Attention-based Feature Enhanced Dehazing Network

Yuanshun Cheng¹, Anhui Tan²,³, Chuansheng Yang²,³, Yueting Yang⁴ and Chao Wang²,³*

¹Tianfu College of SWUFE, Mianyang, Sichuan, 621000, China
²School of Information Engineering, Zhejiang Ocean University, Zhoushan, Zhejiang, 316022, China
³Key Laboratory of Oceanographic Big Data Mining and Application of Zhejiang Province, Zhoushan, Zhejiang, 316022, China
⁴School of Mathematics and Statistics, Beihua University, Jilin, Jilin, 132013, P.R. China
*Corresponding author’s e-mail: wangchao@zjou.edu.cn

Abstract. Presence of haze in images obscures underlying information, which is undesirable in applications requiring accurate environment information. To recover such an image, a dehazing algorithm should enhance the feature information of the background while weakening the feature information of haze. In this paper, we propose an end-to-end attention-based feature enhanced dehazing network (AEDNet), which integrates enhancement strategy and attention mechanism, to achieve haze removal. The network is based on U-Net, which has the advantages of retaining information, obtaining multi-scale features and so on. In the training of the network, pixel loss and perceptual loss are used to preserve feature information and improve the overall quality of results. The extensive evaluation shows that the proposed model performs significantly better than previous dehazing methods on various benchmarks.

1. Introduction
In hazy weather, the atmospheric light reflected from objects will be absorbed and scattered by atmospheric suspended particles, resulting in serious degradation of the image captured by sensors. Image dehazing methods can restore these low-quality images to clear haze-free images, and provide high-quality images for high-level computer vision tasks under bad weather, e.g., automatic driving and criminal investigation. Thus, more and more attention of this field has been drawn to image dehazing.

1.1. Related Work
Many previous works [1,2,3] rely on the following mathematical formula [4] to form haze images:

\[ I(x) = T(x)J(x) + (1 - T(x))A(x), \] (1)

where \(I(x)\) is the observed hazy image, \(J(x)\) is the corresponding clear haze-free image, \(A(x)\) is the global atmospheric light intensity depends on the unknown depth map, \(T(x)\) is the transmission map, and \(x\) is the pixel position. Therefore, hazy images restoration is to first estimate the transmission map and the global atmospheric light intensity and then calculate the results through formula (1).

In addition, some works [5,6] have overcome the problem of color distortion caused by transmission map-based methods. In recent years, some algorithms based on deep CNN directly estimate clean images,
such as [7,8,9], which are data-driven end-to-end methods, have also achieved great success. A feature fusion attention network [9] is proposed to directly restore haze-free images. In order to deal with different channel features and uneven haze distribution, the fusion attention block is composed of a channel attention module and pixel attention module.

1.2. Description of Motivation
Compared with traditional methods, learning-based methods attempt to directly restore the intermediate transmitted image. In order to utilize features at different levels, U-Net [10] and other networks strive to integrate shallow and deep information. Based on U-Net, we employ an enhancement strategy and attention mechanism to selectively capture information from encoder to decoder to restore the haze-free images.

The attention mechanism can process the nonhomogeneous haze well. So we design a new multi-level attention module (MLA) inspired by [9,11].

In general, the contributions of this work are as follows:
- We put forward a new end-to-end attention-based feature enhanced dehazing network AEDNet. This network is superior to the most advanced single image dehazing methods.
- We designed a multi-level attention (MLA) module. Based on the attention mechanism, this module can effectively solve the problem of uneven haze distribution in different pixels and channels.

2. Proposed Method
This chapter focuses on our AEDNet. As shown in figure 1, after the hazy image is input into the network, it will pass through three parts successively. The first part is the encoder ($G_{Enc}$) used for feature extraction, the second part is the feature processing unit ($G_{Res}$), and the last part is the decoder ($G_{Dec}$) used for the haze-free image reconstruction.

![Figure 1: AEDNet Attention-based Feature Enhanced Dehazing Network](image)

2.1. Multi-level Attention Model
MLA, which is mainly composed of feature attention module and residual group, is widely used in the network proposed (see figure 2). In the $G_{Enc}$, MLA is placed behind the dense feature fusion module at
each scale level to perform information weighting and feature fused extraction. And in the $G_{Dec}$, MLA is embedded in feature enhancement modules at each scale level.

In order to deal with the channel feature weighting and uneven haze distribution in the image, the Feature Attention module [9] is proposed. This module treats channels and pixels differently, and the key step is to generate different weights for each channel and pixel of the feature maps.

Feature Attention module consists of Channel Attention (CA) and Pixel Attention (PA), as shown in figure 2 (b). For CA, each channel information first needs to be fused through the global average pooling. And then the feature maps pass through two convolution layers, respectively followed by the Sigmoid and ReLU activation functions.

$$CA_c = \sigma(Conv(\delta(Conv(P(F_c))))),$$

(2)

where $F_c$ is the input, $\sigma$ is the sigmoid function, $\delta$ is the ReLU function, $P$ is the global average pooling, and $CA_c$ is the channel weights. And then we multiply the weights of $CA_c$ with the input $F_c$.

$$F'_c = CA_c \otimes F_c.$$  

(3)

Similar to CA, we implement PA using all operations except global average pooling.

$$PA = \sigma(Conv(\delta(Conv(F^*)))).$$ 

(4)

$$\hat{F} = F^* \otimes PA.$$  

(5)

where $\hat{F}$ is the output of the Feature Attention module (FA).

This module has been shown to be effective in image dehazing, so we directly apply it to our multi-level attention model. Similar to [12], we use the same residual group (containing 3 residual blocks) in the MLA module and 18 residual blocks [13] in the $G_{Res}$.

Figure 2. Multilevel attention module diagram. (a) is the characteristic attention module, (b) is the residual block structure.

2.2. Loss Function
Smooth $\ell_1$ loss and perceptual loss are applied to measure the difference between dehazing results and ground truth:

$$L = \lambda_1 L_1 + \lambda_{per} L_{per},$$

(6)

where $L_1$ and $L_{per}$ represent the $\ell_1$ loss and the perception loss, and $\lambda$ is the weight of each loss.

3. Experiments
Training data: Indoor training dataset (ITS) and outdoor training dataset (OTS) of RESIDE [14].

Testing data: After the training process, we used RESIDE’s synthetic objective testing set (SOTS) as our test dataset.

3.1. The experimental setup
For the data set training of our network, the Adam optimizer is adopted to achieve the training, with the initial learning rate $\alpha = 0.0001$ and momentum is 0.9. In this work, the batch size is set as 8 for better performance. Moreover, the network trained a total of $1 \times 10^5$ iterations, and every 10,000 iterations...
the training process was stopped for validation. Additionally, our experiments are all worked in PyTorch 1.3.1 on an NVIDIA GeForce RTX 2080Ti GPU with 11G video memory.

3.2. The experimental results
In this section, we will demonstrate the effectiveness of our AEDNet method by direct comparison with the most advanced (SOTA) and representative image dehazing methods. The quantitative results in table 1 show that our AEDNet has the best performance both indoors and outdoors.

Table 1. Achievement table of each method of haze removal in Indoor and Outdoor.

| Method           | Indoor PSNR | Indoor SSIM | Indoor PSNR | Indoor SSIM |
|------------------|-------------|-------------|-------------|-------------|
| DCP [1]          | 16.62       | 0.8179      | 19.13       | 0.8148      |
| AOD-Net [7]      | 19.06       | 0.8504      | 20.29       | 0.8765      |
| MSBDN-DFF [12]   | 33.79       | 0.9840      | 30.86       | 0.9871      |
| FFA-Net [9]      | **36.39**   | **0.9886**  | **33.57**   | **0.9871**  |
| AEDNet (Ours)    | **37.52**   | **0.9893**  | **35.14**   | **0.9883**  |

Compared with PSNR and SSIM of the other 4 methods, the higher the value, the better. Red is the highest and blue is the second highest.

Figure 3: Qualitative comparison of the five dehazing methods on the SOTS dataset, where the first three rows are indoor images and the last three rows are outdoor images. The first column is the hazy images and the last column is the corresponding ground truth.

4. Conclusion
We propose an end-to-end enhanced haze removal network with attention, AEDNet, based on U-Net architecture. For the uneven distribution of haze, we propose an MLA module which base on attention.
mechanism. For model training, we use perceptual loss function which is closer to human perception. This method can not only effectively preserve the feature information of the image, but also significantly improve the overall quality of the image. Extensive evaluations show that the proposed model performs well on image dehazing datasets compared to the most advanced methods.

Acknowledgments
This work was supported by the National Natural Science Foundation of China (12001005); Key Project of Natural Science Foundation Joint Fund of Jilin Province (2020122367JC); Science and Technology Development Project Program of Jilin Province (20190303132SF).

References
[1] He, K., Sun, J., & Tang, X. (2010) Single image haze removal using dark channel prior. IEEE transactions on pattern analysis and machine intelligence, 33(12): 2341-2353.
[2] Zhang, H., & Patel, V. M. (2018) Densely connected pyramid dehazing network. In Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3194-3203.
[3] Dong, J., & Pan, J. (2020, August) Physics-based feature dehazing networks. In European Conference on Computer Vision. Springer, Cham. pp. 188-204.
[4] McCartney, E. J. (1976) Optics of the atmosphere: scattering by molecules and particles. New York.
[5] Zhu, Q., Mai, J., & Shao, L. (2014, September) Single Image Dehazing Using Color Attenuation Prior. In BMVC.
[6] Berman, D., & Avidan, S. (2016) Non-local image dehazing. In Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1674-1682.
[7] Li, B., Peng, X., Wang, Z., Xu, J., & Feng, D. (2017) Aod-net: All-in-one dehazing network. In Proceedings of the IEEE international conference on computer vision. pp. 4770-4778.
[8] Liu, X., Ma, Y., Shi, Z., & Chen, J. (2019) Griddehazenet: Attention-based multi-scale network for image dehazing. In Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7314-7323.
[9] Qin, X., Wang, Z., Bai, Y., Xie, X., & Jia, H. (2020, April) FFA-Net: Feature fusion attention network for single image dehazing. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 11908-11915).
[10] Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. (2018) Image super-resolution using very deep residual channel attention networks. In Proceedings of the European conference on computer vision (ECCV). pp. 286-301.
[11] Mei, K., Jiang, A., Li, J., & Wang, M. (2018, December) Progressive feature fusion network for realistic image dehazing. In Asian conference on computer vision. Springer, Cham. pp. 203-215.
[12] Dong, H., Pan, J., Xiang, L., Hu, Z., Zhang, X., Wang, F., & Yang, M. H. (2020) 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.
[13] He, K., Zhang, X., Ren, S., & Sun, J. (2016) Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770-778.
[14] Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., & Wang, Z. (2018) Benchmarking single-image dehazing and beyond. IEEE Transactions on Image Processing, 28(1): 492-505.