UNCERTAINTY-AWARE DIFFUSION MODEL FOR REAL-WORLD IMAGE DEHAZING

YUAN-JIAN QIAO[1], MING-WEN SHAO[1], XIAO-DONG TAN[1], LING-ZHUANG MENG[1]

[1]School of Computer Science and Technology, China University of Petroleum (East China), Qingdao 266580, China
E-MAIL: yjqiao@s.upc.edu.cn, smw278@126.com, s22070002@s.upc.edu.cn, lzhmeng1688@163.com

Abstract:

Diffusion models have recently achieved remarkable success in image dehazing tasks. However, existing methods struggle to handle real-world hazy images due to the neglect of physical properties, i.e., the haze density generally increases with the scene depth and the haze color changes uncertainly. To address this issue, we propose a novel Uncertainty-aware Diffusion model (UnDiff) for real-world image dehazing. Specifically, inspired by the atmospheric scattering model, we elaborate an Uncertainty Degradation Modeling (UDM) pipeline that considers density and color shifts of haze to better suppress the domain gap between synthetic and real images. To adaptively tackle the diverse hazy images, we design a Context-aware Prior Embedding (CPE) module to efficiently learn the contextual information that varies with the scene depth. Benefiting from the above physical guidance mechanism, our UnDiff achieves high-quality restoration on various real-world hazy images. Extensive experiments show that our method exhibits noteworthy superiority over existing methods.

Keywords:

Image dehazing; Physical properties; Uncertainty-aware; Diffusion model; Context-aware attention

1. Introduction

Images taken under hazy scenes not only diminish the visual quality but also affect the performance of downstream practical applications (e.g., autonomous driving and surveillance systems) [1, 2]. To ameliorate this dilemma, the traditional image dehazing algorithms usually leverage prior knowledge to recover the underlying clean image [3, 4]. However, these priors are derived from empirical observations or statistical analysis, leading to unreliability in intricate real-world scenarios. With the development of deep learning techniques, various learning-based image dehazing methods based on Convolutional Neural Networks (CNN) [5, 6] or Visual Transformer (ViT) [7, 8] have been explored to deal with the damaged semantics. Despite the promising performance, these methods cannot effectively recover texture details in severely degraded regions.

Recently, diffusion models have achieved notable progress in the field of image dehazing, which learns high-quality detail reconstruction by iteratively optimizing the reverse denoising process [9, 10]. However, there remain two practical challenges: First, collecting a large number of haze-clean image pairs is extremely difficult. Hence, most existing methods simply employ atmospheric scattering models to synthesize paired data for training, resulting in terrible generalizability in the real world due to the huge domain shift. Second, in realistic environments, the haze density typically increases with the scene depth and the haze color exhibits uncertainty. Nevertheless, existing diffusion-based dehazing methods usually ignore these physical properties, making them struggle to handle intricate real-world hazy images.

To address the above limitations, we propose a novel Uncertainty-aware Diffusion model (UnDiff) for real-world image dehazing. To obtain a large amount of training data that better matches the real haze distribution, we elaborate a Uncertainty Degradation Modeling (UDM) scheme, which reduces the domain gap between the synthetic and real images by introducing uncertain density and color shifts into the atmospheric scattering model. To adaptively handle diverse hazy images, we introduce a pre-trained CLIP model to encode contextual features containing both image content and degradation information. Meanwhile, we design a Context-aware Prior Embedding (CPE) module to guide the diffusion restoration process using contextual information that varies with the scene depth. Benefiting from the above physical modeling and guidance scheme, our UnDiff achieves high-quality restoration for degradation-diverse hazy images in the real world.

The main contributions of this work are summarized below:

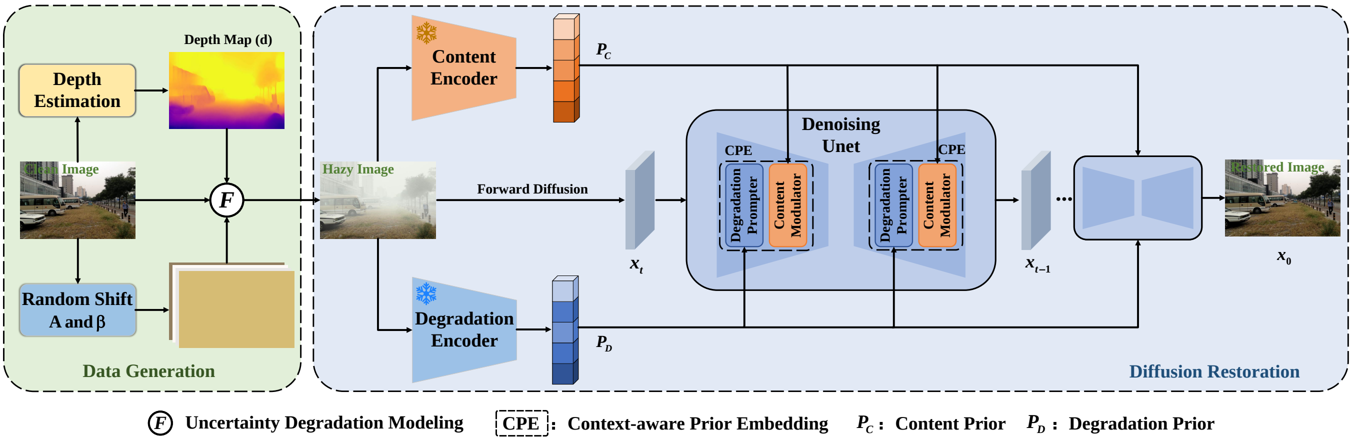


FIGURE 1. Overview of the proposed UnDiff, which encompasses data generation and diffusion restoration processes.

- We investigate the density and color properties of real-world hazy images. By exploiting this nature, we elaborate an Uncertainty Degradation Modeling (UDM) pipeline to generate more realistic training data.
- We propose an Uncertainty-aware Diffusion model (UnDiff) for real-world image dehazing. We also design a Context-aware Prior Embedding (CPE) module within the UnDiff to adaptively tackle diverse hazy images.
- Extensive experiments demonstrate that our method exhibits significant advantages over existing methods on various real-world hazy images.

2. Related Work

2.1. Traditional Methods

To address the ill-posedness of image dehazing tasks, early traditional algorithms attempted to exploit the inherent statistics prior to constrain the solution space and then recovering clean images. For instance, DCP [4] proposes the dark channel prior by observing the dark channel pixel distribution of a large number of haze-free images to effectively remove the haze. Zhu et al. [3] present the color attenuation prior to deal with complex hazy images, which establishes a linear model for modeling the scene depth of a hazy image. Fattal [11] observes that the small patches of pixel values exhibit a one-dimensional color distribution in RGB space and devises a color line prior for haze removal. However, these statistical prior are not applicable to a wide range of situations in complex real-world scenarios, resulting in unsatisfactory recovery performance.

2.2. Learning-based Methods

Benefiting from the advances of deep learning techniques, many learning-based image dehazing approaches have been presented. These methods usually utilize CNN-based or ViT-based architectures to learn the mapping from hazy images to clean ones [5, 6, 12, 7, 8]. For example, AirNet [6] proposes a unified image restoration network to cope with different degradation cases by representing features with contrastive learning. PromptIR [8] utilizes learnable prompts to adapt to different degraded images. Although these methods achieve satisfactory performance, they still suffer from content loss or color shifting when dealing with severe degradation. In recent years, diffusion models have presented strong detail reconstruction capabilities on image dehazing tasks. In particular, WeatherDiff [10] presents a diffusion model with patch learning for image recovery under inclement weather environments. However, these methods fail to take into account the relationship between haze density and scene depth, as well as the color change of the haze, resulting in lousy generalizability in the real world.

3. Method

The overall architecture of UnDiff is shown in Fig. 1, which consists of an uncertainty degradation modeling pipeline that generates diverse hazy images, and a diffusion restoration process that enables flexible cope with degradation-diverse images. Furthermore, we design a context-aware prior embedding module to learn contextual information changing with the scene depth. In the following, we first introduce the basic theory of diffusion models and then describe the above core design in detail.

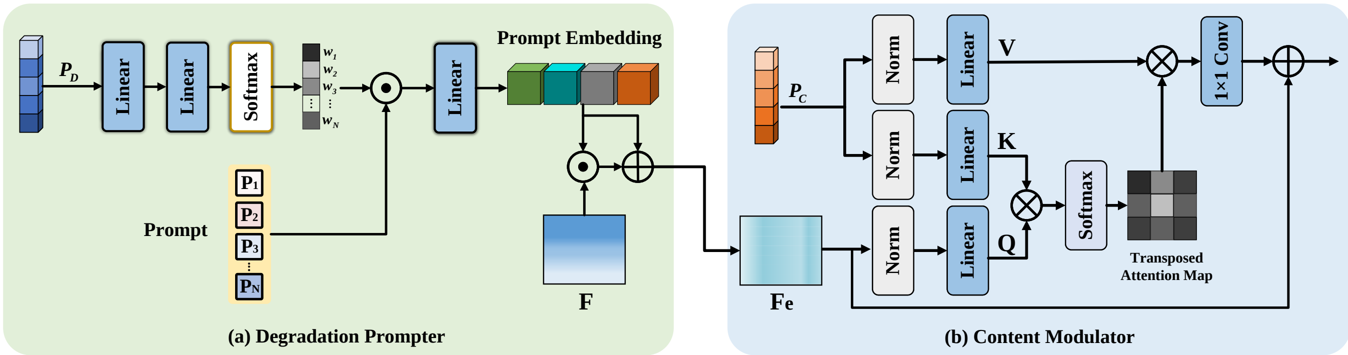


FIGURE 2. Detailed structure of the Context-aware Prior Embedding module (CPE), which contains a degradation prompter and a content modulator.

3.1 Preliminary

Denoising Diffusion Probabilistic Model (DDPM) [13] is a powerful generative model for learning a high-quality mapping from Gaussian noise to original data, which primarily comprises a forward diffusion process and a reverse denoising process. The forward diffusion process starts with the clean data and repeatedly injects Gaussian noise according to the transition kernel $q(x_t|x_{t-1})$ as follows:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I), \tag{1}$$

where β_t are fixed variance schedule corresponding to the diffusion step t. Given an input image x_0, the x_t can be sampled at any timestep using the following formula:

$$x_t = \sqrt{\overline{\alpha}_t}x_0 + \sqrt{1-\overline{\alpha}_t}\epsilon. \tag{2}$$

where ϵ is the Gaussian noise, $\alpha_t = 1-\beta_t$ and $\overline{\alpha}_t = \prod_{i=1}^{t}\alpha_i$. Conversely, the reverse denoising process recovers the clean image x_0 by iteratively optimizing the data distribution, which is expressed as:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t,t), \textstyle\sum_\theta(x_t,t)), \tag{3}$$

where $\mu_\theta(x_t,t)$ denote the mean and $\sum_\theta$ denote the variance that is a time-dependent constants or learnable parameters [13]. By employing a denoising U-Net $\epsilon_\theta(x_t,t)$, the $\mu_\theta(x_t,t)$ can be parameterized as:

$$\mu_\theta(x_t,t) = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(x_t,t)). \tag{4}$$

To train the denoising UNet, a loss function is applied to constrain the noise distribution, which can be formulated as:

$$\mathbb{E}_{x_0,t,\epsilon_t\sim N(0,I)}||\epsilon - \epsilon_\theta(\sqrt{\overline{\alpha}_t}x_0 + \sqrt{1-\overline{\alpha}_t}\epsilon, c, t)||^2 \tag{5}$$

where c represents the conditional information.

3.2 Uncertainty Degradation Modeling

In real-world scenarios, manually collecting a large number of hazy-clean image pairs is laborious and extremely difficult. To obtain large-scale paired data for supervised training, most existing methods utilize original atmospheric scattering models to synthesize hazy data [14]. Despite presenting feasible results, these solutions are not applicable to complex real-world scenarios where the density and color of haze are uncertain. To mitigate this issue, we design a novel depth-guided degradation modeling pipeline to generate more realistic data by introducing density and color shifts. This hazy image modeling scheme can be represented as:

$$I(x) = J(x)e^{-\Delta\beta d(x)} + \Delta A(1-e^{-\Delta\beta d(x)}), \tag{6}$$

where $I(x)$ and $J(x)$ denote the degraded and clean images, respectively. The variable $d(x)$ represents the scene depth, which is estimated in our paper by a pre-trained depth prediction model. The variable β represents the scattering coefficient, which reflects the haze density. We employ $\beta \in [0.4, 1.5]$ to generate hazy images with different densities guided by depth information. In addition, the variable $\Delta A = A + A_b$ represents the uncertain atmospheric lighting, which simulates the color shift that often occurs under hazy scenarios. During the training process, we consider the color change of haze by defining the color bias A_b∈[-0.025, 0.025] and the initial atmospheric light A∈[0.45, 0.98].

Unlike existing methods that simply utilize the original atmospheric scattering model to represent the hazy image formation, we devise an uncertainty degradation modeling pipeline by taking into account the depth-density properties and color variations. With this modeling scheme, our method can better narrow the domain gap between synthetic and real images and enhance the generalizability in real scenarios.

3.3 Context-aware Prior Embedding

Although different methods have been proposed to suppress the haze effect, these methods commonly cannot effectively generalize to different hazy images due to the uncertainty of the degradation presentation [6]. To adaptively cope with unknown hazy images, we leverage the pre-trained CLIP model to learn content and degradation embeddings from the generated hazy images as context prior. Meanwhile, we design a Context-aware Prior Embedding module (CPE) to inject robust contextual information information into the diffusion model for high-quality image dehazing. The CPE primarily encompasses two parts: a degradation prompter and a content modulator, as displayed in Fig. 2.

Specifically, the degradation prompter in CPE first predicts attention-based weights w_i from the degradation embedding P_D, which can be expressed as:

$$w_i = Softmax(L_{linear}(P_D)), \tag{7}$$

where L_{linear} denotes the linear layer and $Softmax$ denotes the Softmax activation function. Following that, these weights w_i are interacted with the learnable prompts P_i to yield the prompt embeddings P_e, which can be formulated as:

$$P_e = L_{linear}(\sum_{i=1}^{K} w_i \otimes P_i), \tag{8}$$

where $\otimes$ denotes linear weighting and K denotes the number of prompt components. The output features of the degradation prompter can be attained by adjusting the input features with the prompt embedding, which can be represented as:

$$P_{e1}, P_{e2} = Split(P_e), \tag{9}$$

$$F_e = F \otimes P_{e1} + P_{e2}. \tag{10}$$

The content modulator interaction prompt feature F_e with the content prior P_C through the self-attention mechanism , which can be represented as:

$$F_{CPE} = F_{Att}(P_e, P_C), \tag{11}$$

where F_{Att} denotes the self-attention operation. With these two designs in the CPE module, the diffusion model can effectively learn the contextual information that varies with scene depth, thus facilitating the recovery quality.

TABLE 1. Quantitative comparison of different methods on the Fattal and SOTS datasets.

Methods	Fattal [11]		SOTS [19]	
	FADE↓	PAQ2PIQ↑	PSNR↑	SSIM↑
DCP [4]	0.481	71.56	19.13	0.814
D4 [20]	<u>0.411</u>	73.13	25.83	0.956
AECRNet [5]	0.431	74.41	23.05	0.896
YOLY [21]	0.479	71.90	14.75	0.857
AirNet [6]	0.433	74.21	23.18	0.900
PromptIR [8]	0.425	<u>74.46</u>	<u>31.31</u>	<u>0.973</u>
Ours	**0.408**	**74.58**	**32.15**	**0.978**

4 Experiments

4.1 Experimental Settings

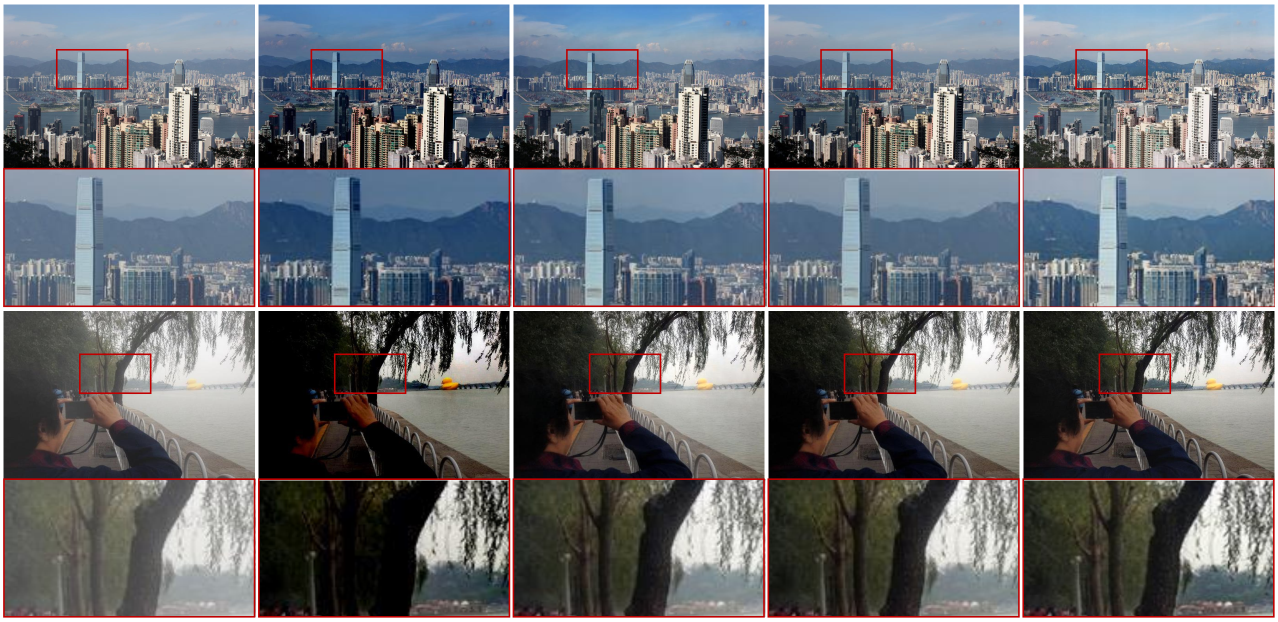
In our experiment, we adopt the AdamW optimizer for 1×10^7 iterations training, and the images are cropped to 256×256 patches. Following cosine annealing strategy [15], the learning rate is gradually decayed from initial $2e^{-4}$ to $1e^{-6}$. Based on the proposed uncertainty degradation modeling scheme, we randomly generate haze data online using clean images from [16]. The depth maps corresponding to the clean images are obtained through a pre-trained depth estimation algorithm. Our model is tested on the Fattal dataset containing 31 real-world images and the SOTS dataset containing 500 synthetic images. We employ the non-reference metrics FADE [17] and PAQ2PIQ [18] to evaluate the Fattal dataset and the PSNR/SSIM metrics for the SOTS dataset.

4.2. Experimental Results

Table. 1 presents the quantitative results on the Fattal [11] and SOTS [19] datasets. As can be seen, our UnDiff outperforms existing methods on both full-reference and non-reference metrics. Compared to PromptIR [8], which also utilizes prompt learning for image recovery, our method obtains a 0.12 PAQ2PIQ gain on the Fattal dataset. Meanwhile, our method achieves 0.74 dB PSNR gain on the SOTS dataset. Fig. 3 presents the visual results of different methods. It can be observed that all comparison methods suffer from varying degrees of haze residue or undesired lighting situations. In contrast, the images recovered by our method show more appropriate brightness and a clear background.

4.3. Ablation Studies

We similarly perform ablation experiments in Table. 2 to verify the role of different components. As can be noticed,



(a) Input (b) YOLY (c) AirNet (d) PromptIR (e) Ours

FIGURE 3. Visual comparison of different methods on Fattal and SOTS datasets.

TABLE 2. The ablation results of our UnDiff.

Methods	UDM	CPE	PSNR↑	SSIM↑
w/o UDM	×	✓	31.94	0.976
w/o DPE	✓	×	31.48	0.973
w/o UDM and DPE	×	×	31.26	0.971
Ours	✓	✓	32.15	0.978

the dehazing performance displays a significant decline after removing the UDM pipeline or the context-aware prior embedding module. This result illustrates the important role of these two designs for the overall model. In addition, as shown in Fig. 4, we visualize the depth maps of different clean images as well as the generated hazy images to demonstrate the effectiveness of our haze modeling scheme.

5. Conclusion

In this paper, we present a novel Uncertainty-aware Diffusion model (UnDiff) for real-world image dehazing. Instead of directly adopting the original atmospheric scattering model, the proposed UnDiff elaborates an Uncertainty Degradation Modeling (UDM) pipeline to synthesize more realistic hazy images. Furthermore, we incorporate a Context-aware Prior Embedding module (CPE) in the diffusion model to achieve adaptive restoration by learning both content and degradation priors. Experiments on the Fattal and SOTS datasets suggest that our approach significantly outperforms existing methods. In future work, we will explore the fusion of different physical priors for image dehazing tasks.

(a) Case1 (b) Case2 (c) Case3

FIGURE 4. Visualization of depth maps and generated hazy images.

Acknowledgements

This work was supported by National Key Research and Development Program of China 2021YFA1000102, the National Natural Science Foundation of China Nos. 62376285, 62272375, 61673396.

References

[1] D. Bogdoll, M. Nitsche, and J. M. Zöllner, "Anomaly detection in autonomous driving: A survey," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 4488–4499.

[2] L. Yi, Q. Zhao, W. Wei, and Z. Xu, "Robust online rain removal for surveillance videos with dynamic rains," *Knowledge-Based Systems*, vol. 222, p. 107006, 2021.

[3] Q. Zhu, J. Mai, and L. Shao, "Single image dehazing using color attenuation prior." in *BMVC*, 2014, pp. 1–10.

[4] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.

[5] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, and L. Ma, "Contrastive learning for compact single image dehazing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10 551–10 560.

[6] B. Li, X. Liu, P. Hu, Z. Wu, J. Lv, and X. Peng, "All-in-one image restoration for unknown corruption," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 452–17 462.

[7] C.-L. Guo, Q. Yan, S. Anwar, R. Cong, W. Ren, and C. Li, "Image dehazing transformer with transmission-aware 3d position embedding," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5812–5820.

[8] V. Potlapalli, S. W. Zamir, S. Khan, and F. S. Khan, "Promptir: Prompting for all-in-one blind image restoration," *arXiv preprint arXiv:2306.13090*, 2023.

[9] B. Xia, Y. Zhang, S. Wang, Y. Wang, X. Wu, Y. Tian, W. Yang, and L. Van Gool, "Diffir: Efficient diffusion model for image restoration," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 13 095–13 105.

[10] O. zdenizci and R. Legenstein, "Restoring vision in adverse weather conditions with patch-based denoising diffusion models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[11] R. Fattal, "Dehazing using color-lines," *ACM Transactions on Graphics*, vol. 34, no. 1, pp. 1–14, 2014.

[12] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "Ffa-net: Feature fusion attention network for single image dehazing," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 11 908–11 915.

[13] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[14] E. J. McCartney, "Optics of the atmosphere: scattering by molecules and particles," *New York*, 1976.

[15] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts. arxiv 2016," *arXiv preprint arXiv:1608.03983*, 2019.

[16] R.-Q. Wu, Z.-P. Duan, C.-L. Guo, Z. Chai, and C. Li, "Ridcp: Revitalizing real image dehazing via high-quality codebook priors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 22 282–22 291.

[17] L. K. Choi, J. You, and A. C. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3888–3901, 2015.

[18] Z. Ying, H. Niu, P. Gupta, D. Mahajan, D. Ghadiyaram, and A. Bovik, "From patches to pictures (paq-2-piq): Mapping the perceptual space of picture quality," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Ppattern Recognition*, 2020, pp. 3575–3585.

[19] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2018.

[20] Y. Yang, C. Wang, R. Liu, L. Zhang, X. Guo, and D. Tao, "Self-augmented unpaired image dehazing via density and depth decomposition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 2037–2046.

[21] B. Li, Y. Gou, S. Gu, J. Z. Liu, J. T. Zhou, and X. Peng, "You only look yourself: Unsupervised and untrained single image dehazing neural network," *International Journal of Computer Vision*, vol. 129, pp. 1754–1767, 2021.