Supplementary Materials: Perceiving and Modeling Density for Image Dehazing

Tian Ye†1, Yunchen Zhang†2, Mingchao Jiang†3, Liang Chen[4], Yun Liu[5], Sixiang Chen[1], and Erkang Chen1⋆

[1] School of Ocean Information Engineering, Jimei University, Xiamen, China
[2] China Design Group Co., Ltd. Nanjing, China
[3] JOYY AI GROUP, Guangzhou, China
[4] Fujian Provincial Key Laboratory of Photonics Technology, Fujian Normal University, Fuzhou, China
[5] College of Artificial Intelligence, Southwest University, Chongqing, China
{201921114031, 201921114013, ekchen}@jmu.edu.cn,
cydiachen@cydiachen.tech, jiangshaoyu1993@gmail.com, cl_0827@126.com,
yunliu@swu.edu.cn

1 Introduction

This document includes additional information, which presents the following information that can be beneficial for the readers:

- The inference time and model complexity comparison.
- The visualization results of proposed Separable Hybrid Attention Module.
- The detailed description of the architectures in ablation studies of the submitted paper.
- Additional visual comparison on real-world and synthetic hazy images.

2 Inference time and model complexity comparison

Table 1. Comparison of Inference time, GMACs (fixed-point multiply accumulate operations performed persecond) and Parameters.

Method	Inf. Time (ms)	GMACs (G)	Params (M)
FFA-Net [4]	486	302.99	4.6M
DMT-Net [3]	192	80.71	54.9M
Ours	231	81.13	18.9M

Our method achieves the best trade-off between parameters and performance. The time reported in the table corresponds to the time taken by each model feed forward an image of dimension 512×512 during the inference stage. We perform

⋆ Corresponding author. †Equal contribution.

all inference testing on an RTX3060 GPU for a fair comparison. Notably, we utilize the *torch.cuda.synchronize()* API function to get accurate feed forward run-time. As we claimed in the paper, FFA-Net is not lightweight enough. To verify this point, we present a comparison in Table.1. FFA-Net, in particular, is computationally complex while having fewer parameters (4.6M) than ours. As a result, our network is **52%** faster than FFA-Net in terms of runtime (**231ms vs. 486ms**). We also observe that our technique outperforms DMT-Net in terms of performance-parameter trade-off (**33.49dB/18.9M vs. 28.53dB/54.9M**).

3 Visualization results of Separable Hybrid Attention Module

In this section, we provide visualization results of our Separable Hybrid Attention Module.

In order to demonstrate the perception ability of the SHA Module, we conduct ablation study on haze images with different levels of haze density, which is provided by the Haze4K dataset. As is shown in Fig. 1, the proposed shallow layers successfully capture the spatial distribution of haze in spatial dimensions.

4 Detailed Architectures In Ablation Studies

In this part, we present the detailed description of model architectures in ablation studies of our submitted paper.

Table 2. The detailed structure of baseline network in ablation study - Effectiveness of Separated Hybrid Attention Module.

Base	SE	ECA	CBA	FA	SWRCA	SHA	MHAB	MHA-C
Conv (3,64)	Conv (3,64)	Conv (3,64)	Conv (3,64)	Conv (3,64)	Conv (3,64)	Conv (3,64)	Conv (3,64)	Conv (3,64)
Conv (64,64)	Conv (64,64)-SE	Conv (64,64)-ECA	Conv (64,64)-ECA	Conv (64,64)-FA	SWRCA	SHA (64,64)	SHA (64,64)	SHA (64,64)
Conv s2 (64,128)	Conv s2 (64,128)	Conv s2 (64,128)	Conv s2 (64,128)	Conv s2 (64,128)	Conv s2 (64,128)	Conv s2 (64,128)	Conv s2 (64,128)	Conv s2 (64,128)
Conv (128,128)	Conv (128,128)-SE	Conv (128,128)-ECA	Conv (128,128)-ECA	Conv (128,128)-FA	SWRCA	SHA(128,128)	SHA(128,128)	SHA(128,128)
Conv s2 (128,256)	Conv s2 (128,256)	Conv s2 (128,256)	Conv s2 (128,256)	Conv s2 (128,256)	Conv s2 (128,256)	Conv s2 (128,256)	Conv s2 (128,256)	Conv s2 (128,256)
Residual Block	Residual Block-SE	Residual Block-ECA	Residual Block-ECA	Residual Block-FA	Residual Block-SWRCA	Residual Block-SHA	MHAB	MHA-C
Residual Block	Residual Block-SE	Residual Block-ECA	Residual Block-ECA	Residual Block-FA	Residual Block-SWRCA	Residual Block-SHA	MHAB	MHA-C
Residual Block	Residual Block-SE	Residual Block-ECA	Residual Block-ECA	Residual Block-FA	Residual Block-SWRCA	Residual Block-SHA	MHAB	MHA-C
Residual Block	Residual Block-SE	Residual Block-ECA	Residual Block-ECA	Residual Block-FA	Residual Block-SWRCA	Residual Block-SHA	MHAB	MHA-C
Up-sample layer	Up-sample layer	Up-sample layer	Up-sample layer	Up-sample layer	Up-sample layer	Up-sample layer	Up-sample layer	Up-sample layer
Conv (128,128)	Conv (128,128)-SE	Conv (128,128)-ECA	Conv (128,128)-ECA	Conv (128,128)-FA	Conv (128,128)-SWRCA	SHA (128,128)	SHA (128,128)	SHA (128,128)
Up-sample layer	Up-sample layer	Up-sample layer	Up-sample layer	Up-sample layer	Up-sample layer	Up-sample layer	Up-sample layer	Up-sample layer
Conv (64,64)	Conv (64,64)-SE	Conv (64,64)-ECA	Conv (64,64)-ECA	Conv (64,64)-FA	Conv (64,64)-SWRCA	SHA (64,64)	SHA (64,64)	SHA (64,64)
Up-sample layer	Up-sample layer	Up-sample layer	Up-sample layer	Up-sample layer	Up-sample layer	Up-sample layer	Up-sample layer	Up-sample layer
Tail module	Tail module	Tail module	Tail module	Tail module	Tail module	Tail module	Tail module	Tail module

5 Visual Comparison on Real-world Hazy Images

Here, we test the performance of different image dehazing methods [2, 4, 1] on real-world hazy images from Web. The visual results are shown in Fig. 4.

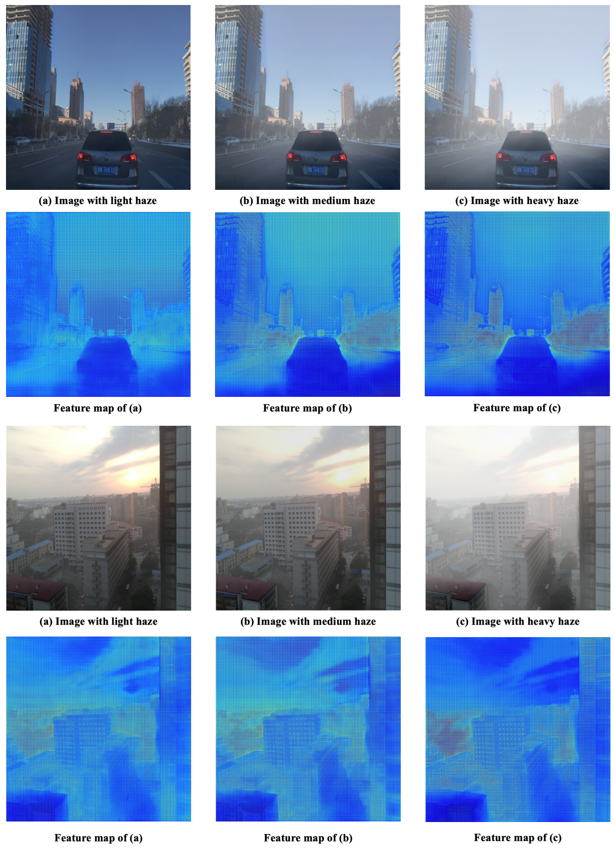

Fig. 1. The visualization results of the last convolution layer in the shallow layer. We calculated the mean values among the channel dimension of feature maps. We use ColorJet to visualize the value of pixels.

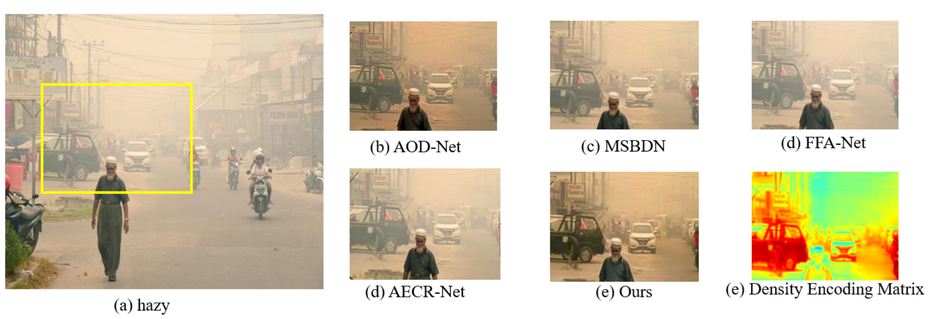

Fig. 2. Visual comparison on the real-world hazy image.

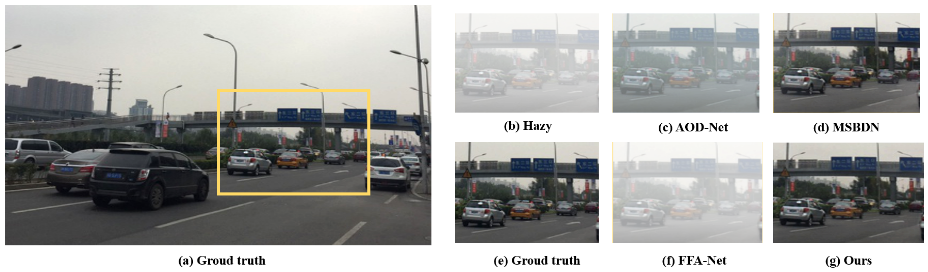

Fig. 3. Visual comparison on the synthetic hazy image.

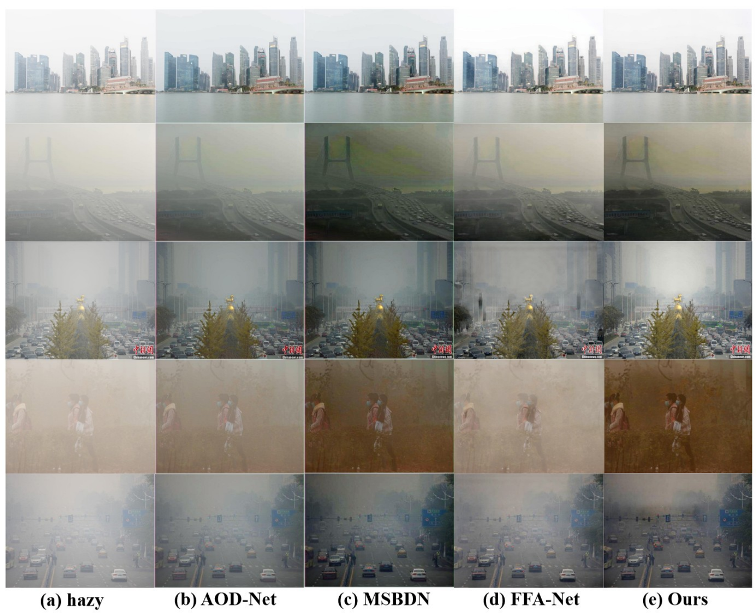

Fig. 4. Compared to other SOTA methods, our method results in a clearer image with the texture closer to the real sense image. The real-world hazy images are from web. The images are best viewed in the full-screen mode.

References

1. 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)
2. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: Aod-net: All-in-one dehazing network. In: Proceedings of the IEEE international conference on computer vision. pp. 4770–4778 (2017)
3. Liu, Y., Zhu, L., Pei, S., Fu, H., Qin, J., Zhang, Q., Wan, L., Feng, W.: From synthetic to real: Image dehazing collaborating with unlabeled real data. In: ACM Multimedia. pp. 50–58. ACM (2021)
4. 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)