Multi-Scale Boosted Dehazing Network with Dense Feature Fusion
Supplementary Material

Hang Dong[1]   Jinshan Pan[2]   Lei Xiang[1]   Zhe Hu[3]   Xinyi Zhang[1]
Fei Wang[1]   Ming-Hsuan Yang[4,5]
[1] College of Artificial Intelligence, Xi'an Jiaotong University
[2] Nanjing University of Science and Technology   [3] Hikvision Research America
[4] University of California, Merced   [5] Google Research
{dhunter1230,sdluran}@gmail.com   zhe.hu@hikvision.com
{xianglei96,jacqueline}@stu.xjtu.edu.cn   wfx@mail.xjtu.edu.cn   mhyang@ucmerced.edu

Overview

In this supplemental material, we provide the proof of **Proposition 1** and experimental verification of **Axiom 1** in Section A and Section B, respectively. To better illustrate the effectiveness of the SOS boosting strategy and dense feature fusion (DFF), we show the learned features of the SOS boosted module and DFF module in Section C. The efficiency analysis in Section E shows that the proposed method is efficient and entails low computational load. In Section F, we further demonstrate the effectiveness of the proposed method on the hazy images when the atmospheric light is not a global constant and the transmission map contains noise. Finally, more quantitative and qualitative comparisons on synthetic datasets and real-world images are provided in Section G.

A. Proof of Proposition 1

In the dehazing problem, a hazy image I can be modeled as

$$I = TJ + (1 - T)A, \qquad (1)$$

where J is the clean scene, T is the transmission map which is inversely proportional to the portion of haze (PoH), and A is the global atmospheric light. The clean scene J and global atmospheric light A are assumed to be constant in the same scene.

For image dehazing, the SOS boosting strategy can be formulated as

$$\hat{J}^{n+1} = g(I + \hat{J}^n) - \hat{J}^n, \qquad (2)$$

where $\hat{J}^n$ denotes the estimated image at the n-th iteration, $g(\cdot)$ is the dehazing approach, and $I + \hat{J}^n$ represents the strengthened image using the hazy input I. Here the portion of haze of the image I in (1) is defined as

$$PoH(I) = (1 - T)\frac{A}{J}, \qquad (3)$$

and it is proportional to $1 - T$ for hazy images of the same scene.

In the manuscript, we propose a proposition that the SOS boosting strategy can facilitate image dehazing performance in terms of Portion of Haze (PoH) under an axiom.

Axiom 1. *The dehazing method g obtains better results in terms of PoH on the images of the same scene but less haze. That is, if J_1 and J_2 are the images of the same scene, and $PoH(J_1) < PoH(J_2)$, then $PoH(g(J_1)) < PoH(g(J_2))$.*

Proposition 1. *Under Axiom 1, the SOS boosting strategy in (2) improves the dehazing performance, as*

$$PoH(\hat{J}^{n+1}) < PoH(\hat{J}^n). \qquad (4)$$

Proof: To prove **Proposition 1**, it is equivalent to show that

$$PoH(g(I + \hat{J}) - \hat{J}) < PoH(\hat{J}), \qquad (5)$$

where $\hat{J} = g(I)$.

By enforcing the physical model (1), we can formulate the dehazed image $\hat{J} = g(I)$ as

$$g(I) = T_g J + (1 - T_g)A, \qquad (6)$$

the strengthened image $I + \hat{J}$ as

$$I + \hat{J} = \frac{T + T_g}{2}(2J) + (1 - \frac{T + T_g}{2})(2A), \qquad (7)$$

and the dehazed image $g(I + \hat{J})$ as

$$g(I + \hat{J}) = T_{boost}(2J) + (1 - T_{boost})(2A), \qquad (8)$$

where T_g and T_{boost} are the transmission map of the dehazed image $g(I)$ and $g(I + \hat{J})$, respectively.

When the dehazing algorithm is effective, the dehazed image would contain less haze than the hazy input, $T_g > T$.

Then we obtain that the strengthened input $I+\hat{J}$ contains less portion of haze than the hazy input I, as $1 - \frac{T+T_g}{2} < 1 - T$. If **Axiom 1** holds, we would have

$$PoH(g(I + \hat{J})) < PoH(g(I)), \qquad (9)$$

and it suggests $T_{boost} > T_g$. Then we can obtain

$$
\begin{aligned}
PoH(g(I + \hat{J}) - \hat{J}) &= (1 - (2T_{boost} - T_g))\frac{A}{J} \\
&< (1 - T_g)\frac{A}{J} \\
&= PoH(g(I)). \qquad \square
\end{aligned}
\qquad (10)
$$

Table A. **Analysis of the portion of haze on the RESIDE dataset [7].** We randomly select 100 scenes from the evaluation set of the RESIDE dataset [7], and then select 3 hazy images of the same scene but different portions of haze. According to the portion of haze, 300 hazy images are grouped into 3 subsets corresponding to small, medium, and large portions of haze. All the evaluated models are trained on the RESIDE dataset with the same settings as mentioned in the manuscript. **Red** and blue texts indicate the best and the second-best performance of each method respectively.

Portion of Haze		Hazy	GFN [14]	PFFNet [12]	DuRN [11]	Ours
Small	PSNR	20.62	22.93	28.23	33.11	35.31
	SSIM	0.861	0.909	0.943	0.973	0.985
Medium	PSNR	14.20	22.89	27.26	30.84	33.78
	SSIM	0.776	0.893	0.923	0.968	0.979
Large	PSNR	10.56	20.56	24.31	27.98	31.60
	SSIM	0.692	0.869	0.900	0.943	0.973
Axiom holds true		—	83%	86%	92%	88%

B. Verification of the axiom for dehazing.

Axiom 1 is similar to the axiom for traditional denoising methods in SOS boosting paper [15]. Although the axiom does not always hold for deep learning-based dehazing methods, we can show that it holds in most situations.

To verify that, we evaluate different dehazing methods on images with different portions of haze. We evaluate the proposed MSBDN-DFF and three state-of-the-art deep dehazing networks, the GFN [14], PFFNet [12], and DuRN [11].

As shown in Table A and Figure A, all the methods perform better on the subset with less portion of haze. We also check if the axiom holds for images of each scene, and present the scene ratios that the axiom holds. The results in the last row of Table A show that the axiom holds in most scenes for all the four methods. We note that similar experiments and observations are presented in [14].

C. Effectiveness of the SOS boosting strategy and DFF

To better illustrate the effectiveness of the proposed SOS boosted modules and the DFF modules, we visualize the learned features in Figure B. As shown in Figure B(d)-(e), the upsampled feature $(\tilde{j}^2) \uparrow_2$ in Figure B(d) contains gridding artifacts because of the limitation of the strided deconvolutional layer in the decoder. We note that the SOS boosted

(a) Small PoH
17.05 / 0.904

(b) GFN [14]
25.79 / 0.948

(c) Ours
33.51 / 0.989

(a) Medium PoH
13.55 / 0.844

(b) GFN [14]
22.07 / 0.922

(c) Ours
30.25 / 0.976

(a) Large PoH
11.09 / 0.782

(b) GFN [14]
16.52 / 0.885

(c) Ours
28.24 / 0.972

Figure A. **Visual results of the hazy images of the same scene but different portions of haze (PoH).** The quantitative results in terms of PSNR and SSIM are provided. According to the qualitative and quantitative results, all the methods perform better on the hazy image with less PoH.
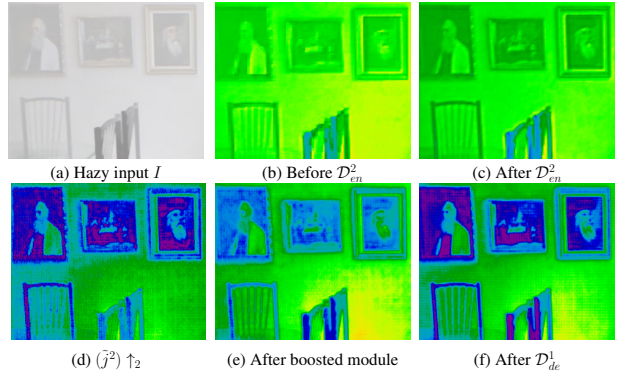

(a) Hazy input I

(b) Before $\mathcal{D}_{en}^2$

(c) After $\mathcal{D}_{en}^2$

(d) $(\tilde{j}^2) \uparrow_2$

(e) After boosted module

(f) After $\mathcal{D}_{de}^1$

Figure B. **Visualization of the learned features.** (b) denotes the input feature of the DFF at the 2-nd level of the encoder $\mathcal{D}_{en}^2$. (c) denotes the enhanced feature $\tilde{i}^2$ by $\mathcal{D}_{en}^2$. (d) is the upsampled feature $(\tilde{j}^2) \uparrow_2$ from the 2-nd level of the decoder. (e) is the boosted feature j^1 by the SOS boosted module at the 1-st level of the decoder. (f) is the enhanced feature $\tilde{j}^1$ by feeding the boosted feature in (e) to the DFF at the 1-st level of the decoder $\mathcal{D}_{de}^1$.

module successfully suppresses the gridding artifacts in Figure B(e), as the structural and spatial information can be transferred from the strengthened feature $(i^1 + (\tilde{j}^2) \uparrow_2)$. The boosted feature j^1 in Figure B(e) will be fed to the DFF module of the decoder for further enhancement.

The effectiveness of the DFF module can also be demonstrated via the visualization of the learned features. The DFF module of the encoder, $\mathcal{D}_{en}^2$, generates the enhanced feature $\tilde{i}^2$ with finer structures (e.g., the backrests of the

Table B. **Quantitative evaluations on the outdoor and indoor sets of the SOTS dataset.** ⋆ denotes the method that is trained on the RESIDE dataset with the same settings as mentioned in the manuscript. **Red** and blue texts indicate the best and the second-best performance respectively.

Methods		DCP [3]	NLD [1]	AODNet [6]	MSCNN [13]	MsPPN* [19]	DcGAN [8]	GFN* [14]	GCANet* [2]	PFFNet* [12]	GDN [10]	DuRN* [11]	Ours*
SOTS Outdoor	PSNR	17.54	18.09	19.74	18.06	30.56	24.85	25.02	26.20	31.11	30.86	33.14	**34.81**
	SSIM	0.848	0.772	0.874	0.821	0.971	0.916	0.916	0.930	0.968	0.982	0.984	**0.986**
SOTS Indoor	PSNR	19.96	16.45	17.86	17.08	29.32	25.89	23.20	30.06	27.32	32.16	30.70	**32.77**
	SSIM	0.870	0.728	0.794	0.801	0.945	0.918	0.882	0.960	0.94	**0.984**	0.976	0.982

chairs in Figure B(c)) by remedying the missing spatial information from high-resolution features. The DFF module of the decoder, $\mathcal{D}_{de}^1$, extracts features from the non-adjacent levels and generates the enhanced feature $\tilde{j}^1$ in Figure B(f). Compared with the input feature j^1, the enhanced feature $\tilde{j}^1$ contains more and clearer details.

D. Differences with ResNet [4] and DenseNet [5].

ResNet [4] and DenseNet [5] are two basic frameworks which have been widely applied to various vision tasks. Instead of simply stacking these two frameworks into U-Net [16], the proposed method is motivated by the SOS boosting algorithm and back-projection technique. The proposed SOS boosted module works differently (in a signal strengthening manner), compared to the residual block. We present theoretical justifications in Section A to the proposed boosted module in the context of dehazing tasks and show that it outperforms the alternative that simply combines U-Net and residual learning (Figure 2 (b) of the manuscript). Furthermore, the DFF module is different from the DenseNet, where we propose a new error feedback mechanism (Figure 3 of the manuscript), replacing the bottleneck layer (concatenation layer and convolutional layer) of the DenseNet, to better fuse multi-scale information. The results in Table 4 of the manuscript (MSBDN-DFF vs. MSBDN-S) show that the error feedback mechanism is more effective (by a large margin 0.77dB) for multi-scale feature fusion than the bottleneck layer. The justifications and experimental results in the manuscript and supplementary material demonstrate that meticulous algorithmic design and integration of components in this work are critical in achieving the state-of-the-art results.

E. Efficiency analysis.

Table C. **Results of the efficiency analysis.** The MSBDN-DFF-S, including four feature levels ($L = 4$) and eight ResBlocks in the feature restoration module G_{Res} ($B = 8$), is a lightweight version of the proposed MSBDN-DFF ($L = 5$ and $B = 18$). ⋆ denotes the method that is trained on the RESIDE dataset with the same settings as mentioned in the manuscript. The testing environment is on a machine with an NVIDIA 2080Ti GPU. **Red** and blue texts indicate the best and the second-best performance respectively.

Method	DCPDN* [18]	GDN [10]	DuRN* [11]	MSBDN-DFF-S*	MSBDN-DFF*
Param	67M	1M	9M	4.5M	31M
Memory	3679M	7099M	4563M	3445M	3885M
Time	0.699s	0.618s	1.14s	0.346s	0.491s
PSNR	28.96	31.51	31.92	32.07	**33.79**

We further evaluate the proposed method against several state-of-the-art methods (DCPDN [18], GDN [10], and DuRN [11]) on 2K images (2560×1440) in terms of the numbers of parameters, GPU memory consumption, and median values of the inference time, as shown in Table C. Their performances on the SOTS datasets are also reported for reference. Overall, the proposed method obtains the best performance and has a faster inference speed than other methods.

Table D. **Quantitative evaluation of the non-uniform A and noisy T.** For evaluation, we select 100 clean images from the outdoor set of the SOTS dataset [7], and synthesize four hazy image datasets using (1) under different settings (global constant A, channel-wise constant A, pixel-wise constant A, and global constant A with noisy T). These four datasets are referred to as Global A, Channel-wise A, Pixel-wise A, and Noisy T. **Red** texts indicate the best performance.

Haze formulation		DCP [3]	GFN [14]	PFFNet [12]	DuRN [11]	Ours
Global A	PSNR	17.55	23.22	29.06	30.79	**32.30**
	SSIM	0.866	0.907	0.958	0.962	**0.971**
Channel-wise A	PSNR	16.93	22.88	28.32	30.10	**31.53**
	SSIM	0.863	0.901	0.953	0.954	**0.968**
Pixel-wise A	PSNR	17.49	23.16	28.62	30.02	**31.75**
	SSIM	0.867	0.907	0.958	0.962	**0.971**
Noisy T	PSNR	16.90	22.75	28.74	30.24	**31.79**
	SSIM	0.865	0.900	0.956	0.954	**0.967**

F. Analysis of the non-uniform A and noisy T.

We also evaluate the proposed method and four dehazing algorithms [3, 14, 12, 11] on the synthetic hazy images when the atmospheric light A is not a global (channel-wise [17] and pixel-wise [9]) constant and the transmission map T contains noise. All the evaluated deep learning-based models are trained on the RESIDE dataset with the same settings as mentioned in the manuscript.

The evaluation results are shown in Table D. The proposed model consistently outperforms other models on all the synthetic datasets. This indicates that the proposed dehazing method is robust to scenarios with non-uniform A and noisy T.

G. More quantitative and qualitative results.

In the manuscript, we have shown that the proposed algorithm performs favorably against state-of-the-art methods on the whole SOTS dataset (including indoor and outdoor images). In the supplemental material, we present the quantitative results on indoor images and quantitative results on outdoor images separately in Table B. Moreover, we provide more qualitative comparisons on synthetic datasets (in Figures C-F) and real-world images (in Figures G-H).

(a1) Ground-truth   (b1) Hazy input   (c1) DCP [3]   (d1) AOD [6]

(e1) GFN [14]   (f1) GCANet [2]   (g1) DuRN [11]   (h1) Ours

(a2) Ground-truth   (b2) Hazy input   (c2) DCP [3]   (d2) AOD [6]

(e2) GFN [14]   (f2) GCANet [2]   (g2) DuRN [11]   (h2) Ours
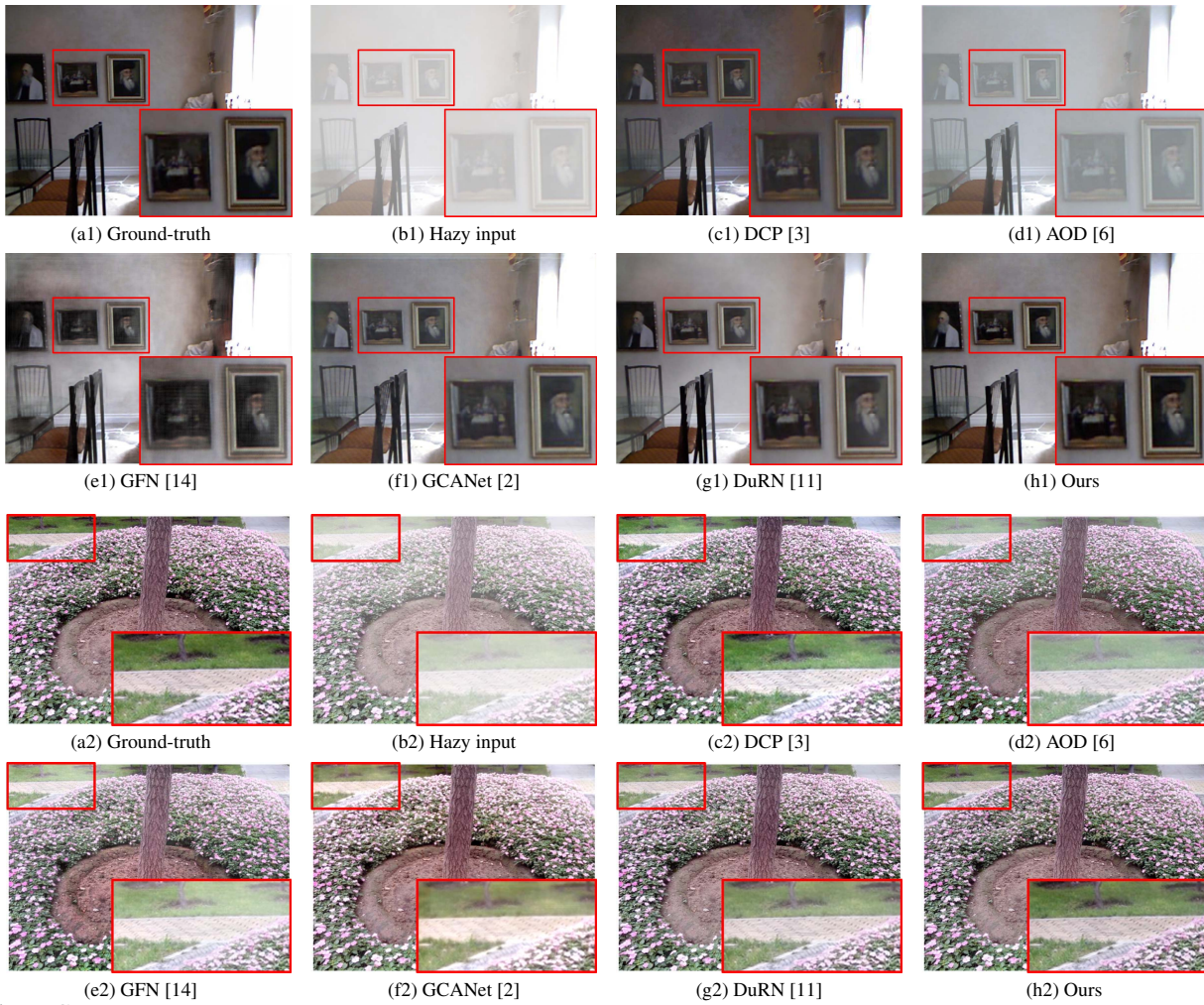
Figure C. **More visual results on the SOTS dataset.** The results in (c1)-(g1) and (c2)-(g2) contain some color distortions and haze residual, while the dehazed image in (h1) and (h2) by our method is much clearer. Best viewed on a high-resolution display.

(a1) Ground-truth

(b1) Hazy input

(c1) DCP [3]

(d1) AOD [6]

(e1) GFN [14]

(f1) GCANet [2]

(g1) DuRN [11]

(h1) Ours

(a2) Ground-truth

(b2) Hazy input

(c2) DCP [3]

(d2) AOD [6]

(e2) GFN [14]

(f2) GCANet [2]

(g2) DuRN [11]

(h2) Ours

Figure D. **Visual results on the HazeRD dataset.** Other dehazing methods fail to remove the haze in the region of the red box, while our method is able to generate clearer images from realistic images with large portion of haze. Best viewed on a high-resolution display.

(a) Ground-truth   (b) Hazy input   (c) DCP [3]   (d) GFN [14]

(e) PFFNet [12]   (f) DuRN [11]   (g) MsPPN [19]   (h) Ours

Figure E. **Visual results on the NTIRE2018-Dehazing dataset.** Our method is able to generate a clearer image with more details from the realistic indoor image. Best viewed on a high-resolution display.

(a) Ground-truth   (b) Hazy input   (c) DCP [3]   (d) GFN [14]

(e) PFFNet [12]   (f) DuRN [11]   (g) MsPPN [19]   (h) Ours
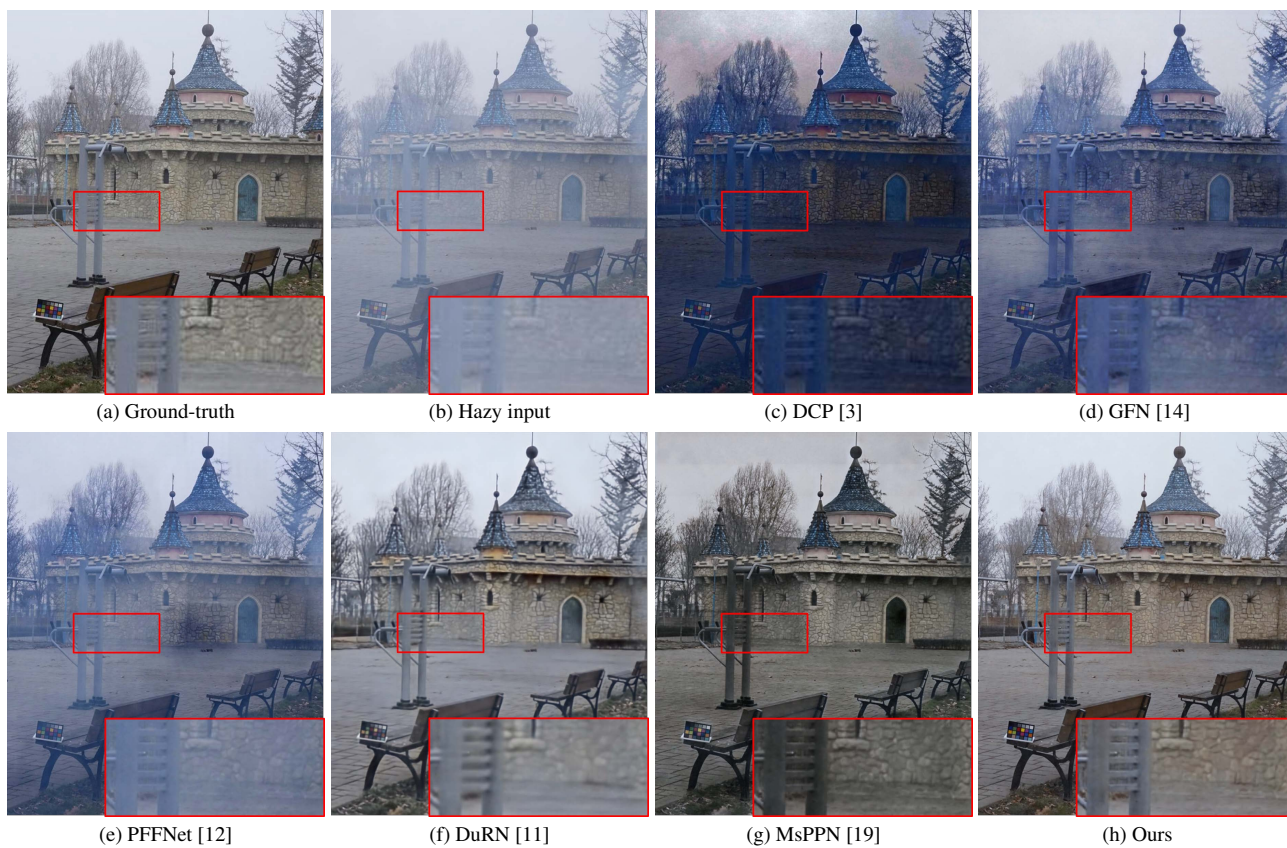
Figure F. **Visual results on the NTIRE2018-Dehazing dataset.** Our method is able to generate a clearer image with less color distortions from the realistic outdoor image. Best viewed on a high-resolution display.

(a1) Hazy input (b1) DCP [3] (c1) GFN [14]

(d1) PFFNet [12] (e1) DuRN [11] (f1) Ours

(a2) Hazy input (b2) DCP [3] (c2) GFN [14]

(d2) PFFNet [12] (e2) DuRN [11] (f2) Ours

Figure G. **More visual results on the real-world images.** The dehazed images by the DCP [3] and GFN [14] contain significant color distortions. The DuRN [11] and PFFNet [12] fail to remove the haze. The proposed method generates clearer images with less color distortions. Best viewed on a high-resolution display.

(a1) Hazy input   (b1) DCP [3]   (c1) GFN [14]

(d1) PFFNet [12]   (e1) DuRN [11]   (f1) Ours

(a2) Hazy input   (b2) DCP [3]   (c2) GFN [14]

(d2) PFFNet [12]   (e2) DuRN [11]   (f2) Ours

Figure H. **More visual results on the real-world images.** The proposed method generates dehazed images with clearer foreground and background. Best viewed on a high-resolution display.

References

[1] Dana Berman, Shai Avidan, et al. Non-local image dehazing. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1674–1682, 2016.

[2] Dongdong Chen, Mingming He, Qingnan Fan, Jing Liao, Liheng Zhang, Dongdong Hou, Lu Yuan, and Gang Hua. Gated context aggregation network for image dehazing and deraining. In *IEEE Winter Conference on Applications of Computer Vision*, pages 1375–1383, 2019.

[3] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2011.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[5] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017.

[6] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *IEEE International Conference on Computer Vision*, pages 4770–4778, 2017.

[7] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Reside: A benchmark for single image dehazing. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018.

[8] Runde Li, Jinshan Pan, Zechao Li, and Jinhui Tang. Single image dehazing via conditional generative adversarial network. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 8202–8211, 2018.

[9] Yu Li, Robby T Tan, and Michael S Brown. Nighttime haze removal with glow and multiple light colors. In *IEEE International Conference on Computer Vision*, pages 226–234, 2015.

[10] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *IEEE International Conference on Computer Vision*, pages 7314–7323, 2019.

[11] Xing Liu, Masanori Suganuma, Zhun Sun, and Takayuki Okatani. Dual residual networks leveraging the potential of paired operations for image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 7007–7016, 2019.

[12] 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.

[13] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *European Conference on Computer Vision*, pages 154–169, 2016.

[14] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3253–3261, 2018.

[15] Yaniv Romano and Michael Elad. Boosting of image denoising algorithms. *Siam Journal on Imaging Sciences*, 8(2):1187–1219, 2015.

[16] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, pages 234–241, 2015.

[17] Matan Sulami, Itamar Glatzer, Raanan Fattal, and Mike Werman. Automatic recovery of the atmospheric light in hazy images. In *IEEE International Conference on Computational Photography*, pages 1–11, 2014.

[18] He Zhang and Vishal M Patel. Densely connected pyramid dehazing network. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3194–3203, 2018.

[19] He Zhang, Vishwanath Sindagi, and Vishal M Patel. Multi-scale single image dehazing using perceptual pyramid deep network. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 902–911, 2018.