RESEARCH ARTICLE

Efficient adaptive feature aggregation network for low-light image enhancement

Canlin Li1*, Pengcheng Gao1, Jinhua Liu2, Shun Song1, Lihua Bi1

1 School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou, China, 2 Shanghai Film Academy, Shanghai University, Shanghai, China

* lcl_zju@aliyun.com

Abstract

Existing learning-based methods for low-light image enhancement contain a large number of redundant features, the enhanced images lack detail and have strong noises. Some methods try to combine the pyramid structure to learn features from coarse to fine, but the inconsistency of the pyramid structure leads to luminance, color and texture deviations in the enhanced images. In addition, these methods are usually computationally complex and require high computational resource requirements. In this paper, we propose an efficient adaptive feature aggregation network (EAANet) for low-light image enhancement. Our model adopts a pyramid structure and includes multiple multi-scale feature aggregation block (MFAB) and one adaptive feature aggregation block (AFAB). MFAB is proposed to be embedded into each layer of the pyramid structure to fully extract features and reduce redundant features, while the AFAB is proposed for overcome the inconsistency of the pyramid structure. EAANet is very lightweight, with low device requirements and a quick running time. We conducted an extensive comparison with some state-of-the-art methods in terms of PSNR, SSIM, parameters, computations and running time on LOL and MIT5K datasets, and the experiments show that the proposed method has significant advantages in terms of comprehensive performance. The proposed method reconstructs images with richer color and texture, and the noises is effectively suppressed.

1 Introduction

In low-light conditions, the captured image suffers from texture and color deviations and contains a lot of noises, which affects the visual perception. Low-light environment also increase the difficulty of high-level vision tasks, such as image detection [1, 2] and visual tracking [3, 4]. Luminance variation is one of the main factors that affect the performance of such task. One solution is to use a capture device with high international standards organization (ISO) to obtain high quality images, however, this method is insufficient and expensive for image improvement. Therefore, image enhancement techniques are often used to enhance low-light images. Researchers have proposed a large number of low-light image enhancement methods, which can be classified into three categories: Histogram equalization-based, Retinex-based, and learning-based.
Histogram equalization is a classical method that enhances the contrast of the whole image by extending the dynamic range of the pixels. The method is simple and convenient, but the enhanced image lacks details. Some improved methods based on HE [5, 6] further improve the visual perception of the image, but still cannot meet the practical demands.

Some works [7–11] combine Retinex theory to decompose the image into a illumination map and a reflection map. The enhanced image is obtained by fusing the processed illumination map and reflection map. However, it is difficult to estimate the illumination map of the image accurately by these methods, because the image has different degrees of degradation in low-light environment. The enhanced images usually have low contrast, blurred details and contain strong noises.

Many learning-based methods for low-light image enhancement have been proposed. For example, some end-to-end methods [12–14] design complex structures to integrate features, but the generated images lack texture and contain noises. Some methods [15, 16] enhance low-light images with the help of physically interpretable Retinex theory. These methods achieve better results than traditional methods but the suppression of noises is inadequate. There are also unsupervised methods [17, 18] that alleviate the need for paired datasets and improve the generalization ability of the methods, but these methods are difficult to train and the generated images have color deviations and noises.

Existing learning-based methods are inefficient in the utilization of features, and the reconstructed images lack detail and have noises. Some methods try to combine pyramid structures to learn features, using a coarse-to-fine strategy to capture fine image details, but the inconsistency of pyramid structures leads to insufficient communication between deep semantic information and shallow fine-grained information. Meanwhile, these methods are complex and require a lot of device resources, which are difficult to be applied to mobile devices or to assist high-level vision tasks. To address these problems, we design an efficient adaptive feature aggregation network (EAANet) for low-light image enhancement. Specifically, we propose the multi-scale feature aggregation block (MFAB) and the adaptive feature aggregation block (AFAB). MFAB contains information aggregation block (IAB) and Dual Attention Block (DAB), during the inference phase, the asymmetric convolution of IAB is fused into a standard convolution, which can enhance the feature representation without increasing the number of parameters, DAB refines the features from both spatial and channel aspects to reduce the feature redundancy. In order to improve the feature scale inconsistency of the pyramid structure, AFAB adaptively selects features from the pyramid structure to establish a connection between the semantic information at the deep layer and the fine-grained information at the shallow layer. In addition, both blocks are lightweight. Through the collaboration in the two blocks, the proposed method reconstructs images with good luminance, color and texture.

The main contributions of this paper are as follows:

1. We propose an efficient adaptive feature aggregation network (EAANet) for low-light image enhancement. Extensive experiments show that the proposed method can effectively reconstruct the color and texture of the image and suppress the image noises.

2. We propose a multi-scale feature aggregation block (MFAB), which improves feature representation by asymmetric convolution and reduces redundant features using a dual attention mechanism. MFAB is a lightweight block and it’s very efficient. In the inference stage, the asymmetric convolution is fused into a standard convolution to further reduce the parameters and computations.

3. In order to address the inconsistency of the pyramid structure, we propose an adaptive feature aggregation block (AFAB). The block enhances the interaction between the shallow
fine-grained information and the deep semantic information of the pyramid to improve the texture and color of the image.

The remainder of this paper is organized as follows. We introduce the related work on low-light image enhancement methods in Section 2. Our proposed EAANet is described in detail in Section 3. We give the experiments and the comparison of results in Section 4. Section 5 is the ablation study of EAANet. Section 6 is the conclusion.

2 Related work

In this section, we introduce the work related to low-light image enhancement methods, including histogram equalization-based methods, Retinex-based methods, and learning-based methods.

2.1 Histogram equalization-based methods

Inspired by Histogram equalization (HE), a large number of improved methods based on HE have been proposed. For example, Pizer et al. [5] proposed an contrast limited adaptive histogram equalization (CLAHE), which restricts each sub-histogram. CLAHE improves the image contrast and suppresses noises, but it first divides the image into multiple block regions, which result in block artifacts in the reconstructed image. To solve the block artifact phenomenon, Srinivasan et al. [19] proposed a histogram equalization using local region stretching, which divides the image into multiple areas according to the gray level, and then histogram equalization is performed for each gray level area. Ibrahim et al. [6] proposed dynamic histogram equalization with preserved luminance, which has a good effect on some locally underexposed images by calculating the global average luminance. However, this method is greatly limited when the image luminance is extremely dark. Although histogram equalization-based methods and their improvement methods are simple and convenient, they usually suffer from image distortion and noises problems.

2.2 Retinex-based methods

Retinex is a theory that imitates human vision, and it is widely used in image enhancement tasks. In recent years, a large number of Retinex-based methods have been proposed. For example, SSR [7], MSR [8] and MSRCR [9], they are pioneering methods. Wang et al. [10] proposed a method to improve the contrast and naturalness of images, but the method still produces severe noises. Cai et al. [11] proposed joint intrinsic-extrinsic prior model (JieP) to protect structural integrity through shape prior. Guo et al. [20] proposed LIME to accelerate the optimization of the illumination map, which greatly reduces the computations. Xu et al. [21] proposed a new detail-preserving variational enhancement model that uses the l1 prior to constraining the illumination map. Retinex-based methods are usually better for image color fidelity, but these methods are poorly adaptable and also prone to artifacts, strong noises, and overexposure. In addition, it is difficult to estimate the illumination map of extremely low-light images with heavy noises.

2.3 Learning-based methods

In recent years, learning-based methods achieved great progress. LLNet [12] was proposed as an early end-to-end method for low-light image enhancement. Wang et al. [13] proposed a globally aware network (GLAD), which incorporates a pyramid structure to extract image features. MIRNet [14] combines a residual structure and an attention mechanism to reconstruct the correct luminance and color. Some methods combine Retinex theory, for example. Wei et al. [15] proposed RetinexNet, which designed a decomposition net and an enhancement net
to decompose and process images. Inspired by Ref [15], Zhang et al. [16] proposed KinD, which designed a light adjustment block to flexibly adjust the luminance. Wang et al. [22] proposed DeepUPE, which combines Retinex and bilateral filtering. Liu et al. [23] proposed RAUNA, which designed a decomposition network with explicit and implicit priors, and taken into account both global and local brightness in the augmented network. Some methods try to learn enhancement results from unpaired data, such as Jiang et al. [24] proposed a degradation-to-refinement generation network (DRGN). Li et al. [17] proposed zero-reference depth curve estimation (Zero-DCE++) for low light image enhancement. Risheng et al. [18] proposed Retinex-inspired unrolling with cooperative prior architecture search (RUAS), which first searches the network architecture through the training set, and then finds the low-light prior structure from the compact search space. These methods further liberate the restriction on the dataset, but they are not stable in training. Compared with traditional methods, learning-based method combined with Retinex theory can better estimate the illumination and reflection maps and have better enhancement performance, but the noises suppression of these methods is still inadequate. Compared with depth Retinex-based methods, the end-to-end methods are more effective in suppressing noises. The end-to-end methods usually have low complexity and have great potential for assisting high-level vision tasks and applying on mobile devices. Recently, Li et al. [25] proposed a luminance-aware pyramid network (LPNet). LPNet is a lightweight end-to-end method, and it has strong competitiveness. LPNet designed a multi-scale contrast feature block for feature extraction and a luminance loss for guiding the model to learn luminance information. However, LPNet uses too many split operations, resulting in insufficient information exchange between channels and the reconstructed image lacks texture. In addition, due to the inconsistency of pyramid structure, the shallow features and deep features of the pyramid structure cannot be fully utilized. This causes color deviations in the reconstructed images by LPNet.

In our work, we propose an efficient adaptive feature aggregation network (EAA Net). EAA Net contains two important blocks, MFAB and AFAB. MFAB utilizes asymmetric convolution and dual attention mechanisms to fully extract features and capture critical features. Compared with LPNet, MFAB is more efficient to extract features. AFAB overcomes the inconsistency of pyramid structure and enhances the communication between shallow fine-grained information and deep semantic information. We describe the proposed EAA Net in Section 3.

3 Method

In this section, we describe the proposed EAA Net. EAA Net use a pyramid structure. Pyramid structure is a common multi-scale structure, and it is one of the main methods for extracting multi-scale information. It is widely used in traditional methods [26], convolutional neural network (CNN) based methods [27–30] and Transformer methods [31]. Pyramid structure has a large perceptual field and can provide rich feature information. EAA Net has three layers, each layer contains multiple MFABs for extracting features, and the output of each layer is incorporated into the shallow layer to guide the shallow layer to learn more refined features. In addition, the output of each layer is also fed into AFAB to establish connections between the features of different layers. As shown in Fig 1, the input image is first mapped as high-dimen-
sional feature through a convolutional layer and then successively downsampled to obtain features at different scales, the operation can be defined as Eqs (1) and (2):

\[ I_1 = \text{Conv3}(I_{input}) \]  

\[ I_n = F_{\text{down}}(I_{n-1}), n = 2, 3 \]
where \(I_{\text{input}}\) is the low-light image and \(I_n\) denotes the corresponding input of the \(n\)th layer. \(F_{\text{down}}\) denotes a down-sampling. \(Conv\) denotes a \(3 \times 3\) convolutional layer. Some critical features will gradually lose with the increase of network depth. Inspired by Ref [32, 33], we construct an aggregation block (AB) to address the problem of losing critical features with the increasing depth of the network. Each AB contains \(n\) multi-scale feature aggregation blocks (MFABs), and the output of each MFAB is concatenated to make full use of the features. We set \(n\) to be 4. To avoid generating a large number of parameters, the concatenated features are first passed through a \(1 \times 1\) convolution layer to adjust the number of channels, and then the features are reconstructed by a \(3 \times 3\) convolutional layer. At the bottom layer, the features first pass through an AB, then the outputs of AB are up-sampled and concatenated with the middle layer features obtained from the previous down-sampling operation. Afterwards, the number of channels is adjusted by \(1 \times 1\) convolutional layer and passed again through the AB. By repeating this operation, the output of the AB for each layer can be obtained, this operation can be defined as Eqs (3)–(5):

\[
F_{\text{AB}} = Conv^3(Conv^1(||MFAB_{\text{out1}}, MFAB_{\text{out2}}, ..., MFAB_{\text{outn}}||))
\]  

\[
I_{\text{out1}} = F_{\text{AB}}(I_1)
\]  

\[
I_{\text{outn}} = F_{\text{AB}}(Conv^1(||F_{\text{up}}(I_{\text{outn-1}}), I_n||)), \quad n = 1, 2
\]

where \(I_{\text{outn}}\) is the output of the \(n\)th layer. |||| is concatenation operation. \(Conv^1\) denotes a \(1 \times 1\) convolutional layer. \(F_{\text{AB}}\) denotes the corresponding operation of AB. \(F_{\text{up}}\) denotes a up-sampling. \(MFAB_{\text{outn}}\) denotes the output of the \(i\)-th MFAB. We describe MFAB in detail in section 3.1. The output of the three layers is fed to AFAB to strengthen the connection between the shallow fine-grained information and the deep semantic information. This operation can be defined as Eq (6):

\[
I_{\text{output}} = Conv^3(Conv^3(F_{\text{AFAB}}(I_{\text{out1}}, I_{\text{out2}}, I_{\text{out3}})))
\]

where \(I_{\text{output}}\) is enhanced image. \(F_{\text{AFAB}}\) denotes the corresponding operation of AFAB. We describe MFAB in detail in section 3.2. The output of AFAB is passed through two \(3 \times 3\) convolutional layers to obtain the final enhanced image. Our method is end-to-end, so that the
whole operation can be defined as Eq (7):

\[ I_{output} = F_{EAANet}(I_{input}) \] (7)

where \( F_{EAANet} \) is the corresponding operation of EAANet. Each down-sampling contains a 2-step \( 3 \times 3 \) convolution and a 1-step \( 3 \times 3 \) convolution. Each up-sampling contains bilinear interpolation and a 1-step \( 3 \times 3 \) convolution. Subsequently, we describe each block of the model and the loss function in details.

3.1 Multi-scale feature aggregation block (MFAB)

Some methods designed novel structures to extract abundant features. For example, MSRN [32] uses convolutional layers with different kernels. LPNet [25] splits features along the channel dimension and then passes through different numbers of convolutional layers, respectively. However, the feature extraction capability of these methods is inefficient. Convolutional layer with large kernel generates a large number of parameters, and the split operation can hinder the information interaction between channels and degrade the feature representation. We proposed the MFAB to better extract and utilize the features. As shown in Fig 2, the feature from the previous layer is divided into two branches by two \( 1 \times 1 \) convolutional layers, respectively. Each branch has only half the number of channels. One branch passes through a \( 3 \times 3 \) convolutional layer, and then waits for a concatenation with another branch. The other branch can
be roughly divided into two parts: An information aggregation block (IAB) and a double attention block (DAB).

Learning-based methods are often necessary to design complex structures to obtain rich feature information. ACNet proposes an efficient method for capturing feature information, it try to enhances the standard convolution kernel using asymmetric convolution. Inspired by ACNet \[34\], we designed an asymmetric convolution block (ACB), which contains 2 asymmetric convolutions and 1 standard convolution. The outputs of the asymmetric convolution and the standard convolution are combined by an element summing operation and Leaky Relu activation is added. For the IAB, the features pass through three ACB. During the training phase, asymmetric convolution block contains asymmetric convolution and standard convolution, the expression can be defined as Eq (8):

\[
P_{ACB} = \delta((U \ast K^{3 \times 1} + U \ast K^{3 \times 3} + U \ast K^{1 \times 3}))
\]

where \(U\) is the feature of the previous layer. \(K\) is convolution kernel. \(\delta\) is Leaky Relu activation.

During the inference phase, we transform the asymmetric convolution to standard convolution to reduce the number of parameters of the network while maintaining performance, and the operation can be defined as Eq (9):

\[
U \ast K^{3 \times 1} + U \ast K^{3 \times 3} + U \ast K^{1 \times 3} = U \ast (K^{3 \times 1} \oplus K^{3 \times 3} \oplus K^{1 \times 3})
\]

where \(\oplus\) is the element-wise addition of the kernel parameters. Receptive field has a great influence on feature extraction, and the receptive fields of the feature output by each ACB are different. We concatenate the output of each asymmetric convolution block to fuse the feature information of different receptive fields.

Low-light images have a lot of noises and the extracted features are often redundant. Attention mechanisms are used for high-level vision tasks such as image classification, image recognition, etc. It can promote models to focus more on valuable features and fade out useless features. Several methods \[25, 35, 36\] have demonstrated that the attention mechanism helps low-level visual models to learn the correct texture and suppress noises. Therefore, we try to use the attention mechanism to reduce feature redundancy. Specifically, we designed parallel spatial and channel attention mechanisms. Color degradation is a challenge for low-light image enhancement. Inspired by Ref \[37\], we added standard deviation (STD) based on global average pooling to the channel attention branch, as shown in Fig 3. The standard deviation is calculated as the degree of dispersion of individual pixels from the mean value of the image. By adding standard deviation, the model can be promoted to pay more attention to edge information, so the color of images can be improved. For the standard deviation, we define \(X = [x_1, x_2, \ldots, x_c]\) as input, whose shape is \(C \times H \times W\), and the expression can be defined as Eq (10):

\[
\varepsilon_c = \sqrt{\frac{1}{HW} \sum_{i,j \in x_c} (x_i^c - \frac{1}{HW} \sum_{i,j \in x_c} x_i^c)^2},
\]

where \(\varepsilon_c\) is the standard deviation of the c-th element of the output. STD can improves the color of images. In order to reduce the computational complexity, existing channel attention mechanisms usually perform channel dimensionality reduction, and this operation lead to the result that the model cannot effectively capture dependencies between channels and lose some important contextual information. Inspired by Ref \[38\], we use 1-dimensional convolution to avoid dimensional reduction. The output of the global average pool is first element-wise summed with \(\varepsilon_o\) and then a one-dimensional convolution is used to avoid channel reduction, and the one-dimensional convolution allows sufficient information interaction between channels. The features are normalized by the Sigmoid activation function and finally the original
features are rescaled by the element multiplication operation. In the spatial attention branch, we do average pooling and maximum pooling for the spatial dimensions, respectively. The output of both pooling ways are concatenated and then the valuable features are learned from the spatial dimension by convolutional layers. Finally, sigmod activation and element multiplication operations are performed to rescale the features. The outputs of spatial attention and channel attention are concatenated and pass through a convolution layer to adjust the channels. Our dual attention block can be defined as Eqs (11)–(13):

$$F_{ch} = U \odot \sigma(Conv1d(GAP(U) + STD))$$  \hspace{1cm} (11)

$$F_{sp} = U \odot \sigma(Conv7(GAP(U) + GMP(U)))$$  \hspace{1cm} (12)

$$F_{DAB} = Conv1(||F_{ch}, F_{sp}||)$$  \hspace{1cm} (13)

where $F_{ch}$ and $F_{sp}$ is channel and spatial attention. $F_{DAB}$ is our dual attention block. $Conv7$ denotes a $7 \times 7$ convolutional layer. $GAP$ is global average pooling. $GMP$ is global max pooling. STD is the standard deviation. Finally, the two branches are concatenated and a residual structure is added to stabilize the gradient propagation. In the spatial attention branch, the convolution kernel size is set as 7. In the channel attention branch, the convolution kernel of the 1-dimensional convolution is set as 3. Due to the combination of asymmetric convolution and 1D convolution, MFAB is very lightweight and also has excellent performance to efficiently extract valuable features and fade out redundant features. The reconstructed images have vivid colors and rich textures.

3.2 Adaptive feature aggregation block (AFAB)

Existing methods often use a pyramid structure to exploit multi-scale features. The pyramid structure incorporates deep semantic information into the shallow layer through
concatenation and element-wise summation operation to guide the model to learn finer features. But in practice, the inconsistency of the pyramid structure hinders the information interaction between shallow fine-grained features and deep semantic information. Liu et al. [39] proposed adaptive spatial feature fusion (ASFF) to address the effect of feature scale inconsistency on single-shot detectors. ASFF assigns weights to features to suppress feature conflicts between different scales, thus improving the scale invariance of features. Yi et al. [40] proposed an adaptive feature selection module (AFSM) for fusing neighboring scale features to improve the image dehazing effect. However, both ASFF and AFSM reduce the channel dimension so as to reduce the number of parameters, which leads to the underutilization of features. AFSM incorporates a channel attention mechanism, but the module only adaptively selects features at neighboring scales, ignoring the association of features at non-neighboring scales. We designed an AFAB, it receives features from each layer of the pyramid structure and adaptively rescales the features at each layer by the multi-branch attention structure. In addition, AFAB combines multi-branch attention structure with 1D convolution to enhance the channel information interaction and fully utilize the valuable information. As shown in Fig 4, $L_1$ is $I_{out_1}$, $L_2$ and $L_3$ are obtained by up-sampling the outputs of $I_{out_2}$ and $I_{out_3}$. The inputs of the three branches are fused by element-wise summation. Then the fused features generate global information by global average pooling. Conv1d means 1-dimensional convolution. The softmax activation values of the three branches are multiplied by the original features on each branch and then fused by element-wise summation operation. The operation of AFAB can be defined as Eqs (14) and (15):

$$S_N = L_N \odot \sigma(Conv1d(GAP(L_1 + L_2 + L_3))), N = 1, 2, 3$$  \hspace{1cm} (14)

$$F_{AFAB} = S_1 + S_2 + S_3$$  \hspace{1cm} (15)

where $S_N$ denotes the output of the three branches of AFAB. $\odot$ denotes element-wise multiplication operation. The AFAB can efficiently establish the dependencies between depth-layers.
semantic information and shallow fine-grained information, making full use of features at different scales. In addition, the block is lightweight and only adds a small number of parameters. The pseudo code of our proposed method is shown in Algorithm 1.

**Algorithm 1:** Efficient Adaptive Feature Aggregation Network for Low-light Image Enhancement

**Input:** Low light \( f^k_i \), Normal light sample set \( g^N_k \).

**Output:** Enhanced image \( f^k_{\text{output}} \).

**Initialization:** Learning rate \( \eta \), batch size \( m \), parameter \( \theta \).

1. for \( k = 1, 2, 3 \) do
   2. \( I_1 \) is extracted by Eq (1);
   3. Extract the \( I_n \) by Eq (2), \( n = 2, 3 \);
   4. \( F_{ACB} \) is calculated by Eq (8);
   5. \( \varepsilon_c \) is calculated by Eq (10), and \( F_{DAB} \) is calculated by Eq (13);
   6. Extract the \( I_{\text{out}} \) by Eq (4);
   7. Extract the \( I_{\text{out}} \) by Eq (5), \( n = 1, 2 \);
   8. \( F_{AFAB} \) is calculated by Eq (15);
   9. Calculate by Eq (6) to update enhanced image \( f^k_{\text{output}} \);
   10. Update parameter \( \theta \) with adam optimizer;
   11. Update Learning rate \( \eta \);
12. end
13. Asymmetric convolution kernels are fused by Eq (9);
14. Obtain final enhanced image \( f^k_{\text{output}} \).

### 3.3 Loss function

We designed a loss function, which consists of two parts: Content loss and perceptual loss.

**Content Loss:** L2 loss tends to produce overly smooth images. The images generated by L1 loss cannot effectively suppress noises. SSIM loss measures the similarity of the output to ground-truth by brightness, contrast, and structure. The images generated by SSIM loss contain less noises and are of good quality. Therefore, we use SSIM loss as content loss. Content loss can be defined as Eqs (16) and (17):

\[
\text{Loss}_c = 1 - \text{SSIM}(x, y)
\]

\[
\text{SSIM}(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}
\]

where \( \text{Loss}_c \) is content loss. \( x, y \) are the generated image and the target image respectively. \( \mu_x, \mu_y \) are the average values of \( x \) and \( y \) respectively. \( \sigma_x^2, \sigma_y^2 \) are the variances of \( x \) and \( y \) respectively. \( \sigma_{xy} \) is the covariance of \( x \) and \( y \). \( C_1, C_2 \) are constants to avoid dividing by zero.

**Perceptual loss:** Johnson et al. [41] proposed a perceptual loss function to optimize visual effects, which is based on the VGG model. We introduce this loss function to improve the image visual quality. We use a pretrained VGG19 model, and the model weights are obtained by training on the ImageNet dataset [42]. Perceptual loss can be defined as Eq (18):

\[
\text{Loss}_p = \frac{1}{C H W_j} \| \phi_i(x) - \phi_i(y) \|
\]

where \( \text{Loss}_p \) is the total loss. \( C, H, W_j \) are the number of channels, height, and width, respectively. \( \phi_i \) is the output of layer \( m \) of the VGG19 network. When we use features from different layers of VGG, the optimization effect on the model will be different, we use the output
of the last convolution of the features layer to calculate the mapping differences. The two losses have a weight ratio of 1.

4 Experiments

4.1 Datasets

In experiments, we use the datasets LOL [15] and MIT5K [43]. The LOL dataset is the first paired dataset collected in real scenes, containing 485 train pairs and 15 test pairs. We follow the previous experience and extract 35 images from the training set for evaluating. The MIT5K dataset is often used by researchers in low-light image enhancement field. In order to make a fair comparison with other methods, we use the Expert C-retouched image as the ground truth. We use 4500 pairs for training and the remaining 500 pairs for evaluating and testing. For the MIT5K dataset, we only consider image enhancement under RGB. The dataset can be made by using Lightroom software, setting the length as 500. The image sizes from the LOL dataset and the MIT5K dataset are 600 × 400 and 500 × 333, respectively.

4.2 Implementation details

We train our network on NVidia GTX 1080ti GPU by using the pytorch framework. The network is trained for 150 epochs. The optimizer uses the Adam optimizer, and the Adam optimizer parameters use the framework default values. We use the cosine annealing strategy and set the initial learning rate as $2 \times 10^{-4}$. The learning rate threshold is set as $2 \times 10^{-6}$. To enhance the robustness of the network, we use random rotation, mirroring for data augmentation. We set the batch size as 16, the patch size as 96 × 96, and set the channel as 32.

4.3 Comparison with typical methods

We compared some recent work on the LOL dataset, including Xu’s method [21], RetinexNet [15], GLAD [13], KinD [16], LPNet [25], Zero-DCE++ [17], RUAS [18], RAUNA [23], DRGN [24]. For the KinD method, we chose an exposure of 5 as the benchmark. Similarly, on the MIT5K dataset, we compared with some recent work including Xu’s method, WhiteBox [44], DPE [45], DeepUPE [22], LPNet, Zero-DCE++, RUAS, RAUNA, DRGN. The code and pre-trained models for the above methods were provided by the original authors. For Zero-DCE++, RUAS and DRGN, since the training datasets are different, we retrain the model on the two datasets to do a fair comparison. In this subsection, we compare three aspects of quantitative comparison, efficiency comparison, and visual comparison.

4.3.1 Quantitative comparison. We use 2 metrics, PSNR and SSIM, for quantitative comparison. PSNR is a commonly used image objective evaluation index, the higher the PSNR value, the higher the image quality. SSIM measures the image quality in terms of luminance, contrast and structure, and the SSIM value is closer to 1, the higher the image quality. Tables 1 and 2 reflect the results of the quantitative comparison of the methods on the LOL and MIT5K datasets.

Table 1. Quantitative comparison on LOL dataset.

| Method     | Xu’s method | RetinexNet | GLAD | KinD | LPNet | Zero-DCE++ | RUAS | RAUNA | DRGN | Ours |
|------------|-------------|------------|------|------|-------|------------|------|-------|------|------|
| PSNR       | 18.52       | 16.77      | 19.72| 20.87| 21.46 | 14.71      | 16.40| 22.49 | 22.11| 22.68|
| SSIM       | 0.631       | 0.559      | 0.704| 0.802| 0.802 | 0.501      | 0.582| 0.765 | 0.821| 0.829|
| Param [M]  | -           | 1.23       | 0.93 | 8.49 | 0.15 | 0.01       | 0.003| 1.85  | 4.71 | 0.35 |
| FLOPs [G]  | -           | 6.79       | 4.37 | 7.44 | 0.77 | 0.24       | 0.06 | 28.41 | 25.66| 1.53 |
| Time [s]   | 2.8715      | 0.3139     | 0.2583| 0.3744| 0.0179| 0.0011     | 0.0041| 0.0950| 0.1372| 0.0182|

https://doi.org/10.1371/journal.pone.0272398.t001
datasets, where the values of PSNR and SSIM are the average of the test sets. We can see that the PSNR and SSIM of Xu’s method, Zero-DCE++, RUAS are not good on both datasets. RAUNA has good PSNR, but SSIM is not well. It can be clearly seen that our method achieved the best PSNR and SSIM on both datasets, which indicates that our method has the best quality of the images. Our method can effectively learn the mapping relationship between low luminance and normal luminance.

4.3.2 Efficiency comparison. Efficiency is one of the important metrics for evaluating models. Existing deep learning methods are usually computationally complex and have long-running time, which greatly limits their usage scenarios. Therefore, lightweight models are gaining more and more attention. We give the number of parameters (Param), the number of floating-point operations (FLOPs) and the running time of different methods for efficiency comparison, as shown in Table 1. FLOPs are commonly used to measure the computational complexity of an algorithm. We compute FLOPs with a patch size of 96×96. Xu’s method is a traditional low-light enhancement method, and its running time is measured on Intel i7–10700 CPU. Other methods are measured on NVidia GTX 1080ti GPU. We can see that RUAS and Zero-DCE++ are lightweight, but the value PSNR and SSIM are not good. Since RAUNA and DRGN are not end-to-end models, their running times are limited. Our method achieves good PSNR and SSIM with a small number of parameters and running time. This shows that our method is lightweight and has a good image enhancement capability.

4.3.3 Visual comparison. We show the visual comparison of some state-of-the-art methods on LOL and MIT5K datasets respectively. The visual comparison results on the LOL dataset are shown in Figs 5 and 6. We can see the luminance of the images reconstructed by Xu’s method are low. The images generated by GLAD and RetinexNet have a lot of noises. The images produced by KinD are too smooth and has some color distortion. LPNet and DRGN lack details. In addition, LPNet cannot handle the mirror area well. The enhanced images
based on Zero-DCE++ and RUAS have a lot of noises and severe color distortion. The images recovered by RAUNA have a small amount of noises. The images generated by our method have rich colors and details while suppressing noises. The visual comparison results on the MIT5K dataset are shown in Figs 7 and 8. It can be seen that the recovered images by White-Box have overexposure. The images reconstructed by DPE and LPNet have low luminance. There are noises and color deviation in the images reconstructed by DeepUPE, RAUNA and DRGN. The enhanced images based on Zero-DCE++ and RUAS have severe distortion. The images generated by our method have better visual perception.

The above visual comparison results reflect the advantages of our method. On two datasets, the enhanced images of our method are more realistic and natural.

5 Ablation study

In this section, we conducted an ablation study on LOL dataset. We will investigate MFAB, the number of MFAB, pyramid structure, and AFAB. The images used in Fig 8 are from the LOL dataset.

5.1 Effectiveness of multi-scale feature aggregation block

We investigate the effect of MFAB, as shown in Table 3. Case 4 is the proposed method EAA-Net. Case 3 removes MFAB and we can see that the value of PSNR and SSIM for the model
have decreased. As shown in Fig 10 (c), the visual results of the model without MFAB are obviously bad. It shows that MFAB has a positive impact on the model. MFAB contains IAB and DAB. We conduct a series of ablation studies to verify the effectiveness of each component.

5.1.1 Effectiveness of information aggregation block (IAB). As shown in Table 4, we use Case 3 as the benchmark data. To verify the effectiveness of IAB, we use three $3 \times 3$ convolutions instead of IAB. Each convolution is followed by an Leaky Relu activation. The results are shown in Case 1 of Table 4. Compared with Case 3, the PSNR and SSIM of Case 1 are significantly lower, which shows the effectiveness of IAB. In addition, we also conducted an ablation study to investigate the effect of ACB on the model, Case 2 use asymmetric convolution instead of standard convolution. We can clearly see that the PSNR and SSIM of Case 2 are higher than that of Case 1. This adequately illustrates the effectiveness of asymmetric convolution. Asymmetric convolution can extract richer features. In the inference phase, asymmetric convolution can be transformed into standard convolution to reduce the computations while maintaining the model performance. The PSNR and SSIM of Case 2 are lower than that of Case 3. This also proves the effectiveness of the aggregation structure. IAB fuses the features from different receptive fields to ensure the maximum utilization of features. We further compare with the feature extraction structure of some classical methods, as shown in Fig 9.

Table 3. Study on EAA Net. Pyramid, AFAB and MFAB are pyramid structure, adaptive feature aggregation block and multi-scale feature aggregation block, respectively.

| Case | Pyramid | AFAB | MFAB | Param [M] | FLOPs [G] | PSNR/SSIM |
|------|----------|------|------|-----------|----------|-----------|
| 1    | ✕        | ✕    | ✓    | 0.38      | 3.53     | 21.87/0.803 |
| 2    | ✓        | ✕    | ✓    | 0.34      | 1.43     | 22.31/0.814 |
| 3    | ✓        | ✓    | ✕    | 0.25      | 1.10     | 21.94/0.807 |
| 4    | ✓        | ✓    | ✓    | 0.35      | 1.53     | 22.68/0.829 |

https://doi.org/10.1371/journal.pone.0272398.t003

Table 4. Study on Information Aggregation Block (IAB). ACB, Aggregation are asymmetric convolutional block, aggregation structure respectively.

| Case | ACB | Aggregation | Param [M] | FLOPs [G] | PSNR/SSIM |
|------|-----|-------------|-----------|----------|-----------|
| 1    | ✕   | ✕           | 0.28      | 1.23     | 22.23/0.811 |
| 2    | ✓   | ✕           | 0.34      | 1.46     | 22.47/0.822 |
| 3    | ✓   | ✓           | 0.35      | 1.53     | 22.68/0.829 |

https://doi.org/10.1371/journal.pone.0272398.t004
The structure in Fig 9(a) is from LPNet. The structure in Fig 9(b) is from MSRN. We replace the IAB with the above structures for comparison experiments respectively, and the results are shown in Table 5. It can be seen that Case 3 achieves the best results. Case 1 is lightweight, but the excessive split operation limits the interaction of channels and leads to the underutilization of features. Case 2 achieves good results, but generates more parameters and computations. The above experiments illustrate that IAB is efficient.

5.1.2 Effectiveness of dual attention block (DAB). We investigated the effectiveness of DAB, as shown in Table 6, and we used Case 4 as the baseline data. Compared with Case 4, the PSNR and SSIM are significantly lower in Case 1, demonstrating that the DAB has a positive effect. To further illustrate the advantages of dual attention block, we investigated the effectiveness of each component for the dual attention block. Case 2 uses only channel attention (CA).

![Fig 9. Feature extraction structure of different methods. (a) Split+Conv3. (b) Conv3+Conv5. (c) IAB.](https://doi.org/10.1371/journal.pone.0272398.g009)

| Case | Module       | Param [M] | FLOPs [G] | PSNR/SSIM   |
|------|--------------|-----------|-----------|-------------|
| 1    | Split+Conv3  | 0.21      | 0.94      | 22.15/0.810 |
| 2    | Conv3+Conv5  | 0.74      | 3.09      | 22.49/0.824 |
| 3    | IAB          | 0.35      | 1.53      | 22.68/0.829 |

https://doi.org/10.1371/journal.pone.0272398.t005
Case 3 adds spatial attention (SA) to CA. We can see that both CA and SA can improve the model. Case 4 is the proposed DAB. We can see that Case 4 has the best performance, which proves the effectiveness of STD. As shown in Fig 10(h) has the best visual effect. The above study verifies the effectiveness of IAB and DAB. IAB can extract features efficiently, while DAB can reduce redundant features.

Table 6. Study on Dual Attention Block (DAB). CA, SA, STD are channel attention, spatial attention, and standard deviation respectively.

| Case | CA | SA | STD | Param [M] | FLOPs [G] | PSNR/SSIM |
|------|----|----|------|-----------|-----------|-----------|
| 1    | ✗  | ✗  | ✗    | 0.34      | 1.48      | 22.13/0.807 |
| 2    | ✓  | ✗  | ✗    | 0.34      | 1.48      | 22.35/0.816 |
| 3    | ✓  | ✓  | ✗    | 0.35      | 1.53      | 22.49/0.816 |
| 4    | ✓  | ✓  | ✓    | 0.35      | 1.53      | 22.68/0.829 |

Fig 10. Visual results of models with different components on LOL dataset. (a) Input. (b) w/o pyramid. (c) w/o MFAB. (d) w/o AFAB. (e) w/o DAB. (f) CA. (g) CA+SA. (h) Ours. (i) GT.

https://doi.org/10.1371/journal.pone.0272398.g010
5.2 Study on the number of multi-scale feature aggregation blocks

In this part, we explored the effect of the number of multi-scale feature aggregation blocks (MFABs). In general, deeper network has more parameters and stronger fitting ability. We provided the results of PSNR, SSIM, parameters, computations and the running time by using different numbers of MFABs, as shown in Table 7. When the number of MFABs gradually increases, the model can obtain better PSNR and SSIM, but the parameters, computations and running time of the model also rise. When the number of MFABs is more than 4, the rising trend of PSNR becomes slow and SSIM appears to decrease, while the parameters and computations of the model increase greatly. In terms of lightweight and efficiency, we set the number of MFABs as 4.

5.3 Effectiveness of network structure

In this part, we performed a series of ablation experiments on network structure, pyramid structure and AFAB. We explored the impact of the pyramid structure, as shown in Table 3. Case 1 removes the pyramid structure and AFAB, retains only the shallow branches, and doubles the number of channels to equal the number of parameters in Case 4 for a fair comparison. We can see that the PSNR and SSIM of Case 1 are significantly lower due to the lack of a pyramid structure. As shown in Fig 10, the visual results of the model without the pyramid structure are significantly worse. The experimental and visual results show that the pyramid structure is effective.

We investigated the effect of AFAB and the results are shown in Table 3. Case 2 removes AFAB, and it can be seen that the PSNR and SSIM of Case 2 are significantly lower. We give the corresponding visual effects, as shown in Fig 10, the images generated by the model using AFAB have better contrast and color. AFAB solves the inconsistency of the pyramid and enables the deep semantic information to better guide the shallow fine-grained information to learn finer textures and colors. To illustrate the advantages of the proposed AFAB, we further compare it with ASFF [39] and AFSM [40]. Since AFSM only performs adaptive fusion of neighboring scale features, we improved it to receive features from each layer of the pyramid. As shown in Table 8, we can see that AFAB obtains the best value of PSNR and SSIM. This indicates that AFAB can better adaptively fuse the features of different scales.

In summary, all parts of the model are effective, and our model achieves good performance with the synergy of each component.

Table 7. Study on the number of Multi-scale Feature Aggregation Blocks (MFABs).

| Number | PSNR/SSIM   | Param [M] | FLOPs [G] | Time [s] |
|--------|-------------|-----------|-----------|----------|
| 2      | 22.23/0.806 | 0.24      | 1.05      | 0.0140   |
| 3      | 22.37/0.811 | 0.30      | 1.29      | 0.0159   |
| 4      | 22.68/0.829 | 0.35      | 1.53      | 0.0182   |
| 5      | 22.71/0.824 | 0.41      | 1.75      | 0.0198   |

Table 8. Study on Adaptive Feature Aggregation Block (AFAB).

| Case | Module | Param [M] | FLOPs [G] | PSNR/SSIM   |
|------|--------|-----------|-----------|-------------|
| 1    | ASFF   | 0.35      | 1.53      | 22.43/0.823 |
| 2    | AFSM   | 0.36      | 1.55      | 22.31/0.820 |
| 3    | AFAB   | 0.35      | 1.53      | 22.68/0.829 |
6 Conclusions
In this paper, we design an efficient adaptive feature aggregation network for low-light image enhancement. We propose two important modules, MFAB and AFAB, to construct the proposed network. MFAB efficiently extracts features using asymmetric convolution and a dual attention mechanism. By using MFAB, the reconstructed image has rich texture and the noises is effectively suppressed. AFAB combines one-dimensional convolution to efficiently rescale the features of each branch, it overcomes the inconsistency of the pyramid structure and improves the luminance, color and texture deviation of the enhanced images. Extensive experiments and ablation studies have shown that the proposed method has significant advantages over state-of-the-art methods. Meanwhile, the method has a quick running time, which has great potential for assisting advanced vision tasks or applying on mobile. In the future, we will further validate the proposed method in more image restoration tasks, such as image de-snowing and de-raining task.

Supporting information
S1 File.
(ZIP)
S2 File.
(RAR)

Author Contributions
Conceptualization: Canlin Li, Pengcheng Gao.
Data curation: Jinhua Liu, Lihua Bi.
Formal analysis: Canlin Li, Pengcheng Gao.
Funding acquisition: Canlin Li.
Investigation: Jinhua Liu, Shun Song, Lihua Bi.
Methodology: Canlin Li, Pengcheng Gao.
Project administration: Canlin Li.
Resources: Shun Song, Lihua Bi.
Software: Canlin Li, Pengcheng Gao, Jinhua Liu.
Supervision: Canlin Li, Pengcheng Gao.
Validation: Canlin Li, Pengcheng Gao, Shun Song, Lihua Bi.
Visualization: Canlin Li, Pengcheng Gao.
Writing – original draft: Canlin Li, Pengcheng Gao, Jinhua Liu.
Writing – review & editing: Canlin Li, Pengcheng Gao, Jinhua Liu.

References
1. Satzoda R. K., Trivedi M. M. Looking at Vehicles in the Night: Detection and Dynamics of Rear Lights. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(12): 4297–4307. https://doi.org/10.1109/TITS.2016.2614545
2. Qian Z. et al. A New Approach to Polyp Detection by Pre-Processing of Images and Enhanced Faster R-CNN. IEEE Sensors Journal, 2021, 21(10): 11374–11381. https://doi.org/10.1109/JSEN.2020.3036005

3. Zhang J., Feng W., Yuan T., et al. SCSTCF: Spatial-Channel Selection and Temporal Regularized Correlation Filters for visual tracking. Applied Soft Computing, 2022, 118: 108485. https://doi.org/10.1016/j.asoc.2022.108485

4. Xia R., Chen Y., Ren B. Improved anti-occlusion object tracking algorithm using Unscented Rauch-Tung-Striebel smoother and kernel correlation filter. Journal of King Saud University—Computer and Information Sciences, 2022: 1319–1578.

5. Pizer S., Johnston R., Ericksen J., et al. Contrast-limited adaptive histogram equalization: speed and effectiveness. Proceedings of the First Conference on Visualisation in Biomedical Computing, 1990: 337–345.

6. Ibrahim H., Pik Kong N. S. Brightness Preserving Dynamic Histogram Equalization for Image Contrast Enhancement. IEEE Transactions on Consumer Electronics, 2007, 53(4): 1752–1758. https://doi.org/10.1109/TCE.2007.4429280

7. Jobson D. J., Rahman Z., Woodell G. A. Properties and performance of a center/surround retinex. IEEE Transactions on Image Processing, 1997, 6(3): 451–462. https://doi.org/10.1109/83.557356 PMID: 18282940

8. Rahman Z., Jobson D. J., Woodell G. A. Multi-scale retinex for color image enhancement. Proceedings of 3rd IEEE International Conference on Image Processing, 1996, 3: 1003-1006.

9. Jobson D. J., Rahman Z., Woodell G. A. A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE Transactions on Image Processing, 1997, 6(7): 965–976. https://doi.org/10.1109/83.597272 PMID: 18282987

10. Wang S., Zheng J., Hu H., et al. Naturalness Preserved Enhancement Algorithm for Non-Uniform Illumination Images. IEEE Transactions on Image Processing, 2013, 22(9): 3538–3548. https://doi.org/10.1109/TIP.2013.2261309 PMID: 23661319

11. Cai B., Xu X., Guo K., et al. A joint intrinsic-extrinsic prior model for retinex. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017: 4020–4029.

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

13. Wang W., Wei C., Yang W., et al. Gladnet: Low-light enhancement network with global awareness. Proceedings of the IEEE International Conference on Automatic Face Gesture Recognition, 2018: 751–755

14. Zamir S. W., Arora A., Khan S., et al. Learning enriched features for real image restoration and enhancement. Proceedings of the European Conference on Computer Vision(ECCV), 2020.

15. Wei C., Wang W., Yang W., et al. Deep retinex decomposition for low-light enhancement. Proceedings of the British Machine Vision Conference, British Machine Vision Association, 2018.

16. Zhang Y., Zhang J., Guo X., Kindling the darkness: A practical low-light image enhancer. Proceedings of the 27th ACM International Conference on Multimedia, 2019: 1632–1640, ACM.

17. Li C., Guo C., Loy C. Learning to enhance low-light image via zero-reference deep curve estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021. https://doi.org/10.1109/TPAMI.2021.3126387

18. Risheng L., Long M., Jiaoz Z., et al. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

19. Srinivasan S., Balram N. Adaptive contrast enhancement using local region stretching. Proceedings of the 9th Asian symposium on information display, 2006: 152–155.

20. Guo X., Li Y., Ling H. LIME: Low-Light Image Enhancement via Illumination Map Estimation. IEEE Transactions on Image Processing, 2017, 26(2): 982–993. https://doi.org/10.1109/TIP.2016.2639450

21. Xu Y., Sun B. A Novel Variational Model for Detail-Preserving Low-Illumination Image Enhancement. Signal Processing, 2022: 108468. https://doi.org/10.1016/j.sigpro.2022.108468

22. Wang R., Zhang Q., Fu C., et al. Underexposed photo enhancement using deep illumination estimation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 6842–6850.

23. Lio X., Xie Q., Zhao Q., et al. Low-light Image Enhancement by Retinex Based Algorithm Unrolling and Adjustment. arXiv preprint arXiv:2202.05972, 2022.

24. Jiang K., Wang Z., Wang Z., et al. Degrade is upgrade: Learning degradation for low-light image enhancement. arXiv preprint arXiv:2103.10621, 2021.
25. Li J., Li J., Fang F., et al. Luminance-Aware Pyramid Network for Low-Light Image Enhancement. IEEE Transactions on Multimedia, 2021, 23: 3153–3165. https://doi.org/10.1109/TMM.2020.3021243
26. Li Z., Shu H. and Zheng C. Multi-Scale Single Image Dehazing Using Laplacian and Gaussian Pyramids. IEEE Transactions on Image Processing, 2021, 30: 9270–9279. https://doi.org/10.1109/TIP.2021.3123551 PMID: 34748490
27. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation. Proceedings of International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.
28. Guo Y., Chen J., Wang J., et al. Closed-Loop Matters: Dual Regression Networks for Single Image Super-Resolution. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 5406-5415.
29. Lim S. and Kim W. DSLR: Deep Stacked Laplacian Restorer for Low-Light Image Enhancement. IEEE Transactions on Multimedia, 2021, 23: 4272–4284. https://doi.org/10.1109/TMM.2020.3039361
30. Jiang K., Wang Z., Yi P., et al. Multi-Scale Progressive Fusion Network for Single Image Dehazing. Proceedings of the 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR), 2020: 8343-8352.
31. Zhang P., Dai X., Yang J., et al. Multi-scale vision longformer: A new vision transformer for high-resolution image encoding Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 2998-3008.
32. Li J., Fang F., Mei K., et al. Multi-scale Residual Network for Image Super-Resolution. Proceedings of the European Conference on Computer Vision (ECCV), 2018.
33. Li J., Tang J., Wu, Residual G. Feature Distillation Network for Lightweight Image Super-Resolution. Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2020.
34. Ding X., Guo Y., Ding G., et al. ACNet: Strengthening the Kernel Skeletons for Powerful CNN via Asymmetric Convolution Blocks. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 1911-1920.
35. Chen Y., Liu L., Phonevilay V., et al. Image super-resolution reconstruction based on feature map attention mechanism. Applied Intelligence (2021) 51:4367–4380. https://doi.org/10.1007/s10489-020-02116-1
36. Tian Y., Wang Y., Yang L., et al. Concatenated Attention Neural Network for Image Restoration. arXiv preprint arXiv:2006.11162, 2020.
37. Hui Z., Gao X., Yang Y., et al. Lightweight Image Super-Resolution with Information Multi-distillation Network. Proceedings of the 27th ACM International Conference on Multimedia (MM’19), 2019:2024–2032.
38. Wang Q., Wu B., Zhu P., et al. Eca-net: Efficient channel attention for deep convolutional networks. Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 11531–11539
39. Liu S., Huang D., Wang Y. Learning spatial fusion for single-shot object detection,2019. arXiv preprint arXiv:1911.09516, 2019.
40. Yi Q., Li J., Fang F., et al. Efficient and Accurate Multi-scale Topological Network for Single Image Dehazing. IEEE Transactions on Multimedia.
41. Johnson J., Alahi A., Li F. Perceptual losses for real-time style transfer and super-resolution. arXiv preprint arXiv:1603.08155, 2016.
42. Deng J., W Dong, Socher R., et al. ImageNet: A large-scale hierarchical image database. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009: 248-255.
43. Bichkovsky V., Paris S., Chan E., et al. Learning photographic global tonal adjustment with a database of input/output image pairs. Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011: 97–104
44. Hu Y., He H., Xu C., et al. Exposure: A white-box photo post-processing framework. ACM Transactions on Graphics, 2018, 37(2):26.1–26.17. https://doi.org/10.1145/3181974
45. Yu S., Wang Y., Kao M., et al. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 6306–6314