Journal of Imaging Science and Technology® 69(6): 1–7, 2025.
© Society for Imaging Science and Technology 2025

PEDNet: Physics-guided Encoder–Decoder Network for Image Dehazing

Yiming Yue
School of Computer Science and Engineering, Shenyang Jianzhu University, No. 25, Hunnan Middle Road, Shenyang, 110168, Liaoning, China

Long Ma
School of Mechanical Engineering, Shenyang Jianzhu University, No. 25, Hunnan Middle Road, Shenyang, 110168, Liaoning, China
E-mail: malong@sjzu.edu.cn

Peng Li
School of Computer Science and Engineering, Shenyang Jianzhu University, No. 25, Hunnan Middle Road, Shenyang, 110168, Liaoning, China

Abstract. *We propose a new convolutional neural network called Physics-guided Encoder–Decoder Network (PEDNet) designed for end-to-end single image dehazing. The network uses a reformulated atmospheric scattering model, which is embedded into the network for end-to-end learning. The overall structure is in the form of an encoder–decoder, which fully extracts and fuses contextual information from four different scales through skip connections. In addition, in view of the uneven spread of haze in the real world, we design a Res2FA module based on Res2Net, which introduces a Feature Attention block that is able to focus on important information at a finer granularity. The PEDNet is more adaptable when handling various hazy image types since it employs a physically driven dehazing model. The efficacy of every network module is demonstrated by ablation experiment results. Our suggested solution is superior to current state-of-the-art methods according to experimental results from both synthetic and real-world datasets.*
Keywords: *dehazing, deep-learning-based, encoder–decoder network, atmospheric scattering model*
© *2025 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.2025.69.6.060503]

1. INTRODUCTION

Haze is a phenomenon of reduced visibility due to tiny particles and gases suspended in the atmosphere. The existence of haze blurs the detailed information of an image and reduces contrast. This creates difficulties in tasks such as image processing and computer vision.

Although dehazing algorithms can improve image quality, completely restoring the true details and colors of an image remains difficult because of the complex differences between hazy and clear images. These differences are influenced by factors such as atmospheric light, transmission, and varying scenes and lighting conditions. Moreover, in different environments, the imaging process of hazy images is affected by various physical factors, making it challenging to fully restore a visual effect that aligns with the real haze-free image.

1.1 *Physical Models*

The atmospheric scattering model [1] is the most classical physical model in image dehazing algorithms. It explains how hazy images are imaged. It is modeled as follows:

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

where $I(x)$ is the input haze image and $J(x)$ is the output dehazed image; the atmospheric light α and transmission map $T(x)$ are unknown parameters. Therefore, according to this physical scattering model, the dehazing of a single image can be transformed into an estimate of the parameters $T(x)$ and α of the haze image.

Since the estimation of atmospheric light and the transmission map of haze images directly impacts the final outcome, accurate parameter estimation is a key factor in the dehazing effect. Some methods [2, 3] estimate the transmission map and atmospheric light separately using priors. However, these prior-based assumptions are often unsuitable, especially in white scenes or with unevenly distributed haze, and incorrect parameter estimation can result in artifacts in the output image. Additionally, separately estimating these parameters using convolutional neural network (CNN) based methods [4, 5] can lead to significant deviations in the results. Meanwhile, deep learning dehazing methods that do not rely on atmospheric scattering models [6, 7] depend solely on synthetic datasets for training, often producing low-quality results when faced with complex, real-world haze images.

To overcome these issues, we propose a new CNN model called Physics-guided Encoder–Decoder Network (PEDNet). This model combines an encoder–decoder architecture [8] with an atmospheric scattering model for end-to-end [9] mapping from hazy to ground-truth images.

Received Nov. 1, 2024; accepted for publication Mar. 7, 2025; published online May 28, 2025. Associate Editor: Gabriele Gianini.
1062-3701/2025/69(6)/060503/7/$25.00

The encoder and the decoder use skip connections, and the Rest2FA module extracts and fuses features at four scales. A reformulated atmospheric scattering model [10] is integrated into the network, addressing the limitations of prior-based methods and ensuring stable dehazing in complex scenarios. Extensive experiments show significant improvements in peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) along with better visual quality.

Our work mainly makes the following contributions:

1. We propose an efficient end-to-end trainable CNN called the PEDNet for image dehazing, which overcomes the challenges of parameter estimation in traditional dehazing methods [2, 11].
2. Unlike the usual U-Net [8] architecture, the skip connections between the encoder and the decoder in the PEDNet allow for better extraction and fusion of both shallow and deep contextual information, preserving image details more effectively.
3. Our Res2FA module introduces a Feature Attention (FA) block consisting of channel and pixel attention, which assigns greater importance to key features and adds flexibility in processing different types of information.
4. The network integrates the atmospheric scattering model, enhancing its generalization capability across various haze images and ensuring more stable dehazing performance.

2. RELATED WORK

Up until now, there have been nearly 30 years of research on single image dehazing methods, and the main methods can be roughly divided into two categories: prior-based methods [3, 4] and deep-learning-based methods [9, 10].

2.1 *Prior-based Image Dehazing Methods*

The biggest challenge in physical model dehazing methods lies in the estimation of parameters while prior-based image dehazing methods estimate the parameters through assumptions summarized from practical observations.

He et al. [4] proposed the Dark Channel Prior (DCP) based on the assumption that "most natural images contain some pixels with very low color intensity." Zhu et al. [12] introduced the Color Attenuation Prior (CAP), assuming that "the brightness and chromaticity of pixels in haze images vary significantly with haze concentration." Berman et al. [2] presented the Non-local Prior, suggesting that "colors in haze-free images can be effectively represented by numerous distinct colors grouped closely in RGB space."

Although prior-based methods have yielded numerous achievements in the area of image dehazing, they have certain limitations. For example, the algorithm accuracy depends on the accuracy of the prior knowledge, leading to the fact that it is not applicable to all cases.

2.2 *Deep-Learning-based Image Dehazing Methods*

Recently, deep-learning-based dehazing methods, with their strong learning ability and adaptive capacity, have shown excellent performance when encountering intricate situations and large-scale data, and show a broader space for development.

Zhang et al. [5] developed DCPDN, which uses a densely connected edge pyramid U-Net to estimate the transmission map and atmospheric light separately. They employed a joint discriminator from the generative adversarial network framework to merge the transmission map with structural information from dehazing results. Qin et al. [7] introduced FFA-Net, which optimizes feature fusion across different levels by using an FA block to give more weight to important features, allowing for adaptive learning of feature weights. Cai et al. [9] created DehazeNet, an end-to-end dehazing system that estimates the transmission map, incorporating a BReLU nonlinear activation function to enhance the clarity of haze-free images.

Different from these methods, we propose a new PEDNet that uses a skip-connected encoder–decoder architecture to fuse multi-scale features of images. Moreover, our dehazing network embedded with a reformulated atmospheric scattering model can better combine physical laws and data-driven learning to provide more stable and accurate dehazing results.

3. PEDNet

In this section, we first reformulate the atmospheric scattering model. Then, we describe PEDNet, which is designed based on this model. As illustrated in Figure 1, the PEDNet establishes the mapping between hazy images and ground-truth images. The network uses an encoder–decoder structure to combine multi-scale contextual information, introduces an FA block to weight and extract important features, and integrates the reformulated atmospheric scattering model for training.

3.1 *Model Reformulation*

The description of the atmospheric scattering model according to Eq. (1) can also be expressed as

$$J(x) = \frac{1}{T(x)} I(x) - \frac{\alpha}{T(x)} + \alpha. \tag{2}$$

Inspired by AOD-Net [10], transforming Eq. (2) can optimize the end-to-end pipeline from hazy images to ground-truth images, which is described as

$$J(x) = G(x)I(x) - G(x) + c, \tag{3}$$

where

$$G(x) = \frac{\frac{1}{T(x)}(I(x) - \alpha) + (\alpha - c)}{I(x) - 1}. \tag{4}$$

In this way, from the original need to calculate atmospheric light α and transmission map $T(x)$ separately, after reformulation, only one parameter $G(x)$ needs to be calculated; c takes the default value of 1 as a constant deviation. The reformulated model not only reduces the difficulty of calculation but also avoids the superposition of

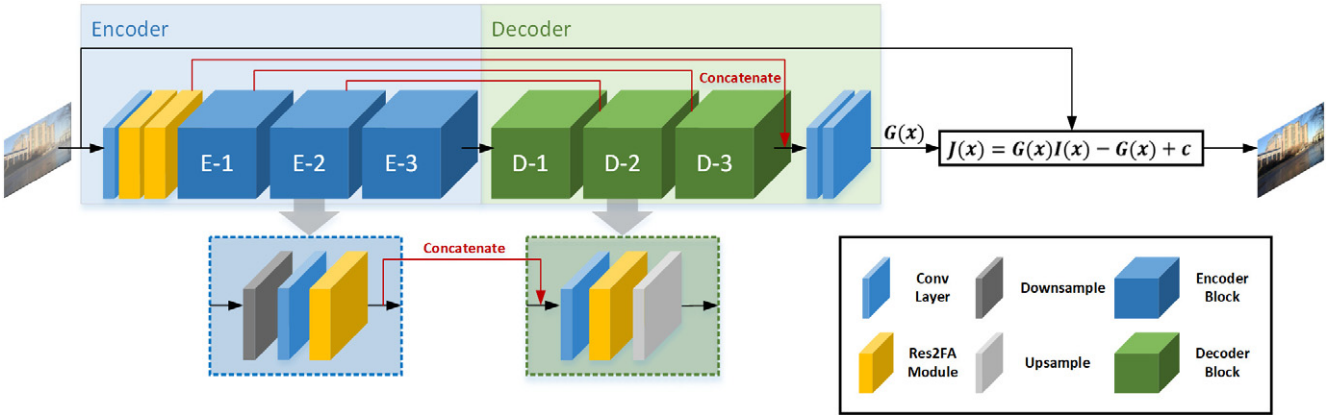

Figure 1. The architecture of our proposed PEDNet.

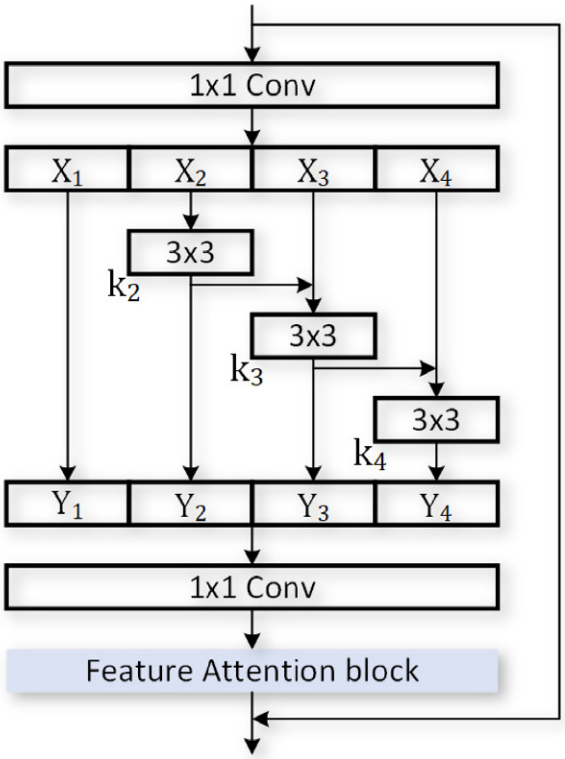

Figure 2. Res2FA module.

errors caused by separate calculations. Since $G(x)$ depends on the input $I(x)$, our goal is to create a deep model that can adapt to different inputs and then train the model by minimizing the loss function between its output and the ground-truth image.

3.2 *Res2FA Module*

Considering that features at different scales may have different levels of importance, in search of an efficient feature extraction module, we designed a Res2FA module inspired by [7, 13]. The Res2FA module, shown in Figure 2, is a CNN that adds the FA block to the Res2Net module.

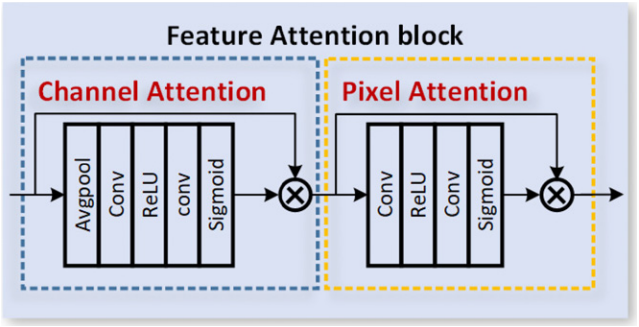

Figure 3. Feature Attention block.

The FA block is made up of channel attention and pixel attention. As depicted in Figure 3, the channel attention is utilized to attend to the input x_c weighted information at different channel features, which is described as

$$CA_c = \sigma(\mathrm{Conv}(\delta(\mathrm{Conv}(\mathrm{Avgpool}(x_c))))), \tag{5}$$

where *Avgpool* denotes global average pooling, *Conv* denotes the convolutional layer, δ denotes the BReLU function, and σ is the sigmoid function.

Then, the input x_c and the channel CA_c are multiplied element-wise to obtain the output F_c after the channel attention:

$$F_c = CA_c \otimes x_c. \tag{6}$$

Given the uneven distribution of haze across various image types, incorporating pixel attention can help the network prioritize features such as densely hazy pixels and high-frequency image regions. This module operates similarly to CA_c. It takes F_c as the input and then passes through a network composed of two convolutional layers, the BReLU function and the sigmoid function:

$$PA = \sigma(\mathrm{Conv}(\delta(\mathrm{Conv}(F_c)))). \tag{7}$$

Subsequently, the input F_c and the output PA undergo an element-wise multiplication operation to obtain the final

output FA of the FA block:

$$FA = F_c \otimes PA. \tag{8}$$

The introduced FA block is able to adaptively learn the weights of the feature channels. Increasing the flexibility in handling various types of information enables the network to prioritize essential features in the dehazing task.

In summary, the Res2FA module incorporates the multi-scale feature representation of Res2Net and the weight adjustment capability of the FA block. In this way, the network can better capture features at different levels and scales, significantly improving the ability to express important features.

3.3 *Encoder–Decoder Network*

Inspired by [14], our PEDNet uses an encoder–decoder network, which has proven effective in tasks like semantic segmentation and object detection. We specifically design this network to enhance dehazing by training on various levels of detail in the image. The encoder captures global and local features while the decoder applies these features to improve image dehazing. Skip connections between the encoder and the decoder facilitate information transfer and accelerate convergence, allowing for efficient combination of multi-scale contextual information and improved dehazing quality and accuracy.

As shown in Fig. 1, the PEDNet consists of an encoder–decoder network and an atmospheric scattering model. The encoder–decoder network facilitates the mapping of the hazy image to $G(x)$, and the dehazing process is ultimately achieved by the reformulated atmospheric scattering model.

This encoder–decoder network can extract features from four different scales. The encoder first uses a 3×3 convolutional layer to capture shallow information and then employs two linked Res2FA modules for finer detail at the current scale. PixelUnShuffle is used for downsampling in the next three encoder blocks without losing the total feature information, followed by a 3×3 convolutional layer and Res2FA module for feature extraction. After each encoder module, the feature map resolution is reduced by a factor of 1, and the channel count doubles.

The decoder network mirrors the encoder network but with a D-1 decoder block using a 1×1 convolutional layer for improved deep features while the last two decoder modules use a 3×3 convolutional layer connected to a Res2FA module and PixelShuffle. The resolution of feature maps doubles with each decoder pass. Finally, the haze image is mapped to $G(x)$ using two 3×3 convolutional layers.

Notably, skip connections between the encoder and the decoder reduce information loss and provide more detailed information, enhancing the quality of the dehazing results.

3.4 *Loss Function*

The mean squared error (MSE) is a simple and widely used loss function in image dehazing tasks. Since there is often a large difference in pixel details between dehazed images and ground-truth images, the loss function MSE can amplify this difference and speed up the training process of network parameters. Although many dehazing algorithms also employ perceptual loss functions, we choose the MSE to optimize the network model. The loss function L_{mse} is defined as follows:

$$L_{\text{mse}} = \frac{1}{N} \sum_{k=1}^{N} \| J_{\text{haze}}^{k} - J_{gt}^{k} \|^2, \tag{9}$$

where J_{haze} is the image after dehazing with the network, J_{gt} is the clear image, and k is the pixel coordinates on the image.

4. EXPERIMENTAL RESULTS

4.1 *Datasets*

Since it is difficult to obtain a substantial quantity of haze images and corresponding ground-truth images in the real world, synthetic haze datasets are often used to train models. RESIDE [15] is a large-scale synthetic dataset that is widely used in the field of image dehazing due to its diversity in evaluation criteria [7, 16]. We use subsets from RESIDE: ITS and OTS for training and SOTS, containing both indoor and outdoor images, for testing.

Moreover, in order to demonstrate the robustness and diversity of PEDNet, we also selected a real-world dataset [11] for testing. Since this dataset does not have corresponding ground-truth images, only a qualitative evaluation of the dehazing effect is performed.

4.2 *Implementation Details*

In the course of our experiment, we verified the feasibility and efficiency of PEDNet in the image dehazing task. The PEDNet parameters were optimized in the training process using the Adam optimizer. The attenuation factors β_1 and β_2 took the default values 0.9 and 0.999 respectively. The initial learning rate and weight falloff were both set at 1×10^{-4}. Due to the large dataset, the network was trained for 50 rounds by default and the results were found to be stable. All experiments were trained with PyTorch and an NVIDIA RTX 4090 GPU with a default batch size of 4.

4.3 *Evaluation of Dehazing Results*

4.3.1 *Evaluation Standard and Competitors*

To evaluate image dehazing methods, we use the PSNR and the SSIM as objective evaluation criteria. The PSNR primarily quantifies the pixel-level differences between the restored and ground-truth images, with higher PSNR values indicating a smaller pixel-level discrepancy, thereby reflecting the accuracy of the dehazing process. The SSIM, on the other hand, assesses the luminance, contrast, and structural similarity between images, offering a more comprehensive evaluation of visual quality and detail restoration, particularly in terms of preserving image structure and texture. The combination of PSNR and SSIM effectively captures the degree of distortion and similarity between images, and these metrics have been widely used in image dehazing quality assessment.

We evaluate the proposed PEDNet against six state-of-the-art dehazing methods on two datasets, which

Table I. Quantitative comparison on SOTS dataset.

Methods	Indoor		Outdoor	
	PSNR	SSIM	PSNR	SSIM
DCP [4]	16.62	0.8179	19.13	0.8148
CAP [12]	19.05	0.8364	21.09	0.8685
AOD-Net [10]	19.06	0.8504	20.29	0.8765
DehazeNet [9]	21.14	0.8472	22.46	0.8514
GFN [16]	22.30	0.8800	21.55	0.8444
4kDehazing [17]	23.35	0.9280	27.55	0.9500
Ours	**27.55**	**0.9415**	**29.23**	**0.9633**

are DCP [4], CAP [12], AOD-Net [10], DehazeNet [9], GFN [16], and 4kDehazing [17]. The DCP and CAP are prior-based methods. The AOD-Net is a deep-learning-based method embedded in the atmospheric scattering model. The DehazeNet outputs transmission maps through a CNN. The GFN and 4kDehazing are direct mapping deep-learning-based methods.

4.3.2 *Evaluation on Synthetic Dataset*

Comparative experiments were conducted on the SOTS dataset [15], which contains 500 indoor and 500 outdoor haze images. For convenience, the metrics of other methods are quoted from [18, 19]. The results in Table I show that our method outperforms six other methods in both PSNR and SSIM. Furthermore, Figure 4 provides a qualitative comparison with six state-of-the-art methods. The first three rows display indoor results, and the last three rows show outdoor results. A comparative analysis reveals that due to the prior assumptions in DCP and CAP, the dehazed images appear darker, lacking detail and showing color distortion. Although AOD-Net and DehazeNet improve brightness, haze residues remain. The GFN and 4kDehazing produce results closer to the ground truth, but our method performs better in edge texture and color fidelity. For instance, our method accurately renders the desktop color in the first image and better preserves the blind texture in the second image. Additionally, our method restores the true sky color more effectively.

4.3.3 *Evaluation on Real-world Dataset*

We further compared the proposed method using a real-world dataset [11]. Since this dataset does not have a corresponding ground-truth image, we only conduct qualitative analysis on this dataset. As shown in Figure 5, our proposed network demonstrates outstanding performance in image dehazing. First, in terms of edge processing, the network is able to achieve smooth and coherent results without obvious artifacts or distortions. For example, our network is able to accurately preserve details and present a natural and smooth effect in the hair part in the initial image and the eaves articulation in the second image. Second, the network also excels in color reproduction. It is able to accurately restore the colors distorted by the haze, making the colors of the image more in line with a real haze-free scene.

In summary, our proposed network demonstrates excellent performance in both quantitative and qualitative evaluation of image dehazing methods. It exhibits smoother and more coherent results in edge processing and is able to accurately reproduce colors in haze-free situations.

Table II. Comparison in different configurations on the SOTS dataset.

Model	AS	PSNR	SSIM
$R=1$	×	23.53	0.8449
$R=2$	×	25.42	0.8813
$R=3$	×	26.51	0.9047
$R=4$	×	27.35	0.9354
PEDNet	√	**28.39**	**0.9524**

4.4 *Ablation Study*

To validate the effectiveness of our proposed PEDNet for single image dehazing, we conducted ablation experiments with different configurations of its backbone module. Since PEDNet focuses on feature extraction at four different scales and owing to the importance of combining atmospheric scattering models, S represents the number of different scales to be fused and AS indicates whether the atmospheric scattering model is embedded. When $S=1$, it means the basic network, including three convolutional layers and two Res2FA modules. When $S=2$, 3, and 4, it means the addition of 1, 2, and 3 groups of encoder–decoder blocks to the basic network respectively. Note that none of the above four networks incorporate the atmospheric scattering model. Only the fifth PEDNet embeds it in the model.

The SOTS dataset of RESIDE was selected for the testing. The experimental details are the same as previously described. The ultimate experimental outcomes are derived from the mean of the measurements obtained on the SOTS indoor and outdoor haze images. As shown in Table II, each additional scale increases the PSNR by 1.4 db on average and the SSIM by nearly 3%. The PEDNet, which incorporates the atmospheric scattering model, increases the PSNR by 1.04 db compared with the fourth method. Obviously, the encoder–decoder block and the atmospheric scattering model have a significant impact on the PSNR and the SSIM.

5. CONCLUSION

In this paper, we propose a new single image dehazing method called the PEDNet, which is a combination of an encoder–decoder network and an atmospheric scattering model. Image features are extracted from four different scales in the network, and then the features of different scales are fully integrated through skip connections. In addition, multiple Res2FA backbone modules have been introduced to focus on important haze information at a more fine-grained level. Finally, the reformulated atmospheric scattering model

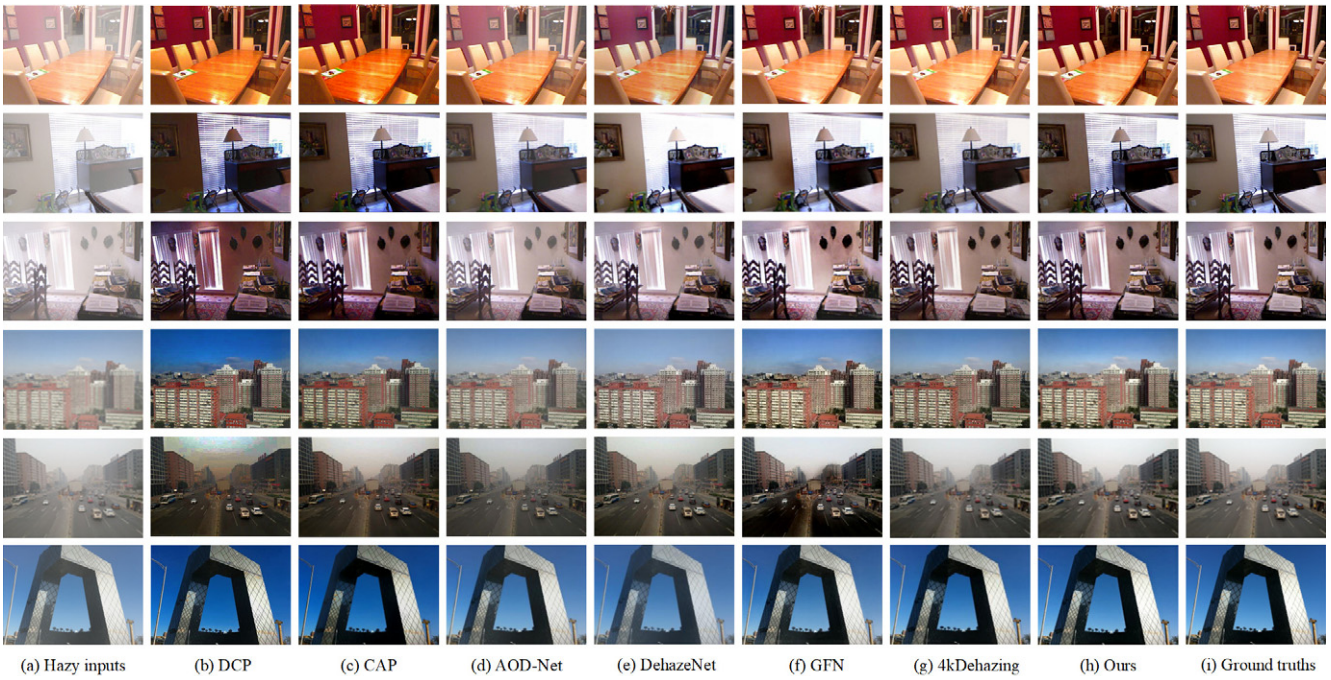

Figure 4. Qualitative comparison on SOTS dataset from RESIDE.

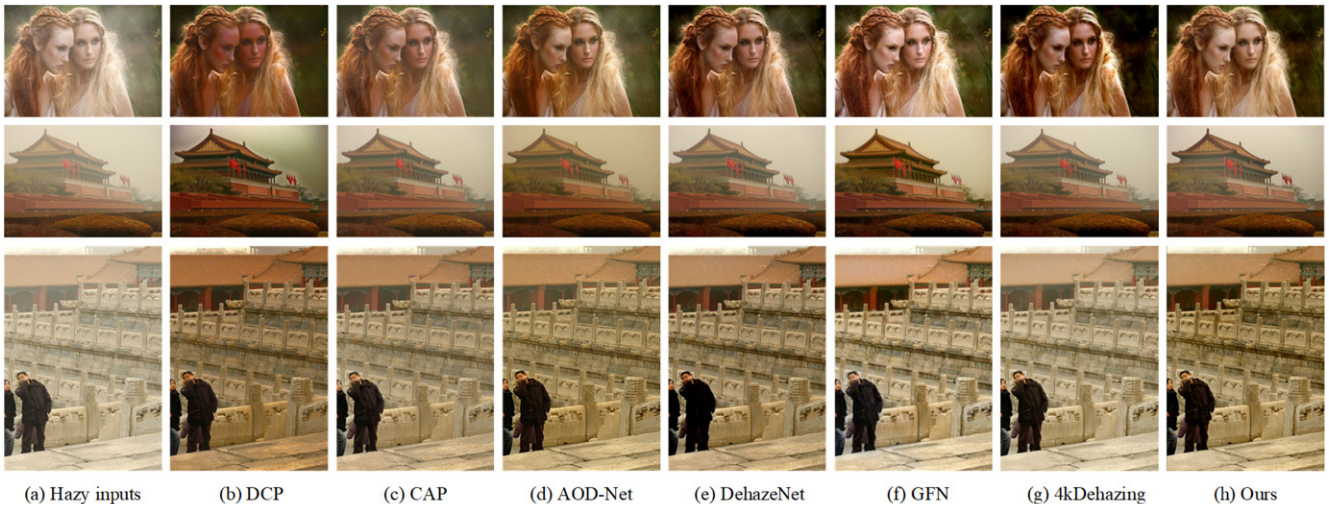

Figure 5. Qualitative comparison on real-world dataset.

is integrated to complete the end-to-end mapping. We evaluate the PEDNet against six state-of-the-art methods in multiple aspects on synthetic and real-world datasets, and the results indicate that our proposed method has a superior dehazing effect.

REFERENCES

[1] E. J. McCartney, *Optics of the Atmosphere: Scattering by Molecules and Particles* (Wiley, New York, 1976).

[2] D. Berman, T. Treibitz, and S. Avidan, "Non-local image dehazing," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2016), pp. 1674–1682.

[3] J.-P. Tarel and N. Hautiere, "Fast visibility restoration from a single color or gray level image," *2009 IEEE 12th Int'l. Conf. on Computer Vision* (IEEE, Piscataway, NJ, 2009), pp. 2201–2208.

[4] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," IEEE Trans. Pattern Anal. Mach. Intell. **33**, 2341–2353 (2010).

[5] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2018), pp. 3194–3203.

[6] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and G. Hua, "Gated context aggregation network for image dehazing and deraining," *2019 IEEE Winter Conf. on Applications of Computer Vision (WACV)* (IEEE, Piscataway, NJ, 2019), pp. 1375–1383.

[7] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: Feature fusion attention network for single image dehazing," *Proc. AAAI Conf. on Artificial Intelligence* (AAAI, Palo Alto, California, USA, 2020), Vol. 34, pp. 11908–11915.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: Proc., Part III 18th Int'l.*

Conf (Springer International Publishing, Cham, Switzerland, 2015), pp. 234–241.

[9] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," IEEE Trans. Image Process. **25**, 5187–5198 (2016).

[10] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-one dehazing network," *Proc. IEEE Int'l. Conf. on Computer Vision* (IEEE, Piscataway, NJ, 2017), pp. 4770–4778.

[11] R. Fattal, "Dehazing using color-lines," ACM Trans. Graph. (TOG) **34**, 1–14 (2014).

[12] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," IEEE Trans. Image Process. **24**, 3522–3533 (2015).

[13] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2net: A new multiscale backbone architecture," IEEE Trans. Pattern Anal. Mach. Intell. **43**, 652–662 (2019).

[14] M. Z. Alom, C. Yakopcic, T. M. Taha, and V. K. Asari, "Nuclei segmentation with recurrent residual convolutional neural networks based U-Net (r2U-Net)," *NAECON 2018 - IEEE National Aerospace and Electronics Conference* (IEEE, Piscataway, NJ, 2018), pp. 228–233.

[15] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," IEEE Trans. Image Process. **28**, 492–505 (2018).

[16] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M.-H. Yang, "Gated fusion network for single image dehazing," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2018), pp. 3253–3261.

[17] Z. Zheng, W. Ren, X. Cao, X. Hu, T. Wang, F. Song, and X. Jia, "Ultra-high-definition image dehazing via multi-guided bilateral learning," *2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Piscataway, NJ, 2021), pp. 16180–16189.

[18] J. Gui, X. Cong, Y. Cao, W. Ren, J. Zhang, J. Zhang, J. Cao, and D. Tao, "A comprehensive survey and taxonomy on single image dehazing based on deep learning," ACM Comput. Surv. **55**, 1–37 (2023).

[19] Y. Qu, Y. Chen, J. Huang, and Y. Xie, "Enhanced pix2pix dehazing network," *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2019), pp. 8160–8168.