Feature Dense Relevance Network for Single Image Dehazing

Yun Liang[1] , **Enze Huang**[1] , **Zifeng Zhang** [1] , **Zhuo Su**[2,3,⋆] and **Dong Wang**[1]
[1] Guangzhou Key Laboratory of Intelligent Agriculture, College of Mathematics and Informatics, South China Agricultural University
[2] School of Computer Science and Engineering, Sun Yat-sen University
[3] Research Institute of Sun Yat-sen University in Shenzhen
yliang@scau.edu.cn, 20203162017@stu.scau.edu.cn, tszfeng@stu.scau.edu.cn, suzhuo3@mail.sysu.edu.cn, bluesky@scau.edu.cn

Abstract

Existing learning-based dehazing methods do not fully use non-local information, which makes the restoration of seriously degraded region very tough. We propose a novel dehazing network by defining the Feature Dense Relevance module (FDR) and the Shallow Feature Mapping module (SFM). The FDR is defined based on multi-head attention to construct the dense relationship between different local features in the whole image. It enables the network to restore the degraded local regions by non-local information in complex scenes. In addition, the raw distant skip-connection easily leads to artifacts while it cannot deal with the shallow features effectively. Therefore, we define the SFM by combining the atmospheric scattering model and the distant skip-connection to effectively deal with the shallow features in different scales. It not only maps the degraded textures into clear textures by distant dependence, but also reduces artifacts and color distortions effectively. We introduce contrastive loss and focal frequency loss in the network to obtain a realistic and clear image. The extensive experiments on several synthetic and real world datasets demonstrate that our network surpasses most of the state-of-the-art methods.

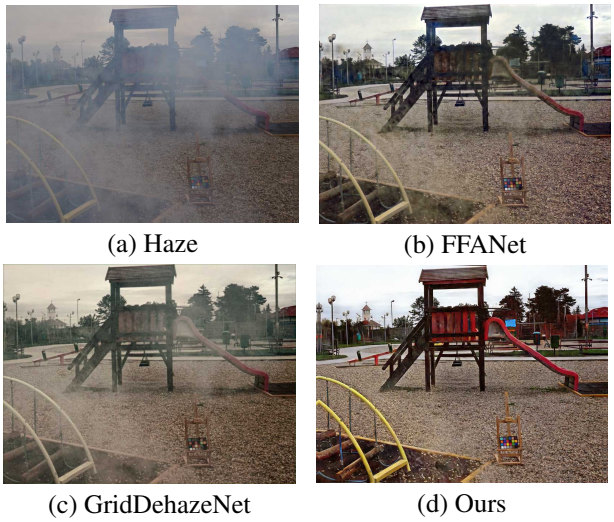
1 Introduction

Haze is a traditional atmospheric phenomenon where suspended particles obscure the clarity and reduce the contrast in the subject. Haze might cause problems for high-level visual tasks such as object detection, object recognition, etc. Therefore, image dehazing is a hot topic in computer vision. It aims to restore clear images from hazy images. The mathematical relation between the hazy image I and the clear image J is usually described as the atmosphere scattering model by

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

where T, A, x denote transmission maps, global atmospheric light and pixel position, respectively. Because of the indefinition of T and A, image dehazing is ill-posed.

⋆Corresponding author: Zhuo Su



(a) Haze (b) FFANet (c) GridDehazeNet (d) Ours

Figure 1: Comparisons of different methods in non-homogeneous hazy scene. The present methods easily cause color loss (such as the missing red color of the slide in (b)) and under dehazing as (c). Our result restores the color and fine texture (such as the distant wall).

Early dehazing methods rely on some assumptions [Zhang *et al.*, 2019]. They use statistics and other techniques to estimate transmission maps and atmospheric light. However, the assumptions do not always hold in real scenes. Thus, these methods easily lead to distortions and halo.

With the development of deep learning, many neural network approaches have been proposed to predict clear images or the transmission maps and atmospheric light to compute dehazed images. Without assumptions, the present learning-based methods remove haze better. But they have two problems: **1)** cannot use distant similar features fully [Zhou *et al.*, 2020] which are very important for dehazing [Nie *et al.*, 2021]. This problem often leads to under dehazing especially in non-homogeneous hazy scenes. **2)** cannot handle the shallow features effectively which deviate greatly from clear images [Li *et al.*, 2017b]. This problem leads to artifacts easily in dehazed images and limits the restoration of the fine textures with serious degradation.

To cope with the above problems, we propose a feature dense relevance network by defining the Feature Dense Relevance module (FDR) and Shallow Feature Mapping module (SFM). The FDR boosts the local feature by global in-

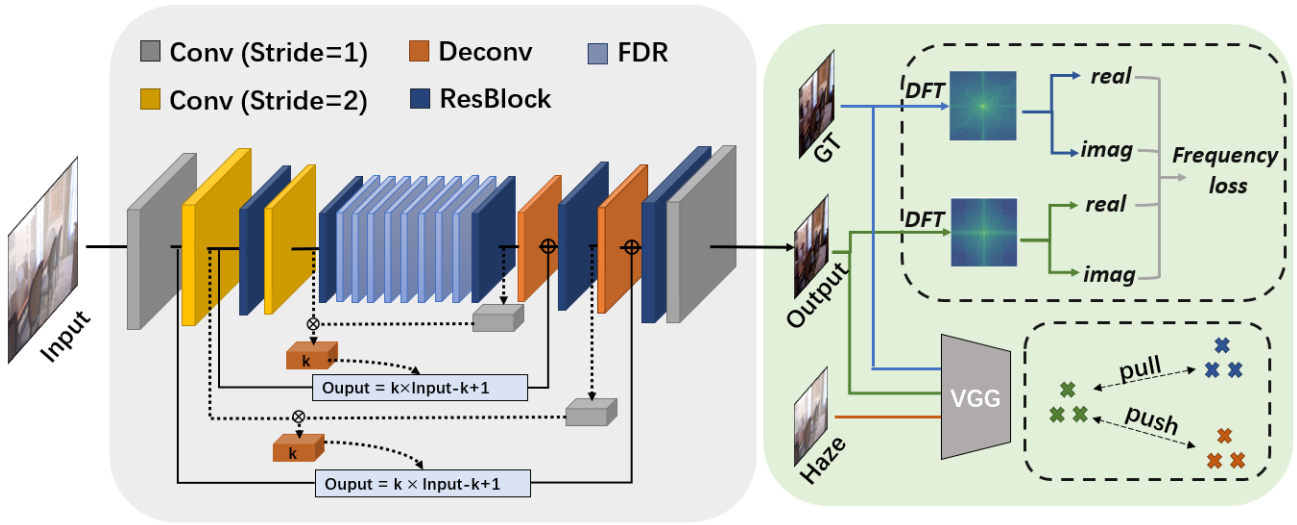


Figure 2: Overview of feature dense relevance dehazing network. The left box is network structure and the right box represents the optimization. In left box, the encoding module includes left dark blue blocks and yellow blocks. The decoding module includes right dark blue blocks and orange blocks. The light blue block is FDR. Our SFM includes the formula and the lines linking the formula with the network. In the right box, the dashed box above represents the focal frequency loss and the below one explains the principle of contrastive learning.

formation to enhance the feature from encoding module. At the same time, the shallow feature from encoding module is mapped to obtain clear texture feature by the SFM. At last, decoding module constructs the dehazed image based on clear texture features and the enhanced feature from the FDR. To optimize the network, we introduce contrastive loss [Wu *et al.*, 2021] to provide more reference for the network, and the focal frequency loss [Jiang *et al.*, 2021] to improve color accuracy in dehazing as Figure 1. Overall, our contributions are as follows:

- We propose a feature dense relevance network based on FDR and SFM for dehazing as Figure 2. The FDR with global receptive field aims to construct the dense relationship between different local features. It has the ability to recover non-homogeneous degraded local areas using non-local information.
- Our SFM is defined based on skip-connection and physical model to achieve shallow features mapping. It makes dehazed images much more realistic while the seriously degraded fine textures are mapped into clear textures.
- Contrastive learning is introduced in our dehazing network to learn the difference between hazy images and clear images for efficiently dehazing. Besides, we utilize the focal frequency loss in the network to obtain dehazed images with more accurate color and details.

2 Related Work

Current single image dehazing methods can be divided into prior-based methods and learning-based methods.

Prior-based dehazing methods. Many such methods calculate transmission maps and atmospheric light based on some prior assumptions [Raanan and Fattal, 2014; Berman *et al.*, 2016; Tan, 2008; Zhang *et al.*, 2019]. He et al. [He *et al.*, 2011] define the dark channel in clear images to estimate the transmission maps. On the basis of dark channel prior, Hu et al. [Hu *et al.*, 2019] assume that the atmospheric light in global and local is equal and define the equation to compute the transmission maps. Despite sometimes prior-based methods remove haze well, they rely on assumptions and easily cause halo and distortions when assumptions cannot hold.

Learning-based dehazing methods. With the development of deep learning, a number of methods apply CNNs for dehazing. Without essential assumptions, these methods break down the limitation of prior. Generally, learning-based methods are classified into the methods to predict the transmission maps or atmospheric light and end-to-end methods. Deep neural network is firstly used in predicting the transmission maps or atmospheric light [Ren *et al.*, 2016]. Cai et al. [Cai *et al.*, 2016] begin to use CNN to estimate the transmission maps. Zhang et al. [Zhang and Patel, 2018] define a densely connected pyramid network to predict the transmission maps and atmospheric light. However, such methods always cause distortions because they ignore the relationship between the atmospheric light and transmission maps.

In order to reduce the dependence of physical model, end-to-end methods [Chen *et al.*, 2019; Shyam *et al.*, 2021; Qin *et al.*, 2019] are designed to estimate clear images directly. Li et al. [Li *et al.*, 2017a] robustly estimate clear image directly by predicting the joint parameter of the transmission maps and atmospheric light. Dong et al. [Dong *et al.*, 2020a] propose a multi-scale boosted dehazing network using error-feedback strategy to restore the texture. Although end-to-end methods have better robustness, they often have under dehazing in complex scenes for lacking flexibility. To solve this problem, attention is applied in end-to-end methods. Liu et al. [Liu *et al.*, 2019] propose a GridDehazeNet by using channel attention which treats different channels with

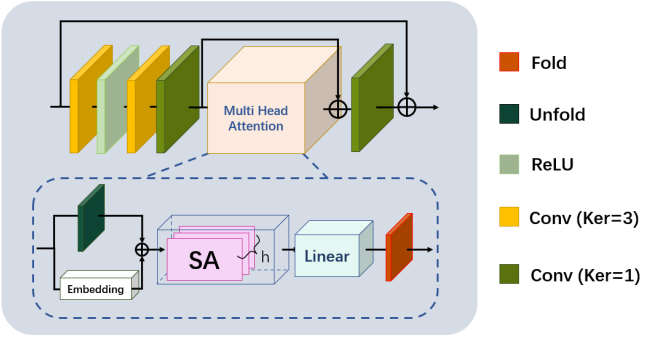


Figure 3: The structure of FDR. Multi-head attention consists of SAs. SA represents self-attention and h is the number of heads. FDR first uses convolution to extract features and expands three-dimensional features into two-dimensional sequences, then uses multi-head attention to construct the dense relevance of features.

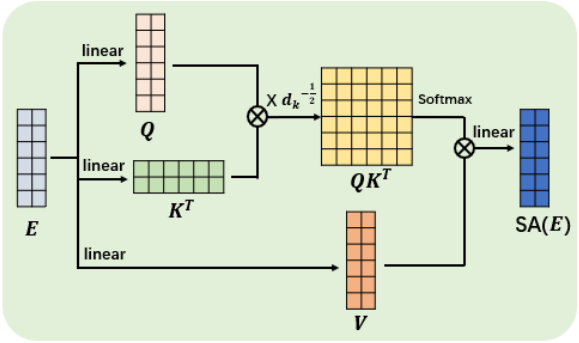


Figure 4: Overview of self-attention (SA). Q, K, V are generated by linear function according to input sequence E. QK^T enables each element in sequence E to affect each other.

different importance. Channel attention enriches the channel information to increase the robustness in complex scenes. By fusing channel attention and pixel attention, Wu et al. [Wu *et al.*, 2021] design a AECR-Net to treat different areas distinctively to improve the dehazing. However, the convolution operations can not take advantage of the information outside local region. Therefore, the present end-to-end methods often bring artifacts and distortions. Thus, it is important for dehazing to establish the relationship between different local features in the global.

At present, transformer is widely applied in many visual tasks [Liu *et al.*, 2021; Zheng *et al.*, 2021]. Multi-head attention in transformer is able to establish dense relevance in feature space [Liang *et al.*, 2021]. It can enhance the local feature by effectively utilizing non-local information. Inspired by transformer, we propose a feature dense relevance network for image dehazing. The network aims to establish dense relationship between different local features globally. It also includes a module based on atmospheric scattering model to successfully map shallow features. Our method breaks the bottleneck of the current dehazing network and produces more realistic and clearer dehazed images.

3 Method

Our feature dense relevance network is constructed by defining the FDR, the SFM, encoding module and decoding module. Additionally, we introduce focal frequency loss and contrastive loss to bring more references for the optimization of network which finally produce more realistic dehazed images. The network is shown in Figure 2.

3.1 Feature Dense Relevance

The existing learning-based methods often treat the local features as the processing unit. Actually, according to the imaging principle, a large number of similar local features can be extracted in the global scope of an image. These similar features provide valuable reference for each other in dehazing. For example, the local features describing light-degraded regions provide important clues for restoring the heavy-degraded regions with similar features. Therefore, the similarity relationship and the mutual support among local features is important for dehazing. This paper proposes a Feature Dense Relevance module (FDR) via multi-head attention as shown in Figure 3. It efficiently constructs similarity relationship and supports to adjust local features by non-local information for the generation of natural and clear images.

Our multi-head attention with global receptive field is able to establish the dense relationship between different local features. The multi-head attention is constructed by self-attentions and one self-attention supports the calculation of one head shown as Figure 4. One head is formulated as:

$$Atten(E) = Softmax(\frac{Q(E)K(E)^T}{\sqrt{d_k}})V(E), \quad (2)$$

where E is the input sequence. Q, K, V denote the linear operations to generate query, key and value. d_K is the dimension of K. By multiplying the matrix of self-attention with the query and the key, we construct the dense relevance between each element in sequence E. $V(E)$ represents the mapping of each element itself when it can not be associated with other elements. This mapping can enhance the element by itself. In multi-head attention, different heads learn supports between different local features in the global range to make the dehazing network more expressive.

We propose the FDR according to the above multi-head attention. As shown in Figue 3, the principle of FDR is as follows: firstly, we use convolution to extract feature maps and apply $n \times m$ sliding window to obtain a sequence from them. The features in a window are expanded as a $1 \times n \times m$ element of the sequence. Significantly, the sequence contains all information in the feature space and different element represent different local features. Secondly, we use the multi-head attention (as the SAs in Figure 3) to construct the similarity relationship between each element and all elements in the sequence to obtain the dense relevance of local features in the global range. To make up for the loss of spatial information caused by serialization, we design a learnable embedding encoder to generate location embeddings as the Embedding in Figure 3. Finally, the output of the multi-head attention is handled by Linear and then reshaped to the feature maps as the output of FDR.

We design a sub-network using eight FDRs shown as the light blue block in Figure 2 to achieve the dense relevance

inside the features. In the sub-network, the eight FDRs are connected one by one and the output feature maps of previous FDR are used as the input of the next one. This deep structure enhances the feature maps gradually which leads to excellent robustness. For each FDR owning one multi-head attention, the sub-network has eight multi-head attentions. The head numbers of them are respectively set to [1, 2, 8, 16, 16, 8, 2, 1]. The embedding encoder of FDR is defined by convolution, average pooling and unfold operation.

3.2 Shallow Feature Mapping

The shallow feature is seriously degraded by haze. Existing methods often use distant skip-connection to preserve the texture of an image (such as edges and corners). However, they usually bring artifacts because the shallow features deviate greatly from clear images. Hence, the mapping from degraded shallow feature to clear feature is crucial for detail reconstruction in dehazing. This paper proposes a Shallow Feature Mapping module (SFM) which introduces physical model into distant skip-connection. It maps fine textures with heavy degradation to clear textures and deal with shallow features effectively to obtain a clear image without artifacts.

The SFM is defined by the physical model for dehazing from AOD-Net [Li *et al.*, 2017a]. The physical model is:

$$J(x) = K(x)I(x) - K(x) + b, \tag{3}$$

where K, I, J respectively denote the joint parameters of atmospheric light and transmission map, hazy image and clear image. b is the constant deviation and x is the pixel position. Different from the direct skip-connection, the SFM first learns the joint parameter K in the feature space. Then the shallow features are mapped according to formula (3) and will be used to restore the texture details.

There are three scales of features in our network: the small scale, the medium scale and the large scale. We obtain the joint parameter K on the smaller scale features and apply it to the mapping of larger scale features in turn. We calculate the joint parameter as follows. First we define the small-scale/medium-scale feature obtained by the encoding module as F_e. Second, we define the feature of the same scale in the decoding module as F_d. Finally, we define equation (4) to calculate K according to formula (3). The reason is, in the same scale, F_e is the feature closest to the hazy image and F_d is the feature closest to the clear image.

$$K = \frac{F_d - b}{F_e - 1}, \tag{4}$$

We set $b = 1$. To prevent gradient explosion, we use convolution and sigmoid to approximate the result of $\frac{1}{F_e - 1}$. After obtaining K, we upsample K and apply it to map the adjacent larger-scale shallow feature using formula (3).

As shown in Figure 2, we use two SFMs in the network. Each of them is defined by a formula and the lines linking the formula with the network. We take the outermost SFM as an example. First, we use the output of the first downsampling as F_e (the output of the leftmost yellow block). Second, we set the input of the latest upsample as F_d (the input of the rightmost red block). Finally, we use F_e and F_d to obtain K which will be upsampled next. With the K, we map the outermost shallow feature (the output of the leftmost grey block) by formula (3) into the decoding module to generate a clear image with more fine textures.

3.3 Loss Function

We use the following four loss functions to construct the overall loss: L1 loss, perception loss, contrastive loss and focal frequency loss.

L1 Loss. L1 loss constructs the L1 distance between the GroundTruth (GT) and the dehazed image to describe the loss of the network. It makes the output of our network closer to the GT in the raw space to generate a clear image.

Perception Loss. We compare the feature difference between the GT and dehazed image using pre-trained VGG layer by layer. This loss ensures the features of dehazed images closer to the features of the GT in multi-dimension space. It enables the network to restore the structure of clear images.

Contrastive Loss. Contrastive loss is expressed as:

$$L_c(\theta) = \sum_{i=1}^{n} \frac{||V_i(I_{gt}) - V_i(D(I_{haze}))||_2}{||V_i(I_{haze}) - V_i(D(I_{haze}))||_2}, \tag{5}$$

where V_i denotes VGG feature extractor on the i^{th} layer. D refers to the dehazing network and θ is the parameters. I_{haze} is the hazy image and I_{gt} is the clear image. We set the I_{gt} as positive sample and the I_{haze} as negative sample. Contrastive loss can make the dehazed image close to the positive sample and away from the negative sample. The use of negative samples provides more references for the network to improve the training efficiency and dehazing performance.

Focal Frequency Loss. Focal frequency loss constrains the network in frequency domain. It not only improves the accuracy in frequency, but also pays more attention to recover the seriously degraded areas. It is expressed as:

$$L_f(\theta) = \frac{1}{ab}\rho|T(I_{gt}) - T(D(I_{haze}))|^2, \tag{6}$$

where a, b represent the height and the width of an image respectively. T is the function to obtain complex frequency value which separates the real part and imag part by fourier transform (DFT) and Euler's formula [Jiang *et al.*, 2021]. ρ denotes the weight by:

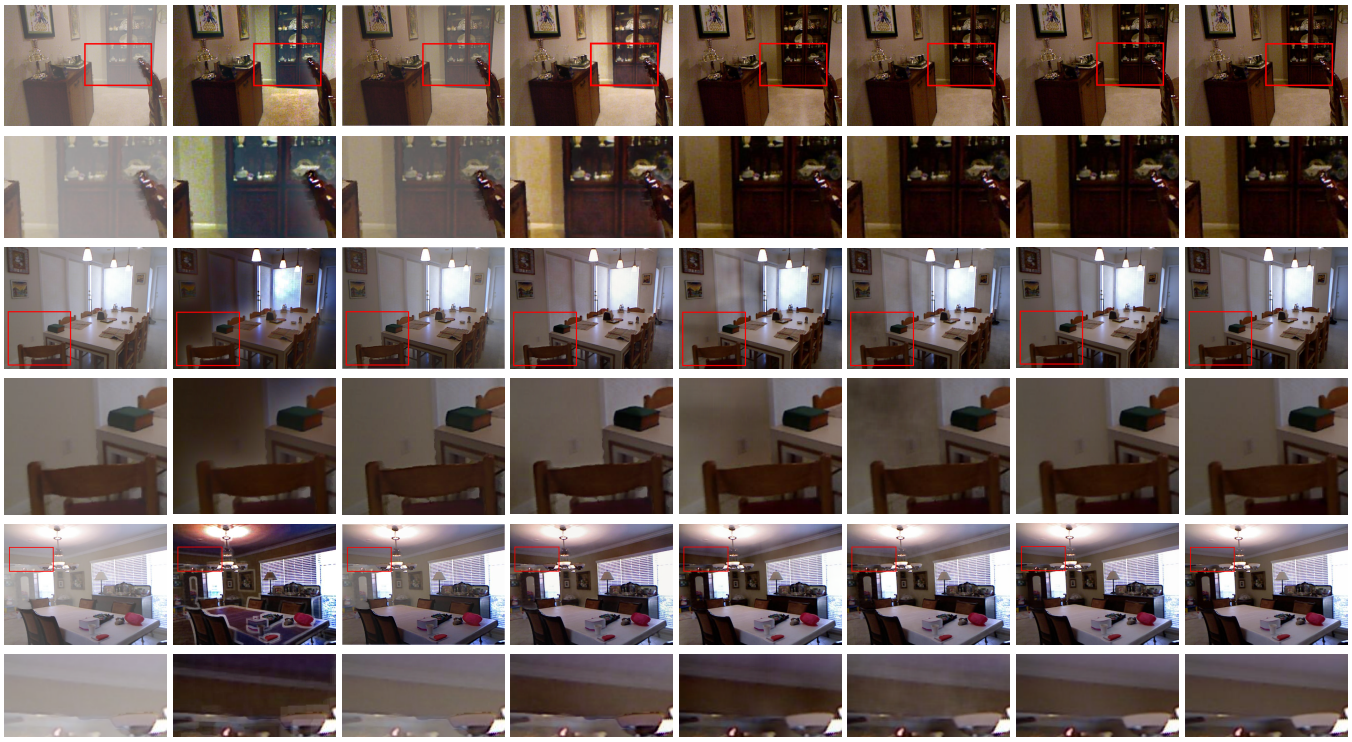
$$\rho = |T(I_{gt}) - T(D(I_{haze}))|^{\alpha}, \tag{7}$$

where α is the scaling factor for flexibility (we set α=1). ρ adjusts the weight dynamically depending on different degradation to improve the performance in the severely degraded areas. ρ is normalized into the range [0, 1].

Our total loss function is: $L = L_1 + \gamma_1 L_p + \gamma_2 L_c + \gamma_3 L_f$, where L_1 and L_P mean L1 loss and perceived loss. The γ denotes the weight. According to [Li *et al.*, 2019], we train the network with the settings: γ_1=0.03, γ_2=0.03, γ_3=0.02.

4 Experiments

The initial learning rate is set as 0.0001 and we adopt the cosine annealing strategy [Qin *et al.*, 2019] to adjust the learning rate. The Adam optimizer is used whose betas parameter



(a) Haze (b) DCP (c) AOD (d) EPDN (e) GDN (f) FFANet (g) Ours (h) GT

Figure 5: Qualitative comparisons of SOTS indoor dataset. The plural row is the area in the red box of the singular row. In the indoor synthetic scenes, our method dehazes the image better than most of the others which usually bring artifacts.

DataSet	DCP	AODNet	GCA	GDN	FD-GAN	MSBDN	FFANet	ACER-Net	Ours
SOTS	16.61/0.83	19.82/0.81	30.23/0.89	32.16/0.95	22.15/0.89	33.79/0.98	36.39/0.98	37.17/0.99	37.35/0.98
O-Haze	16.55/0.68	15.13/0.60	16.55/0.68	20.97/0.70	21.68/0.73	17.15/0.33	21.37/0.74	-	24.52/0.82
Dense-Haze	12.20/0.39	12.63/0.39	10.99/0.39	13.67/0.39	14.94/0.45	14.72/0.37	14.60/0.49	15.80/0.46	15.31/0.55
NH-Haze	11.63/0.48	11.54/0.47	12.31/0.50	15.69/0.64	19.87/0.73	16.28/0.48	17.89/0.70	19.88/0.71	20.19/0.77

Table 1: Quantitative comparisons (PSNR/SSIM) of our network and other methods. The best and second best are shown by red and green.

is remained default. Besides, for data augmentation, we use 240×240 window to randomly cut the training images and rotate them randomly. Our network is trained by RTX 3090.

4.1 Datasets

RESIDE. RESIDE [Li *et al.*, 2018] includes synthetic and real-world hazy/clear images which are collected indoor and outdoor. Its Indoor Training Set (ITS) contains 1,399 clean images and 13,990 hazy images generated by different atmosphere light and transmission maps. Its Synthetic Objective Testing Set (SOTS) contains 500 indoor and 500 outdoor test images. In this paper, we train the dehazing networks on all the images of ITS and test them on all the images of SOTS.

NTIRE2018/2019/2020. The O-Haze [Ancuti *et al.*, 2018], Dense-Haze [Ancuti *et al.*, 2019] and NH-Haze [Ancuti *et al.*, 2020] are from NTIRE 2018, 2019 and 2020 respectively. Each of them has its own training set and test set. We separately use their training sets to train the networks and use the corresponding test sets to evaluate the dehazing performance.

4.2 Comparison with State-of-the-art Methods

Quantitative comparisons. We compare our method with eight dehazing methods, including AODNet [Li *et al.*, 2017a], DCP [He *et al.*, 2011], GCANet [Chen *et al.*, 2019], GDN [Liu *et al.*, 2019], MSBDN [Dong *et al.*, 2020a], FD-GAN [Dong *et al.*, 2020b], FFANet [Qin *et al.*, 2019] and ACER-Net [Wu *et al.*, 2021]. Except ACER-Net, the compared methods are trained by our experiment enviroment and the optical model. The quantitative comparisons are shown in Table 1. Our network performs better than most of others on the SOTS indoor dataset with PSNR of 37.35 db and SSIM of 0.98. For the other datasets, our method outperforms most of the top algorithms. Compared with the eight dehazing methods, our method is always ranked as the first or the second.

Indoor qualitative comparisons. Figure 5 shows the qualitative comparisons on the indoor dataset. Our method maintains the color consistency of objects when other methods create artifacts (see the wall in the row 4). Besides, our dehazed image is realistic while the others lead to distortions (see the

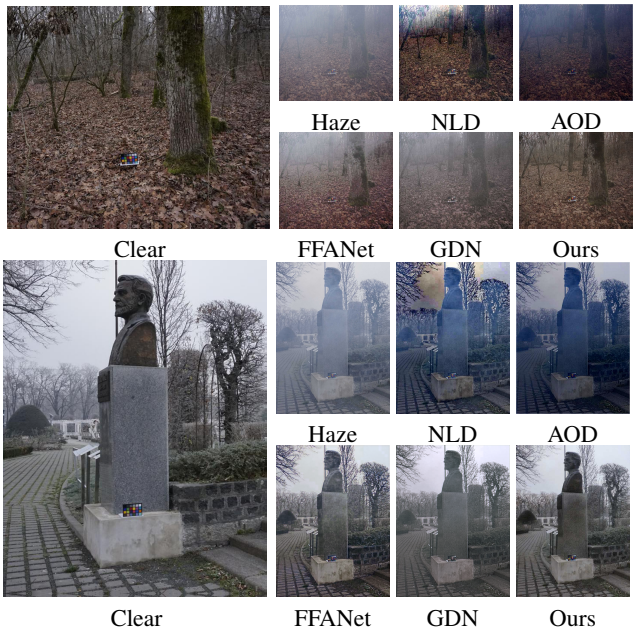


Figure 6: Qualitative comparisons of O-Haze. In the real-world outdoor scenes, our method obtains a more realistic image than others.

ceiling in the row 6). Overall, the results of our method are visually closer to the GT than other methods.

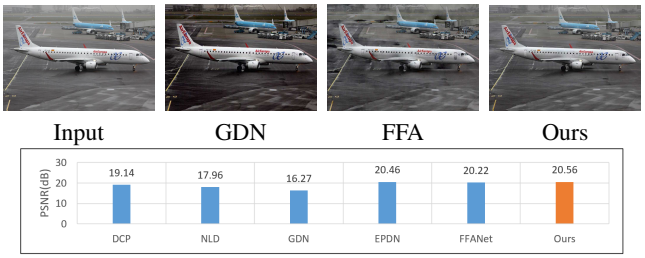
Outdoor qualitative comparisons. As shown in Figure 6, in outdoor complex scenes, our network removes haze best while others lead to under dehazing (see the statue). We restore the details well while the others are unsatisfactory (see the trees in the top left of the woods), because we establish mutual support between different local features. The outdoor comparison shows that our method has great robustness.

Generalization. We train the methods on ITS, and test them on the real-world dataset and the SOTS outdoor dataset to verify the generalization. As Figure 7, the top row shows our method produces more realistic than the others in generalizing to real-world dataset. The down row shows that our method performs best in generalizing to outdoor dataset.

4.3 Ablation Study

We define a baseline consisting of decoding module, encoding module and FDRs. We add different modules to the baseline for ablation study: **(1) Baseline+SC:** Skip connections is added between encoding module and decoding module. **(2) Baseline+SFM:** SFM is added into baseline. **(3) Baseline+SFM+CL:** Contrastive loss is used for training the network of (2). **(4) Baseline+SFM+CL+FL (Ours):** focal frequency loss is added on (3). **(5) FANet:** We replace the FDR in **Ours** with FAblock [Qin *et al.*, 2019] which also contains attention mechanism to explore the effect of FDR.

Effectiveness of SFM. The SFM effectively enhances the shallow features of the network. As shown in Table 2, adding SFM can improve the performance with an increase of PSNR 0.33db from the baseline to the baseline+SFM. In order to further explore whether the physical model works, we replace the SFM with SC to do comparison. SFM improve the performance with an increase of PSNR 0.95db from baseline+SC to baseline+SFM.



Comparison on outdoor dataset.

Figure 7: Experiments about the generalization. Qualitative comparisons of real-world is shown on the top row and the quantitative analysis on outdoor dataset is shown on the down row.

Model	CL	FL	PSNR	SSIM
Baseline	-	-	34.92	0.97
Baseline+SC	-	-	34.30	0.97
Baseline+SFM	-	-	35.25	0.97
Baseline+SFM	✓	-	36.89	0.98
FANet	✓	✓	35.78	0.97
Baseline+SFM	✓	✓	**37.35**	**0.98**

Table 2: Result of ablation study in SOTS dataset. CL means contrastive loss and FL means focal frequency loss.

Effectiveness of focal frequency loss. The focal frequency loss effectively reduces the distortion of the dehazed image. by comparing from Baseline+SFM+CL to ours, the PSNR is 0.46db higher than without focal frequency loss.

Effectiveness of FDR. FAblock can also achieve the three dimensional attention mechanism, but it is based on convolution and can not build the correlation between local features in the global range. The PSNR of using FDR is 1.57db higher than that of FAblock. Experiments show that the FDR is effective, and necessary to establish dense feature relevance.

5 Conclusion

This paper proposes a feature dense relevance network for image dehazing. This network is defined by the FDR and the SFM. The FDR aims to construct the dense relationship between each local feature in the global range. It enables different local regions to support each other to efficiently restore the seriously degraded regions in complex scenes. At the same time, the SFM is proposed based on physical model and distant skip-connection to effectively recover the fine textures by preserving and mapping the shallow features. A large number of experiments in real-world and synthetic datasets show that, our method produces more realistic and clearer images than most of the state-of-the-art dehazing methods based on the quantitative and qualitative evaluations.

Acknowledgments

This research was supported by the Key R&D Project of Guangzhou No.202206010091, the Science and Technology Planning Project of Guangdong Province No. 2019A050510034, the Guangzhou Key Laboratory of Intelligent Agriculture No.201902010081, the Shenzhen Science and Technology Program No.JCYJ20200109142612234 and the Guangdong Basic and Applied Basic Research Foundation No.2021A1515012313.

References

[Ancuti *et al.*, 2018] Codruta O Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In *CVPRW*, pages 754–762, 2018.

[Ancuti *et al.*, 2019] Codruta O Ancuti, Cosmin Ancuti, Mateu Sbert, and Radu Timofte. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In *ICIP*, pages 1014–1018. IEEE, 2019.

[Ancuti *et al.*, 2020] Codruta O Ancuti, Cosmin Ancuti, and Radu Timofte. Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In *CVPRW*, pages 444–445, 2020.

[Berman *et al.*, 2016] D. Berman, T. Treibitz, and S. Avidan. Non-local image dehazing. In *CVPR*, 2016.

[Cai *et al.*, 2016] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet. *IEEE Transactions on Image Processing*, 2016.

[Chen *et al.*, 2019] D. Chen, M. He, Q. Fan, J. Liao, and G. Hua. Gated context aggregation network for image dehazing and deraining. In *WACV*, 2019.

[Dong *et al.*, 2020a] H. Dong, J. Pan, L. Xiang, Z. Hu, and M. H. Yang. Multi-scale boosted dehazing network with dense feature fusion. In *CVPR*, 2020.

[Dong *et al.*, 2020b] Yu Dong, Yihao Liu, He Zhang, Shifeng Chen, and Yu Qiao. Fd-gan: Generative adversarial networks with fusion-discriminator for single image dehazing. In *AAAI*, volume 34, pages 10729–10736, 2020.

[He *et al.*, 2011] K. He, S. Jian, Fellow, IEEE, and X. Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 33(12):2341–2353, 2011.

[Hu *et al.*, 2019] Hai-Miao Hu, Hongda Zhang, Zichen Zhao, Bo Li, and Jin Zheng. Adaptive single image dehazing using joint local-global illumination adjustment. *IEEE Transactions on Multimedia*, 22(6):1485–1495, 2019.

[Jiang *et al.*, 2021] Li. Jiang, B. Bai, W. Wu, and C. Chen. Focal frequency loss for image reconstruction and synthesis. In *ICCV*, 2021.

[Li *et al.*, 2017a] B. Li, X. Peng, Z. Wang, J. Xu, and F. Dan. Aod-net: All-in-one dehazing network. In *ICCV*, 2017.

[Li *et al.*, 2017b] Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, and Ming-Hsuan Yang. Universal style transfer via feature transforms. *Advances in neural information processing systems*, 30, 2017.

[Li *et al.*, 2018] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018.

[Li *et al.*, 2019] Lerenhan Li, Yunlong Dong, Wenqi Ren, Jinshan Pan, Changxin Gao, Nong Sang, and Ming-Hsuan Yang. Semi-supervised image dehazing. *IEEE Transactions on Image Processing*, 29:2766–2779, 2019.

[Liang *et al.*, 2021] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *ICCV*, pages 1833–1844, 2021.

[Liu *et al.*, 2019] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *ICCV*, 2019.

[Liu *et al.*, 2021] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *ICCV*, 2021.

[Nie *et al.*, 2021] Ying Nie, Kai Han, Zhenhua Liu, An Xiao, Yiping Deng, Chunjing Xu, and Yunhe Wang. Ghostsr: Learning ghost features for efficient image super-resolution. *arXiv preprint arXiv:2101.08525*, 2021.

[Qin *et al.*, 2019] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *AAAI*, 2019.

[Raanan and Fattal, 2014] Raanan and Fattal. Dehazing using color-lines. *ACM Transactions on Graphics (TOG)*, 34(1), 2014.

[Ren *et al.*, 2016] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *ECCV*, pages 154–169. Springer, 2016.

[Shyam *et al.*, 2021] Pranjay Shyam, Kuk-Jin Yoon, and Kyung-Soo Kim. Towards domain invariant single image dehazing. In *AAAI*, volume 35, pages 9657–9665, 2021.

[Tan, 2008] R. T. Tan. Visibility in bad weather from a single image. In *CVPR*, 2008.

[Wu *et al.*, 2021] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *CVPR*, pages 10551–10560, 2021.

[Zhang and Patel, 2018] He Zhang and Vishal M. Patel. Densely connected pyramid dehazing network. In *CVPR*, pages 3194–3203, 2018.

[Zhang *et al.*, 2019] Yunfeng Zhang, Ping Wang, Qinglan Fan, Fangxun Bao, Xunxiang Yao, and Caiming Zhang. Single image numerical iterative dehazing method based on local physical features. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(10):3544–3557, 2019.

[Zheng *et al.*, 2021] Sixiao Zheng, Jiachen Lu, Hengshuang Zhao, Xiatian Zhu, Zekun Luo, Yabiao Wang, Yanwei Fu, Jianfeng Feng, Tao Xiang, Philip HS Torr, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In *CVPR*, pages 6881–6890, 2021.

[Zhou *et al.*, 2020] Shangchen Zhou, Jiawei Zhang, Wangmeng Zuo, and Chen Change Loy. Cross-scale internal graph neural network for image super-resolution. *Advances in neural information processing systems*, 33:3499–3509, 2020.