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

Rui-Qi Wu1  Zheng-Peng Duan1  Chun-Le Guo1*  Zhi Chai2  Chongyi Li3
1VCIP, CS, Nankai University  2Hisilicon Technologies Co. Ltd.
3S-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 Real Image Dehazing network via high-quality Codebook Priors (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://rqwu.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:

$$I(x) = J(x)t(x) + A(1 - t(x)),$$  \hspace{1cm} (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 $\beta$.

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
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 Real Image Dehazing network via high-quality Codebook Priors (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 \( \alpha \) and the statistical distribution gap of HQPs activation in Sec. 4.3. By adjusting \( \alpha \), the distribution of HQPs activation can be shifted. Moreover, we present a theoretically feasible solution to obtain the optimal \( \alpha \) by minimizing the Kullback-Leibler Divergence of two probability distributions. More significantly, the value of \( \alpha \) 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 $G_{vq}$ of VQGAN. We also design the Controllable HQPs Matching (CHM) operation for real domain adaptation by re-calculating the distance $d^*_{z_k}$ 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.

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 autoencoder [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 concise-
ness, we represent Eq. (1) as \( I(x) = \mathcal{P}(J(x), t(x), A) \). The formation of the hazy image can be written as:

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

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 theJPEG artifacts. It is desirable to remove such artifacts while dehazing. \( \text{JPEG}(\cdot) \) 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 \( Z \) with HQPs and the corresponding decoder \( G_{vq} \) (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 \( E_{vq} \) 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 \( Z \in \mathbb{R}^{K \times n} \) and then obtain the discrete representation \( z_{ij}^q \), which can be written as:

\[
z_{ij}^q = \mathcal{M}(\hat{z}_{ij}) = \arg\min_{z_k \in Z} ||\hat{z}_{ij} - z_k||_2, \tag{3}
\]

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 \( G_{vq} \):

\[
x' = G_{vq}(z^q) = G_{vq}(\mathcal{M}(E_{vq}(x))), \tag{4}
\]

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 \( G_{vq} \) is suboptimal. It is intuitive that our next step is to train an encoder \( E \) that can help priors matching, and a decoder \( G \) 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 \( E \). 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 \( F^{vq}_i \) from \( G_{vq} \) with the features \( F^i \) from \( G \), which can be written as:

\[
F^i_w = \text{DCONV}(F^{vq}_i, \text{CONV}({\text{Concat}}(F^{vq}_i, F^i))), \tag{5}
\]
where $F^i_w$ 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 $F^i_w$ 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:

$$F^i = F^i + \frac{\sum F^i_w}{\sum F^i_w} F^i.$$ (6)

### 4.3. Controllable HQPs Matching Operation

**Observation 2.** Our RIDCP achieves relatively satisfactory results with the help of $E$ and $G$. 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 $E$ and the matching operation $M(\cdot)$. Since it is difficult to retrain $E$ on real hazy images without reference images, defining a new matching operation $M'(\cdot)$ 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:

$$M'(\tilde{z}) = \arg \min_{\tilde{z} \in \mathbb{R}^K} (F(\tilde{f}_k, \alpha) \times ||\tilde{z} - z_k||),$$ (7)

where $F(\tilde{f}_k, \alpha)$ 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 $\alpha$. There are three objectives in the design of $F$: 1) Since higher $\tilde{f}_k$ means less activation is needed, $F$ should be monotonic with $\tilde{f}_k$ thus ensuring consistent trend adjustment. 2) $F(0, \alpha) \equiv 1$ 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 $\alpha$, e.g., $\forall f_1 > f_2, \forall \alpha_1 > \alpha_2 \rightarrow F(\tilde{f}_1, \alpha_1) > F(\tilde{f}_2, \alpha_2)$. Coincidentally, the exponential function has these properties, thus $F$ can be formulated as:

$$F(\tilde{f}_k, \alpha) = e^{\alpha \times \tilde{f}_k}.$$ (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 $\alpha$.** Our method is able to control the HQPs matching based on the above strategy. The final goal is to find a suitable $\alpha$ to adapt the network to real domain. According to the law of large numbers, the frequencies $f_h^k, f_c^k$ can be substituted for the corresponding probabilities $P_e(x = z_k), P_h(x = z_k | \alpha)$. 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 $P_e(x = z_k)$ and $P_h(x = z_k | \alpha)$, which is also the maximum likelihood estimation of $\alpha$:

$$\hat{\alpha} = \arg \min_{\alpha} KL(P_e || P_h)$$

$$= \arg \min_{\alpha} \sum_{i=1}^{K} P_e(x = z_i) \log \frac{P_e(x = z_i)}{P_h(x = z_i | \alpha)}$$

$$= \arg \max_{\alpha} \sum_{i=1}^{K} P_e(x = z_i) log \frac{P_h(x = z_i | \alpha)}{P_h(x = z_i)}$$

$$= \arg \max_{\alpha} \prod_{i=1}^{K} P_e(x = z_i)P_h(x = z_i | \alpha).$$ (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
effect on the results. Note that, $\alpha$ may not be the determined choice for all cases. One can flexibly adjust $\alpha$ 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 $256 \times 256$, 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-
Figure 6. Visual comparison on Fattal’s data [12].

Table 1. Quantitative comparison and user study on RTTS dataset. ‘US’ shows the percentage of votes in the user study.

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

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.

Table 1. Quantitative comparison and user study on RTTS dataset. ‘US’ shows the percentage of votes in the user study.

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. Ablation Study. The degree of real domain adaptation is controlled by parameter $\alpha$ in Eq. (8) and one can adjust the final result flexibly by adjusting $\alpha$. Thus, we are curious about what influence the different $\alpha$ will have. As Figure 7 shows, the value of $\alpha$ 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 $\alpha$ 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
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.

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


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


