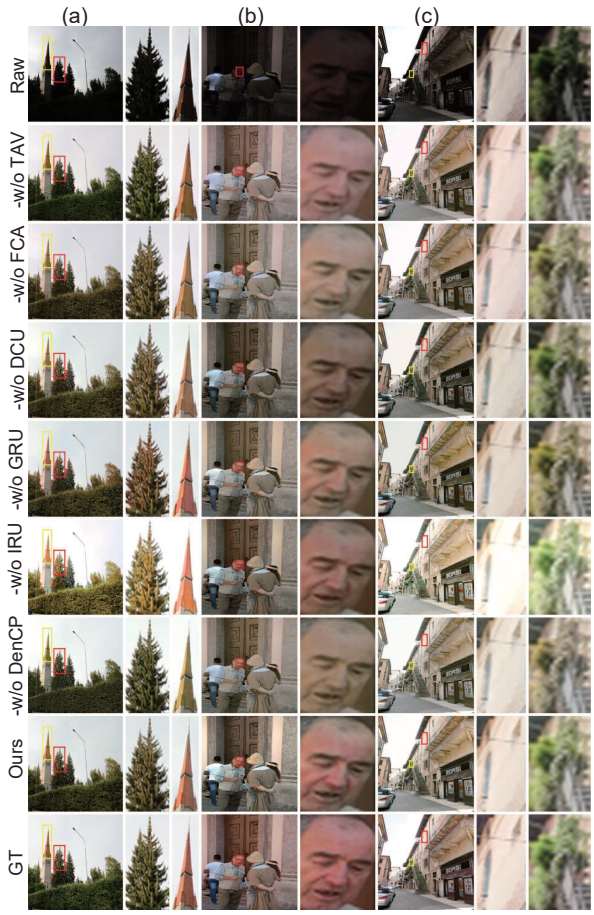

Fig. 13 Comparison of different configuration networks, with (a)–(c) from the LOL dataset

Table 4 shows the index performance and runtime of different configuration networks. -w/o FCA has the worst performance on PSNR and SSIM, indicating that FCA creates good improvement on the color and structure of the network. -w/o IRU has the worst performance on FSIM and LOE, indicating that IRU has played a certain role in the extraction of feature details and brightness information. -w/o GRU has the worst performance on MSE, indicating that GRU provides an effective adaptive memory function for network pixel information. The network of complete components achieves the best performance in each index, which represents the advantage of each component to the network. In terms of running efficiency, -w/o FCA and -w/o DenCP achieve the highest and second highest rates, respectively. In addition, we compare the efficiency of DenCP with those of the classical models CNN, UNet, and ResNet. In the same environment, the single image runtime is 0.4266 s for CNN, 0.4333 s for UNet, 0.4674 s for ResNet, and 0.4933 s for DenCP. It can be seen that DenCP is not as efficient as these classical models. Although FCA and DenCP make the network operate less efficiently, we still choose to keep both modules owing to their advantages.

3.5 Weakness analysis

When the low-light image displays color deviation in the underwater environment, our method highlights the over-correction ability to enhance image brightness while removing the influence of other colors. This process suppresses the overly colorful image, thus leading to the unsatisfactory ENTROPY index of our method on the DICM dataset, but at the same time, the image visibility is effectively improved. Therefore, our method achieves the optimal BRISQUE value. In Fig. 14, we show the comparison between the raw image and all results for the hard mix image. The compared methods achieve the ideal ENTROPY index on DICM. It can be seen that our method produces the least color blocks in the hard mix image, which represents lack of content richness, but also proves that our method has good ability to correct chaotic colors and makes the overall image colors more balanced.

Table 4 The index performance and runtime of networks with different configurations on the LOL dataset

Configuration	PSNR	SSIM	FSIM	LOE	MSE	Runtime (s)
-w/o TAV	22.060	0.9224	0.9593	50.481	0.003 05	0.4933
-w/o FCA	21.876	0.9198	0.9601	<u>50.235</u>	0.003 14	**0.3600**
-w/o DCU	22.692	0.9268	<u>0.9642</u>	51.209	0.002 83	0.4766
-w/o GRU	22.479	0.9252	0.9613	57.304	0.003 21	0.4537
-w/o IRU	22.162	0.9212	0.9589	59.866	0.002 90	0.4600
-w/o DenCP	<u>22.900</u>	<u>0.9301</u>	0.9638	55.861	<u>0.002 62</u>	<u>0.3933</u>
Ours	**23.414**	**0.9674**	**0.9674**	**49.136**	**0.002 46**	0.4933

The optimal values are in bold and the suboptimal values are underlined

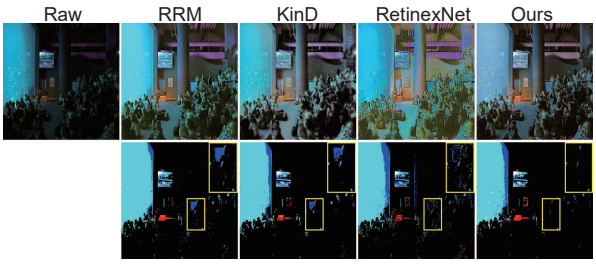

Fig. 14 The comparison between different method results and raw images for the hard mix image. The upper part is the original image and the enhancement results of each method, and the lower part is the corresponding hard mix detail enlargement. More color blocks in the figure mean richer information in the results

4 Conclusions

In this paper, we propose a recurrent network based on FCA and design a two-stage attention mechanism consisting of self-attention and FCA, which focuses on low-light regions and important channels. The main body of the network consists of three units: difference concern, gate recurrent, and iterative residual. The difference concern unit network learns the difference between the learning results and the FCA output. GRU memorizes the global information of multiple stages with lower computational cost. The iterative residual unit extracts in-depth features by shared parameters. In addition, we design a DenCP to extract features from the inverse image of a low-light image to enhance the color vividness of each iteration output.

Benefiting from the multiple iterations of network learning, the proposed network has outstanding quality improvement for low-light images. Through extensive experiments on public datasets, the proposed method shows good brightness and visibility enhancement in all methods. The optimal iteration number and the effectiveness of each component are validated through ablation studies. Considering that the proposed method will suppress the original bright colors when facing the low-light scene with uneven color, we will apply FCA in the improved network structure in future work to enrich the saturation and visual expression of the image. The new network structure will be applied to low-light video enhancement or other challenging scenes, providing an excellent quality improvement solution for information in adverse environments.

Contributors

Jinjiang LI designed the research. Zhixiong HUANG designed the software. Zhen HUA and Linwei FAN processed the data. Zhixiong HUANG drafted the paper. Jinjiang LI helped organize the paper. Zhixiong HUANG and Linwei FAN revised and finalized the paper.

Compliance with ethics guidelines

Zhixiong HUANG, Jinjiang LI, Zhen HUA, and Linwei FAN declare that they have no conflict of interest.

Data availability

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

References

Abdullah-Al-Wadud M, Kabir H, Dewan MAA, et al., 2007. A dynamic histogram equalization for image contrast enhancement. *IEEE Trans Consum Electron*, 53(2):593-600. https://doi.org/10.1109/TCE.2007.381734

Aradi S, 2022. Survey of deep reinforcement learning for motion planning of autonomous vehicles. *IEEE Trans Intell Transp Syst*, 23(2):740-759. https://doi.org/10.1109/TITS.2020.3024655

Bychkovsky V, Paris S, Chan E, et al., 2011. Learning photographic global tonal adjustment with a database of input/output image pairs. IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.97-104. https://doi.org/10.1109/CVPR.2011.5995413

Celik T, Tjahjadi T, 2011. Contextual and variational contrast enhancement. *IEEE Trans Image Process*, 20(12):3431-3441. https://doi.org/10.1109/TIP.2011.2157513

Chen BH, Wu YL, Shi LF, 2019. A fast image contrast enhancement algorithm using entropy-preserving mapping prior. *IEEE Trans Circ Syst Video Technol*, 29(1):38-49. https://doi.org/10.1109/TCSVT.2017.2773461

Cheng HD, Shi XJ, 2004. A simple and effective histogram equalization approach to image enhancement. *Dig Signal Process*, 14(2):158-170. https://doi.org/10.1016/j.dsp.2003.07.002

Cho K, van Merriënboer B, Gulcehre C, et al., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. Proc Conf on Empirical Methods in Natural Language Processing, p.1724-1734. https://doi.org/10.3115/v1/D14-1179

Guo CL, Li CY, Guo JC, et al., 2020. Zero-reference deep curve estimation for low-light image enhancement. IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.1780-1789. https://doi.org/10.1109/CVPR42600.2020.00185

Guo XJ, Li Y, Ling HB, 2017. LIME: low-light image enhancement via illumination map estimation. *IEEE Trans Image Process*, 26(2):982-993. https://doi.org/10.1109/TIP.2016.2639450

Hao SJ, Han X, Guo YR, et al., 2020. Low-light image enhancement with semi-decoupled decomposition. *IEEE Trans Multim*, 22(12):3025-3038. https://doi.org/10.1109/TMM.2020.2969790

Hochreiter S, Schmidhuber J, 1997. Long short-term memory. *Neur Comput*, 9(8):1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang G, Liu Z, van der Maaten L, et al., 2017. Densely connected convolutional networks. IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.2261-2269. https://doi.org/10.1109/CVPR.2017.243

Huang ZX, Li JJ, Hua Z, et al., 2022. Underwater image enhancement via adaptive group attention-based multiscale cascade transformer. *IEEE Trans Instrum Meas*, 71:5015618. https://doi.org/10.1109/TIM.2022.3189630

Jiang YF, Gong XY, Liu D, et al., 2021. EnlightenGAN: deep light enhancement without paired supervision. *IEEE Trans Image Process*, 30:2340-2349. https://doi.org/10.1109/TIP.2021.3051462

Jung E, Yang N, Cremers D, 2020. Multi-frame GAN: image enhancement for stereo visual odometry in low light. Proc 3rd Annual Conf on Robot Learning, p.651-660.

Kingma DP, Ba J, 2014. Adam: a method for stochastic optimization. Proc 3rd Int Conf on Learning Representations.

Lee C, Lee C, Kim CS, 2012. Contrast enhancement based on layered difference representation. 19th IEEE Int Conf on Image Processing, p.965-968. https://doi.org/10.1109/ICIP.2012.6467022

Lee C, Lee C, Kim CS, 2013. Contrast enhancement based on layered difference representation of 2D histograms. *IEEE Trans Image Process*, 22(12):5372-5384. https://doi.org/10.1109/TIP.2013.2284059

Li CL, Tang SQ, Yan JW, et al., 2020. Low-light image enhancement based on quasi-symmetric correction functions by fusion. *Symmetry*, 12(9):1561. https://doi.org/10.3390/sym12091561

Li JJ, Feng XM, Hua Z, 2021. Low-light image enhancement via progressive-recursive network. *IEEE Trans Circ Syst Video Technol*, 31(11):4227-4240. https://doi.org/10.1109/TCSVT.2021.3049940

Li L, Wang RG, Wang WM, et al., 2015. A low-light image enhancement method for both denoising and contrast enlarging. IEEE Int Conf on Image Processing, p.3730-3734. https://doi.org/10.1109/ICIP.2015.7351501

Li MD, Liu JY, Yang WH, et al., 2018. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Trans Image Process*, 27(6):2828-2841. https://doi.org/10.1109/TIP.2018.2810539

Li PL, Liang JL, Zhang MH, 2021. A degradation model for simultaneous brightness and sharpness enhancement of low-light image. *Signal Process*, 189:108298. https://doi.org/10.1016/j.sigpro.2021.108298

Lim KL, Jiang XD, Yi CY, 2020. Deep clustering with variational autoencoder. *IEEE Signal Process Lett*, 27:231-235. https://doi.org/10.1109/LSP.2020.2965328

Liu L, Ouyang WL, Wang XG, et al., 2020. Deep learning for generic object detection: a survey. *Int J Comput Vis*, 128(2):261-318. https://doi.org/10.1007/s11263-019-01247-4

Liu RS, Ma L, Zhang JA, et al., 2021. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.10556-10565. https://doi.org/10.1109/CVPR46437.2021.01042

Liu YJ, Wang ZN, Zeng Y, et al., 2021. PD-GAN: perceptual-details GAN for extremely noisy low light image enhancement. IEEE Int Conf on Acoustics, Speech and Signal Processing, p.1840-1844. https://doi.org/10.1109/ICASSP39728.2021.9413433

Loh YP, Chan CS, 2019. Getting to know low-light images with the exclusively dark dataset. *Comput Vis Image Underst*, 178:30-42. https://doi.org/10.1016/j.cviu.2018.10.010

Lore KG, Akintayo A, Sarkar S, 2017. LLNet: a deep autoencoder approach to natural low-light image enhancement. *Patt Recogn*, 61:650-662. https://doi.org/10.1016/j.patcog.2016.06.008

Lv FF, Li Y, Lu F, 2021. Attention guided low-light image enhancement with a large scale low-light simulation dataset. *Int J Comput Vis*, 129(7):2175-2193. https://doi.org/10.1007/s11263-021-01466-8

Ma L, Liu RS, Zhang JA, et al., 2022. Learning deep context-sensitive decomposition for low-light image enhancement. *IEEE Trans Neur Netw Learn Syst*, 33(10):5666-5680. https://doi.org/10.1109/TNNLS.2021.3071245

Mittal A, Soundararajan R, Bovik AC, 2013. Making a "completely blind" image quality analyzer. *IEEE Signal Process Lett*, 20(3):209-212. https://doi.org/10.1109/LSP.2012.2227726

Peng T, Su LL, Zhang RH, et al., 2020. A new safe lane-change trajectory model and collision avoidance control method for automatic driving vehicles. *Expert Syst Appl*, 141:112953. https://doi.org/10.1016/j.eswa.2019.112953

Ren WQ, Liu SF, Ma L, et al., 2019. Low-light image enhancement via a deep hybrid network. *IEEE Trans Image Process*, 28(9):4364-4375. https://doi.org/10.1109/TIP.2019.2910412

Ren XT, Li MD, Cheng WH, et al., 2018. Joint enhancement and denoising method via sequential decomposition. IEEE Int Symp on Circuits and Systems, p.1-5. https://doi.org/10.1109/ISCAS.2018.8351427

Shiau YH, Chen PY, Yang HY, et al., 2015. A low-cost hardware architecture for illumination adjustment in real-time applications. *IEEE Trans Intell Transp Syst*, 16(2):934-946. https://doi.org/10.1109/TITS.2014.2347701

Singh H, Kumar A, Balyan LK, et al., 2017. A novel optimally gamma corrected intensity span maximization approach for dark image enhancement. 22nd Int Conf on Digital Signal Processing, p.1-5. https://doi.org/10.1109/ICDSP.2017.8096035

Singh N, Bhandari AK, 2021. Principal component analysis-based low-light image enhancement using reflection model. *IEEE Trans Instrum Meas*, 70:70:5012710. https://doi.org/10.1109/TIM.2021.3096266

Wang LW, Liu ZS, Siu WC, et al., 2020. Lightening network for low-light image enhancement. *IEEE Trans Image Process*, 29:7984-7996. https://doi.org/10.1109/TIP.2020.3008396

Wang QL, Wu BG, Zhu PF, et al., 2020. ECA-Net: efficient channel attention for deep convolutional neural networks. IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.11531-11539. https://doi.org/10.1109/CVPR42600.2020.01155

Wang SH, Luo G, 2018. Naturalness preserved image enhancement using a priori multi-layer lightness statistics. *IEEE Trans Image Process*, 27(2):938-948. https://doi.org/10.1109/TIP.2017.2771449

Wang SH, Zheng J, Hu HM, et al., 2013. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Trans Image Process*, 22(9):3538-3548. https://doi.org/10.1109/TIP.2013.2261309

Wang W, Sun N, Ng MK, 2019. A variational gamma correction model for image contrast enhancement. *Inv Probl Imag*, 13(3):461-478. https://doi.org/10.3934/ipi.2019023

Wang YF, Liu HM, Fu ZW, 2019. Low-light image enhancement via the absorption light scattering model. *IEEE Trans Image Process*, 28(11):5679-5690. https://doi.org/10.1109/TIP.2019.2922106

Wei C, Wang WJ, Yang WH, et al., 2018. Deep retinex decomposition for low-light enhancement. British Machine Vision Conf, Article 155.

Wu XM, Liu XH, Hiramatsu K, et al., 2017. Contrast-accumulated histogram equalization for image enhancement. IEEE Int Conf on Image Processing, p.3190-3194. https://doi.org/10.1109/ICIP.2017.8296871

Xie EZ, Ding J, Wang WH, et al., 2021. DetCo: unsupervised contrastive learning for object detection. IEEE/CVF Int Conf on Computer Vision, p.8372-8381. https://doi.org/10.1109/ICCV48922.2021.00828

Xu CR, Peng ZZ, Hu XZ, et al., 2020. FPGA-based low-visibility enhancement accelerator for video sequence by adaptive histogram equalization with dynamic clip-threshold. *IEEE Trans Circ Syst I Regul Papers*, 67(11):3954-3964. https://doi.org/10.1109/TCSI.2020.3010634

Xu K, Yang X, Yin BC, et al., 2020. Learning to restore low-light images via decomposition-and-enhancement. IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.2278-2287. https://doi.org/10.1109/CVPR42600.2020.00235

Xu YD, Yang C, Sun BB, et al., 2021. A novel multi-scale fusion framework for detail-preserving low-light image enhancement. *Inform Sci*, 548:378-397. https://doi.org/10.1016/j.ins.2020.09.066

Yan XA, Liu Y, Jia MP, 2020a. Multiscale cascading deep belief network for fault identification of rotating machinery under various working conditions. *Knowl-Based Syst*, 193:105484. https://doi.org/10.1016/j.knosys.2020.105484

Yan XA, Liu Y, Xu YD, et al., 2020b. Multistep forecasting for diurnal wind speed based on hybrid deep learning model with improved singular spectrum decomposition. *Energy Conv Manag*, 225:113456. https://doi.org/10.1016/j.enconman.2020.113456

Yang B, Cao XL, Yuen C, et al., 2021. Offloading optimization in edge computing for deep-learning-enabled target tracking by Internet of UAVs. *IEEE Int Things J*, 8(12):9878-9893. https://doi.org/10.1109/JIOT.2020.3016694

Yang WH, Wang WJ, Huang HF, et al., 2021a. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Trans Image Process*, 30:2072-2086. https://doi.org/10.1109/TIP.2021.3050850

Yang WH, Wang SQ, Fang YM, et al., 2021b. Band representation-based semi-supervised low-light image enhancement: bridging the gap between signal fidelity and perceptual quality. *IEEE Trans Image Process*, 30:3461-3473. https://doi.org/10.1109/TIP.2021.3062184

Ying ZQ, Li G, Ren YR, et al., 2017. A new low-light image enhancement algorithm using camera response model. IEEE Int Conf on Computer Vision Workshops, p.3015-3022. https://doi.org/10.1109/ICCVW.2017.356

Yu SY, Zhu H, 2019. Low-illumination image enhancement algorithm based on a physical lighting model. *IEEE Trans Circ Syst Video Technol*, 29(1):28-37. https://doi.org/10.1109/TCSVT.2017.2763180

Zamir SW, Arora A, Khan S, et al., 2020. Learning enriched features for real image restoration and enhancement. Proc 16th European Conf on Computer Vision, p.492-511. https://doi.org/10.1007/978-3-030-58595-2_30

Zhang L, Zhang L, Mou XQ, et al., 2011. FSIM: a feature similarity index for image quality assessment. *IEEE Trans Image Process*, 20(8):2378-2386. https://doi.org/10.1109/TIP.2011.2109730

Zhang TL, Li JJ, Fan H, 2022. Progressive edge-sensing dynamic scene deblurring. *Comput Visual Media*, 8(3):495-508. https://doi.org/10.1007/s41095-021-0246-4

Zhang YH, Zhang JW, Guo XJ, 2019. Kindling the darkness: a practical low-light image enhancer. Proc 27th ACM Int Conf on Multimedia, p.1632-1640. https://doi.org/10.1145/3343031.3350926