EDMFEN: Edge detection-based multi-scale feature enhancement Network for low-light image enhancement

Canlin Li[1]*, Shun Song[2]*, Pengcheng Gao[1], Wei Huang[1], and Lihua Bi[1]
[1] School of Computer Science and Technology, Zhengzhou University of Light Industry
Zhengzhou, 450000 China
[2] School of Information Engineering, Henan Vocational University of Science and Technology
Zhoukou, 466000 China
[e-mail: lcl_zju@aliyun.com, hnzktkss@163.com]
*Corresponding authors: Canlin Li, Shun Song

Received November 30, 2023; revised February 4, 2024; accepted March 13, 2024; published April 30, 2024

Abstract

To improve the brightness of images and reveal hidden information in dark areas is the main objective of low-light image enhancement (LLIE). LLIE methods based on deep learning show good performance. However, there are some limitations to these methods, such as the complex network model requires highly configurable environments, and deficient enhancement of edge details leads to blurring of the target content. Single-scale feature extraction results in the insufficient recovery of the hidden content of the enhanced images. This paper proposed an edge detection-based multi-scale feature enhancement network for LLIE (EDMFEN). To reduce the loss of edge details in the enhanced images, an edge extraction module consisting of a Sobel operator is introduced to obtain edge information by computing gradients of images. In addition, a multi-scale feature enhancement module (MSFEM) consisting of multi-scale feature extraction block (MSFEB) and a spatial attention mechanism is proposed to thoroughly recover the hidden content of the enhanced images and obtain richer features. Since the fused features may contain some useless information, the MSFEB is introduced so as to obtain the image features with different perceptual fields. To use the multi-scale features more effectively, a spatial attention mechanism module is used to retain the key features and improve the model performance after fusing multi-scale features. Experimental results on two datasets and five baseline datasets show that EDMFEN has good performance when compared with the state-of-the-art LLIE methods.

Keywords: low-light image enhancement, deep learning, edge detection, multi-scale feature enhancement, spatial attention mechanism.

This work was supported in part by the Science and Technology Planning Project of Henan Province under Grant 242102211003 and 212102210097.

http://doi.org/10.3837/tiis.2024.04.009 **ISSN : 1976-7277**

1. Introduction

Low-light image enhancement (LLIE) is a major challenge in the task of computer vision. Low-light images have some defects, such as blurred target content, low brightness, poor contrast, and darker colors. These defects have a great impact on tasks like target detection, recognition, and image segmentation. Therefore, to recover image details and compensate for image brightness and contrast, LLIE techniques are widely studied. The mainstream LLIE techniques include traditional LLIE methods and deep learning-based LLIE methods.

Traditional LLIE methods. Traditional LLIE methods can divide into two main categories: histogram equalization-based (HE) methods [1-2] and Retinex theory-based methods [3-6]. Many HE methods are designed for LLIE in the early days. Yang et al [1] proposed the adaptive Contrast Enhancement (ACE), which uses local information to highlight details and textures, thus sharpening the edges of the image. ACE enriches the spatial structure of the image but generates a lot of noise and takes a long time to compute. Brightness preserving dynamic histogram equalization (BPDHE) [2] was proposed that maintained luminance is not adapted to low-light images as it must maintain luminance. HE methods are simple and fast, however, they usually lead to over-enhancement. Recently, the methods based on the Retinex theory have been widely used in LLIE. Wang et al. [3] proposed a Naturalness preserved enhancement (NPE) for non-uniformly illuminated images. NPE method enhances image details while still maintaining naturalness. The weighted variance model proposed by simultaneous reflectance and illumination estimation (SRIE) [4] to estimate the illumination and reflectance of the observed image can suppress noise to a certain extent. Multi-scale fusion (MF) [5] decomposed the image into reflectance image and illumination image, where the latter is processed under different illumination levels, these processed images are appropriately weighted fusion to obtain the final illumination map. The reflectance image compensates this illumination image to obtain an enhanced image. Guo et al [6] proposed a simple but effective illumination map estimation (LIME) that allows the illuminance mapping to be enhanced accordingly. This illuminance mapping estimated the illuminance of each pixel by its maximum value in the R, G, and B channels, coupled with a structured refinement of the initial illuminance mapping. Retinex achieved a balance among color constancy, edge enhancement and dynamic range compression, thus allowing adaptive enhancement of a wide range of different types of images. However, it may cause color distortion and artifacts in the enhanced images.

Deep learning-based LLIE methods. Deep learning-based LLIE methods are received more and more attention due to their non-linear feature learning capability. Deep learning-based LLIE methods are mainly classified into supervised learning, unsupervised learning, and zero-reference learning. Supervised learning methods usually compare the enhanced results with the labeled image for supervising the network training process. Kin Gwn Lore et al. [7] first combined CNN with LLIE techniques and proposed a depth-based self-encoder approach to make the brighter parts of the image adaptively not over-amplified in the high dynamic range. RetinexNet [8] applied Retinex theory to CNN to estimate and adjust the illumination map to achieve LLIE. This method can achieve good results, but there may be a certain amount of noise in the enhancement results. Therefore, KinD [9] designed a network same as RetinexNet by adding a noise-canceling recovery network to improve image fidelity. DeepUPE [10] learnt the mapping relationship between the illumination map and the image by extracting local and global features. Li et al. [11] proposed a luminance-aware pyramid network (LPNet). The network learnt the luminance between the input image and the GT image by adding light-sensitive losses to progressively supervise the refinement of the

illumination in the branches. Models for unsupervised learning can be trained on unpaired datasets. EnlightenGAN [12] is the first unsupervised method that employ a multi-scale discriminator with a self-regularising loss function. Zero-sample learning methods can learn enhancement from test images only. Zero-DCE [13] took low-light images as input and high-order curves as output and dynamically adjusted the curves as input at the pixel level to obtain enhanced images. An improved and simpler of Zero-DCE is proposed called Zero-DCE++ [14], which maintains the performance of Zero-DCE while providing faster inference. Pan et al. [15] proposed a simple and effective video moment retrieval network, which is trained by a bottom-up method, possesses robustness and can locate the target moment in untrimmed videos based on natural language queries. In [16], it shows how deep learning to extract features, reinforcement learning to optimize decisions, and world models to predict environments can improve learning and decision-making. The advantages of these integrated methods are the ability to handle complex tasks, adapt to dynamic environments, and demonstrate strong performance and adaptability in various domains. Ma et al. [17] described a method called "Motion Stimulation" for combinatorial action recognition. The method utilizes motion stimulation to enhance the representation of action features and captures the temporal information of the action by applying the stimulation to video frames of different periods. Fu et al. [18] proposed a method called "Recurrent Thrifty Attention Network" for remote sensing scene recognition. The method effectively captures key features in remote sensing images by introducing a recurrent attention mechanism and is optimized in terms of computational efficiency. Ma et al. [19] proposed the SCI method, which is a self-calibrated illumination learning framework for low-light image enhancement. It features efficiency, flexibility, and high quality. Zheng et al. [20] proposed the Semantic method, a semantic-guided zero-shot low-light enhancement network that exhibits characteristics of zero-shot learning, efficiency, and semantic preservation. Wu et al. [21] proposed the URetinex-Net method, which enhances low-light images through deep unrolling networks and implicit prior regularization modeling. It exhibits adaptability and high efficiency. Fan et al.[22] proposed an image enhancement network (HWMNet) based on an improved hierarchical model: M-Net+. Using semi-wavelet attention blocks on M-Net+ to enrich features in the wavelet domain. The normalized flow model proposed by Wang et al.[23] can effectively model the one-to-many relationship between low-light image enhancement and normal exposure images. By learning to map the distribution of normally exposed images into a Gaussian distribution, the method is better able to model the conditional distribution of normally exposed images, thereby providing better quantitative and qualitative results, including improved exposure illumination, reduced noise and artifacts and enhanced color.

Although these methods are effective for LLIE, there are still deficiencies. Current deep learning-based methods do not take sufficient account of edge information. However, the edge information is important for the image enhancement results due to the presence of texture details at the edges of images. Overly complex network models are too demanding in terms of execution efficiency and performance. Therefore, to address these issues, an edge detection-based multi-scale feature enhancement network is proposed for LLIE (EDMFEN) in this paper. The network combines an edge detection module (EDM) with a multi-scale feature enhancement module (MSFEM). The edge information extracted by EDM is used to refine the original image, and then the extracted edge information is fused with the output of MSFEB to enrich the enhanced image detail and texture structure. High-quality images with sharp edges can be reconstructed with minimal pixel loss.

The main contributions of this paper can be summarized as follows:

- We proposed a lightweight network for LLIE, EDMFEN, which first acquires edge information from the original images, then acquires features from multi-scale images and fuses them to reconstruct the enhanced images.
- An EDM constructed by the Sobel operator is introduced to extract edge information from original images, and the spatial structure details are complemented to acquire high-quality enhanced images.
- An MSFEM with an attention mechanism is proposed. In this module, multi-branch parallel computing is used to obtain image features with different perceptual fields, and an attention mechanism is introduced to obtain contrast information in the image to make the obtained features more representative.

2. Related works

2.1 Light weight Network

Complex network models often have high requirements for hardware devices. How to solve the storage and computation problems of complex networks for deep learning is the key to applying these methods to real-world scenarios. lightweight network models are currently becoming a research trend for various tasks in computer vision. SqueezeNet [24] reduces parameters by 1×1 Conv in the Squeeze module and expands the number of channels with the Expand module to retain more features with fewer channels. The depth-separable convolution is used in MobileNet [25] to convolve different channels separately, and then point-by-point convolution is used to fuse the features, greatly reducing the amount of computation and parameters. The core of ShuffleNet [26] is group convolution and channel reorganization, which reduces computational effort while maintaining model performance. These lightweight networks usually have a small amount of computation and a number of parameters, as well as have high execution efficiency.

2.2 Edge detection

Edge is an important part of images as well as the basis for computer vision tasks. Traditional edge detection methods using differential operators are more sensitive to noise in the image and fast but obtain incomplete information about the structure. The Canny detection [27] detects good closure and continuity of edges, however, the execution is less efficient. Roberts operator [28] is an operator that uses local difference operators to find edges. It uses the difference between two adjacent pixels in the diagonal direction to approximate the gradient amplitude to detect edges. The effect of detecting vertical edges is better than oblique edges, and the positioning accuracy is high. It is sensitive to noise and cannot suppress the influence of noise. The Prewitt operator [29] is an edge detection operator that detects the position and intensity of edges by performing convolution operations in the horizontal and vertical directions in the image. It can effectively capture edge features in both horizontal and vertical directions and performs well in edge detection tasks, but is more sensitive to noise. The Sobel detection [30] is a simple and computationally compact method that has a smoothing effect on noise. Using Sobel detection can be obtained accurate edge detail information and a rich texture structure from the image.

2.3 Attention mechanism

Recently, most attention blocks have been proposed to focus on deep weighted non-matching,

emphasizing essential information features and suppressing useless features. Hu et al. [31] proposed the Squeeze and Stimulate block to perform recalibration of the feature responses of channels by modeling the inter-dependencies between them. Considering the importance of the positional relationship of each pixel, a non-local Network (NLNet) [32] is proposed to compute the interaction between any two positions ignoring their distance. GCNet [33] simplified the NLNet framework by using query-independent attention maps for all positions. Fu et al. [34] designed a dual-attention network that contains spatial and channel modules to combine local features with global features.

3. Proposed method

EDMFEN improves the edge details and clarity of low-light images and enhances the overall contrast of the image by fusing the edge information and multi-scale features of the image. The edge detection block is used to extract the edge information of the image, which can help enhance the details and edges of the image to make it clearer. In the multi-scale feature extraction block, the multi-scale information is extracted, and combined with the ECA attention mechanism, the weights of features on different scales can be adaptively assigned according to their importance, thereby enhancing important information in the image. The edge information of the image is continuously injected into the multi-scale information to better restore the details of the image. The framework of EDMFEN is shown in **Fig. 1**.

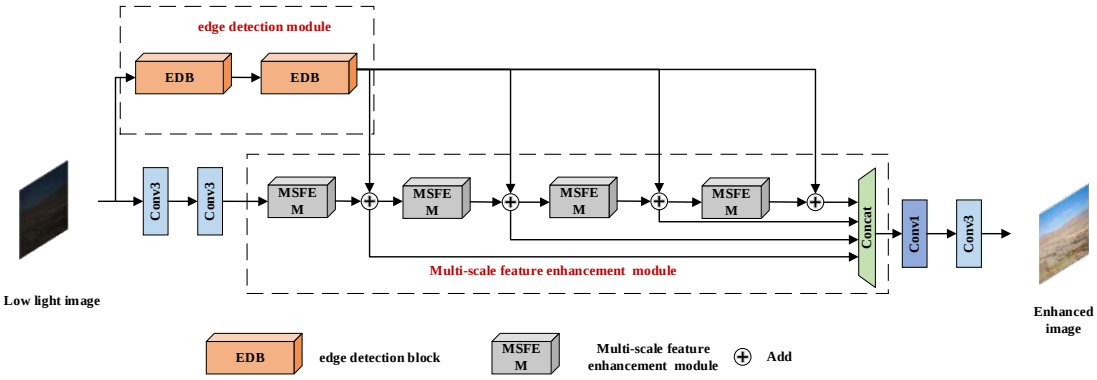

Fig. 1. The framework of the EDMFEN

Two EDBs are used to extract edge information, while four MSFEMs are used to extract multi-scale features.

Define the input image as I which is processed by two 3×3 Convs to obtain the feature image I_o. The output in the edge extraction branch can be written as (1):

$$I_1 = F_{EDB}\left(F_{EDB}\left(I\right)\right) \tag{1}$$

where F_{EDB} denotes the EDB for the edge extraction section, and I_1 is the output of EDM. For multi-scale features extraction, refining the output of each MSFEM in the branch can be defined as (2):

$$E_n = \begin{cases} F_{MSFEM}\left(I_0\right) & n=1 \\ F_{MSFEM}\left(E_{n-1}+I_1\right) & n=2,3,...,n \end{cases} \tag{2}$$

where F_{MSFEM} denotes the operation of MSFEM, E_n represent the output of *n-th* MSFEM. Each intermediate output of the MSFEM is added to the next MSFEM by adding element-by-

element to the output of the edge extraction module to gradually guide the feature extraction module.

To make use of the image features more efficiently, a 1×1 Conv is used to aggregate these multi-scale features. In this way, the integrity of the hierarchical information can be retained with a relatively small number of parameters and computational effort. Finally, a 3×3 Conv is used to obtain the enhanced images.

In the training process, given a training data set $\left\{I_{in}^{m}, I_{gt}^{m}\right\}_{m=1}^{M}$, the total loss function is minimized by comparing the real image with the output image of the model, which can be expressed as (3):

$$\hat{\theta}=\arg\min_{\theta}\frac{1}{M}\sum_{m=1}^{M}L_{total}\left(F_{EDMFEN}\left(I_{in}^{m}\right), I_{gt}^{m}\right) \tag{3}$$

where I_{in}^{m} and I_{gt}^{m} represent the input and the target image respectively, θ denotes parameters of the EDMFEN, $F_{EDMFEN}(.)$ represents the EDMFEN, $L_{total}(.)$ represents the total loss function to minimize the difference between target image and enhanced image, the loss function is described in the next section 3.3.

3.1 Edge detection module

Edges are an important part of images and contain a large amount of spatial structure information. Extracting edge information can retain the structural information of the image. In low-light image enhancement, edge detection algorithms can be used to extract edge information in the image, and then the brightness, contrast, color, and other attributes of the image can be adjusted based on the edge information to achieve better enhancement effects. There is a certain amount of noise in the edges of most images, which may produce artifacts, so an edge detection branch consisting of an edge detection module is introduced in EDMFEN to extract edge information of the image. The Sobel operator method is simple and has a small amount of calculation. It can not only produce better detection results but also have a smooth suppression effect on noise. Therefore, the edge detection block provides more accurate edge direction information and restores rich texture information of the image, so we introduced the Sobel operator in EDMFEN to extract the edge information [35]. The structure of EDB is shown in **Fig. 2**.

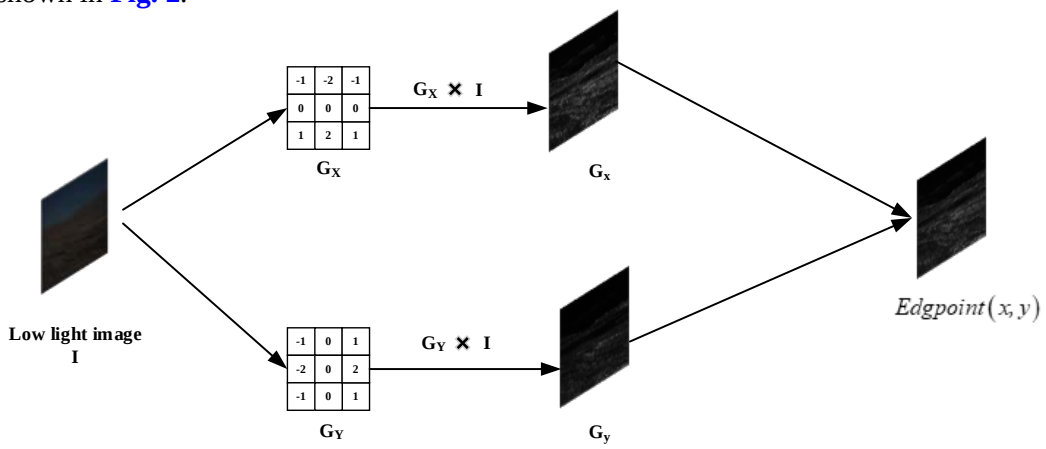

Fig. 2. The structure of EDB

The Sobel operator is used to calculate a gray-scale approximation to the image luminance function. The Sobel operator can generate the corresponding gradient vector at any point in the image. The Sobel operator detects image edges based on the gradient method. The components of the gradient method represent the rate of change of pixel values with distance from the x and y directions. The Sobel operator takes the derivative of the input image pixels and finds the point with the largest derivative to locate the edge. In a discrete image, the pixel spacing between two points can be considered as 1, based on the top, bottom, left, and right neighborhood of the edge point. The weights obtained are the convolution kernels, and the two 3×3 matrices, G_X in the horizontal direction and G_Y in the vertical direction, the horizontal and vertical luminance difference approximations can be obtained by convoluting the image with G_X and G_Y. Where G_x and G_y represent the image gray values detected by the horizontal and vertical edges, respectively, then the gradient at each pixel is estimated, and the edge point is found for each point in the image by combining the results of the convolution. As shown in (4):

$$Edgpoint(x,y)=\sqrt{G_x^{\,2}+G_y^{\,2}} \tag{4}$$

The Sobel detection has good detection results and smooth noise processing effect. It can provide more accurate edge direction information and restore rich texture information of images.

3.2 Multi-scale feature enhancement module (MSFEM)

Some research work [36] [37] shows that multi-scale framework can extract features of different receptive fields, and these features represent rich information in images. Therefore, we proposed a multi-scale feature extraction module to increase the detailed information of the image. MSFEM is designed to obtain rich features, comprising an MSFEB and an attention mechanism block. The MSFEB is used to extract multi-scale features. In addition, to prevent the loss of shallow features when extracting deep information, a jump connection is added after the attention mechanism block. High-contrast, detail-rich image features can be obtained by using MSFEM. The structure of MSFEM is shown in **Fig. 3**.

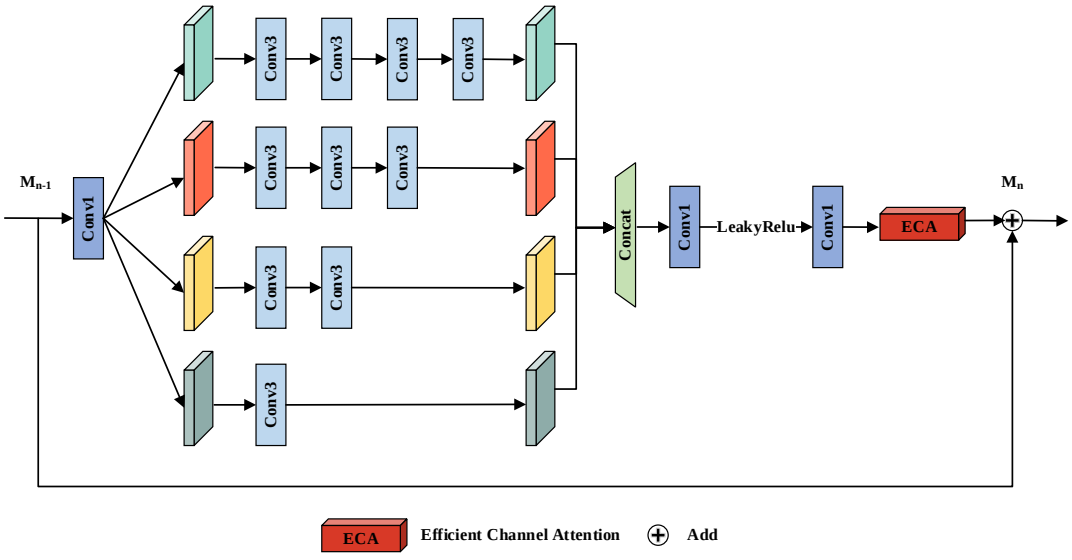

Fig. 3. The structure of MSFEM

3.2.1 Multi-scale feature extraction block (MSFEB)

To obtain rich features, a multi-scale feature extraction module is designed to obtain features of different scales through different receptive fields. In **Fig. 3**, the input of the network is prepossessed by a 1×1 Conv and then split along channel dimensional to divide into four groups that have the same channels. Each group uses a different number of 3×3 Convs to extract multi-scale features. Two 3×3 Convs and a 5×5 Conv have the same field of perception, three 3×3 Convs, and a 7×7 Conv have the same field of perception, and four 3×3 Convs have the same field of perception as a 9×9 Conv. However, compared to the direct use of 5×5, 7×7, and 9×9 Convs, using combined 3×3 Convs can reduce the number of parameters, retaining a larger receptive field in the meanwhile. However, using a grouped model hinders the flow of information between groups and weakens the feature representation. To facilitate feature fusion among groups, after group exchange, the obtained feature maps at different scales are stacked using contact. The stacked information is integrated using 1×1 Conv, and then the leakyRelu activation function is applied to filter the valid features. The channels are downscaled using 1×1 Conv to restore to the same channel number as input, thereafter the features are weighted using the attention mechanism module.

3.2.2 ECANet

The core idea of the attention mechanism in illumination image enhancement is to highlight key areas in the image to enhance image quality. Reduce the influence of noise and other interference factors, emphasize necessary information features, and suppress useless features. In the field of deep learning, the performance of a network relies to receive and process a large amount of data. However, at certain moments, only a small part of the data is the most important. In this case, the attention mechanism is very suitable. In [31], a block is proposed to squeeze and expand channels, thus establishing the interdependence of all channels. The attention mechanism further handles multi-scale feature enhancement by weighting key features in the image to obtain weighted feature maps. ECANet [38] can effectively improve model performance by avoiding dimensionality reduction and increasing interactions between adjacent channels. In ECANet, input features are processed through global average pooling layers and 1×1 Convs to aggregate channel features and improve information flow. The structure is shown in **Fig. 4**. The input feature map is acquired through global average pooling to obtain a 1×1 Conv feature vector, and then the feature vector is passed through a one-dimensional convolution and activation function with a convolution kernel size of 5 to obtain the weight factor. Finally, the weight factor is multiplied by the input feature maps to obtain feature maps with an attention mechanism. The application of ECA in this section can effectively promote multi-scale feature fusion, resulting in more representative feature maps.

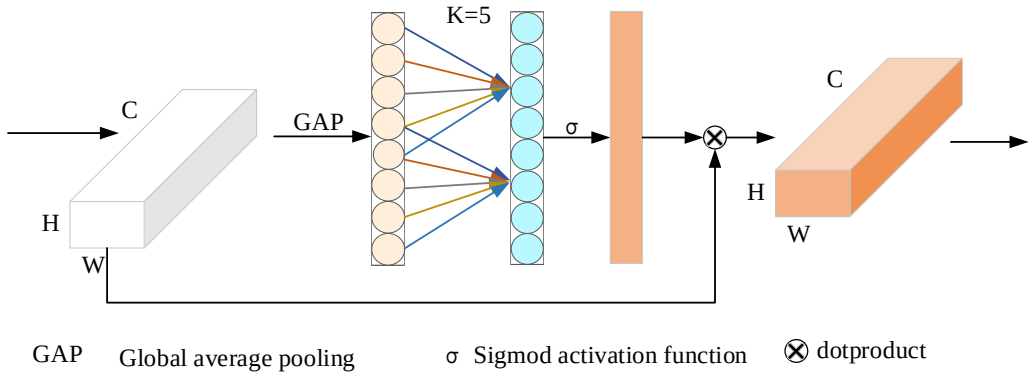

Fig. 4. The Network Structure of Attention Mechanism ECA

3.3 Loss function

Given a training dataset to train the network model, a suitable loss function is also essential. We designed a hybrid loss strategy, which is combined SSIM loss with VGG loss, to improve the quality of enhanced images in spatial detail and edge structure.

SSIM loss: SSIM can evaluate image quality, which is measured from three parts: image brightness, contrast, and structure, as shown in (5) and (6):

$$SSIM = \left[l\left(I,\hat{I}\right) \right]^{\alpha} \left[c\left(I,\hat{I}\right) \right]^{\beta} \left[s\left(I,\hat{I}\right) \right]^{\gamma} \tag{5}$$

$$\begin{cases} l\left(I,\hat{I}\right) = \dfrac{2\sigma_I \sigma_{\hat{I}} + C_1}{\sigma_I^2 + \sigma_{\hat{I}}^2 + C_1} \\ c\left(I,\hat{I}\right) = \dfrac{2\mu_I \mu_{\hat{I}} + C_2}{\mu_I^2 + \mu_{\hat{I}}^2 + C_2} \\ s\left(I,\hat{I}\right) = \dfrac{\mu_{I\hat{I}} + C_3}{\mu_I \ \mu_{\hat{I}} + C_3} \end{cases} \tag{6}$$

where $\hat{I}$ is the enhanced image. $l\left(I,\hat{I}\right)$, $c\left(I,\hat{I}\right)$, and $s\left(I,\hat{I}\right)$ are the difference in luminance, contrast, and structure between I and $\hat{I}$. α, β, and λ represent the proportion of different characteristics in the SSIM measure, σ_I and $\sigma_{\hat{I}}$ are the mean values of I and $\hat{I}$ respectively, μ_I and $\mu_{\hat{I}}$ are the variance of I and $\hat{I}$ respectively, $\mu_{I\hat{I}}$ is the covariance between I and $\hat{I}$, C_1, C_2, and C_3 are constants. We use multi-scale SSIM loss function to enhance the structural similarity of images. L_{SSIM} can be expressed as (7):

$$L_{SSIM} = 1 - SSIM \tag{7}$$

The larger the value of SSIM, the less loss in the image and the more similar the structure.

Perceived loss: We use VGG perceived loss to calculate the loss of the spatial dimension of the image, using the semantic information of the image to improve the spatial detail of the enhanced images. As shown in (8):

$$L_{VGG} = \frac{1}{WHC} \sum_{x=1}^{W} \sum_{y=1}^{H} \sum_{s=1}^{C} \left\| \hat{I}_{x,y,s} - I_{x,y,s} \right\|^2 \tag{8}$$

where W, H, and C represent the width, height, and channel of the image.

Total loss: The final total loss of the network model is the weighted sum of perceived loss and SSIM loss, L_{total} can be expressed as (9):

$$L_{total} = \lambda_s L_{SSIM} + \lambda_v L_{VGG} \tag{9}$$

where λ_s and λ_v denote the weights corresponding to L_{SSIM} and L_{VGG}

4. Experiments results

4.1 Experimental setup

To verify the effectiveness of the EDMFEN, we designed and analyzed the experimental results on LOL [39] and MIT5K [40] datasets. To ensure the fairness of the experiments, all methods are experimented on a PC with an NVIDIA GeForce RTX 2070 8GB GPU. Adam is used as an optimizer with the model default parameters. We set the batch size to 16, the input image size for each batch to 96×96, and set 0.0002 as the initial learning rate. For all experimental methods, the parameter settings and code details are consistent.

4.2 Image Dataset

The experiments are trained and tested on LOL and MIT5K datasets. For the LOL dataset, 450 pairs of images are used for training and validation, 50 pairs of images are used for testing. For the MIT5K dataset, 4500 pairs of images are selected for training and validation, 500 pairs of images for testing. In addition, to verify the robustness of EDMFEN, we tested five benchmark datasets MEF [41], NPE [3], VV [42], DICM [43] and LIME [4].

4.3 Compare with the latest methods

To demonstrate the advantages of EDMFEN, we compared EDMFEN with traditional methods MF [6], NPE [3], SRIE [5], LIME [4] and deep learning KinD [9], LPNet [11], Zero-DCE++ [14], SCI [19], semantic [20], URetinex-Net [21] in each of the following aspects.

4.3.1 Quantitative comparison

EDMFEN is compared quantitatively with several current methods of image enhancement. Two evaluation indicators with references, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), are chosen to evaluate network performance. PSNR measures the noise level after image enhancement, and it is the mean square error between the enhanced image and the original image. SSIM indicates the structural similarity of images, including brightness, contrast, and structure. The use of PSNR and SSIM can provide a more comprehensive and objective evaluation of the quality of enhanced images. **Table 1** reflects that on the LOL and MIT5K datasets, our network achieves good performance, outperforming other methods and illustrating the superiority of our approach. The five benchmark datasets of DICM, LIME, MEF, NPE and VV are tested on LOL using a pre-training model, and Natural Image Quality Evaluation (NIQE) is chosen to evaluate the performance of the enhanced images, with lower values of NIQE indicating better performance. In **Table 2**, our model EDMFEN shows a significant advantage over the other methods. Specifically, our method works best on the DICM, LIME, MEF, NPE and VV datasets, according to NIQE. The generalization ability of EDMFEN model is well demonstrated.

4.3.2 Efficiency comparison

Table 1 summarizes the empirical results, generated on the LOL dataset with image size 512× 512, and processing speed is compared with metrics (Param, FLOPs, Time). The best results in black bold show that the proposed EDMFEN outperforms others in terms of Param and FLOPs metrics. The second best result was obtained on the Time metrics. MF, NPE, SRIE, and LIME do not involve network parameters, so we are not considering them. It is worth mentioning that LPNet is a lightweight network, and our metrics far surpass it. Due to Zero-DCE++ using only a 7-layer end-to-end network, we are weaker in terms of runtime. We can conclude that the proposed model, while simple, exhibits significantly faster speeds compared to competitors. This is attributed to fine-tuning small networks and substantial downsampling operations that greatly reduce the dimensionality of the inference feature space.

Table 1. Numerical results of LOL and MIT5K datasets

Methods	LOL					MIT5K	
	Param [M]	FLOPs [G]	Time [s]	PSNR	SSIM	PSNR	SSIM
MF	-	-	1.38	18.74	0.67	17.48	0.78
NPE	-	-	6.11	17.20	0.53	17.21	0.77
SRIE	-	-	2.45	14.25	0.54	19.48	0.79
LIME	-	-	3.71	17.37	0.55	14.54	0.75
KinD	8.49	7.44	1.75	20.38	0.80	21.84	0.79
LPNet	0.77	0.15	0.65	21.70	0.78	24.55	0.90
Zero-DCE++	1.29	0.35	**0.02**	17.42	0.76	20.21	0.80
SCI	3.14	0.069	0.17	20.83	0.82	20.46	0.83
Semantic	2.37	0.12	0.08	20.60	0.79	19.38	0.67
URetinex-Net	12.29	7.31	1.96	21.33	0.84	21.92	0.82
Ours	**0.48**	**0.05**	0.39	**23.08**	**0.82**	**25.27**	**0.92**

Table 2. Numerical results of the five benchmark datasets of DICM, LIME, MEF, NPE and VV

Methods	DICM	LIME	MEF	NPE	VV
	NIQE				
MF	3.49	4.07	3.49	4.11	2.93
NPE	3.45	4.11	3.53	4.15	3.03
LIME	3.47	4.09	3.56	4.19	2.79
SRIE	3.62	4.05	3.45	4.14	3.23
KinD	4.17	4.67	3.83	4.31	3.06
LPNet	3.94	4.35	4.26	4.16	3.52
Zero-DCE++	4.04	4.19	3.82	4.31	3.85
SCI	3.40	4.02	4.01	4.05	2.88
Semantic	3.32	4.05	3.91	4.49	3.41
URetinex-Net	3.36	3.96	3.52	4.08	2.90
Ours	**3.28**	**3.95**	**3.27**	**4.03**	**2.78**

4.3.3 Subjective evaluation

To validate the effectiveness of our proposed EDMFEN, the qualitative evaluation of the proposed method and different methods is conducted in detail based on the subjective visual comparison on two different datasets LOL and MIT5K. **Fig. 5** shows the images obtained with several methods of enhancement on the dataset LOL. SRIE produces an underexposed image, and LIME is a little more exposed, but still underexposed compared to the other methods. Meanwhile, SRIE and LIME do not recover the pattern details and the color of the images very well. The MF and NPE exposures are moderate, but the images they produce contain a lot of noise. Specifically, it is particularly noticeable in the cup mouth area in the zoomed-in region. The KinD method works better, but it may introduce some artifacts, which result in losing details of the image. In detail, it is particularly noticeable in the zoomed-in region. In terms of the image as a whole, the LPNet method has a small deficiency in color recovery compared to the KinD method, but as can be seen from the enlarged areas, there is a high level of detail reproduction in the image. In addition, the images generated by Zero-DCE have distortion due to overexposure. SCI and Semantic have higher brightness, the color of the former is lost and color shift occurs, the texture details of the latter image are insufficient, and the brightness of URetinex-Net is darker and local details are lost. Compared with these above methods, the color of the pattern in the picture generated by our proposed EDMFEN is closer to GT and more texture details are recovered in the picture.

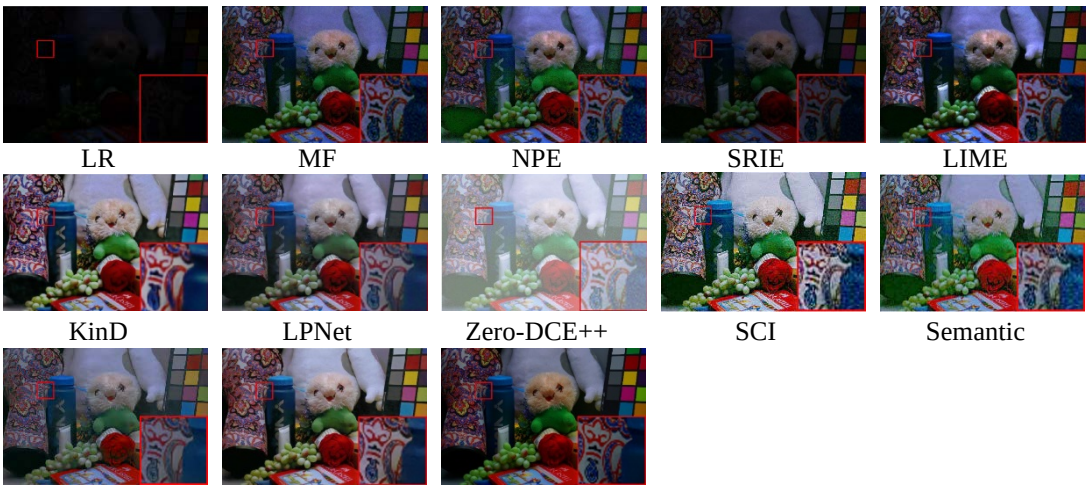

Fig. 5. The subjective visual in LOL dataset

In addition, **Fig. 6** shows the experimental results obtained on the dataset MIT5K. The MF and LIME are overexposed resulting in distorted images. The NPE and SRIE methods yield images with moderate exposure. Specifically, in the zoomed-in area, it can be seen that the roof has good detail recovery, but there is a slight color distortion. The KinD and LPNet methods yield more detailed images compared to NPE and SRIE, but the color recovery of the image in KinD is inadequate. The color recovery in LPNet is better, but there is some noise in the images as seen in the zoomed-in areas. The brightness of SCI and Semantic is higher, and the detail recovery of SCI is better, but there is color shift. From the enlarged view of Semantic, the texture details of the image are lost too much, and the brightness of URetinex-Net is darker, and the local details are lost too much. Compared with these selected methods, our proposed EDMFEN is closest to GT, with natural colors, rich detail, and high quality.

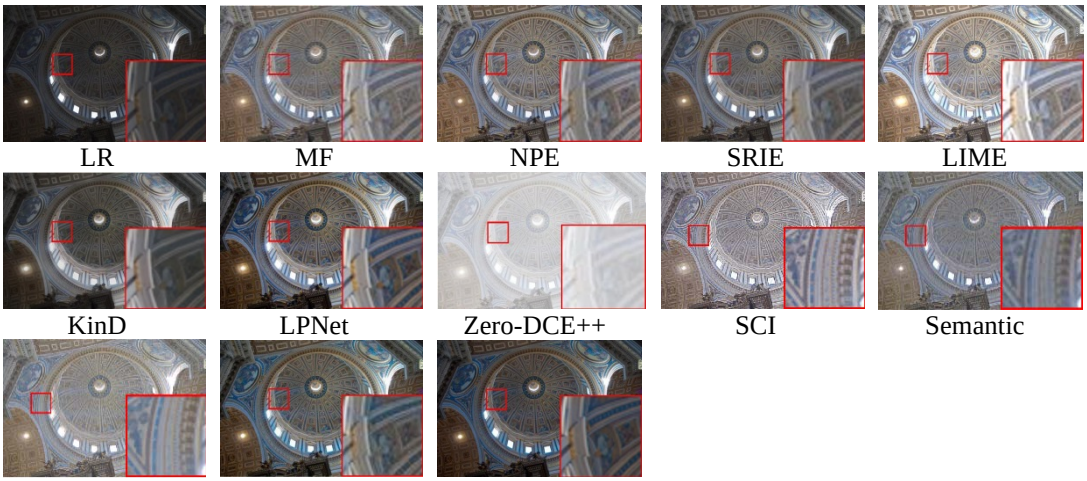

Fig. 6. The subjective visual in MIT5K dataset

To demonstrate the generalization ability and robustness of our proposed EDMFEN, we show the experiment results of our method on DICM, LIME, MEF, NPE and VV benchmark datasets in **Fig. 7**. Our method is generally well compatible with different datasets, achieving high-quality detail recovery and brightness enhancement.

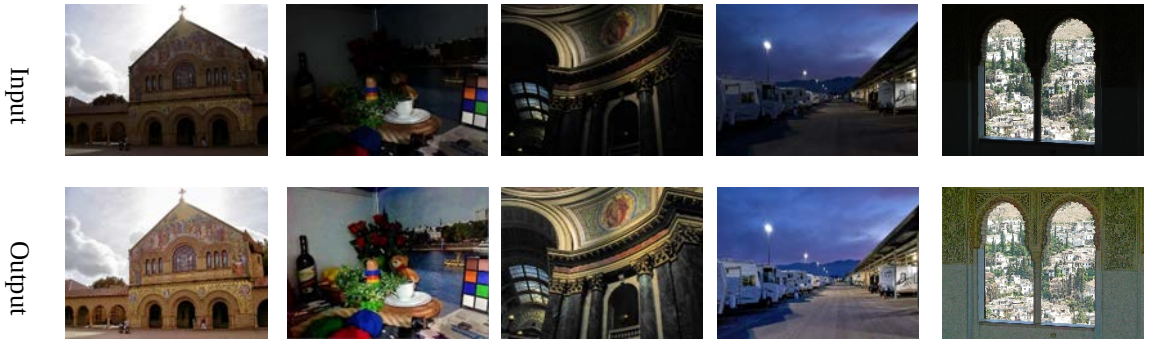

Fig. 7. The experiment results of our method in DICM, LIME, MEF, NPE and VV benchmark datasets

Comparing the proposed method with different methods on benchmark datasets can show the advantage on generalization ability of proposed method. Due to space constraint, we demonstrated the enhanced results from different methods on VV benchmark dataset, since VV is a higher resolution dataset with more details in the images. Our model also shows an obvious advantage over those selected methods according to the enhanced results. To show specific details, we framed a small part of the image and place a four-fold enlargement of it in the lower right corner of the image. In **Fig. 8**, the reference image is the LR image. According to all the results, it is found that each method had some degree of effect on the enhancement of the image. In the traditional four methods, the NPE enhanced image is highly exposed and noisy. The SRIE method is underexposed and does not have a significant effect on the restoration of hidden details in the image. MF has the most significant outcome, but it also

enhances the noise in the image. As can be seen in **Fig. 8**, the deep learning methods work better compared to the traditional methods. LPNet and KinD have similar results. Specifically, the KinD method is less well exposed, but LPNet has better luminance reproduction, and both have higher reproduction of hidden details in the image. Zero-DCE++ has a more pronounced brightness enhancement of the image, but it is overexposed and produces some distortion. There is less noise in the SCI image, but there is color shift in the image. Semantic has better detail recovery, but the image details are not rich enough. The brightness of URetinex-Net is darker and the image details are lost too much. Overall, our proposed method tends to be more stable in terms of brightness, saturation, and structure enhancement, and the restoration of hidden details in the image is noticeable. This demonstrates the good generalization ability of EDMFEN.

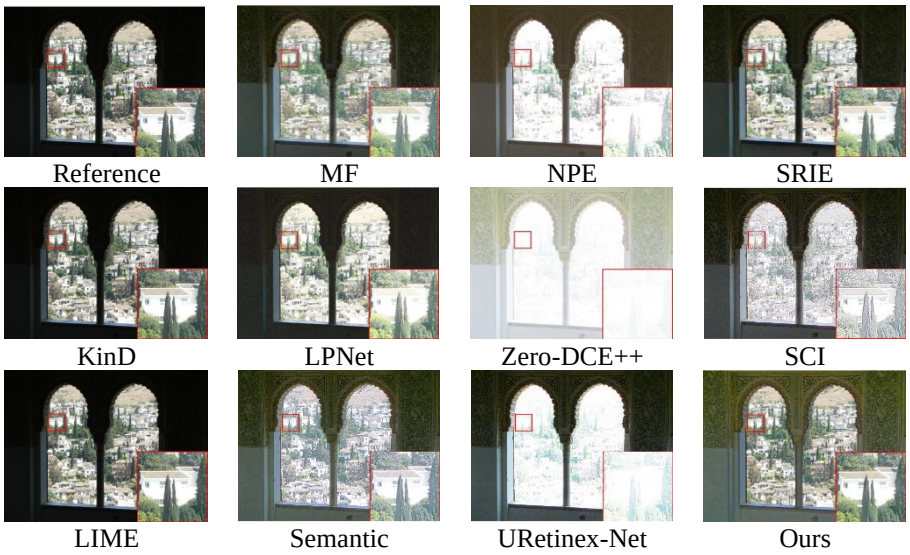

Fig. 8. VV benchmark dataset comparison chart

5. Ablation experiment

To demonstrate the importance of each module for EDMFEN, ablation experiments are carried out. We study the importance of EDM, MSFEM and attention mechanism module to the modular network framework. Ablation experiments are performed on LOL dataset by removing one module at a time.

Case1 represents the experiment without MSFEB, Case2 represents the experiment without attention module, Case3 represents the experiment without EDB, and Case4 represents the complete experiment EDMFEN. In **Table 3**, by comparing Case1 and Case4, using multi-scale feature extraction has a significant effect on LLIE, because multi-scale features can better represent the rich information in the image. In addition, the comparison results between Case 2 and Case 4 illustrate the impact of the attention module on the enhancement effect. Using the attention module can filter unnecessary redundant information in the image and make full use of key features. Finally, Case3 and Case4 illustrate that incomplete edge information also has a great impact on the results, and edge information has a great impact on the enhancement of image spatial details. By comparing these ablation experiments, we can conclude that the settings of each module have a certain impact on image enhancement, and the aggregation of all modules enables the model to achieve optimal performance.

Table 3. Ablation experiments on dataset LOL

Case	EDB	MSFEB	Attention	PSNR	SSIM
Case1	√			18.28	0.65
Case2	√	√		18.99	0.69
Case3		√	√	21.01	0.71
Case4	√	√	√	**23.08**	**0.82**

6. Conclusion

This paper proposed an EDMFEN method which consists of two modules, one module MSFEM for multi-scale features enhancement and the other module EDM for obtaining edge information of the image. During the multi-scale features enhancement process, edge information is continuously injected, and useful information generated from shallow features is sent directly to the end of the MSFEB. The framework produces more representative image features and aggregates the extracted shallow and deep information together, then combined with a spatial attention mechanism that allows features to be more focused on key spatial content, thereby improving the performance. Numerous comparative and ablation experiments have shown that our network is subjectively and objectively superior to the most advanced methods.

References

[1] P. M. Narendra, and R. C. Fitch, “Real-Time Adaptive Contrast Enhancement For Imaging Sensors,” *Airborne Reconnaissance III. International Society for Optics and Photonics*, vol. 137, pp. 31-37, 1978. Article (CrossRef Link)

[2] Ibrahim H, and Kong N S P, “Brightness preserving dynamic histogram equalization for image contrast enhancement,” *IEEE Trans. Consum. Electron*, vol. 53, no. 4, pp. 1752–1758, 2007. Article (CrossRef Link)

[3] S. Wang, J. Zheng, H. Hu, and B. Li, “Naturalness preserved enhancement algorithm for non-uniform illumination images,” *TIP*, vol. 22, no. 9, pp. 3538–3548, 2013. Article (CrossRef Link)

[4] Fu, X., Zeng, D., Huang, Y., Zhang, X. P., and Ding, X, “A Weighted Variational Model for Simultaneous Reflectance and Illumination Estimation," in *Proc. of CVPR*, pp. 2782-2790, 2016.

[5] Fu, X., Zeng, D., Huang, Y., Liao, Y., Ding, X., and Paisley, J, “A fusion-based enhancing method for weakly illuminated images," *Signal Processing*, vol. 129. pp. 82-96, 2016. Article (CrossRef Link)

[6] Guo, X., Y. Li, and H. Ling, "LIME: Low-Light Image Enhancement via Illumination Map Estimation,” *IEEE Trans. Image Process*, vol. 26, no. 2, pp. 982-993, 2017. Article (CrossRef Link)

[7] K. G. Lore, A. Akintayo, and S. Sarkar, “LLNet: A deep au-toencoder approach to natural low-light image enhancement,” *PR*, vol. 61, pp. 650–662, 2017. Article (CrossRef Link)

[8] Wei, C., Wang, W., Yang, W., and Liu, J., "Deep retinex decomposition for low-light enhancement,” *arXiv preprint arXiv*, 2018. Article (CrossRef Link)

[9] Zhang, Y., J. Zhang, and X. Guo, "KinDling the Darkness: A Practical Low-light Image Enhancer,” in *Proc. of the 27th ACM International Conference on Multimedia*, pp. 1632–1640, 2019. Article (CrossRef Link)

[10] R. Wang, Q. Zhang, C.W. Fu, X. Shen, W.S. Zheng, and J. Jia, “Underexposed photo enhancement usign deep illumi-nation estimation,” in *Proc. of CVPR*, pp. 6849–6857, 2019.

[11] Li, J., Li, J., Fang, F., Li, F., and Zhang, G., "Luminance-aware Pyramid Network for Low-light Image Enhancement," *IEEE Trans Multimedia*, pp. 3153-3165, 2020.
[12] Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., and Wang, Z, "EnlightenGAN: Deep Light Enhancement Without Paired Supervision," *IEEE Trans. Image Process*, vol. 30, pp. 2340-2349, 2021. Article (CrossRef Link)
[13] Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., and Cong, R., "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. of CVPR*, pp.1780–1789, 2020.
[14] Li, C., Guo, C., and Loy, C. C, "Learning to enhance low-light image via zero-reference deep curve estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 8, pp. 4225-4238, 2022. Article (CrossRef Link)
[15] Pan W, Zhao Z, Huang W, et al., "Video moment retrieval with noisy labels," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1-13, 2022. Article (CrossRef Link)
[16] Matsuo Y, LeCun Y, Sahani M, et al., "Deep learning, reinforcement learning, and world models," *Neural Networks*, vol. 152, pp. 267-275, 2022. Article (CrossRef Link)
[17] Ma, L., Zheng, Y., Zhang, Z., Yao, Y., Fan, X., and Ye, Q., "Motion Stimulation for Compositional Action Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 5, pp. 2061-2074, May 2023. Article (CrossRef Link)
[18] Fu L., Zhang D., Ye Q., "Recurrent thrifty attention network for remote sensing scene recognition," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 10, pp. 8257-8268, 2021. Article (CrossRef Link)
[19] Ma L, Ma T, Liu R, et al., "Toward fast, flexible, and robust low-light image enhancement," in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5637-5646, 2022.
[20] Zheng S, Gupta G., "Semantic-guided zero-shot learning for low-light image/video enhancement," in *Proc. of the IEEE/CVF Winter conference on applications of computer vision*, pp. 581-590, 2022.
[21] Wu W, Weng J, Zhang P, et al., "Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement," in *Proc. of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5901-5910, 2022.
[22] Fan C M, Liu T J, Liu K H., "Half wavelet attention on M-Net+ for low-light image enhancement," in *Proc. of 2022 IEEE International Conference on Image Processing (ICIP)*, IEEE, pp. 3878-3882, 2022. Article (CrossRef Link)
[23] Wang Y, Wan R, Yang W, et al., "Low-light image enhancement with normalizing flow," in *Proc. of the AAAI conference on artificial intelligence*, Vol. 3, no. 36, pp. 2604-2612, 2022. Article (CrossRef Link)
[24] Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., and Keutzer, K., "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," *arXiv preprint arXiv*, 2016. Article (CrossRef Link)
[25] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., and Adam, H, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv preprint arXiv*, 2017. Article (CrossRef Link)
[26] Zhang, X., Zhou, X., Lin, M., and Sun, J., "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices," in *Proc. of CVPR*, pp.6848-6856, 2017.
[27] Ding, L., and Goshtasby, A., "on the Canny Edge Detector," *Pattern recognition*, vol.34, no.3, pp. 721-725, 2001. Article (CrossRef Link)
[28] Roberts R E., Attkisson C C., Rosenblatt A., "Prevalence of psychopathology among children and adolescents," *American journal of Psychiatry*, vol. 155, no. 6, pp. 715-725, 1998. Article (CrossRef Link)
[29] Prewitt J M S., "Object enhancement and extraction," *Picture processing and Psychopictorics*, pp. 75-149, 1970. Article (CrossRef Link)
[30] Raman, M., and H. Aggarwal, "Study and Comparison of Various Image Edge Detection Techniques," *International Journal of Image Processing*, vol. 3, no. 1, pp. 1-11, 2009.

[31] Hu, J., Shen, L., and Sun, G, “Squeeze-and-Excitation Networks," in *Proc. of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7132-7141, 2018. Article (CrossRef Link)
[32] Wang, X., Girshick, R., Gupta, A., and He, K, “Non-local Neural Networks," in *Proc. of CVPR*, pp.7794-7803, 2018.
[33] Cao, Y., Xu, J., Lin, S., Wei, F., and Hu, H, "GCNet: Non-Local Networks Meet Squeeze-Excitation Networks and Beyond," in *Proc. of IEEE ICCV*, 2019.
[34] Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., and Lu, H, “Dual Attention Network for Scene Segmentation," in *Proc. of CVPR*, pp.3146-3154, 2019.
[35] Xie, S., and Tu, Z, “Holistically-Nested Edge Detection," in *Proc. of ICCV*, pp.1395-1403, 2015.
[36] Zheng H., Chen J., Chen L., et al., “Feature enhancement for multi-scale object detection,” *Neural Processing Letters*, vol. 51, pp. 1907-1919, 2020. Article (CrossRef Link)
[37] Pandey G., Ghanekar U., “Single image super-resolution using multi-scale feature enhancement attention residual network,” *Optik*, vol. 231, pp. 166359, 2021. Article (CrossRef Link)
[38] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., and Hu, Q, “Supplementary material for ‘ECA-Net: Efficient channel attention for deep convolutional neural networks,” Tech. Rep.
[39] Wei, C., Wang, W., Yang, W., and Liu, J., "Deep retinex decomposition for low-light enhancement,” *arXiv preprint arXiv*, 2018. Article (CrossRef Link)
[40] Bychkovsky, V., Paris, S., Chan, E., and Durand, F, “Learning photographic global tonal adjustment with a database of input/output image pairs,” in *Proc. of CVPR*, pp.97–104, 2011.
[41] Lee, C., Lee, C., Lee, Y. Y., and Kim, C. S, “Power-constrained contrast enhancement for emissive displays based on histogram equalization,” *TIP*, vol. 21, no. 1, pp. 80– 93, 2012. Article (CrossRef Link)
[42] Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., and Cong, R, “Zero-reference deep curve estimation for low-light image enhancement,” in *Proc. of CVPR*, pp.1780–1789, 2020.
[43] Lee, C., Lee, C., and Kim, C. S, “Contrast enhancement based on layered difference representation of 2d histograms,” *TIP*, vol. 22, no. 12, pp. 5372–5384, 2013. Article (CrossRef Link).

Canlin Li received the B.S. degree in computer science from National University of Defense Technology, China, in 1998, and the M.S. degree in computer science and technology from Zhejiang University, China, in 2004, and the Ph.D. degree in computer science and engineering from Shanghai Jiao Tong University, China, in 2010. He is currently an Associate Professor and M.S. Tutor with School of Computer Science and Technology, Zhengzhou University of Light Industry. His research interests include image processing, pattern recognition, and artificial intelligence, visual media computing.

Shun Song received the B.S. degrees from Zhengzhou University of Light Industry. Zhengzhou, China, in 2020, and the M.S degree with Zhengzhou University of Light Industry. Zhengzhou, China, in 2023. Current teacher at Henan Vocational University of Science and Technology. Her research interests include image processing, and deep learning.

Pengcheng Gao received the B.S. degrees from xinlian college of Henan Normal University, Henan, China, in 2019. He is currently working toward the M.S degree with Zhengzhou University of Light Industry, Henan, China. His current research interests include image processing and artificial intelligence.

Wei Huang (Member, IEEE) was born in Henan, China, in 1982. He received the Ph.D. degree in pattern recognition and intelligence systems from the Nanjing University of Science and Technology (NUST), Nanjing, China, in 2015. He is currently an Associate Professor with School of Computer Science and Technology, Zhengzhou University of Light Industry. His current research interests include pansharpening, hyperspectral classification, image processing and machine learning.

Lihua Bi received the M.S. degree from Guangxi University for Nationalities of China, Nanning, China. She is currently working at Zhengzhou University of Light Industry, Zhengzhou, China. Her main research interests include image processing, data processing and pattern recognition.