



Hazy Removal via Graph Convolutional with Attention Network

Bin Hu¹ · Zhuangzhuang Yue¹ · Mingcen Gu¹ · Yan Zhang¹ · Zhen Xu¹ · Jinhang Li¹

Received: 19 August 2021 / Revised: 17 October 2022 / Accepted: 2 April 2023 / Published online: 11 April 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Most deep learning based single image dehazing methods use convolutional neural networks (CNN) to extract features, however CNN can only capture local features. To address the limitations of CNN, We propose a basic module that combines CNN and graph convolutional network (GCN) to capture both local and non-local features. The basic module consist of a CNN with triple attention modules (CAM) and a dual GCN module (DGM). CAM that combines the channel attention, spatial attention and pixel attention is designed to earn more weight from important local features. DGM combines spatial coherence computing and channel correlation computing to extract non-local information. The architecture of the network is similar to U-Net, and skip connections used in the symmetrical network can pass the image details from shallow layers to deep layers. Experimental results in several datasets indicate that the proposed method outperforms the state-of-the-arts both quantitatively and qualitatively.

Keywords Graph convolutional network · Attention · Image dehazing · Deep learning

1 Introduction

Image dehazing is a typical low-level image processing problem in the real world. Since there exist infinite feasible solutions, it is a highly ill-posed problem, and it has become a hot topic in the field of image restoration. The atmosphere scattering model [1, 2] is a simple yet effective method to solve the problem.

$$I(x) = J(x)t(x) + A(1 - t(x)) \quad (1)$$

where $I(x)$ is the hazy image and $J(x)$ is the clear image, A is the global atmospheric light which represents the intensity of the scattered light of the scene, and $t(x)$ is the transmission map which describes the attenuation in intensity.

Let the clear image $J(x)$ be the output, formula (1) can be re-written as:

$$J(x) = \frac{1}{t(x)}I(x) - A\frac{1}{t(x)} + A \quad (2)$$

From formula (2), we can observe that the way to restore the clear image $J(x)$ is to estimate A and $t(x)$. However, only

the hazy image $I(x)$ is known, it is difficult to restore the clear image $J(x)$.

In recent decades, lots of techniques have been proposed to remove the hazy from images, and significant progress has been achieved. Generally, single image dehazing methods can be categorized into two classes: model-driven and data-driven. Early works are mostly based on the physic model such as the atmosphere scattering model [1], and those methods usually try to design hand-crafted features to estimate A and $t(x)$ via computing formula (2), or explore prior knowledge to deal with the problem [3]. However, those methods are easily sensitive to image variations such as changes in viewpoints, illumination, and scenes [4].

Recently, data-driven methods based on deep learning have become the dominant techniques to solve the low-level image processing [4–8] and the high-level computer vision [9, 10]. The data-driven methods using deep learning directly regress the intermediate transmission matrix or the final clear images due to the massive training data and powerful computing power [5, 6]. Compared to traditional model-driven methods, the data-driven methods based on deep learning achieve superior performance with robustness.

Most deep learning methods use convolutional neural networks (CNNs) as the backbones, although great progress has been made in single image dehazing, CNN can only capture the local spatial image feature but lack in broad contextual

✉ Bin Hu
hubin@ntu.edu.cn

¹ School of Information Science and Technology, Nantong University, Nantong, Jiangsu, China

information. Dilated convolution [11] is proposed to obtain larger receptive field, however it is still a convolution operation, the feature captured from dilated convolution is still local spatial information. While graph convolutional network (GCN) is proved to extract long-range contextual features [12], such as non-local net [13], which is widely used in image and video applications [14, 15]. However, there are still few works to apply GCNs into image dehazing.

Recently, various attention mechanisms are proposed to extract local information, including spatial neighbor information [13], channel-wise and pixel-wise feature. However, haze in image is unevenly distributed, and the network with single attention mechanism cannot make full use of the information from the image.

In summary, we propose a network with GCN to address the limitations of CNNs, also, we introduce a CNN module with multi-attention mechanisms to gain more information from the image. The proposed end-to-end network combines GCN and multi-attention CNN, and it can extract both local and broad contextual information.

The follows are our contributions.

- 1) A CNN with triple attention modules (CAM) is proposed, and the CAM combines the channel attention, spatial attention and pixel attention in channel-wise, spatial-wise and pixel-wise to earn more weight from important local features. The dilated convolution is used to obtain larger receptive field.
- 2) A dual GCN module (DGM) is proposed, and the DGM combines spatial coherence computing and channel correlation computing to extract non-local information.
- 3) Our network is an end-to-end network and is easy to implement. The experiment results show that our work achieves superior performance in comparison with the state-of-the-arts on both synthetic and real- world data sets.

2 Related Work

In general, most single image dehazing works can be categorized into model-driven and data-driven two classes. The atmospheric scattering model is a most widely used data-driven method, and the works based on the model follow the similar three steps: (1) estimating the transmission map $t(x)$ by the hazy image samples; (2) estimating the global atmospheric light A using empirical methods; (3) computing the clear image $J(x)$ according to formula (3).

Early methods often require multiple images from the same scene under different conditions [2, 13–15] to estimate transmission map $t(x)$. In the different weather conditions, researchers took several images of the same scene [2] or

different angles [16] to estimate $t(x)$. However, these methods do not work when there only exist one image for a scene.

Fattal [17] proposed a refined image formation model to estimate the scene transmission and surface shading by separating the hazy image into regions of constant albedo. But the method can only deal with the images that contain a slight haze and it requires time-consuming computations. He [3] discovered the dark channel prior (DCP) and used the soft-matting operations to estimate the transmission matrix, and the method is more reliable and simple, followed by many successors. But when the color of the scene objects are similar to the atmospheric light, the DCP is found to be unreliable, and DCP is computationally expensive. Gibson et al. [18] proposed a standard median filter to improve the DCP computing speed. Martin et al. [19] adopted Markov Random Fields for image restoration.

Recently, with the great success of deep learning in diverse computer vision tasks, the data-driven de-hazing approaches using CNNs become popular. The CNN methods can directly learn $t(x)$ or restore clear image from massive data. Li et al. [20] first proposed a novel end-to-end lightweight CNN called AOD-Net by formulating the atmosphere light and transmission map in one matrix to generate clear images directly. But the architecture of AOD-Net is too simple, the results are not well. Cai et al. [6] designed a more complex network called DehazeNet to generate the clear image by estimating atmospheric light from the hazy image also, and the results are better than AOD-Net. Similarly, different architectures are designed generate clear image, such as multi-scale CNN [4, 21], which can generate a coarse-scale transmission matrix and then gradually refined it. Liu et al. propose a mesh network structure for image dehazing [22]. Dong et al. [23] proposed a boosted U-Net based on boosting and error feedback. Liu et al. proposed a double residual connection [24] to perform the dehazing. Chen et al. proposed an encoder-decoder network called GCANet to fusion feature in different layers [25]. Besides, generative models such as GAN [8] and diffusion-driven [26] are used for image dehazing. However, the above networks are mostly based on CNN and cannot extract broad contextual feature.

Graph Convolutional Networks have been widely used in many high-level computer vision tasks to extract contextual information. For image and video, the most widely used form of GCNs is the non-local network [13]. In recent years, the GCNs have been applied to capture the global contextual information [27]. However, there are still few works to apply GCNs into image dehazing.

Attention mechanisms are widely used in both high level computer vision and low level computer vision tasks, and its main idea is to capture long-range inter-dependencies in channel-wise, spatial-wise or pixel-wise. GridDehazeNet [22] used a channel-wise attention mechanism to make the

network more flexible information exchange and aggregation. FFA-Net [28] combined channel-wise and pixel-wise attention to capture more information, and it proved that multi-attention is feasible and effective.

3 Proposed Method

In this section, we mainly discuss the detail of the proposed graph convolution with attention network, which is a trainable end-to-end network and has no reliance on the atmosphere scattering model. The architecture of our network looks like the U-Net [29], shown in Fig. 1. The skip connection used in the symmetrical network can pass the image details from shallow layers to deep layers. The network takes the hazy image X as the input and the clear image Y as the predicted result. The network consists of two convolution layers for pre-processing, several basic units and two convolution layers for reconstructing output. The pre-processing and reconstruction layers are designed standard 3×3 convolutional operations. The basic unit contains CAM and DGM, as shown in Fig. 1.

3.1 CNN with Attention Module

In our framework, a CNN with triple attention modules (CAM) is proposed, the architecture of basic CAM is depicted in Fig. 2, it consists of two dilated convolution layers with 3×3 kernel size, residual learning and an attention block, the first dilated convolution layer with $DF = 1$ is activated by ReLU, and the DF of the second dilated convolution layer is set to 3. A global residual learning connects the input feature and the output feature. With the local residual learning and global residual learning, the low-frequency regions from the input features can be learned by the skip connection.

Dilated convolutions can increase the receptive field without increasing parameters, the output of dilated convolution is defined as:

$$F_{DF} = K_{DF} * F_{in} + b \tag{3}$$

Where F_{DF} and F_{in} are output features and input features, respectively, DF is the dilation factor and K is the convolution kernel size.

The attention block (AB) combines the channel attention, pixel attention and spatial attention, which can provide additional flexibility in dealing with non-local and local information, and can expand the representational ability of CNNs. and the architecture of AB is depicted in Fig. 3. The “S” in the figure means sigmoid activation function, and the “C” means concatenation operation.

3.1.1 Channel Attention

Generally, a network uses a set of convolutional layer to express the neighboring spatial dependencies within local receptive fields. However, the global spatial patterns also need to be considered under the complicated non-uniform condition. When the neighborhoods of the image contain strong hazy component, the contextual information from clear regions may be required. Recently, the channel attention module [30] is developed to capture a richer non-local and overall feature by modeling channel interdependencies. Thus, we propose the channel attention module to extract non-local context features.

The channel attention module mainly concerns that different channel features have totally different weighted information. Firstly, a global average pooling is used to capture the channel-wise global spatial features.

$$g_c = H_p(F_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_c(i, j) \tag{4}$$

where H_p means the global average pooling function, $X_c(i, j)$ is the value of c -th channel of input X_c at position (i, j) . And the dimension of the feature changes from $C \times H \times W$ to $C \times 1 \times 1$, C denotes the channels, and $H \times W$ is the size of the feature map.

Then two dilated convolution layers, which are activated by a ReLU function and a sigmoid function respectively are applied to get the weights of the different channels, and DF of the first dilated convolution layers is set to 1, and the second is set to 3.

$$C_f = \sigma(\text{DConv}(\delta(\text{DConv}(g_c)))) \tag{5}$$

where σ stands for the sigmoid function, and δ is the ReLU function.

Finally, the weight of the channel F_c^* is computed by element-wise multiplying the input F_{input} and C_f .

$$F_c^* = F_{input} \odot C_f \tag{6}$$

3.1.2 Pixel Attention

Considering that the hazy image distribution is variant on the different image pixels, we further learn the spatially variant properties of the hazy images in an adaptive way by the pixel attention module. The pixel attention is applied to get weights from pixel, which makes the network pay more attention to informative features, such as thick-hazed pixels and high-frequency image region.

The architecture of pixel attention module is depicted in Fig. 3, it consists of two dilated convolution layers with ReLU and sigmoid activation function.

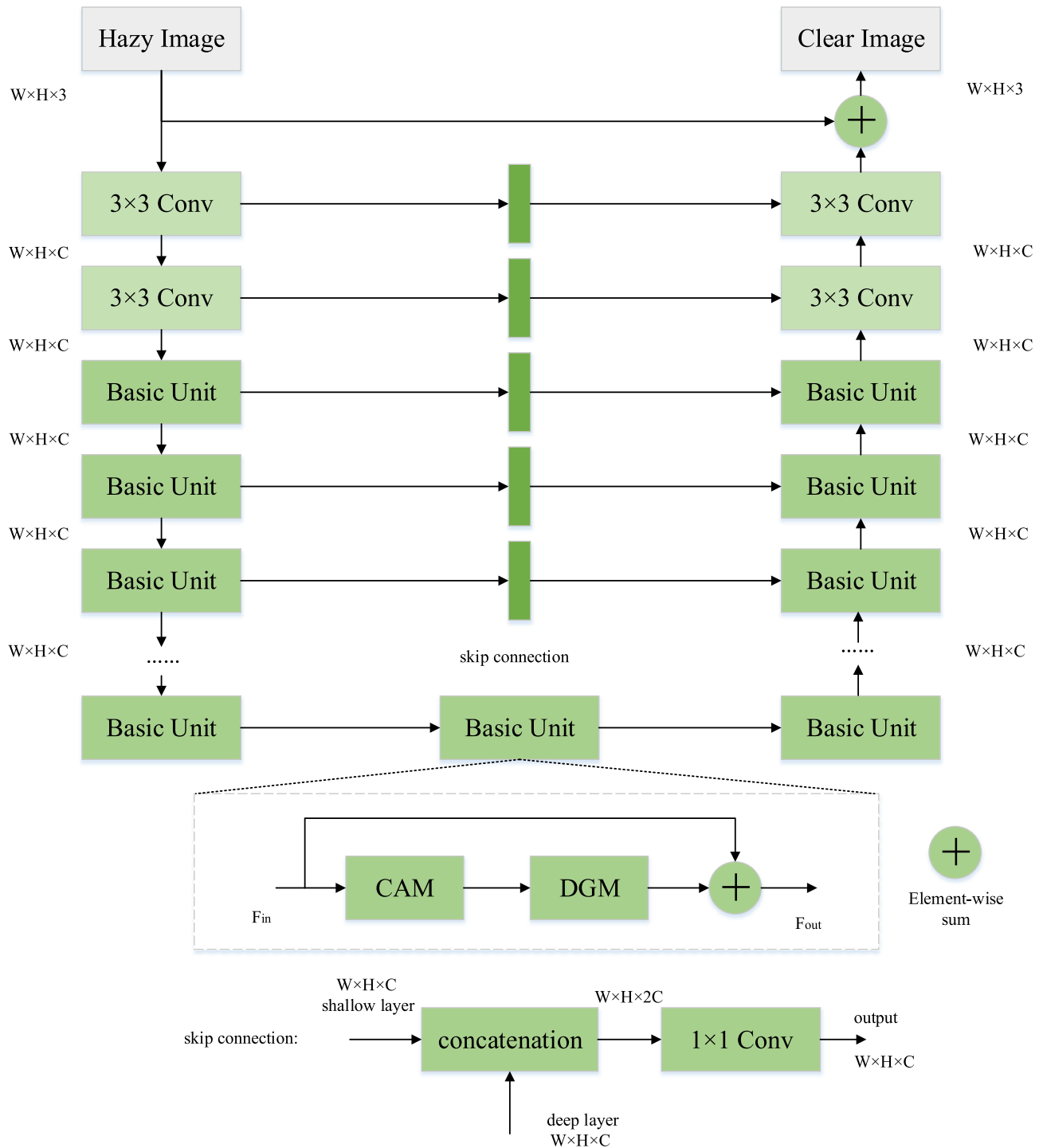


Figure 1 The architecture of the network. The network is symmetrical, it consists of two convolution layers for pre-processing, several basic units and two convolution layers for reconstructing output. The skip connections are used between shallow layers and deep layers.

$$C_p = \sigma(\text{DConv}(\delta(\text{DConv}(F_c^*)))) \tag{7}$$

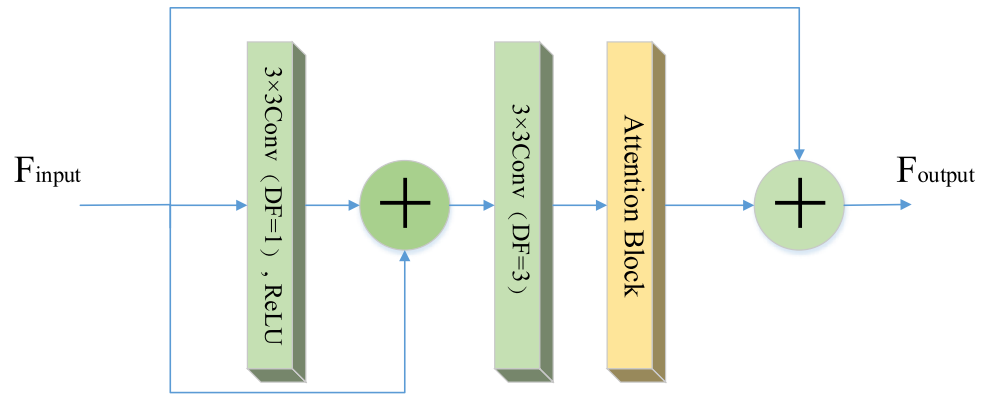
Then, we element-wise multiply F_c^* and C_p as the output of the channel-pixel attention map:

$$F_{CP} = F_c^* \odot C_p \tag{8}$$

3.1.3 Spatial Attention

Spatial attention is designed to exploit the spatial attention map from the input convolutional features F_{input} . The spatial attention module first applies global average pooling on F_{input} along the channel dimensions and outputs a

Figure 2 CAM.



feature map $f \in \mathbb{R}^{H \times W}$, the feature f then is passed through a dilated convolution layer with $DF = 1$ and sigmoid activation to get the spatial attention feature $f_{SA} \in \mathbb{R}^{H \times W}$.

Finally, the spatial attention map f_{SA} and channel-pixel attention map F_{CP} are concatenated, and then the concatenated feature map passed through a convolution layer with 1×1 kernel size to obtain the attention map.

3.2 Dual GCN Module

Although dilated convolution and attention mechanism are used in the CAM, it is still a convolution operation essentially, the feature captured from CAM is still lack in contextual information. To address the limitations of CAM, we adopt a dual GCN module to capture the contextual features

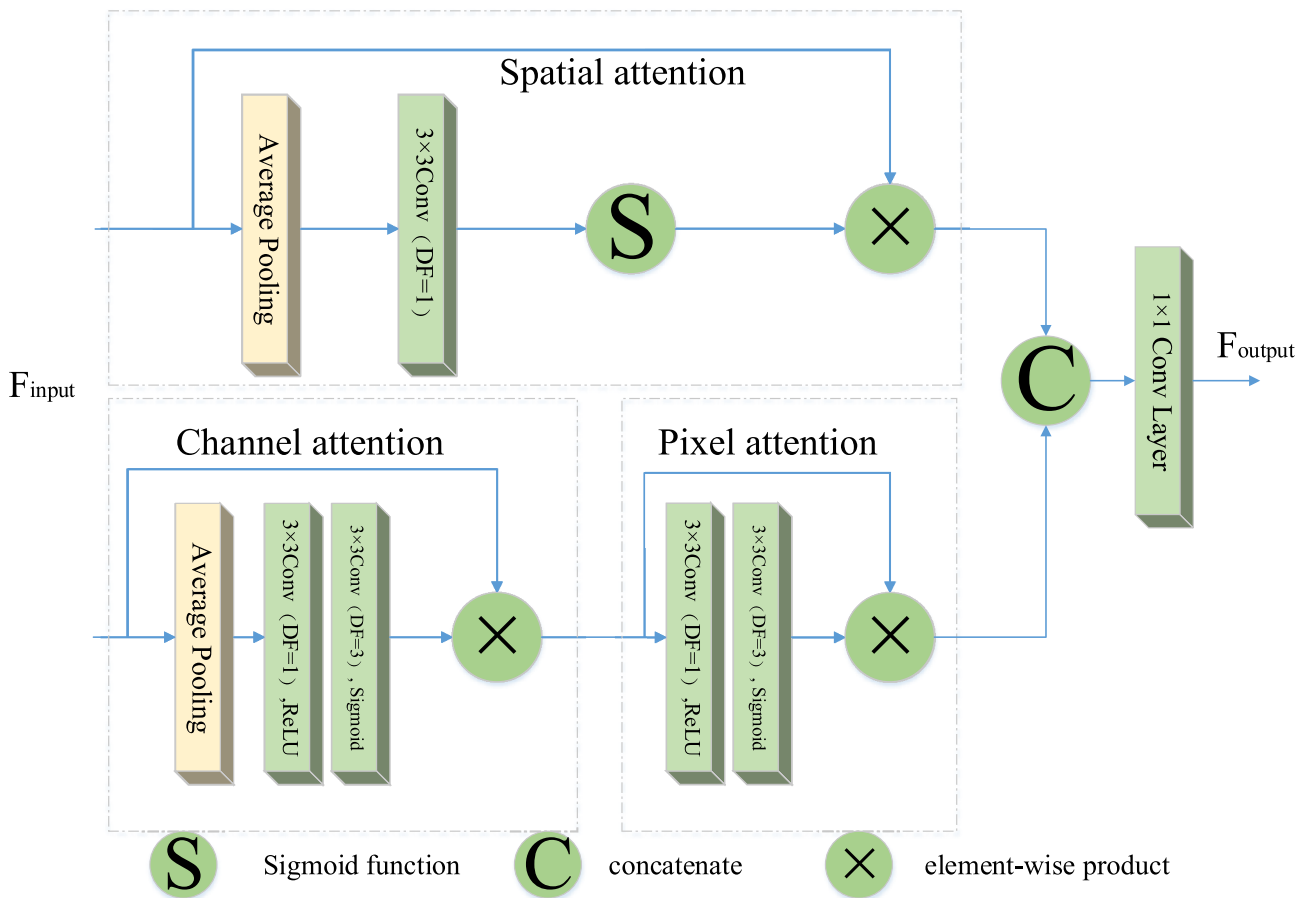
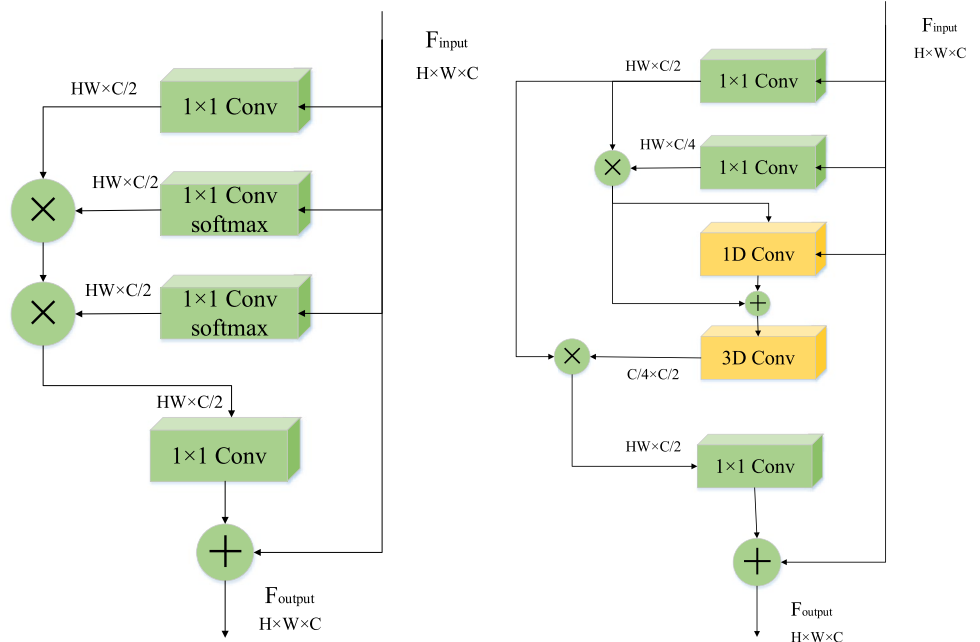


Figure 3 Attention block.

Figure 4 spatial GCN and channel GCN, the “1 × 1 Conv” means 1 × 1 convolution operation which is used to change the dimension of feature, exchange information from different channels and add non-linear activation. The “1D Conv” is dilated convolution layers with DF=1 to obtain larger receptive field. “×” means the matrix multiplication operation.



for image de-hazing. The GCN module contains spatial GCN operation and channel GCN operation. The spatial GCN [14] is designed to explore global spatial information between pixels. The channel GCN [31] is derived from the channels of feature map to explore the global information between channels. With the dual GCN module, the global spatial and channel information are captured.

Spatial GCN is designed to explore global spatial information which contains the relationship between one pixel and all other pixel in the feature map. According to [15], let $F \in \mathbb{R}^{H \times W \times C}$ be the input feature map, where H and W are the height and width of the features map F and C denotes the channels number, the GCN operation is defined as:

$$F_{GCN} = AFW \tag{9}$$

Where A and W are the adjacency matrix and the weight matrix, respectively. The pixels are the nodes of the graph, the information are passed between all the nodes, and the non-local information are extracted. The spatial GCN is

depicted in Fig. 4. As the figure shows, the input feature is processed by three convolution layer with 1×1 kernel size, and the channel size is reduced from C to $C/2$, a *softmax* operation is used in the last two convolution layer to avoid numerical instabilities [32]. With a local shortcut, the output of spatial GCN F_{sGCN} is defined as:

$$F_{sGCN} = F_{in} + A_{sGCN}F_sW_{sGCN} \tag{10}$$

where $F_s = conv(F_{in})$, A_{sGCN} is the adjacency matrix, which is calculated by matrix multiplication operation,

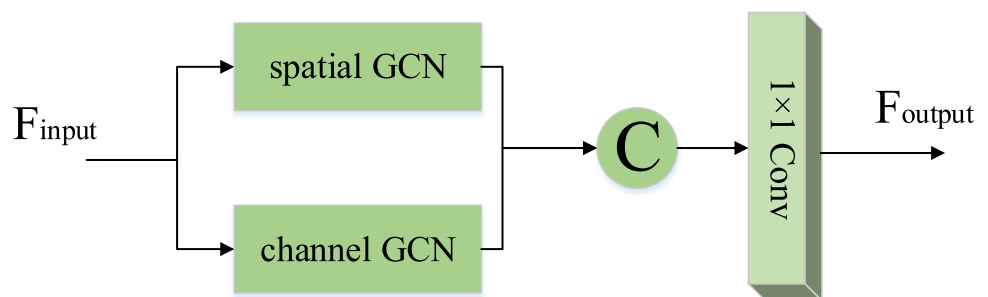
$$A_{sGCN} = conv(F_{in}) \otimes conv(F_{in})^T \tag{11}$$

and W_{sGCN} can be seen as the weight matrix, which is used to perform a hidden-to-output operation by a 1×1 convolution layer.

Channel GCN is designed to capture the channel correlations between the feature maps, the channel GCN is defined as:

$$F_{cGCN} = F_{in} + \phi(\hat{F}_c) = F_{in} + \phi(A_{cGCN}F_cW_{cGCN}) \tag{12}$$

Figure 5 The Dual GCN module, the features from two GCNs are concatenated.



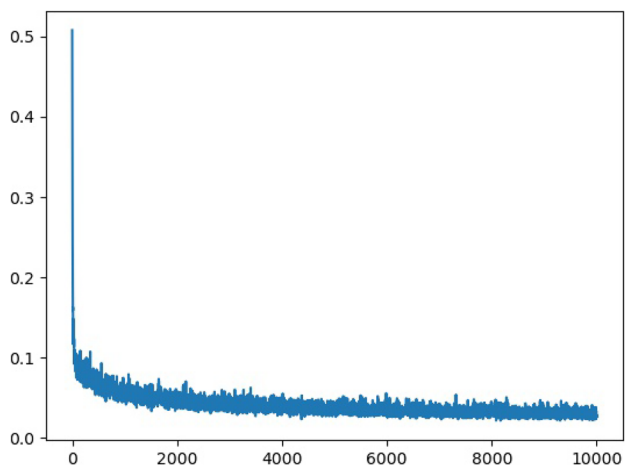


Figure 6 Loss curve on SOTS.

where A_{cGCN} can be considered as the adjacency matrix, W_{cGCN} is the weight matrix, and $\phi(\cdot)$ is used to perform a hidden-to-output operation by a 1×1 convolution layer.

Two 1×1 convolution layers $\zeta(\cdot)$ and $\kappa(\cdot)$ are adopted on the input F_{in} to aggregate different channel features, where $\zeta(F_{in}) \in \mathbb{R}^{HW \times C/2}$ and $\kappa(F_{in}) \in \mathbb{R}^{HW \times C/4}$. Then we multiply the two matrix to get the output:

$$F_c = \text{softmax}(\kappa(F_{in})^T \otimes \zeta(F_{in})) \quad (13)$$

The output feature $F_c \in \mathbb{R}^{C/4 \times C/2}$ contains $C/4$ nodes whose dimension is $C/2$, also a *softmax* operation is used to avoid numerical instabilities. An identity matrix I is used to propagate the nodes [33], and the \hat{F}_c can be calculated by:

$$\hat{F}_c = (I + A_{cGCN})(\kappa(F_{in})^T \otimes \zeta(F_{in}))W_{cGCN} \quad (14)$$

the adjacency matrix $A_{cGCN} \in \mathbb{R}^{C/4 \times C/4}$ and the weights matrix $W_{cGCN} \in \mathbb{R}^{C/2 \times C/2}$ are implemented by two dilated convolution layers with $DF=1$ and $DF=3$, respectively. Since the size of graph \hat{F}_c is $C/4 \times C/2$, a 1×1 convolution layer $\phi(\cdot)$ is used to reshape the output size to $HW \times C$.

Dual GCN module consist of the spatial GCN and the channel GCN, The architecture is depicted in Fig. 5, and the “C” means concatenation operation. We concatenate the F_{sGCN} and F_{cGCN} , and then the concatenated feature map passed through a convolution layer with 1×1 kernel size to exchange information from different channels, and finally output the global information.

3.3 Loss Function

Researchers have proposed lots of loss functions to deal with image de-hazing, such as perceptual loss, Mean squared error (MSE), GAN loss, L_2 loss and smooth L_1 loss. Since the smooth L_1 loss performs better PSNR and SSIM metrics in many image restoration tasks [34], we adopt the loss function to train our network also:

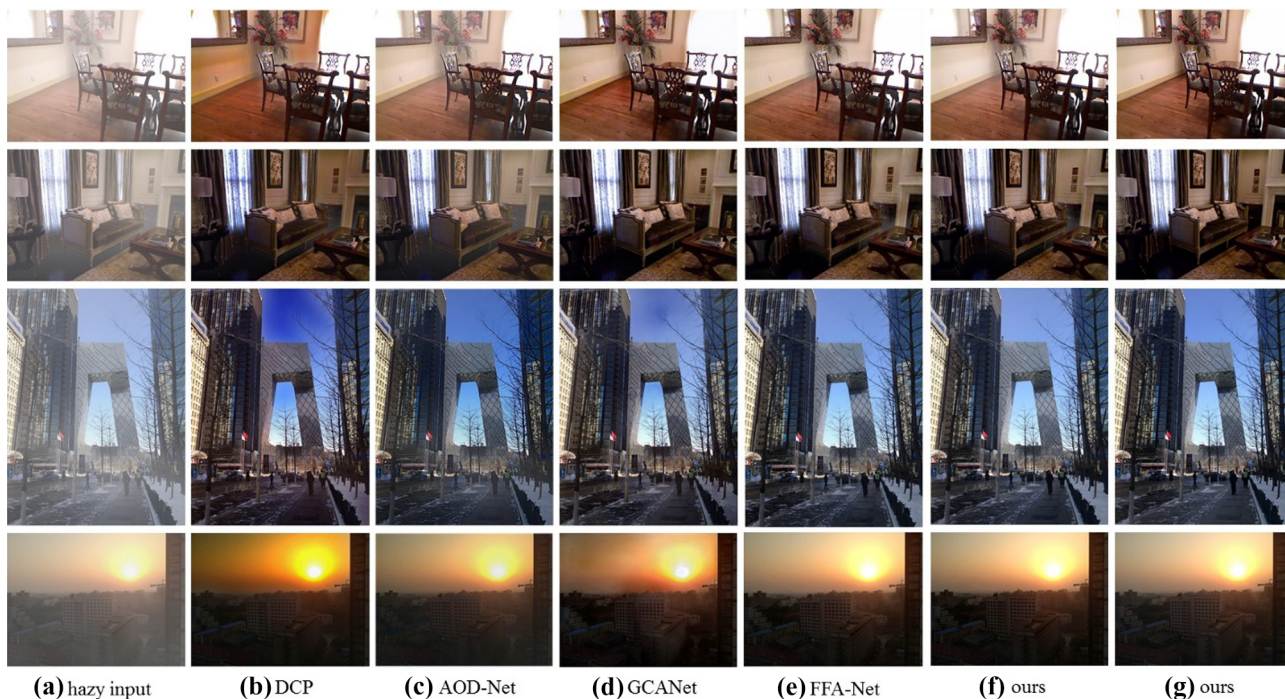


Figure 7 Qualitative comparisons on SOTS.

Table 1 Quantitative comparisons.

Method	Indoor		Outdoor		Dense-Haze		NH-HAZE	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DCP [3]	16.62	0.8179	19.13	0.8148	10.06	0.3856	10.57	0.5196
AOD-Net [20]	19.06	0.8504	20.29	0.8765	13.14	0.4144	15.40	0.5693
GCA-Net [25]	30.23	0.9800	28.68	0.9712	13.21	0.4253	17.23	0.6138
GFN [21]	22.30	0.8800	21.55	0.8444	13.96	0.4274	16.87	0.5317
DehazeNet [6]	21.14	0.8472	22.46	0.8514	13.84	0.4252	16.62	0.5283
GridDehazeNet [22]	32.16	0.9836	30.86	0.9819	13.31	0.3681	13.80	0.5370
FFA-Net [28]	36.39	0.9886	33.57	0.9849	14.39	0.4524	19.87	0.6915
Ours	37.01	0.9912	34.69	0.9903	15.36	0.4853	19.96	0.7216

$$L_s = \frac{1}{N} \sum_{x=1}^N \sum_{i=1}^3 F_s(\hat{J}_i(x) - J_i(x)) \quad (15)$$

where

$$F_s(e) = \begin{cases} 0.5e^2, & \text{if } |e| < 1, \\ |e| - 0.5, & \text{otherwise.} \end{cases} \quad (16)$$

$\hat{J}_i(x)$ and $J_i(x)$ stand the intensity of the i th color channel of pixel x in the de-hazed image and hazy image, respectively, and N is the pixel count of the image.

$A \in [0.7, 1.0]$ and scatter parameters $t \in [0.6, 1.8]$. The Outdoor Training Set (OTS) contains 8477 clean images and 296,695 hazy images that generated by the clean images with $A \in [0.8, 1.0]$ and $t \in [0.04, 0.2]$. The Synthetic Objective Testing Set (SOTS) of RESIDE is used for testing, and the SOTS contains 500 indoor images and 500 outdoor images. And the Real-world Task-driven Testing Set (RTTS) of RESIDE contains 4,322 real-world hazy images crawled from the web. The images of Dense-Haze, NH-HAZE and real-world dataset [33] are also real-world hazy images, and we evaluate the robustness of our method in the real-world.

4 Experiment Results

4.1 Datasets

We evaluate our method on several datasets including RESIDE [35], Dense-Haze [36], NH-HAZE [37] and real-world dataset [38]. RESIDE is a new benchmark for image de-hazing to from large-scale training sets, the dataset contains synthetic hazy images in both in-door and outdoor scenarios from depth dataset [39] and stereo datasets [40]. After data cleaning, the Indoor Training Set (ITS) of RESIDE contains 1399 clean images and 13,990 hazy images that generated by the clean images with global atmosphere light

4.2 Training Settings and Implementations

We resize the size of training images to 240×240 , and randomly rotate the images by $90, 180, 270$ degrees and horizontal flip the images for data augmentation. We choose the Adam optimizer for accelerated training, where β_1 and β_2 take the default values of 0.9 and 0.999, respectively. The number of Basic Unit is set to 11. We adopt the cosine annealing strategy [41] to adjust the learning rate η_t from the initial value $\eta = 1 \times 10^{-4}$, to 0 by following the cosine function:

$$\eta_t = \frac{1}{2} \left(1 + \cos\left(\frac{t\pi}{T}\right) \right) \eta \quad (17)$$

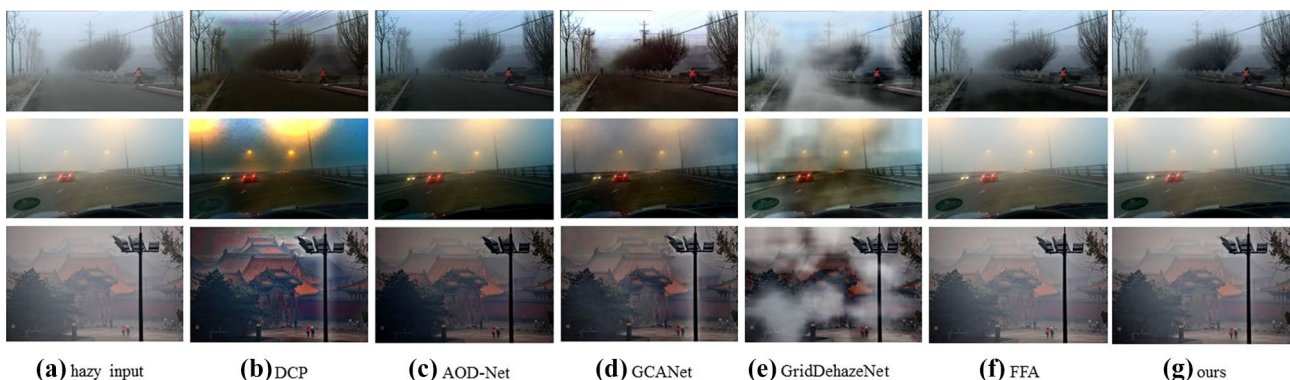


Figure 8 Qualitative comparisons on RTTS.

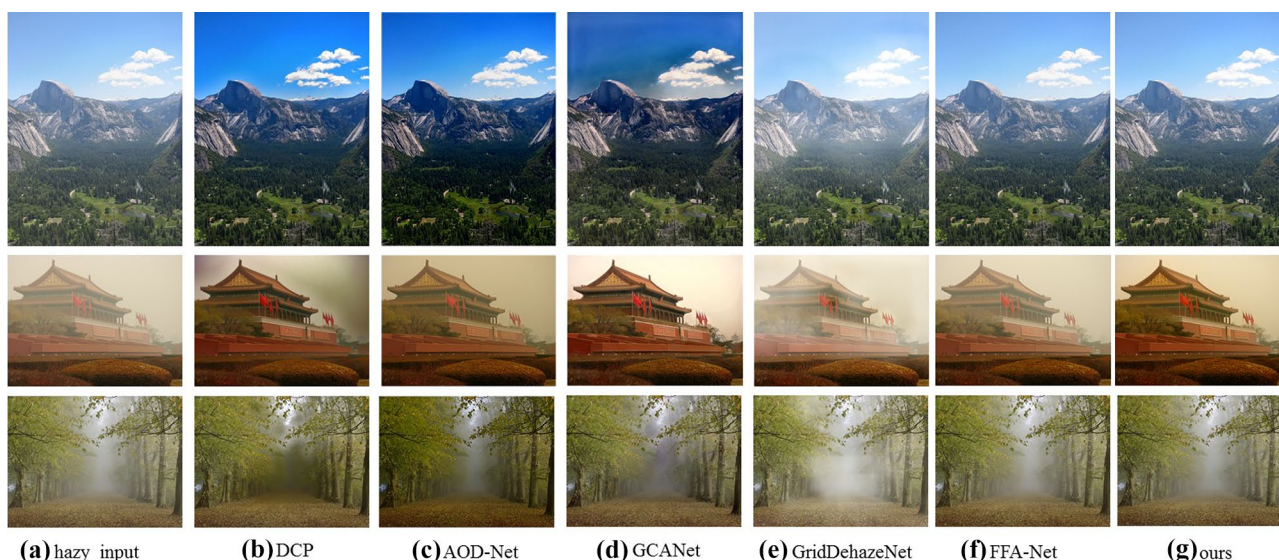


Figure 9 Qualitative comparisons on the real-world dataset.

where T is the total number of training batches, and t is the current training batch. We implement the network on Pytorch, we can observe that the loss drops fast, the model converges easily as show in Fig. 6, where X axis indicates training steps, and Y indicates loss value.

4.3 Results and Analysis

In this section, we will compare our network with previous state-of-the-art image de-hazing algorithms including the DCP [3], AOD-Net [1], GCANet [25], GFN [21], DehazeNet [6], GridDehazeNet [22] and FFA-Net [28] both quantitatively and qualitatively. Among these methods, DCP method is a prior-based method which is regarded as the baseline in single image de-hazing, and the others are all data-driven methods based on CNN. Peak signal to noise ratio (PSNR) and structure similarity (SSIM) are used for quantitative assessment of the de-hazed outputs. The quantitative comparison results are shown in Table 1.

From Table 1, it can be seen that the value of PSNR and SSIM of our proposed is better than the other methods. Compared to the FFA-Net which only capture the local information by attention net, our result of PSNR improves up to about 1%.

The qualitative comparisons of visual effect on SOTS are shown in Fig. 7. We select four images from the outdoor dataset and the indoor dataset respectively, and the upper

two rows are indoor results, the left two rows are outdoor results. The first column is the hazy input and the last column is the ground-truth, and the middle columns are de-hazed results from DCP, AOD-Net, GCANet, FFA-Net and our network respectively. From the results, we can find that the DCP method suffers from severe color distortion extremely, especially the blue sky and the halo of the sun in the outdoor images, also DCP loses some details. AOD-Net cannot remove all the hazy regions from the hazy image because of its simple network architecture, and the brightness value of the output is lower than others. GCANet also performs not well at the blue sky and the halo of the sun. FFA-Net performs as well as ours on SOTS. The images recovered from our network are almost entirely in line with real scene information, especially, the restoration of blue sky and halo images is much better.

We further give the qualitative comparisons on RTTS [35] and real-world dataset [38] in Figs. 8 and 9, respectively, the models used are all trained on RESIDE, and the results are largely consistent with those on the SOTS dataset. The DCP and GCANet still suffer from severe color distortions, and AOD-Net can't remove the haze completely and the output images are of low-brightness. GridDehazeNet also can't remove the haze completely and produce some white spots. Compares to our results, FFA-Net performs not well either such as the second image in Fig. 9. All the methods cannot remove the hazy far away such as the end of the road in the

Table 2 Comparisons of model complexity.

Method	AOD-Net	GCANet	DehazeNet	FFA	GridDehazeNet	ours
Parameters (10^6)	0.002	0.7	0.008	4.7	3.3	2.1
FLOPs (10^9)	0.2	24	0.9	530	124	246

Table 3 Ablation studies on SOTS.

CAM			DGM		PSNR
Chanel attention	Pixel attention	Spatial attention	Spatial GCN	Chanel GCN	
✓			✓		35.82
✓			✗		34.78
✗			✓		34.83
✓			✓	✗	35.42
✓			✗	✓	35.48
✓	✓	✗	✓		35.46
✓	✗	✓	✓		35.39
✗	✓	✓	✓		35.37

first row of Fig. 8 and the last row of Fig. 9, but other methods suffer from severe color distortions in the hazy regions. Above all, our method is capable of outperforming the other methods in image details and color fidelity in general.

Furthermore, we provide the model complexity comparison with SOTA methods using total parameter number and floating point operations (FLOPs) and the results are reported in Table 2. The total parameter number reflects the memory required, and FLOPs reflects the computation required. AOD-Net shows a clear advantage because of its simplest network. Compared with the SOTA methods, our net performs best but not cost the most.

4.4 Ablation Studies

In this section, we present ablation experiments to discuss the CAM and DGM of our network. The factors below are mainly concerned: (1) CAM and DGM in the basic unit; (2) the attention modules in CAM; (3) the GCN modules in DGM. Evaluation is performed on SOTS outdoor dataset with the same training epoch of each model, and the images as training input are cropped to 120×120 , and the other parameters are set the same as Section 4.2. The results are shown in Table 3. From the results, we can observe that the PSNR achieves a best value when CAM and DGM are used completely, and when some modules are not used in the Basic Unit, the PSNR value decreased.

5 Conclusion

In this work, we propose a simple yet effective network which combines CNN and GCN for image de-hazing. The network uses a CNN module with triple attention to extract local spatial information and a dual GCN module to extract broad contextual information. The CNN part combines the channel attention, spatial attention and pixel attention to earn more weight from important local features. The GCN part combines spatial

coherence computing and channel correlation computing to extract non-local information. The results in several datasets show that the proposed network outperforms the state-of-the-arts and has a powerful advantage in the restoration of image detail and color fidelity.

Acknowledgements This work is supported by Jiangsu industry research project BY2020552.

Data Availability All data generated or analyzed during this study are included in this published article [35–38].

References

- McCartney, E. J., & Hall, F. (1976). Optics of the atmosphere: Scattering by molecules and particles. *Physics Today*, 30, 76–77.
- Narasimhan, S., & Nayar, S. (2000). Chromatic framework for vision in bad weather. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 598–605). IEEE Press
- He, K., Sun, J., & Tang, X. (2011). Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12), 2341–2353.
- Ren, W., Pan, J., Zhang, H., & Yang, M. H. (2020). Single image dehazing via multiscale convolutional neural networks with holistic edges. *International Journal of Computer Vision*, 128(1), 240–259.
- Wang, H., Xie, Q., Wu, Y., Zhao, Q., et al. (2020). Single image rain streaks removal: A review and an exploration. *International Journal of Machine Learning and Cybernetics*, 11, 853–872.
- Cai, B., Xu, X., Jia, K., Qing, C., & Tao, D. (2016). DehazeNet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11), 5187–5198.
- Kim, G., Ha, S., & Kwon, J. (2018). Adaptive patch based convolutional neural network for robust dehazing. In *IEEE International Conference on Image Processing (ICIP)* (pp. 2845–2849). IEEE Press.
- Zhang, X. (2021). Research on remote sensing image de-haze based on GAN. *Journal of Signal Processing Systems*, 94, 305–313.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). IEEE Press.
- Cao, X., Zhou, F., Xu, L., Meng, D., Xu, Z., & Paisley, J. (2018). Hyperspectral image classification with Markov random fields and a convolutional neural network. *IEEE Transactions on Image Processing*, 27(5), 2354–2367.
- Yu, F., & Koltun, V. (2016). *Multi-scale context aggregation by dilated convolutions*. In ICLR.
- Kipf, T. N., & Welling, M. (2017). *Semi-supervised classification with graph convolutional networks*. In ICLR.
- Wang, X., Girshick, R., Gupta, A., & He, K. (2018). Non-local neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 7794–7803). IEEE Press.
- Zha, Z. J., Liu, J., Chen, D., & Wu, F. (2020). Adversarial attribute-text embedding for person search with natural language query. *IEEE Transactions on Multimedia*, 22(7), 1836–1846.
- Zhu, Y., Zha, Z. J., Zhang, T., Liu, J., & Luo, J. (2020). A structured graph attention network for vehicle reidentification. In *ACM MM*.
- Treibitz, T., & Schechner, Y. (2009). Polarization: Beneficial for visibility enhancement?. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 525–532). IEEE Press.
- Fattal, R. (2008). Single image dehazing. *ACM Transactions on Graphics*, 27(3), 72.
- Gibson, K. B., Vo, D., & Nguyen, T. (2012). An investigation of dehazing effects on image and video coding. *IEEE Transactions on Image Processing*, 21(2), 662–673.

19. Pleschberger, M., & Schrunner, S. (2020). An explicit solution for image restoration using Markov Random Fields. *Journal of Signal Processing Systems*, 92(2), 257–267.
20. Li, B., Peng, X., Wang, Z., Xu, J., & Feng, D. (2016). Aod-net: All-in-one dehazing network, in *Proceedings of the IEEE International Conference on Computer Vision (CVPR)* (pp. 4770–4778). IEEE Press.
21. Ren, W., Ma, L., Zhang, J., Pan, J., Cao, X., Liu, W., & Yang, M. H. (2018). Gated fusion network for single image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3253–3261.
22. Liu, X., Ma, Y., Shi, Z., & Chen, J. (2019). GridDehazeNet: Attention-based multi-scale network for image dehazing. *IEEE International Conference on Computer Vision (ICCV)* (pp. 7313–7322). IEEE Press.
23. Dong, H. (2020). Multi-scale boosted dehazing network with dense feature fusion. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2157–2167). IEEE Press.
24. Liu, X., Sukanuma, M., Sun, Z., & Okatani, T. (2019). Dual residual networks leveraging the potential of paired operations for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7007–7016).
25. Chen, D., He, M., Fan, Q. (2019). Gated context aggregation network for image dehazing and deraining. *2019 IEEE winter conference on applications of computer vision (WACV)* (1375–1383). IEEE.
26. Ally, N., Nombo, J., Ibwe, K., et al. (2021). Diffusion-driven image denoising model with texture preservation capabilities. *Journal of Signal Processing Systems*, 93, 937–949.
27. Yu, W., Huang, Z., Zhang, W., Feng, L., & Xiao, N. (2019). Gradual network for single image de-raining. In *ACMMM*.
28. Qin, X., Wang, Z., Bai, Y., Xie, X., & Jia, H. (2020). *FFA-Net: Feature fusion attention network for single image dehazing*. AAAI.
29. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: convolutional networks for biomedical image segmentation. In *MICCAI*.
30. Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 7132–7141). IEEE Press.
31. Fu, X., Qi, Q., Zhu, Y., Ding, X., & Zha, Z. J. (2021). Rain streak removal via dual graph convolutional network. AAAI.
32. Chen, Y., Kalantidis, Y., Li, J., Yan, S., & Feng, J. (2018). *A²-nets: Double attention networks*. In *NeurIPS*.
33. Chen, Y., Rohrbach, M., Yan, Z., Shuicheng, Y., Feng, J., & Kalantidis, Y. (2019). *Graph-based global reasoning networks*. In *CVPR*.
34. Lim, B., Son, S., Kim, H., Nah, S., & Lee, K. M. (2017). Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Rork-shops (CVPRW)* (pp. 136–144). IEEE Press.
35. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., & Wang, Zh. (2019). Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1), 492–505.
36. Codruta, O., Ancuti, C., & Ancuti (2019). *Mateu Sbert, and Radu Timofte. Dense haze: A benchmark for image dehazing with dense-haze and haze-free images*. In *ICIP*.
37. Ancuti, C. O., Ancuti, C., & Timofte, R. (2020). *NH-HAZE: An image dehazing benchmark with nonhomogeneous hazy and haze-free images*. *CVPRW*.
38. Fattal, R. (2014). Dehazing using color-lines. *ACM Transactions on Graphics*, 34(1), 1–14.
39. Silberman, N., Hoiem, D., Kohli, P., & Fergus, R. (2012). Indoor segmentation and support inference from RGBD images. In *European Conference on Computer Vision (ECCV)* (pp. 746–760). Springer-Verlag.
40. Scharstein, D., & Szeliski, R. (2017). High-accuracy stereo depth maps using structured light. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. I-I). IEEE Press.
41. He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., & Li, M. (2019). Bag of tricks for image classification with convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 558–567). IEEE Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.