RIDCP: Revitalizing Real Image Dehazing via High-Quality Codebook Priors

Rui-Qi Wu[1] Zheng-Peng Duan[1] Chun-Le Guo[1*] Zhi Chai[2] Chongyi Li[3]
[1]VCIP, CS, Nankai University [2]Hisilicon Technologies Co. Ltd.
[3]S-Lab, Nanyang Technological University
{wuruiqi, adamduan0211}@mail.nankai.edu.cn, guochunle@nankai.edu.cn,
chaizhi2@huawei.com, chongyi.li@ntu.edu.sg

Abstract

*Existing dehazing approaches struggle to process real-world hazy images owing to the lack of paired real data and robust priors. In this work, we present a new paradigm for real image dehazing from the perspectives of synthesizing more realistic hazy data and introducing more robust priors into the network. Specifically, (1) instead of adopting the de facto physical scattering model, we rethink the degradation of real hazy images and propose a phenomenological pipeline considering diverse degradation types. (2) We propose a **R**eal **I**mage **D**ehazing network via high-quality **C**odebook **P**riors (RIDCP). Firstly, a VQGAN is pre-trained on a large-scale high-quality dataset to obtain the discrete codebook, encapsulating high-quality priors (HQPs). After replacing the negative effects brought by haze with HQPs, the decoder equipped with a novel normalized feature alignment module can effectively utilize high-quality features and produce clean results. However, although our degradation pipeline drastically mitigates the domain gap between synthetic and real data, it is still intractable to avoid it, which challenges HQPs matching in the wild. Thus, we re-calculate the distance when matching the features to the HQPs by a controllable matching operation, which facilitates finding better counterparts. We provide a recommendation to control the matching based on an explainable solution. Users can also flexibly adjust the enhancement degree as per their preference. Extensive experiments verify the effectiveness of our data synthesis pipeline and the superior performance of RIDCP in real image dehazing. Code and data are released at https://rq-wu.github.io/projects/RIDCP.*

1. Introduction

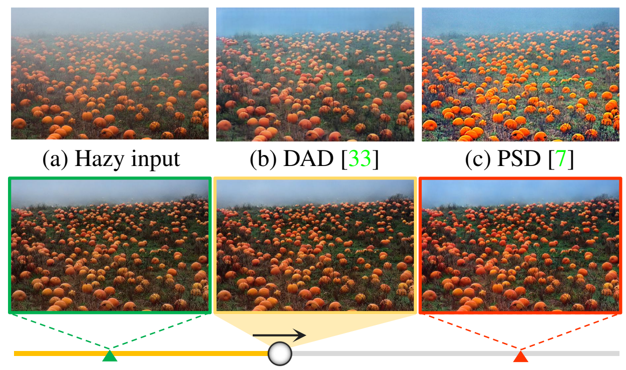
Image dehazing aims to recover clean images from their hazy counterparts, which is essential for computational photography and high-level tasks [20, 32]. The hazy image formulation is commonly described by a physical scattering model:

$$I(x) = J(x)t(x) + A(1 - t(x)), \tag{1}$$

where $I(x)$ denotes the hazy image and $J(x)$ is its corresponding clean image. The variables A and $t(x)$ are the global atmosphere light and transmission map, respectively. The transmission map $t(x) = e^{\beta d(x)}$ depends on scene depth $d(x)$ and haze density coefficient β.

Given a hazy image, restoring its clean version is highly ill-posed. To mitigate the ill-posedness of this problem, various priors, e.g., dark channel prior [16], color attenuation prior [44], and color lines [12] have been proposed in existing traditional methods. Nevertheless, the statistical priors cannot cover diverse cases in real-world scenes, leading to suboptimal dehazing performance.

With the advent of deep learning, image dehazing has achieved remarkable progress. Existing methods either adopt deep networks to estimate physical parameters [5, 21, 31] or directly restore haze-free images [10, 15, 27, 30, 40]. However, image dehazing neural networks perform limited generalization to real scenes, owing to the difficulty in



(a) Hazy input (b) DAD [33] (c) PSD [7]

(d) Ours

Figure 1. Visual comparisons on a typical hazy image. The proposed method generates cleaner results than other two state-of-the-art real image dehazing approaches. The enhancement degree of our result can be flexibly adjusted by adopting different parameters in the real-domain adaptation phase. The image with a golden border is the result obtained under our recommended parameter.

*Corresponding author

collecting large-scale yet perfectly aligned paired training data and *solving the uncertainty of the ill-posed problem without robust priors* . Concretely, 1) collecting large-scale and perfectly aligned hazy images with the clean counterpart is incredibly difficult, if not impossible. Thus, most of the existing deep models use synthetic data for training, in which the hazy images are generated using Eq. (1), leading to the neglect of multiple degradation factors. There are some real hazy image datasets [2, 3] with paired data, but the size and diversity are insufficient. Moreover, these datasets deviate from the hazy images captured in the wild. These shortcomings inevitably decrease the capability of deep models in real scenes. 2) Real image dehazing is a highly ill-posed issue. Generally, addressing an uncertain mapping problem often needs the support of priors. However, it is difficult to obtain robust priors that can cover the diverse scenes of real hazy images, which also limits the performance of dehazing algorithms. Recently, many studies for real image dehazing try to solve these two issues by domain adaptation from the perspective of data generation [33,39] or priors guidance [7,23], but still cannot obtain desirable results.

In this work, we present a new paradigm for real image dehazing motivated by addressing the above two problems. To obtain large-scale and perfectly aligned paired training data, we rethink the degradation of hazy images by observing amounts of real hazy images and propose a novel data generation pipeline considering multiple degradation factors. In order to solve the uncertainty of the ill-posed issue, we attempt to train a VQGAN [11] on high-quality images to extract more robust high-quality priors (HQPs). The VQGAN only learns high-quality image reconstruction, so it naturally contains the robust HQPs that can help hazy features jump to the clean domain. The observation in Sec. 4.1 further verifies our motivation. Thus, we propose the **R**eal **I**mage **D**ehazing network via high-quality **C**odebook **P**riors (RIDCP). The codebook and decoder of VQGAN are fixed to provide HQPs. Then, RIDCP is equipped with an encoder that helps find the correct HQPs, and a new decoder that utilizes the features from the fixed decoder and produces the final result. Moreover, we propose a novel Normalized Feature Alignment (NFA) that can mitigate the distortion and balance the features for better fusion.

In comparison to previous methods [6, 14, 43] that introduce codebook for image restoration, we further design a unique real domain adaptation strategy based on the characteristics of VQGAN and the statistical results. Intuitively, we propose Controllable HQPs Matching (CHM) operation that replaces the nearest-neighbour matching by imposing elaborate-designed weights on the distances between features and HQPs during the inference phase. The weights are determined by a controllable parameter α and the statistical distribution gap of HQPs activation in Sec. 4.3. By adjusting α, the distribution of HQPs activation can be shifted. Moreover, we present a theoretically feasible solution to obtain the optimal α by minimizing the Kullback-Leibler Divergence of two probability distributions. More significantly, the value of α can be visually reflected as the enhancement degree as shown in Figure 1(d), and users are allowed to adjust the dehazing results as per their preference. Our CHM is effective, flexible, and explainable.

Compared with the state-of-the-art real image dehazing methods, *e.g.*, DAD [33] and PSD [7], only the proposed RIDCP can effectively process the hazy images captured in the wild while generating adjustable results, which are shown in Figure 1. The contributions of our work can be summarized as follows.

- We present a new paradigm to push the frontier of deep learning-based image dehazing towards real scenes.
- We are the first to leverage the high-quality codebook prior in the real image dehazing task. The controllable HQPs matching operation is proposed to overcome the gap between synthetic and real domains and produce adjustable results.
- We re-formulate the degradation model of real hazy images and propose a phenomenological degradation pipeline to simulate the hazy images captured in the wild.

2. Related Work

2.1. Single Image Dehazing

Image Dehazing. The early attempts at single image dehazing consider estimating the parameters of the atmosphere scattering model presented in Eq. (1) by priors on haze-free images [4, 12, 16, 35, 44], These methods have achieved impressive results. However, the handcrafted priors based on empirical observations are hard to perform well in diverse scenarios. For example, the assumption of DCP [16] is not available in the sky region. The proposed method obtains the priors of high-quality images by pre-training a discrete codebook on large-scale datasets, which is more reliable and comprehensive.

With the development of deep learning techniques, how to use data-driven ideology to remove haze gains a lot of attention. At the early stage, many studies [5, 21, 31] try to adopt convolutional neural networks (CNNs) to estimate the parameters of the degradation model in Eq. (1). In addition, in order to avoid accumulated errors in parameters estimation, some end-to-end networks [10, 15, 27, 30, 40] are proposed to directly estimate the haze-free image. The above learning-based methods have achieved excellent performance on synthetic datasets. However, their significant performance drop on real-world data urgently needs to be solved.

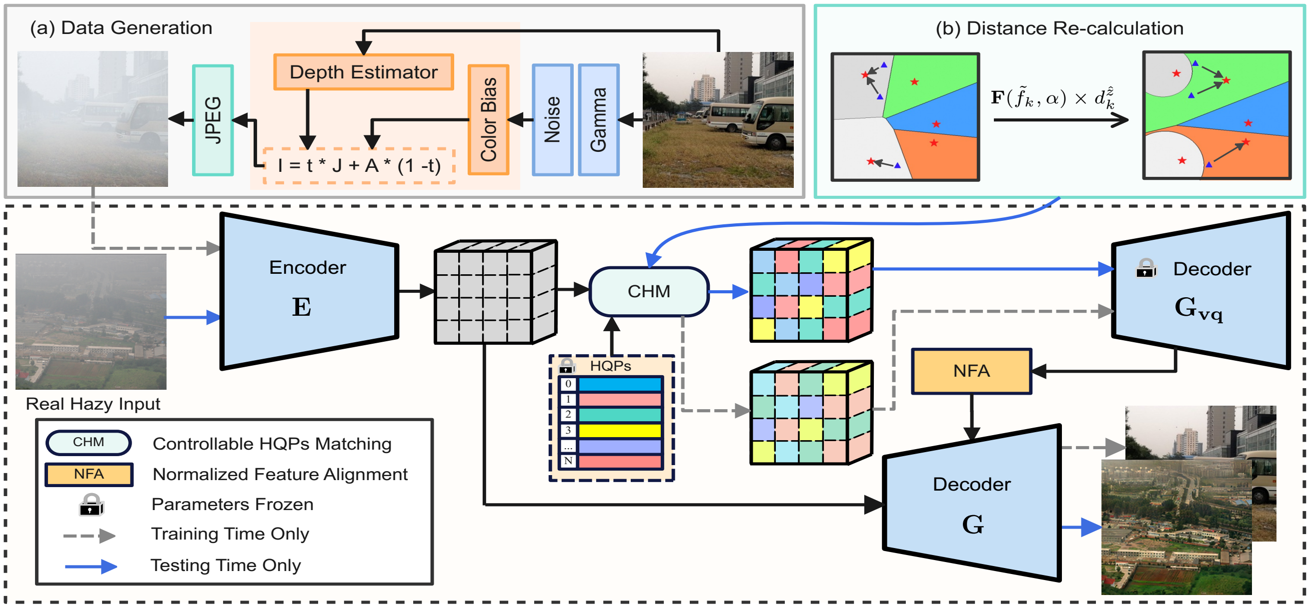


Figure 2. Overview of our RIDCP. During the training phase, we train the dehazing network on the data synthesized by our data generation pipeline, as illustrated in (a). The network is based on the pre-trained HQPs codebook and the corresponding decoder $\mathbf{G}_{vq}$ of VQGAN. We also design the Controllable HQPs Matching (CHM) operation for real domain adaptation by re-calculating the distance $d_k^{\hat{z}} = ||\hat{z} - z_k||$ between features and HQPs. (b) represents the distance re-calculation with two Voronoi diagrams, where the colored cells indicate belonging to better HQPs and the gray cells vice versa. Triangles represent features and star points represent HQPs. It can be seen that after the distance recalculation points that originally belonged to the gray cells are forced to be assigned to the colored cells by our CHM.

Real Image Dehazing. Recently, some works pay attention to real image dehazing. One research line is to utilize GANs [13] for generating hazy data that fits the real haze domain. Shao *et al.* [33] design a domain adaptation strategy based on the framework of CycleGAN [?]. Yang *et al.* [39] propose an unpaired dehazing framework named D4. It can estimate the scene depth of hazy images and generate hazy data with different thicknesses to benefit dehazing model training. However, GANs are easy to produce artifacts in the generated results, which is harmful to training models. Another research line aims to introduce prior knowledge by loss functions or network architectures. Li *et al.* [23] propose a semi-supervised pipeline that adopts prior-based loss functions to train networks on the real dataset. PSD [7] adds a physical-based sub-network on the pre-trained dehazing model and further proposes a prior loss committee to fine-tune the network on real-world data in an unsupervised manner. Nevertheless, directly using handcrafted priors cannot avoid the inherent flaws of prior-based methods. In our study, we investigate overcoming the weaknesses of both types of real image dehazing methods by proposing a novel data generation pipeline and exploiting the latent high-quality priors.

2.2. Discrete Codebook Learning

Recently, a vector-quantized auto-encoder framework was proposed in VQ-VAE [36], which learns a discrete codebook in latent space. The discrete representation effectively addresses the "posterior collapse" issue in auto-encoder [19] architecture. VQGAN [11] further improves the perceptual quality of reconstructed results by introducing adversarial supervision for codebook learning. The learned discrete codebook helps boost the performance in many low-level vision tasks including face restoration [14, 43] and image super-resolution [6]. Gu *et al.* [14] introduce the vector quantization technique to face restoration and design a parallel decoder to achieve a balance between visual quality and fidelity. Zhou *et al.* [14] cast blind face restoration as a code prediction task, and propose a Transformer-based prediction network to replace the nearest-neighbor matching operation for better matching the corresponding code. FeMaSR [6] extends the discrete codebook learning to blind super-resolution. Motivated by the exciting performance of these approaches, we are the first to leverage the high-quality codebook prior for real image dehazing. A novel and controllable HQPs matching operation is proposed to further bridge the gap between our synthetic data and real data, which is inevitable for real scenes.

3. Data Preparation for Real Image Dehazing

Redesigning the pipeline of data generation has been demonstrated as an effective way for solving real-world low-level vision tasks [37, 38, 41]. Based on these works, we consider various degradation factors when synthesizing paired data for training the dehazing network, which can mitigate the domain gap with the real data. For conciseness, we represent Eq. (1) as $I(x) = \mathcal{P}(J(x), t(x), A)$. The formation of the hazy image can be written as:

$$I(x) = JPEG(\mathcal{P}(J(x)^{\gamma} + \mathcal{N}, e^{\beta \hat{d}(x)}, A + \Delta A)). \quad (2)$$

The details of Eq. (2) are introduced as follows:
Poor light condition. $\gamma \in [1.5, 3.0]$ is a brightness adjustment factor and $\mathcal{N}$ is the Gaussian noise distribution. These two components can simulate poor light conditions that frequently occur in hazy weather.
Transmission map. As a key parameter in the degradation model, we adopt the depth estimation algorithm [18] to estimate depth map $d(x)$ and use $\beta \in [0.3, 1.5]$ to control the haze density.
Colorful haze. To obtain diverse hazy images, the color bias of atmosphere light is considered, which is implemented by a three-channel vector $\Delta A \in [-0.025, 0.025]$. The range of A is in the range of $[0.25, 1.0]$.
JPEG compression. We observe that dehazing algorithms amplify the JPEG artifacts. It is desirable to remove such artifacts while dehazing. $JPEG(\cdot)$ denotes JPEG compression in the final results.

We select 500 clean images to build the paired data, and the hazy data is generated on-the-fly during the training phase. Additionally, low light and JPEG compression appear with 50% probability in the proposed pipeline.

4. Methodology

The key idea of our work is to adopt a discrete codebook that introduces high-quality priors (HQPs) into the dehazing network. The overall framework of the proposed method is illustrated in Figure 2. The training phase can be divided into two stages. In the first training stage, we pre-train a VQGAN [11] on high-quality data, obtaining a latent discrete codebook $\mathcal{Z}$ with HQPs and the correspondence decoder $\mathbf{G}_{vq}$ (Sec. 4.1). In the second stage, our RIDCP based on the pre-trained VQGAN is trained on hazy images generated by the proposed synthesis pipeline (Sec. 4.2). Moreover, in order to help the network find more accurate code, we propose a controllable adjustment feature matching strategy based on code activation distribution on high-quality images (Sec. 4.3). Besides, the details of training objectives can be found in supplementary materials.

4.1. Latent Codebook for High-quality Priors

We first introduce how VQGAN works briefly. Given a high-quality image patch x, which is the input of the VQGAN encoder $\mathbf{E}_{vq}$ and the corresponding outputs are the latent features $\hat{z}$. Then each "pixel" $\hat{z}_{ij}$ of $\hat{z}$ will be matched to the nearest HQPs in codebook $\mathcal{Z} \in \mathbb{R}^{K \times n}$ and then obtain the discrete representation z^q_{ij}, which can be written as:

$$z^q_{ij} = \mathcal{M}(\hat{z}_{ij}) = arg \min_{z_k \in \mathcal{Z}} (||\hat{z}_{ij} - z_k||_2), \quad (3)$$

(a) Hazy input (b) Reconstruction

Figure 3. Result reconstructed by the pre-trained VQGAN. The haze is removed but distorted textures are introduced.

where K denotes the codebook size, n is the channel number of $\hat{z}$, and $\mathcal{M}(\cdot)$ represents the matching operation. Finally, the input x is reconstructed by $\mathbf{G}_{vq}$:

$$x' = \mathbf{G}_{vq}(z^q) = \mathbf{G}_{vq}(\mathcal{M}(\mathbf{E}_{vq}(x))), \quad (4)$$

where x' is the reconstructed result.
Observation 1. To understand the potential of the HQPs in the codebook and better utilize it, we made some observations on the results reconstructed by the pre-trained VQGAN. As illustrated in Figure 3, our VQGAN can remove the thin haze and recover vivid color for the hazy input without fine-tuning. We analyze that using HQPs in a matching manner can replace the degraded feature so help it jump to the high-quality domain. However, the dehazing ability of VQGAN is limited due to the difficulty in matching the correct code. Moreover, some distorted textures are produced because of the information loss during the vector-quantized phase. Thus, directly adopting the features from $\mathbf{G}_{vq}$ is suboptimal. It is intuitive that our next step is to train an encoder $\mathbf{E}$ that can help priors matching, and a decoder $\mathbf{G}$ that can utilize the features reconstructed from HQPs.

4.2. Image Dehazing via Feature Matching

Based on the observation in Sec. 4.1, image dehazing is decoupled into two sub-tasks: matching correct code and removing texture distortion.
Encoder for HQPs Matching. We follow SwinIR [25] which shows its powerful feature extraction ability for image restoration to design our encoder $\mathbf{E}$. Specifically, the shallow feature extraction head consists of a stack of residual layers [17] and $4\times$ downsamples the features. Then 4 residual swin transformer blocks [28] are followed, which serve as the deep feature extraction module.
Decoder with Normalized Feature Alignment. We propose the Normalized Feature Alignment (NFA) to help the decoder utilize the features reconstructed from HQPs. Firstly, VQGAN tends to decrease results' fidelity due to the information loss brought by vector-quantized operation [14, 43]. Our solution is to eliminate the distortion by the guidance of features before HQPs matching. Specifically, in ith layer, we adopt the deformable convolution [9] to align the features F^i_{vq} from $\mathbf{G}_{vq}$ with the features F^i from $\mathbf{G}$, which can be written as:.

$$F^i_w = DCONV(F^i_{vq}, CONV(Concat(F^i_{vq}, F^i))), \quad (5)$$

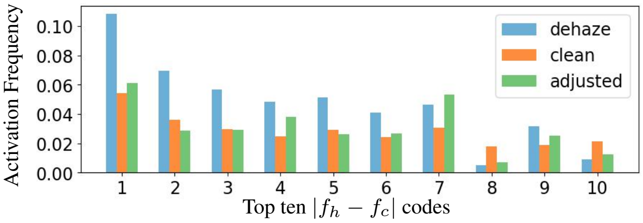


Figure 4. Code activation frequencies under different situations. 'dehaze' denotes inputting real hazy images to dehazing network and 'clean' denotes feeding clean images into the pre-trained VQGAN network. 'adjusted' denotes our RIDCP equipped with the CHM under the recommended parameter with real hazy inputs.

where F_w^i denotes the features after warping and $DCONV$ is the deformable convolutional layer. $CONV$ is the convolutional layer for offset generation. In addition, we notice that the ratio of the values of F_w^i and F^i is not stable, resulting in an inadequate combination. Thus, we balance the contributions of each by forcing them to be in the same order of magnitude, which can be written as:

$$F^i = F^i + \frac{\sum F^i}{\sum F_w^i} F_w^i. \tag{6}$$

4.3. Controllable HQPs Matching Operation

Observation 2. Our RIDCP achieves relatively satisfactory results with the help of $\mathbf{E}$ and $\mathbf{G}$. However, there are still limitations, *e.g.*, low color saturation in some challenging real data. Rather than the HQPs that already show a strong capability in reconstructing vivid results (see Observation 1), the main reason is the difficulty in finding correct HQPs, which is caused by the domain gap between synthetic data and real data. Although the domain gap is drastically reduced by our synthesis pipeline than previous works [22, 42], it is still impossible to cover all real-world hazy conditions by our pipeline.

To verify our claim, we made an observation as follows. We randomly collect 200 high-quality clean images as input of the pre-trained VQGAN and compute the activation frequency $f_c \in \mathbb{R}^K$ of each code. Similarly, 200 real hazy images are fed to the dehazing network to compute the frequency $f_h \in \mathbb{R}^K$. Figure 4 illustrates the activation frequencies of the codes with the top ten largest differences between f_h and f_c. We can see a significant distribution shift. The observation proves that the unavoidable domain gap results in a divergent matching for HQPs. Thus, HQPs still have unexplored potential.

Controllable Matching via Distance Re-calculation. Based on the above observation, it is indispensable to match better HQPs when encountering real hazy images, *i.e.*, priors with high frequency on clear images. Two components can affect the HQPs matching, which are the encoder $\mathbf{E}$ and the matching operation $\mathcal{M}(\cdot)$. Since it is difficult to retrain $\mathbf{E}$ on real hazy images without reference images, defining a new matching operation $\mathcal{M}'(\cdot)$ sounds like a reasonable solution. We propose Controllable HQPs Matching (CHM) that re-calculates distances by assigning different weights during matching phase. The CHM can be written as:

$$\mathcal{M}'(\hat{z}) = arg \min_{z_k \in Z} (\mathbf{F}(\tilde{f}_k, \alpha) \times ||\hat{z} - z_k||), \tag{7}$$

where $\mathbf{F}(\tilde{f}_k, \alpha)$ is the function to generate weights based on the frequency difference $\tilde{f}_k = f_h^k - f_c^k$ and adjusted by a parameter α. There are three objectives in the design of $\mathbf{F}$: 1) Since higher $\tilde{f}_k$ means less activation is needed, $\mathbf{F}$ should be monotonic with $\tilde{f}_k$ thus ensuring consistent trend adjustment. 2) $\mathbf{F}(0, \alpha) \equiv 1$ so that HQPs with the same frequencies on clear and hazy data are not adjusted. 3) The degree of adjustment can be controlled monotonically by α, *e.g.*, $\forall \tilde{f}_1 > \tilde{f}_2, \forall \alpha_1 > \alpha_2 \rightarrow \frac{\mathbf{F}(\tilde{f}_1, \alpha_1)}{\mathbf{F}(\tilde{f}_2, \alpha_1)} > \frac{\mathbf{F}(\tilde{f}_1, \alpha_2)}{\mathbf{F}(\tilde{f}_2, \alpha_2)}$. Coincidentally, the exponential function has these properties, thus $\mathbf{F}$ can be formulated as;

$$\mathbf{F}(\tilde{f}_k, \alpha) = e^{a \times \tilde{f}_k}. \tag{8}$$

Figure 2(b) adopts two Voronoi diagrams to simulate the changes occurring in the high-dimensional space during feature matching. As we can see, the points originally belonging to gray cells are matched to the colored cells after distance re-calculation, *i.e.*, finding better HQPs.

Possible Solution of the Recommended $\boldsymbol{\alpha}$. Our method is able to control the HQPs matching based on the above strategy. The final goal is to find a suitable α to adapt the network to real domain. According to the law of large numbers, the frequencies f_c^k, f_h^k can be substituted for the corresponding probabilities $P_c(x = z_k), P_h(x = z_k|\alpha)$. The gap between the dehazing results and the clean domain can be represented by the difference between the two probability distributions. Thus, the real domain adaptation problem is transferred into calculating an optimal parameter $\hat{\alpha}$ that can minimize the forward Kullback-Leibler Divergence of $P_c(x = z_k)$ and $P_h(x = z_k|\alpha)$, which is also the maximum likelihood estimation of α:

$$\begin{aligned}
\hat{\alpha} &= arg \min_{\alpha} KL(P_c||P_h) \\
&= arg \min_{\alpha} \sum_{i=1}^{K} P_c(x = z_i) \log \frac{P_c(x = z_i)}{P_h(x = z_i|\alpha)} \\
&= arg \max_{\alpha} \sum_{i=1}^{K} P_c(x = z_i) \log P_h(x = z_i|\alpha) \\
&= arg \max_{\alpha} \prod_{i=1}^{K} P_c(x = z_i) P_h(x = z_i|\alpha).
\end{aligned} \tag{9}$$

We use a binary search algorithm to iteratively find the approximate optimal solution for $\hat{\alpha}$. The final determination is $\hat{\alpha} = 21.25$ and higher precision calculations have little

(a) Hazy image (b) MSBDN [10] (c) Dehamer [15] (d) PSD [7] (e) D4 [39] (f) DAD [33] (g) RIDCP

Figure 5. Visual comparison on RTTS dataset [22].

effect on the results. **Note that, $\hat{\alpha}$ may not be the determined choice for all cases. One can flexibly adjust α according to their preference.**

5. Experiments

5.1. Datasets

High-quality Datasets. In order to obtain high-quality results from the pre-trained HQPs, the VQGAN needs to be trained on large-scale datasets containing high-resolution and texture-sharp images. In our work, we use DIV2K [1] and Flickr2K [26] (containing $4,250$ images) to train the first stage. Both datasets are widely used in high-quality reconstruction tasks [6, 24, 25].

Real Haze Datasets. We qualitatively and quantitatively evaluate our dehazing network on the RTTS dataset [22], which contains over $4,000$ real hazy images with diverse scenes, resolutions, and degradation issues. Besides, we use Fattal's dataset [12] that includes 31 classic real hazy cases for further visual comparison.

5.2. Implementation Details

For both VQGAN and RIDCP training, we use Adam optimizer with default parameters ($\beta_1 = 0.9, \beta_2 = 0.99$). The learning rate is fixed to 0.0001 during the training phase and the batch size is set to 16. For data augmentation, we randomly resize and crop the input into a size of 256×256, and flip it with a half probability. During the first training stage, our VQGAN is pre-trained on DIV2K and Flickr2K for 350K iterations. Then, the proposed RIDCP is trained on the data generated by the proposed synthesis pipeline for 10K iterations. All experiments are implemented with PyTorch framework on 4 NVIDIA V100 GPUs. The code implemented by MindSpore framework is also provided.

5.3. Comparison with State-of-the-Art Methods

We compare the performance of the proposed method with several state-of-the-art dehazing approaches. The experiments are designed from both quantitative and qualitative perspectives. Moreover, we also conduct a user study to verify the subjective performance of our method.

Quantitative Comparison. Since there is no ground-truth image in real hazy datasets, we use some non-reference metrics for quantitative comparison. We first adopt the Fog Aware Density Evaluator (FADE) [8] for haze density estimation. In addition, two widely-used image quality assessment metrics, BRISQUE [29] and NIMA [34] are also included. The quantitative comparison is conducted on RTTS dataset with two dehazing methods (MSBDN [10] and Dehamer [15]) that achieve outstanding performance on synthetic hazy image datasets [22], and three real dehazing ap-

(a) Hazy image (b) MSBDN [10] (c) Dehamer [15] (d) PSD [7] (e) D4 [39] (f) DAD [33] (g) RIDCP

Figure 6. Visual comparison on Fattal's data [12].

Table 1. Quantitative comparison and user study on RTTS dataset. Red indicates the best and blue indicates the second best. 'US' shows the percentage of votes in the user study.

Method	FADE↓	BRISQUE↓	NIMA↑	US↑
Hazy image	2.484	37.011	4.3250	0.030
MSBDN [10]	1.363	28.743	4.1401	0.046
Dehamer [15]	1.895	33.866	3.8663	0.041
DAD [33]	1.130	32.727	4.0055	0.143
PSD [7]	0.920	25.239	4.3459	0.105
D4 [39]	1.358	33.206	3.7239	0.079
RIDCP	0.944	18.782	4.4267	0.556

proaches (DAD [33], PSD [7], D4 [39]). The results are illustrated in Table 1. The proposed RIDCP achieves the best in terms of BRISQUE and NIMA, which gains 25.57% and 1.86% improvements, respectively. For FADE, our method is ranked second slightly below the PSD. However, as shown in Figure 5, PSD tends to produce over-enhanced results, which leads to an inaccurate evaluation. Overall, RIDCP achieves the best results on quantitative metrics, and subsequent experiments will further prove our superiority.

Qualitative Comparison. We perform the qualitative comparison on RTTS and Fattal's datasets, which is shown in Figure 5 and Figure 6. We can observe that Dehamer, MSBDN, and D4 cannot process the real hazy images well. PSD can produce bright results but the dehazing ability is limited. DAD is effective in haze removal, while its results suffer from color bias and dark tone. Our method generates the best perceptual results in terms of brightness, colorfulness, and haze residue compared to other methods. More results can be found in supplementary materials.

User Study. We conduct a user study to evaluate the proposed method subjectively against other methods. We randomly select 100 images from RTTS dataset for comparison and invite 5 experts with image processing background and 5 naive observers as volunteers. Before the user study, we give the observer three tips: 1) The primary concern is whether the haze is removed, especially the dense haze in the distance. 2) Pay attention to whether the natural color is recovered. 3) A good method should generate artifact-free results. Afterward, the images are displayed to the observer group by group. Each group contains the input image and the results generated by different methods. The observer is required to choose the best one after at least 10 seconds of observation. We statistic the percentage of each method selected as the best and the final scores are listed in Table 1. The proposed RIDCP achieves the highest score and is well ahead of the second place, further demonstrating our method's superior dehazing ability.

5.4. Ablation Study

In order to verify the effectiveness of each key component, we conduct a series of ablation experiments. Generally, we discuss the effectiveness of CHM, NFA, and the phenomenological degradation pipeline in this section.

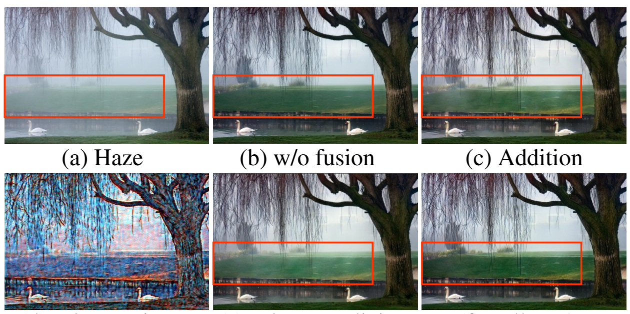
Influence of Adjustment Parameter. The degree of real domain adaptation is controlled by parameter α in Eq. (8) and one can adjust the final result flexibly by adjusting α. Thus, we are curious about what influence the different α will have. As Figure 7 shows, the value of α and the image enhancement effect belong to a linear relationship. More visual-pleasing even over-enhanced results can be produced when $\alpha > 0$. Interestingly, we can obtain under-enhanced results if α is adjusted in the opposite direction.

Effectiveness of NFA. Our NFA can help remove the distorted textures caused by feature matching while preserving the useful information reconstructed from HQPs. The NFA can be divided into two key parts: warping operation based on deformable convolution and normalize-based addition. To analyze the role of each part, we propose 4 variants to replace NFA, which are: *1) Without any fusion operation. 2) Adding directly. 3) Normalized addition without warping. 4) Warping and direct addition.* Figure 8 shows a set of comparisons. Observing the grass area in red boxes, the result of variant 1 is dark and remaining thin haze. Variants 2 and 4 also have non-homogeneous fog residues in some areas. Variant 3 generates obvious artifacts because forcing normalizing unaligned features to the same order of magnitude and adding them together makes the network difficult to train. Only the full NFA achieves the best in brightness and haze removal.

Effectiveness of the Phenomenological Degradation Pipeline. To prove that our proposed degradation pipeline

(a) Hazy image (b) $\alpha = -60.0$ (c) $\alpha = 0.0$ (d) $\alpha = 21.25$ (e) $\alpha = 60.0$

Figure 7. Results under different adjustment degrees. The proposed CHM allows users to adjust the degree of enhancement from low ($\alpha = -60.0$) to high ($\alpha = 60.0$). The recommended value ($\alpha = 21.25$) facilitates the network to generate the most natural results (surrounded by the blue box). Zoom in for the best view.



(a) Haze (b) w/o fusion (c) Addition

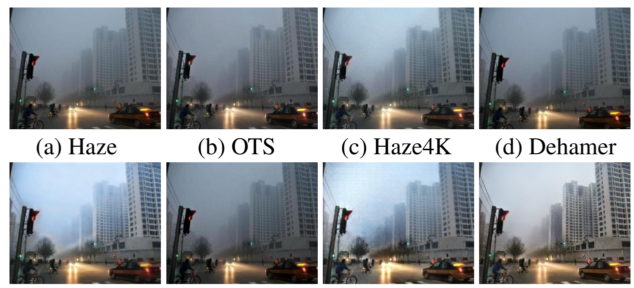
(d) w/o warping (e) w/o normalizing (f) Full NFA

Figure 8. Ablation results of the proposed NFA.

for paired data generation can boost the capabilities of haze removal, we retrain our RIDCP on two widely-used synthetic datasets, which are OTS [22] and Haze4K [42]. Notably, Haze4K is post-processed by DAD [33]. Moreover, we replace the training set from OTS with our synthetic data for transformer-based Dehamer and CNN-based MSBDN, thus demonstrating that it can generally bring gains. The comparison results are illustrated in Figure 9. We can observe that our dehazing network cannot remove the haze under the training of OTS and Haze4K. Besides, Dehamer and MSBDN can generate results with less haze and higher brightness with the help of our training data. However, they still struggle in color recovery compared to our method, which also demonstrates the effectiveness of HQPs and our adaptation strategy.

6. Discussion

Conclusion. In this paper, we present a novel paradigm to revitalize deep dehazing networks towards the real world. Our proposed phenomenological degradation pipeline synthesizes more realistic hazy data, which achieves significant gains in haze removal. Based on our observations and analysis, we introduce the high-quality priors in VQGAN to the dehazing network and progressively leverage their power, which finally builds our real image dehazing network via high-quality codebook priors (RIDCP). Extensive experiments show the superiority of the proposed paradigm.



(a) Haze (b) OTS (c) Haze4K (d) Dehamer

(e) Dehamer† (f) MSBDN (g) MSBDN† (h) Ours

Figure 9. Ablation results of the proposed data generation pipeline. We retrain our dehazing network on OTS (b) and Haze4K (c) to verify the effectiveness of our generation pipeline. Dehamer and MSBDN are also retrained on our synthetic data, which are marked by †.

Limitations and Future Work. In the process of doing our work on RIDCP, we observed that there are still some difficulties that are urgent to be addressed. We leave the challenges here and hope that future work can address them

- Existing dehazing methods including RIDCP can not process non-homogeneous haze well.
- Dehazing based on enhancement fashion is limited. Generative ability should be introduced for recovering extremely dense haze.
- We found that difficult to benchmark dehazing methods in quantitative fairly. Robust metrics for evaluating the quality of dehazing results are also needed.

Acknowledgements. This work is funded by the National Key Research and Development Program of China (NO.2018AAA0100400), Fundamental Research Funds for the Central Universities (Nankai University, NO.63223050), China Postdoctoral Science Foundation (NO.2021M701780). We are also sponsored by CAAI-Huawei MindSpore Open Fund.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 126–135, 2017. 6

[2] Codruta O Ancuti, Cosmin Ancuti, Mateu Sbert, and Radu Timofte. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1014–1018. IEEE, 2019. 2

[3] Codruta O Ancuti, Cosmin Ancuti, and Radu Timofte. 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 (CVPRW)*, pages 444–445, 2020. 2

[4] Dana Berman, Shai Avidan, et al. Non-local image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1674–1682, 2016. 2

[5] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing (TIP)*, 25(11):5187–5198, 2016. 1, 2

[6] Chaofeng Chen, Xinyu Shi, Yipeng Qin, Xiaoming Li, Xiaoguang Han, Tao Yang, and Shihui Guo. Real-world blind super-resolution via feature matching with implicit high-resolution priors. In *Proceedings of the 30th ACM International Conference on Multimedia (ACM MM)*, pages 1329–1338, 2022. 2, 3, 6

[7] Zeyuan Chen, Yangchao Wang, Yang Yang, and Dong Liu. Psd: Principled synthetic-to-real dehazing guided by physical priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7180–7189, 2021. 1, 2, 3, 6, 7

[8] Lark Kwon Choi, Jaehee You, and Alan Conrad Bovik. Referenceless prediction of perceptual fog density and perceptual image defogging. *IEEE Transactions on Image Processing (TIP)*, 24(11):3888–3901, 2015. 6

[9] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 764–773, 2017. 4

[10] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2157–2167, 2020. 1, 2, 6, 7

[11] Patrick Esser, Robin Rombach, and Bjorn Ommer. Taming transformers for high-resolution image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12873–12883, 2021. 2, 3, 4

[12] Raanan Fattal. Dehazing using color-lines. *ACM Transactions on Graphics (TOG)*, 34(1):1–14, 2014. 1, 2, 6, 7

[13] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C Courville, and Yoshua Bengio. Generative adversarial nets. In *Neural Information Processing Systems (NeruIPS)*, 2014. 3

[14] Yuchao Gu, Xintao Wang, Liangbin Xie, Chao Dong, Gen Li, Ying Shan, and Ming-Ming Cheng. Vqfr: Blind face restoration with vector-quantized dictionary and parallel decoder. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022. 2, 3, 4

[15] Chun-Le Guo, Qixin Yan, Saeed Anwar, Runmin Cong, Wenqi Ren, and Chongyi Li. Image dehazing transformer with transmission-aware 3d position embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5812–5820, 2022. 1, 2, 6, 7

[16] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 33(12):2341–2353, 2010. 1, 2

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

[18] Mu He, Le Hui, Yikai Bian, Jian Ren, Jin Xie, and Jian Yang. Ra-depth: Resolution adaptive self-supervised monocular depth estimation. *arXiv preprint arXiv:2207.11984*, 2022. 4

[19] Geoffrey E Hinton and Richard Zemel. Autoencoders, minimum description length and helmholtz free energy. *Advances in Neural Information Processing Systems*, 6, 1993. 3

[20] Shih-Chia Huang, Trung-Hieu Le, and Da-Wei Jaw. Dsnet: Joint semantic learning for object detection in inclement weather conditions. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 43(8):2623–2633, 2020. 1

[21] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 4770–4778, 2017. 1, 2

[22] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing (TIP)*, 28(1):492–505, 2019. 5, 6, 8

[23] Lerenhan Li, Yunlong Dong, Wenqi Ren, Jinshan Pan, Changxin Gao, Nong Sang, and Ming-Hsuan Yang. Semi-supervised image dehazing. *IEEE Transactions on Image Processing (TIP)*, 29:2766–2779, 2019. 2, 3

[24] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3867–3876, 2019. 6

[25] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1833–1844, 2021. 4, 6

[26] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for sin-

gle image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 136–144, 2017. 6

[27] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 7314–7323, 2019. 1, 2

[28] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 10012–10022, 2021. 4

[29] Anish Mittal, Anush K Moorthy, and Alan C Bovik. Blind/referenceless image spatial quality evaluator. In *2011 Conference Record of the Forty Fifth Asilomar Conference on Oignals, Systems and Computers (ASILOMAR)*, pages 723–727. IEEE, 2011. 6

[30] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 2020. 1, 2

[31] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 154–169. Springer, 2016. 1, 2

[32] Christos Sakaridis, Dengxin Dai, Simon Hecker, and Luc Van Gool. Model adaptation with synthetic and real data for semantic dense foggy scene understanding. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 687–704, 2018. 1

[33] Yuanjie Shao, Lerenhan Li, Wenqi Ren, Changxin Gao, and Nong Sang. Domain adaptation for image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2808–2817, 2020. 1, 2, 3, 6, 7, 8

[34] Hossein Talebi and Peyman Milanfar. Nima: Neural image assessment. *IEEE Transactions on Image Processing (TIP)*, 27(8):3998–4011, 2018. 6

[35] Robby T Tan. Visibility in bad weather from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2008. 2

[36] Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in Neural Information Processing Systems*, 30, 2017. 3

[37] Hong Wang, Zongsheng Yue, Qi Xie, Qian Zhao, Yefeng Zheng, and Deyu Meng. From rain generation to rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14791–14801, 2021. 3

[38] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 1905–1914, 2021. 3

[39] Yang Yang, Chaoyue Wang, Risheng Liu, Lin Zhang, Xiaojie Guo, and Dacheng Tao. Self-augmented unpaired image dehazing via density and depth decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2037–2046, 2022. 2, 3, 6, 7

[40] Tian Ye, Mingchao Jiang, Yunchen Zhang, Liang Chen, Erkang Chen, Pen Chen, and Zhiyong Lu. Perceiving and modeling density is all you need for image dehazing. *arXiv preprint arXiv:2111.09733*, 2021. 1, 2

[41] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 4791–4800, 2021. 3

[42] Zhuoran Zheng, Wenqi Ren, Xiaochun Cao, Xiaobin Hu, Tao Wang, Fenglong Song, and Xiuyi Jia. Ultra-high-definition image dehazing via multi-guided bilateral learning. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16180–16189. IEEE, 2021. 5, 8

[43] Shangchen Zhou, Kelvin C.K. Chan, Chongyi Li, and Chen Change Loy. Towards robust blind face restoration with codebook lookup transformer. In *Neural Information Processing Systems (NeruIPS)*, 2022. 2, 3, 4

[44] Qingsong Zhu, Jiaming Mai, and Ling Shao. Single image dehazing using color attenuation prior. In *British Machine Vision Conference (BMVC)*. Citeseer, 2014. 1, 2