

Image defogging based on Multi-input and Multi-scale UNet

Zhengchun Lin

Guangdong Polytechnic Normal University

Qingxing Luo (✉ 1085380367@qq.com)

Guangdong Polytechnic Normal University

Yunzhi Jiang

Guangdong Polytechnic Normal University

Jing Wang

Guangdong Polytechnic Normal University

Siyuan Li

Guangdong Polytechnic Normal University

Gongwen Cheng

Shenzhen Chace Network Information Technology Co., Ltd

genrang Zheng

Zhongshan Polytechnic

Research Article

Keywords: image dehazing, Convolutional neural network, Feature fusion, image restoration, image processing

Posted Date: April 13th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1510599/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The coarse-to-fine image defogging strategy has been widely used in the structural design of individual image defogging networks. In the traditional method, multi-scale input image subnets are superimposed, so that the sharpness of the image is gradually improved from the bottom subnet to the top subnet, which inevitably leads to the loss of image details. Toward a fast and accurate dehazing network design, we revisit the coarse-to-fine strategy and present a multi-input and multi-scale U-Net (MIMS-UNet). The MIMS-UNet has two distinct features. On the one hand, the single-encoder of MIMS-UNet adopts multi-input and multi-scale image, which increases the computation amount but greatly improves the network performance. On the other hand, codec structures with context blocks are used to capture context information and recover more details. The experimental results show that the proposed method achieves good results in both quantification and visualization. Compared with the existing methods, the proposed network can achieve ideal results of defogging and effectively avoid color distortion after defogging.

1. Introduction

Due to the absorption and scattering of light by airborne particles, foggy images acquired by imaging equipment have problems such as decreased contrast, color distortion and loss of details, and thus affect the application of images in subsequent tasks, such as object recognition and scene understanding. At the same time, it is not conducive to image feature extraction and recognition, and reduces the effectiveness of outdoor vision system. Therefore, image defogging has important research significance in the field of computer vision.

According to the physical scattering models [1, 2, 3], the hazing process is usually expressed as:

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

$$t(x) = e^{-\beta(\lambda) d(x)}, \quad (2)$$

where $I(x)$ and $J(x)$ are the observed hazy image and the haze-free scene radiance, A is global atmospheric light which represents the intensity of ambient light, and $t(x)$ indicates scene transmission describing the portion of light that is not scattered and reaches the camera sensors. $d(x)$ and $\beta(\lambda)$ represent scene depth and atmospheric scattering parameters, respectively.

However, it is often difficult to estimate the transmission image from foggy images. Early prior based methods attempted to estimate transmission images by using the statistical characteristics of clear images, such as non-local prior [4] and color attenuation prior [5]. However, these images have a large prior error, resulting in serious color distortion and contrast reduction in the restored images. At present, with the improvement of computer computing power, dehazing methods based on convolutional neural networks have become the main stream of research. These methods are effective and superior to prior-based algorithms, and have significant performance improvement. However, most of the current methods of defogging are to directly remove the observed image, ignoring the destruction of texture details in the

process of defogging, resulting in the phenomenon of noise amplification and color distortion after the image defogging. In addition, it usually minimizes the mean square error between foggy images and fog-free images, which is easy to lose high-frequency image details. In foggy conditions, this method may cause transition smoothness and produce artifacts for some regions with rich texture boundaries. In addition, the receptive field of traditional CNN is relatively small, and expanding the receptive field by deepening the network structure will lead to high resource consumption.

In this paper, MIMS-UNet is used to solve the above problems. We propose to use a network to process multi-scale images and then fuse multi-input information to compensate for the lost high-frequency image details. In view of the relatively small receptive field of traditional CNN, the context module is introduced, which not only does not increase the network structure, but also can further expand the receiving domain and capture multi-scale information. Compared with the most advanced methods, our method can achieve good performance while maintaining relatively small computational overhead.

The contributions of this work are summarized as follows:

We propose a new MIMS-UNet network for effective dehazing. The network can extract relevant features from fogged image content and recover details and textures from fogged images.

We propose to use encoder-decoder structure with context block to capture multi-scale information and defog from coarse to fine.

Through a lot of experiments, this model has better performance than other advanced algorithms.

2. Related Work

In general, there are three methods for image defogging: the method based on image enhancement algorithm based on signal processing principle, the method based on physical model [10] and the method of deep learning. Methods based on image enhancement [6, 7] do not consider the reason of image degradation and improve visual effect by enhancing image contrast, such as histogram equalization algorithm [8] and Retinex[9] algorithm. However, such methods do not take into account the root cause of image degradation, and images after defogging are prone to information loss. In addition, the application scenarios of this method are often limited.

The method based on the physical model is based on the atmospheric scattering model, and the transmission diagram is estimated and the atmospheric light pair model is solved according to different prior information, so as to realize the image defogging. Among these methods, the most widely studied and applied [11] is the defogging method based on dark channel prior, and the representative works in this regard include [38, 11, 40, 24]. Specifically, Tan [38] proposed a method of maximum contrast defogging, which observed that a clear image often had a higher contrast than the corresponding image with fog, and the image obtained after defogging would have color distortion. He et al. [9] used dark channel prior (DCP) to estimate transmittance in local areas. This method is based on the following

assumption: in at least one color channel, pixel values in haze-free patches are close to zero, but severe block effect exists in the transmittance obtained. Subsequent work improved the efficiency and performance of the DCP method [13, 14, 15, 16, 17]. Zhu et al. [12] proposed to recover the depth information of the image based on color attenuation prior, and then estimate the transmission image. Berman et al. [24] assumed that hundreds of different colors could well approximate the colors of clear images, and then carried out image defogging based on this prior. Although these methods have been shown to be effective for image defogging, their performance is inherently limited because the assumed priors are not suitable for all real images. Although significant progress has been made in physical model-based defogging methods, the simplified physical model cannot accurately estimate the transmission image and atmospheric light parameters, and the prior information is not universally applicable. For example, the processing ability of image sky region is poor, and the phenomenon of incomplete defogging and background information loss is easy to occur.

In recent years, methods based on deep learning have become the mainstream of research. This method uses convolutional neural network to restore clear images without considering the specific causes of image degradation. Cai et al. [41] proposed the end-to-end defogging model DehazeNet to estimate the transmission graph and construct the network according to the prior information in the traditional defogging method. Ren et al. [32] proposed an end-to-end Gated Fusion Network (GFN) based on the idea of image Fusion. The Network first estimated the corresponding weight graphs of three input images, and then weighted Fusion of the three input images guided by the weight graphs to obtain fog-free images. Ren et al. [26] proposed a coarse-to-fine strategy to estimate the initial transmittance. Zhang et al. [37] proposed a dense Connected Pyramid Dehazing Network (DCPDN) to estimate the medium transport map. Qu et al. [30] proposed an enhanced pix2pix de-fogging net to directly learn the mapping relationship between fog and nothing. Pairs of images are not required for training, and the recovered fog-free images are relatively clear and natural. Li et al. [20] proposed the parameters of the physical scattering model of atmospheric light by integrating the transmission diagram and the model, and designed AOD-Net to estimate the parameters based on CNN. The above direct end-to-end method can improve the image restoration quality to a certain extent, but the over-fitting phenomenon often occurs in the learning process and the original style characteristics of the image are easily ignored, resulting in incomplete or excessive defogging in some areas and color distortion in the restored image.

3. Proposed Method

This section describes the details of our MIMS-UNet framework. First, let's outline our approach. The architecture of MIMS-UNet is based on MSBDN[18]. In order to achieve efficient defogging, MSBDN[18] is greatly improved. Then the image MIMS module and image defogging module are introduced in detail. The overall architecture of the proposed MIMS-UNet is shown in Fig. 1.

3.1. multi-input multi-scale Dehazing Network

This paper designs a multi-input and multi-scale defogging network based on CNN, which is independent of atmospheric scattering model and can realize end-to-end defogging directly from foggy images to fog-free images. Firstly, the backbone of MINS-UNet is a UNet network. It has been proved that multi-scale images can better process fog of different degrees in images. A variety of CNN based de-fogging methods also adopt this idea and take foggy images of different scales as the input of each sub-network.

In our MIMS-UNet, each Dense Feature Fusion module takes a foggy image of different scales as input. That is to say, in addition to extracting the reduced features from the DFF above, we also extract the features from the reduced fog image, and then combine the two. By using complementary information of small size features and feature information obtained from small size images, our algorithm can effectively remove fog.

In order to make smaller foggy images (B2, B3, B4, B5) can be smoothly input into the network, they are encoded by SCM module and then fed into DFF module and multi-scale information fusion as the input of each layer. It can keep the original color of the image well while defogging, and overcome the problems of incomplete defogging and color distortion after defogging.

We first used shallow convolution module[21] (SCM) to extract features from the downsampling image, as shown in Fig. 2. We use the 3×3 and 1×1 convolution stacks. We connect the last 1×1 layer feature to the input B_K , and then An additional 1×1 convolution layer is used to further refine the characteristics of the join. The output of SCM at the K^{th} level is represented as SCM_k^{out} , and here we use SCM for the second and third and fourth levels as shown in Fig. 1.

3.2. Context block

Multiscale information is important for image dehazing tasks; therefore, the downsampling operation is usually adopted in networks. But, When the image resolution is too low, the image structures are destroyed, and information is lost. It is not conducive to reconstruction of features. In order not to further reduce the space resolution, and increase the receiving domain and embedded multi-scale features, we are introduced in the context of smallest scales between the encoder and decoder, which contains different parallel convolution, the inflation rate rather than the sample, so that we can without any increase in the number of arguments or destroy the structure of expanding receptive field. The background block[33] obtained good results in the image fragment [42] and the blur removal task [43]. The four expansion rates are 1, 2, 3, and 4. Features are extracted from different acceptance domains and the output is estimated by fusion (as shown in Fig. 3). This is useful for estimating offsets from larger acceptance domains. In our network, we first use 1×1 convolution to compress the feature channel to further reduce the running time and simplify the operation. In the fusion process, 1×1 convolution output channel is used to fuse with the original input feature. Meanwhile, in order to prevent information blocking, jump connection is adopted between the input and output. Context module incorporates rich hierarchical context information, which is beneficial to image defogging.

3.3. Implementations

The method in this paper is implemented based on Pytorch framework, and NVIDIA 3090 GPU is used to train defogging network in Ubuntu environment. The initial learning rate was set as 10^{-4} , batchsize was set as 8, Adam optimizer with momentum decay index $\beta_1=0.9$ and $\beta_2=0.999$ was used for optimization, and the number of iterations was 100. Use MSE as a loss function to limit network output and ground truth.

4. Experimental Results

In this section, we will introduce and analyze the experimental methods and results. First, the experiment in this paper is based on training and testing in the RESIDE dataset (OTS and ITS), and the results are compared with other algorithms.

4.1. Datasets

RESIDE dataset. The RESIDE dataset [19] contains both synthesized and real-world hazy/clean image pairs of indoor and outdoor scenes. The Japanese training set contains 1399 sharp images and 13990 misty images generated from the corresponding sharp images. For a fair comparison, we use the training set provided by MSBDN[18], In order to compare with the latest methods, PSNR and SSIM are used in this paper, and a comprehensive comparison test is carried out in the Comprehensive Objective Test Set (SOTS), which contains 500 indoor and 500 outdoor images. In this paper, comparative experiments are conducted with the current superior defogging methods, including the classical defogging method DCP [11] and the recent deep learning-based defogging method. The network in this paper still shows good defogging ability, and the method in this paper performs well in image details and color fidelity, and the results are shown in Table 1.

NTIRE2018-Dehazing challenge dataset. The Ntire2018-Dehazing Challenge dataset includes an outdoor dataset (O-HAZE[29]) and an indoor dataset(I-HAZE [25]). The indoor data set [25] includes 35 pairs of foggy and fog-free images of indoor family environment, and the outdoor data set[29] includes 45 pairs of foggy and fog-free images of various outdoor scenes. We cut the images of the two data sets into a size of 544×400 as a real fog test set to test the stability of our network.

4.2. Performance Evaluation

We evaluated our method based on manual prior (DCP[11] and NLD[24]) and deep convolutional neural network (AOD[20], MSCNN[26], MsPPN[27], DcGAN[28], GFN[32], GCANet[22], PFFNet[34], GDN[35], DuRN[36], MSBDN[18], FDU[44], DA_dahazing[31], GridDehazeNet[39], EPDN[30]).

Table 1
Quantitative comparison (Average PSNR/SSIM)
of the dehazing results on synthetic datasets.

Method	PSNR	SSIM
DCP[11]	18.75	0.859
NLD[24]	17.27	0.752
(ICCV'17)AODNet[20]	18.8	0.834
(CVPR'19)MSCNN[29]	17.57	0.811
MSPPN[30]	29.94	0.958
DcGAN[31]	25.37	0.917
(CVPR'18)GFN[32]	24.11	0.899
GCANet[33]	28.13	0.945
PFFNet[34]	29.22	0.954
(CVPR'19)GDN[35]	31.51	0.983
(CVPR'19)DuRN[36]	31.92	0.981
Method	PSNR	SSIM
(CVPR'19)EPDN[47]	23.82	0.891
(ICCV'19)GridDehazeNet[45]	32.16	0.984
(CVPR'20) DA_dahazing[46]	27.76	0.932
(CVPR'20)MSBDN[18]	33.79	0.984
(ECCV'20)FDU[44]	32.68	0.976
our	34.45	0.985

We also give a quantitative comparison of the results of defogging in Table 2. Compared with classical defogging algorithms such as DCP[11] and NLD[24], our results are far superior to these classical algorithms in quantitative (PSNR and SSIM). Meanwhile, compared with the most advanced fog removal algorithm [44, 46] in recent years, our results are still superior. As shown in Table 2, this method obtained the highest PSNR and SSIM values in SOTS. In addition, compared with MSBDN[18], The PSNR and SSIM of STOS data set increased by 0.66 dB and 0.001, respectively.

From Fig. 4, we can observe that DCP[11] algorithm has serious color distortion, and the result does not look close to reality. AOD[20] only removes a small amount of fog, and the overall tone of the restored image is white. We can see that GFN [32] also suffers from color distortion in some cases, and the results after defogging appear darker than our method. Although the direct estimation of sharp images based on

the end-to-end trainable network [22, 23] has better results than other indirect methods. we proposed method is superior to these algorithms in both PSNR and SSIM, but these networks do not optimize the defogging problem well, there is still the problem of incomplete defogging, and the defogging image is not clear enough. The algorithm based on MSBDN[18] is not clear enough in some details. Compared with these methods, the image restored by our algorithm has sharper structure and details and is closer to ground truth. Our algorithm produces better visual results.

After testing our effectiveness in the SOTS test set, we would RESIDE in the donatreside data collection and test our network on the O-haze [29] and I-haze [25] datasets in order to verify the stability and generalization of our network. The results are shown in Figs. 5 and 6. The test results were not as good as those in the SOTS test set, but were largely free of fog on the images. The comparison results of the above experiments can reflect the superiority of the proposed algorithm using multi-input and multi-scale input. This technical scheme makes the algorithm in this paper show good performance in both overall and local defogging effects.

4.3. Ablation Study and Analysis

In this section, in order to verify the effectiveness of the proposed module, we conducted a laugh ablation experiment(as shown in Fig. 7). In order to make a fair comparison, all the experimental methods mentioned below were trained using the same setup as the proposed algorithm.

Table 2
Ablation study of different components

Base	√	√	√
MIMS	×	√	√
Context block	×	×	√
PSNR	33.79	34.43	34.45
SSIM	0.984	0.984	0.985

Ablation on the MIMS. When the MIMS module was introduced into the basic network, the network performance was greatly improved. After the quality evaluation of the experiment was carried out on the SOTS test set, the PSNR result (as shown in Table 2) increased by 0.64dB. which fully proves the validity of the proposed module.

Ablation on the context block. The context block complements the down-sampling operation to capture larger field information. We can observe that performance is improved when context blocks are introduced, and the details that were lost after the image was defogged are nicely fixed.

We also carried out the visualization experiment of ablation experiment, as shown in Fig. 6. (c) represents the decontamination fog fruit diagram before adding our module. With the successive addition of MIMS and CB modules, the image after decontamination becomes clearer and the recovery of details is getting

better and better. In conclusion, ablation studies show that our proposed multi-input multi-scale module and context background module are useful for improving the model's defogging effect and recovering more detailed information.

Conclusion

Aiming at the problems of existing image defogging algorithms, such as dependence on prior information, loss of feature information and fuzzy details after defogging, we proposed an effective multi-input and multi-scale U-Net defogging network. The network is based on encoder-decoder architecture, and the study shows that the proposed module can effectively solve the problem of image defogging, while preserving the original color of the image well. The method proposed in this paper effectively solves the problem of incomplete defogging or color distortion caused by excessive defogging. Through comparative tests, the overall performance of the proposed algorithm is better than the existing algorithm, and the local performance of the algorithm is better.

Declarations

Conflict of interest statement. We declare that we have no conflict of interest

References

1. McCartney, E.J., & Hall, F.F. (1976). Optics of the Atmosphere: Scattering by Molecules and Particles. *Physics Today*, 30, 76-77. <https://doi.org/10.1063/1.3037551>
2. Narasimhan, S.G., Nayar, S.K. Vision and the Atmosphere. *International Journal of Computer Vision* 48, 233–254 (2002). <https://doi.org/10.1023/A:1016328200723>
3. Li, Y., You, S., Brown, M.S., & Tan, R.T. (2017). Haze Visibility Enhancement: A Survey and Quantitative Benchmarking. *Vis. Image Underst.*, 165, 1-16. <https://doi.org/10.1016/j.cviu.2017.09.003>
4. Berman, T. Treibitz and S. Avidan, "Non-local Image Dehazing," 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1674-1682, <https://doi.org/10.1109/CVPR.2016.185>.
5. Zhu, J. Mai and L. Shao, "A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior," in *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3522-3533, Nov. 2015. <https://doi.org/10.1109/TIP.2015.2446191>
6. Xu, G. Zhai, X. Wu and X. Yang, "Generalized Equalization Model for Image Enhancement," in *IEEE Transactions on Multimedia*, vol. 16, no. 1, pp. 68-82, Jan. 2014, <https://doi.org/10.1109/TMM.2013.2283453>.
7. Jiang, B., Woodell, G.A. & Jobson, D.J. Novel multi-scale retinex with color restoration on graphics processing unit. *J Real-Time Image Proc* 10, 239–253 (2015). <https://doi.org/10.1007/s11554-014-0399-9>

8. Xu H T ,Zhai G T, Wu X L , et al. Generalized equalization model for image enhancement[J]. *IEEE Transactions on Multimedia* ,2014, 16(1):68-82.
9. Liu H B, Yang J,Wu ZP, et al. A fast single i.mage dehazing method based on dark channel prior and Retinex theory[J]. *Acta Automatica Sinica*, 2015, 41(7):1264-1273.
<https://doi.org/10.16383/j.aas.2015.c140748>
10. Narasimhan, S.G., Nayar, S.K. Vision and the Atmosphere. *International Journal of Computer Vision* 48, 233–254 (2002). <https://doi.org/10.1023/A:1016328200723>
11. He, J. Sun and X. Tang, "Single Image Haze Removal Using Dark Channel Prior," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341-2353, Dec. 2011, <https://doi.org/10.1109/TPAMI.2010.168>.
12. Zhu, J. Mai and L. Shao, "A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior," in *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3522-3533, Nov. 2015, <https://doi.org/10.1109/TIP.2015.2446191>.
13. Tarel and N. Hautière, "Fast visibility restoration from a single color or gray level image," 2009 IEEE 12th International Conference on Computer Vision, 2009, pp. 2201-2208, <https://doi.org/10.1109/ICCV.2009.5459251>.
14. Meng, Y. Wang, J. Duan, S. Xiang and C. Pan, "Efficient Image Dehazing with Boundary Constraint and Contextual Regularization," 2013 IEEE International Conference on Computer Vision, 2013, pp. 617-624, <https://doi.org/10.1109/ICCV.2013.82>.
15. Li, R. T. Tan and M. S. Brown, "Nighttime Haze Removal with Glow and Multiple Light Colors," 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 226-234, <https://doi.org/10.1109/ICCV.2015.34>.
16. Nishino, K., Kratz, L. & Lombardi, S. Bayesian Defogging. *Int J Comput Vis* 98, 263–278 (2012). <https://doi.org/10.1007/s11263-011-0508-1>
17. Yang, D., Sun, J. (2018). Proximal Dehaze-Net: A Prior Learning-Based Deep Network for Single Image Dehazing. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) *Computer Vision – ECCV 2018*. ECCV 2018. *Lecture Notes in Computer Science()*, vol 11211. Springer, Cham. https://doi.org/10.1007/978-3-030-01234-2_43
18. Dong et al., "Multi-Scale Boosted Dehazing Network With Dense Feature Fusion," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 2154-2164, <https://doi.org/10.1109/CVPR42600.2020.00223>.
19. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., & Wang, Z. (2017). RESIDE: A Benchmark for Single Image Dehazing. *ArXiv*, abs/1712.04143.
20. Li, X. Peng, Z. Wang, J. Xu and D. Feng, "AOD-Net: All-in-One Dehazing Network," in 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017 pp. 4780-4788. <https://doi.org/10.1109/ICCV.2017.511>
21. Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, Sung-Jea Ko; Rethinking Coarse-To-Fine Approach in Single Image Deblurring *Proceedings of the IEEE/CVF International Conference on*

- Computer Vision (ICCV), 2021, pp. 4641-4650 <https://doi.org/10.48550/arXiv.2108.05054>
22. Chen et al., "Gated Context Aggregation Network for Image Dehazing and Deraining," 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), 2019, pp. 1375-1383, <https://doi.org/10.1109/WACV.2019.00151>.
 23. Liu, M. Suganuma, Z. Sun and T. Okatani, "Dual Residual Networks Leveraging the Potential of Paired Operations for Image Restoration," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 7000-7009, <https://doi.org/10.1109/CVPR.2019.00717>.
 24. Berman, T. Treibitz and S. Avidan, "Non-local Image Dehazing," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1674-1682, <https://doi.org/10.1109/CVPR.2016.185>.
 25. Ancuti, C., Ancuti, C.O., Timofte, R., De Vleeschouwer, C. (2018). I-HAZE: A Dehazing Benchmark with Real Hazy and Haze-Free Indoor Images. In: Blanc-Talon, J., Helbert, D., Philips, W., Popescu, D., Scheunders, P. (eds) Advanced Concepts for Intelligent Vision Systems. ACIVS 2018. Lecture Notes in Computer Science(), vol 11182. Springer, Cham. https://doi.org/10.1007/978-3-030-01449-0_52
 26. Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, MH. (2016). Single Image Dehazing via Multi-scale Convolutional Neural Networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science(), vol 9906. Springer, Cham. https://doi.org/10.1007/978-3-319-46475-6_10.
 27. Zhang, V. Sindagi and V. M. Patel, "Multi-scale Single Image Dehazing Using Perceptual Pyramid Deep Network," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018, pp. 1015-101509, <https://doi.org/10.1109/CVPRW.2018.00135>.
 28. Li, J. Pan, Z. Li and J. Tang, "Single Image Dehazing via Conditional Generative Adversarial Network," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 8202-8211, <https://doi.org/10.1109/CVPR.2018.00856>.
 29. Ancuti, C., Ancuti, C.O., Timofte, R., De Vleeschouwer, C. (2018). I-HAZE: A Dehazing Benchmark with Real Hazy and Haze-Free Indoor Images. In: Blanc-Talon, J., Helbert, D., Philips, W., Popescu, D., Scheunders, P. (eds) Advanced Concepts for Intelligent Vision Systems. ACIVS 2018. Lecture Notes in Computer Science(), vol 11182. Springer, Cham. https://doi.org/10.1007/978-3-030-01449-0_52
 30. Qu, Y. Chen, J. Huang and Y. Xie, "Enhanced Pix2pix Dehazing Network," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 8152-8160, <https://doi.org/10.1109/CVPR.2019.00835>.
 31. Shao, L. Li, W. Ren, C. Gao and N. Sang, "Domain Adaptation for Image Dehazing," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 2805-2814, <https://doi.org/10.1109/CVPR42600.2020.00288>.
 32. Ren et al., "Gated Fusion Network for Single Image Dehazing," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 3253-3261, <https://doi.org/10.1109/CVPR.2018.00343>.

33. Chang, Meng and Li, Qi and Feng, Huajun and Xu, Zhihai. Spatial- Adaptive Network for Single Image Denoising. In European Conference on Computer Vision, 2020. https://doi.org/10.1007/978-3-030-58577-8_11
34. Kangfu Mei, Aiwen Jiang, Juncheng Li, and Mingwen Wang. Progressive feature fusion network for realistic image dehazing. In Asian Conference on Computer Vision, pages 203–215, 2018. https://doi.org/10.1007/978-3-030-20887-5_13
35. Liu, Y. Ma, Z. Shi and J. Chen, "GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 7313-7322, <https://doi.org/10.1109/ICCV.2019.00741>.
36. Liu, M. Suganuma, Z. Sun and T. Okatani, "Dual Residual Networks Leveraging the Potential of Paired Operations for Image Restoration," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 7000-7009, <https://doi.org/10.1109/CVPR.2019.00717>.
37. Zhang and V. M. Patel, "Densely Connected Pyramid Dehazing Network," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 3194-3203, <https://doi.org/10.1109/CVPR.2018.00337>.
38. T. Tan, "Visibility in bad weather from a single image," 2008 IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1-8, <https://doi.org/10.1109/CVPR.2008.4587643>.
39. Liu, Y. Ma, Z. Shi and J. Chen, "GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 7313-7322, <https://doi.org/10.1109/ICCV.2019.00741>.
40. Fattal, R. (2014). Dehazing Using Color-Lines. ACM Transactions on Graphics (TOG), 34, 1 - 14. <https://doi.org/10.1145/2651362>
41. Cai, X. Xu, K. Jia, C. Qing and D. Tao, "DehazeNet: An End-to-End System for Single Image Haze Removal," in IEEE Transactions on Image Processing, vol. 25, no. 11, pp. 5187-5198, Nov. 2016, <https://doi.org/10.1109/TIP.2016.2598681>.
42. Chen, L.C., Papandreou, G., Schroff, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587 (2017)
43. Zhou, J. Zhang, W. Zuo, H. Xie, J. Pan and J. S. Ren, "DAVANet: Stereo Deblurring With View Aggregation," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 10988-10997, <https://doi.org/10.1109/CVPR.2019.01125>.
44. Jiangxin Dong and Jinshan Pan. Physics-based feature de-hazing networks. In European Conference on Computer Vision, pages 188–204, 2020. https://doi.org/10.1007/978-3-030-58577-8_12

Figures

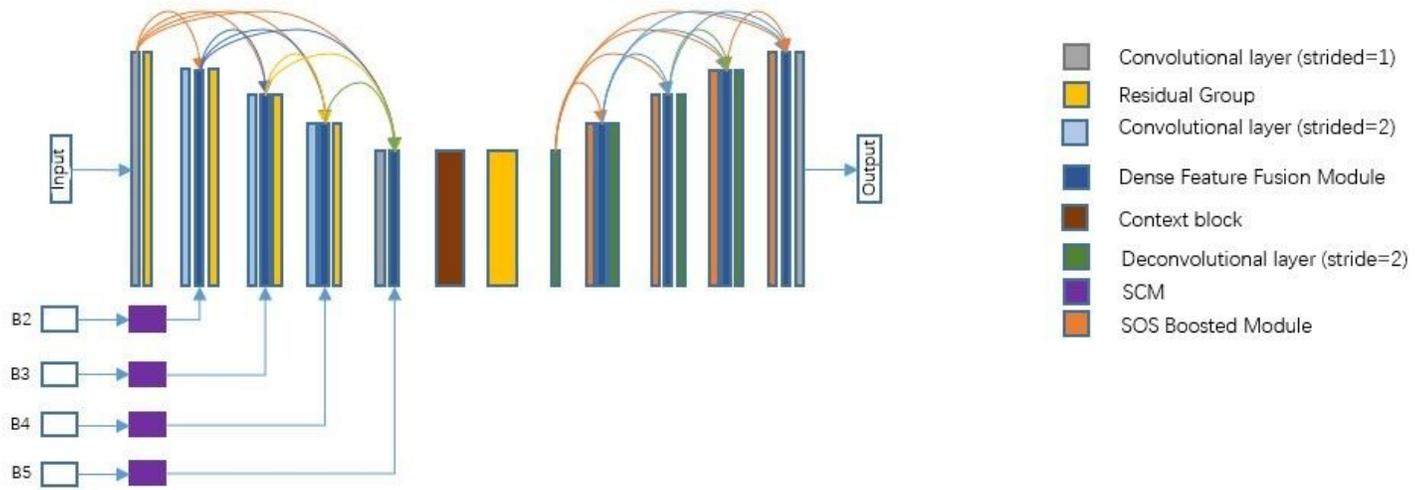


Figure 1

The framework of our proposed multi-input multi-scale U-net (MIMS-UNet) dehazing network.

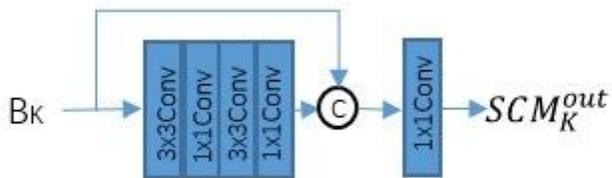


Figure 2

SCM module structure

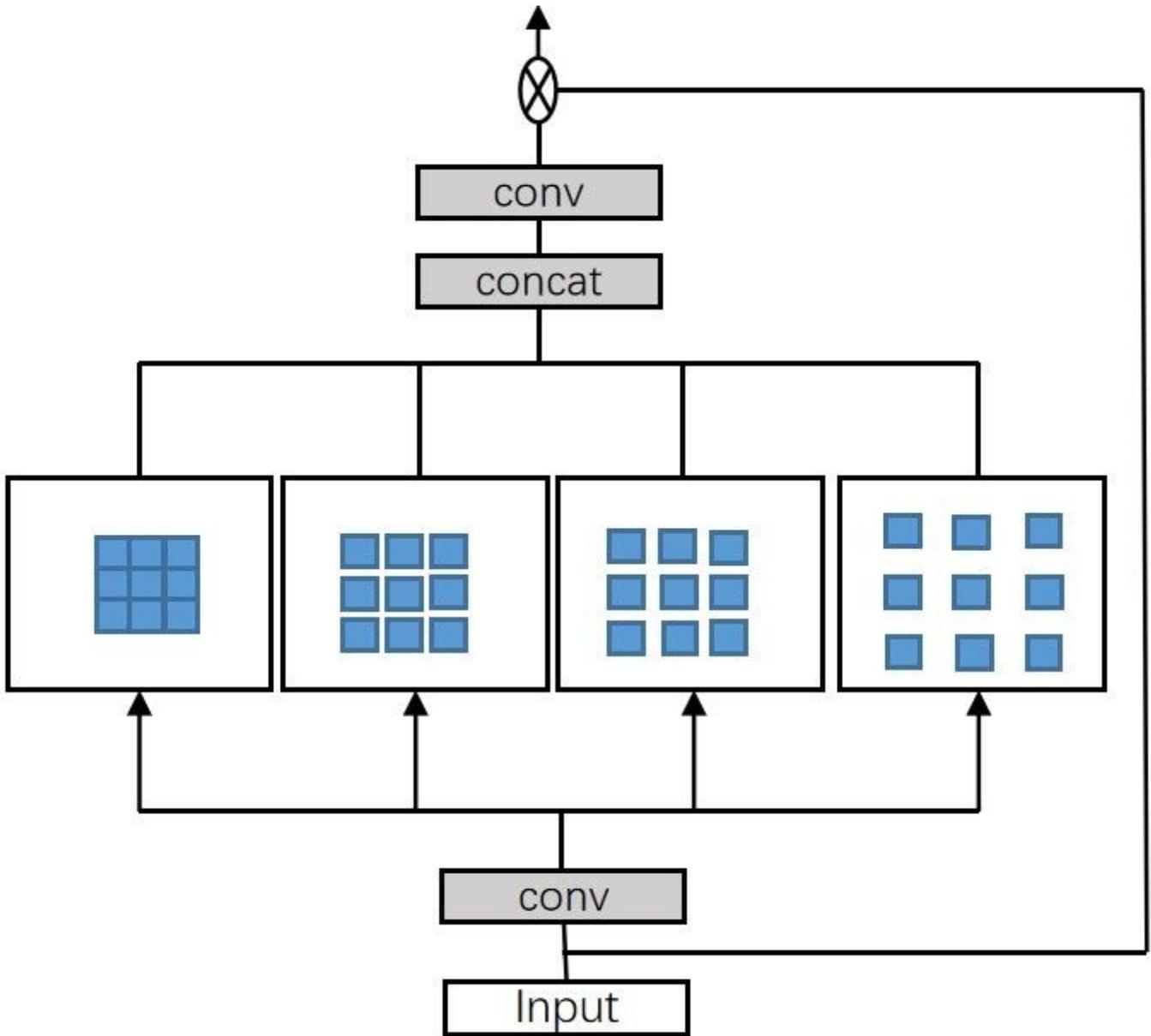


Figure 3

The architecture of the context block.

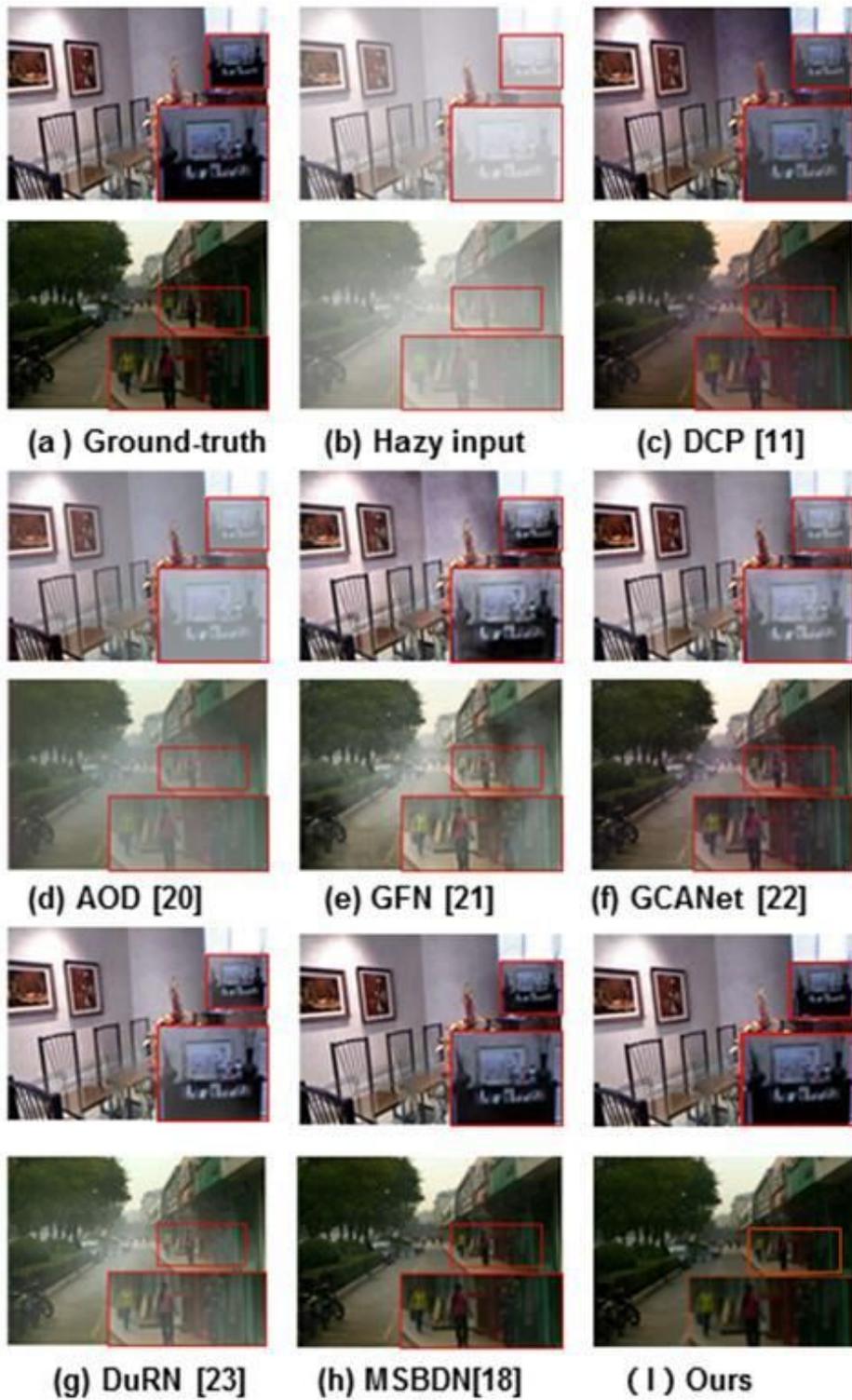


Figure 4

Visual results on the SOTS dataset.



Figure 5

Visual results on the O-HAZE dataset.

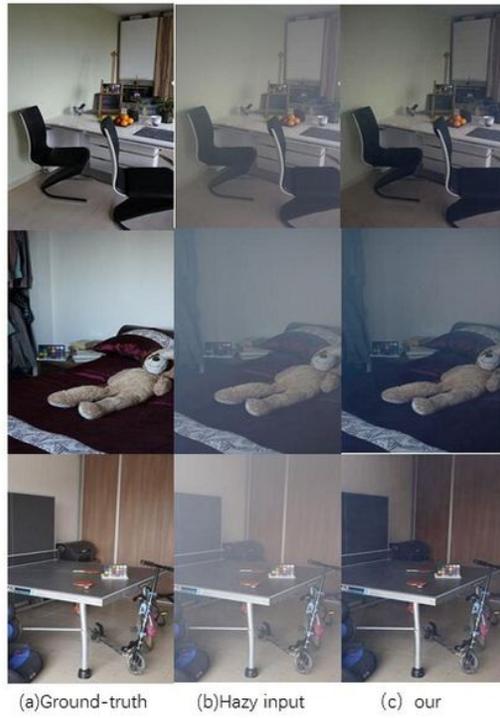


Figure 6

Visual results on the I-HAZE dataset.



Figure 7

Visualization of ablation results. (a) represents a clear image; (b) represents foggy images; (c) represents that our module image is not added; (d) Represents joining the MINS module; (d) represents the addition of MINS and background modules;