Unleashing the Potential of the Semantic Latent Space in Diffusion Models for Image Dehazing

Zizheng Yang, Hu Yu, Bing Li, Jinghao Zhang, Jie Huang, and Feng Zhao*

MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition,
University of Science and Technology of China
{yzz6000, yuhu520, bing0123, jhaozhang, hj0117}@mail.ustc.edu.cn,
fzhao956@ustc.edu.cn

Abstract. Diffusion models have recently been investigated as powerful generative solvers for image dehazing, owing to their remarkable capability to model the data distribution. However, the massive computational burden imposed by the retraining of diffusion models, coupled with the extensive sampling steps during the inference, limit the broader application of diffusion models in image dehazing. To address these issues, we explore the properties of hazy images in the semantic latent space of frozen pre-trained diffusion models, and propose a Diffusion Latent Inspired network for Image Dehazing, dubbed DiffLI^{2}D. Specifically, we first reveal that the semantic latent space of pre-trained diffusion models can represent the content and haze characteristics of hazy images, as the diffusion time-step changes. Building upon this insight, we integrate the diffusion latent representations at different time-steps into a delicately designed dehazing network to provide instructions for image dehazing. Our DiffLI^{2}D avoids re-training diffusion models and iterative sampling process by effectively utilizing the informative representations derived from the pre-trained diffusion models, which also offers a novel perspective for introducing diffusion models to image dehazing. Extensive experiments on multiple datasets demonstrate that the proposed method achieves superior performance to existing image dehazing methods.

Keywords: Image Dehazing · Diffusion Models· Latent Space

1 Introduction

Image dehazing aims to recover a clean image from its hazy counterpart, which is critical to high-level vision tasks such as image classification [19, 26, 45] and object detection [17,32,40]. It is challenging and ill-posed due to the infinite possible solutions for a given hazy image. Conventional methods utilize the physical scattering model [35] to estimate the clean images. With the development of deep learning, convolution neural network (CNN) and Transformer-based methods have achieved great success in image dehazing [7, 31, 52, 62, 64, 66]. Recently,

* Corresponding Author.

diffusion models [22, 48] exhibit great impressive performance in image generation, and achieve unprecedented success in downstream tasks, such as image editing [4, 20] and personalization [14, 33, 42]. Meanwhile, the diffusion models also significantly broaden the scope of possibilities for image dehazing.

The prevailing approach to applying diffusion models for image dehazing is to re-train a diffusion model that is conditioned on the hazy image from scratch [34, 63]. These methods utilize the hazy image as the condition, and concatenate it with the noise map, which aims to implicitly guide the diffusion models to predict the corresponding clean image during the reverse process. Such paradigm requires re-training the entire diffusion models, which typically costs massive time and computation resources. On the other hand, the potential time-consuming sampling in the reverse process also limits their application.

To address the above issues, we try to investigate the potential of diffusion models for image dehazing from a new perspective: *"Can we directly leverage the rich knowledge contained in pre-trained diffusion models, instead of re-training diffusion models from scratch?"* To this end, we investigate the properties of hazy images within the semantic latent space of frozen pre-trained diffusion models. Previous works [23,28,37] have discovered that the semantic latent space (named *h-space*) has nice properties for high-level semantic manipulation. It is essential to investigate whether the *h-space* also exhibits properties necessary for low-level image dehazing. Specifically, we discover that, as the diffusion time-step changes, the *h-space* representations of hazy images undergo a gradual transformation, transitioning from primarily encoding the underlying image contents to increasingly capturing the haze characteristics. Fig. 1 describes the properties and Sec. 4.1 provides detailed analysis. Note that our exploration of *h-space* for image dehazing is different from previous works that focus on the intermediate outputs during the diffusion process. To the best of our knowledge, it is the first attempt to explore the potential of the semantic latent space in pre-trained diffusion models towards image dehazing.

The aforementioned observation promotes us to leverage the informative *h-space* representations to facilitate image dehazing. To this end, we propose a new framework, called the Diffusion Latent Inspired network for Image Dehazing (DiffLI^{2}D), which aims to integrate the *h-space* representations for effective image dehazing. The DiffLI^{2}D adopts a hierarchical architecture similar to U-Net [41], enabling it to learn multi-scale features for image dehazing [52, 57, 64]. Specifically, to facilitate the content recovery of hazy images, we design a content integration module (CIM), which assists the DiffLI^{2}D in restoring image contents by utilizing the content representations derived from *h-space*. Furthermore, for better haze removal, a haze-aware enhancement (HAE) module is developed. It leverages the haze representations obtained from *h-space* as guidance, enabling DiffLI^{2}D to remove the haze from the input hazy images effectively.

Moreover, the proposed DiffLI^{2}D does not require re-training any diffusion models, and circumvents the time-consuming reverse sampling process. Compared with existing diffusion model-based methods [57, 61], our DiffLI^{2}D costs less computation resources.

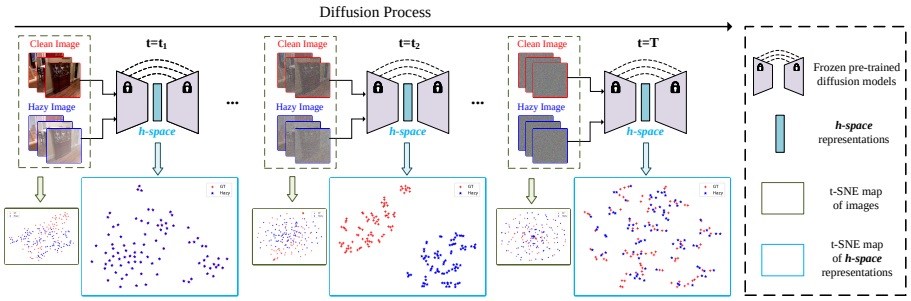


Fig. 1: Distributions of hazy images and their corresponding clean images in *h-space* and image space at different time-step t during the diffusion process. When the time-step t is small (*i.e.*, $t = t_1$), the *h-space* features with the same underlying image content are tightly clustered together, while those with different image contents are separated. When t is large (*i.e.*, $t = t_2$), the *h-space* features of hazy and clean images are distinguished. Note that when t becomes too large (*i.e.*, $t = T$), the distribution of hazy and clean *h-space* features becomes chaotic and irregular. The t-SNE maps in image space are also presented for comparison. Please zoom in for better view.

We summarize our main contributions as follows:

- To the best of our knowledge, this is the first attempt to explore the *h-space* of diffusion models for image dehazing. Additionally, we propose the DiffLI^{2}D framework for image dehazing through leveraging the informative representations derived from *h-space*.
- Our findings reveal a transition in the *h-space* representations of hazy images, shifting from encoding the image contents to capturing haze characteristics, as the diffusion time-step changes.
- Considering the properties of *h-space* representations, we develop two modules, namely CIM and HAE, to facilitate the content recovery and haze removal in DiffLI^{2}D by leveraging the features derived from *h-space*.
- Extensive experiments demonstrate the superiority of our method. Moreover, the DiffLI^{2}D requires less computation resources, since it avoids re-training diffusion models and time-consuming reverse sampling process.

2 Related Work

2.1 Image Dehazing

Image dehazing aims to recover a clean image from its hazy version. Conventional approaches use physical scattering model [35], and try to regularize the solution space with various image priors [3, 12, 18]. However, these hand-crafted image priors may not be reliable. Recently, deep learning-based methods have dominated the image dehazing algorithms [7, 38, 59, 60, 62, 66, 68]. For example, AOD-Net [29] tries to recover the clean images by reformulating the physical

scattering model. AECR-Net [55] introduces contrastive regularization to image dehazing. FSDGN [62] attempts to recover clean images through both spatial and frequency domains. Recently, transformer [49] is also introduced to image dehazing task and has achieved great success [6, 31, 39, 52, 64].

2.2 Diffusion Models

Diffusion model [22, 46], as a newly emerged generative model, has achieved remarkable progress in image generation [10] and various downstream tasks, like image editing [4, 20] and personalization [14, 27, 33, 42]. Taking the DDPM [22] as an example, it constructs a Markov chain, and trains a denoising network, which aims to accurately fit target distributions. Current diffusion models-based image restoration methods can be divided into two categories. The first one is to re-train a diffusion model from scratch [34, 44, 61, 63], which often demands massive computation resources and time. The second one is to guide the pre-trained diffusion models to generate target images by constraining the reverse sampling [9, 47], avoiding re-training diffusion models. However, the time-consuming reverse sampling and the need for accurate degradation process limit their applications. The work [28] explores the properties of the semantic latent space in pre-trained diffusion models (*i.e.*, *h-space*) for high-level semantic manipulation. Despite this, the characteristics of *h-space* for low-level image restoration are yet to be explored.

3 Preliminary: Diffusion Models

In this paper, we follow the DDPM [22], and briefly introduce the key points in diffusion models. Concretely, it consists of a T-steps forward process that gradually adds Gaussian noise to the input image x_0, and a reverse process that learns to generate images by progressively denoising.

In the forward process, for any $t \in [0, T]$, we can get the current state x_t:

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

where x_t is the noisy image at time-step t, β_t is the variance schedule [22], and $\mathbf{I}$ is the identity matrix. Through the reparameterization, we can get x_t given x_0:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)\mathbf{I})), \tag{2}$$

where $\alpha_t = 1 - \beta_t$, and $\bar{\alpha}_t = \prod_{s=0}^{t} \alpha_s$.

During the reverse process, the diffusion models aim to estimate the previous state x_{t-1} from the current state x_t. We can get the posterior distribution $p(x_{t-1}|x_t, x_0)$ through the Bayes' theorem:

$$p(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \mu_t(x_t, x_0), \sigma_t^2\mathbf{I}), \tag{3}$$

where the mean $\mu_t(x_t, x_0) = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon)$, and the variance $\sigma_t^2 = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$. DDPM leverages a neural network ϵ_θ to estimate the noise ϵ in $\mu_t(x_t, x_0)$. For any time-step $t \in [0, T]$, we can get the loss function defined in [22]:

$$L(\theta) = \|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon)\|_2^2. \tag{4}$$

In the reverse process, the DDPM utilizes the iterative sampling from the posterior distribution to get the x_{t-1}. This allows the DDPM to generate a sample $x_0 \sim q(x_0)$ from a pure Gaussian noise $x_T \sim \mathcal{N}(0, \mathbf{I})$, where $q(x_0)$ denotes the data distribution of the training dataset.

4 Method

In this section, we provide a detailed introduction to our method. We first investigate the properties of *h-space* representation at different time-step in Sec. 4.1. And then, we introduce the delicately designed DiffLI^{2}D in Sec. 4.2.

4.1 H-Space Investigation for Image Dehazing

The denoising neural network ϵ_θ in Eq. 4 is commonly implemented as U-Net in diffusion models. The bottleneck of the frozen pre-trained U-Net, also known as the *h-space*, has been demonstrated to be rich in semantics and can be utilized for high-level semantic manipulation [28]. This inspires us to investigate the potential of *h-space* representation for low-level image dehazing.

To simplify expression, let's define some variables first. Given a hazy image x and its corresponding clean (ground-truth) counterpart y, we can get their noisy version at time-step t through the Eq. 2, denoted as x_t and y_t, respectively, where $t \in [0, T]$. Following [10], the T is set to 1000. Note that $x = x_0$, and $y = y_0$. By feeding the x_t and y_t into the frozen pre-trained diffusion models ϵ_θ, we can further obtain the corresponding *h-space* features, represented as h_t^{haz} and h_t^{cle}, respectively.

Investigating H-Space. To explore the relation between hazy and clean images in *h-space*, we propose a decoder to map the *h-space* features back to images, as shown in Fig. 2. Specifically, for each time-step t, we train a corresponding decoder D_t, which aims to map the h_t^{cle} to the noise-free clean image y. It can be formulated by:

$$\mathcal{L}_t = \|D_t(h_t^{cle}) - y\|_1 \tag{5}$$

where $\|\cdot\|_1$ denotes the L1 regularization. Note that we only use the *h-space* features h_t^{cle} corresponding to the clean images y_t to train D_t, while the hazy images and their *h-space* features do not participate in training D_t.

After that, we feed the h_t^{haz} to the trained D_t, and obtain the corresponding reconstruction results, which is $r_t^{haz} = D_t(h_t^{haz})$. The h_t^{cle} is also sent to the D_t for comparison, which is $r_t^{cle} = D_t(h_t^{cle})$. Interestingly, we find that the r_t^{haz} represents different characteristics of the original hazy image x, as t changes. As shown in Fig. 2, when t is small (*e.g.*, $t = t_1$), $r_{t_1}^{haz}$ focuses on representing the contents of the original image, making it very similar to $r_{t_1}^{cle}$. As t progressively increases (*e.g.*, when $t = t_2$), $r_{t_2}^{haz}$ shifts from primarily representing the image

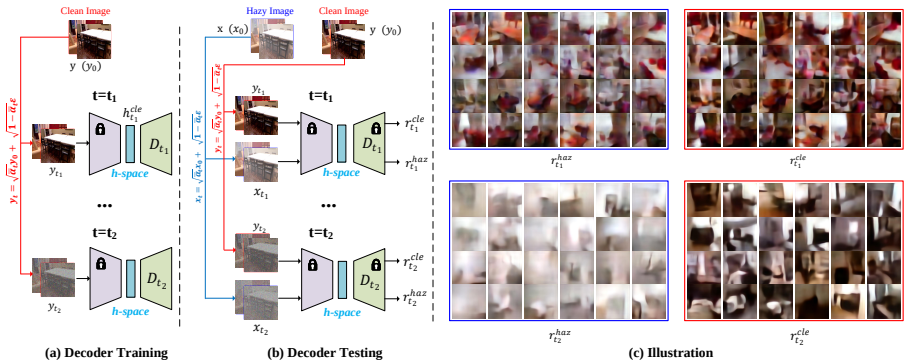


Fig. 2: Illustration of our investigation of the *h-space*. (a) Decoder training: for each time-step t, we train a decoder D_t to reconstruct the noise-free clean image y from the *h-space* feature h_t^{cle}. Note that the decoders are trained with clean noisy images y_t only. (b) Decoder testing: we send both hazy and clean *h-space* features h_t^{haz} and h_t^{cle} to the trained D_t, and obtain r_t^{haz} and r_t^{cle}, respectively. (c) Illustration of r_t^{haz} and r_t^{cle} at different t. As the time-step t changes, the r_t^{haz} represents different components of the original hazy image x^{haz}, which further indicates the different representation of *h-space* features at different t.

content to reflecting the haze characteristics in x, which makes it significantly different from $r_{t_2}^{cle}$. Based on the above observations, we can deduce the conclusion: when the time-step t is small, the *h-space* feature of the hazy image primarily represents the content of the image; as t increases, the *h-space* features shift its emphasis towards representing the haze characteristics of the hazy image. We provide more analysis and implementation details in **Appendix**.

To further verify our conclusion, we present the t-SNE maps illustrating the *h-space* features of hazy-clean image pairs at different t, as shown in Fig 1. We also show their t-SNE maps in image space for comparison. We can see that, when t is small, the h_t^{haz} and its corresponding clean counterpart h_t^{cle} are tightly clustered together, while the (h_t^{haz}, h_t^{cle}) pairs with different content are separated individually, which indicates that both h_t^{haz} and h_t^{cle} represent the image content. In contrast, when t is large, the *h-space* features of hazy images are clustered together, and those of clean images are clustered together separately, showing that the h_t^{haz} represents the haze characteristics in images. As a comparison, in the image space, the x_t and y_t are distributed irregularly at all time-step t. It is noteworthy that when t is too large (*e.g.*, $t = T$), both the *h-space* and the image space exhibit irregular patterns. This is because the sufficiently large noise removes both content and haze in the original hazy image.

Discussion. Many works [8, 50] have proven that diffusion models generate images in a coarse-to-fine manner during the reverse process, and we attribute the properties exhibited by *h-space* to this. At the early steps of the diffusion process (*i.e.*, t is small), the diffusion models focus on fine-grained details, which enables the *h-space* features to fully perceive the background content of the

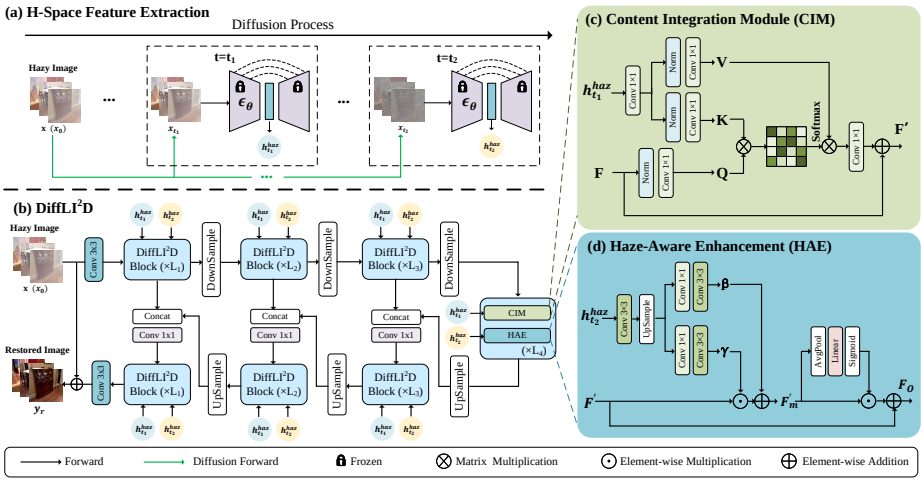


Fig. 3: Architecture of the proposed DiffLI^{2}D. (a) Given a hazy image x, we first get the noisy versions x_{t_1} and x_{t_2} through Eq. 2, and then obtain the *h-space* features $h_{t_1}^{haz}$ and $h_{t_2}^{haz}$ by sending x_{t_1} and x_{t_2} into the frozen pre-trained diffusion model. (b) Architecture of the dehazing network, which comprises multiple blocks arranged in the U-Net structure. Each block consists of a Content Integration Module (CIM) and a Haze-Aware Enhancement (HAE) module, where the former is designed to leverage the $h_{t_1}^{haz}$ to facilitate the content recovery, while the latter utilizes the $h_{t_2}^{haz}$ as guidance for effective haze removal.

image. In contrast, at the late steps of the diffusion process (*i.e.*, t is large), the diffusion models concentrate more on the coarse attributes, which allows the *h-space* features to represent the foreground haze of the image. More discussion is provided in **Appendix**.

4.2 Architecture of the DiffLI^{2}D

The observations in Sec. 4.1 promote us to utilize the *h-space* features to facilitate image dehazing. For all *h-space* features h_t^{haz} ($t \in [0, T]$), the $h_{t_1}^{haz}$ (*i.e.*, $t = t_1$) and $h_{t_2}^{haz}$ (*i.e.*, $t = t_2$) are the most representative, where the former captures the underlying content of x, while the latter characterizes the haze attributes, as discussed in Sec. 4.1. The selection of t_1 and t_2 will be described in Sec. 5.3. Based on this, we propose a Diffusion Latent Inspired network for Image Dehazing (DiffLI^{2}D). As shown in Fig. 3, the DiffLI^{2}D comprises multiple DiffLI^{2}D blocks arranged in the U-Net structure. Each block consists of a Content Integration Module (CIM) that leverages $h_{t_1}^{haz}$ for image content recovery, and a Haze-Aware Enhancement (HAE) module which utilizes $h_{t_2}^{haz}$ for haze removal.

Content Integration Module. To effectively leverage the *h-space* feature $h_{t_1}^{haz}$ for content recovery, we propose the Content Integration Module (CIM), as illustrated in Fig. 3(c). Specifically, given the intermediate feature $F \in \mathbb{R}^{\hat{H} \times \hat{W} \times \hat{C}}$

and the *h-space* feature $h_{t_1}^{haz} \in \mathbb{R}^{H' \times W' \times C'}$, we can get $h_{t_1}^{haz'} = W_c h_{t_1}^{haz}$, where $h_{t_1}^{haz'} \in \mathbb{R}^{H' \times W' \times \hat{C}}$, and W_c is a 1×1 convolution kernel. Then, F is projected into query $Q = W_Q F$, and $h_{t_1}^{haz'}$ is projected into key $K = W_K h_{t_1}^{haz'}$ and value $V = W_V h_{t_1}^{haz'}$. W_Q, W_K, and W_V are all implemented by 1×1 convolution kernel. After that, we reshape $Q \in \mathbb{R}^{\hat{H}\hat{W} \times \hat{C}}$, and $K, V \in \mathbb{R}^{H'W' \times \hat{C}}$, and obtain final output F' formulated by:

$$F' = W_O \cdot Softmax(\frac{QK^T}{\sqrt{\hat{C}}}) \cdot V + F, \tag{6}$$

where W_O is implemented by 1×1 convolution kernel. Similar with transformer [49], the multi-head mechanism is introduced to the CIM. Through the interaction, the CIM encourages the DiffLI^{2}D to fully explore the correspondence between F and $h_{t_1}^{haz}$, and further enables the DiffLI^{2}D to dynamically capture the informative content representations in $h_{t_1}^{haz}$ for effective content recovery.

Haze-Aware Enhancement. We further design a Haze-Aware Enhancement (HAE) module to utilize $h_{t_2}^{haz}$ as guidance for haze removal, as shown in Fig. 3(d). Given the *h-space* feature $h_{t_2}^{haz}$, we first use it as the guidance to dynamically enhance the input feature F' in a SFT [51] manner, which is:

$$F'_m = W_\gamma h_{t_2}^{haz} \cdot F' + W_\beta h_{t_2}^{haz}. \tag{7}$$

After that, we further modulate the integrated feature F'_m from channel dimension, which is:

$$F_o = \sigma(W_L \cdot AvgPool(F'_m)) \cdot F'_m + F', \tag{8}$$

where σ denotes the Sigmoid function, W_L is a linear layer, and $AvgPool$ means the average pooling operation. Through this, the HAE adapts DiffLI^{2}D to the haze characteristics within $h_{t_2}^{haz}$, which dynamically modulates the input features under the guidance of $h_{t_2}^{haz}$ for haze removal and further enhancement.

The optimization objective for training DiffLI^{2}D is defined as:

$$\mathcal{L} = \|y_r - y\|_1 \tag{9}$$

where y_r denotes the restored image. Note that the pre-trained diffusion model ϵ_θ is frozen, and do not participate in the optimization.

Note that our DiffLI^{2}D avoids re-training diffusion models [24, 63] and the time-consuming reverse sampling process [9, 47]. Instead, by leveraging the *h-space* features $h_{t_1}^{haz}$ and $h_{t_2}^{haz}$ as guidance, the DiffLI^{2}D can effectively recover clean images. More detailed comparisons are discussed in **Appendix**.

5 Experiments

In this section, we conduct comprehensive experiments to verify the effectiveness of the proposed DiffLI^{2}D. More experimental results and details can be found in **Appendix** for further reference.

Table 1: Performance comparisons with state-of-the-art dehazing methods across synthetic (*i.e.*, SOTS) and real-world (*i.e.* Dense-Haze and NH-HAZE) dehazing datasets. The superscript * means diffusion model-based method for image dehazing. The best results are marked as **bold** and the second ones are masked by underline.

Method	SOTS			Dense-Haze			NH-HAZE			#Params
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	
DCP [18]	15.09	0.765	0.069	10.03	0.386	0.605	10.57	0.520	0.399	-
DehazeNet [5]	20.64	0.800	0.242	13.84	0.425	0.637	16.62	0.524	0.529	0.01M
AOD-Net [29]	19.82	0.818	0.099	13.14	0.414	0.599	15.40	0.569	0.495	0.002M
FFA-Net [38]	36.39	0.989	0.005	14.39	0.452	0.498	19.87	0.692	0.365	4.68M
MSBDN [11]	33.79	0.984	0.029	15.37	0.486	0.536	19.23	0.706	0.292	31.35M
SwinIR [31]	24.93	0.932	0.049	12.20	0.510	0.639	16.15	0.623	0.479	0.91M
AECR-Net [55]	37.17	0.990	0.007	15.80	0.466	0.537	19.88	0.707	0.278	2.61M
MPRNet [65]	32.14	0.983	0.011	13.82	0.519	0.620	17.88	0.631	0.368	15.74M
Restormer [64]	38.43	0.989	0.009	15.17	0.557	0.629	18.32	0.635	0.355	26.13M
Dehamer [15]	36.63	0.988	0.005	16.62	0.560	0.480	**20.66**	0.684	0.230	132.50M
IR-SDE* [34]	33.82	0.984	0.014	12.03	0.508	0.485	12.59	0.520	0.361	537.21M
ResShift* [63]	29.06	0.950	0.017	13.67	0.517	0.576	16.26	0.625	0.327	114.65M
DiffLI^{2}D* (Ours)	**40.33**	**0.992**	**0.004**	**16.97**	**0.584**	**0.406**	20.29	**0.738**	**0.217**	8.63M

5.1 Implementation Details

Datasets. We evaluate our method on both synthetic and real-world datasets. For synthetic scene, the RESIDE [30] is utilized for training and testing. Specifically, the subset Indoor Training Set (ITS) of RESIDE is used for training. It consists of 13,990 hazy images, which are generated from 1,399 clean images. The subset Synthetic Objective Testing Set (SOTS) of RESIDE includes 500 indoor and 500 outdoor hazy images. We choose the indoor part for testing. For real-world scene, the Dense-Haze [1] and NH-HAZE [2] are adopted. Both datasets have 55 hazy-clean image pairs, 50 of which are utilized for training and 5 of which are utilized for testing.

Training Details. The proposed DiffLI^{2}D is trained by Adam optimizer, where β_1 and β_2 are set to 0.9 and 0.999, respectively. The total training epoch is set to 1200. The initial learning rate is set to 2×10^{-4}, and decreases with a factor of 0.5 every 300 epochs. The mini-batch is set to 40, and the images are resized, cropped to 128×128 with being flipped horizontally randomly. In our experiments, we choose the unconditional DDPM model pre-trained on ImageNet [10] as the ϵ_θ. The whole model is trained with one 3090Ti GPU using PyTorch framework. More implementation details are provided in **Appendix**.

Evaluation Metrics. To assess the performance, we adopt three different metrics, including PSNR, SSIM [53] and LPIPS [67]. PSNR and SSIM are employed to quantify the fidelity of the restored images – the higher their values, the better the restoration quality. LPIPS is utilized to measure the perceptual difference and visual quality, with lower values indicating better performance.

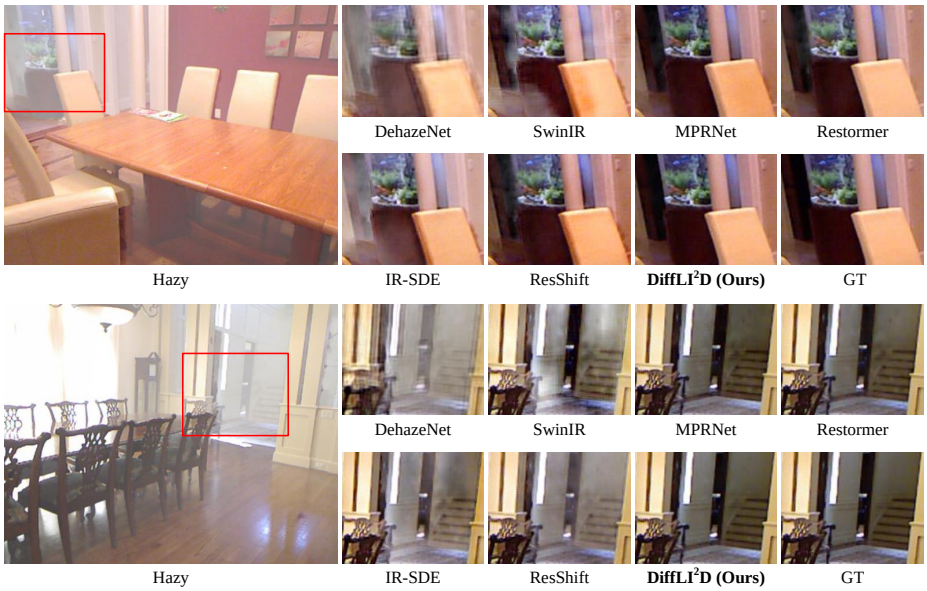


Fig. 4: Qualitative results of different methods for dehazing on SOTS dataset. Our method is shown in **bold**.

5.2 Evaluation on Image Dehazing

Experiment Results on Synthetic Dataset. Tab. 1 shows the comparison results between the DiffLI^{2}D and existing dehazing methods on SOTS dataset. We can see that DiffLI^{2}D outperforms Restormer and Dehamer with less parameters. Additional, the lower LPIPS scores indicate that the image restored by DiffLI^{2}D are better aligned with human visual system. It is noteworthy that, compared with existing diffusion model-based methods (*e.g.*, IR-SDE), our method not only achieves superior results but also avoids the re-training diffusion models. Moreover, our method circumvents the potential time-consuming sampling during the inference. We also shows the qualitative comparison in Fig. 4. As we can see, our method can recover image details and remove haze more effectively.

Experiment Results on Real-World Datasets. We further evaluate the proposed DiffLI^{2}D on Dense-Haze [1] and NH-HAZE [2] dataset. Tab. 1 shows that the DiffLI^{2}D outperforms or achieves at least comparable performance to compared image dehazing methods across two datasets. Fig. 5 and Fig. 6 illustrate the qualitative results on Dense-Haze and NH-HAZE, respectively. It can be seen that our method can recover haze-free images with better visual effects.

5.3 Ablation Study

In this section, we perform comprehensive ablation studies to demonstrate the effectiveness of our designs in the proposed DiffLI^{2}D. More results of ablation studies are provided in **Appendix**.

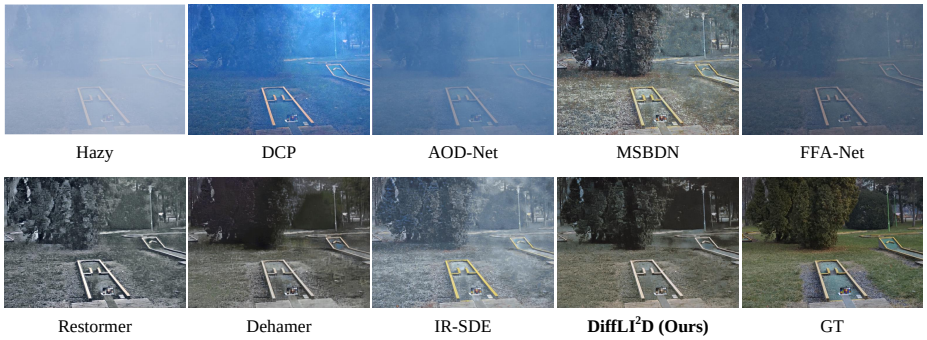


Fig. 5: Qualitative results of different methods for dehazing on Dense-Haze dataset.

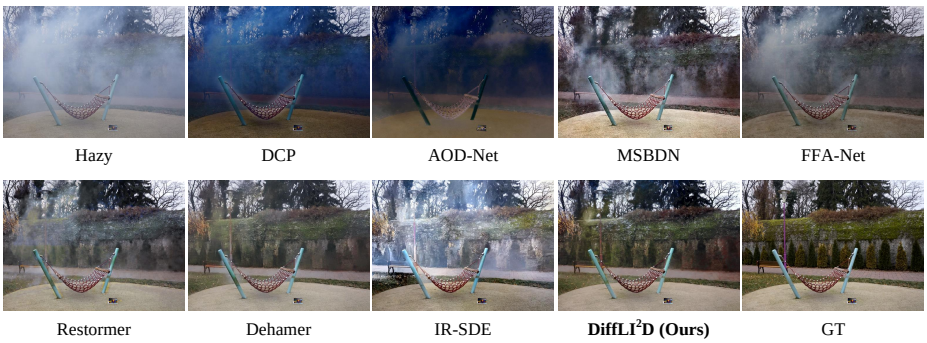


Fig. 6: Qualitative results of different methods for dehazing on NH-HAZE dataset.

Choice of t_1 and t_2 for image dehazing. As discussed in Sec. 4.1, the *h-space* features exhibit different characteristics at different time-step t. In fact, as t increases, the *h-space* features undergo a gradual and continuous transformation, transitioning from primarily representing the contents to mainly reflecting the haze properties. We show this transformation in Fig. 7. This implies that, compared with other time-steps, the *h-space* feature h_0^{haz} corresponding to $t = 0$ is the most representative of the underlying image content. So in our experiments, t_1 is set to 0. As illustrated in Fig. 8(a), $t_1 = 0$ is the best choice.

We further investigate the influence of the choice of t_2 for image dehazing. We find that the DiffLI^{2}D achieve optimal results for image dehazing when t_2 is around 500. As described in Fig. 8(b), $t_2 = 500$ outperforms $t_2 = 100$ by 1.99dB in terms of PSNR, which also surpasses $t_2 = 600$ by 2.07dB in PSNR. This is consistent with the observation in Fig. 7. When t_2 is relatively small (*e.g.*, $t_2 = 300$), the *h-space* features $h_{t_2}^{haz}$ are still intertwined with $h_{t_2}^{cle}$, and cannot effectively represent the haze characteristics. When t_2 is around 500, the $h_{t_2}^{haz}$ are clustered together, which can capture and represent the haze attributes effectively. Note that whether for t_1 or t_2, when they are too large (*e.g.*, $t_1 = 1000$ or $t_2 = 1000$), the image dehazing performance of DiffLI^{2}D experiences a

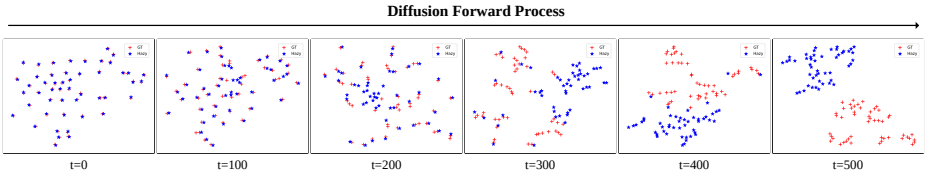
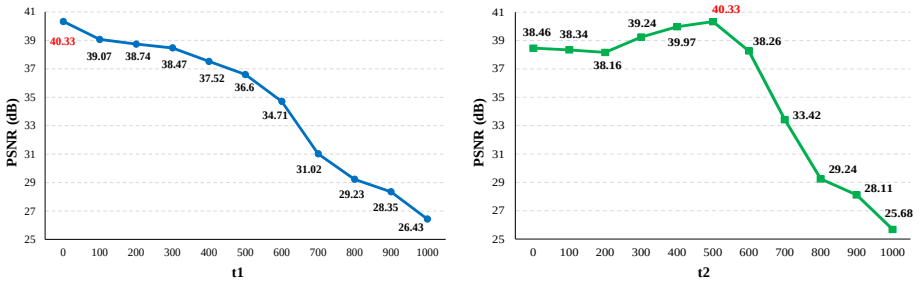


Fig. 7: Transformations of the *h-space* feature distributions of hazy-clean image pairs, as the time-step t changes.



(a) Choice of t1 for image dehazing (a) Choice of t2 for image dehazing

Fig. 8: Performance (PSNR) comparison of different choice of t_1 and t_2 for image dehazing on SOTS dataset.

significant decline. This can be attributed to the excessive noise that erases a considerable amount of information from the original hazy image x, making the *h-space* feature less effective in guiding the dehazing process.

Effectiveness of the CIM and HAE. Our DiffLI^{2}D consists of two key modules, the content integration module (CIM) and the haze-aware enhancement (HAE) module. The former is designed to facilitate the image content recovery with $h^{haz}_{t_1}$, while the latter aims to enhance haze removal using $h^{haz}_{t_2}$. To evaluate the benefits of them, we design several variants as shown in Tab. 2. Among them, the "DiffLI^{2}D w/o CIM" represents that we replace the CIM in each DiffLI^{2}D block with self-attention module that has similar number of parameter as CIM. We do a similar operation for the "DiffLI^{2}D w/o HAE". The "Baseline" means that both CIM and HAE are replaced.

As we can see, "DiffLI^{2}D w/o CIM" and "DiffLI^{2}D w/o HAE" outperform "Baseline" by 1.56dB, 1.72dB, 0.53dB and 1.79dB, 1.83dB, 1.05dB in terms of PSNR on SOTS, Dense-Haze, NH-HAZE dataset, respectively. With both two modules, the "DiffLI^{2}D" achieves 40.33dB, 16.97dB, 20.29dB on SOTS, Dense-Haze, NH-HAZE dataset, which demonstrates that the CIM and HAE are complementary and both vital to the DiffLI^{2}D, jointly resulting in a superior performance in image dehazing.

Effectiveness of the *H-Space* features. To further evaluate the effectiveness of the *h-space* for image dehazing, we compare it with other feature spaces. Specifically, we compare *h-space* with feature spaces of different layers in the

Table 2: Ablation results of several variants of DiffLI^{2}D for image dehazing on SOTS, Dense-Haze, and NH-HAZE dataset. The PSNR is utilized for evaluation.

Model	CIM	HAE	SOTS	Dense-Haze	NH-HAZE
Baseline	×	×	36.65	14.32	18.11
DiffLI^{2}D w/o CIM	×	√	38.21	16.04	18.64
DiffLI^{2}D w/o HAE	√	×	38.44	16.15	19.16
DiffLI^{2}D	√	√	**40.33**	**16.97**	**20.29**

Table 3: Performance (PSNR) comparisons between *H-Space* and other feature spaces for image dehazing on SOTS, Dense-Haze, and NH-HAZE datasets.

Method	Layer Comparison			Network Comparison		
	Layer-E	Layer-D	H-Space	VGG16	ResNet50	H-Space
SOTS	36.52	39.10	**40.33**	27.83	25.49	**40.33**
Dense-Haze	13.86	16.44	**16.97**	12.87	12.52	**16.97**
NH-HAZE	17.53	19.88	**20.29**	16.23	15.62	**20.29**

same diffusion model, and we also compare *h-space* with feature spaces of other networks. Tab. 3 shows the comparison results. For different layers, "Layer-E" and "Layer-D" means that we replace the *h-space* representations $h_{t_1}^{haz}$ and $h_{t_2}^{haz}$ with those features extracted from the encoder and decoder of the same pre-trained diffusion model. As we can see, compared with features extracted from the encoder and decoder, the *h-space* features achieves 3.81dB, 3.11dB, 2.76dB and 1.23dB, 0.53dB, 0.41dB improvement in terms of PSNR on SOTS, Dense-Haze, and NH-HAZE dataset, respectively. For different networks, we choose the commonly-used VGG16 [45] and ResNet50 [19] as comparison. Note that, for fair comparison, both VGG16 and ResNet50 are pre-trained on ImageNet [43], which is the same dataset used to pre-train the diffusion model ϵ_θ employed in our study. The "VGG16" and "ResNet50" in Tab. 3 denote that we replace the *h-space* features $h_{t_1}^{haz}$ and $h_{t_2}^{haz}$ with corresponding features extracted from the feature spaces of VGG16 and ResNet50, respectively. As we can see, *h-space* features drive from pre-trained diffusion model can facilitate the image dehazing more effectively than those extracted from other pre-trained models.

5.4 The Generalization Capabilities

To evaluate the generalization capabilities of our method, we further evaluate the effectiveness of our method on more low-level image restoration tasks. Specifically, we further test our method on low-light image enhancement on LOL-v1 [54] and LOL-v2 dataset [58]. We show the comparison results about low-light image enhancement on Tab. 4. FID [21] metric is introduced for further evaluation. We also show the qualitative results in Fig. 9. As we can see, the proposed DiffLI^{2}D can also achieve superior performance when handling other low-level image restoration tasks, which demonstrates that the capabilities of DiffLI^{2}D are

Table 4: Performance comparisons with state-of-the-art low-light image enhancement methods on LOL-v1 and LOL-v2 Real dataset. The superscript * denotes the diffusion model-based image restoration method. The best results are marked as **bold** and the second ones are masked by underline.

Method	LOLv1				LOLv2-Real			
	PSNR	SSIM	LPIPS	FID	PSNR	SSIM	LPIPS	FID
RetinexNet [54]	16.77	0.462	0.417	126.27	17.72	0.652	0.436	133.91
Zero-DCE [16]	14.86	0.562	0.372	87.24	18.06	0.580	0.352	80.45
EnlightenGAN [25]	17.61	0.653	0.372	94.70	18.68	0.678	0.364	84.04
URetinex-Net [56]	19.97	0.828	0.267	62.38	21.13	0.827	0.208	49.84
Uformer [52]	19.00	0.741	0.354	109.35	18.44	0.759	0.347	98.14
Restormer [64]	20.61	0.792	0.288	73.00	21.36	0.835	0.314	63.69
WeatherDiff* [36]	16.30	0.786	0.277	65.61	15.87	0.801	0.272	65.82
ResShift* [63]	19.23	0.735	0.225	61.21	20.41	0.704	0.218	60.72
IR-SDE* [34]	12.90	0.557	0.486	175.33	12.53	0.511	0.453	157.08
GDP* [13]	13.93	0.630	0.445	95.16	13.15	0.527	0.421	97.54
DiffLI^{2}D* (Ours)	**23.30**	**0.849**	**0.136**	**55.88**	**22.35**	**0.874**	**0.186**	**42.49**

Fig. 9: Qualitative results of different methods for low-light image enhancement.

not limited to image dehazing. More discussion, analysis, visual results and comparison results about other image restoration tasks are provided in **Appendix**.

6 Conclusion

In this paper, we investigate the semantic latent space of frozen pre-trained diffusion models for image dehazing, and reveal that the features in the semantic latent space can effectively represent the content and haze characteristics of hazy images, as the time-step changes. We also propose a Diffusion Latent Inspired network for Image Dehazing (DiffLI^{2}D), which uses the semantic latent features of frozen pre-trained diffusion models for effective image dehazing. Extensive experiments on multiple datasets demonstrate the effectiveness of our method.

Acknowledgments. This work was supported by the Anhui Provincial Natural Science Foundation under Grant 2108085UD12. We acknowledge the support of GPU cluster built by MCC Lab of Information Science and Technology Institution, USTC.

References

1. Ancuti, C.O., Ancuti, C., Sbert, M., Timofte, R.: Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In: 2019 IEEE international conference on image processing (ICIP). pp. 1014–1018. IEEE (2019)
2. Ancuti, C.O., Ancuti, C., Timofte, R.: Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 444–445 (2020)
3. Berman, D., Avidan, S., et al.: Non-local image dehazing. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1674–1682 (2016)
4. Brooks, T., Holynski, A., Efros, A.A.: Instructpix2pix: Learning to follow image editing instructions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18392–18402 (2023)
5. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: An end-to-end system for single image haze removal. IEEE transactions on image processing **25**(11), 5187–5198 (2016)
6. Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., Ma, S., Xu, C., Xu, C., Gao, W.: Pre-trained image processing transformer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12299–12310 (2021)
7. Chen, W.T., Ding, J.J., Kuo, S.Y.: Pms-net: Robust haze removal based on patch map for single images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11681–11689 (2019)
8. Choi, J., Lee, J., Shin, C., Kim, S., Kim, H., Yoon, S.: Perception prioritized training of diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11472–11481 (2022)
9. Chung, H., Kim, J., Mccann, M.T., Klasky, M.L., Ye, J.C.: Diffusion posterior sampling for general noisy inverse problems. arXiv preprint arXiv:2209.14687 (2022)
10. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Advances in neural information processing systems **34**, 8780–8794 (2021)
11. 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)
12. Fattal, R.: Single image dehazing. ACM transactions on graphics (TOG) **27**(3), 1–9 (2008)
13. Fei, B., Lyu, Z., Pan, L., Zhang, J., Yang, W., Luo, T., Zhang, B., Dai, B.: Generative diffusion prior for unified image restoration and enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9935–9946 (2023)
14. Gal, R., Alaluf, Y., Atzmon, Y., Patashnik, O., Bermano, A.H., Chechik, G., Cohen-Or, D.: An image is worth one word: Personalizing text-to-image generation using textual inversion. arXiv preprint arXiv:2208.01618 (2022)
15. Guo, C.L., Yan, Q., Anwar, S., Cong, R., Ren, W., Li, C.: Image dehazing transformer with transmission-aware 3d position embedding. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5812–5820 (2022)
16. Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R.: Zero-reference deep curve estimation for low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1780–1789 (2020)

17. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969 (2017)
18. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE transactions on pattern analysis and machine intelligence **33**(12), 2341–2353 (2010)
19. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
20. Hertz, A., Mokady, R., Tenenbaum, J., Aberman, K., Pritch, Y., Cohen-Or, D.: Prompt-to-prompt image editing with cross attention control. arXiv preprint arXiv:2208.01626 (2022)
21. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems **30** (2017)
22. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
23. Jeong, J., Kwon, M., Uh, Y.: Training-free content injection using h-space in diffusion models. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 5151–5161 (2024)
24. Jiang, H., Luo, A., Han, S., Fan, H., Liu, S.: Low-light image enhancement with wavelet-based diffusion models. arXiv preprint arXiv:2306.00306 (2023)
25. Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., Wang, Z.: Enlightengan: Deep light enhancement without paired supervision. IEEE transactions on image processing **30**, 2340–2349 (2021)
26. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Communications of the ACM **60**(6), 84–90 (2017)
27. Kumari, N., Zhang, B., Zhang, R., Shechtman, E., Zhu, J.Y.: Multi-concept customization of text-to-image diffusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1931–1941 (2023)
28. Kwon, M., Jeong, J., Uh, Y.: Diffusion models already have a semantic latent space. arXiv preprint arXiv:2210.10960 (2022)
29. 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)
30. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single-image dehazing and beyond. IEEE Transactions on Image Processing **28**(1), 492–505 (2018)
31. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1833–1844 (2021)
32. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)
33. Liu, Z., Feng, R., Zhu, K., Zhang, Y., Zheng, K., Liu, Y., Zhao, D., Zhou, J., Cao, Y.: Cones: Concept neurons in diffusion models for customized generation. arXiv preprint arXiv:2303.05125 (2023)
34. Luo, Z., Gustafsson, F.K., Zhao, Z., Sjölund, J., Schön, T.B.: Image restoration with mean-reverting stochastic differential equations. arXiv preprint arXiv:2301.11699 (2023)
35. McCartney, E.J.: Optics of the atmosphere: scattering by molecules and particles. New York (1976)

36. Özdenizci, O., Legenstein, R.: Restoring vision in adverse weather conditions with patch-based denoising diffusion models. IEEE Transactions on Pattern Analysis and Machine Intelligence (2023)
37. Park, Y.H., Kwon, M., Choi, J., Jo, J., Uh, Y.: Understanding the latent space of diffusion models through the lens of riemannian geometry. Advances in Neural Information Processing Systems **36** (2024)
38. 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)
39. Qiu, Y., Zhang, K., Wang, C., Luo, W., Li, H., Jin, Z.: Mb-taylorformer: Multi-branch efficient transformer expanded by taylor formula for image dehazing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12802–12813 (2023)
40. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems **28** (2015)
41. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)
42. Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., Aberman, K.: Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22500–22510 (2023)
43. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. International journal of computer vision **115**, 211–252 (2015)
44. Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M.: Image super-resolution via iterative refinement. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(4), 4713–4726 (2022)
45. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
46. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: International conference on machine learning. pp. 2256–2265. PMLR (2015)
47. Song, J., Vahdat, A., Mardani, M., Kautz, J.: Pseudoinverse-guided diffusion models for inverse problems. In: International Conference on Learning Representations (2022)
48. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:2011.13456 (2020)
49. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
50. Wang, J., Yue, Z., Zhou, S., Chan, K.C., Loy, C.C.: Exploiting diffusion prior for real-world image super-resolution. arXiv preprint arXiv:2305.07015 (2023)
51. Wang, X., Yu, K., Dong, C., Loy, C.C.: Recovering realistic texture in image super-resolution by deep spatial feature transform. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 606–615 (2018)

52. Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., Li, H.: Uformer: A general u-shaped transformer for image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 17683–17693 (2022)
53. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)
54. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560 (2018)
55. Wu, H., Qu, Y., Lin, S., Zhou, J., Qiao, R., Zhang, Z., Xie, Y., Ma, L.: Contrastive learning for compact single image dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10551–10560 (2021)
56. Wu, W., Weng, J., Zhang, P., Wang, X., Yang, W., Jiang, J.: Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5901–5910 (2022)
57. Xia, B., Zhang, Y., Wang, S., Wang, Y., Wu, X., Tian, Y., Yang, W., Van Gool, L.: Diffir: Efficient diffusion model for image restoration. arXiv preprint arXiv:2303.09472 (2023)
58. Yang, W., Wang, W., Huang, H., Wang, S., Liu, J.: Sparse gradient regularized deep retinex network for robust low-light image enhancement. IEEE Transactions on Image Processing **30**, 2072–2086 (2021)
59. Yang, Y., Wang, C., Liu, R., Zhang, L., Guo, X., Tao, D.: Self-augmented unpaired image dehazing via density and depth decomposition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2037–2046 (2022)
60. Ye, T., Zhang, Y., Jiang, M., Chen, L., Liu, Y., Chen, S., Chen, E.: Perceiving and modeling density for image dehazing. In: European Conference on Computer Vision. pp. 130–145. Springer (2022)
61. Yi, X., Xu, H., Zhang, H., Tang, L., Ma, J.: Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. arXiv preprint arXiv:2308.13164 (2023)
62. Yu, H., Zheng, N., Zhou, M., Huang, J., Xiao, Z., Zhao, F.: Frequency and spatial dual guidance for image dehazing. In: European Conference on Computer Vision. pp. 181–198. Springer (2022)
63. Yue, Z., Wang, J., Loy, C.C.: Resshift: Efficient diffusion model for image super-resolution by residual shifting. arXiv preprint arXiv:2307.12348 (2023)
64. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5728–5739 (2022)
65. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14821–14831 (2021)
66. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3194–3203 (2018)
67. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018)

68. Zheng, Y., Zhan, J., He, S., Dong, J., Du, Y.: Curricular contrastive regularization for physics-aware single image dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5785–5794 (2023)