



End-to-end residual attention mechanism for cataractous retinal image dehazing

Defu Qiu^{a,b}, Yuhu Cheng^{a,b}, Xuesong Wang^{a,b,*}

^a Engineering Research Center of Intelligent Control for Underground Space, Ministry of Education, China University of Mining and Technology, Xuzhou, 221116, China

^b School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China

ARTICLE INFO

Article history:

Received 23 October 2021

Revised 20 March 2022

Accepted 24 March 2022

Keywords:

Cataractous

Dehaze

Retinal image

Skip connection

Residual attention mechanism

ABSTRACT

Background and objective: Cataract is one of the most common causes of vision loss. Light scattering due to clouding of the lens in cataract patients makes it extremely difficult to image the retina of cataract patients with fundus cameras, resulting in a serious decrease in the quality of the retinal images taken. Furthermore, the age of cataract patients is generally too old, in addition to cataracts, the patients often have other retinal diseases, which brings great challenges to experts in the clinical diagnosis of cataract patients using retinal imaging.

Methods: In this paper, we present the End-to-End Residual Attention Mechanism (ERAN) for Cataractous Retinal Image Dehazing, which it includes four modules: encoding module, multi-scale feature extraction module, feature fusion module, and decoding module. The encoding module encodes the input cataract haze image into an image, facilitating subsequent feature extraction and reducing memory usage. The multi-scale feature extraction module includes a hole convolution module, a residual block, and an adaptive skip connection, which can expand the receptive field and extract features of different scales through weighted screening for fusion. The feature fusion module uses adaptive skip connections to enhance the network's ability to extract haze density images to make haze removal more thorough. Furthermore, the decoding module performs non-linear mapping on the fused features to obtain the haze density image, and then restores the haze-free image.

Results: The experimental results show that the proposed method has achieved better objective and subjective evaluation results, and has a better dehazing effect.

Conclusion: We proposed ERAN method not only provides visually better images, but also helps experts better diagnose other retinal diseases in cataract patients, leading to better care and treatment.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

Cataract is an age-related disease, and with the aging of the world population, the prevalence of cataract may continue to increase [1]. It is difficult for human experts or computer-aided diagnosis systems to diagnose cataracts based on blurred retinal images. Although the blurred image of cataract retinal image obtained by scanning laser confocal imaging method is small, and this imaging method can provide better retinal imaging results for cataract patients, compared with fundus camera, this imaging device is expensive and not popular, especially in rural areas [2–3]. Therefore, the development of an algorithm to enhance the contrast and readability of cataract retinal images has important practical significance and clinical value, especially for retina-related health screen-

ing items, such as screening for diabetic retinopathy [4]. Although there are retinal image enhancement methods based on contrast constrained adaptive histogram equalization and HSV color space brightness gain matrix to enhance the contrast of cataract retinal images, these methods are independent of the degeneration mechanism of retinal images, and are not specifically designed for blurred retinal images of cataracts, which greatly reduces the enhancement effect and universality [5–6].

The existing single-image dehaze methods can be briefly divided into two categories: image dehaze methods based on traditional prior knowledge [7], and image dehaze methods based on deep learning. Before 2016, image dehaze algorithms based on traditional prior knowledge dominated the mainstream. Fattle et al. [8] proposed a physical model method for estimating the reflectivity of the scene based on prior knowledge and principal component analysis. He et al. [9] proposed a dark channel prior method (DCP) by analyzing the different performance of the clear image

* Corresponding author.

E-mail address: wangxuesong@cumt.edu.cn (X. Wang).

and the haze image in the color channel, and achieved a significant improvement in the dehazing effect. When the color of the ground and other objects is similar to the atmospheric light intensity, the restored image will produce serious color distortion. In order to solve the problems of the dark channel method, many improved DCP-based algorithms have been proposed and achieved good dehazing effects. However, algorithms based on traditional prior knowledge are strongly dependent on the choice of prior knowledge. When the imaging environment changes, such methods often cannot handle all haze images well.

With the development of deep learning and the proposal of large-scale benchmark datasets, increasing number of scholars try to use deep learning tools to directly predict atmospheric transmission maps or ultimately restore images [10–11]. Cai et al. [12] proposed the DehazeNet method, using a convolutional neural network to estimate the atmospheric transmission map of the haze image, and achieved a certain performance improvement. Ren et al. [13] proposed a multi-scale deep dehaze network (MSCNN), which used a multi-scale neural network to estimate the transmission map from coarse to fine. Li et al. [14] modified the classic atmospheric scattering model and used a lightweight convolutional neural network (AOD-Net) to directly predict the final restored image, avoiding the performance degradation caused by incorrectly estimating the atmospheric transmission map. Zhang et al. [15] proposed a dense pyramid dehaze network (DCPDN), which embedded the formula of the atmospheric scattering model into the constructed neural network, and estimated the two parameters of the atmospheric light value and the transmission map in the network. Liu et al. [16] designed a multi-scale neural network (GridDehazeNet) based on the attention mechanism, which realized multi-scale feature fusion and further improved the effect of single image dehaze. Chen et al. [17] proposed a gated context aggregation network (GCANet), introduced smooth hole convolution and designed a gated fusion network.

However, these algorithms cannot be directly applied to the cataract retinal images. One of the main reasons is: for supervised deep learning, we lack the blurred retinal images of cataract patients and the corresponding clear retinal images after cataract surgery. This is the biggest obstacle that hinders the application of deep convolutional neural networks in cataract retinal image dehazing [18]. In addition, even if we can obtain such image pairs, it is difficult to ensure that the retinal images taken before and after the operation are completely matched at the pixel level. At the same time, due to the poor quality of cataract retinal image and the distortion of the optical path, achieving the registration of this type of image pair at the pixel level also faces a huge challenge.

In addition, Zhu et al. [19] proposed a generative adversarial networks (GAN) structure called CycleGAN, which realizes the training of unpaired image data by applying the defined cyclic consistency loss, and is widely used in natural scene image dehaze without paired training images. However, this loss focuses more on the main and global features of the image, which is different from images in natural scenes, and the retina images often contain many small-scale features [20]. Therefore, when the CycleGAN network is applied to retinal imaging of cataracts, many obvious vascular artifacts will often be produced in the result after dehaze.

2. Relate work

2.1. Haze image imaging model

In the field of natural scene image, the imaging model of haze image is as follows [21–22]:

$$I(x) = D(x)t(x) + L[1 - t(x)], \quad (1)$$

where x represents the pixel of the image, $I(x)$ represents the natural scene image of the haze image, $D(x)$ represents the clear image that needs to be restored, $t(x)$ represents the transmittance of the system, and L represents the global atmospheric light composition.

In the field of cataract retinal imaging, the corresponding imaging optical model is as follows [23]:

$$I(x) = \alpha \cdot L \cdot r(x) \cdot t(x) + L[1 - t(x)], \quad (2)$$

where x represents the pixels of the image, $I(x)$ represents the retinal image of the cataract, L is the fixed illumination of the camera, $r(x)$ is the retinal system including all the distortions related to the eye optics and the fundus camera, as well as the influence of the slight curvature of the retinal plane, $t(x)$ is the transmission of the system rate, and α represents the attenuation of retinal illumination caused by cataracts.

In the above-mentioned model, the solution of the α value is relatively cumbersome. On the one hand, the coefficient needs to be calculated using the clear image of the patient before the cataract. However, in practice, it is difficult to obtain the image. At the same time, under different lighting conditions, the obtained α value may also be different. On the other hand, the value of α is a constant, and the magnitude of the value has a linear effect on the output result, so we can ignore the effect of α in the solution process by multiplying the final output result by a linear coefficient. Considering the above factors, Cheng et al. [24] then proposed a reduced model, that is, the above formula is completely consistent with the imaging model of haze images in natural scenes. Therefore, the dehaze of the cataract retinal image has a certain degree of compatibility with the dehaze of the natural scene image.

2.2. Residual learning

In the process of the development and research of convolutional neural networks, due to the increasing number of classification problems and the increasing difficulty of recognition, people have increasingly higher requirements for the depth of convolutional neural network layers. The residual network [25], born in 2015, spans two neural layers through a skip connection. As shown in Fig. 1, this skip connection can achieve this multi-layer identity mapping, and learn the following layers of the deep network into an identity mapping. Hence, the network can maintain its feature expression ability while widening its depth, and subtly solve the problem of gradient disappearance or gradient explosion caused by the deepening of the number of layers. Assuming that one layer of the network is $H(x)$, and the residual block of the residual network can be expressed as:

$$H(x) = F(x) + x, \quad (3)$$

where x is the parameter value passed by the network, $F(x)$ is the residual value, and $H(x)$ is the predicted value. Only $F(x) = 0$ can form an identity mapping $H(x) = 0$, and the residual structure expression is:

$$x_{l+1} = x_l + F(x_l, W_l), \quad (4)$$

Through recursion, the expression of neurons in any depth layer can be obtained:

$$x_L = x_l + \sum_{i=1}^{L-1} F(x_i, W_i), \quad (5)$$

The structure of the neural network is expressed as:

$$y = F(x, W_i) + W_s x, F = W_2 \delta(W_1 x), \quad (6)$$

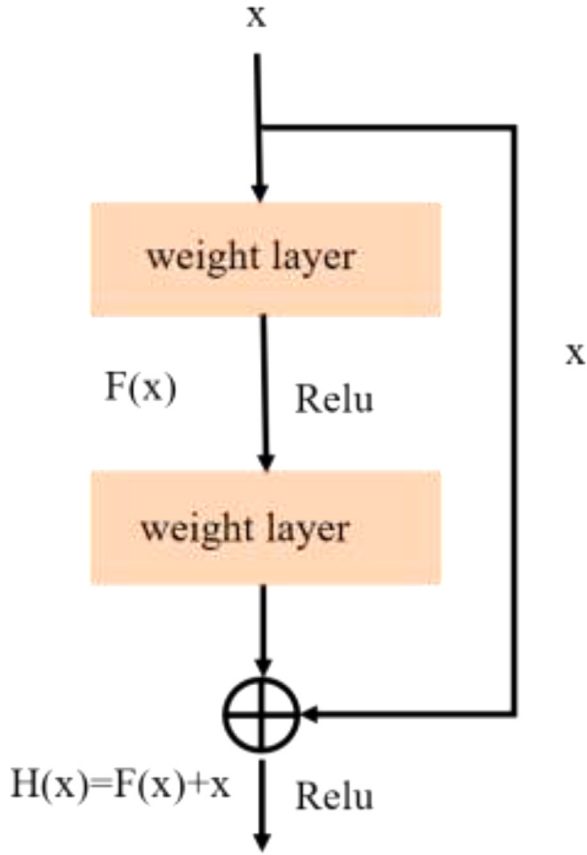


Fig. 1. Structural framework of residual learning.

3. Methodology

Owing to the encoding-decoding structure, we have achieved great success in low-level vision tasks such as image super-resolution. This paper adds the encoding-decoding structure to the end-to-end residual attention thought as the basic framework of the model. As shown in Fig. 2, the network consists of four parts: an encoding module, a multi-scale feature extraction module, a feature fusion module, and a decoding module. The input of the

model is a haze image, which passes through four modules in turn to get the final dehaze image.

3.1. Encoding module

The encoding module converts the input retinal haze map into a feature image, as shown in Fig. 2, which consists of two convolution blocks and a down-sampling convolution block, the size of the convolution kernel is 3×3 , and the step size of the convolution layer is 1, the fill is set to 1, and the number of input channels is 4. The step size of the down-sampling convolution block is 2, and the number of output channels is 64. The encoding module is expressed as:

$$F_0(x) = \text{Conv}_3(\text{Conv}_2(\text{Conv}_1(I(x)))) \quad (7)$$

where $F_0(x)$ represents the encoded feature image.

3.2. Multi-scale feature extraction module

In order to expand the receptive field and extract features of different scales without reducing the resolution, we designed a multi-scale feature extraction module. As shown in Fig. 2, the module uses three groups of feature extraction modules with different dilated rates to extract different receptive field information to obtain feature maps of different scales. The multi-scale feature extraction module is expressed as:

$$\begin{cases} F_1(x) = \text{FEM}_1(F_0(x)) \\ F_2(x) = \text{FEM}_2(F_1(x)), \\ F_3(x) = \text{FEM}_3(F_2(x)) \end{cases} \quad (8)$$

where $F_1(x)$, $F_2(x)$, and $F_3(x)$ respectively represent the extracted features of different scales. Each feature extraction module (FEM_i , $i = 1, 2, 3$) includes one multi-scale feature extraction module (MFEM_i), two residual blocks (RB_1 , RB_2) with the same structure, and one adaptive skip connection (ASC_i). The feature extraction module is expressed as:

$$\text{FEM}_i(x) = \text{ASC}_i(\text{RB}_2(\text{RB}_1(\text{FEM}_i(F_{i-1}(x)))) + F_{i-1}(x), \quad (9)$$

where $F_{i-1}(x)$ represents the feature of the current input.

The feature extraction module is shown in Fig. 3, which includes two cavity convolution modules and skip connections. Firstly, we use a separate shared convolutional layer [27] with a

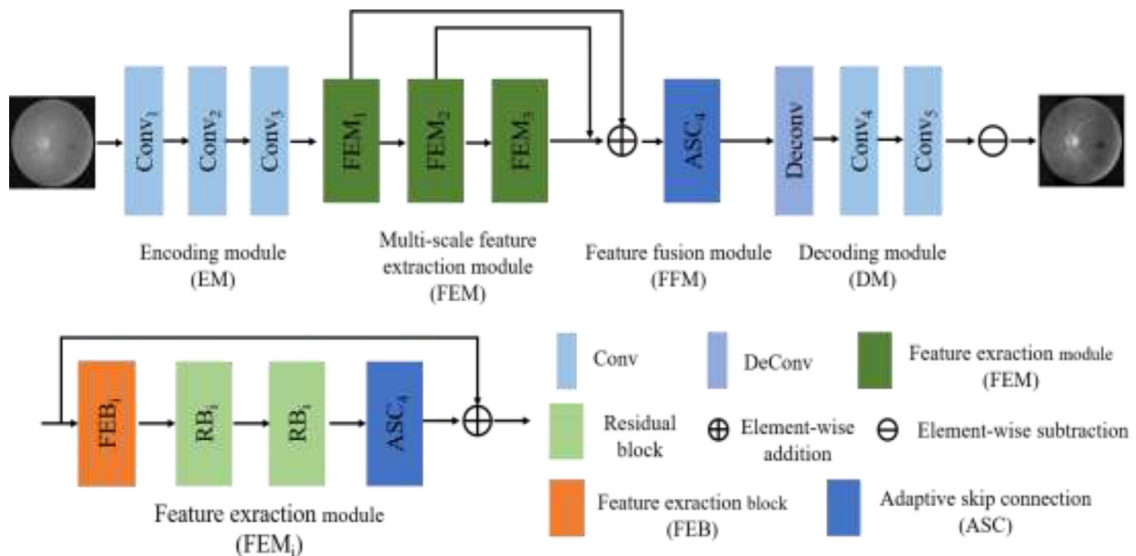


Fig. 2. Structural framework of ERAN for cataractous retinal image dehazing processing.

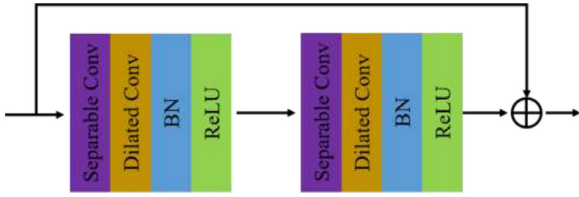


Fig. 3. Structure diagram of feature extraction module (FEM).

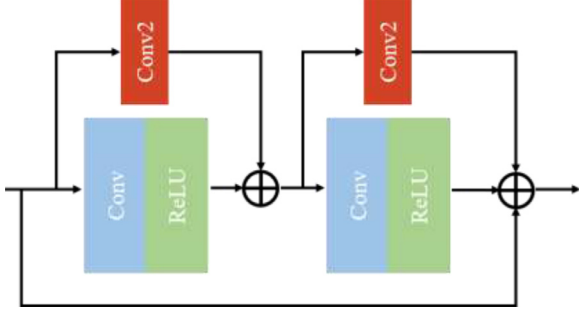


Fig. 4. Structure diagram of residual block (RB).

convolution kernel size of $(2r-1) \times (2r-1)$, a step size of 1, and an output channel of 64 to enhance the dependency between the input units and avoid grid artifacts. Then we use a convolution kernel with a size of 3×3 , a step size of 1, and a cavity convolution layer with an output channel of 64 to expand the receptive field. Finally, the batch normalization (BN) [28] layer and the ReLU activation function layer are used to accelerate the model convergence. At the same time, skip connections are used to avoid the disappearance of the gradient. The dilated ratio of FEB_i is 1, 3, and 5. The feature extraction module is expressed as:

$$FEB_i(x) = FE_2(FE_1(F_{i-1}(x))) + F_{i-1}(x), \quad (10)$$

where $FE_i(\cdot)$, $i = 1, 2$ represents the hole convolution function.

As shown in Fig. 4, the residual block adopts a skip connection method to improve the learning ability of the network, and the residual block contains two convolutional blocks and local skip connections. In additions, the convolutional block contains a convolutional layer with a convolution kernel size of 3×3 , a step size of 1, and an output channel of 64, and a ReLU activation function layer. In order to further extract the features of the current scale, the feature extraction module is followed by two residual blocks to ensure the integrity of the features of the current scale. The residual block RB(x) is expressed as:

$$RB_i(x) = CB_2(CB_1(F_{i-1}(x))) + F_{i-1}(x), \quad (11)$$

where $CB_i(\cdot)$, $i = 1, 2$ represents the hole convolution function.

The adaptive skip connection structure is shown in Fig. 5. Firstly, perform global average pooling without dimensionality reduction. Secondly, a convolutional layer with a convolution kernel size of 3×3 is adaptively determined to realize cross-channel in-

formation interaction, and then the Sigmoid function is used to obtain the weighted value of the feature map [29]. Finally, use the feature map weight value to adjust the input feature map, and output the weighted feature map. The adaptive skip connection is expressed as:

$$ASC_i(x) = \delta(\text{Conv}(g_i(F_{i-1}(x)))) \otimes F_{i-1}(x), \quad (12)$$

where $g_i(\cdot)$ represents the global average pooling function, $\text{Conv}(\cdot)$ represents the convolution function with a convolution kernel size of 3×3 , and δ represents the Sigmoid function.

3.3. Feature fusion module

In order to obtain a large enough receptive field and obtain more context information, the down-sampling operation is often used in the encoder, but the down-sampling operation often brings about the problem of loss of the detailed information of the feature map. In the ERAN network, the encoder and the decoder often adopt a skip connection method to enhance the features in the decoder. There are two main forms of skip connection. One is to merge the feature information in the encoder with the feature information in the corresponding decoder in the form of channel addition. The other is to stack the feature information in the encoder with the feature information in the corresponding decoder, and then use the stacked features for subsequent processing [26].

Both two skip connection methods have the effect of supplementing details, but at the same time they also introduce a lot of useless information. In response to this problem, we designed an adaptive skip connection method, as shown in Fig. 4. This method mainly includes the following four steps.

- 1) Preliminary integration of features. The feature information F_E in the encoder and the feature information F_D in the corresponding decoder are preliminarily merged in the manner of channel addition.
- 2) Feature extraction. A convolutional layer with a convolution kernel size of 3 is used to extract the fused features.
- 3) Channel weighting. The channel attention mechanism is used to weight each channel of the extracted features to further extract useful feature information.
- 4) Enhanced features. The channel-weighted feature is fused with the feature information F_D in the decoder in step 1 in a channel addition manner to obtain an enhanced feature F_S .

In order to effectively fuse multi-scale features, the feature fusion module first stacks feature of different scales. Then the stacked multi-scale features are compressed through a convolutional layer with a convolution kernel size of 3×3 . The number of input channels is 192 and the number of output channels is 64. The stacked and compressed features are expressed as:

$$F_{cat}(x) = \text{Conv}_{cat}(\text{cat}(F_1(x), F_2(x), F_3(x))), \quad (13)$$

where Conv_{cat} represents the compressed convolution function, and $\text{cat}(\cdot)$ represents the splicing function. After stacking and

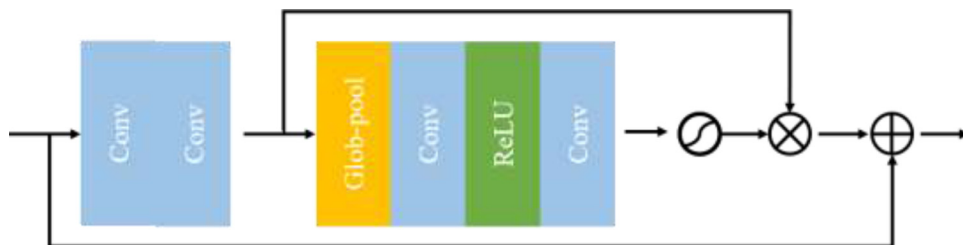


Fig. 5. Structure diagram of adaptive skip connection module (ASC).

compressing features of different scales, the high-efficiency channel attention mechanism is used again to dynamically adjust its channel weights, learn rich context information, and suppress redundant information, and enhance the network's ability to extract haze density images to make dehaze more thorough. The feature fusion module is expressed as:

$$F_c(x) = ASC_4(F_{cat}(x)), \quad (14)$$

3.4. Decoding module

In the ERAN network, the decoding module and the encoding module are symmetrical. The decoding module contains 1 deconvolution block and two convolution blocks, and the padding is set to 1. The size of the convolution kernel of the deconvolution block is 4×4 , the step size is 2, and the number of input channels is 64. The size of the convolution kernel of one convolution kernel block is 3×3 , and the step size is 1, and the size of the convolution kernel of the other convolution kernel block is 1×1 , and the output channel is 3. The decoding module performs nonlinear mapping on the fused features to obtain the haze density image, and the decoding module is expressed as:

$$I_h(x) = Conv_5(Conv_4(DeConv(F_c(x)))) \quad (15)$$

Finally, the restored image without haze is expressed as:

$$J(x) = I(x) - I_h(x), \quad (16)$$

3.5. Loss function

In order to obtain a good model, it is necessary to design an appropriate loss function to measure the degree of agreement between the predicted result of the model and the true value in the training data. Since the current evaluation index for image dehaze is mainly the objective evaluation index PSNR, most dehaze networks use the Mean Square Error (MSE) loss [30] or a combination of MSE loss, perception loss, and generation confrontation loss as the loss of the model. Function to obtain a better quality of objective evaluation. In additions, in order to reduce the amount of calculation and improve the training speed of the model, we use a simple mean square error loss to supervise each pixel of the output image. The loss function of the model is:

$$L_{MSE} = \frac{1}{N} \sum_{x=1}^N \sum_{i=1}^3 \|J_i(x) - I_i(x)\|_2^2, \quad (17)$$

where $J_i(x)$ and $I_i(x)$ respectively represent the pixel value of the i th color channel of pixel x in the dehaze image and the real haze-free image, and N is the total number of pixels in a single channel.

4. Experiment and analysis

4.1. Dataset and training details

In order to compare with other models, the training of the model is carried out on the RESIDE public dataset [31], which consists of a synthetic indoor haze image dataset and a synthetic outdoor haze image dataset. For indoor image dehaze, we directly use the Indoor Training Dataset (ITS) in RESIDE as the training dataset. The dataset consists of 1339 original images and 13,990 haze images. The haze images are composed of original images. The image is synthesized using an atmospheric scattering model, where the global atmospheric light value ranges from 0.8 to 1.0, and the atmospheric scattering parameter ranges from 0.04 to 0.2. For outdoor image dehaze, some of the original images in the Outdoor Training Dataset (OTS) in RESIDE have the same labels as the test

dataset. These images may affect the experimental results and cannot truly reflect the performance of the model. Therefore, we used the GridDehaze network [16] to train the OTS cleaned dataset. The dataset contains a total of 296,695 synthetic haze images and 8447 original images.

In addition, we used the Synthetic Objective Testing Dataset (SOTS) in RESIDE as the test dataset to evaluate the proposed model objectively and subjectively. The test data includes a total of 500 indoor composite images and 500 outdoor composites. image. In the selection of evaluation indicators, we use are the common peak signal to noise ratio (PSNR) [32] and structural similarity index (SSIM) [33] of retinal image dehazing. Based on the difference between the pixels of two images, the PSNR provides an objective index to measure the similarity of two images, which is widely used in the evaluation of the quality of SR reconstructed images. The SSIM combines the brightness, contrast, and structure of the image to evaluate the similarity of two images. It is an evaluation index that is more in line with the human visual system, and the larger the value, the more similar the two images, the maximum value of SSIM is 1.

$$PSNR = 10 \log \frac{255^2 \times W \times H \times C}{\sum_{i=1}^W \sum_{j=1}^H \sum_{k=1}^C [\bar{x}(i, j, k) - x(i, j, k)]^2}, \quad (18)$$

where W, H, C respectively represent the width, length, and number of channels of the image, \bar{x} represents the SR image, x represents the original image.

$$SSIM(X, Y) = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)}, \quad (19)$$

where μ_X, μ_Y represent the mean value of X , and Y , respectively, σ_X, σ_Y represent the value of image X and Y , respectively. σ_{XY} represents the covariance X , and Y , C_1, C_2 are constants.

Moreover, in order to verify the effect of our algorithm on retinal image reconstruction, we reconstructed it on the public Fundus Image Registration (FIRE) dataset [34]. The FIRE is a retinal fundus image dataset, which contains 129 fundus retinal images, which are combined into 134 image pairs by different features. Moreover, the image was jointly constructed by Papageorgiou Hospital and Aristotle University of Thessaloniki. The Nidek AFC-210 fundus camera was used to capture the image. The camera acquires an image with a resolution of 2912×2912 pixels and a FOV of 45° in the x and y dimensions.

In this paper, we used the Pytorch, the operating environment is Ubuntu 18.04, and the model training is done on NVIDIA 3080 GPU. During training, the image is processed in a random cropping manner. The image block size of the model input is 240×240 . The ADAM optimizer with a batch size of 24 is used to train the model. The momentum parameters β_1 and β_2 adopt the default values of 0.9, and 0.009, respectively. The initial learning rate of the model is 0.001. For ITS training, 100 iterations are performed, and the learning rate of every 20 iterations decays to half of the previous. For OTS training, 10 iterations were performed, and the learning rate was reduced to half of the previous one every two iterations.

4.2. Qualitative results

In order to fully verify the effectiveness of the proposed model, we compared several advanced traditional image removal algorithms and deep learning-based image removal algorithms, including DCP, AOD-Net, DehazeNet, PFFNet, and GCANet methods. Table 1 is the value obtained by the objective evaluation indicators of each algorithm under the SOTS test dataset. Among them, the larger the value of PSNR and SSIM, the better the image effect.

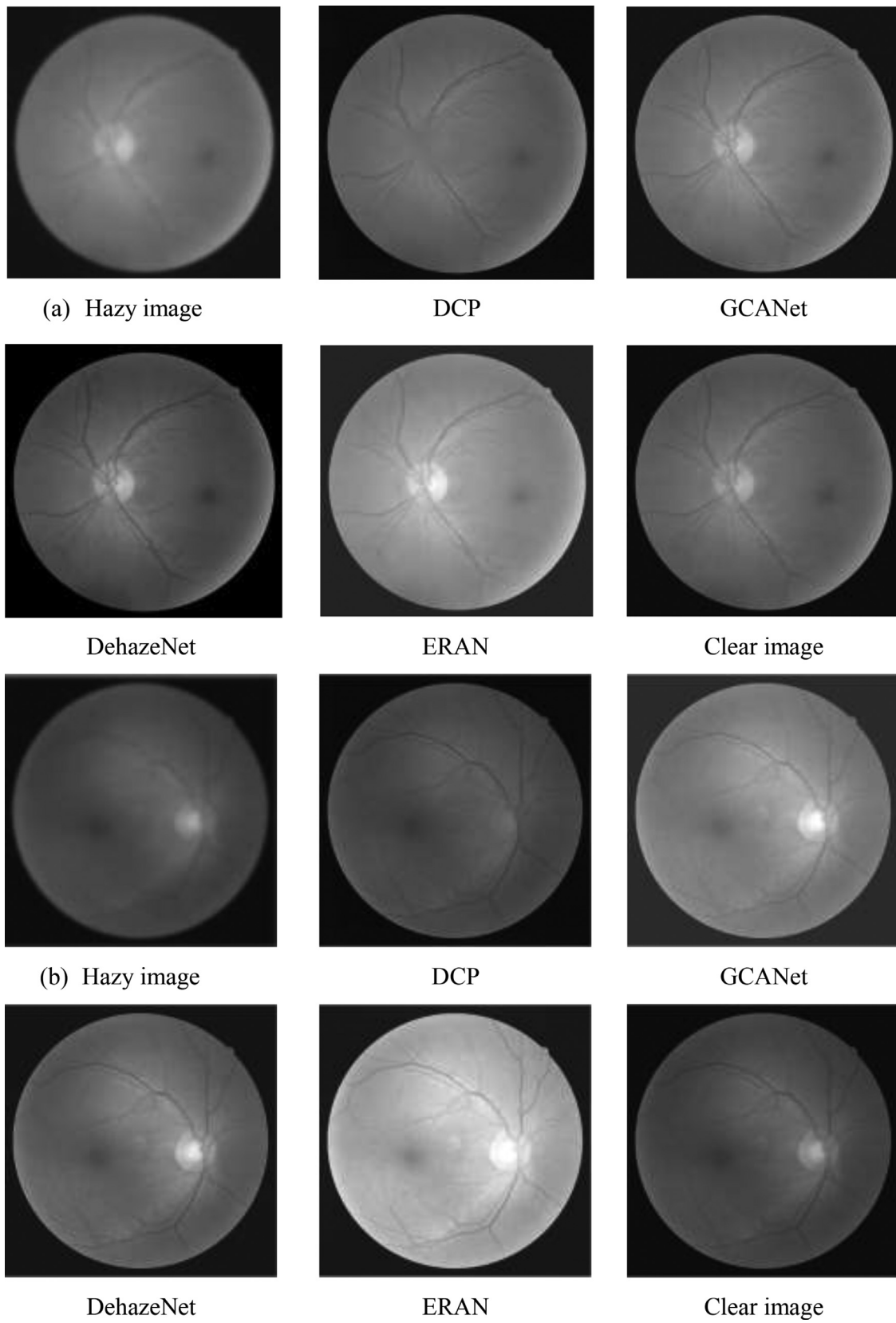


Fig. 6. Experimental results of the cataractous retinal image dehazing.

Table 1

Comparison of PSNR and SSIM tested on synthetic hazy images for different dehazing methods.

Algorithm	SOTS Indoor images		SOTS Outdoor images	
	PSNR	SSIM	PSNR	SSIM
DCP	16.164	0.854	19.147	0.861
AOD-Net	20.158	0.816	24.142	0.918
DehazeNet	22.303	0.881	21.554	0.844
PFFNet	29.513	0.956	25.802	0.893
GCANet	30.134	0.967	28.343	0.957
ERAN(Ours)	33.448	0.979	30.501	0.964

It can be seen from Table 1 that for the peak signal-to-noise ratio of SOTS Indoor images, DCP, AOD-Net, DehazeNet, and PFFNet methods are all below 30. The GCANet method is 0.521 dB higher than PFFNet method, which is a great improvement over the previous four methods. Compared with the GCANet method, the PSNR of our ERAN method is improved by 3.314 dB. In additions, for the SSIM, none of the first three methods reached 0.900, and our method had the highest SSIM, indicating that our method was due to other advanced comparison methods.

Subjective evaluation can directly and quickly distinguish the pros and cons of the restored images, and we select different cataract retinal images for analysis and comparison. From the comparison of the methods in Fig. 6, it is found that the overall effect of DCP method for cataract dehazing is dark, and the effect is not ideal. Moreover, the dehazing effect of the AOD-Net method is improved, but some images are oversaturated. The DehazeNet method focuses on extracting the content characteristics of the image, resulting in the restored haze-free image is blurry and the color is gray. However, the details of the retinal image recovered by our ERAN method are kept well, which effectively solves the problem of the loss of blurred details after the DehazeNet method is used to remove the haze.

5. Conclusion

In this paper, we propose the End-to-End Residual Attention Mechanism (ERAN) for Cataractous Retinal Image Dehazing. The network design is based on the encoding-decoding structure, which facilitates subsequent feature extraction and reduces memory usage. The multi-scale feature extraction and feature fusion module are used to effectively extract and fuse features of different scales. At the end of each module, an adaptive skip connection is used to adaptively make the salient features have greater weight, suppress redundant information, and enhance the performance of the network to restore retinal cataract density images, so that the retina can remove cataracts more thoroughly. Through comparative experiments, it can be concluded that the proposed network has better dehaze ability. For future implementation, we will further improve the network and increase the speed of retinal image recognition in order to apply this technology to the field of medical images.

Ethical approval

No ethics approval is required.

Declaration of Competing Interest

The authors declare that there is no conflict of interests in this article.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China under Grant Nos. 61976215 and 62176259.

References

- [1] M. Badar, M. Haris, A. Fatima, Application of deep learning for retinal image analysis: a review[J], Comput. Sci. Rev. 35 (2020) 100203.
- [2] A. Miele, C. Fumagalli, G. Abbruzzese, et al., Biometric refractive error after cataract and retina surgery: a systematic review and a benchmark proposal[J], Eye (2021).
- [3] A. Raj, A.K. Tiwari, M.G. Martini, Fundus image quality assessment: survey, challenges, and future scope[J], IET Image Proc. 13 (8) (2019) 1211–1224.
- [4] E. Peli, T. Peli, Restoration of retinal images obtained through cataracts[J], IEEE Trans. Med. Imaging 8 (4) (1989) 401–406.
- [5] P.A. Asbell, I. Dualan, J. Mindel, et al., Age-related cataract[J], Lancet North Am. Ed. 365 (9459) (2005) 599–609.
- [6] D.E. Romo-Bucheli, U. Schmidt-Erfurth, H. Bogunovic, End-to-end deep learning model for predicting treatment requirements in neovascular amd from longitudinal retinal oct imaging[J], IEEE J. Biomed. Health Inform. 24 (12) (2020) 3456–3465.
- [7] R. Fattal, Single image dehazing[J], ACM Trans. Graph. 27 (3) (2008) 1–9.
- [8] K. He, J. Sun, X. Tang, Guided image filtering[J], IEEE Trans. Pattern Anal. Mach. Intell. 35 (6) (2012) 1397–1409.
- [9] K. He, J. Sun, X. Tang, Single Image Haze Removal Using Dark Channel Prior[J], IEEE Trans. Pattern Anal. Mach. Intell. 33 (12) (2011) 2341–2353.
- [10] D. Qiu, L. Zheng, J. Zhu, et al., Multiple improved residual networks for medical image super-resolution[J], Future Gener. Comput. Syst. (2021) 200–208.
- [11] D. Qiu, Y. Cheng, X. Wang, et al., Gradual back-projection residual attention network for magnetic resonance image super-resolution[J], Comput. Methods Programs Biomed. 208 (2021) 106252.
- [12] B. Cai, X. Xu, K. Jia, et al., DehazeNet: an end-to-end system for single image haze removal[J], IEEE Trans. Image Process. 25 (11) (2016) 5187–5198.
- [13] W. Ren, L. Si, Z. Hua, et al., in: Single Image Dehazing Via Multi-scale Convolutional Neural Networks[C]//European Conference on Computer Vision, Springer, Cham, 2016, pp. 154–169.
- [14] B. Li, X. Peng, Z. Wang, et al., in: AOD-Net: All-in-One Dehazing Network[C]//2017 IEEE International Conference on Computer Vision (ICCV), IEEE, 2017, pp. 4770–4778.
- [15] Z. He, V.M. Patel, in: Densely Connected Pyramid Dehazing Network[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2018, pp. 3194–3203.
- [16] X. Liu, Y.R. Ma, Z.H. Shi, et al., in: GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing[C]//Processing of IEEE International Conference on Computer Vision, IEEE, Seoul, Korea (South), 2019, pp. 7313–7322.
- [17] D. Chen, M. He, Q. Fan, et al., in: Gated Context Aggregation Network for Image Dehazing and Deraining[C]//2019 IEEE Winter Conference on Applications of Computer Vision, 2019, pp. 1375–1383.
- [18] D. Qiu, Y. Cheng, X. Wang, et al., Cardiac magnetic resonance images super-resolution via multi-channel residual attention networks, Comput. Math. Methods Med. (2021) 1–8.
- [19] J.Y. Zhu, T. Park, P. Isola, et al., in: Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE international Conference on Computer Vision, 2017, pp. 2223–2232.
- [20] D. Qiu, Y. Cheng, X. Wang, et al., Progressive U-Net residual network for computed tomography images super-resolution in the screening of COVID-19, J. Radiat. Res. Appl. Sci. 14 (1) (2021) 369–379.
- [21] S.G. Narasimhan, S.K. Nayar, Vision and the Atmosphere[J], Int. J. Comput. Vis. 48 (3) (2002) 233–254.
- [22] Y.Z. Su, Z.G. Cui, C. He, et al., Prior guided conditional generative adversarial network for single image dehazing - ScienceDirect[J], Neurocomputing 423 (2021) 620–638.
- [23] E. Peli, T. Peli, Restoration of retinal images obtained through cataracts[J], IEEE Trans. Med. Imaging 8 (4) (1989) 401–406.
- [24] J. Cheng, Z. Li, Z. Gu, et al., Structure-Preserving Guided Retinal Image Filtering and Its Application for Optic Disk Analysis[J], IEEE Trans. Med. Imaging 37 (11) (2018) 2536–2546.
- [25] K. He, X. Zhang, S. Ren, et al., in: Deep Residual Learning for Image Recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, Las Vegas, 2016, pp. 770–778.
- [26] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation[C], Int. Conf. Med. Image Comput. Comput.-Assisted Interv. (2015) 234–241.
- [27] Z. Wang, S. Ji, S. in: moothed dilated convolutions for improved dense prediction [C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 2486–2495.
- [28] S. Ioffe, C. Szegedy, in: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[C]//Proceedings of IEEE International Conference Machine Learning, IMLS, Lille, France, 2015, pp. 1–11.
- [29] Q. Wang, B. Wu, P. Zhu, et al., in: ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020, pp. 11534–11542.

- [30] J. Johnson, A. Alahi, F.F. Li, in: Perceptual losses for real-time style transfer and super-resolution[C]//Proceedings of 2016 European Conference on Computer Vision, Springer, Amsterdam, The Netherlands, 2016, pp. 694–711.
- [31] B.Y. Li, W.Q. Ren, D.P. Fu, et al., Benchmarking single-image dehazing and beyond[J], IEEE Trans. Image Process. 28 (1) (2019) 492–505.
- [32] K. Zhang, W. Zuo, L. Zhang, in: Learning a Single Convolutional Super-Resolution Network for Multiple Degradations[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018, pp. 3262–3271.
- [33] C. Hernandez-Matas, X. Zabulis, A. Triantafyllou, et al., FIRE: fundus Image Registration Dataset, Model, Artif. Intell. Ophthalmol (1) (2017) 16–28.
- [34] V.B. Gupta, N. Chitranshi, J.D. Haan, et al., Retinal changes in Alzheimer's disease— Integrated prospects of imaging, functional and molecular advances[J], Prog. Retin. Eye Res. 82 (8) (2020) 100899.