Density-Aware Diffusion Model for Efficient Image Dehazing

Ling Zhang[1,2] , Wenxu Bai[1] , Chunxia Xiao[†3]

[1]School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, China
[2]Hubei Key Laboratory of Intelligent Information Processing and Realtime Industrial System, Wuhan University of Science and Technology, Wuhan, China
[3]School of Computer Science, Wuhan University, Wuhan, China
zhling@wust.edu.cn, bench@wust.edu.cn, cxxiao@whu.edu.cn

Abstract

Existing image dehazing methods have made remarkable progress. However, they generally perform poorly on images with dense haze, and often suffer from unsatisfactory results with detail degradation or color distortion. In this paper, we propose a density-aware diffusion model (DADM) for image dehazing. Guided by the haze density, our DADM can handle images with dense haze and complex environments. Specifically, we introduce a density-aware dehazing network (DADNet) in the reverse diffusion process, which can help DADM gradually recover a clear haze-free image from a haze image. To improve the performance of the network, we design a cross-feature density extraction module (CDEModule) to extract the haze density for the image and a density-guided feature fusion block (DFFBlock) to learn the effective contextual features. Furthermore, we introduce an indirect sampling strategy in the test sampling process, which not only suppresses the accumulation of errors but also ensures the stability of the results. Extensive experiments on popular benchmarks validate the superior performance of the proposed method. The code is released in https://github.com/benchacha/DADM.

CCS Concepts

• *Computing methodologies* → *Image dehazing; Density-aware; Diffusion model;*

1. Introduction

Images captured in hazy environments often become blurred due to the interference of haze, affecting the visibility of the images. Through image dehazing techniques, we can significantly improve the clarity and visual quality of images, providing support for some related computer vision tasks, such as autonomous driving [CMC*24, CWC*23], surveillance systems [KDW*23, RCV*23, LZZ*24], object detection [CLS*18, LZ23, LLW*22], and image segmentation [ZSQ*17, CPK*17, CLZX21]. Clear images can improve the accuracy and effectiveness of these tasks. Therefore, image dehazing is a necessary and important task to enhance the quality of the image.

Traditional image dehazing methods are mostly based on prior knowledge, such as dark channel prior (DCP) [HST10] and color attenuation prior (CAP) [ZMS15]. However, these methods have poor generalization ability. Subsequently, deep learning technologies have made significant progress in image dehazing [QWB*20, WQL*21, CXJ*16, SHQD23, GYA*22], demonstrating better generalization performance. They can obtain good results when dealing with regular haze images, but the performance decreases when facing dense haze images. CNN-based methods [QWB*20,

WQL*21, DPX*20, CXJ*16, CRCK23] are limited by the finite receptive field and difficult to handle long-distance dependencies, resulting in haze residual or local distortion, as shown in Figure 1(d). While Transformer-based methods [SHQD23, GYA*22, QZW*23] can capture global information, they relatively weaken the utilization of local features. They may lead to the loss of features in regions with dense haze, generating results with detail distortion, as shown in Figure 1(e).

Recently, the diffusion model [SSDK*20] has been widely utilized in the field of image generation due to the strong ability for modeling the distribution of image pixels. However, since the existing diffusion models [LGZ*23, YHZ*23] do not take into account the variability of haze density in images, the straightforward application of these diffusion models for image dehazing may not yield the desired results, especially for dense haze images. As shown in Figure 1(f), artifacts or distortions may appear in the dehazing result, making the appearance of certain regions seem unrealistic.

To address the aforementioned issues, we introduce a density-aware diffusion model (DADM) for image dehazing, which uses the density of the haze as guidance information to help the network reconstruct a clear image. Figure 2 presents the framework of the proposed DADM. Based on the IR-SDE framework [LGZ*23], we train the diffusion model DADM to learn the mapping from haze images to haze-free images. In the reverse diffusion process in

---

[†] Corresponding author.

(a) Input   (b) Our result   (c) GT

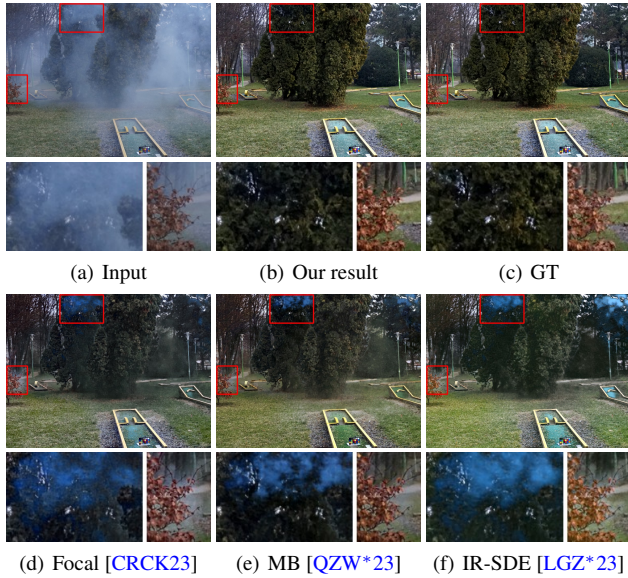(d) Focal [CRCK23]   (e) MB [QZW*23]   (f) IR-SDE [LGZ*23]

Figure 1: *Image dehazing. (a) is the haze image from NH-Haze dataset. Due to the dense haze in the image, existing image dehazing methods may suffer from haze residual or texture distortion, such as (d), (e) and (f). Considering the density of haze in the image, our method can produce a more desirable result (b).*

DADM, we propose a density-aware dehazing network (DADNet) to estimate the noise in the input image and remove the noise (haze layer) in the image. In the testing process, we take the haze image as input and gradually remove the haze through the pre-trained DADNet to get a haze-free image.

To improve the performance of DADNet, we introduce a cross-feature density extraction module (CDEModule) to extract the haze density in the image. We use the extracted haze density as auxiliary information to provide more precise and comprehensive guidance for DADNet. Furthermore, we propose a density-guided feature fusion block (DFFBlock) to fuse the features based on the haze density information. Specifically, we use a density-aware feature extraction unit (DFEUnit) to extract key features and a haze-aware feature optimization unit (DFEUnit) to learn effective contextual information for feature reconstruction.

During the testing process, the diffusion model gradually restores the haze image to a clear haze-free image based on the mapping learned during training. However, since the density of haze in images varies from scene to scene, the number of time steps required to restore these images to clarity will also vary. When the dehazing result reaches an optimal value during the testing process, continuing the denoising process may mistake the non-haze features as noise and remove them, resulting in a degradation of the final dehazing image, as shown in Figure 4. To address this challenge, we introduce an indirect sampling strategy, which effectively reduces errors in the sampling process and improves the accuracy and stability of the testing process.

In summary, our main contributions are as follows:

- We present a density-aware diffusion model (DADM) for image

dehazing, which takes full advantage of the haze density in the image and produce high-quality dehazing results.
- We introduce a density-aware dehazing network in the reverse diffusion process to remove the haze in the image, in which we design a CDEModule to extract the haze density and a DFFBlock to learn effective contextual information.
- We design an indirect sampling strategy to effectively reduce errors in the sampling process, which can improve the accuracy and stability of the testing process.

Experimental results on the four public datasets show that our proposed methods outperforms other SOTAs methods. Beside, our method shows better visual effects on images with dense haze and complex scenes.

2. Related Work

Traditional Dehazing Method. Traditional methods often rely on image priors and physical models to achieve image dehazing [HST10, ZMS15, JR13, XG12, LYX17]. Among them, the dark channel prior algorithm (DCP) [HST10] estimated the haze density by calculating the dark channel of the image, achieving the dehazing effect. Additionally, the color attenuation prior (CAP) [ZMS15] algorithm recovered the original image by estimating the depth of the scene. The histogram equalization algorithm [JR13] equalized the histogram of the input image, resulting in a uniform distribution of pixel values across the grayscale range. However, traditional methods have certain limitations, with relatively weak generalization capabilities. With the rapid development of deep learning techniques, methods based on neural networks have gradually become the mainstream, bringing significant advancements in image dehazing.

CNN-based Image Dehazing. Convolutional Neural Networks (CNNs) utilize deep convolutions to extract image features for image dehazing [CXJ*16, DPX*20, QWB*20, WQL*21, ZLZX23]. Early attempts such as DehazeNet [CXJ*16] focused on local features through end-to-end learning, laying the foundation for image dehazing. Subsequent methods like MSBDN [DPX*20] adapted better to different haze intensities and scenes through convolutions and deconvolutions, achieving significant progress. FFA-Net [QWB*20] introduced feature fusion and attention mechanisms, further improving the effect of image dehazing. AECR-Net [WQL*21] applied contrastive learning to image dehazing task, enhancing the haze removal capabilities. These methods show that with clever networks and effective feature utilization, CNN-based methods can effectively improve the quality and robustness of image dehazing. However, CNNs have difficulties in handling long-range dependencies, thus still exhibiting limitations when processing global information and complex scenes.

Transformer-based Image Dehazing. Recently, transformer models have also been introduced to handle image dehazing [ZAK*22, SHQD23, GYA*22, QZW*23], which can solve the problem of long-range dependencies. Restormer [ZAK*22] effectively captured the interactions between distant pixels in images through multi-head attention and feed-forward networks. DehazeFormer [SHQD23] proposed a new feature fusion scheme for image dehazing. Dehamer [GYA*22] combined CNN and transformer, using

the features extracted by transformer to modulate the features of CNN. MB-TaylorFormer [QZW*23] utilized Taylor expansion to achieve linear computational complexity, significantly improving computational efficiency. These transformer-based methods have improved the effect of image dehazing to some extent. However, when facing dense haze images, the features in heavy haze regions are disturbed, making the extracted global information inaccurate, resulting in color shift and image distortion in the results.

Diffusion-based Image Dehazing. As a powerful image generation framework, diffusion models [HJA20, FLP*23, WYYZ23, YZTL23] have received widespread attention. IR-SDE [LGZ*23] introduced the concept of mean regression stochastic differential equations (SDE) [SSDK*20], enhancing image restoration quality by probabilistically modeling adaptation to different noise and blur conditions. DehazeDDPM [YHZ*23] introduced conditional DDPM and worked with physical modeling to solve the task of dense haze image dehazing. GDP [FLP*23] leveraged a pre-trained denoising diffusion probabilistic model (DDPM) [HJA20] for high-fidelity recovery and enhancement in an unsupervised manner. However, when processing dense haze images with large differences in haze density, these methods often lead to incomplete haze removal due to their limited adaptability to changes in haze density. Thus, we utilize the density information of haze images to guide help the diffusion model removing shadows in the image.

3. Motivation

When the density of haze is dense, the fine textures, colors and contrast in the scene are often severely distorted, which significantly increases the difficulty of image dehazing. Existing dehazing methods are often difficult to accurately capture and recover the content in regions with dense haze, resulting in haze residue and texture degradation. Thus, we need to find more advanced and accurate haze removal algorithm to ensure that the detail and color information in the image can be maximally recovered while the haze is completely removed.

Based on the powerful denoising capability of the diffusion model [LCS*21, MFZ*23, FLP*23, YL24], we consider utilizing the diffusion model for the image dehazing. The diffusion model is a generative model that approximates the real data distribution by gradually adding noise (forward diffusion) and subsequently removing noise (reverse diffusion). In image dehazing, this step-by-step approximation assists in gradually removing the haze while preserving the details and textures of the image, generating clearer and more natural dehazing results.

Haze is composed of tiny water droplets or aerosol particles, which form haze layers of different density at different atmospheric depths. However, learning-based image dehazing methods tend to ignore the variability of haze exhibited between different regions, adopting instead an undifferentiated processing approach. When dealing with dense and varied haze images, these methods may lead to artifacts in the results, such as local over-enhancement or color distortion.

To more accurately model and remove haze layers from images, we propose a density-aware diffusion model (DADM) for image

dehazing. DADM utilizes the density of haze in the image as guiding information to help the network accurately achieve dehazing while maintaining the realism of the result. By introducing the density-aware mechanism, our DADM can handle various variable haze images, especially those with dense haze.

4. Diffusion Model for Image Dehazing

In this paper, we use IR-SDE [LGZ*23] framework to implement our DADM. For image dehazing, the haze image can be considered as noisy data, while the clear haze-free image is the target data. By training a diffusion model, we can learn a mapping from haze images to haze-free images. The diffusion model contains a forward diffusion process and a reverse diffusion process. In testing, we can take the haze image as input and gradually remove the haze through the reverse diffusion process of the model, ultimately obtaining a haze-free image.

Forward diffusion. The forward diffusion process gradually degrades a high-quality haze-free image x_0 into noisy data x_T by adding noise step by step, and x_t is the intermediate state at the time step $t \in \{1, \ldots, T\}$, as shown in Figure 2(a). In our task, x_T is the original haze image I_{haze} with white Gaussian noise $\epsilon \sim \mathcal{N}(1, \lambda^2)$. Based on the IR-SDE [LGZ*23], the forward diffusion process $\{x_t\}_0^T$ is that,

$$dx = \theta_t(\mu - x)dt + \sigma_t dw, \tag{1}$$

where θ_t and σ_t are parameters related to time step t, respectively representing the degradation rate and the size of the disturbance noise. w is a standard Brownian motion. μ is the original haze image.

According to IR-SDE, we can calculate the intermediate state x_t directly from the inital x_0, which can be expressed as:

$$x_t = \mu + (x_0 - \mu)e^{-\bar{\theta}_t} + \sqrt{v_t}\epsilon_t, \tag{2}$$

where ϵ_t is a standard Gaussian noise, and $\epsilon_t \sim \mathcal{N}(0, I)$. $\bar{\theta}_t = \int_0^t \theta_z \, dz$, and $v_t = \lambda^2(1 - e^{-2\bar{\theta}_t})$. To ensure a closed-form solution, IR-SDE sets $\sigma_t^2/\theta_t = 2\lambda^2$. λ^2 is a fixed variance, which is set to 50 in our experiments.

Based on Eq. (2), we can compute the result $\tilde{x}_0$ from x_t at any time, which is necessary in the sampling process we propose later, that is

$$\tilde{x}_0 = \frac{x_t - \sqrt{v_t}\tilde{\epsilon}_t - \mu}{e^{-\bar{\theta}_t}} + \mu. \tag{3}$$

where $\tilde{\epsilon}_t$ is the predicted noise produced during the reverse diffusion process.

Reverse diffusion. By simulating the reverse of the forward diffusion process, we can learn a conditional reverse diffusion process to denoise data from noisy data x_T to a less noise state x_t at time step t, as shown in Figure 2(a). The reverse diffusion process based on IR-SDE is that,

$$dx = \left[\theta_t(\mu - x) - \sigma_t^2 \nabla \log p_t(x)\right]dt + \sigma_t d\hat{w}, \tag{4}$$

where $\hat{w}$ represents the reverse-time standard Brownian motion. $\nabla \log p_t(x)$ is the ground-truth score function, given by $\nabla \log p_t(x) = -\frac{\epsilon_t}{\sqrt{v_t}}$.
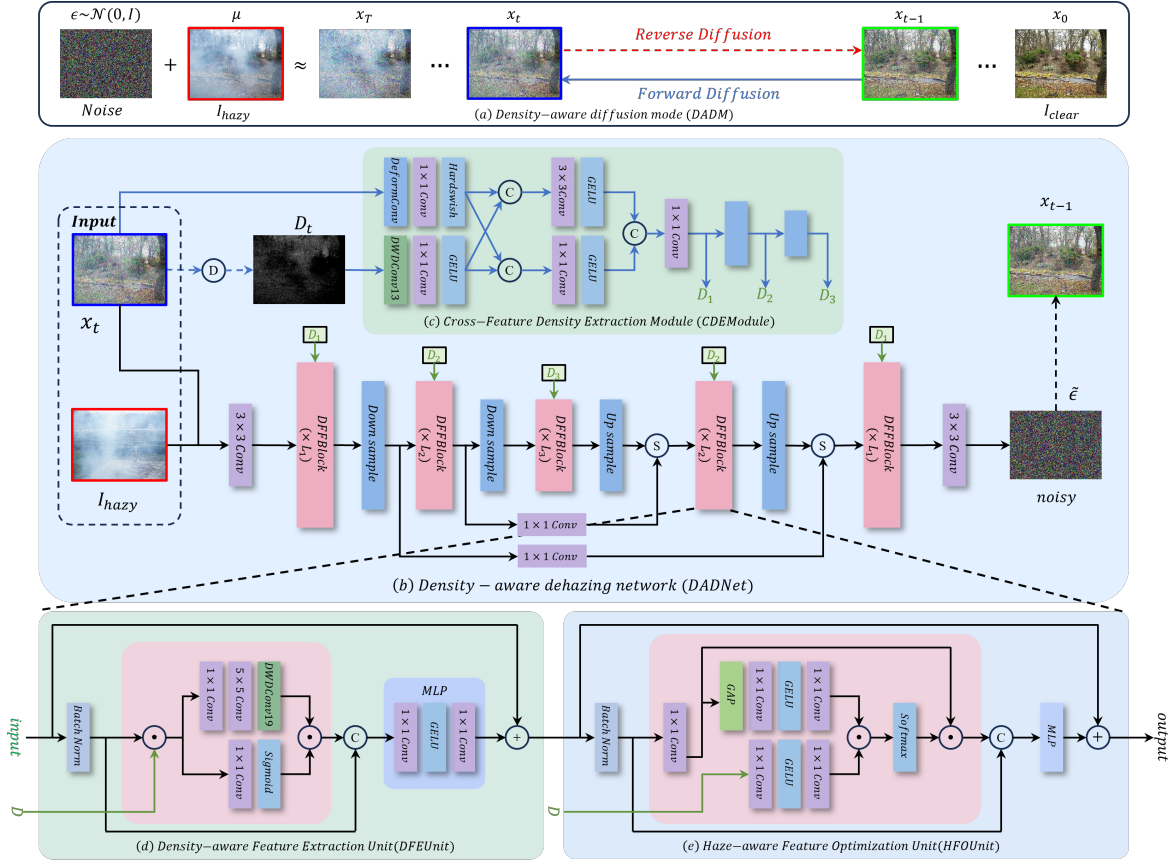
Figure 2: *Overview of our DADM. We recover haze-free images from haze images by adopting the IR-SDE model (a). In the reverse diffusion process, we introduce a density-aware dehazing network (DADNet) to estimate the noise from the input image and remove the haze in the image (b). The CDEModule (c) in DADNet is used to optimize the dark channel map and obtain the accurate haze density for the image, helping DADNet produce high-quality results. Moreover, we design a DFFBlock to fuse effective features and learn the contextual information for feature reconstruction, which contains a DFEUnit (d) and a HFOUnit (e).*

Then, we can calculate the intermediate sampling result $\tilde{x}_{t-1}$ from x_t:

$$
\begin{aligned}
\tilde{x}_{t-1} &= x_t - \mathrm{d}x_t \\
&= x_t - \left[\theta_t(\mu - x_t) - \sigma_t^2\left(-\frac{\tilde{\epsilon}_t}{\sqrt{v_t}}\right)\right]\mathrm{d}t + \sigma_t \mathrm{d}\hat{w},
\end{aligned}
\tag{5}
$$

Note that, we use the L1 loss between the intermediate state $\tilde{x}_{t-1}$ and the ideal state x_{t-1}^* to train the dehazing network, which is described as:

$$
L_{diff} = \mathbb{E}\left[\left\|\tilde{x}_{t-1} - x_{t-1}^*\right\|\right],
\tag{6}
$$

where x_{t-1}^* is the intermediate state at time step $t-1$ in the reverse diffusion process, and is calculated as,

$$
x_{t-1}^* = \frac{1 - e^{-2\bar{\theta}_{i-1}}}{1 - e^{-2\bar{\theta}_i}} e^{-\theta_i'}(x_t - \mu) + \frac{1 - e^{-2\theta_i'}}{1 - e^{-2\bar{\theta}_i}} e^{-\bar{\theta}_{i-1}}(x_0 - \mu) + \mu,
\tag{7}
$$

where $\theta_i' = \int_{i-1}^i \theta_t \mathrm{d}t = \theta_t \mathrm{d}t$.

5. Density-aware Dehazing Network

In the reverse diffusion process of our DADM, we introduce a density-aware dehazing network (DADNet) to estimate the noise in the input image and recover a haze-free image from a haze image (considered as a noisy state). Figure 2 (b) shows the framework of our DADNet, which is encoder-decoder structure. We use x_t and the original haze image I_{hazy} as the input of the dehazing network, and the output is the noise $\tilde{\epsilon}_t$ of x_t. The encoder and decoder are three-level structures, and each level consists of L density-guided feature fusion blocks (DFFBlocks). Each level of the decoder fuses features from the encoder through SK-Fusion [SHQD23]. We also introduce a cross-feature density extraction module (CDEModule) to extract the haze density information in the image, which uses x_t and its dark channel map as input. The extracted haze density information is used as auxiliary information to help DADNet better estimate the noise $\tilde{\epsilon}_t$ in the image.

5.1. Cross-Feature Density Extraction Module

In haze image, the density of haze is different in different regions. However, the current diffusion-based methods do not fully consider
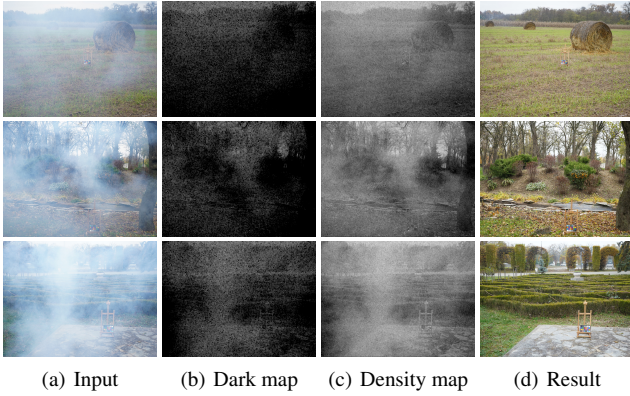
| (a) Input | (b) Dark map | (c) Density map | (d) Result |

Figure 3: *The haze density of the image. (b) is the dark channel map produced by DCP. (d) is our dehazing results by using the estimated density maps (c) .*

such variation in haze density, resulting in artifacts such as color distortion and detail degradation, as shown in Figure 1(f). To address this problem, we first estimate the haze density in the image. Then, our network can adaptively process regions based on the haze density.

The dark channel map [HST10] is usually interpreted as the haze density in the image. However, since we add noise to the image, the noise will also presents a very low pixel value in the dark channel map. Therefore, the dark channel value of noise areas is also very low (as shown in Figure 3(b)), resulting in the inability to extract accurate haze density information. Thus, the dark channel cannot be directly used as haze density to guide the network for dehazing.

To address above issues, we introduce a cross-feature density extraction module (CDEModule) to optimize the dark channel map and obtain the accurate haze density for the image, as shown in Figure 3(c). Our CDEModule is a dual-branch structure, with the image branch dedicated to extracting features from the input image, and the dark channel branch focused on extracting features from the dark channel map. To more accurately estimate the haze density, the two branches integrate the features of each other. In our experiments, we use DCP algorithm [HST10] to compute the dark channel map for x_t.

Figure 2(c) illustrates the architecture of the proposed CDEModule. The image branch use x_t as input, and the dark channel branch uses D_t as input. In the image branch, we first use a dynamic convolutional layer to extract the structural features of x_t. Then, we use a 1×1 convolutional layer to transform the features and enhance their expressive power. Next, we employ a hardswish activation function to obtain the preliminary representation B_1 of structural features. In the dark channel branch, due to the dark channel map often contains small local features, we first utilize a dilated convolution with a large convolution kernel to extract features. Then, we employ a 1×1 convolutional layer and a *GELU* function to get the preliminary representation B_2 of the dark channel features.

To optimize the utilization of features, we cross-utilize the features extracted from the two branches. We concatenate B_1 and B_2. Then, we use a 3×3 convolutional layer and a *GELU* function to obtain the final structural features $\bar{B}_3$, and employ a 1×1 convolu-

tional layer and a *GELU* function to obtain the final dark channel features $\bar{B}_4$. We concatenate B_3 and B_4, and utilize a 1×1 convolution to reduce the dimensionality and reconstruct the final haze density feature D for x_t.

To ensure the reverse process is effective, we perform two downsampling operations on D and obtain haze density information at three scales, namely D_1, D_2 and D_3. This ensures that the subsequent network can fully utilize haze density features at different scales, providing more precise and comprehensive guidance for the dehazing process.

5.2. Density-guided Feature Fusion Block

In real scenes, the density of haze in an image is often inconsistent. This inconsistency brings a great challenge in extracting local and global information, especially for dense haze. For example, inaccurate local information may result in artifacts with pseudo-structures. Inaccurate global information, on the other hand, may lead to color distortion and contrast imbalance, resulting in inconsistent appearance in the image.

To address the above problems, we introduce a density-guided feature fusion block (DFFBlock) to extract and fuse features from images. Our DFFBlock contains a density-aware feature extraction unit (DFEUnit) and a haze-aware feature optimization unit (HFOUnit). DFEUnit aims to suppresses noise and redundant information and select the key features in the image, improving the robustness of the model. HFOUnit is used to learn the effective contextual information for feature reconstruction.

Density-aware Feature Extraction Unit. To effectively reduce errors in the feature extraction caused by haze, we introduce a density-aware feature extraction unit (DFEUnit) to extract features. Particularly, we construct a branch that acts as an gating controller to filter the features. Through the gating mechanism, our DFEUnit selectively retains key information and eliminates some useless features like noise, thus significantly improving the effectiveness of feature extraction.

Figure 2(d) illustrates the architecture of our DFEUnit. We first perform batch normalization for the input feature F and obtain feature $\bar{F}$. Then, we perform element-wise multiplication on $\bar{F}$ and the extracted haze density feature D at the corresponding scale to obtain a mixed feature, which is fed into a two-branch structure that contains a feature extraction branch and a gated branch. The feature extraction branch applies a 1×1 convolution, a 5×5 convolution and a dilated convolution with dilation rate of 3 and a kernel size of 7×7 to extract features F_1. The gated branch uses a 1×1 convolution and a sigmoid function to obtain the corresponding gating information F_2. The gating information F_2 represents the information retention ratio at the corresponding position. We perform element-wise multiplication on F_1 and F_2 to control the feature output of the feature extraction branch. Thus, we can utilize the haze density to focus on important information and eliminate useless features.

We concatenated the output features from the two-branch structure with $\bar{F}$, and feed into a *MLP* module to get F_{MLP}. The *MLP* module contains a 1×1 convolution, a *GELU* activation function and a 1×1 convolution. It can perform dimensionality reduction

in the channel dimension to ensure that F_{MLP} have the same size as F. Finally, we perform element-wise addition on F_{MLP} and F to obtain the output feature F_{unit1} for DFEUnit, that is

$$F_{unit1} = MLP(concat(F_1 \times F_2, \bar{F})) + F, \tag{8}$$

where MLP represents the MLP module.

Haze-aware Feature Optimization Unit. Haze in the image can obscure some of the information in the scene. The region with dense haze is a damaged area, and the color and detail information in that region has been lost. We usually use contextual information to recover the content of these regions. However, low density haze only partially obscures the color and detail information of the scene. The features in these regions easily lead to feature bias when extracting global information, resulting in color or texture distortion in the haze-free results.

To solve this problem, we introduce a haze-aware feature optimization unit (HFOUnit) to extract effective contextual information from the image. To better learn the effective features in the image, we learn a haze-aware attention map, which represents the haze weights of the image. With the learned haze-aware attention map, HFOUnit can mitigate the effect of haze on image feature extraction and better capture long range dependencies in images. Our HFOUnit uses F_{unit1} produced by DFEUnit as input.

Figure 2(e) illustrates the pipeline of the proposed HFOUnit. Similar to DFEUnit, we first perform batch normalization for F_{unit1} to get feature $\bar{F}_{unit1}$. Then, we perform a convolution on $\bar{F}_{unit1}$ and get feature F_3. To improve the utilization of the features, we import F_3 into a channel attention unit to obtain a channel attention A_1, which can help focus on channels that are more important for the task. The channel attention unit consists of a global average pooling, a 1×1 convolution, a *ReLU* activation function and a 1×1 convolution. On the other hand, the haze density feature D are fed into a pixel attention unit to obtain pixel-level attention A_2. The pixel attention unit consists of a 1×1 convolution, a *ReLU* activation function and a 1×1 convolution. We perform element-wise multiplication on A_1 and A_2 to get the global attention A_{global} by using a softmax function. We perform element-wise multiplication on A_{global} and F_3 to obtain the haze-aware contextual features F_4, which is concatenated with $\bar{F}_{unit1}$. Next, we use the MLP module to optimize the concatenated features for fusion and dimension reduction.

Finally, we perform element-wise addition on the optimized features and F to obtain the output feature F_{unit2} for HFOUnit, that is

$$F_{unit2} = MLP(concat(A_{global} \times F_3, \bar{F}_{unit1})) + F. \tag{9}$$

6. Test Sampling Process

6.1. Problem

During the application of diffusion model, we usually predict a sequence of T time steps to gradually generate and refine the target image. During the testing process, we utilize the pre-trained model to perform step-by-step dehazing, and this process is set to undergo T iterations. However, considering the differences in the density
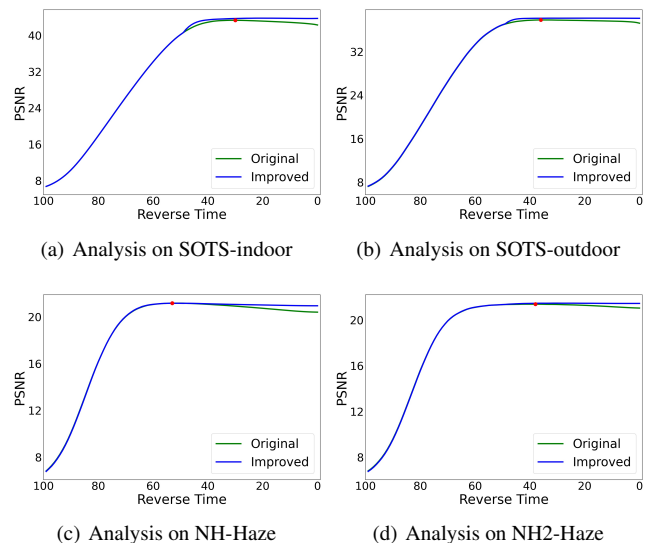

(a) Analysis on SOTS-indoor   (b) Analysis on SOTS-outdoor

(c) Analysis on NH-Haze   (d) Analysis on NH2-Haze

Figure 4: *Analysis of the test sampling process on four datasets. We compare the PSNR values of the recovered $\tilde{x}_0$ and I_{clear} at each time step, using both the original sampling method (green line) and our improved indirect sampling strategy (blue line). The higher the value is, the more similar $\tilde{x}_0$ and I_{clear} are. The red dot indicates the optimal value. We observe that our indirect sampling process is more stable in the second stage.*

and distribution of haze in different scenes, the number of dehazing steps (or denoising steps) required to achieve complete haze removal may vary for different images. Thus, there is a problem when using diffusion models in the testing process: **Is the final result optimal after T time steps of optimization?**
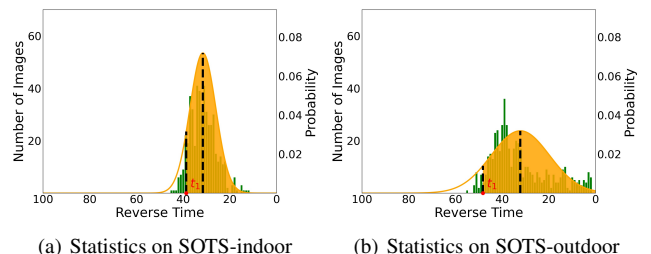

(a) Statistics on SOTS-indoor   (b) Statistics on SOTS-outdoor

Figure 5: *The statistics of PSNR between the recovered $\tilde{x}_0$ and I_{clear} at each time step on two datasets. The orange curve is the corresponding normal distribution curves based on their mean and variance. the area under the orange curve from t_1 to $t = 0$ is 90% of the total area.*

To delve the effectiveness of each time step in the dehazing process, we derive the intermediate haze removal results at each time step and evaluate them on four datasets. For a more intuitive illustration, we use the PSNR value to evaluste the quality of dehazing result at each step. As shown in Figure 4, we observe that the PSNR value peaks at a certain intermediate time step and then shows a gradual decreasing trend. This indicates that the dehazing results have achieved an optimum at some intermediate time point, while
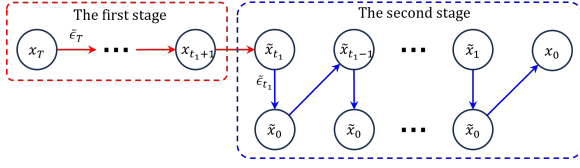
Figure 6: *Sampling process for testing. The first stage uses the general sampling strategy of the diffusion model. However, the second stage uses our indirect sampling strategy.*

the image quality may instead degrade as the dehazing process continues. So, when using the diffusion model for dehazing, the final result after T time steps may not be optimal.

The diffusion model is initially designed for image generation, and its final output lacks explicit labels as benchmarks, thus introducing a degree of randomness. However, for image dehazing task, we expect the generated haze-free images to be as close to the ground-truth as possible, with clear evaluation standards. Image dehazing task pursue the accuracy of the recovered images. After the diffusion model achieves optimal dehazing performance, if the denoising process continues, the network may treat non-haze features as noise and remove them. As the process continues, the accumulation of errors caused by excessive processing may lead to a gradual degradation of the results at a certain time step. As a result, after reaching the optimal value, the PSNR value shows a gradual decreasing trend in Figure 4.

In summary, although the denoising process of the diffusion model can improve the quality of the image, it cannot guarantee that the final dehazing result will be optimal.

6.2. Indirect Sampling Strategy

To reduce errors caused by excessive processing, we need to optimize the test sampling process, which aims to enhance the quality and stability of the final dehazing images. From Figure 4, we can observe that the sampling process is divided into two stages. In the first stage, the sampling results show a gradual trend of positive optimization. However, in the second stage, the sampling results may gradually decline after reaching the optimal value. To this end, we introduce an indirect sampling strategy to improve the sampling process in the second stage.

First, we need to determine the start time t_1 of the second stage. For a better analysis, we evaluate all the sampling results at various time steps using the PSNR values against the ground truth. Figure 5 statistics the PSNR distribution on SOTS-indoor and SOTS-outdoor datasets. Based on the mean value t_2 and the variance σ of the distribution, we can plot the corresponding normal distribution curves $c \sim \mathcal{N}(t_2, \sigma^2)$. Next, we set a criterion that the area under the corresponding normal distribution curve from t_1 to $t = 0$ (the shaded area in Figure 5) is 90% of the total area. We set it up this way to ensure that the time step of the optimal value is between t_1 to $t = 0$. Subsequently, we conduct a statistical count of t_1 on six test datasets. The t_1 values recorded for SOTS-indoor, SOTS-outdoor, NH-Haze, NH2-Haze, Rain100H and Rain100L are 38, 48, 50, 56, 65, and 49, respectively. In our experiments, we adopt the average value of the six values about t_1 as the value of t_1, that is $t_1 = 51$.
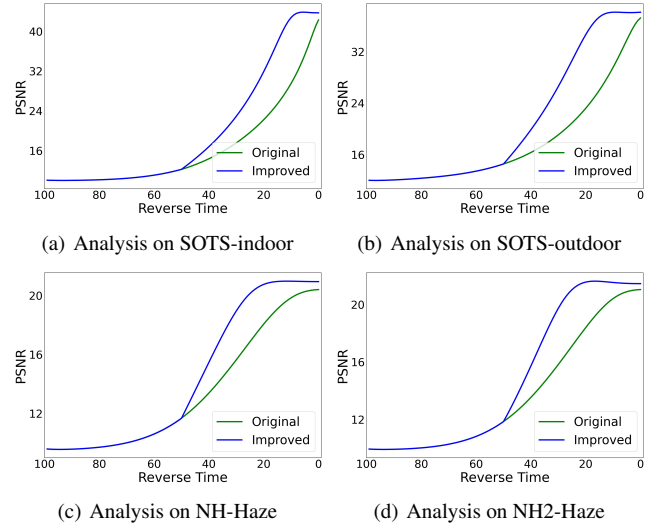
(a) Analysis on SOTS-indoor

(b) Analysis on SOTS-outdoor

(c) Analysis on NH-Haze

(d) Analysis on NH2-Haze

Figure 7: *Analysis of x_t on four datasets. We compare the PSNR values of x_t and I_{clear} at each step, using both the original sampling method (green line) and our improved indirect sampling strategy (blue line). At the same time step, the higher the value is, the closer x_t is to the ground-truth. The better values obtained by our strategy indicate that the indirect sampling strategy reduces the error during the sampling process.*

We use t_1 as the dividing time point. $t > t_1$ is the first stage, while $t \leq t_1$ is the second stage. These two stages use different sampling methods. As shown in Figure 6, when $t > t_1$, we use the original sampling way to obtain x_{t-1} by using Eq. (5) directly. When $t \leq t_1$, we first use the predicted noise $\tilde{\epsilon}$ and Eq. (3) to compute the current haze removal result $\tilde{x}_0$. Then we use Eq. (7) to get $\tilde{x}_{t-1}$ based on the computed $\tilde{x}_0$. This sampling process is shown in Algorithm 1.

To validate the effectiveness of our indirect sampling strategy, we conduct a comparative analysis of the sampling results obtained by the original sampling method and our improved sampling method. We evaluated $\{x_t\}_0^T$ acquired by both methods on four test datasets. As shown in Figure 7, we evaluate the PSNR vaules between x_t and ground-truth for each time step. From the results, we can clearly see that x_t obtained through the indirect sampling strategy exhibits superior performance in terms of PSNR. This result demonstrates that our improved indirect sampling method effectively reduces errors during the sampling process.

In addition, we evaluate the dehazing results x_0 for each time step based on Eq. (3), as shown in Figure 4. As can be observed that, the original sampling method shows a decreasing trend in PSNR value after reaching the optimal value. However, the improved sampling method exhibits a relatively smooth trend in the later part of the second stage, with more stable numerical values. This indicates that our indirect sampling strategy is more robust in obtaining the final dehazing results, improving the reliability of data sampling.

To sum up, with our indirect sampling strategy, we can not only effectively suppress the accumulation of errors but also ensure the accuracy of the results, delivering more realistic visual experience for users.

Table 1: *Quantitative comparisons of state-of-the-art dehazing methods on four datasets in terms of PSNR and SSIM. "-" indicates invalid, and the best and second-best results are highlighted in bold and underlined respectively.*

Methods	Venue&Year	SOTS-indoor		SOTS-outdoor		NH-Haze		NH-Haze2	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DCP [HST10]	TPAMI 2010	16.62	0.818	19.13	0.815	-	-	11.68	0.648
DehazeNet [CXJ*16]	TIP 2016	19.82	0.821	24.75	0.927	16.62	0.524	11.77	0.621
MSBDN [DPX*20]	CVPR 2020	33.67	0.985	33.48	0.982	19.23	0.706	20.11	0.800
FFA-Net [QWB*20]	AAAI 2020	36.39	0.989	33.57	0.984	14.39	0.452	20.00	0.823
AECR-Net [WQL*21]	CVPR 2021	37.17	0.990	-	-	19.88	0.717	20.68	0.828
DehazeFormer [SHQD23]	CVPR 2022	40.05	<u>0.996</u>	34.95	0.984	19.02	0.760	<u>21.19</u>	0.872
Dehamer [GYA*22]	CVPR 2022	36.63	0.988	35.18	0.986	<u>20.66</u>	0.684	19.18	0.794
SCANet [CTJK23]	IJCAI 2023	40.40	<u>0.996</u>	38.01	**0.995**	-	-	-	-
FocalNet [CRCK23]	ICCV 2023	40.82	<u>0.996</u>	37.71	**0.995**	20.43	<u>0.790</u>	20.59	0.874
C2P-Net [ZZH*23]	CVPR 2023	42.56	0.995	36.68	0.990	-	-	<u>21.19</u>	0.833
MB-TaylorFormer [QZW*23]	ICCV 2023	<u>42.64</u>	0.994	<u>38.09</u>	0.991	17.24	0.608	19.33	0.675
IR-SDE [LGZ*23]	ICML 2023	32.80	0.986	-	-	19.40	0.585	19.12	0.754
DIACMPN [ZZL24]	CVPR 2024	42.18	**0.997**	36.56	0.993	-	-	-	-
Original sampling	PG 2024	42.28	**0.997**	37.23	<u>0.994</u>	20.40	0.775	21.04	<u>0.879</u>
Improved sampling	PG 2024	**43.71**	**0.997**	**38.13**	0.995	**20.94**	**0.809**	**21.45**	**0.893**

Algorithm 1 Indirect sampling method

1: **Input**: haze image μ, denoising network $\tilde{\epsilon}_\phi$, number of implicit sampling steps T, start time t_1, and initial parameters ϕ
2: Sample $\epsilon \sim \mathcal{N}(0, I)$
3: **for** $t = T$ to 0 **do**
4:   **if** $t \le t_1$ **then**
5:     $\tilde{\epsilon}_t = \tilde{\epsilon}_\phi(x_t, \mu, t)$
6:     $\tilde{x}_0 = \frac{x_t - \sqrt{v_t}\tilde{\epsilon}_t - \mu}{e^{-\bar{\theta}_t}} + \mu$
7:     $\tilde{x}_{t-1} = \frac{1 - e^{-2\bar{\theta}_t - 1}}{1 - e^{-2\bar{\theta}_i}} e^{-\theta'_i}(x_t - \mu) + \frac{1 - e^{-2\theta'_t}}{1 - e^{-2\bar{\theta}_i}} e^{-\bar{\theta}_{i-1}}(\tilde{x}_0 - \mu) + \mu$
8:   **else**
9:     $\tilde{\nabla} \log p_t(x) = \frac{\tilde{\epsilon}_t}{\sqrt{v_t}}$
10:     $x_{t-1} = x_t - \left[\theta_t(\mu - x_t) - \sigma_t^2 \tilde{\nabla} \log p_t(x)\right] \mathrm{d}t$
11:   **end if**
12: **end for**
13: **return** x_0

---

7. Experiments

7.1. Experiment Settings

Implementation Details. Our method is implemented in PyTorch, which is trained on a NVIDIA GeForce RTX 3090. The model is trained using the Lion optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.99$. We use the learning rate adjustment strategy of cosine annealing, and the initial learning rate is set to $4e-5$ and gradually decays to $2e-6$. The total time steps T is set to 100. Furthermore, we augment the training data using random horizontal flipping and random rotation.

Dataset. We train our model and evaluate the performance on ITS, OTS [LRF*18], NH-Haze and NH2-Haze datasets [AAVT21]. ITS contains 13990 image pairs and OTS contains 313950 image pairs. Both SOTS-indoor and SOTS-outdoor datasets [LRF*18] contain 500 image pairs, and they are synthetic datasets. NH-Haze and NH-Haze2 datasets are real datasets with large differences in haze density. NH-Haze dataset contains 25 image pairs and NH-Haze contains 55 image pairs.

Metric. We utilize the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) [WBSS04] between the produced image and the ground-truth to assess the performance of the method.

7.2. Comparison with State-of-the-arts

Quantitative Comparison. To verify the effectiveness of our method, we compare our method with various state-of-the-art dehazing methods, such as DCP [HST10], DehazeNet [CXJ*16], MSBDN [DPX*20], FFA-Net [QWB*20], AECR-Net [WQL*21], SCANet [CTJK23], Focal-Net [CRCK23] and C2P-Net [ZZH*23], DehazeFormer [SHQD23], Dehamer [GYA*22], MB-TaylorFormer [QZW*23], and IR-SDE [LGZ*23]. Table 1 concludes the comparison results on four datasets. Compared with other methods, our proposed method achieves the best performance for all metrics on the four datasets, clearly demonstrating the effectiveness of our method.

Visual Comparison. Figure 8 provides some visual dehazing results on synthetic haze images. It can be seen, when conditions in the image do not satisfy the dark channel assumption, DCP [HST10] results in poor dehaze results, as shown in Figure 8(b). By not considering the utilization of global features, FFA-Net [QWB*20] and AECR-Net [WQL*21] may have dehazed some non-haze features as well, such as dust on the floor, as shown in Figure 8(d, e). MSBDN [DPX*20] and Dehamer [GYA*22] may appear to over-process haze in some areas, resulting in color artifacts, as shown in Figure 8(c, f). However, our method effectively removes haze in the image without artifacts, as shown in Figure 8(g).

Figure 9 provides some visual dehazing results on real images to further verify the superiority of our method. The inaccurate atmospheric light may lead to color deviations in the results of DCP [HST10], as shown in Figure 9(b). Due to the limitation of CNN, AECR [WQL*21] may produce color distortion in regions with dense haze, as shown in Figure 9(c). The results of DehazeFormer
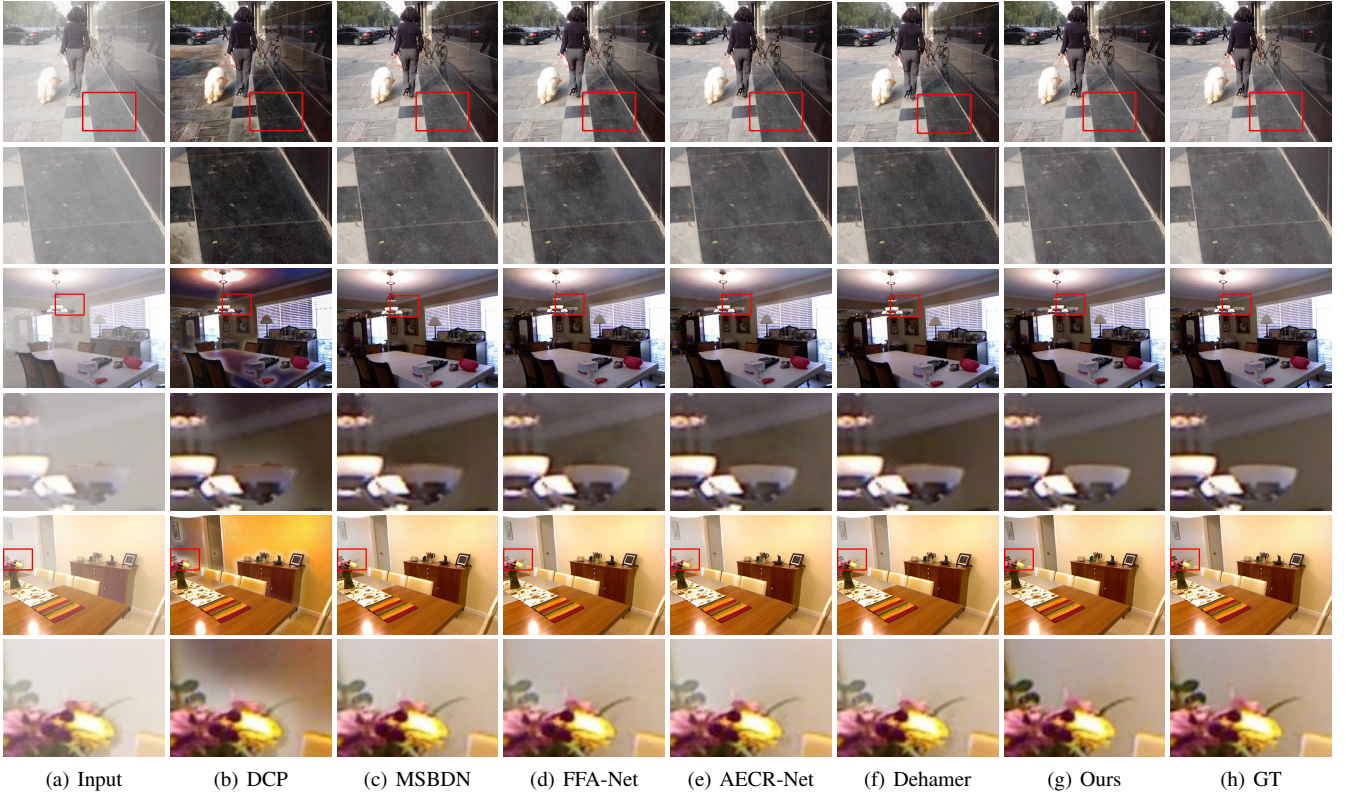
| (a) Input | (b) DCP | (c) MSBDN | (d) FFA-Net | (e) AECR-Net | (f) Dehamer | (g) Ours | (h) GT |

Figure 8: *Visual comparison with state-of-the-art dehazing methods on synthetic images.*

[SHQD23], Dehamer [GYA*22] and MB-Taylorformer [QZW*23] may have haze residue for dense haze images as the transformer-based structures are unable to extract the accurate structural information in regions with dense haze, as shown in Figure 9(d, e, g). Focal-Net [CRCK23] often suffer from haze residual for dense haze images, as shown in Figure 9(f). Comparatively, with the guidance of the density of haze, our results have little artifacts and are similar to the ground-truth images, as shown in Figure 9(i).

To verify the performance of our DADM, we retrain the diffusion model IR-SDE [LGZ*23] on our training dataset. Figure 11 gives some visual comparison. From the results, we can observe that, since IR-SDE does not take into account the density changes of haze in the image, it leads to haze residual in the results. Comparatively, results produced by our method have little artifacts.

7.3. Ablation Study

Ablation of Density-aware Dehazing Network. To further evaluate the performance of each component in our density-aware dehazing network in DADM, we conduct ablation experiments on two datasets using six variants (with or without specific module). These seven variations are as follows:
(1) $DADM_1$: replace CDEModule with a 3×3 convolution.
(2) $DADM_2$: without haze density as guidance.
(3) $DADM_3$: replace the dark channel map with gray level image as the input of CDEModule.
(4) $DADM_4$: replace DFFBlock with BatchNorm+Conv+GELU+

Conv.
(5) $DADM_5$: DFFBlock without DFEUnit.
(6) $DADM_6$: DFFBlock without HFOUnit.
(7) $DADM_7$: replace the density map with the dark channel map.

We train the six variants on the ITS dataset and OTS dataset using the same training strategy and test on SOTS-indoor and SOTS-outdoor datasets. The results are smmarized in Table 2. It can be observed from the results that, (1) the guidance of haze density can help improve the performance of the method and our CDEModule is effective; (2) DFEUnit and HFOUnit are necessary to ensure the high-quality dehazing results; (3) all the components together produces the best results. We also provide some visual results in Figure 10, from which we can observed our DADNet with all the components achieves better performance and is more realistic in terms of details.

Ablation of Indirect Sampling Strategy. To verify the effectiveness of the proposed indirect sampling strategy, we use the indirect sampling strategy and the original sampling strategy to test our DADM and IR-SDE [LGZ*23], respectively. Table 3 summarizes the comparison results. From the results, we can observe that, IR-SDE and our DADM using our indirect sampling strategy both obtain better values that that using the original sampling strategy. These results show that our indirect sampling strategy is useful.

In our indirect sampling strategy, we set the start time $t_1 = 51$ to determine the second stage. To verify the effectiveness of this value, we apply four different values of t_1 to implement the indirect

Zhang Ling & Bai Wenxu & Xiao Chunxia / Density-Aware Diffusion Model for Efficient Image Dehazing

|(a) Input | (b) DCP | (c) AECR | (d) DehazeFormer | (e) Dehamer | (f) Focal | (g) MB | (h) Ours | (i) GT |

Figure 9: *Visual comparison with state-of-the-art dehazing methods on real images.*

| (a) Input | (b) $DADM_1$ | (c) $DADM_2$ | (d) $DADM_3$ | (e) $DADM_4$ | (f) $DADM_5$ | (g) $DADM_6$ | (h) $DADM_7$ | (i) Ours | (j) GT |
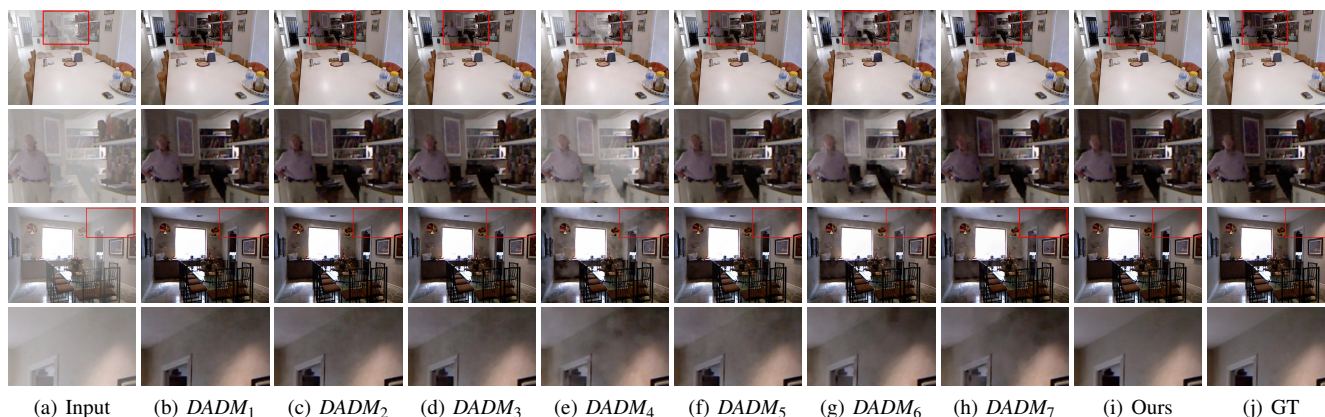
Figure 10: *Visual comparison for ablation study.*

(a) Input   (b) IR-SDE   (c) Our result   (d) GT
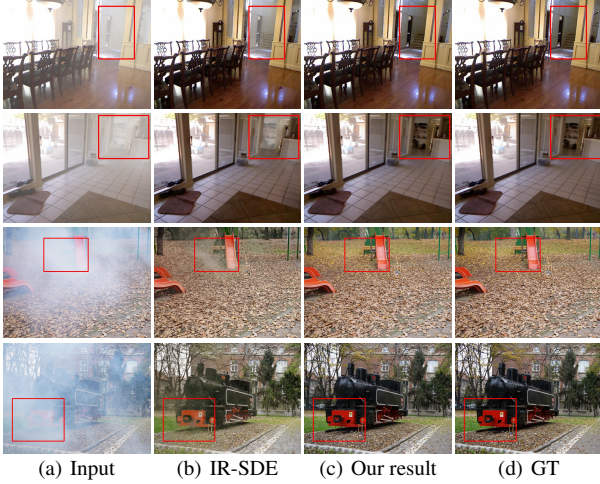
Figure 11: *Visual comparison with IR-SDE.*

Table 2: *Quantitative results of ablation study on SOTS-indoor and SOTS-outdoor datasets.*

Methods	SOTS-indoor		SOTS-outdoor	
	PSNR	SSIM	PSNR	SSIM
$DADM_1$	39.93	0.993	35.51	0.993
$DADM_2$	39.31	0.994	35.13	0.992
$DADM_3$	40.54	0.994	36.44	0.993
$DADM_4$	26.76	0.940	28.19	0.989
$DADM_5$	39.03	0.992	35.43	0.993
$DADM_6$	28.58	0.957	30.57	0.990
$DADM_7$	37.26	0.983	34.62	0.987
Our method	**42.28**	**0.997**	**37.23**	**0.994**

sampling strategy. Table 2 concludes the results, from which we can observe that our method obtain the best values when $t_1 = 51$.

Limitation. Our method can effective remove haze in the image. However, when the whole image contains very dense haze, our method has difficulty in completely removing haze from the image due to the inability to extract effective features, as shown in the Figure 12.
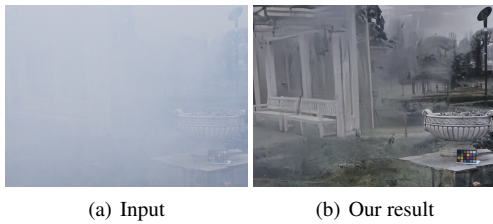

(a) Input   (b) Our result

Figure 12: *Limitation.*

8. Conclusion

In this paper, we propose a novel density-aware diffusion model (DADM) for image dehazing, which utilizes the density of the haze as guidance information to help the network reconstruct a clear image. Our DADM takes the haze image as input and gradually re-

Table 3: *Quantitative comparisons using different test sampling strategy. +Original represents the method using the original sampling strategy, and +Improved represents the method using the proposed indirect sampling strategy.*

methods	Rain100L		Rain100H	
	PSNR	SSIM	PSNR	SSIM
IR-SDE + Original	38.30	0.985	31.65	0.904
IR-SDE + Improved	38.59	0.985	32.32	0.927
Our DADM + Original	38.45	0.985	31.86	0.911
Our DADM + Improved	**38.72**	**0.986**	**32.52**	**0.934**

Table 4: *Quantitative results of ablation study about t_1 in the test sampling process.*

t_1	SOTS-indoor		SOTS-outdoor		NH-Haze		NH2-Haze	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
71	43.57	**0.997**	37.92	**0.995**	20.34	0.777	21.13	0.882
61	43.65	**0.997**	38.02	**0.995**	20.78	0.787	21.37	0.890
41	**43.73**	**0.997**	37.98	**0.995**	20.90	0.798	21.32	0.889
31	43.57	**0.997**	37.64	0.994	20.86	0.790	21.24	0.886
51	43.71	**0.997**	**38.13**	**0.995**	**20.94**	**0.809**	**21.45**	**0.893**

move the haze through the pre-trained DADNet to get a haze-free image. In the reverse diffusion process in DADM, we propose a density-aware dehazing network (DADNet) to estimate the noise in the input image and remove the noise (haze layer). To improve the performance of the network, we also propose a density-guided feature fusion block (DFFBlock) to learn effective contextual information for feature reconstruction. Furthermore, we introduce an indirect sampling strategy in the test sampling process, which not only suppress the accumulation of errors but also ensure the stability of the results. Extensive experiments demonstrate the superiority of the proposed method.

In the future, we would like to optimize our model to better adapt to varying degrees of haze density. In addition, we will investigate more efficient algorithms to reduce computational time and resource consumption, striving for a more optimized solution.

Acknowledgments

References

[AAVT21] ANCUTI C. O., ANCUTI C., VASLUIANU F.-A., TIMOFTE R.: Ntire 2021 nonhomogeneous dehazing challenge report. In *CVPR* (2021), pp. 627–646. 8

[CLS*18] CHEN Y., LI W., SAKARIDIS C., DAI D., VAN GOOL L.: Domain adaptive faster r-cnn for object detection in the wild. In *CVPR* (2018), pp. 3339–3348. 1

[CLZX21] CHEN Z., LONG C., ZHANG L., XIAO C.: Canet: A context-aware network for shadow removal. In *ICCV* (2021). 1

[CMC*24] CUI C., MA Y., CAO X., YE W., ZHOU Y., LIANG K., CHEN J., LU J., YANG Z., LIAO K.-D., ET AL.: A survey on multimodal large language models for autonomous driving. In *IEEE/CVF*

Winter Conference on Applications of Computer Vision (2024), pp. 958–979. 1

[CPK*17] CHEN L.-C., PAPANDREOU G., KOKKINOS I., MURPHY K., YUILLE A. L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence 40*, 4 (2017), 834–848. 1

[CRCK23] CUI Y., REN W., CAO X., KNOLL A.: Focal network for image restoration. In *ICCV* (2023), pp. 13001–13011. 1, 2, 8, 9

[CTJK23] CUI Y., TAO Y., JING L., KNOLL A.: Strip attention for image restoration. In *International Joint Conference on Artificial Intelligence, IJCAI* (2023), vol. 2. 8

[CWC*23] CHEN L., WU P., CHITTA K., JAEGER B., GEIGER A., LI H.: End-to-end autonomous driving: Challenges and frontiers. *arXiv preprint arXiv:2306.16927* (2023). 1

[CXJ*16] 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 (2016), 5187–5198. 1, 2, 8

[DPX*20] 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 *CVPR* (2020), pp. 2157–2167. 1, 2, 8

[FLP*23] 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 *CVPR* (2023), pp. 9935–9946. 3

[GYA*22] GUO C.-L., YAN Q., ANWAR S., CONG R., REN W., LI C.: Image dehazing transformer with transmission-aware 3d position embedding. In *CVPR* (2022), pp. 5812–5820. 1, 2, 8, 9

[HJA20] HO J., JAIN A., ABBEEL P.: Denoising diffusion probabilistic models. *Advances in neural information processing systems 33* (2020), 6840–6851. 3

[HST10] HE K., SUN J., TANG X.: Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence 33*, 12 (2010), 2341–2353. 1, 2, 5, 8

[JR13] JUN W. L., RONG Z.: Image defogging algorithm of single color image based on wavelet transform and histogram equalization. *Applied mathematical sciences 7*, 79 (2013), 3913–3921. 2

[KDW*23] KESHAVIAH A., DIAMOND M. B., WADE M. J., SCARPINO S. V., AHMED W., AMMAN F., ARUNA O., BADILLA-AGUILAR A., BAR-OR I., BERGTHALER A., ET AL.: Wastewater monitoring can anchor global disease surveillance systems. *The Lancet Global Health 11*, 6 (2023), e976–e981. 1

[LCS*21] LIANG J., CAO J., SUN G., ZHANG K., VAN GOOL L., TIMOFTE R.: Swinir: Image restoration using swin transformer. In *ICCV* (2021), pp. 1833–1844. 3

[LGZ*23] 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). 1, 2, 3, 8, 9

[LLW*22] LIU Z., LI Y., WANG W., LIU L., CHEN R.: Mesh total generalized variation for denoising. *IEEE Transactions on Visualization and Computer Graphics 28*, 12 (2022), 4418–4433. 1

[LRF*18] 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 (2018), 492–505. 8

[LYX17] LIAO B., YIN P., XIAO C.: Efficient image dehazing using boundary conditions and local contrast. *Computers Graphics 70*, feb. (2017), 242–250. 2

[LZ23] LING ZHANG YINGHAO HE Q. Z. Z. L. X. Z. C. X.: Document image shadow removal guided by color-aware background. In *CVPR* (2023). 1

[LZZ*24] LIU Z., ZHAO Y., ZHAN S., LIU Y., CHEN R., HE Y.: Pcdnf: Revisiting learning-based point cloud denoising via joint normal filtering. *IEEE Transactions on Visualization and Computer Graphics 30*, 8 (2024), 5419–5436. 1

[MFZ*23] MEI Y., FAN Y., ZHANG Y., YU J., ZHOU Y., LIU D., FU Y., HUANG T. S., SHI H.: Pyramid attention network for image restoration. *Computer Vision 131*, 12 (2023), 3207–3225. 3

[QWB*20] QIN X., WANG Z., BAI Y., XIE X., JIA H.: Ffa-net: Feature fusion attention network for single image dehazing. In *AAAI* (2020), vol. 34, pp. 11908–11915. 1, 2, 8

[QZW*23] 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 *ICCV* (2023), pp. 12802–12813. 1, 2, 3, 8, 9

[RCV*23] RAWAT R., CHAKRAWARTI R. K., VYAS P., GONZÁLES J. L. A., SIKARWAR R., BHARDWAJ R.: Intelligent fog computing surveillance system for crime and vulnerability identification and tracing. *International Journal of Information Security and Privacy (IJISP) 17*, 1 (2023), 1–25. 1

[SHQD23] SONG Y., HE Z., QIAN H., DU X.: Vision transformers for single image dehazing. *IEEE Transactions on Image Processing 32* (2023), 1927–1941. 1, 2, 4, 8, 9

[SSDK*20] 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). 1, 3

[WBSS04] 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 (2004), 600–612. 8

[WQL*21] WU H., QU Y., LIN S., ZHOU J., QIAO R., ZHANG Z., XIE Y., MA L.: Contrastive learning for compact single image dehazing. In *CVPR* (2021), pp. 10551–10560. 1, 2, 8

[WYYZ23] WANG Y., YU J., YU R., ZHANG J.: Unlimited-size diffusion restoration, 2023. 3

[XG12] XIAO C., GAN J.: Fast image dehazing using guided joint bilateral filter. *Visual Computer 28*, 6-8 (2012), 713–721. 2

[YHZ*23] YU H., HUANG J., ZHENG K., ZHOU M., ZHAO F.: High-quality image dehazing with diffusion model. *arXiv preprint arXiv:2308.11949* (2023). 1, 3

[YL24] YUANZHEN LI FEI LUO C. X.: Diffusion-fof: Single-view clothed human reconstruction via diffusion-based fourier occupancy field. In *CVPR* (2024). 3

[YZTL23] YANG P., ZHOU S., TAO Q., LOY C. C.: Pgdiff: Guiding diffusion models for versatile face restoration via partial guidance, 2023. 3

[ZAK*22] ZAMIR S. W., ARORA A., KHAN S., HAYAT M., KHAN F. S., YANG M.-H.: Restormer: Efficient transformer for high-resolution image restoration. In *CVPR* (2022), pp. 5728–5739. 2

[ZLZX23] ZHANG L., LONG C., ZHANG X., XIAO C.: Exploiting residual and illumination with gans for shadow detection and shadow removal. *ACM transactions on multimedia computing communications and applications 19*, 3 (2023), 120.1–120.22. 2

[ZMS15] ZHU Q., MAI J., SHAO L.: A fast single image haze removal algorithm using color attenuation prior. *IEEE transactions on image processing 24*, 11 (2015), 3522–3533. 1, 2

[ZSQ*17] ZHAO H., SHI J., QI X., WANG X., JIA J.: Pyramid scene parsing network. In *CVPR* (2017), pp. 2881–2890. 1

[ZZH*23] ZHENG Y., ZHAN J., HE S., DONG J., DU Y.: Curricular contrastive regularization for physics-aware single image dehazing. In *CVPR* (2023), pp. 5785–5794. 8

[ZZL24] ZHANG Y., ZHOU S., LI H.: Depth information assisted collaborative mutual promotion network for single image dehazing. *arXiv preprint arXiv:2403.01105* (2024). 8