Frequency Domain-Based Diffusion Model for Unpaired Image Dehazing

Supplementary Material

In this supplementary material, Sec. A1 provides the motivation analysis of the reconstructed amplitude in the frequency domain. Sec. A2 illustrates the detailed architecture of the dehazing network and analyzes its efficiency. Sec. A3 describes the detailed training and inference algorithms. Sec. A4 highlights the advantages of using the unpaired data training paradigm in this work. Sec. A5 and Sec. A6 describe the dataset used and the training details, respectively. Sec. A7 analyzes the limitations of our method. Finally, Sec. A8 shows more quantitative and qualitative comparison results.

A1. Motivation Analysis

As described in Sec. 1 of the main paper, the illumination contrast is manifested by the amplitude spectrum, while the texture structure information is manifested by the phase spectrum. Haze mainly affects the illumination contrast, while the structural information is immune to haze degradation. That is to say, that haze degradation is transferred with the exchange of the amplitude spectrum. Therefore, the key motivation of our FrDiff is to make the diffusion model (DM) reconstruct the amplitude of haze-free images in the unpaired training setting.

To justify this motivation, we follow the assumption of the classical dark channel prior [19], and analyze it by studying the statistical properties of the dark channel before and after the exchange of the amplitude spectrum. Specifically, as shown in Fig. 2(a) of the main paper, we use the FFT and IFFT operations to replace the amplitude spectrum of the haze images (*i.e.*, Hazy) with the amplitude spectrum of the clear images (*i.e.*, Clear) on the SOTS-Indoor [28] dataset. By doing so, we can obtain a set of synthetic clear images (*i.e.*, SynClear). Then, we follow [28] using the patch with the size of 15×15 to calculate their dark channels.

Fig. A1 shows several example images and the corresponding dark channels. Figs. A1(a), (c), and (e) are the Hazy images, the Clear images, and the SynClear images, respectively. Figs. A1(b), (d), and (f) are their corresponding dark channels. Visually, the intensity of the dark channel is a rough approximation of the thickness of the haze. The dark channel of a Hazy image will have higher intensity in regions with denser haze (see Fig. A1(b)). As the amplitude spectrum of the Hazy image is replaced with the amplitude spectrum of the Clear image, the dark channels of the SynClear image are closer to the dark channels of the Clear image (see Figs. A1(d) and (f)). This proves that the data properties of the synthetic clear image are closer to those of the clear image.

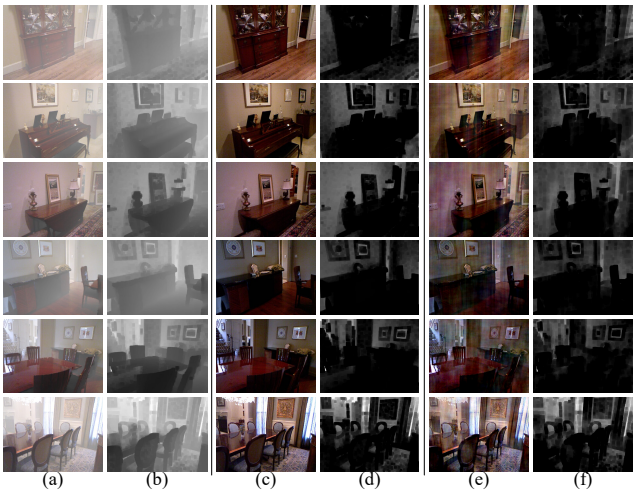

Figure A1. Example images in SOTS-Indoor [28] dataset. (a), (c), and (e) are the haze images (*i.e.*, Hazy), the clear images (*i.e.*, Clear), and the synthetic clear images (*i.e.*, SynClear), respectively. (b), (d), and (f) are their corresponding dark channels. The dark channel of the SynClear image is more similar to the Clear image indicating that the replacement of the amplitude spectrum is effective in removing the haze.

In addition, Figs. A2(a), (b), and (c) are the histogram of the pixel intensities in all of the dark channels of the haze images (*i.e.*, Hazy), the clear images (*i.e.*, Clear), and the synthetic clear images (*i.e.*, SynClear) in the dataset, respectively. Figs. A2(d), (e), and (f) are the corresponding histograms of the average intensity of each dark channel (each bin stands for 16 intensity levels). We have the following observations:

- 1) The intensity of the pixels in the dark channel of the Hazy images mostly falls between 50-200 (see Fig. A2(a)). In contrast, the Clear and SynClear images have mostly zero values and about 75 percent of the pixels are below 25 (see Figs. A2(b) and (c)).
- 2) The average intensity of each dark channel of the Hazy images also mostly falls between 100-150 (see Fig. A2(d)), while the average intensity of each dark channel for both Clear and SynClear images is below 100 (see Figs. A2(e) and (f)).

In general, this statistic not only provides very strong support for the dark channel prior [19], but also demonstrates that the illumination contrast (*e.g.*, Haze) is manifested by the amplitude spectrum, and replacing the amplitude spectrum from the clear image can effectively remove the haze.

Therefore, this unique frequency domain property pro-

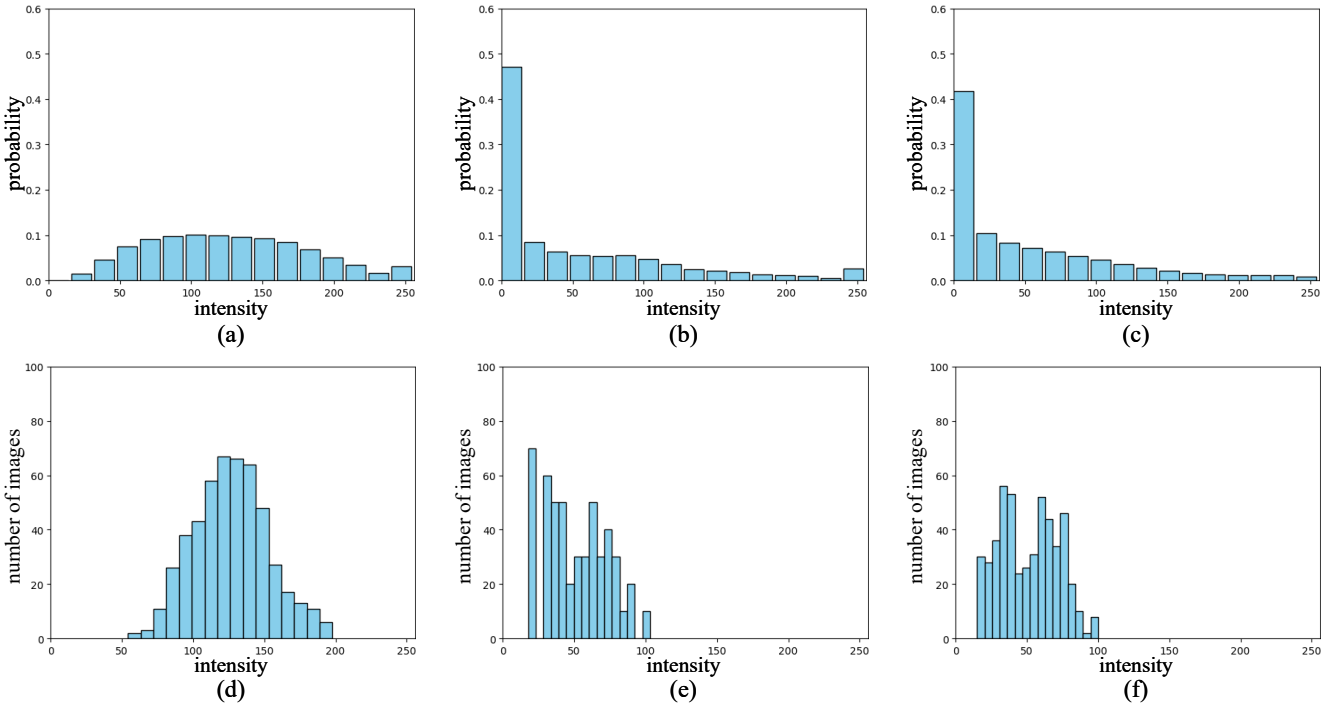

Figure A2. Statistics of the dark channels. (a), (b), and (c) are the histogram of the pixel intensities in all of the dark channels of the haze images (*i.e.*, Hazy), the clear images (*i.e.*, Clear), and the synthetic clear images (*i.e.*, SynClear) in the dataset, respectively. (d), (e), and (f) are the corresponding histogram of the average intensity of each dark channel (each bin stands for 16 intensity levels). The dark channel statistics indicate that the synthetic clear image obtained by replacing the amplitude spectrum is closer to the clear image distribution.

vides inspiration for haze removal. Our key insight of FrDiff, *i.e.*, allowing DM to learn the amplitude of haze-free images in the unpaired training setting, is justified and has great potential.

A2. Architecture Details

As described in Sec. 3.1 of the main paper. To ensure the non-local feature capture capability and the reconstruction capability of the dehazing process, we adopt NAFNet [6], a simple UNet-based architecture, as our dehazing network.

Specifically, we illustrate the detailed architecture of the lightweight U-Net as shown in Fig. A3. The entire network is based on the U-Net architecture. We follow the existing approach [6] to extract features by stacking some NAFNet's Blocks on each scale, where the number of basic blocks is marked. In each scale, we also incorporate the Frequency Compensation Layer (see Fig. 4(a) of the main paper) in the contracting path (up-to-down) of UNet and input amplitude residual z into these layers. The dehazed image is outputted through the expanding path (down to up) of UNet. The parameters of the dehazing network are 8.69M.

In addition, we use a neural network consisting of 5 stacked ResBlocks, denoted as ϵ_θ, to estimate the noise. The purpose of using ResBlocks as the denoising network is to ensure the same resolution of inputs and outputs while

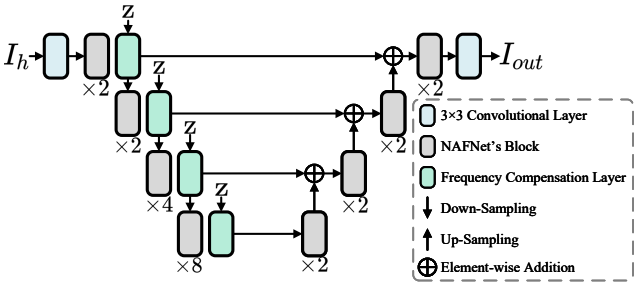

Figure A3. Network structure of dehazing network.

minimizing the model parameters. The parameters of the denoising network are 0.07M and the FLOPs required per iteration is about 4.5G.

On 3090 GPU with 256 × 256 input, the runtime is 0.04s. Future we'll further accelerate the denoising process by one-step distillation.

A3. Algorithm

The first and second stage training algorithms for FrDiff are shown in Alg. A1 and Alg. A2, respectively. The inference algorithm for FrDiff is shown in Alg. A3.

Algorithm A1 FrDiff Training: Stage One

Input: ARE, dehazing network.
Output: Trained dehazing network.

for I_h, I_c **do**
  $\mathrm{z} = \mathrm{ARE}(I_h, I_c)$. (paper Eqs. (1)-(3))
  $I_{out} = \mathrm{DehazingNetwork}(I_h, \mathrm{z})$
  Calculate $\mathcal{L}_{s1}$ loss (paper Eq. (6)).
end for
Output the trained dehazing network.

A4. Advantages of Unpaired Training

The field of unsupervised training for image dehazing can be divided into pseudo-label-based approaches [8, 9, 17, 47, 57] and unpaired training based approaches [7, 15, 33, 40, 41, 53, 54, 60].

The pseudo-label-based approaches [8, 9, 17, 47, 57] focus on mining and modeling the characteristics of real haze, and usually improves its performance in real-world haze scenarios by introducing an additional re-hazing pipeline. However, the design of a reasonable pipeline critically hinges on domain expertise. Without expert insights, the performance of such methods may falter due to the inherent challenges in capturing domain-specific characteristics. In contrast, unpaired training-based approaches [7, 15, 33, 40, 41, 53, 54, 60] aims to directly learn the mapping of the haze domain to the clear domain through tailored training strategies, and thus more concise and promising.

A5. More Dataset Details

We evaluate our method on widely-used RESIDE [28], I-HAZE [1], and Fattal's [18] datasets, which cover synthetic, artificial, and real-world images. Specifically, the RESIDE [28] dataset contains several subsets: (a) ITS, which contains 13,990 synthetic indoor hazy-clear pairs. (b) SOTS-Indoor and SOTS-Outdoor, which contain 500 synthetic indoor/outdoor hazy-clear pairs. (c) HSTS-Synth and HSTS-Real, which contain 10 synthetic hazy-clear pairs and 10 real-world hazy image without ground-truth images, respectively. (d) URHI, which contains over 4,000 real hazy image without ground-truth images. I-HAZE [1] dataset contains 35 artificial hazy-clear pairs produced by professional haze generators. Fattal's [18] dataset includes 31 real-world hazy images in various scenes.

For fair comparisons, we follow the previous works [54, 60] to use ITS dataset from RESIDE as the training set. We validate the performance of FrDiff on synthetic data using the SOTS-Indoor, SOTS-Outdoor, and HSTS-Synth datasets with ground truth. We validate the performance of FrDiff on real-world data using the HSTS-Real, Fattal's, and URHI datasets without ground truth.

Algorithm A2 FrDiff Training: Stage Two

Input: ARE, trained dehazing network,, denoising network, $\beta_t (t \in [1, T])$.
Output: Trained denoising network, trained dehazing network.

Init: $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_T = \prod_{i=0}^{T} \alpha_i$.
Init: The dehazing network copies the parameters of trained dehazing network.
for I_h, I_c **do**
  $\mathrm{z} = \mathrm{ARE}(I_h, I_c)$. (paper Eqs. (1)-(3))
  Diffusion Process:
  We sample z_T by $q(\mathrm{z}_T \mid \mathrm{z}) = \mathcal{N}(\mathrm{z}_T; \sqrt{\bar{\alpha}_T}\mathrm{z}, (1 - \bar{\alpha}_T)\mathbf{I})$ (paper Eq. (7))
  Denoising Process:
  $\hat{\mathrm{z}}_T = \mathrm{z}_T$
  $A_h = \mathrm{FFT}(\text{Conv-Block}(I_h))$
  for $t = T$ to 1 **do**
    $\hat{\mathrm{z}}_{t-1} = \frac{1}{\sqrt{\alpha_t}}(\hat{\mathrm{z}}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\hat{\mathrm{z}}_t, A_h, t)) + \sqrt{1-\alpha_t}\epsilon_t$ (paper Eq. (10))
  end for
  $\hat{\mathrm{z}} = \hat{\mathrm{z}}_0$
  $I_{out} = \mathrm{DehazingNetwork}(I_h, \hat{\mathrm{z}})$
  Calculate $\mathcal{L}_{s2}$ loss (paper Eq. (11)).
end for
Output the trained denoising network and trained dehazing network.

Algorithm A3 FrDiff Inference

Input: Trained denoising network, trained dehazing network, $\beta_t (t \in [1, T])$, hazy images I_h.
Output: Dehazed images I_{out}.

Init: $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_T = \prod_{i=0}^{T} \alpha_i$.
Denoising Process:
Sample $\mathrm{z}_T \sim \mathcal{N}(0, 1)$
$\hat{\mathrm{z}}_T = \mathrm{z}_T$
$A_h = \mathrm{FFT}(\text{Conv-Block}(I_h))$
for $t = T$ to 1 **do**
  $\hat{\mathrm{z}}_{t-1} = \frac{1}{\sqrt{\alpha_t}}(\hat{\mathrm{z}}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\hat{\mathrm{z}}_t, A_h, t)) + \sqrt{1-\alpha_t}\epsilon_t$ (paper Eq. (10))
end for
$\hat{\mathrm{z}} = \hat{\mathrm{z}}_0$
$I_{out} = \mathrm{DehazingNetwork}(I_h, \hat{\mathrm{z}})$
Output dehazed images I_{out}.

A6. More Training Details

During training, we follow existing works [10, 59] to adopt a two-stage strategy to optimize our model.

As shown in Fig. 3(a) of the main paper, in stage one, we utilize the proposed amplitude residual encoder (ARE) to obtain amplitude residual z, and train the dehazing net-

work for haze removal. In this case, the z is directly fed into the dehazing network without involving the diffusion and denoising processes. It is infeasible to supervise model training with the strict pixel-level loss function under unpaired data setting. Therefore, we follow existing works [40, 53, 54], using the same adversarial loss $\mathcal{L}_{GAN}$ and patchwise contrast loss $\mathcal{L}_{PatchNCE}$ to encourage the dehazed results have the same domain distribution as the clear images, and the hyper-parameters λ_{GAN} and $\lambda_{PatchNCE}$ are all set to 1.

As shown in Fig. 3(b) of the main paper, in stage two, we joint train diffusion model and dehazing network from stage one to reconstruct the amplitude residual for enhancing the dehazing process. In this stage, the z first adds noise to output z_T through the diffusion process, and then reconstructs the amplitude residual $\hat{z}$ through multiple denoising processes. We additionally include the diffusion loss $\mathcal{L}_{diff}$ based on the $\mathcal{L}_{GAN}$ and $\mathcal{L}_{PatchNCE}$, and the hyper-parameter λ_{diff} is set to 1.

In both training stages, we use Adam [26] optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$, learning rate is 1×10^{-4}. The epoch number is 200. The batch size is 8 and the input patch size is 256×256 and augments the data with random horizontal and vertical flips. The input size of 256×256 is randomly cropped from all training images in an unpaired learning procedure.

A7. Limitation

Although our FrDiff can reconstruct frequency domain features well for handling haze, when the resolution of the input image is larger, the resolution of the amplitude residuals reconstructed by the diffusion model also increases. This means that the computational effort of the diffusion model will increase. Therefore, it is expected to make the diffusion model learn a set with a fixed number of amplitude spectral priors to reconstruct the frequency domain features so as to avoid increasing computational costs significantly.

In addition, removing the haze by reconstructing the spectrum ignores the local spatial differences of the haze, and the future promises to further enable the model to handle spatially varying haze of different thicknesses.

A8. More Results

In this section, we first provide more quantitative results and spectral visualisations to validate the effectiveness of FrDiff. Then, we analyze the sensitivity of the hyper-parameters in the loss function. Finally, we show more visualization results.

Results on O-HAZE [2] dataset. To further demonstrate the generalization ability of our FrDiff, we add comparisons on O-HAZE [2] in Tab. A1. Experimental results show that FrDiff achieves higher performance compared to prevailing unpaired data-based training methods.

Methods	D^4[60]	UCL-Dehaze[53]	ODCR[54]	FrDiff
PSNR	16.92	15.57	17.46	18.22
SSIM	0.607	0.566	0.632	0.648
FADE	1.358	1.715	0.970	0.744
MUSIQ	63.042	62.432	63.388	63.981

Table A1. Quantitative comparison with SOTA methods on O-HAZE [2] dataset.

Spectral Visualizations To prove the reliability of FrDiff, we present the spectral visualizations in Fig. A4. FrDiff can effectively reconstruct low-frequency illuminance (red circle) and eliminate the high-frequency spectral bias (red arrow). In addition, we measure the L2 similarity(↓) of the frequency features with GT to show the superiority (ODCR:6.257, FrDiff:2.156).

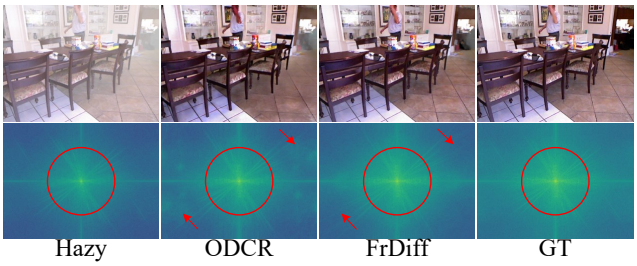

Figure A4. Visual results on SOTS-Indoor [28] and SOTS-Outdoor [28] datasets. Zoom in to see better visualization.

Effect of Hyper-parameters. To explore the influence of hyper-parameters used in Eq. (11). We discuss the different λ_{GAN}, $\lambda_{PatchNCE}$, and λ_{diff} as shown in Tab. A2. It is worth noting that each time we let only one hyper-parameter change and set the remaining two to 1. Proper λ_{GAN} can provide effective supervision of model training. The impact of $\lambda_{PatchNCE}$ is insignificant since the haze-unrelated information is introduced at the same time. The performance is positively correlated with the λ_{diff}, demonstrating the ability of the denoising network to reconstruct magnitude residuals. However, a larger λ_{diff} may lead to instability in training. After a trade-off between performance and training stability, we choose 1 as the value of λ_{GAN}, $\lambda_{PatchNCE}$, and λ_{diff}.

More Visual Results. To further verify the effectiveness of our method, we show more comparison results among the proposed FrDiff and other advanced methods on these benchmarks. The results on **SOTS-Indoor** [28] dataset

λ_{GAN}	PSNR	$\lambda_{PatchNCE}$	PSNR	λ_{diff}	PSNR
0.1	36.39	0.1	36.43	0.1	36.31
1	36.54	1	36.54	1	36.54
10	36.53	10	36.50	10	36.58

Table A2. Results of different λ_{GAN}, $\lambda_{PatchNCE}$, and λ_{diff} on SOTS-Indoor [28] dataset.

are shown in Figs. A5 and A6. The results on **SOTS-Outdoor** [28] dataset are shown in Figs. A7 and A8. The results on **HSTS-Synth** [28], **Fattle's** [18], and **URHI** [28] dataset are shown in Fig. A9, Fig. A10, and Fig. A11, respectively.

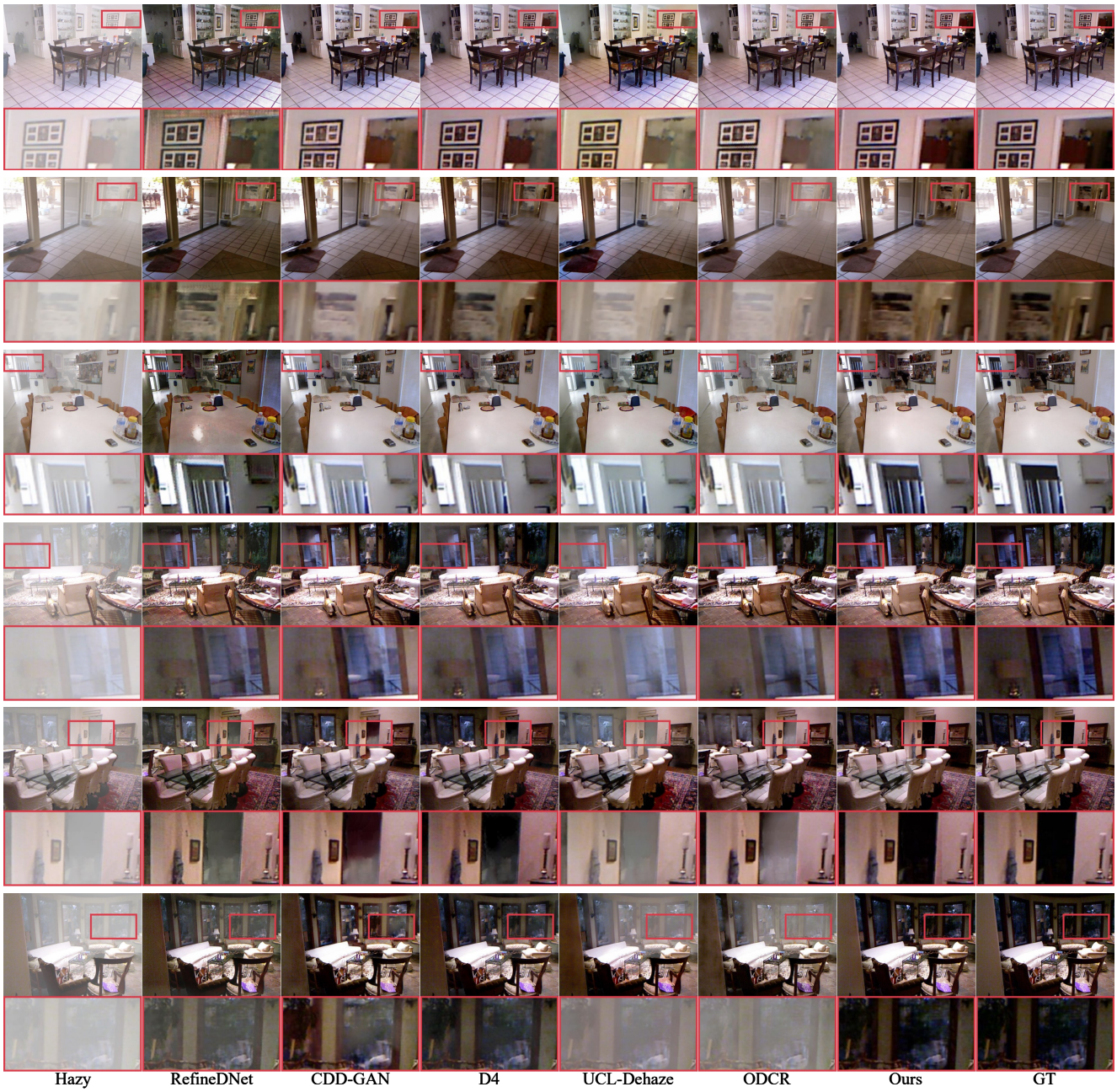

Figure A5. Visual results on SOTS-Indoor [28] dataset. The method is shown at the bottom of each case. Zoom in to see better visualization.

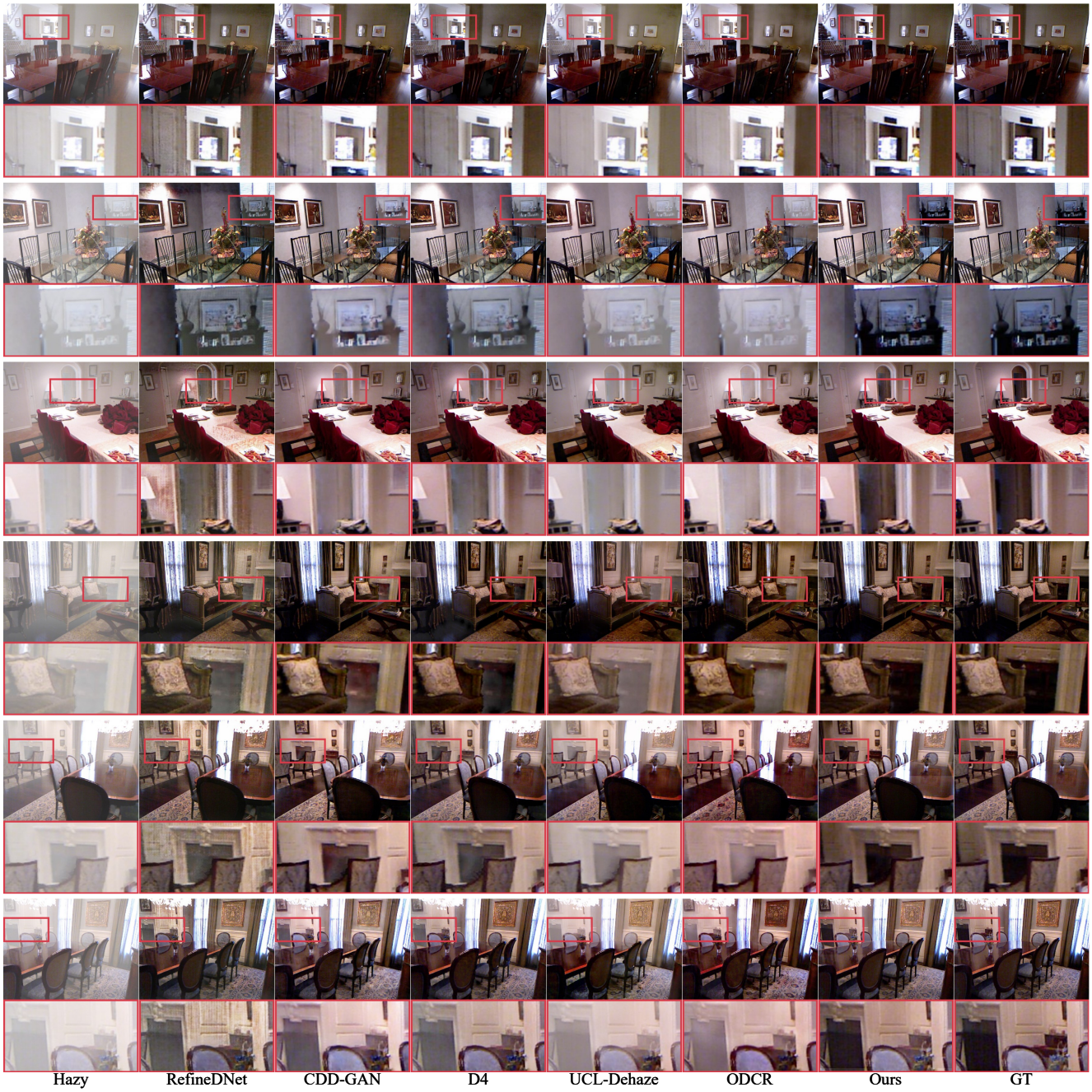

Figure A6. Visual results on SOTS-Indoor [28] dataset. The method is shown at the bottom of each case. Zoom in to see better visualization.

Figure A7. Visual results on SOTS-Outdoor [28] dataset. The method is shown at the bottom of each case. Zoom in to see better visualization.

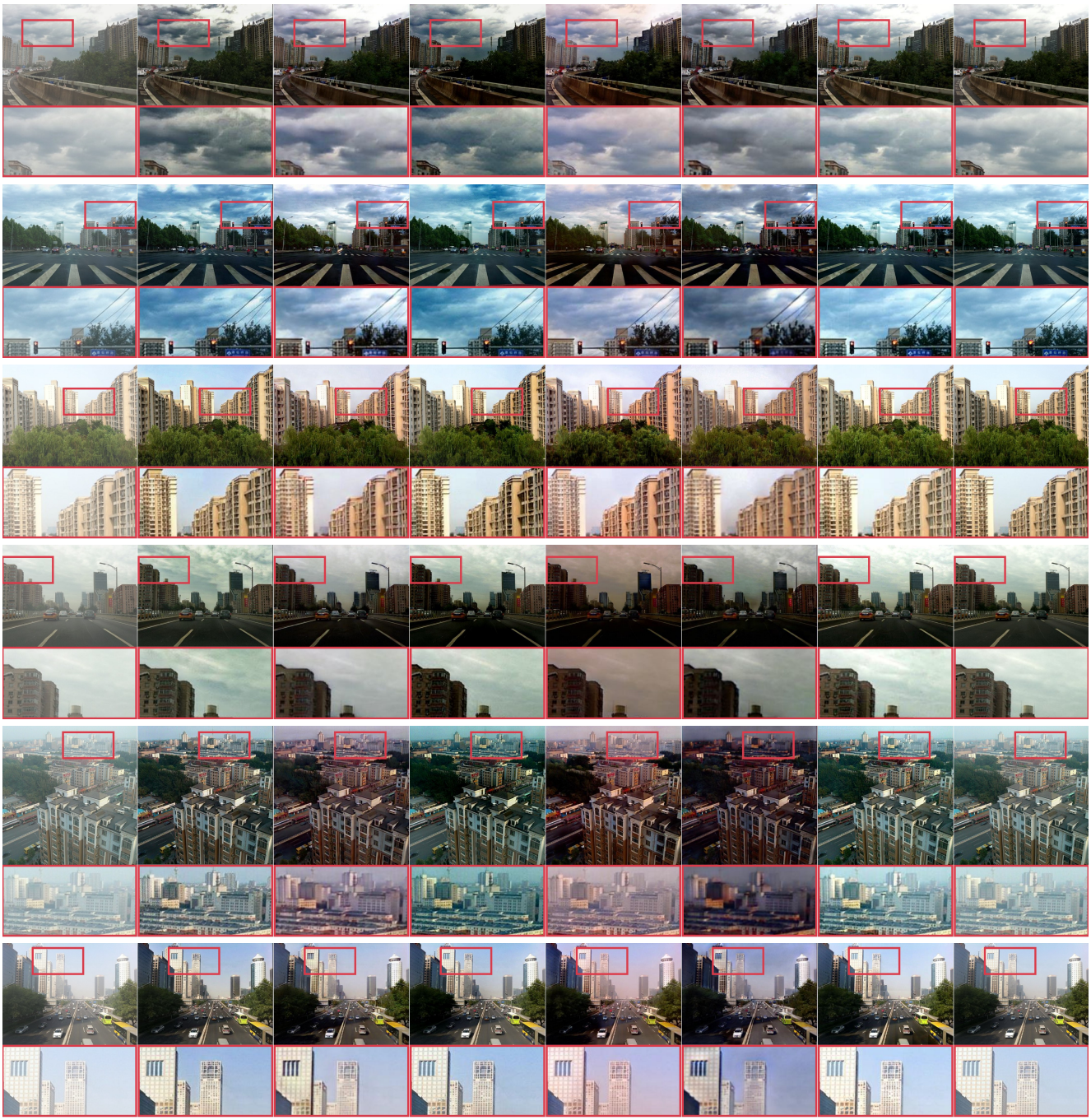

Figure A8. Visual results on SOTS-Outdoor [28] dataset. The method is shown at the bottom of each case. Zoom in to see better visualization.

Figure A9. Visual results on HSTS-Synth [28] dataset. The method is shown at the bottom of each case. Zoom in to see better visualization.

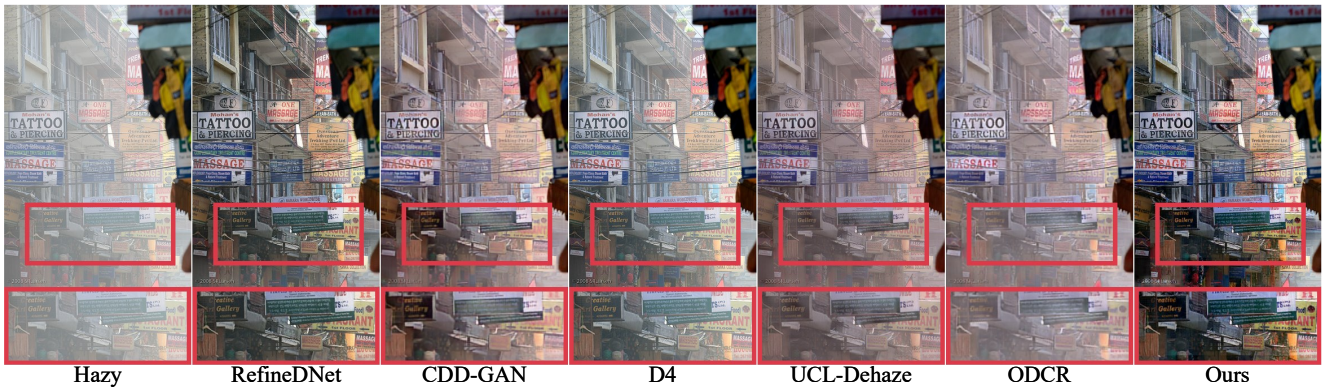

Figure A10. Visual results on Fattle's [18] dataset. The method is shown at the bottom of each case. Zoom in to see better visualization.

Figure A11. Visual results on URHI [28] dataset. The method is shown at the bottom of each case. Zoom in to see better visualization.

References

[1] Cosmin Ancuti, Codruta O Ancuti, Radu Timofte, and Christophe De Vleeschouwer. I-HAZE: A dehazing benchmark with real hazy and haze-free indoor images. In *ACIVS*, pages 620–631. Springer, 2018. 5, 6, 3

[2] Codruta O Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. O-HAZE: a dehazing benchmark with real hazy and haze-free outdoor images. In *CVPRW*, pages 754–762, 2018. 4

[3] Dana Berman, Shai Avidan, et al. Non-local image dehazing. In *CVPR*, pages 1674–1682, 2016. 1

[4] Dana Berman, Tali Treibitz, and Shai Avidan. Single image dehazing using haze-lines. *IEEE TPAMI*, 42(3):720–734, 2018. 1, 2

[5] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. DehazeNet: An end-to-end system for single image haze removal. *IEEE TIP*, 25(11):5187–5198, 2016. 2

[6] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *ECCV*, pages 17–33. Springer, 2022. 3, 2

[7] Xiang Chen, Zhentao Fan, Pengpeng Li, Longgang Dai, Caihua Kong, Zhuoran Zheng, Yufeng Huang, and Yufeng Li. Unpaired deep image dehazing using contrastive disentanglement learning. In *ECCV*, pages 632–648. Springer, 2022. 1, 3, 6

[8] Zeyuan Chen, Yangchao Wang, Yang Yang, and Dong Liu. PSD: Principled synthetic-to-real dehazing guided by physical priors. In *CVPR*, pages 7180–7189, 2021. 3, 6

[9] Zixuan Chen, Zewei He, Ziqian Lu, Xuecheng Sun, and Zhe-Ming Lu. Prompt-based test-time real image dehazing: a novel pipeline. In *ECCV*, pages 432–449. Springer, 2024. 3, 6

[10] Zheng Chen, Yulun Zhang, Ding Liu, Jinjin Gu, Linghe Kong, Xin Yuan, et al. Hierarchical integration diffusion model for realistic image deblurring. *NeurIPS*, 36, 2024. 2, 3, 5

[11] Lark Kwon Choi, Jaehee You, and Alan Conrad Bovik. Referenceless prediction of perceptual fog density and perceptual image defogging. *IEEE TIP*, 24(11):3888–3901, 2015. 7

[12] Qili Deng, Ziling Huang, Chung-Chi Tsai, and Chia-Wen Lin. HardGAN: A haze-aware representation distillation gan for single image dehazing. In *ECCV*, pages 722–738. Springer, 2020. 2

[13] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In *CVPR*, pages 2157–2167, 2020. 1, 2

[14] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2020. 8

[15] Deniz Engin, Anil Genç, and Hazim Kemal Ekenel. Cycle-Dehaze: Enhanced cyclegan for single image dehazing. In *CVPRW*, pages 825–833, 2018. 1, 3

[16] Yuchen Fan, Jiahui Yu, Yiqun Mei, Yulun Zhang, Yun Fu, Ding Liu, and Thomas S Huang. Neural sparse representation for image restoration. *NeurIPS*, 33:15394–15404, 2020. 8

[17] Chengyu Fang, Chunming He, Fengyang Xiao, Yulun Zhang, Longxiang Tang, Yuelin Zhang, Kai Li, and Xiu Li. Real-world image dehazing with coherence-based pseudo labeling and cooperative unfolding network. *NeurIPS*, 37: 97859–97883, 2025. 3, 6

[18] Raanan Fattal. Dehazing using color-lines. *ACM TOG*, 34 (1):1–14, 2014. 1, 2, 5, 6, 7, 3, 10

[19] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE TPAMI*, 33(12): 2341–2353, 2010. 1, 2, 6

[20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 5

[21] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *NeurIPS*, 33:6840–6851, 2020. 3, 5

[22] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *CVPR*, pages 7132–7141, 2018. 8

[23] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. *NeurIPS*, 28, 2015. 8

[24] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *NeurIPS*, 35: 23593–23606, 2022. 2, 3

[25] Diederik P Kingma. Auto-encoding variational bayes. In *ICLR*, 2014. 5

[26] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6, 4

[27] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. AOD-Net: All-in-one dehazing network. In *ICCV*, pages 4770–4778, 2017. 2, 6

[28] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE TIP*, 28(1):492–505, 2018. 1, 5, 6, 7, 8, 3, 4, 9, 10

[29] Boyun Li, Yuanbiao Gou, Shuhang Gu, Jerry Zitao Liu, Joey Tianyi Zhou, and Xi Peng. You only look yourself: Unsupervised and untrained single image dehazing neural network. *IJCV*, 129:1754–1767, 2021. 6

[30] Chongyi Li, Chun-Le Guo, Man Zhou, Zhexin Liang, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Embedding fourier for ultra-high-definition low-light image enhancement. In *ICLR*, 2023. 1

[31] Chengyang Li, Heng Zhou, Yang Liu, Caidong Yang, Yongqiang Xie, Zhongbo Li, and Liping Zhu. Detection-friendly dehazing: Object detection in real-world hazy scenes. *IEEE TPAMI*, 45(7):8284–8295, 2023. 1

[32] Guangyuan Li, Chen Rao, Juncheng Mo, Zhanjie Zhang, Wei Xing, and Lei Zhao. Rethinking diffusion model for multi-contrast mri super-resolution. In *CVPR*, pages 11365–11374, 2024. 2, 3, 5

[33] Wei Liu, Xianxu Hou, Jiang Duan, and Guoping Qiu. End-to-end single image fog removal using enhanced cycle con-

sistent adversarial networks. *IEEE TIP*, 29:7819–7833, 2020. 1, 3

[34] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *ICCV*, pages 7314–7323, 2019. 1, 2

[35] Yang Liu, Jinshan Pan, Jimmy Ren, and Zhixun Su. Learning deep priors for image dehazing. In *ICCV*, pages 2492–2500, 2019. 2

[36] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *CVPR*, pages 11461–11471, 2022. 2, 3

[37] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE TIP*, 21(12):4695–4708, 2012. 7

[38] Srinivasa G Narasimhan and Shree K Nayar. Vision and the atmosphere. *IJCV*, 48:233–254, 2002. 1

[39] Alan V Oppenheim and Jae S Lim. The importance of phase in signals. *Proceedings of the IEEE*, 69(5):529–541, 1981. 1, 2, 4

[40] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *ECCV*, pages 319–345. Springer, 2020. 1, 3, 5, 6, 4

[41] Yuanjian Qiao, Mingwen Shao, Leiquan Wang, and Wangmeng Zuo. Learning depth-density priors for fourier-based unpaired image restoration. *IEEE TCSVT*, 2023. 1, 3

[42] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. FFA-Net: Feature fusion attention network for single image dehazing. In *AAAI*, pages 11908–11915, 2020. 1, 2, 6

[43] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In *CVPR*, pages 8160–8168, 2019. 6

[44] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *ECCV*, pages 154–169. Springer, 2016. 1, 2, 6

[45] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *CVPR*, pages 3253–3261, 2018. 2

[46] Christos Sakaridis, Dengxin Dai, Simon Hecker, and Luc Van Gool. Model adaptation with synthetic and real data for semantic dense foggy scene understanding. In *ECCV*, pages 687–704, 2018. 1

[47] Yuanjie Shao, Lerenhan Li, Wenqi Ren, Changxin Gao, and Nong Sang. Domain adaptation for image dehazing. In *CVPR*, pages 2808–2817, 2020. 2, 3

[48] Nikolay Skarbnik, Yehoshua Y Zeevi, and Chen Sagiv. *The importance of phase in image processing*. Technion-Israel Institute of Technology, Faculty of Electrical Engineering, 2009. 1, 2, 4

[49] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *ICLR*, 2021. 5

[50] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE TIP*, 32:1927–1941, 2023. 2

[51] Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *NeurIPS*, 30, 2017. 8

[52] A Vaswani. Attention is all you need. *NeurIPS*, 2017. 5, 8

[53] Yongzhen Wang, Xuefeng Yan, Fu Lee Wang, Haoran Xie, Wenhan Yang, Xiao-Ping Zhang, Jing Qin, and Mingqiang Wei. Ucl-dehaze: Towards real-world image dehazing via unsupervised contrastive learning. *IEEE TIP*, 2024. 1, 3, 5, 6, 4

[54] Zhongze Wang, Haitao Zhao, Jingchao Peng, Lujian Yao, and Kaijie Zhao. Odcr: Orthogonal decoupling contrastive regularization for unpaired image dehazing. In *CVPR*, pages 25479–25489, 2024. 1, 3, 5, 6, 4

[55] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *ECCV*, pages 3–19, 2018. 8

[56] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *CVPR*, pages 10551–10560, 2021. 2

[57] Rui-Qi Wu, Zheng-Peng Duan, Chun-Le Guo, Zhi Chai, and Chongyi Li. RIDCP: Revitalizing real image dehazing via high-quality codebook priors. In *CVPR*, pages 22282–22291, 2023. 2, 3, 6

[58] Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance discrimination. In *CVPR*, pages 3733–3742, 2018. 8

[59] Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. DiffIR: Efficient diffusion model for image restoration. In *ICCV*, pages 13095–13105, 2023. 2, 3, 5

[60] Yang Yang, Chaoyue Wang, Risheng Liu, Lin Zhang, Xiaojie Guo, and Dacheng Tao. Self-augmented unpaired image dehazing via density and depth decomposition. In *CVPR*, pages 2037–2046, 2022. 1, 3, 6, 4

[61] Tian Ye, Sixiang Chen, Wenhao Chai, Zhaohu Xing, Jing Qin, Ge Lin, and Lei Zhu. Learning diffusion texture priors for image restoration. In *CVPR*, pages 2524–2534, 2024. 2, 3

[62] Hu Yu, Naishan Zheng, Man Zhou, Jie Huang, Zeyu Xiao, and Feng Zhao. Frequency and spatial dual guidance for image dehazing. In *ECCV*, pages 181–198. Springer, 2022. 1, 3

[63] Hu Yu, Jie Huang, Kaiwen Zheng, and Feng Zhao. High-quality image dehazing with diffusion model. *arXiv preprint arXiv:2308.11949*, 2023. 2

[64] He Zhang and Vishal M Patel. Densely connected pyramid dehazing network. In *CVPR*, pages 3194–3203, 2018. 1, 2

[65] Shiyu Zhao, Lin Zhang, Ying Shen, and Yicong Zhou. Refinednet: A weakly supervised refinement framework for single image dehazing. *IEEE TIP*, 30:3391–3404, 2021. 6

[66] Yu Zheng, Jiahui Zhan, Shengfeng He, Junyu Dong, and Yong Du. Curricular contrastive regularization for physics-aware single image dehazing. In *CVPR*, pages 5785–5794, 2023. 2

[67] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, pages 2223–2232, 2017. 6

[68] Qingsong Zhu, Jiaming Mai, and Ling Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE TIP*, 24(11):3522–3533, 2015. 1, 2