Computers, Materials & Continua
DOI: 10.32604/cmc.2023.031444
Article

Tech Science Press

CLGA Net: Cross Layer Gated Attention Network for Image Dehazing

Shengchun Wang[1], Baoxuan Huang[1], Tsz Ho Wong[2], Jingui Huang[1,*] and Hong Deng[1]

[1]Hunan Normal University, Changsha, 410006, China
[2]Z-emotion Pty Ltd., Victoria, 3178, Australia
*Corresponding Author: Jingui Huang. Email: hjg@hunnu.edu.cn
Received: 18 April 2022; Accepted: 12 June 2022

Abstract: In this paper, we propose an end-to-end cross-layer gated attention network (CLGA-Net) to directly restore fog-free images. Compared with the previous dehazing network, the dehazing model presented in this paper uses the smooth cavity convolution and local residual module as the feature extractor, combined with the channel attention mechanism, to better extract the restored features. A large amount of experimental data proves that the defogging model proposed in this paper is superior to previous defogging technologies in terms of structure similarity index (SSIM), peak signal to noise ratio (PSNR) and subjective visual quality. In order to improve the efficiency of decoding and encoding, we also describe a fusion residual module and conduct ablation experiments, which prove that the fusion residual is suitable for the dehazing problem. Therefore, we use fusion residual as a fixed module for encoding and decoding. In addition, we found that the traditional defogging model based on the U-net network may cause some information losses in space. We have achieved effective maintenance of low-level feature information through the cross-layer gating structure that better takes into account global and subtle features. We also present the application of our CLGA-Net in challenging scenarios where the best results in both quantity and quality can be obtained. Experimental results indicate that the present cross-layer gating module can be widely used in the same type of network.

Keywords: Deep learning; dehazing; image restoration; end to end

1 Introduction

Fog is a natural weather phenomenon. It contains turbid media such as smoke and dust, which blur and degrade images. In the topic of advanced machine vision, what should be processed are clear images instead of blurred ones, and the restoration of blurred images is a preprocessing step of advanced topics. Therefore, the defogging processing of the haze image has become a trending topic. The purpose of image defogging is to restore a clean scene from a foggy image.

Most defogging methods are based on physical scattering models [1–3], which are expressed as:

$$I(x) = J(x)\, t(x) + A(x)\,(1 - t(x)) \tag{1}$$

where I is the observed blurred image, x is the pixel position, J is the clear image, A is the atmospheric light, and t is the medium transmission image, which can be expressed as $\mathrm{t(x)} = \mathrm{e} \wedge (-\beta \mathrm{d(z)})$, with β and d(x) representing atmospheric scattering parameters and scene depth, respectively.

Most of the initial defogging methods are based on prior knowledge, using the statistical characteristics of the fog-free image to approximate the transmission image t(x) or atmospheric light A, and then restore the final clean image J(x) according to the formula. The dark channel prior (DCP) [4] proposed by He et al. is based on the fact that most areas of the fog-free image have very low values in at least one channel. Meng et al. [5] obtains clearer images through boundary constraints and contextual regularization. Color attenuation prior (CAP) is performed on blurred images to establish a linear model of scene depth, and then the model parameters are learned and supervised. Zhu et al. [6] creates a linear model to simulate the scene depth of the hazy image, estimating the fog density by the difference between saturation and brightness. However, estimating the transmission map and atmospheric light from the blurred image is an ill-posed problem. In different scenarios, this prior knowledge may not match the actual situation, resulting in inaccurate transmission map estimation and inaccurate restored images.

With the rise of deep learning, researchers have put efforts on neural networks to restore fog-free images, including residual learning ResNet [7,8], DehazeNet [9], multi-scale convolutional neural networks (CNN) [10], and pyramid dehazing network [11–13]. Compared with the traditional methods, deep learning methods try to directly retrieve the transmission image or the final fog-free image for end-to-end defogging, using a large number of training data sets. With the development of big data, they gradually achieved better performance and robustness.

This paper proposes an end-to-end cross-layer gated attention network for image dehazing.

In the past, the CNN-based defogging network processed the features uniformly, but in a single image, the fog density is often not uniform, as the weights of mist vary significantly from region to region. DCP pointed out that some pixels have extremely low values on at least one channel, which implies that different channel features may have feature information with different degrees of importance. So, the attention mechanism for the channel appeared later [14,15]. Therefore, we adopt a lightweight channel pixel attention module to weight image features of different regions and channels, which expands the network's learning ability and generalization ability for different features, and enhances the stability of the network. The emergence of ResNet has become a milestone in deep learning. We use fusion residuals in series with each layer of encoding/decoding to build a basic module. The multi-layer residual allows the network to efficiently extract dense fog and high-frequency features in a simple encoding and decoding module. Multiple residual learning supplements the multi-scale feature [16,17] map with a certain degree of low-frequency information. This eventually leads to an efficient network without low frequency information degradation.

Multi-scale boosted dehazing network (MSBDN) [18] advances the use of Strengthen-Operate-Subtract (SOS) boosting strategy to apply the coded information to the generation of the decoded information. More structural and spatial information is included in the image restoration module to enhance the accuracy of defogging. Since the encoding module information is too low-level, direct additions may bring too much unimportant information to calculation and thus lead to non-optimal results. The cross-layer gating mechanism presented in Section 3.1 allows high-level information of the feature extraction module to control the degree of participation of low-level information of the encoding module in decoding. Since the encoding module and the feature extraction module are unidirectional, the data is continuously processed, and the information becomes more and more advanced. Therefore, we create a Cross-Layer Gated Attention (CLGA) mechanism. After up-sampling and convolutional gating, the higher-level information is used to determine the lower-level information, and the correlation between different levels of information is employed, therefore we can strengthen the more important part of the low-level information to participate in the decoding module.

In the feature extraction stage, the gradual deepening of the network often weakens the shallow information [19]. In order to identify and fuse features at different levels, we adopt a feature fusion structure to divide the feature extraction module into different levels of information according to different levels. And according to the gated sub-network proposed by GCA-net [20], the importance of different levels of information is determined, and the fusion is carried out according to the corresponding importance weight. Then the network uses the feature attention module to dynamically adjust the fusion feature weights, expand high-frequency information and suppress redundant information, and enhance the network's ability to extract different haze features.

Our primary contribution is to introduce a new type of dehazing network, CLGA-Net that uses separate expansion convolution and local residuals as the feature extraction module, and takes advantage of fusion of multi-level feature information to better learn the importance of different features. In Sections 3.2 and 3.3, the usefulness of channel pixel attention and dense residuals [21] for image dehazing algorithms is exhibited with the simplicity and efficiency. Additionally, the cross-layer gated attention mechanism shows its capability for dehazing detailed parts of images and great potential in more symmetrical networks.

2 Related Work

Single image defogging is an inverse recovery process of physical corrosion defined by Eq. (1). Due to the unknown transmission image and atmospheric light, this is an ill-posed problem. All techniques proposed so far can be roughly divided into two categories (see Fig. 1): traditional methods based on prior knowledge and modern methods based on machine learning. The most obvious difference between the two is that the former uses human statistical prior knowledge of known clear images to estimate transmission maps or atmospheric light, while the latter directly regresses through automatic learning of a large number of hazy clear image pairs.

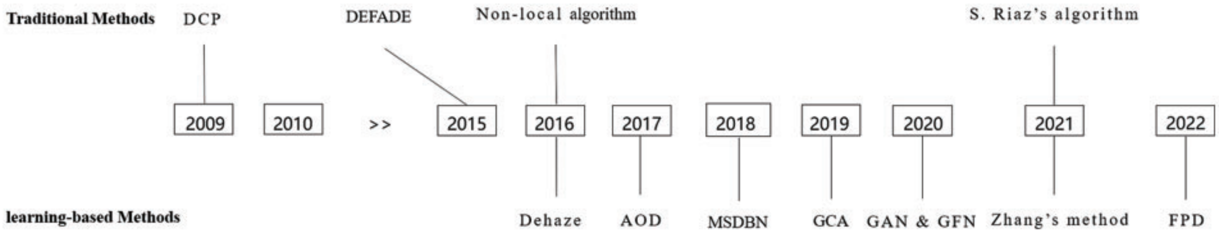

Figure 1: Dehazing methods proposed in recent years

2.1 Traditional Methods

In traditional prior-based methods, a large amount of various statistical prior information has been used to compensate for the loss of information in the process of image destruction. The dark channel prior(DCP) [4,17,22] estimates the transmission image and atmospheric light by more effective calculation of the transmission matrix, and uses Eq. (1) to restore the fog-free image; Berman et al. [23] put forward a non-local prior to describe the features of a clean image. The algorithm relies on the color of a haze-free image. The color is well approximated by hundreds of different colors, forming tight clusters in the RGB space. Choi et al. [24] found their algorithm on a priori statistics of twelve kinds of physical information such as brightness, information entropy, and color saturation. The white balance image and contrast enhancement image are refined by Laplacian multi-scale algorithm to process the weight map of various physical quantities. Thus, the DEFADE dehazing method is proposed; Riaz et al. [25] adopted an efficient scaling technique for transmission map estimation in 2021. A coarse transmission map is estimated by using the minimum of different size patches. Then a cascaded fast guided filter refines the transmission map. All of these schemes have achieved a certain degree of success, but methods based on a prior can not adapt to all situations, such as large-area sky images and outdoor high-brightness images.

2.2 learning-Based Methods

In recent years, applying learning-based methods to dehazing has become an interest. This kind of method takes advantage of a neural network to solve the problem by learning a large amount of prior knowledge of training sets, estimating the transmission matrix, investigating the numerical difference between the clear image and the foggy image, and saving the network weight information. For example, AOD-Net [26] reconstructs the physical scattering model and generalizes the two unknowns in the atmospheric scattering model to one unknown therefore the loss in the process of defogging is alleviated. Dehaze-Net [9] processes blurred images by estimating the transmission matrix, but inaccurate transmission mapping estimation reduces the dehazing effect of the model. The idea of applying GAN network to denoising is usually analogic to the use of a generator for generating denoising images, and a discriminator for judging denoising effects [27]. GFN-Net [28] employees a dual-branch [29] convolutional neural network to extract basic features and recovery features respectively. However, the gated fusion of the features obtained from these two branches leads to feature confusion easily, and the probability of loss explosion during the training process is relatively high. Instead of evaluating the transmission matrix, GCA-Net [21] achieves defogging by estimating the difference between a clear image and a blurred image, which greatly improves the quality. The MSDBN-DFF [18] network derives from U-Net [30], and uses back projection technology to realize the effective integration of multi-layer features in space and reduce the loss of low-level information. The FPD network [12] applies the FPN network structure in the field of object detection for dehazing, which can effectively integrate high-level and low-level semantics. Zhang et al. [31] combine the dehazing algorithm with an iterative fine-matching algorithm derived from motion structure to perform 3D reconstruction of dehazing images to improve the accuracy and accuracy of dehazing.

3 Method

In this section, we describe the structure of our end-to-end cross-layer gated attention network CLGA-Net. As shown in Fig. 2, this network consists of three parts–encoding, feature extraction and decoding. The input of the network is a hazy image, which is encoded into a feature map through the encoding part, then inputted to the feature extraction module. During three Base Feature Extraction module operations, the process keeps the size and channel of the feature map unchanged. The outputs of the three modules are spliced and convolutionally fused according to channels, and the fused features will be an input to the decoding part to obtain the final restored fog-free image. In addition, a dense fusion module, which consists of 3 identical residual layers, is added after the encoding layer and before the decoding layer. Each feature extraction module combines local residual learning and attention mechanisms. Attention mechanisms can simultaneously change the weight of features on pixels and channels.

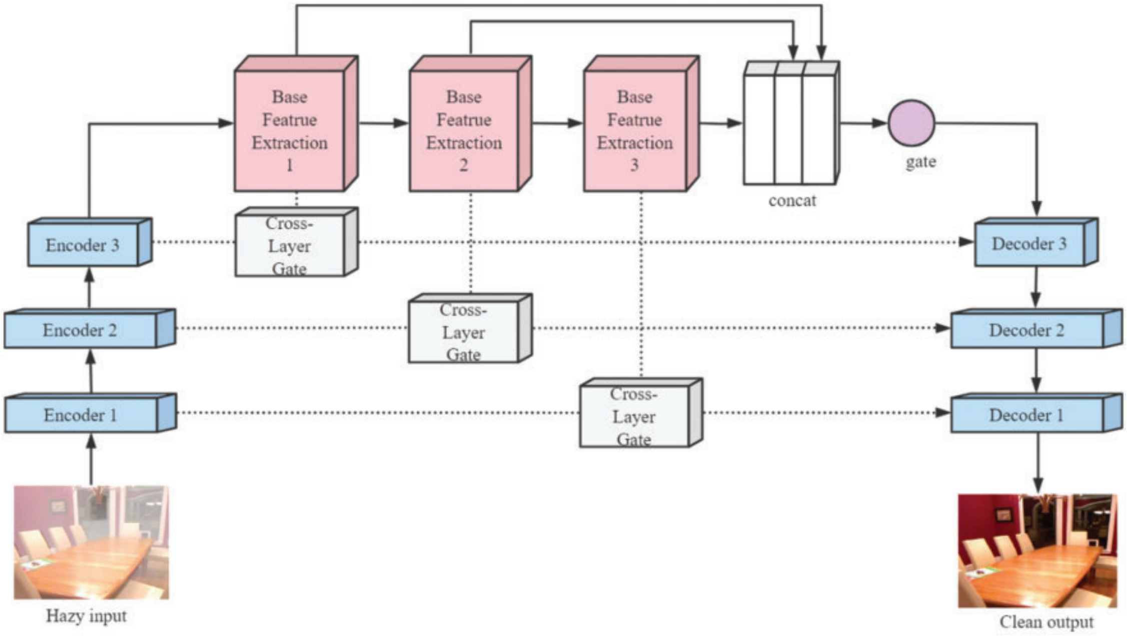

Figure 2: Network structure

3.1 Cross Layer Gating Attention

In most of the symmetric networks used for image defogging, the encoding and decoding part have the same number of layers. By regarding the decoding part as a clear image recovery module of the network, the output of each layer generated by the decoding part should be closer to a fog-free image. Information of the encoding part participating in the generation of the decoding part is effective for dehazing [9]. In this paper, we modify this participation process and add a cross-layer gate (CLG) mechanism between encoding and decoding (see Fig. 3).

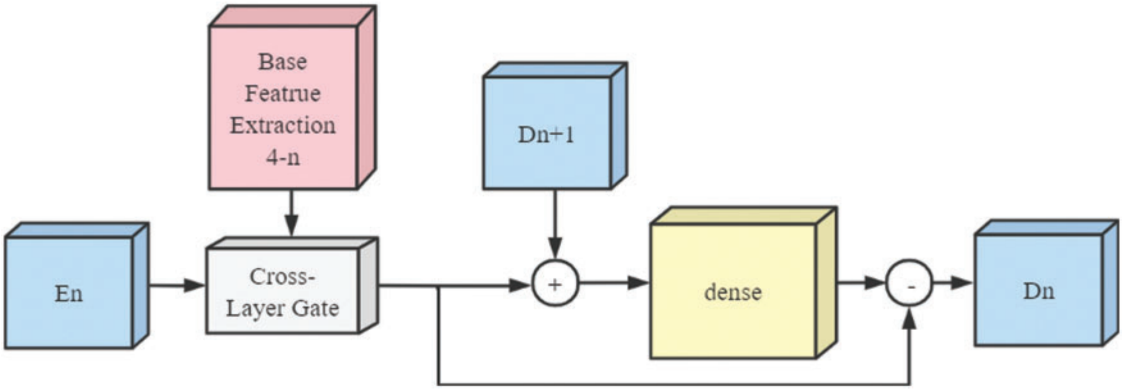

Figure 3: CLGA module structure

We believe that the feature map of the feature extraction part is a high-level feature obtained by gradually encoding the input image through the encoding layer. The semantics behind are becoming more and more abundant, which makes the detection of small targets more difficult. Each feature value of the high-level features covers the features of the corresponding areas of the low-level features. If low-level features participate in the generation of the decoding layer, the importance of each pixel should be distinguished to ensure that the information involved in the decoding is more effective. Therefore, the degree of participation of each low-level feature is determined according to the value of the high-level feature. Increasing the weight for important low-level features, and reducing the weight for unimportant features avoid spending a lot of resources on unnecessary calculations.

In short, we no longer simply add the output of the encoding layer to the generation of the decoding layer. Instead, the feature matrix obtained by the feature extraction module is treated as decision information. We propose a gating mechanism where the feature matrix output by the feature extraction module is an input to the gating mechanism after up sampling. The output is a single channel weight matrix. The output of the encoding layer is multiplied by the weight matrix to determine the participation of the coding layer information in the generation process of the decoding layer.

$$D_n = dense\left(\tilde{E}_n + D_{n-1}\right) - \tilde{E}_n \tag{2}$$

$$\tilde{E}_n = CLG\left(B_{4-n} \uparrow\right) * E_n \tag{3}$$

where, E_n represents the output of the nth layer of the encoding part, and $B_{4-n} \uparrow$ is the upsampled feature of the output of the 4-nth feature extraction module. The Cross-Layer Gate has only one convolutional layer, and the kernel size is 3∗3. Its input is the feature matrix of the feature extraction module, and the output is a single-channel weight matrix (1∗H∗W). In the mechanism we construct, we multiply the output of the encoding module with the feature matrix output by the deep feature extraction module through the calculation of the gating mechanism. That is, the deeper the feature matrix determines the degree of participation of the shallower coding layer information in the decoding layer. The calculated result is added to the output of the previous decoding layer and calculated by the dense residual module, the encoding layer information involved in the calculation is subtracted to obtain the *n*th layer of the final decoding module, namely *Dn*.

For completeness, we also present two alternatives for the CLG module. They are expressed as

$$D_n = dense\left(\tilde{E}_n + D_{n-1}\right) + \tilde{E}_n \tag{4}$$

and

$$D_n = dense\left(\tilde{E}_n + D_{n+1}\right) - D_{n+1} \tag{5}$$

respectively.

Experiments show that in formula 4 the basic features of the encoding layer get involved in the generation of the decoding layer too much, confusing the extraction of effective features. Eq. (5) subtracts the decoding information of the upper layer, and the information of the decoding layer is not fully utilized. Therefore, we choose the CLG mechanism represented by Eqs. (2) and (3).

Experiments have proved that our mechanism can make the network to better restore the physical information of the image, such as color, saturation, texture and other features.

3.2 Channel Pixel Attention Block

In order to enhance the network's ability to learn features of different channels and different regions, we put forward an attention mechanism that is able to change the weights of features on the channels and pixels simultaneously.

$$H_p\left(F_c\right) = \frac{1}{H}\sum_{i=1}^{H} X_c\left(i,j\right) \tag{6}$$

$$W_p\left(h_c\right) = \frac{1}{W}\sum_{i=1}^{W} X_c\left(i,j\right) \tag{7}$$

where Xc (i, j) represents the value of the c^{th} channel Xc at position (i, j), and H_p, W_p are adaptive global average pooling function. The feature shape is changed from $C \times H \times W$ to $C \times H \times 1$ and $C \times 1 \times W$. In order to obtain the weights of different channels and different pixels, the feature matrix first changes the weight in the H direction through the pooling function H_p, then changes the weight in the W direction through the W_p pooling function, after that convolution and activation are performed.

$$h_c = \delta\left(Conv\left(H_p\left(F_c\right)\right)\right) * F_c \tag{8}$$

$$F_c = \delta\left(Conv\left(W_p\left(h_c\right)\right)\right) * h_c \tag{9}$$

where δ is the sigmoid function. Following each pooling, the input element-wise multiplies with the weight obtained, and finally gets the output of the channel pixel attention block. This module is different from the previous simple channel attention or spatial attention. One call of the module can realize the transformation of the channel weights twice, and can realize the transformation of the pixel weights (once in the H direction and once in the W direction) simultaneously. The model is light in structure and simple in calculation, giving excellent effect.

3.3 Feature Extraction Block

The feature extraction part is composed of 3 basic feature extraction blocks. In fact, we suggest that the number of feature extraction modules should be equal to the number of coding layers as much as possible to maintain the stability of the CLG module. Each module is composed of 1 smooth hole convolution, 2 residual modules and 1 channel pixel attention module.

GCA proposes to add a corresponding additional convolutional layer before the hole convolution to increase the dependency between image pixels, so as to avoid losing the continuity of information. However, the accumulation of several convolutions of holes will cause the network to ignore the size information of the object itself. In other words, although GCA solves the grid artifacts issue caused by cavity convolution, cavity convolution itself is used to expand the receptive field without loss of coverage, that is, it has a certain effect on large objects, but for small objects not friendly. Considering the efficient application of ResNet in various networks, as shown in Fig. 4, we decide to learn the local residuals in the hole convolution part, and add the input information of the module to the output of the smooth hole convolution module to avoid the loss of information due to small objects in the image.

The residual module used in our network is mainly composed of two convolutional layers and a matrix addition. The residual module does not change the size or channel of the feature matrix during the entire calculation, but only learns the residual after multiple convolutions to optimize the network's extraction of fog features. The mathematical expression of the residual module is as follows:

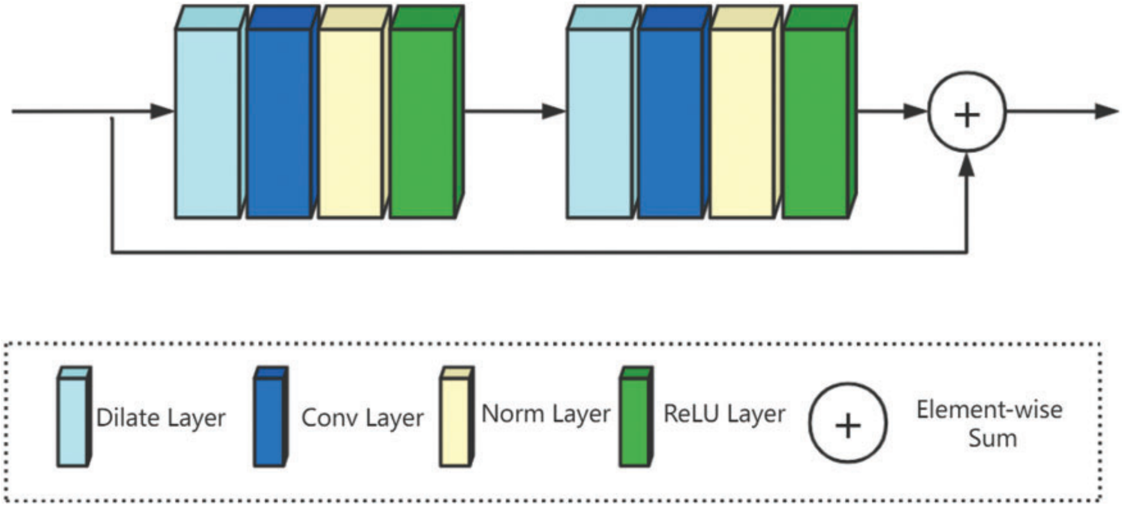

Figure 4: Smooth dilated residual block structure

$$R_1 = \sigma\left(Conv\left(Pad\left(x\right)\right)\right) \tag{10}$$

$$out = 0.1 * Conv\left(Pad\left(R_1\right)\right) + x \tag{11}$$

where x is the input of the residual module, and Pad is the ReflactionPad2d function, which fills the input. σ is the ReLU function. After two iterations of padding and convolution, the computed result is multiplied by 0.1 and added to the input matrix. We set the convolution result to be multiplied by 0.1 in order to learn the smallest possible residual and reduce the learning difficulty.

In the feature extraction module, we regard the output of the three Base Feature Extraction Blocks as three levels of feature information from shallow to deep, and connect the three types of information according to the channel direction. In addition, we adopt the Channel Pixel Attention mechanism to obtain adaptive learning weights and multiply the feature information to fuse three different levels of features. Therefore, we can ensure that the low-level information is passed to the deep level. Through channel attention, the network can pay more attention to physical information such as dense fog areas and high-frequency textures.

3.4 Loss Function

Mean squared error (MSE) or other L2 Loss is the most commonly employed loss function in defogging networks. However, Lim et al. [21] pointed out that L1 loss provides better performance on PSNR and SSIM metrics than L2 loss in training for many images restoration tasks. L2 loss is more sensitive to parts with larger errors, while L1 is a linear function. The error penalty of different sizes is the same, and the color and brightness are better preserved. Through comparative experiments, we also proved that by adding L1 Loss to the loss function, the defogging effect can better restore the color information and texture information of the image. Therefore, we suggest a linear combination of MSE Loss and L1 Loss as the loss function of the network.

$$L = \alpha L_{mse} + (1 - \alpha) L_1 \tag{12}$$

where α is a constant. After conducting considerable parameter tuning tests, we empirically set the value of α as 0.84. In order not to affect the convergence and ensure the stability of the network, we choose SmoothL1Loss as the L1 loss function.

4 Experiments

4.1 Experimental Implementation Details

In our experiments, we verify the effectiveness of CLGA-Net dehazing. We train and evaluated CLGA-Net on a public data set, and compared the experimental results with previous methods. CLGA-Net uses Adam optimizer to train for 100 epochs, the default initial learning rate is 0.001, and every 40 epochs, the learning rate is reduced to 10% of the original. Then we take the best experimental result as the final result of the experiment. We run experiments with the PyTorch framework on the GTX 1080ti graphics card, with all batch sizes set to 4.

4.2 Dataset Setup

We found that the data set used by the previous defogging network is synthesized formed on the atmospheric scattering model of Eq. (1), and only this specific data set is evaluated. For the sake of objectivity, we use RESIDE, a dehazing evaluation data set provided by Google. Its test set and training set consist of a large amount of depth and stereo data. Li et al. [32] used different evaluation indicators to evaluate the existing dehazing algorithms and compared them in more detail. Although their test data set includes indoor and outdoor images, they only report quantitative results for the indoor portion. On this basis, we select indoor datasets and outdoor datasets and conduct qualitative comparisons.

We select 6995 pairs of indoor images and 1000 pairs of outdoor images from the RESIDE dataset as training sets, and keep the image size unchanged. SOTS is a test subset of RESIDE, containing 500 indoor hazy images and 500 outdoor hazy images. All our methods for comparison are trained on the selected RESIDE dataset and evaluated on the SOTS test set. At the same time, we select several real-world hazy images from the Internet to evaluate the dehazing effect of various networks in real-world images.

4.3 Evaluation Metrics

For quantitative evaluation, we adopt SSIM and PSNR to evaluate the effect of the dehazing algorithm. SSIM is an indicator of the structural similarity of two images. The higher the structural similarity between the two images, the closer the SSIM is to 1. PSNR is a statistical indicator based on the gray value of image pixels. The higher the PSNR, the better the image restoration.

The equations for the two indicators are shown in Eqs. (13)–(15):

$$\mathrm{SSIM}(x, y) = \frac{\left(2\mu_x\mu_y + c_1\right)\left(\sigma_{xy} + c_2\right)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)} \tag{13}$$

where x, y are the two images to be compared, μ is the average of the image gray levels, σ is the standard deviation, σ_{xy} and is the covariance of x and y.

$$PSNR = 20 * log_{10}\left(\frac{MAX_x}{\sqrt{MSE}}\right) \tag{14}$$

and

$$MSE = \frac{1}{mn}\left(\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[x\left(i,j\right) - y\left(i,j\right)\right]^2\right) \tag{15}$$

MSE is the square of the image x and y residual values, and m∗n is the size of the two images. MAX_x is the maximum value that represents the color range. If each pixel is represented by 8 bits, then it is 255.

4.4 Quantitative and Qualitative Evaluation for Image Dehazing

We compare our network with the previous state-of-the-art image defogging methods quantitatively and qualitatively.

As is shown in Tab. 1, we used the six most famous dehazing models for quantitative evaluation: DCP, CAP, AOD-Net, Dehaze-Net, GFN-Net and GCA-Net. The first two means are traditional methods based on prior knowledge, and the latter four means are based on learning. For convenience, all results except GCA in Tab. 1 are directly quoted from [21]. For the latest dehazing method GCA, they also report the results of the RESIDE SOTS indoor and outdoor data set in the paper. We choose PSNR and SSIM introduced above, the two most widely used indicators in defogging tasks. It can be seen that CLGA-Net gives much better performance than all previous dehazing methods and reaches a considerable level in average prediction time.

Table 1: PSNR and SSIM results and average prediction time on google reside indoor dataset

	DCP	CAP	AOD-Net	Dehaze-Net	GFN-Net	GCA-Net	Ours
PSNR	16.62	19.05	19.06	21.14	22.30	30.23	**30.93**
SSIM	0.82	0.84	0.85	0.86	0.88	0.94	**0.96**
Average prediction time	2.531 s	2.581 s	1.078 s	2.618	10.861 s	**0.066 s**	0.079 s

We show the dehazing effect on the indoor and outdoor hazy image datasets in Figs. 5 and 6. Combining the observation of the two images, we find that the selected methods have a certain regularity in the dehazing effect of indoor or outdoor hazy images. We can observe that the dehazed images of DCP and CAP are dark, and the color distortion is severe. The result of AOD-Net dehazing is not ideal and there is still a lot of fog residue and a little color distortion. The effect of GFN-Net is not advantageous, and the local fog of the image cannot be restored. In comparison with previous networks, although GCA-Net generates better results in terms of color reproduction, there is a certain gap with the Ground True image e.g., low saturation. It can be seen from this figure that our network provides the best dehazing effect, while yet completely eliminating the haze. It maintains the outline detail and clarity of the object, and preserves the original color and brightness.

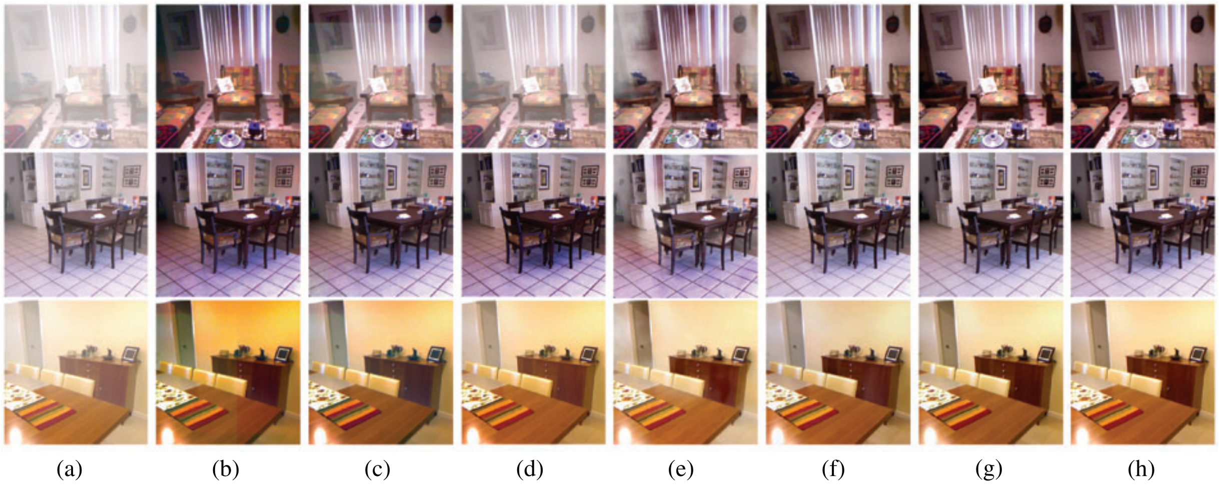

Figure 5: Indoor hazy images results. (a) Hazy, (b) DCP, (c) CAP, (d) AOD-Net, (e) GFN-Net, (f) GCA-Net, (g) Ours, (h) GT

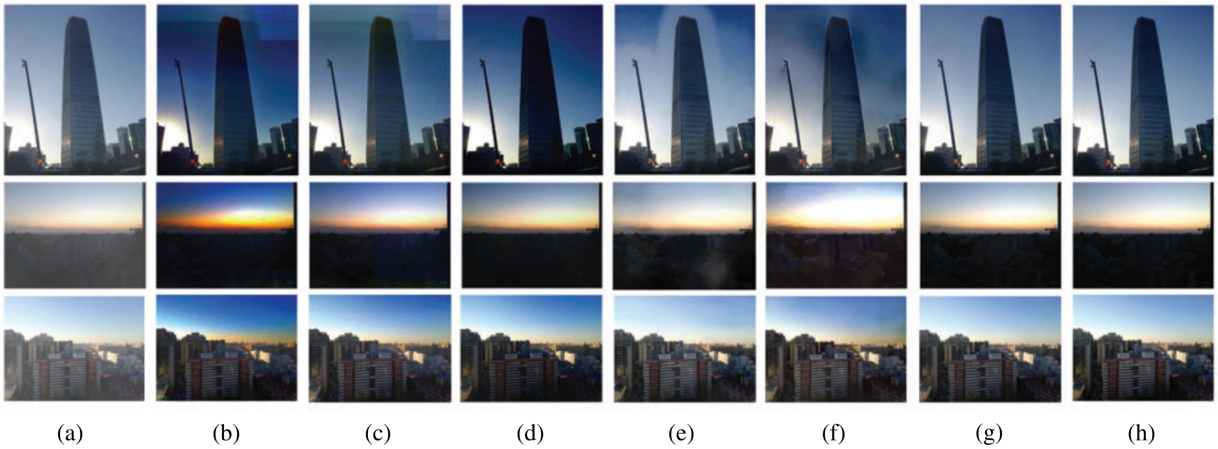

Figure 6: Outdoor hazy images results. (a) Hazy, (b) DCP, (c) CAP, (d) AOD-Net, (e) GFN-Net, (f) GCA-Net, (g) Ours, (h) GT

The dehazing effect on the real hazy images is shown in Fig. 7. The fog density of a real outdoor scene rises with the increase of the depth of field, and the task of dehazing the sky part is more difficult. Traditional dehazing methods in the light of prior knowledge are completely helpless for real hazy images. However, the dehazing effect of the method based on machine learning in the real scene cannot achieve the dehazing effect of the synthetic hazy images. It can be seen from the figure that our network can uniformly dehaze the image while keeping the image clean and tidy, and there is no local pollution causing subjective visual deficiencies. Moreover, when a large area of white scene appears in the image, the dehazing technique is particularly important to distinguish the white area from the hazy area. As can be seen from Fig. 8, our network is able to identify white areas accurately (such as reflections on water, snow, sky, etc.) and remove haze precisely.

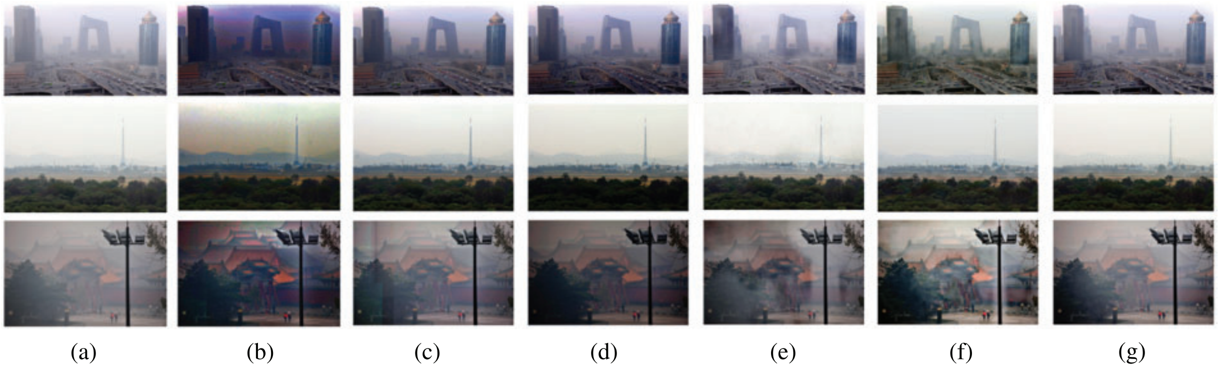

Figure 7: Real hazy images results. (a) Hazy, (b) DCP, (c) CAP , (d) AOD-Net, (e) GFN-Net, (f) GCA-Net, (g) Ours

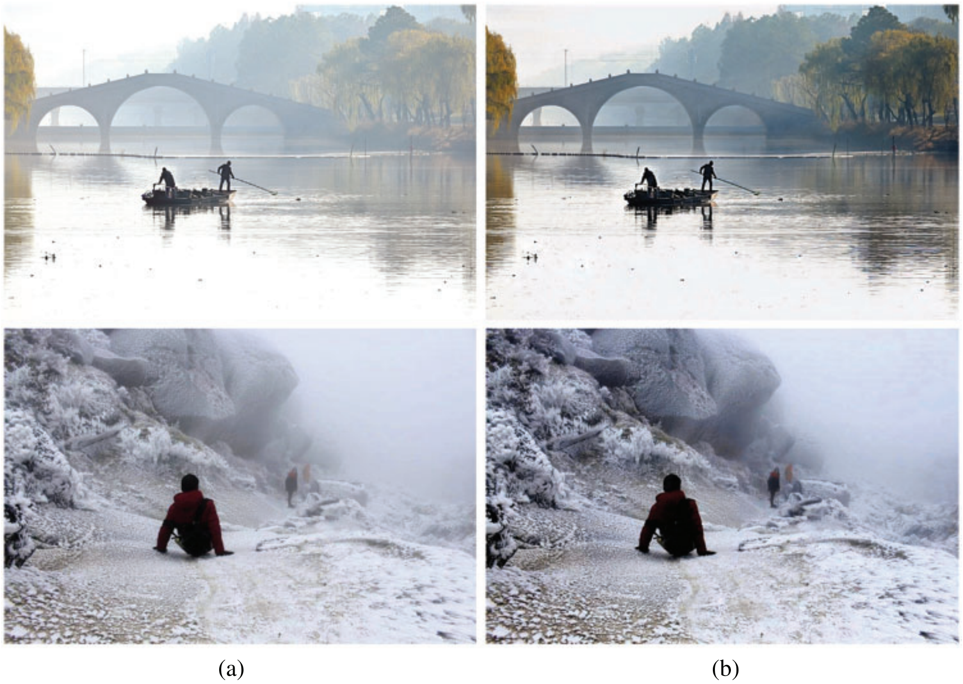

Figure 8: Large area white scenery image dehazing results. (a) Hazy, (b) Ours

4.5 Effectiveness of CLGA-Net Structure

In order to prove the superiority of our CLGA network structure, we conduct ablation experiments on different modules of the network. We mainly focus on several components: Cross-Layer Gating mechanism, dense residual module, channel pixel attention mechanism. We use this to evaluate 4 different network configurations on image defogging, adding a component incrementally each time, and compare the evaluation results of dehazing. As is shown in Tab. 2, the network performance is poor without any components. Each time a component is added, the performance continues to improve, and the evaluation results gradually increase. Therefore, we can prove that the combination of our proposed components is effective for dehazing tasks.

Table 2: Detailed ablation analysis of each component under different training configurations

Dense residual		√	√	√
CLG block			√	√
Channel pixel attention				√
PSNR	28.17	29.28	30.24	30.93

We apply the CLG module to other symmetric networks to prove the high applicability of our module. We select three more classic symmetric networks-GCA Net and U-Net, and the MSBDN network emerged in 2020. Fig. 9 shows the horizontal structure between MSBDN-Net and U-Net. It can be seen that the Strengthen-Operate-Subtract (SOS) module used by MSBDN-Net simply adds the output of the encoding layer and the input of the decoding layer. U-Net uses the matrix concatenation method to apply the information of the encoding layer. GCA-Net does not have any connection or computation between the encoding part and the decoding part. To show that our CLG module is beneficial for dehazing, we add the CLG module to these three networks, and compare them with the original network without CLG module. Three layers of encoding and decoding are selected in each network, and a CLG module is added to each layer. Using the same training framework, hardware and software facilities, the model effects of the six networks with training 100 epochs are shown in the Tab. 3:

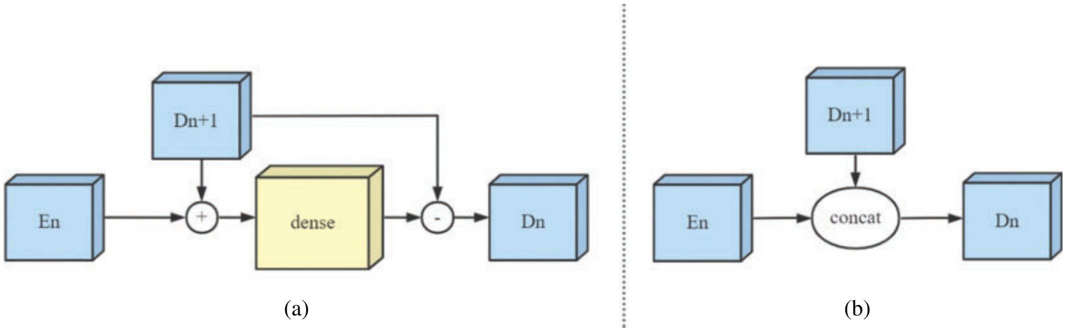

Figure 9: (a) SOS module in MSBDN-Net, (b) Concatenate module in U-Net

Table 3: Comparison of the dehazing effects of three symmetry networks with and without the CLG module

	PSNR		SSIM	
GCA Net	25.98		0.82	
GCA Net with CLG block	27.78	↑6.9%	0.84	↑2.4%
U-Net	28.62		0.84	
U-Net with CLG block	29.63	↑3.5%	0.87	↑3.5%
MSBDN Net	27.15		0.91	
MSBDN Net with CLG block	27.22	↑0.2%	0.93	↑2.2%

It can be seen from the Tab. 3 that after adding the CLG module to the three symmetrical networks, compared with the original network without adding the CLG module, both PSNR and SSIM indicators are improved, and the effect is particularly obvious on the GCA network. Taking an indoor hazy image as an example, Fig. 10 shows the dehazing effect of GCA-Net and GCA-Net with CLG module on the same image. We can clearly see from the enlarged area that the restoration of the light-colored background by the network with CLG module is closer to the real image, the saturation is higher, the texture information is restored more delicately, and the dehazed image is clearer.

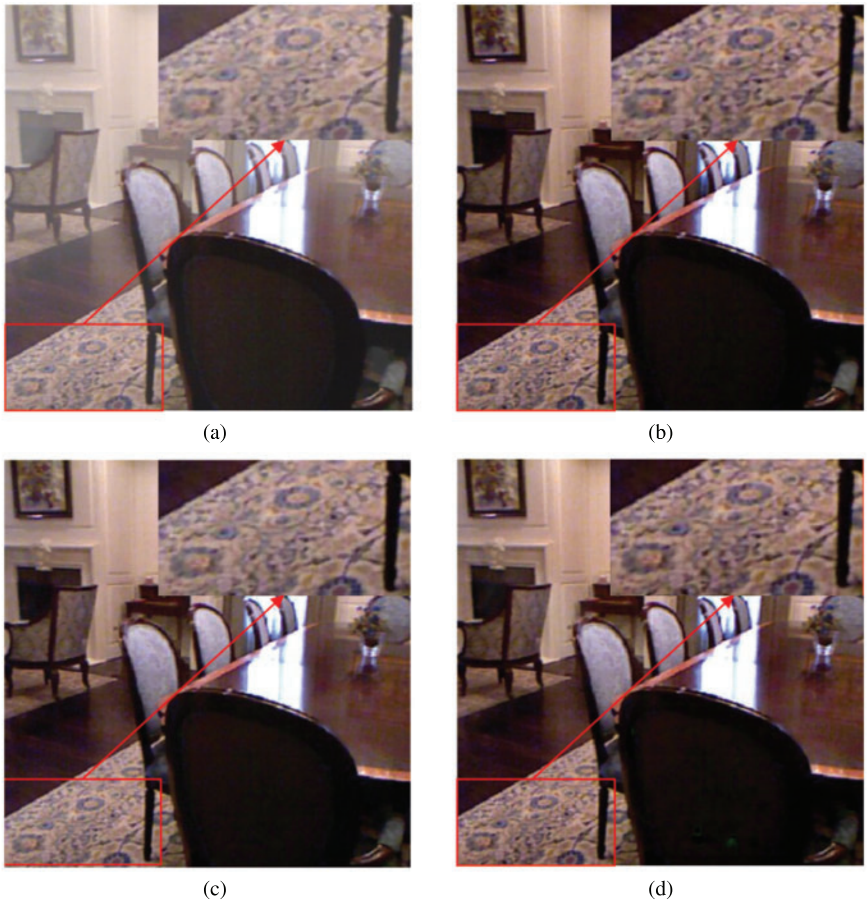
(a)
(b)
(c)
(d)

Figure 10: An example of dehazing. (a) Hazy, (b) GT, (c) GCA-Net, (d) GCA-Net with CLG block

To verify that CLG outperforms the SOS module and U-net's concatenation module, and to explain the reasons for not choosing the cross-layer gating mechanism represented by Eqs. (4) and (5) proposed in Section 3.1, we also conduct module ablation experiments. We replace the CLG modules in the proposed network with the SOS and Concatenation modules shown in Fig. 9 and the concatenated modules represented by Eqs. (4) and (5), respectively, and train them in the same environment. The results are shown in Tab. 4. It can be seen that after the CLG was replaced, the test results dropped by 2 to 6 percentage points. Due to the existence of dense residual and Channel pixel attention, the test results are still kept at a good level. This proves that compared with the above four modules, CLG is more suitable as a lateral connection mechanism between encoding and decoding.

Table 4: Comparison of CLG, SOS and concatenation modules in CLGA network architecture

	PSNR		SSIM	
CLGA Net	30.93		0.96	
CLGA Net with SOS block	29.84	↓3.5%	0.92	↓4.1%
CLGA Net with concatenation block	28.78	↓7.0%	0.90	↓6.2%
CLGA Net with Eq. (4)	30.37	↓1.8%	0.96	-
CLGA Net with Eq. (5)	30.02	↓2.3%	0.95	↓0.7%

In addition to achieving better results in test metrics, CLG can also do better in subjective vision. The SOS module simply adds the output of the encoding layer to the input of the decoding layer, followed by dense residual calculation (a linear operation in the horizontal direction), then the output of the decoding layer is subtracted out. As is shown in Fig. 11, the dehazing effect of an indoor hazy image is taken as an example. When using the previous SOS module, the color of the restored fog-free image is darker, the texture of the object is blurred, and the foggy part is difficult to remove. In contrast, our CLG mechanism addresses these limitations and dehazes more thoroughly without mistaking light-colored backgrounds for hazy areas. In the meantime, it can maintain texture information and restore the original color of the image.

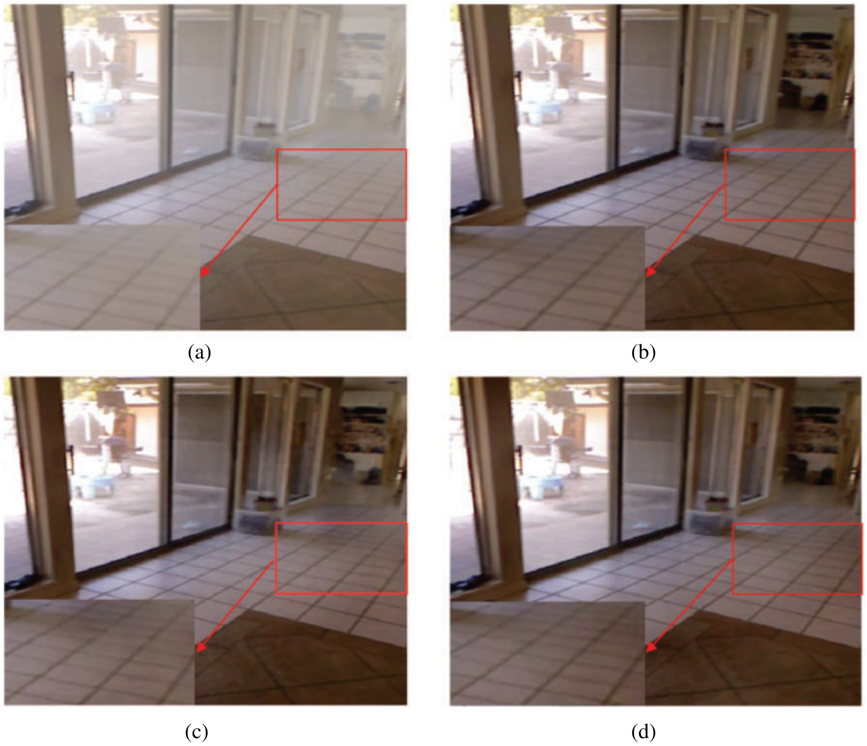
(a)
(b)
(c)
(d)

Figure 11: An example of image dehazing. (a) Hazy, (b) GT, (c) Previous SOS, (d) CLG

5 Conclusion

This paper proposes an end-to-end cross-layer gated attention network for image dehazing. In order to learn shallow feature information more efficiently, a cross-layer gating enhancement mechanism and a dense residual module are adopted. In addition, for the purpose of paying attention to the difference of features in channels and spaces, we investigate the channel pixel attention mechanism. Although our network structure is simpler, it is more efficient than past methods. Our CLG module has more advantages in dehazing and can be widely applied to other symmetric networks. Our network shows more capability of maintaining image details and color fidelity. We would like to improve the CLGA-Net model to have less loss in training and make a more accurate judgment of the fog in the image. We also hope to utilize incremental learning and migration learning to improve the speed and accuracy of the model. The effective modules in CLGA-Net could be also expected to advance the research of visual tasks such as rain removal, deblurring, and noise reduction.

Although our model achieves good results on metrics, the limitations of CLGA-net still exist. How to make the model reach the top level in prediction time, and how to make the output of the feature extraction module have reliable decision-making after up-sampling, these are the problems that need to be further improved. Optimizing the structure of CLG and selecting an appropriate up-sampling scheme are the contents of our follow-up research.

Funding Statement: This work is supported by the Key Research and Development Program of Hunan Province (No.2019SK2161) and the Key Research and Development Program of Hunan Province (No.2016SK2017).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] E. J. McCartney, "Optics of the atmosphere: Scattering by molecules and particles," in *NYJW*, New York, NY, USA, pp. 698–699, 1976.

[2] R. T. Tan, "Visibility in bad weather from a single image," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, IEEE Conf. on*, Anchorage, AK, USA, pp. 1–8, 2008.

[3] S. G. Narasimhan and S. K. Nayar, "Chromatic framework for vision in bad weather," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Hilton Head, SC, USA, pp. 598–605, 2000.

[4] K. He, J. Sun and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.

[5] G. Meng, Y. Wang, J. Duan, S. Xiang and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *Proc. of the IEEE Int. Conf. on Computer Vision*, Sydney, Australia, pp. 617–624, 2013.

[6] Q. Zhu, J. Mai and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Trans. Image Processing*, vol. 24, no. 11, pp. 3522–3533, 2015.

[7] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 770–778, 2016.

[8] S. Xie, R. Girshick, P. Dollár, Z. Tu and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, pp. 1492–1500, 2017.

[9] B. Cai, X. Xu, K. Jia, C. Qing and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.

[10] W. Ren, S. Liu, H. Zhang, J. Pan., X. Cao *et al.,* "Single image dehazing via multiscale convolutional neural networks," in *European Conf. on Computer Vision*, Amsterdam, The Netherlands, Springer, pp. 154–169, 2016.

[11] H. Zhang, V. Sindagi and V. M. Patel, "Multiscale single image dehazing using perceptual pyramid deep network," in *IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, Salt Lake City, UT, USA, pp. 902–911, 2018.

[12] S. Wang, P. Chen, J. Huang and T. H. Wong, "Fpd net: Feature pyramid dehazenet," *Computer Systems Science and Engineering*, vol. 40, no. 3, pp. 1167–1181, 2022.

[13] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan *et al.,* "Feature pyramid networks for object detection," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, pp. 2117–2125, 2017.

[14] J. Hu, L. Shen and G. Sun, "Squeeze-and-excitation networks," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 7132–7141, 2018.

[15] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo *et al.,* "ECA-Net: Efficient channel attention for deep convolutional neural networks," arXiv 1910.03151, 2019.

[16] C. O. Ancuti and C. Ancuti, "Single image dehazing by multi-scale fusion," *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3271–3282, 2013.

[17] B. Xie, F. Guo and Z. Cai, "Improved single image dehazing using dark channel prior and multi-scale retinex," in *Proc. of the IEEE Conf. on Intelligent System Design and Engineering Application*, Changsha, CS, China, pp. 848–851, 2010.

[18] D. Hang, J. Pan, L. Xiang, Z. Hu, X. Zhang *et al.,* "Multi-scale boosted dehazing network with dense feature fusion," arXiv preprint arXiv:2004.13388, 2020.

[19] C. Song, X. Cheng, Y. X. Gu, B. J. Chen and Z. J. Fu, "A review of object detectors in deep learning," *Journal on Artificial Intelligence*, vol. 2, no. 2, pp. 59–77, 2020.

[20] D. Chen, M. He, Q. Fan, J. Liao and L. Zhang, "Gated context aggregation network for image dehazing and deraining," in *2019 IEEE Winter Conf. on Applications of Computer Vision (WACV)*, Waikoloa Village, HI, USA, pp. 1375–1383, 2019.

[21] B. Lim, S. Son, H. Kim, S. Nah and K. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, Honolulu, HI, USA, pp. 136–144, 2017.

[22] H. Xu, J. Guo, Q. Liu and L. Ye, "Fast image dehazing using improved dark channel prior," in *Information Science and Technology (ICIST), 2012 Int. Conf. on*, Wuhan, China, IEEE, pp. 663–667, 2012.

[23] D. Berman and S. Avidan, "Non-local image dehazing," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 1674–1682, 2016.

[24] L. Choi, J. You and A. C. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3888–3901, 2015.

[25] S. Riaz, M. W. Anwar, I. Riaz, H. Kim, Y. Nam *et al.,* "Multiscale image dehazing and restoration: An application for visual surveillance," *Computers, Materials & Continua*, vol. 70, no. 1, pp. 1–17, 2022.

[26] B. Li, X. Peng, Z. Wang, J. Xu and D. Feng, "Aod-net: All-in-one dehazing network," in *Proc. of the IEEE Int. Conf. on Computer Vision*, Venice, Italy, pp. 4770–4778, 2017.

[27] J. Ouyang, Y. He, H. Tang and Z. Fu, "Research on denoising of cryo-em images based on deep learning," *Journal of Information Hiding and Privacy Protection*, vol. 2, no. 1, pp. 1–9, 2020.

[28] X. Zhang, H. Dong, Z. Hu, W. Lai, F. Wang *et al.,* "Gated fusion network for degraded image super resolution," *International Journal of Computer Vision*, vol. 128, no. 6, pp. 1699–1721, 2020.

[29] R. Chen, L. Pan, C. Li, Y. Zhou, A. Chen *et al.,* "An improved deep fusion CNN for image recognition," *Computers, Materials & Continua*, vol. 65, no. 2, pp. 1691–1706, 2020.

[30] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," arXiv preprint arXiv:1505.04597, 2015.

[31] J. Zhang, X. Qi, S. H. Myint and Z. Wen, "Deep-learning-empowered 3d reconstruction for dehazed images in IOT-enhanced smart cities," *Computers, Materials & Continua*, vol. 68, no. 2, pp. 2807–2824, 2021.

[32] B. Li, W. Ren, D. Fu, D. Tao, D. Feng *et al.,* "Reside: A benchmark for single image dehazing," arXiv preprint arXiv:1712.04143, 2017.