MSF²DN: Mutil Scale Feature Fusion Dehazing Network with Dense Connection

Guangfa Wang[1], Xiaokang Yu[✉]

College of Computer Science and Technology, Qingdao University, China
{2020020635, xyu}@qdu.edu.cn

Abstract. Single image dehazing is a challenging problem in computer vision. Previous work has mostly focused on designing new encoder and decoder in common network architectures, while neglecting the connection between the two. In this paper, we propose a multi scale feature fusion dehazing network based on dense connection, MSF²DN. The design principle of this network is to make full use of dense connection to achieve efficient reuse of features. On the one hand, we use a dense connection inside the base module of the encoder-decoder to fuse the features of different convolutional layers several times, and on the other hand, we design a simple multi-stream feature fusion module which fuses the features of different stages after uniform scaling and feeds them into the base module of the decoder for enhancement. Numerous experiments have demonstrated that our network outperforms the existing state-of-the-art networks.

Keywords: Image dehazing · Feature fusion · Dense connection.

1 Introduction

In computer vision and computer graphics, high-quality images are the basis for advanced vision tasks, so obtaining clear image from hazy image has received a great deal of attention in the last two decades. Fig.1 shows an example of dehazing.

To describe the haze, an atmospheric scattering model [29, 30] was proposed, which has the following equation:

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

where I is the hazy image received by the camera, J denotes a clear scene radiance, A represents the global atmospheric light, which is usually a constant, and t is a transmission map, which can be expressed as $t = e^{-\beta d(x)}$, d is the distance from the camera to the scene radiation, β is the atmospheric scattering coefficient, x describes the pixel position.

It is clear that this is an ill-posed problem, and in order to get a haze-free image, many statistical priors [5, 18, 19, 28, 37, 44] appear. These priors can solve this problem to some extent, but unfortunately, the robustness of these hand-designed priors is not strong and can't work in complex scenarios.

To avoid the limitations of prior methods, data-driven deep learning has been applied to dehazing with quite good result. Early researchers used convolutional neural networks(CNNs) [6, 34, 40, 41] to estimate the exact transmission map and atmospheric value, however, obtaining the transmission map and atmospheric value from a single hazy image is not an easy task, and the atmospheric scattering model itself does not fully describe the cause of haze, so such a strategy has been gradually abandoned in recent years. Later researchers preferred to build a trainable end-to-end CNNs [7, 10, 11, 16, 24–27, 33, 39, 43] to obtain haze-free image directly from hazy image, and such a strategy achieved better result. In continuous development, we found that simply stacking convolutional layers does not solve the dehazing problem well, so past work tends to develop a new network structure, such as Dilated Network [7], U-Net [10], Grid Network [25], and CycleGAN [35], or build a new enhancement module [10, 31] to solve this problem, these algorithms do solve this problem well, but they all almost ignore the full use of features. MSBDN [10] is aware of this problem and uses the dense feature fusion module(DFF) to exploit features, but it is limited to the exploitation of features inside the encoder and decoder, ignoring the connection of features between the two.

In this work, we propose a multi scale feature fusion dehazing network based on dense connection, MSF^{2}DN. Our core principle is to make full use of the dense connection to achieve full utilization of the features in the U-shaped network architecture. In the base module of the encoder, we design a generic feature extraction module, MLFF, which incorporates shallow features into deep features to enhance feature delivery as well as reuse. In the base module of the decoder, in order to be able to apply dense connection to get fully enhanced features, we extend the base module of FFANet [31] and design a new enhancement module, DFFD. In each base module of the encoder and decoder, we add a pixel attention module and channel attention module [31] to enhance the performance of CNNs. At the network architecture level, the features at different scales are upsampled or downsampled to unify to the same scale, feeding into the multi stream feature fusion module(MSFF) we designed for simple processing and then feeding into the decoder base module for enhancement. We embed residual groups in the network to allow the model to learn to bypass unimportant information and focus the main performance on the dense haze region. Extensive experiments have shown that the performance of our method outperforms the state-of-the-art network. The contributions of this paper are summarized as follows:

– A base module, MLFF, based on dense connection and attention mechanisms, is designed for feature passing and reuse, ensuring a diversity of deep features.
– An effective feature enhancement module, DFFD, is extended to enable the network to adaptively assign more weights to important features using the channel attention module and the pixel attention module.
– A multi scale feature fusion network based on dense connection is designed to realize the effective utilization of features in the existing network architecture, and excellent experimental results are obtained.
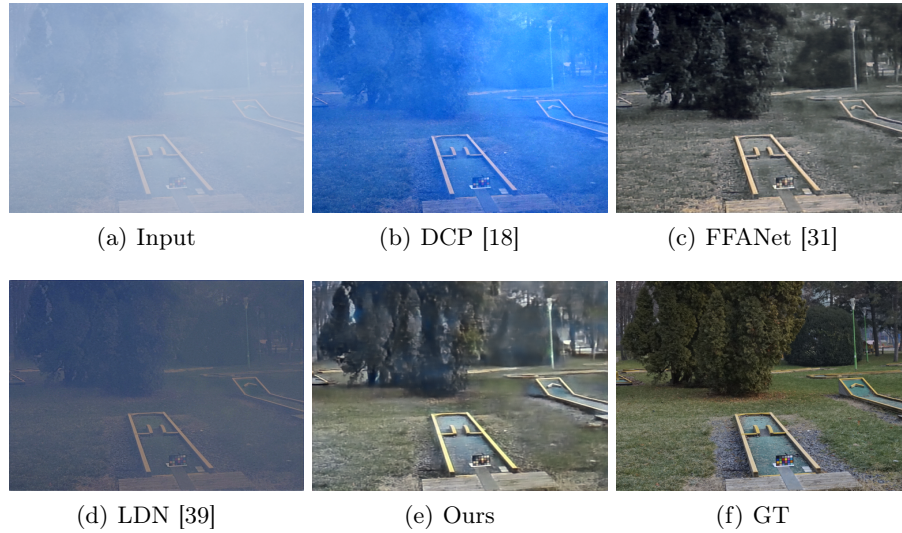
(a) Input (b) DCP [18] (c) FFANet [31]

(d) LDN [39] (e) Ours (f) GT

Fig. 1. Haze removal.

2 Related Work

2.1 Single image dehazing

Image dehazing is an important research topic in computer vision, aiming to obtain a haze-free image from a hazy image. Current dehazing methods can be divided into two categories: prior-based image dehazing [5, 18, 19, 28, 37, 44] and learning-based image dehazing [6, 7, 10, 11, 16, 24–27, 33, 34, 39–41, 43].

Prior-based Image Dehazing A priori-based dehazing methods rely on the atmospheric scattering model [29, 30], which is an ill-posed problem, so researchers have solved this problem by mathematical statistics to obtain common features of hazy images or clear images, forming a prior or hypotheses. Tan [38] gets a clear image by maximizing local contrast, He *et al.* [18] design the dark channel prior by counting the clear image features, and obtaining the clear image based on this prior. Berman *et al.* [5] distribute the image pixels in the RGB color space and obtain the clustering relationship between clear image pixels and hazy image pixels to develop a non-local dehazing algorithm. Kim *et al.* [19] further develops dark channel prior [18] and use different functions to enhance image saturation to obtain clear images. Unfortunately, all these methods can only be used for a certain class of scenes, and they are all less robust in complex scenes, for example, dark channel prior [18] is ineffective in the sky.

Learning-based Image Dehazing With the rise of CNNs in the high-level vision task, such as object detection, recognition and related tasks [12, 13, 32],

CNNs are also subsequently used in image dehazing. Early CNN-based dehazing networks [6, 34, 40, 41] still adopt a multi-stage strategy, using CNNs to accurately estimate the transmission map and atmospheric value, and then substitute into an atmospheric scattering model to obtain the scene radiation. However, if the transmission map is not properly estimated, high-quality dehazing results will not be obtained. Therefore, the mainstream dehazing networks nowadays are end-to-end networks [7, 10, 11, 16, 24–27, 33, 39, 43] that do not require the estimation of transmission maps and atmospheric value, these networks are directly from the hazy image to get a clear image. Such strategy often yields better results in synthetic atlases, but due to the lack of a theoretical basis, they can still be applied to reality with unnatural results, therefore, RefineNet [43] and PSDNet [8], which combine the prior and CNNs have emerged, but such methods still have much room for improvement as of now.

2.2   Dense connection

As the network deepens, the network suffers from gradient vanishing/explosion problems as well as degradation problems. To solve this problem, many works, such as ResNet [15], Highway Network [36], FractalNets [21], have been proposed. All these works reveal the principle that creating a short path from lower to higher levels is beneficial to solving the degradation problem. Inspired by the above work, DenseNet [17] was proposed, which connects the network layers directly to ensure maximum information flow between the network layers. For each layer, the feature maps of all previous layers are its input, and its output feature maps are the input of all subsequent layers. Following such simple connection rules, DenseNet [17] naturally integrates the properties of identity mappings, deep supervision, and diversified depth, and it can be a good feature extractor.

2.3   Multi scale feature fusion

Many CNNs used to solve vision tasks often use downsampling operations to reduce the number of parameters, but bring the problem of feature loss. In order to allow the network to get enough features, feature fusion(feature concatenation [42], dense connection [16, 17]) is often adopted in the network design process. However, it is not enough to simply fuse the features to extract valid information. To exploit features between adjacent layers, Liu *et al.* [25] design grid-Net, but the portability of such a design is too weak. Dong *et al.* [10] design the DFF, which effectively fuses features from multiple scales, but this work only pays attention to the connection within the encoder and decoder.

3   Proposed Method

We propose a multi-scale feature fusion dehazing network with dense connection, MSF^{2}DN, which is extended from a U-shaped network and is end-to-end

trainable. There are three major components: the multi-level feature fusion module(MLFF), the double feature fusion dehazing module(DFFD), and the multi-stream feature fusion module(MSFF). The specific network structure is shown in Fig.2.
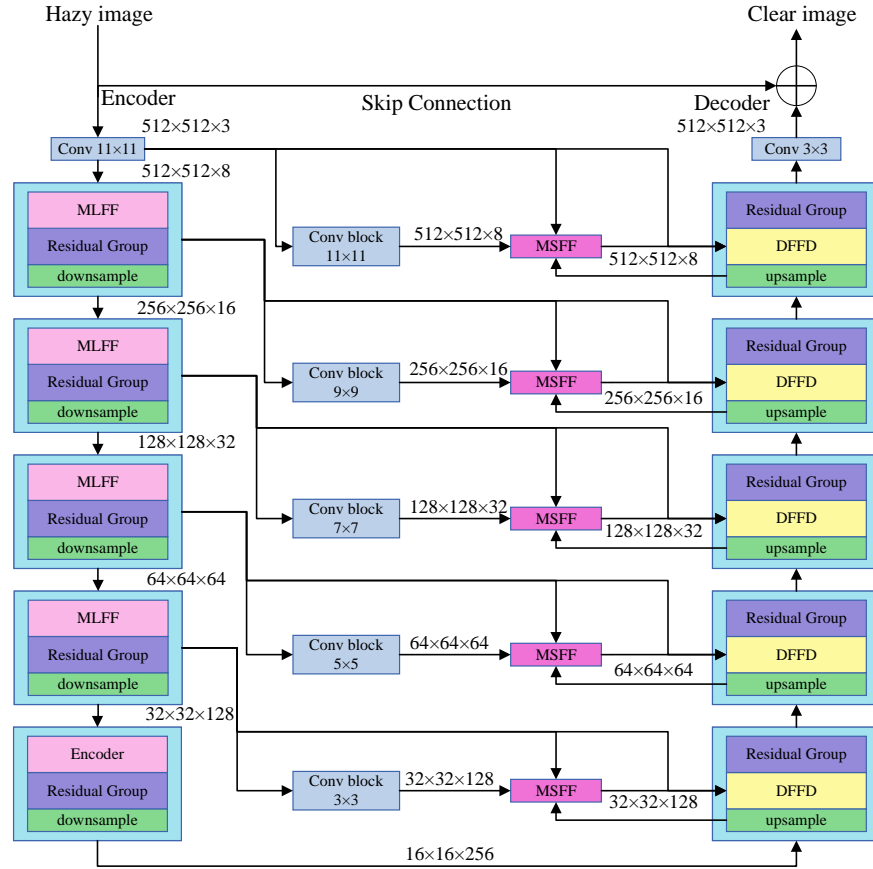

Fig. 2. Overall architecture of MSF^{2}DN.

3.1 Multi-level Feature Fusion Module

Previous work has usually used dense connections between base modules, but this approach still results in feature loss. In order to reduce the feature loss, we apply the dense connection inside the base module and thus design a multi-level feature fusion module(MLFF). As shown in Fig.3, it consists of 4 convolutional layers, 2 attention modules, 1 activation function, and dense connections. Dense connections allow the network to skip unimportant information, such as thin

haze or even clear areas, allowing most of the network's weights to be focused on the hard-to-solve dence haze areas. Our encoder consists of five MLFF modules, each stage of which downsamples features with a convolutional layer with a stride of 2.

Experiments have proven that adding dense connections within the base module of the encoder can indeed improve the performance of the network.
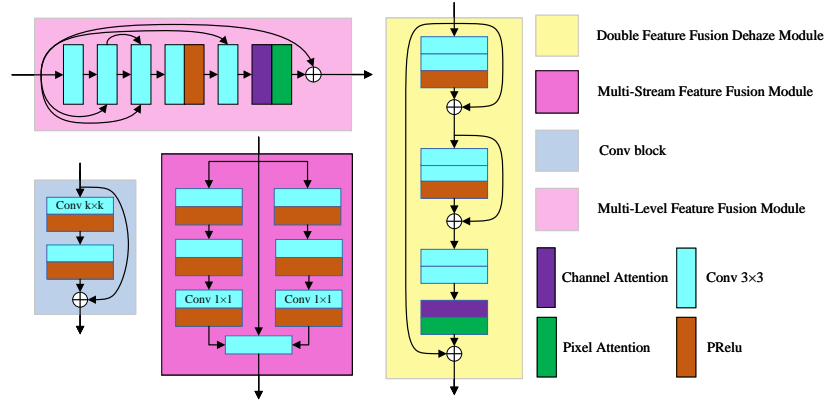

Fig. 3. Schematic diagram of the important module of MSF^{2}DN.

3.2   Double Feature Fusion Dehazing Module

FFANet's [31] base module has relatively powerful enhancements to features, but FFANet [31] does not make good use of the module, stacking too many base blocks. We extend a double feature fusion dehazing module(DFFD) in order to allow the combination of the base block and the dense connection, as shown in Fig.3, which consists of 6 convolutional layers, 2 attention modules, 1 activation function, and the dense connection, which allows the network to focus on the important information while augmenting the features.

The decoder consists of five DFFD modules, each stage of which upsamples features with a deconvolutional layer with a stride of 2.

3.3   Multi-stream Feature Fusion Module

There is a very large amount of work in the field of image dehazing that involves skipping connection, but it is relatively simple to send features from the encoder base module to the corresponding decoder base module, ignoring the connection of features between the two stages. Referring to the connection of NBNet [9] architecture, we redesigned a simple multi-stream feature fusion module to process the features after upsampling and downsampling, aiming to obtain the features

under different sense fields and fuse the features of the encoder base module and decoder base module. Our aim is to fuse the features of both phases and let the features obtained by the encoder base module guide the same-size decoder base module to reconstruct the image structure information. The specific detail is shown in Fig.3.

3.4 Implementations Details

As shown in Fig.2, in the encoder, each base block contains MLFF, Residual group, and downsample, and in the decoder, each base block contains Residual group, DFFD, and upsample, where the residual group contains 4 residual blocks [15] to deepen the network. The activation function is PRelu [14] and not every convolution function is connected after the activation function, as configured in Fig.3. The convolution radius of the first convolutional layer of the network is set to 11. In addition, different sizes of convolutional kernels, 11, 9, 7, 5, and 3, are set at each stage of the network according to the feature map size of each layer in order to expect that a larger perceptual field can alleviate the localization of convolution, and the radius of the convolutional kernels of all other convolutional layers is 3.

We choose Mean Square Error(MSE) as the loss function, written as:

$$\mathcal{L}(G, x, y) = ||x - G(y)||_2 \tag{2}$$

where $G(\cdot)$, x, and y represent MSF^2DN, clear image, and hazy image, respectively.

The overall training process includes 100 epochs, and the batch size is 16. the optimizer is ADAM [20], β_1 is 0.9, β_2 is 0.999. The initial learning rate is set to 10^{-4}, and the learning rate is multiplied by 0.1 for every 25 epochs. All the experiments are conducted on one NVIDIA V100 GPU. **Our code and trained model is available at** `https://github.com/Bruce-WangGF/MSFFDN`.

4 Experimental Results

4.1 Datasets

We compare our method with state-of-the-art(SOTA) methods in two categories: synthetic [23] and real-world datasets [1–4].

Sythetic Datasets. RESIDE [23] was proposed by Li, B. *et al.* There are five subsets, where Indoor Training Set(ITS), Outdoor Training Set(OTS), Synthetic Objective Testing Set(SOTS) are synthetic sets, Real World task-driven Testing Set(RTTS) is the real-world dataset and Hybrid Subjective Testing Set(HSTS) includes both real-world and synthetic set. The datasets include a variety of indoor and outdoor scenes under daylight. In order to maintain the generalization ability of the model, we selected 22,000 images from ITS and OTS, of which 15,000 were randomly selected from ITS and 7,000 were randomly selected from OTS. The SOTS includes 500 indoor images and 500 outdoor hazy images and their corresponding clear images, which we choose as the testing set.

Real-world Datasets To demonstrate the superiority of our method in the real world, we chose four real-world datasets: Dense HAZE [1], NH-HAZE [2], O-HAZE [3], I-HAZE [4]. They are from the NTIRE image dehazing challenge. Due to the small number of images, for each dataset, we choose 5 images as a validation set, 5 images as a testing set, and the rest of the images are expanded to 10,000 by cropping.

(a) Input   (b) DCP [18]   (c) ADN [22]   (d) GDN [25]   (e) FFANet [31]

(f) MSBDN [10]   (g) LDN [39]   (h) RDN [43]   (i) Ours   (j) GT
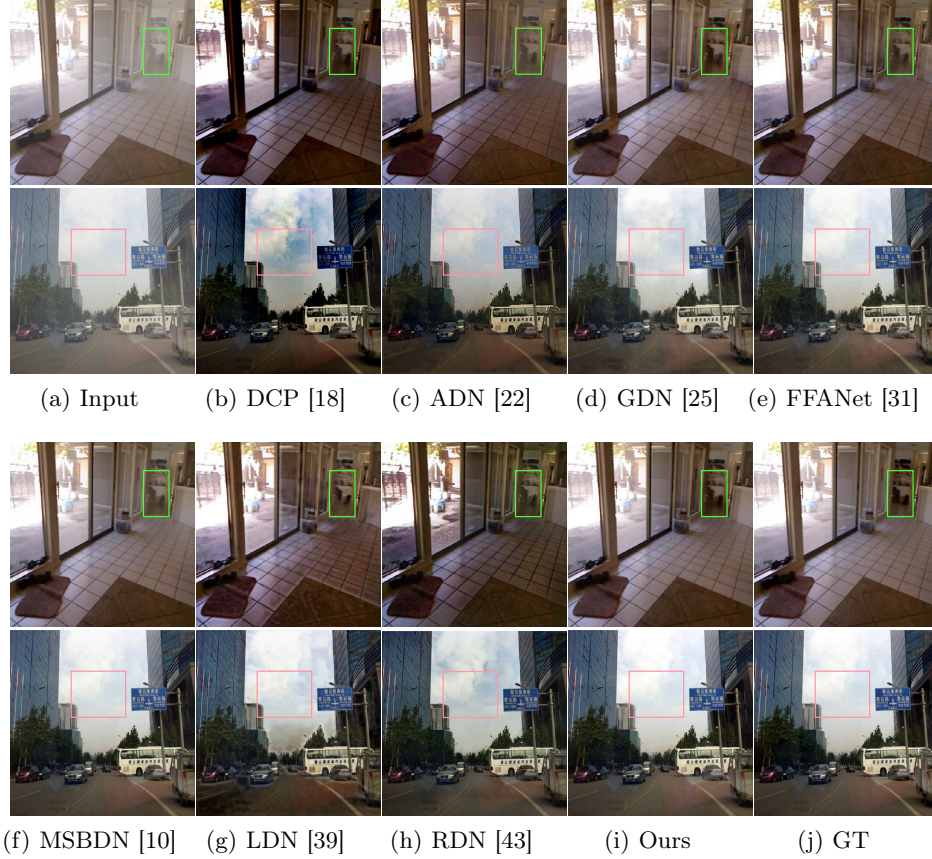
Fig. 4. Visual results comparison on SOTS [23] dataset. The green and pink window indicate the visible difference between each method.

4.2   Evaluation metrics and comparisons

In the field of image dehazing, Peak Signal to Noise Ratio(PSNR) and Structural Similarity index(SSIM) are often used as image quality evaluation metrics, we also adopt these two metrics.

Table 1. Quantitative evaluations on the benchmark dehazing datasets. Red texts and blue texts indicate the best and the second-best performance respectively.

	Matrix	DCP[18]	AodNet[22]	GDN[25]	FFANet[31]	MSBDN[10]	LDN[39]	RDN[43]	Ours
SOTS	PSNR	18.58	20.10	24.46	24.39	27.08	21.27	24.12	27.17
	SSIM	0.818	0.828	0.886	0.879	0.915	0.832	0.933	0.915
I-HAZE	PSNR	13.66	19.66	18.83	19.65	22.05	19.76	13.96	24.52
	SSIM	0.688	0.855	0.826	0.872	0.869	0.860	0.741	0.909
O-HAZE	PSNR	17.54	20.00	23.86	24.31	24.15	20.53	16.92	25.19
	SSIM	0.761	0.798	0.804	0.809	0.822	0.794	0.740	0.858
NH	PSNR	14.91	16.50	19.55	20.42	20.95	16.93	12.37	22.04
	SSIM	0.674	0.633	0.767	0.794	0.796	0.656	0.539	0.803
Dence	PSNR	14.15	15.50	15.82	18.46	18.27	15.67	12.15	18.37
	SSIM	0.552	0.498	0.576	0.629	0.603	0.506	0.426	0.622

4.3   Performance Analysis

We will analyze the advantages and disadvantages of our algorithm with other SOTA algorithms [10, 18, 22, 25, 31, 39, 43] on a synthetic dataset and four real-world datasets in both quantitative and qualitative aspects.

Analysis on synthetic datasets As shown in Table 1, under the same training environment and training method, our method obtained the highest PSNR and the second-highest SSIM compared to other SOTA methods. Compared to MSBDN [10], our method has a 0.09 dB improvement in PSNR. Note that we downloaded two versions of SOTS from Github, our model and MSBDN [10] reach **34.21**dB PSNR and **34.00**dB PSNR respectively in the other version. We also give the qualitative analysis of our method and other SOTA methods, as shown in Fig.4. We can observe that the DCP [18] performs relatively well, but the overall image is dark, and the RDN [43] with fused DCP [18] and CNNs has the same problem. The ADN [22], LDN [39], and MSBDN [10] all have the problem of incomplete dehazing at the green window. The GDN [25] has better results compared to the above methods, but the haze still appears in the lower right corner of the green window. Our method has a natural style for this image, and it is similar to the real image in both high and low-frequency regions. In the pink box, our model and FFANet [31] have the most natural result in the sky.

Analysis on real-world datasets The advantage of our method over its performance on synthetic datasets is mainly in real-world datasets, where we divide four real datasets into two categories: thin datasets(O-HAZE [3], I-HAZE [4]) and thick datasets(NH-HAZE [2], Dence HAZE [1]). As shown in Table 1, our method obtains the best performance on O-HAZE [3] and I-HAZE [4].

We also present the dehazing results for both datasets in Fig.5. We can observe that for the I-HAZE [4] example, the results of DCP [18], and RDN [43] are very dark and lose a lot of details. The performance of ADN [22], and LDN [39] is poor and the images appear gray. The results of GDN [25] have incongruent image tones and appear noisy. FFANet [31] has a halo effect at the green window. MSBDN [10] has a brighter image compared to GT as a whole, and the color saturation of the color palette is low as seen in the blue window.

Our method is also brighter compared to GT, but better than MSBDN [10], and the overall style is more natural and similar to GT.
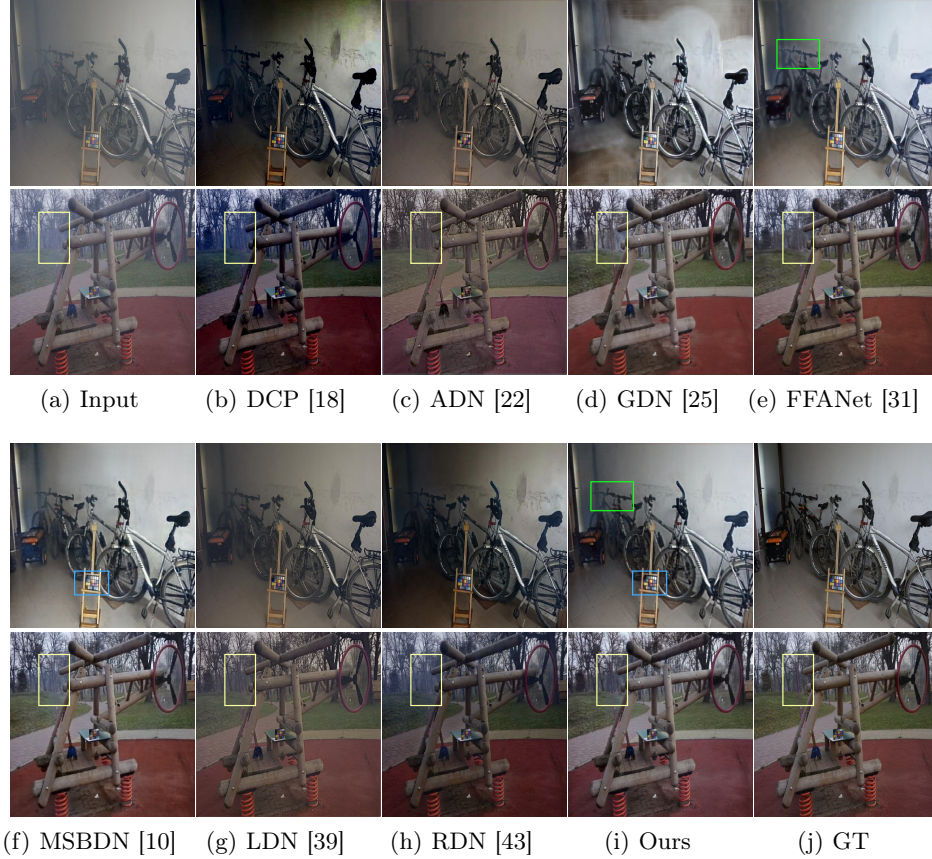

(a) Input   (b) DCP [18]   (c) ADN [22]   (d) GDN [25]   (e) FFANet [31]

(f) MSBDN [10]   (g) LDN [39]   (h) RDN [43]   (i) Ours   (j) GT

Fig. 5. Haze removal in I-HAZE [4] and O-HAZE [3].

In the case image of O-HAZE [3], all previous methods also present roughly the same problems as I-HAZE [4], and we give the main differences in the yellow window. As you can see, except for ADN [22], FFANet [31], LDN [39], and our method, none of the other methods can get the haze inside the window nowhere. FFANet [31] is also not as good as our method for the scenery at the yellow window.

How to dehaze dense hazy images in the real world has been a very tricky problem, and existing work does not yield a better performance due to the very small number of samples. Our method makes further progress on the dense hazy dataset. As shown in Table 1, our method obtains the best performance on NH-HAZE [2] with its PSNR and SSIM of 22.04 dB, 0.803, and the second-

best MSBDN [10] with its PSNR and SSIM of 20.95dB and 0.796, respectively. On the Dence HAZE [1] dataset, both of our methods obtain the second-best performance with PSNR and SSIM lagging behind FFANet [31] by 0.09dB and 0.007, respectively.
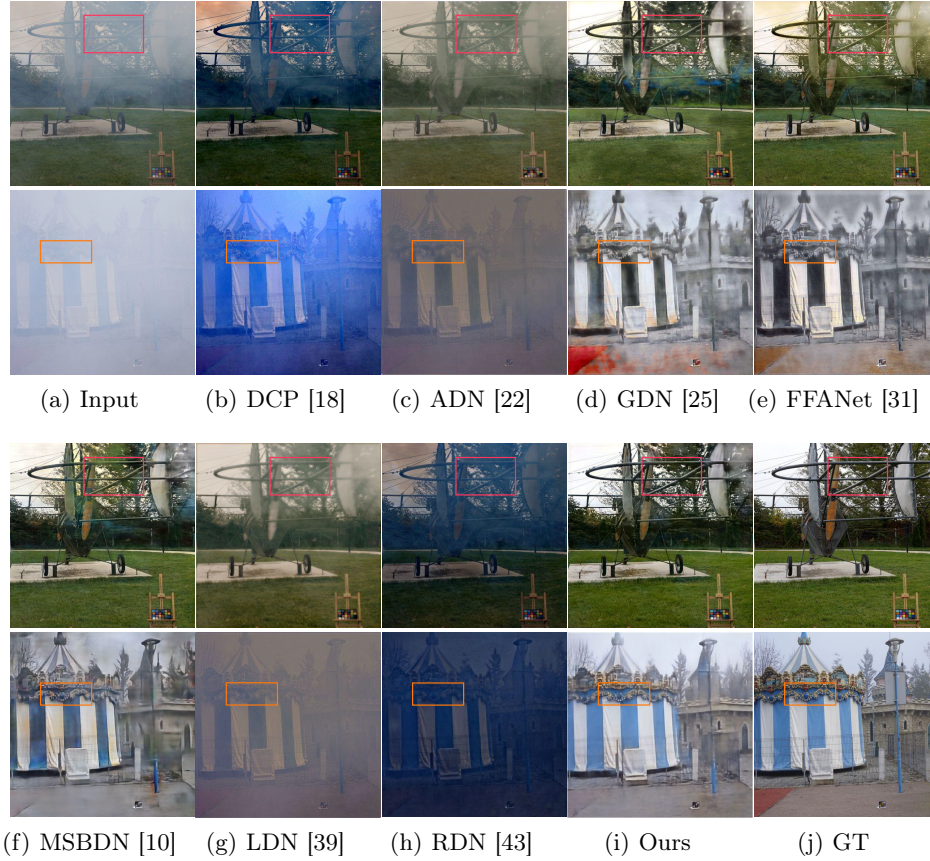

(a) Input   (b) DCP [18]   (c) ADN [22]   (d) GDN [25]   (e) FFANet [31]

(f) MSBDN [10]   (g) LDN [39]   (h) RDN [43]   (i) Ours   (j) GT

Fig. 6. Haze removal in NH-HAZE [2] and Dence HAZE [1].

Fig.6 shows the results of SOTA methods and our method in NH-HAZE [2] and Dence HAZE [1], and it can be seen that both the DCP [18] and the RDN [43] results combining DCP [18] and CNNs are very poor and show severe color bias. All methods show different degrees of dehazing incompleteness, especially ADN [22], LDN [39] is the most serious. We can observe that although FFANet [31] obtains a high matrix performance, it has a very poor visual effect. For the scenes in the red and orange window, the visual effect of the result obtained by our method completely surpasses that of the previous SOTA methods.

Although the four real-world datasets can simulate haze formation to a large extent, they are ultimately not obtained in a naturally occurring haze environment and therefore do not fully represent the formation of haze in the real world. We have taken several images of naturally occurring haze in the real world to compare our method with the SOTA methods. As you can see, our method still works, especially in the first image, where our method clearly surpasses the other SOTA methods in terms of sensory processing of the sky.
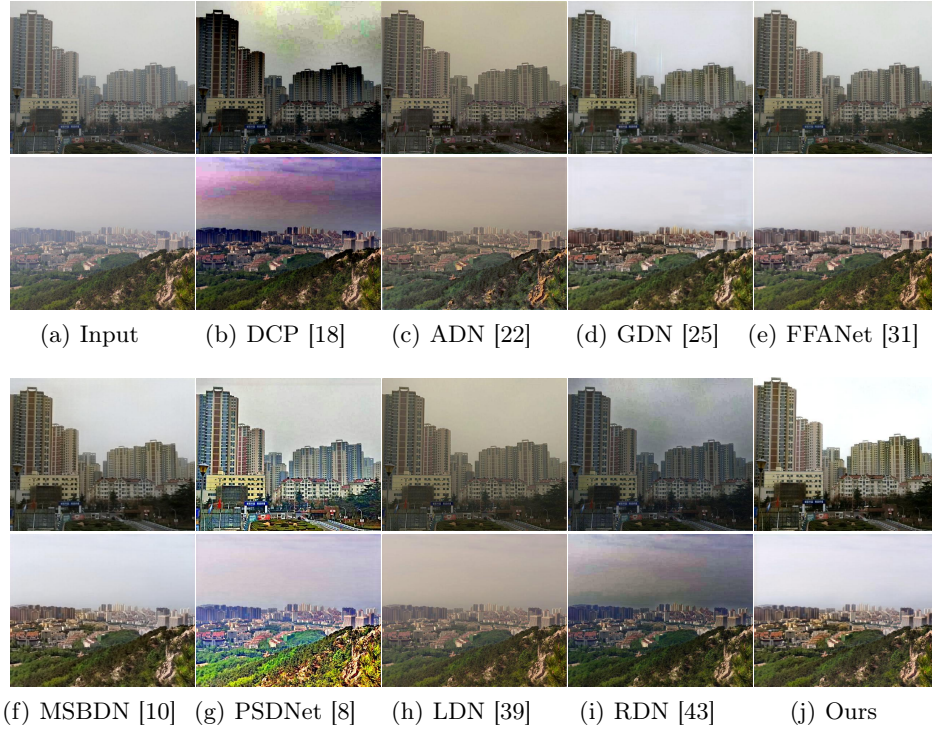

(a) Input   (b) DCP [18]   (c) ADN [22]   (d) GDN [25]   (e) FFANet [31]

(f) MSBDN [10]   (g) PSDNet [8]   (h) LDN [39]   (i) RDN [43]   (j) Ours

Fig. 7. Haze removal in real-world.

4.4   Model complexity and inference time

We give the number of parameters of the model, the inference time for different image sizes, and the FLOPs of the model to measure the complexity of the model (input image size 512×512×3) in Table 2, using an RTX 3090 GPU. × in the table indicates that the model requires more than 24GB of video memory.

Compared to the suboptimal MSBDN [10], the number of parameters and FLOPs of our network are about $\frac{1}{4}$ and $\frac{1}{3}$ respectively, and the inference time is reduced accordingly. It can be seen that for 2048×2048 resolution image, our

model can still inference normally, which indicates that our network uses less memory than MSBDN [10], FFANet [31], and GDN [25].

Table 2. Comparison of the number of parameters, FLOPs, and inference time.

	ADN[22]	GDN[25]	FFANet[31]	MSBDN[10]	LDN[39]	RDN[43]	Ours
Param(MB)	0.002	0.960	4.680	31.35	0.030	68.56	7.670
FLOPs(GB)	0.460	0.45	1.46	97.77	7.890	238.1	31.46
128×128(s)	0.004	0.073	0.267	0.304	0.006	−	0.110
256×256(s)	0.006	0.076	0.307	0.313	0.008	0.206	0.118
512×512(s)	0.014	0.100	0.486	0.333	0.019	0.216	0.140
1024×1024(s)	0.042	0.216	×	0.422	0.056	0.420	0.245
2048×2048(s)	0.204	×	×	×	0.245	1.236	0.736

4.5   Ablation Study

To demonstrate the effectiveness of our method, we perform ablation experiments to analyze each module, including MLFF, DFFD, MSFF, residual group, and dense connection.

As shown in Table 3, ✓means that the module has been used, × means that the module has been removed directly, and − means that a simple module has been used instead of the module. In this section, we use the Conv block replacing MLFF and the FFA module [31] replacing DFFD.

The training configuration of the above network is consistent with section 3.4, and the training and testing sets used for the ablation experiments are I-HAZE [4], the performance of the above model is shown in Table 3.

Effect of MLFF MLFF is a general-purpose feature extractor that enables feature reuse and forwards pass without adding additional parameters, compared to simple stacked convolutional layers, using only dense connections. As can be seen in Table 3, our proposed MLFF improves the PSNR and SSIM from 23.92dB, 0.904 to 24.52dB, 0.909.

Effect of DFFD The FFA module [31] has already achieved greater success in dehazing task, and we inherited this idea and further enhanced its capability by proposing the DFFD to make it a more effective dehazing module. As can be seen from Table 3, we have obtained an improvement of 1.13dB PSNR and 0.008 SSIM using the DFFD to replace the FFA module [31], which is sufficient proof that our proposed DFFD module is effective.

Effect of MSFF As opposed to simply feeding features to the decoder, we propose a simple multi-stream processing module to fully fuse features from different sources and let the network select the important features, and our strategy is experimentally proven to be effective.

Effect of residual group Previous experience tells us that sufficiently deep networks generally give better performance, so we add the residual group. Experiments have proven this to be effective. As shown in Table 3, the network obtains a PSNR, SSIM improvement of 1.03dB and 0.045.

Effect of dense connection As shown in Table 3, the PSNR and SSIM of the model without dense connection is only 21.31 dB and 0.853 respectively. After adding dense connections to each base module, PSNR and SSIM reach the best, which shows that dense connections do allow the network features to be passed on effectively and can make the encoder a good feature extractor.

Table 3. Ablation experiments on the effectiveness of each module.

MLFF	MSFF	DFFD	Dense Connection	Residual group	PSNR	SSIM
−	✓	✓	✓	✓	23.92	0.904
✓	×	✓	✓	✓	23.86	0.905
✓	✓	−	✓	✓	23.39	0.901
✓	✓	✓	×	✓	21.31	0.853
✓	✓	✓	✓	×	23.49	0.864
✓	✓	✓	✓	✓	24.52	0.909

5   Conclusion

In this paper, we propose a novel MSF^{2}DN for single image dehazing, which consists of a multi level feature fusion module(MLFF), a double feature fusion dehazing module(DFFD), extended from an FFA block using a dense connection and attention mechanism, and a multi stream feature fusion module to deal with the features from different stages. Dense connections allow the network to skip relatively unimportant features such as thin haze or even clear areas, focusing the main performance of the network on dense haze areas. The ablation experiments prove that our proposed module is helpful for the improvement of network performance. We evaluate MSF^{2}DN on a synthetic dataset and real-world datasets and demonstrate that MSF^{2}DN outperforms existing SOTA methods.

References

1. Ancuti, C.O., Ancuti, C., Sbert, M., Timofte, R.: Dense haze: A benchmark for image dehazing with dense-haze and haze-free images. In: arXiv (2019)
2. Ancuti, C.O., Ancuti, C., Timofte, R.: Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2020)
3. Ancuti, C.O., Ancuti, C., Timofte, R., Vleeschouwer, C.D.: O-haze: A dehazing benchmark with real hazy and haze-free outdoor images. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2018)
4. Ancuti, C.O., Ancuti, C., Timofte, R., Vleeschouwer, C.D.: I-haze: a dehazing benchmark with real hazy and haze-free indoor images (2018)
5. Berman, D., Treibitz, T., Avidan, S.: Non-local image dehazing. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
6. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: An end-to-end system for single image haze removal. IEEE Transactions on Image Processing **25**(11), 5187–5198 (2016)
7. Chen, D., He, M., Fan, Q., Liao, J., Zhang, L., Hou, D., Yuan, L., Hua, G.: Gated context aggregation network for image dehazing and deraining. In: 2019 IEEE winter conference on applications of computer vision (WACV). pp. 1375–1383. IEEE (2019)
8. Chen, Z., Wang, Y., Yang, Y., Liu, D.: Psd: Principled synthetic-to-real dehazing guided by physical priors. In: Computer Vision and Pattern Recognition (2021)
9. Cheng, S., Wang, Y., Huang, H., Liu, D., Liu, S.: Nbnet: Noise basis learning for image denoising with subspace projection (2020)
10. Dong, H., Pan, J., Xiang, L., Hu, Z., Zhang, X., Wang, F., Yang, M.H.: Multi-scale boosted dehazing network with dense feature fusion. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2157–2167 (2020)
11. Engin, D., Gen, A., Ekenel, H.K.: Cycle-dehaze: Enhanced cyclegan for single image dehazing. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2018)
12. Girshick, R.: Fast r-cnn. Computer Science (2015)
13. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. IEEE Transactions on Pattern Analysis & Machine Intelligence (2017)
14. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: CVPR (2015)
15. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
16. He, Z., Sindagi, V., Patel, V.M.: Multi-scale single image dehazing using perceptual pyramid deep network. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2018)
17. Huang, G., Liu, Z., Laurens, V., Weinberger, K.Q.: Densely connected convolutional networks. In: IEEE Computer Society (2016)
18. Kaiming, He, Jian, Sun, Xiaoou, Tang: Single image haze removal using dark channel prior. IEEE Transactions on Pattern Analysis & Machine Intelligence (2011)
19. Kim, S.E., Park, T.H., Eom, I.K.: Fast single image dehazing using saturation based transmission map estimation. IEEE Transactions on Image Processing **29**, 1985–1998 (2019)

20. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. Computer Science (2014)
21. Krizhevsky, A., Sutskever, I., Hinton, G.: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems **25**(2) (2012)
22. Li, B., Peng, X., Wang, Z., Xu, J., Dan, F.: Aod-net: All-in-one dehazing network. In: 2017 IEEE International Conference on Computer Vision (ICCV) (2017)
23. Li, B., Ren, W., Fu, D., Tao, D., Wang, Z.: Reside: A benchmark for single image dehazing (2017)
24. Li, R., Pan, J., Li, Z., Tang, J.: Single image dehazing via conditional generative adversarial network. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
25. Liu, X., Ma, Y., Shi, Z., Chen, J.: Griddehazenet: Attention-based multi-scale network for image dehazing (2019)
26. Liu, X., Suganuma, M., Sun, Z., Okatani, T.: Dual residual networks leveraging the potential of paired operations for image restoration. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
27. Mei, K., Jiang, A., Li, J., Wang, M.: Progressive feature fusion network for realistic image dehazing (2018)
28. Meng, G., Wang, Y., Duan, J., Xiang, S., Pan, C.: Efficient image dehazing with boundary constraint and contextual regularization. In: Proceedings of the 2013 IEEE International Conference on Computer Vision (2013)
29. Narasimhan, S.G., Nayar, S.K.: Chromatic framework for vision in bad weather. In: IEEE Computer Society Conference on Computer Vision & Pattern Recognition (2000)
30. Narasimhan, S.G., Nayar, S.K.: Vision and the atmosphere. International Journal of Computer Vision **48**(3), 233–254 (2002)
31. Qin, X., Wang, Z., Bai, Y., Xie, X., Jia, H.: Ffa-net: Feature fusion attention network for single image dehazing. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 11908–11915 (2020)
32. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis & Machine Intelligence **39**(6), 1137–1149 (2017)
33. Ren, W., Ma, L., Zhang, J., Pan, J., Cao, X., Liu, W., Yang, M.H.: Gated fusion network for single image dehazing. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (2018)
34. Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.H.: Single image dehazing via multi-scale convolutional neural networks. In: European conference on computer vision. pp. 154–169. Springer (2016)
35. Shao, Y., Li, L., Ren, W., Gao, C., Sang, N.: Domain adaptation for image dehazing. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2808–2817 (2020)
36. Srivastava, R.K., Greff, K., Schmidhuber, J.: Training very deep networks. Computer Science (2015)
37. Sulami, M., Glatzer, I., Fattal, R., Werman, M.: Automatic recovery of the atmospheric light in hazy images. In: IEEE International Conference on Computational Photography (2014)
38. Tan, R.T.: Visibility in bad weather from a single image. In: 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), 24-26 June 2008, Anchorage, Alaska, USA (2008)

39. Ullah, H., Muhammad, K., Irfan, M., Anwar, S., Sajjad, M., Imran, A.S., de Albuquerque, V.H.C.: Light-dehazenet: a novel lightweight cnn architecture for single image dehazing. IEEE Transactions on Image Processing **30**, 8968–8982 (2021)
40. Yang, X., Xu, Z., Luo, J.: Towards perceptual image dehazing by physics-based disentanglement and adversarial training. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 32 (2018)
41. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3194–3203 (2018)
42. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (2018)
43. Zhao, S., Zhang, L., Shen, Y., Zhou, Y.: Refinednet: A weakly supervised refinement framework for single image dehazing. IEEE Transactions on Image Processing **30**, 3391–3404 (2021)
44. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. IEEE Transactions on Image Processing **24**(11), 3522–3533 (2015)