Towards Real-world Image Dehazing: A Tailored Dehazing Method and A High-Quality Dataset

Yong Liu*, Qingji Dong*, Chao Zhu, Yu Guo, Fei Wang†
National Key Laboratory of Human-Machine Hybrid Augmented Intelligence,
National Engineering Research Center of Visual Information and Applications,
Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University
Xi'an, China
yongliu@stu.xjtu.edu.cn, dongqingji@stu.xjtu.edu.cn, zhu_chao@stu.xjtu.edu.cn,
yu.guo@xjtu.edu.cn, wfx@mail.xjtu.edu.cn

Abstract

Existing dehazing methods face challenges in generalization due to the lack of paired real-world training data and tailored models. Recently, some semi-supervised/unsupervised schemes have been explored and achieve impressive performance. However, their performance still depends heavily on synthetic training data and the introduced prior-based strong constraints do not always hold. In this paper, we first propose a Prior-compensated Multi-stage Dehazing Network (PMDN), which can learn different levels of real-world haze distribution through multi-stage progressive learning. To utilize prior knowledge effectively, we introduce a Prior-based Feature Compensation Module (PFCM), guiding intermediate results with an adaptive weight. Additionally, we propose a MixCut Consistent Dehazing (MCCD) strategy to mix paired and derived images using a cross-cutting scheme, reinforcing dehazing through consistency principles. We also introduce RealHQ-HAZE, a new dataset with 200 collected real-world hazy images, corresponding 200 carefully rendered haze-free images, and additional 1000 varicolored hazy images transferred from the collected images. Extensive experiments demonstrate the effectiveness of our dataset and the superior performance of PMDN over existing state-of-the-art dehazing methods.

Keywords:* *Image Dehazing, Image Enhancement, Prior-based learning, Semi-supervised learning

1. Introduction

Single image dehazing has long been a thriving research topic over the past two decades, which aims to remove the haze from images captured in real-world scenarios and restore visually pleasing results. According to the physical model of the haze process [27, 30], the formation of haze can be modeled by:

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

where I, J, t, A and x denotes the hazy image, the haze-free image, the transmission map, the global atmospheric light, and the pixel position, respectively.

With the powerful learning capability of CNN, learning-based methods have shown an explosive popularity in the field of single image dehazing [7, 40, 32, 11, 13, 26, 31, 18]. Due to the difficulty in obtaining real-world hazy/haze-free image pairs, these methods mostly focus on the synthetic training dataset [24] with a large domain gap to real-world data, and do not generalize well to real-world hazy images. To seek further breakthroughs, some researchers have turned to construct small-scale real-world image haze datasets [2, 3, 1, 4] by simulating real-world haze scenes with professional haze machines. However, due to limited haze distribution and model over-fitting, methods based on these datasets tend to suffer from the drop in performance. Moreover, the lack of various lighting and color tones on synthetic training datasets also limits the performance of those methods in diverse haze scenes.

To address the above problems, some semi-supervised methods [25, 33, 9, 39] and unsupervised methods [37, 22, 16, 40] have been proposed. They introduced unpaired real-world hazy images into the model training process through the prior-based constraints, such as dark channel loss[16], total variation loss[33] and CLAHE reconstruction loss[9]. Despite varying degrees of progress in these methods, their performance still heavily depends on synthetic training data in the supervised stage and unrobust enforcing prior constraints in the unsupervised stage. In addition, these image dehazing models ignore tailor-made designs for real-world image dehazing, leaving much room for further exploration.

*Equal contribution
†Corresponding author

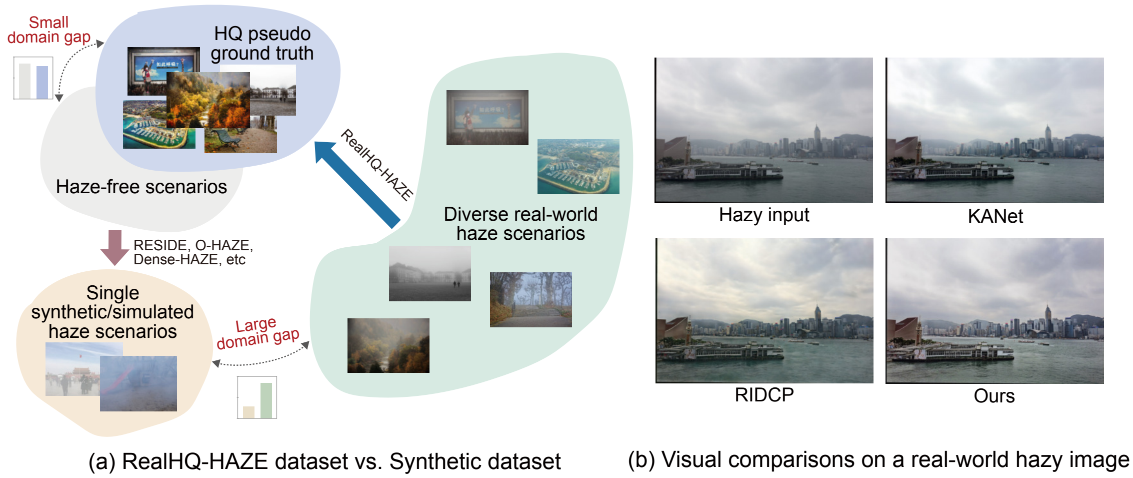

Figure 1. Overview of the proposed dataset and method in this work. (a) We explore the inverse generation of high-quality pseudo-GTs from diverse real-world haze scenarios, and our RealHQ-HAZE has a smaller domain gap compared to existing synthetic/simulated dehazing datasets; (b) The proposed PMDN trained on RealHQ-HAZE can completely remove the haze in the image and achieve visually pleasing results, which significantly outperform other image dehazing methods.

In this paper, we introduce a new real-world image dehazing dataset, called RealHQ-HAZE, to replace existing datasets that have a large domain gap with real-world hazy images. Rather than seeking a unified paradigm for diverse and unknown real-world haze reproduction, we employ a multi-model guided ground truth (GT) generation strategy to efficiently convert real-world hazy images into high-quality pseudo-GTs. Specifically, the introduced strategy utilises dehazing model committee, simple image enhancement techniques and style transfer model, respectively, to gradually remove the haze on the image and obtain paired visually realistic pseudo-GTs with negligible domain gap to real-world haze-free images. Moreover, in order to further expand the practicality of the dataset for real-world scenarios, we also generate 1000 derived hazy images using a diffusion-based data augmentation technique. These images exhibit a wide range of lighting conditions and color tones. Examples of RealHQ-HAZE can be seen in Figure 1(a), Figure 4, Figure 6, and supplementary material. Comparative analysis demonstrate that the proposed RealHQ-HAZE is more effitive against existing image dehazing datasets in real-world dehazing task.

In addition, from the perspective of deep models for real-world image dehazing, we propose a Prior-compensated Multi-stage Dehazing Network (PMDN), which incorporates two well-established principles, i.e., progressive learning and prior-based feature compensation, to address the real-world image dehazing problem. First, to enhance the network's ability to learn real-world haze types and distributions, our proposed PMDN employs a multi-stage progressive learning strategy to constrain the intermediate dehazed results by different weights. Inspired by [12, 44], we introduce the physical model of the haze process into the deep feature space at each stage. Second, instead of considering the existing image prior as a strong constraint and optimizing the model in an unsupervised manner, we propose a Prior-based Feature Compensation Module (PFCM), which integrates the image priors as a non-learnable embedding into each stage of PMDN to efficiently guide the network to produce the intermediate haze-free results. In addition, to boost the adaptability of the model in general haze scenarios, we propose a MixCut Consistent Dehazing (MCCD) strategy. Specifically, we use paired images and relevant derived images from our RealHQ-HAZE datasets to train the network according to the dehazing consistency principle and the novel cross-cutting scheme. We use supervision constraints to reasonably transfer the uncertainty of derived data to paired data. Extensive experiments show that the proposed PMDN can effectively remove the haze in real-world hazy images while restoring visually pleasing results. Visual comparisons on a typical hazy image can be seen in Figure 1(b).

In summary, our main contributions are as follows:

- We propose a new real-world image dehazing dataset called RealHQ-HAZE, which contains high-quality hazy/pseudo-GT pairs and additional derived hazy images with diverse lighting and color tones, which can further facilitate the research on real-world image dehazing.
- We propose a novel Prior-compensated Multi-stage Dehazing Network (PMDN), which tackles the real-world image dehazing problem by incorporating progressive learning, prior-based feature compensation, as well as the MixCut Consistent Dehazing strategy.

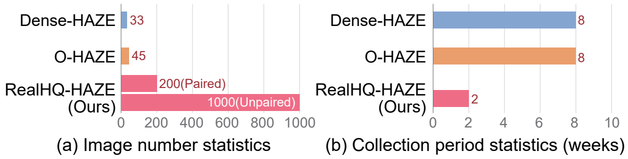

Figure 2. Comparison of different real-world image dehazing datasets. Our RealHQ-HAZE dataset has 4 times more image pairs and additional 1000 varicolored hazy images but takes about 1/4 time against [2, 4], which is efficient and easily implementable.

- Extensive experiments show that the proposed RealHQ-HAZE performs well in real-world image dehazing, and the proposed PMDN can produce impressive results with sharp details and natural tones.

2. Related Work

Recent years have witnessed an explosive spread of studies focusing on the single image dehazing task. Here, we refer the readers to for a comprehensive review of the most relevant datasets and recent methods.

Image Dehazing Datasets. To realize the learning and removal of the haze on the image by the CNN model, RESIDE [24] is a widely used large-scale benchmark that synthesizes hazy images by setting random atmospheric lights and scatter coefficients. However, the generated hazy images often exhibit large domain gap from real-world hazy images, leading to poor generalization ability of models in real-world scenarios. Furthermore, the real-world hazy images it contains are outdated and low-resolution, which seriously deviates from the trajectory of modern photography, resulting in suboptimal performance of the dehazing model.

To address this problem, a series of works [4, 1, 2, 3] attempted to collect small-scale real-world datasets using specialized haze machines. Since the absence of wind was the most difficult parameter to satisfy, these works took up to 8 weeks to collect datasets with less than 50 pairs of images. This is inefficient and hard to reproduce, as can be seen from the comparison in Figure 2. In addition, because the haze machine can only simulate scenes with limited scene depth and simple haze distributions, the applicability of models based on these datasets is also limited. Therefore, it is significant to build a new high-quality real-world image dehazing dataset.

Image Dehazing Methods. The rapid development of neural networks is greatly advancing the research of single image dehazing. Qin et al. [31] proposed a feature fusion attention network that combines channel attention with pixel attention, taking into account different channel-wise weighted information and uneven pixel-wise haze distribution. Yang et al. [41] proposed a self-augmented image dehazing framework that explores the scatter coefficient and depth information contained in hazy and haze-free images. Although significant progress has been achieved, these methods always fail to process natural hazy images due to the large domain gap between synthetic/simulated training data and real-world haze data. To revitalize the performance of learning-based image dehazing models in real-world haze scenarios, Li et al. [25], Shao et al. [33] exploited the prior properties of clean images (e.g., dark channel prior and image gradients) to constrain the additional unpaired real-world hazy images. More recently, Chen et al. [9] proposed a principled synthetic-to-real dehazing framework that integrates multiple physical priors into a prior loss committee to fine-tune the pre-trained model. Wu et al. [39] introduced the high-quality codebook priors in VQ-GAN [14] to the dehazing network, which uses the controllable HQPs matching operation to bridge the gap between synthetic and real domains. Wang et al. [37] introduced a method projecting image features to orthogonal space to reduce the relevance between features, and proposed a self supervised network to assign orthogonal features to haze-related and unrelated components. Feng et al. [15] construct a new framework by making efforts from imaging perspective, structural modules, and training strategies to improve adaptablity in real-world environments. Nevertheless, we find that these methods lack special network structure design that toward to real-world image dehazing and strong image priore constraints cannot cover diverse real-world haze scenarios well. In this paper, we aim to investigate datasets and methods dedicated to real-world image dehazing task.

3. RealHQ-HAZE Dataset

3.1. Diverse Real-World Hazy Images

To guarantee that the proposed dataset contains a variety of haze scenarios, we carefully selected 200 real-world hazy images taken by professional cameras from the Internet. Specifically, these images include a variety of scene types such as country, city, forest, lake, mountain, and aerial photography. Due to the different weather conditions at the time of shooting, the haze distribution of these images is also different and diverse, which provides the possibility for the learning-based model to better understand the haze in the real world. Rather than previous low-resolution real-world hazy images, our collected images have a high-resolution from 720P to 2K resolutions, which are widely used in our lives. Examples of these images are given in the green panel in Figure 1(a). In addition, we also collected 100 high-quality real-world hazy images RealHQ-Test from Interent, RESIDE [24], and BeDDE [43] as a test set for RealHQ-HAZE, called RealHQ-Test, to replace the existing outdated and low-resolution evaluation set.

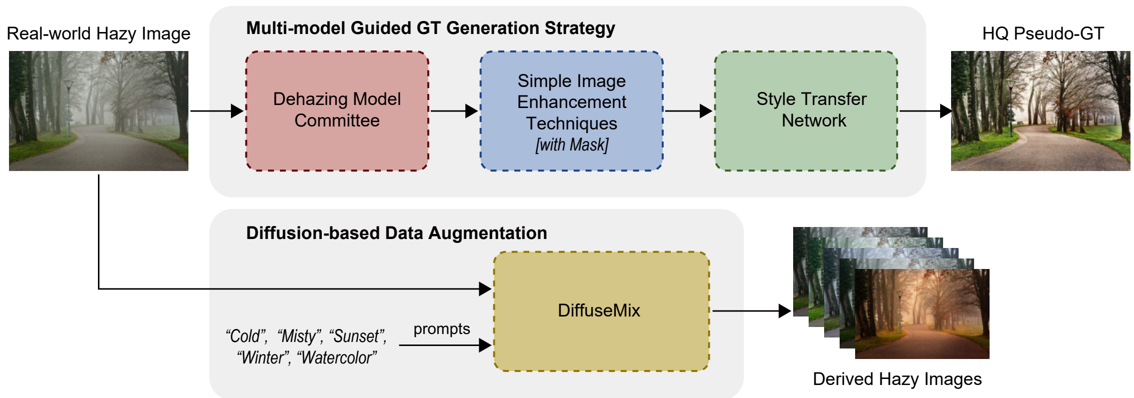

Figure 3. The details of dataset creation process, including the Multi-model Guided GT Generation Strategy and Hazy Images Augmentation Process. We use a Dehazing Model Committee, Simple Image Enhancement Techniques, and a Style Transfer Network (STN), respectively, to generate high-quality pseudo-GTs with clearly visible details and natural scenes. We utilize InstruxtPix2Pix to generate derived hazy images with a variety styles.

3.2. High-Quality Pseudo-GTs

As shown in Figure 3, we use a multi-model guided GT generation strategy to inversely obtain the corresponding pseudo-GT. Specifically, we first select three existing state-of-the-art image dehazing models, i.e., EPDN [32], DAD [33] and PSD [9], and combine them into a Dehazing Model Committee. We aim to initially remove the haze on the raw image by Dehazing Model Committee, and get a relatively clear intermediate result. Due to the different architectural designs and training strategies, each image dehazing model has its advantages in dehazing ability and color restoration. For example, DAD removes most of the haze from the image using its well-designed domain adaptation paradigm. However, the results suffer from a significant loss of image detail, which can be compensated by the dehazing results of EPDN. On the other hand, PSD provides brighter images while removing haze, which can help alleviate the overall darkening problem observed in EPDN. To mix these dehazed results into one image based on their respective advantages, we use the well-known Photoshop to generate different masks for different results, and realize the image mixing by $Y = \sum_{i=1}^{3} X_i \otimes Mask_i$.

To further bridge the domain gap between the mixed images and the real-world haze-free images, we adjust the corresponding image properties (such as exposure, texture, sharpening, and clarity) in Photoshop. Finally, we train a simplified Style Transfer Network (STN) CycleGAN [45] to bring the pseudo-GT to more realistic haze-free images. Note that our final RealHQ-HAZE is carefully selected from over 600 sets of paired hazy images and pseudo-GTs. More detailed description can be found in the supplementary material.

As can be seen from the Figure 4, the generated pseudo-GT is clear in terms of haze removal and vivid in terms of color restoration. In addition, we perform quantitative comparisons on different real-world datasets in Figure 2. It can be seen that the proposed dataset has more than 4 times image pairs and additional 1000 varicolored hazy images but takes only 1/4 time against than [4, 2], which is easier to implement.

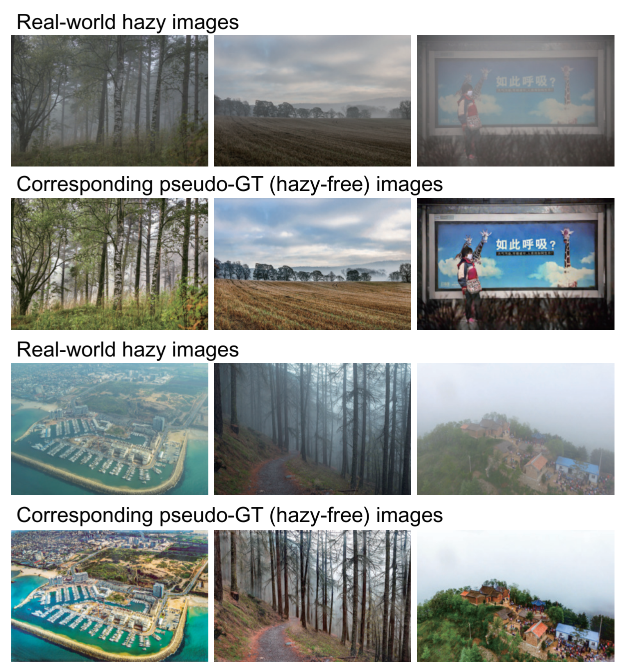

Figure 4. Examples of different haze scenarios and corresponding high-quality pseudo-GTs in the proposed RealHQ-HAZE dataset.

3.3. Diffusion-based Hazy Image Augmentation

Considering the potential limitations of our initial dataset, i.e., the lack of illumination and color tones in real-

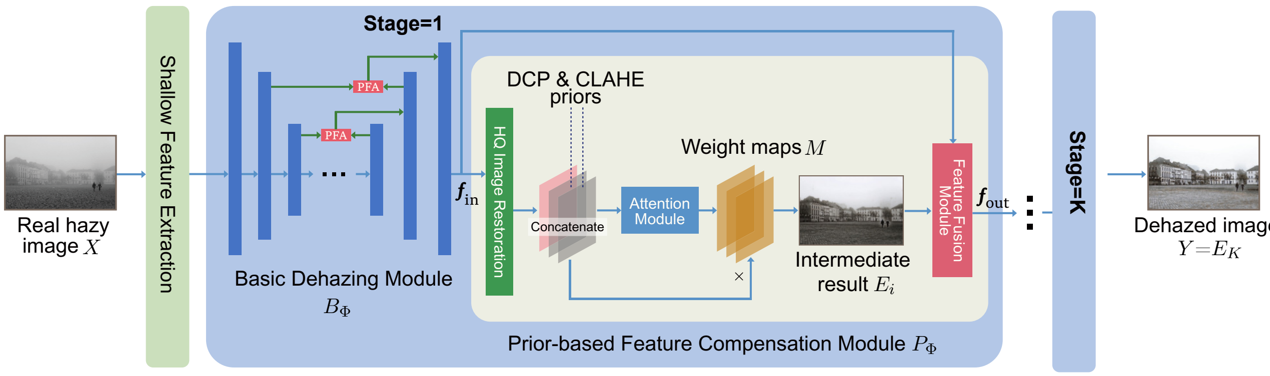

Figure 5. The overall architecture of the proposed Prior-compensated Multi-stage Dehazing Network (PMDN). Our method progressively removes haze from images via multi-stage learning. In each stage, we integrate the shallow and deep features through efficient Physical-driven Feature Aggregation (PFA) and use the well-known image priors to guild the generated intermediate results in the proposed Prior-based Feature Compensation Module (PFCM).

world haze scenes, and to further expand the diversity of our dataset, we employ DiffuseMix [20], a diffusion-based image augmentation method, to generate diverse derived hazy images in different environmental conditions. These images have similar scenes to the collected hazy images, but are not pixel-wise equivalent to HQ Pseudo GTs due to the randomness of the diffusion models in content generation.

Specifically, as shown in Figure 3, the image augmentation process consists of a pretrained diffusion model that takes a prompt p_j from a predefined set of prompts. we choose a set of 5 prompts, including [“`Cold`”, “`Misty`”, “`Sunset`”, “`Winter`”, “`Watercolor`”]. These prompts are selected because of their generic nature and applicability to a wide variety of images. Along with every prompt p_j, each hazy image X_i would be produced an augmented counterpart image X_{ij}. The image editing process in existing diffusion models is usually open-ended, requiring textual prompts to obtain various image-to-image or text-to-image transformations. In our case, as the goal is to achieve a slightly modified but not too different version of X_i. To preserve the visual information without significant alterations by the diffusion model, similar to DiffuseMix, we utilize InstruxtPix2Pix [6] to generate images. This scheme not only enhances the diversity of our dataset but also ensures that our data more accurately represents the wide range of hazy conditions that can be encountered in real-world scenarios. Considering transferred images are not pixel-wise equivalent to our Pseudo-GTs, we treat them as derived images. Figure 6 shows some examples of transferred hazy images.

Figure 6. Examples of derived hazy images transferred from the collected images in the proposed RealHQ-HAZE dataset.

4. Proposed Method

To build an efficient dehazing model for real-world haze scenarios, we define the dehazing task as a multi-stage progressive reconstruction problem. To achieve this, we propose a Prior-compensated Multi-stage Dehazing Network (PMDN). Figure 5 illustrates the overall architecture of the proposed PMDN.

4.1. Basic Architecture

Given a real-world hazy image X, we first pass it into a shallow feature extraction unit to extract shallow features and then fed the shallow features into K continuously dehazing stage. In each stage, a basic dehazing module B_Φ is used to generate preliminary dehazed features and a Prior-based Feature Compensation Module (PFCM) P_Φ is proposed to guild the generated intermediate result.

As for basic dehazing module, inspired by [8] with effective performance in image restoration task, we use the basic NAFNet's block to extract profound features from the initial shallow input at each down-sampling and up-sampling stage in the basic dehazing module B_Φ. Furthermore, unlike several previous works that use simple skip connections to integrate the shallow features into the deep features, we introduce the physics model of the haze process into the deep feature space and apply the Physical-driven Feature Aggregation (PFA) to fuse the features from these two layers for effective feature preserving. In PFCM, more prior knowledge is introduced to assist the learning of the dehazing model, and a predicted intermediate result E_i can be obtained and then be used to further refine the previously dehazed features $\boldsymbol{f}_{in}$. In this way, we can get the final dehazed result E_K (i.e., Y) through K continuously dehazing stages.

4.2. Physical-driven Feature Aggregation

To enhance the interpretability of the model in the feature space dehazing process, we propose a Physical-driven Feature Aggregation (PFA) scheme to replace the raw skip connection in B_Φ. Specifically, we first reformulate the Equation 1 as,

$$J(x) = I(x)V(x) - B(x) + I(x) \tag{2}$$

where $V(x) = 1/t(x) - 1$ and $B(x) = (1/t(x) - 1) \cdot A$. Then we apply a feature extractor to the Equation 2 and get,

$$k \otimes J = k \otimes (I \odot V) - k \otimes B + k \otimes I \tag{3}$$

where $\otimes$ denotes the convolution operator and $\odot$ denotes the Hadamard product. By using $\mathbf{K}$, $\mathbf{I}$, $\mathbf{V}$, and $\mathbf{B}$ denote the matrix-vector forms of k, I, V, and B, respectively, we can represent the Equation 3 in matrix-vector form,

$$\mathbf{KJ} = \mathbf{KVI} - \mathbf{KB} + \mathbf{KI} \tag{4}$$

Furthermore, the matrix $\mathbf{KV}$ can be decomposed into the product of $\mathbf{EK}$ and the Equation 4 can be rewritten as,

$$\mathbf{KJ} = \mathbf{E}(\mathbf{KI}) - \mathbf{KB} + \mathbf{KI} \tag{5}$$

By doing a few matrix-vector form transformations, we get,

$$\widetilde{\mathbf{Y}} = \mathbf{E}\widetilde{\mathbf{X}} - \widetilde{\mathbf{B}} + \widetilde{\mathbf{X}} \tag{6}$$

We consider the downsampling layer and its output in B_Φ as the haze feature extractor and $\widetilde{\mathbf{X}}$, respectively. In the training process, we drive the unsampling layer to learn $\mathbf{E}$ and $\widetilde{\mathbf{B}}$. Finally, we obtain deep features $\widetilde{\mathbf{Y}}$ that are irrelevant to haze from Equation 6 and pass them to the subsequent decoder unit. The final output of the PFA operations can be formulated as,

$$\boldsymbol{f}_{\uparrow i+1} = \mathrm{PFA}\left(\boldsymbol{f}_{\downarrow i}, \boldsymbol{f}_{\uparrow i}\right) = \boldsymbol{e}_{\uparrow i} \odot \boldsymbol{f}_{\downarrow i} - \boldsymbol{b}_{\uparrow i} + \boldsymbol{f}_{\downarrow i} \tag{7}$$

where $\boldsymbol{f}_{\uparrow i}$ = Concatenation $[\boldsymbol{e}_{\uparrow i}, \boldsymbol{b}_{\uparrow i}]$, $\boldsymbol{f}_{\uparrow i}$ and $\boldsymbol{f}_{\downarrow i}$ are feature maps generated by upsampling layer and downsampling layer, respectively. Our PFA allows B_Φ to efficiently use shallow features and deep features to inversely generate haze-irrelevant features at different scales and depths, which is critical for real-world image dehazing.

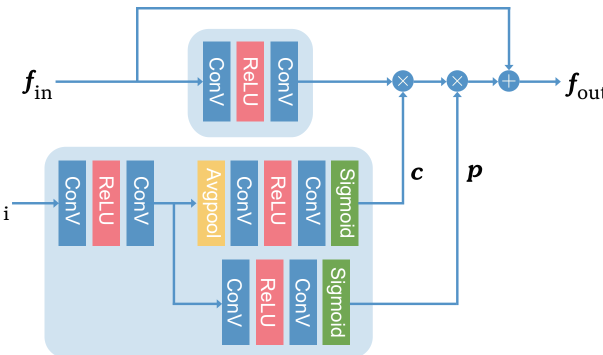

Figure 7. The details of Feature Fusion Module (FFM) in the proposed PFCM.

4.3. Prior-based Feature Compensation Module

Account for the complex and varied weather conditions, light conditions and color tones that contribute to real-world haze, it is necessary to introduce more known prior knowledge to assist the learning of the dehazing model. We propose Prior-based Feature Compensation Module (PFCM) and introduce two well-grounded image priors, i.e., Dark Channel Prior [17] and Contrast Limited Adaptive Histogram Equalization, to guide the dehazing module B_Φ to generate more favorable features. Specifically, we first obtain the predicted result I_{model} from B_Φ through the HQ image restoration unit. Then, I_{model} is concatenated with the introduced image priors (i.e., I_{DCP} and I_{CLAHE}). Next, the adaptive weight maps $M \in \mathbb{R}^{3\times H\times W}$ can be generated by feeding the concatenated image $I \in \mathbb{R}^{(3\times 3)\times H\times W}$ to an attention module. Here, we get the intermediate haze-free result E_i by element-wise multiply the concatenated image I and the weight maps M,

$$E_i = I_{model} \odot M_1 + I_{DCP} \odot M_2 + I_{CLAHE} \odot M_3 \tag{8}$$

where $\odot$ operator denotes element-wise multiply. For this intermediate result E_i, we perform supervised constraint with the GT. Finally, we use E_i to recalibrate the output features of B_Φ by the Feature Fusion Module (FFM). The structure of the FFM is illustrated in Figure 7. Based on the feature fusion attention method [31], we obtain $\boldsymbol{c} \in$

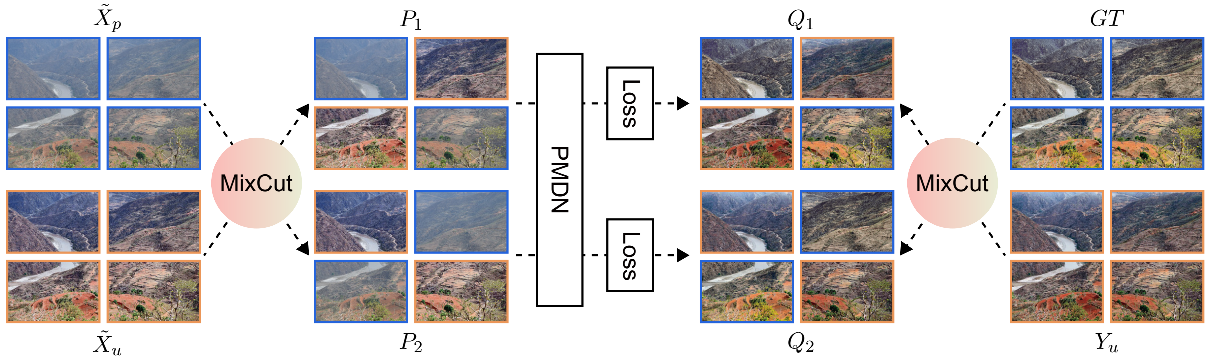

Figure 8. The details of the introduced MixCut scheme. We randomly select a set of images from all the results of the first dehazing and perform the MixCut operation on them to get the second input of PMDN. The target images are obtained in the same way. Note that for the paired data, we use the existing Y_{GT}; for the derived data, we use the final result Y_u of the first dehazing.

$\mathbb{R}^{C\times 1\times 1}$ and $\boldsymbol{p} \in \mathbb{R}^{1\times H\times W}$ through the channel attention mechanism and the pixel attention mechanism, respectively. The compensated features $\boldsymbol{f}_{out}$ at the current stage can be formulated as Equation 9 and propagated to the next stage for further processing.

$$\boldsymbol{f}_{out} = (ConVs(\boldsymbol{f}_{in}) \odot c) \odot p + \boldsymbol{f}_{in} \tag{9}$$

4.4. MixCut Consistent Dehazing

To further improve the performance of the proposed PMDN, we introduce derived part of the RealHQ-HAZE dataset $\{X_u\}$ besides the paird part $\{X_p, Y_{GT}\}$ to train the network and propose a novel MixCut Consistent Dehazing (MCCD) scheme.

Specifically, given a set of the real-world haze inputs $\{X_p, X_u\}$ (denoted as $\{E_{p0}, E_{u0}\}$), we first get the intermediate results $\{E_{pi}, E_{ui}\}(i = 1, 2, .., K)$, where i is the sequence number of the dehazing stage, and get the corresponding final haze-free images $Y_p = E_{pK}$ and $Y_u = E_{uK}$. For the intermediate results $\{E_{pi}, E_{ui}\}$, it is not difficult to find that all $E_i(i = 0, 1, ..., K)$ are hazy images with different degrees of degradation under the same underlying haze-free scenes, which we refer to as the dehazing consistency principle. As MSBDN [11] pointed out that the dehazing method achieves better results in terms of PoH (Portion of Haze) on the hazy images of the same scene but less haze, our goal is to dehaze the images $\widetilde{X}_p \in \{E_{pi}\}$ and $\widetilde{X}_u \in \{E_{ui}\}$ again through the dehazing consistency principle to boost the adaptive ability of the PMDN.

As shown in Figure 8, in order to alleviate the uncertainty caused by derived $\widetilde{X}_u$, we introduce a MixCut scheme to ingeniously mix $\widetilde{X}_p$ and $\widetilde{X}_u$ and get P_1 and P_2. It is noted that the derived hazy images in our RealHQ-HAZE dataset are transferred from the paired hazy images. Therefore, compared to directly mixing with other real hazy images, using our transferred hazy images enables the consistency of the visual content in the mixed images. We then pass them into the proposed PMDN to generate corresponding dehazed results Q_1 and Q_2. Based on the dehazing consistency principle that under the same underlying haze-free scenes, we expect the results predicted from X_u and $\widetilde{X}_u$ to be consistent. Therefore, we regard the final results of the network Y_u as the target haze-free image of $\widetilde{X}_u$ and get the corresponding target haze-free image of Q_1 and Q_2 through perform same MixCut operation on Y_{GT} and Y_u.

As for loss functions, we use L1 loss to constrain the proposed MixCut Consistent Dehazing strategy, as well as PMDN. Specifically, in multi-stage progressive supervised dehazing process, we constraint the dehazed results E_{pi} at each stage:

$$L_p = \sum_{i=1}^{K} \omega_i \left| E_{pi} - Y_{GT} \right| \tag{10}$$

where ω_i are the weights. In semi-supervised Mix-Cut Consistent Dehazing process, we perform L1 loss on the re-dehazed results.

$$L_h = \sum_{j=1}^{2} \left| P_j - Q_j \right| \tag{11}$$

The total loss function can be expressed as,

$$L = L_p + \kappa L_h \tag{12}$$

where κ is the positive weight.

5. Experiments

5.1. Implementation Details

For training, we use the proposed RealHQ-HAZE dataset containing 200 image pairs and 1000 derived hazy images. We randomly crop all the images to 256 × 256 and adopt the Adam optimizer [21] (β_1 = 0.9 and β_2 = 0.999) to optimize our model. The proposed PMDN consists of three stages (i.e., K = 3). We train the PMDN for 200 epochs with

Figure 9. Visual comparisons of the performance of models trained on different datasets. The first row shows the hazy image and the qualitative results of models trained on the synthetic OTS dataset [24]. The second row shows the qualitative results of models trained on our RealHQ-HAZE dataset. Models trained on our RealHQ-HAZE dataset produce haze-free images with higher brightness and clarity. Best viewed with zooming in.

Table 1. Quantitative evaluations on RealHQ-Test set. The model trained on RealHQ-HAZE derived hazy images achieves better results.

Methods	BRISQUE↓	NIMA↑	FADE↓	NIQE↓	CLIP-IQA↑
Ours (Collected unpaired hazy images)	9.369	4.596	1.020	3.359	0.782
Ours (RealHQ-HAZE derived hazy images)	**8.974**	**4.806**	**0.951**	**3.286**	**0.795**

an initial learning rate of 10^{-4}, and the batch size is set to 4. We set $\omega_1 = 0.4, \omega_2 = 0.7, \omega_3 = 1$ and $\kappa = 0.5$. We perform quantitative and qualitative experimental evaluations of the models on the RealHQ-Test set. Both the shallow feature extraction unit and the HQ image restoration unit in our PMDN consist of only one convolution layer. The code and the proposed RealHQ-HAZE dataset will be made publicly available.

5.2. RealHQ-HAZE dataset vs. Synthetic dataset

Dehazing performance. To demonstrate the advantages of the RealHQ-HAZE dataset, we further compare the performance of five dehazing models (i.e., EPDN [32], MSBDN [11], FFA-Net [31], PSD [9], and our PMDN) trained on the synthetic OTS dataset [24] and our RealHQ-HAZE dataset, respectively. The OTS dataset is a synthetic outdoor hazy dataset in RESIDE dataset [24]. The qualitative comparison is shown in Figure 9. Benefiting from the high-quality hazy/haze-free sample pairs in our RealHQ-HAZE dataset, models trained on it achieve outstanding performance in terms of both details and brightness. It is easy to see that the results restored by the models trained on our RealHQ-HAZE dataset have brighter details, sharper edges and a more realistic visual perception. Further comparative analysis can be found in the supplementary material.

To demonstrate the effectiveness of the derived hazy images in RealHQ-HAZE dataset, we select 1000 unpaired real-world hazy images from the URHI [24], the LIVE Image Defogging Database [10] and the Internet to train our network using our MixCut Consistent Dehazing (MCCD) scheme. We employ five no-reference based Image Quality Assessment (IQA) metrics to quantitatively compare the results with the origin results. When training on collected unpaired hazy images, the MixCut operation concatenates two content unrelated images, which restricts the training performance. And the lack of illumination and color tones in collected hazy images also causes limited effectiveness. As shown in Table 1, we can see that the model using the derived hazy images in RealHQ-HAZE dataset trained achieves better score.

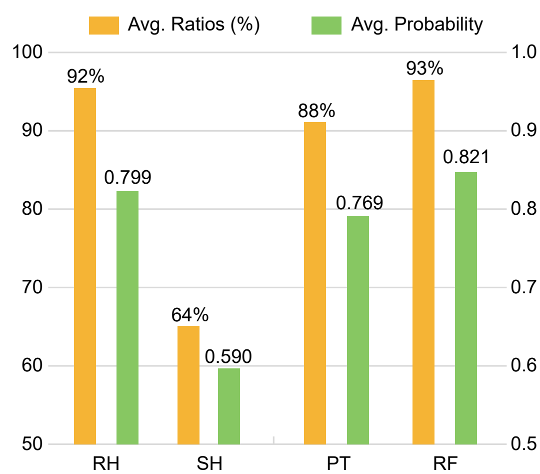

Figure 10. The average ratios and probability that the images of RH and SH are classified as real-world hazy images and PT and RF are classified as real-world haze-free images.

Image classification. To verify the advantage of the proposed RealHQ-HAZE dataset, we train a binary classification network following [42] based on the pre-trained DenseNet-121 [19]. Specifically, we use a collection of 500

Table 2. Quantitative evaluations on RealHQ-Test set. Our proposed method achieves substantially better results. Best and second best results are **highlighted** and underlined, respectively.

Methods	BRISQUE↓	NIMA↑	FADE↓	NIQE↓	CLIP-IQA↑
Hazy	20.713	4.217	3.027	4.063	0.557
AOD-Net [23]	19.424	4.303	2.662	3.978	0.635
EPDN [32]	25.144	4.548	2.276	4.947	0.717
FFA [31]	16.786	4.414	1.409	3.642	0.727
MSBDN [11]	10.365	3.808	1.135	3.571	0.743
AECRNet [38]	23.628	4.525	1.120	4.028	0.648
PSD [9]	11.557	4.003	**0.818**	3.312	0.709
DAD [33]	25.824	4.213	1.206	4.213	0.728
D^4 [41]	12.684	3.989	1.110	3.841	0.680
RIDCP [39]	27.943	3.649	1.899	4.485	0.626
KANet [15]	**8.048**	4.746	2.535	3.680	0.660
Ours	8.974	**4.806**	0.951	**3.286**	**0.795**

hazy images and another set of 500 haze-free images introduced in [10] as the training set and the Binary Cross Entropy (BCE) loss is adopted to optimize the network parameters.

In the testing stage, we use the trained model to classify the following sets of images:

- RH: 200 real-world hazy images in RealHQ-HAZE;
- SH: 200 synthetic hazy images in OTS [24];
- PT: 200 high-quality pseudo-GTs in RealHQ-HAZE;
- RF: 200 real-world haze-free images in OTS [24].

The average ratios and probability of different test sets are presented in Figure 10. Due to the diversity and complexity of real-world haze, only 64% of SH are classified as real-world hazy images, which is 28% lower than RH. In contrast, most of the pseudo-GT (PT) are classified as real-world haze-free images, and the domain gap with RF is only 5%. The classification results show that our dataset is more suitable for real-world image dehazing task.

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

To demonstrate the effectiveness of the proposed PMDN, we perform a series of quantitative and qualitative experiments with several state-of-the-art dehazing methods.

Quantitative evaluation. As shown in Table 2, we first visualize the quanitative results on RealHQ-Test set of our method against the following methods: AOD-Net [23], EPDN [32], FFA [31], MSBDN [11], AECRNet [38], PSD [9], DAD [33], D^4 [41], RIDCP [39] and KANet [15]. Note that we retrain AOD-Net, EPDN, FFA, MSBDN, AECRNet, and PSD with original settings on our RealHQ-HAZE dataset for fair comparison. DAD, D^4, RIDCP, and KANet are domain adaptation, self-augmented, and real data-based methods, respectively. Since a paired GT reference does not exist for real-world hazy image, we used five well-known no-reference based Image Quality Assessment (IQA) metrics to quantify the quality of the dehazed images.

Specifically, BRISQUE [28] and NIQE [29] are two blind image quality metrics, which can assess the overall quality of the images from the perspective of image structure and aesthetics. Our method achieves the best NIQE score and obtains the second-best BSRIQUE score, slightly below that of KANet. NIMA [29] evaluates images from a technical and aesthetic point of view based on a deep object detection neural network and our method outperforms other methods again. The Fog-Aware Density Evaluator (FADE) [10] is a metric to evaluate the haze density of images. Since it neglects to consider the color restoration and detail recovery of the images, it often leads to inaccurate estimation, which makes our method inferior to PSD. CLIP-IQA [35] is a new metric introduced to assess both the quality perception (look) and abstract perception (feel) of images through the rich visual language prior encapsulated in CLIP [36]. To distinguish visually pleasing sharp images from low-quality images with haze residue, blurred details or noise, we use the antonym [“`Sharp photo.`”, “`Not sharp photo.`”] as the prompt of the CLIP model. As a result, our method is far superior to other methods. Overall, our method achieves the competitive scores and shows great superiority in real-world image dehazing.

For further fair comparison, we evaluate our method on two commonly used real-world datasets NH-HAZE [3] and NH-HAZE 2 [5]. NH-HAZE is a non-homogeneous realistic dataset with pairs of real hazy and corresponding haze-free images, containing 55 outdoor scenes. NH-HAZE 2 is an artificial dataset for the NTIRE 2021 competition,

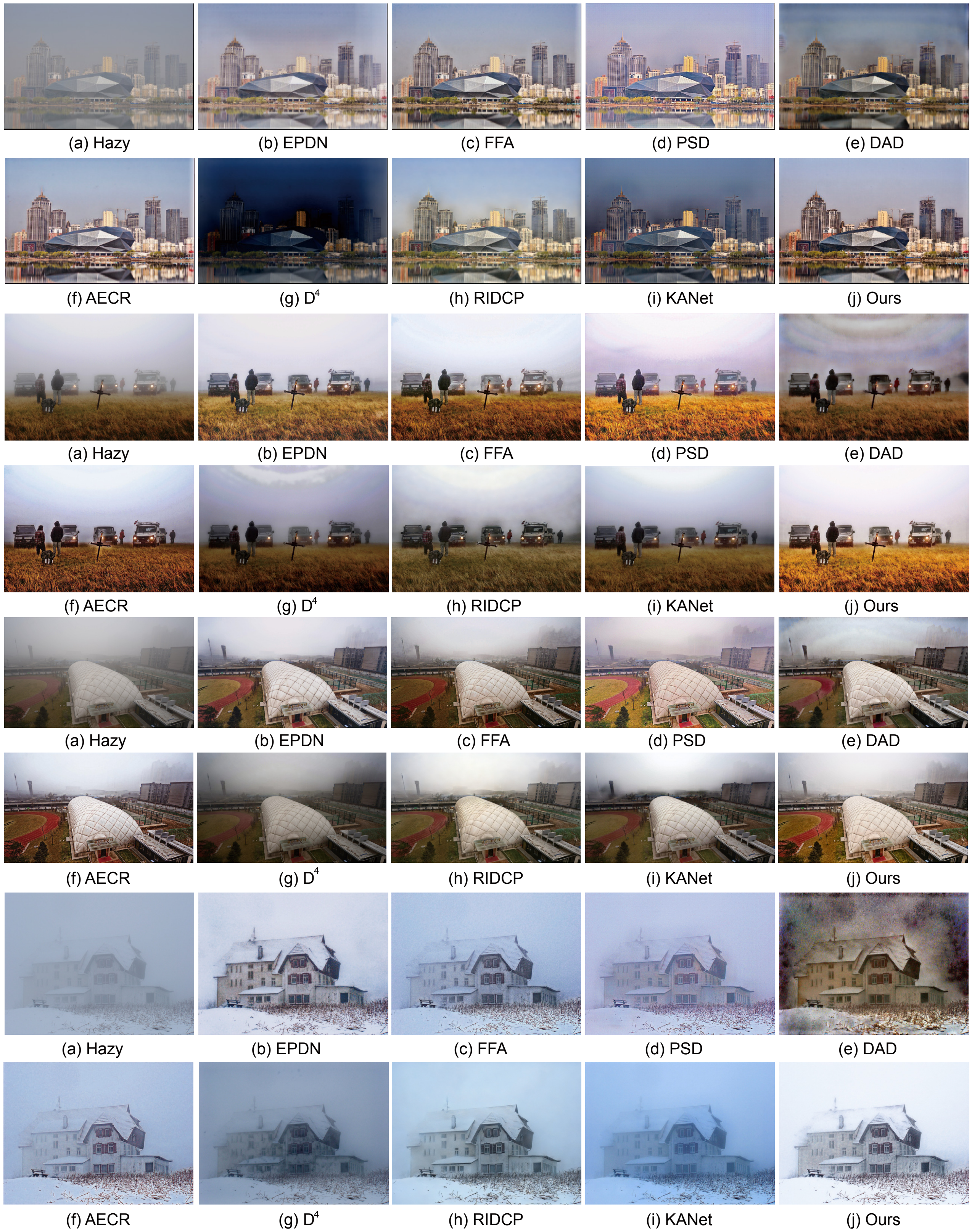

Figure 11. Visual comparisons of different dehazing methods on real-world hazy images, from which we can observe that our method produces images with natural colors and sharp outlines. Best viewed with zooming in.

Table 3. Quantitative evaluations on NH-HAZE [3] and NH-HAZE 2 [5] set. Best and second best results are **highlighted** and underlined, respectively.

Methods	NH-HAZE [3]		NH-HAZE 2 [5]	
	PSNR(dB)↑	SSIM↑	PSNR(dB)↑	SSIM↑
Hazy	13.33	0.508	13.07	0.602
AOD-Net [23]	14.05	0.529	13.73	0.628
EPDN [32]	14.24	0.637	14.24	0.660
FFA [31]	14.67	0.635	14.50	<u>0.709</u>
MSBDN [11]	14.86	<u>0.638</u>	14.62	0.705
AECRNet [38]	14.32	0.616	14.06	0.686
PSD [9]	13.98	0.577	14.10	0.679
DAD [33]	<u>15.08</u>	0.623	**14.74**	0.693
D^4 [41]	14.59	0.570	14.20	0.666
RIDCP [39]	14.17	0.600	14.24	0.674
KANet [15]	13.69	0.571	14.01	0.643
Ours	**15.09**	**0.643**	**14.74**	**0.711**

which consists of 25 pairs of non-homogeneous hazy images and clear images. Considering these two datasets have paired GT reference, commonly used image quality evaluation metrics PSNR (dB) and SSIM can be employed to evaluate the dehazing performance. The quantitative comparison of different methods on NH-HAZE and NH-HAZE 2 is recorded in Table 3, it can be found that our method achieves the best results, which further verifies the effectiveness of our method in dehazing real-world hazy images.

Qualitative evaluation. We conduct qualitative comparisons with several image dehazing methods in Figure 11. As we can see, the results of EPDN [32] exhibit a notable amount of colored noise and artifacts, failing to yield clean results. the results of FFA [31] is insufficient in color reproduction and contains noise that affects perception. PSD [9] and RIDCP [39] do not completely remove the haze in the images. The results of AECRNet [38] exhibit high contrast and oversharpening, leading to an unnatural representation of the images. The results dehazed by DAD [33] and D^4 [41] tend to darken the images and fails to restore the degraded details. Especially, the results of DAD on the second and fourth images are the worst, and the perception is greatly reduced. Although KANet [15] gains the best BRISQUE score in quantitative comparison, it tends to lean towards darker color tones and can't get visual pleasing results. In addition, all these compared methods are prone to stratification in the sky region, which is very evident on the second picture. In contrast, our method produces visually pleasing results, achieving the best performance in terms of haze removal, brightness, and detail recovery. More results are included in the supplementary material.

Model complexity and inference time. In addition to the comparison of model dehazing performance, we also compare the complexity and inference efficiency of different dehazing models, which are necessary for real-world dehazing methods but always ignored in previous works. The results

Table 4. Comparison of parameters and latency. Note that latency is measured with images resized to 1080×1920 on a single Quadro RTX 8000 GPU.

Methods	DAD [33]	D^4 [41]	RIDCP [39]	KANet [15]	Ours
#Param(M)	52.06	22.90	29.48	55.66	3.79
Latency(s)	0.688	1.833	2.949	0.829	0.736

are reported in Table 4. It can be seen that our model parameters are much less than those of several other methods. Because of the introduction of DCP and CLAHE, our method is slightly slower than DAD [33], but much faster than the latest D^4 [41], RIDCP [39] and KANet [15].

5.4. Ablation Studies

To verify the effectiveness of significant components proposed in our overall framework, we perform ablation studies with four configurations:

(1) **w/o PFA.** Replace Physics-driven Feature Aggregation Module with simple residual connection.

(2) **w/o PFCM.** The model is trained without prior-based image fusion operation in PFCM.

(3) **w/o MCCD.** Remove the MixCut Consistent Dehazing during the semi-supervised training process.

(4) **w/o full-stage progressive dehazing.** Reduce the dehazing stages during inference, referred to as **Stage1** or **Stage2** in the following.

Qualitative evaluation. We visualize the qualitative results in Figure 12. It can be seen that without using PFA to introduce the physical model of the haze process into the deep feature space, we can see in Figure 12 (b) that our model cannot decompose and learn the real-world haze distribution well, resulting in haze residue. Without prior-based feature compensation, our model fails to generate clear and bright image, as shown in Figure 12 (c). In addition, we verify the effectiveness of MixCut Consistent Dehazing (MCCD) in the comparison of Figure 12 (d) and Figure 12 and it is obvious that MCCD greatly boosts the dehazing effect of our model on real-world hazy images.

Furthermore, we provide an example in Figure 12 (f) to illustrate the effectiveness of multi-stage progressive dehazing. It can be seen that the haze is removed step by step. Note that when increasing the dehazing stages to more than three, there is only a slight performance gain, but at a higher computational cost. Therefore, we set the number of stages as 3 (i.e. K=3) in our work.

Quantitative evaluation. We further operate the quantitative evaluation. We choose NIMA [34] and NIQE [29] metrics to perform the evaluation. As shown in Table 5, each module we designed in our PMDN has the potential to improve the NIMA metric and NIQE metric, demonstrating the effectiveness of physics-driven feature aggregation and the prior knowledge of dehazing process. It also can be seen that the application of MCCD during the semi-supervised

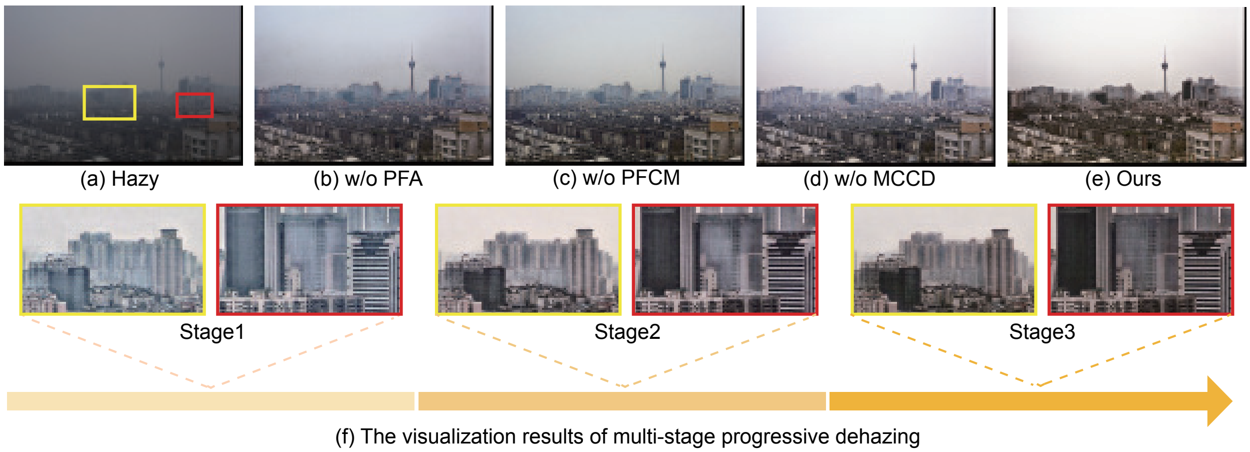

Figure 12. Different configurations and corresponding qualitative results in the ablation study. Each module we designed has the potential to increase the dehazing performance. Best viewed with zooming in.

Table 5. Quantitative ablation study. Incorporating the modules we designed can increase the NIMA [29] and NIQE [29] metrics score.

	Hazy	w/o PFA	w/o PFCM	w/o MCCD	Stage1	Stage2	Ours
NIMA ↑	4.217	4.453	4.380	4.497	4.195	4.311	**4.806**
NIQE ↓	4.063	3.501	3.620	3.471	3.825	3.524	**3.286**

training process can enhance the model's performance.

Moreover, we evaluate each stage's dehazing performance of multi-stage progressive dehazing process. It can be observed that two IQA metrics are improved with the increase in stages, validating the capability of multi-stage progressive dehazing in PMDN.

5.5. Limitation

Although our method provides a more effective solution for the real-world image dehazing task, it occasionally exhibits failures, especially in dense haze scenes. It is unable to produce satisfactory results for a severely degraded scene that the texture structure information in the scene is severely obscured by haze. As generative models continue to mature, developing stronger haze removal algorithms is an important avenue for future research.

6. Conclusion

In this paper, we proposed the Prior-compensated Multi-stage Dehazing Network (PMDN) for real-world image dehazing task with high performance. We defined dehazing as a multi-stage progressive reconstruction problem and introduced a Prior-based Feature Compensation Module (PFCM) to guide the network by non-learnable prior embeddings. To boost the adaptability of the proposed PMDN in diverse real-world haze scenarios, we proposed MixCut Consistent Dehazing (MCCD) strategy based on a cross-cutting scheme and dehazing consistency principle. In addition, we proposed a new real-world image dehazing dataset called RealHQ-HAZE, containing various real-world hazy/high-quality pseudo-GT pairs and additional transferred varicolored hazy images. Extensive experiments show that our RealHQ-HAZE is more suitable for real-world image dehazing task and our PMDN performs favorably against the state-of-the-art image dehazing methods.

7. Acknowledgement

This work was supported in part by the National Major Science and Technology Projects of China under Grant 2009XJTU0016.

References

[1] C. Ancuti, C. O. Ancuti, R. Timofte, and C. D. Vleeschouwer. I-haze: a dehazing benchmark with real hazy and haze-free indoor images. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 620–631. Springer, 2018. 1, 3

[2] C. O. Ancuti, C. Ancuti, M. Sbert, and R. 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. 1, 3, 4

[3] C. O. Ancuti, C. Ancuti, and R. 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*, pages 444–445, 2020. 1, 3, 9, 11

[4] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer. O-haze: a dehazing benchmark

with real hazy and haze-free outdoor images. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 754–762, 2018. 1, 3, 4

[5] C. O. Ancuti, C. Ancuti, F.-A. Vasluianu, and R. Timofte. Ntire 2021 nonhomogeneous dehazing challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 627–646, 2021. 9, 11

[6] T. Brooks, A. Holynski, and A. A. Efros. Instructpix2pix: Learning to follow image editing instructions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18392–18402, 2023. 5

[7] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016. 1

[8] L. Chen, X. Chu, X. Zhang, and J. Sun. Simple baselines for image restoration. *arXiv preprint arXiv:2204.04676*, 2022. 6

[9] Z. Chen, Y. Wang, Y. Yang, and D. Liu. Psd: Principled synthetic-to-real dehazing guided by physical priors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7180–7189, 2021. 1, 3, 4, 8, 9, 11

[10] L. K. Choi, J. You, and A. C. Bovik. Referenceless prediction of perceptual fog density and perceptual image defogging. *IEEE Transactions on Image Processing*, 24(11):3888–3901, 2015. 8, 9

[11] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, and M.-H. Yang. Multi-scale boosted dehazing network with dense feature fusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2157–2167, 2020. 1, 7, 8, 9, 11

[12] J. Dong and J. Pan. Physics-based feature dehazing networks. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings*, pages 188–204. Springer, 2020. 2

[13] Y. Dong, Y. Liu, H. Zhang, S. Chen, and Y. Qiao. Fd-gan: Generative adversarial networks with fusion-discriminator for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 10729–10736, 2020. 1

[14] P. Esser, R. Rombach, and B. Ommer. Taming transformers for high-resolution image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12873–12883, 2021. 3

[15] Y. Feng, L. Ma, X. Meng, F. Zhou, R. Liu, and Z. Su. Advancing real-world image dehazing: perspective, modules, and training. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 3, 9, 11

[16] A. Golts, D. Freedman, and M. Elad. Unsupervised single image dehazing using dark channel prior loss. *IEEE transactions on Image Processing*, 29:2692–2701, 2019. 1

[17] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010. 6

[18] M. Hong, J. Liu, C. Li, and Y. Qu. Uncertainty-driven dehazing network. 2022. 1

[19] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017. 8

[20] K. Islam, M. Z. Zaheer, A. Mahmood, and K. Nandakumar. Diffusemix: Label-preserving data augmentation with diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27621–27630, 2024. 5

[21] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 7

[22] B. Li, Y. Gou, S. Gu, J. Z. Liu, J. T. Zhou, and X. Peng. You only look yourself: Unsupervised and untrained single image dehazing neural network. *International Journal of Computer Vision*, 129:1754–1767, 2021. 1

[23] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng. An all-in-one network for dehazing and beyond. *arXiv preprint arXiv:1707.06543*, 2017. 9, 11

[24] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018. 1, 3, 8, 9

[25] L. Li, Y. Dong, W. Ren, J. Pan, C. Gao, N. Sang, and M.-H. Yang. Semi-supervised image dehazing. *IEEE Transactions on Image Processing*, 29:2766–2779, 2019. 1, 3

[26] Y. Liu and X. Hou. Local multi-scale feature aggregation network for real-time image dehazing. *Pattern Recognition*, 141:109599, 2023. 1

[27] E. J. McCartney. Optics of the atmosphere: scattering by molecules and particles. *New York*, 1976. 1

[28] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, 21(12):4695–4708, 2012. 9

[29] A. Mittal, R. Soundararajan, and A. C. Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012. 9, 11, 12

[30] S. G. Narasimhan and S. K. Nayar. Vision and the atmosphere. *International journal of computer vision*, 48(3):233–254, 2002. 1

[31] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11908–11915, 2020. 1, 3, 6, 8, 9, 11

[32] Y. Qu, Y. Chen, J. Huang, and Y. Xie. Enhanced pix2pix dehazing network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8160–8168, 2019. 1, 4, 8, 9, 11

[33] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang. Domain adaptation for image dehazing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2808–2817, 2020. 1, 3, 4, 9, 11

[34] H. Talebi and P. Milanfar. Nima: Neural image assessment. *IEEE transactions on image processing*, 27(8):3998–4011, 2018. 11

[35] J. Wang, K. C. Chan, and C. C. Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 2555–2563, 2023. 9

[36] S. Wang, Y. Duan, H. Ding, Y.-P. Tan, K.-H. Yap, and J. Yuan. Learning transferable human-object interaction detector with natural language supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 939–948, 2022. 9

[37] Z. Wang, H. Zhao, J. Peng, L. Yao, and K. Zhao. Odcr: Orthogonal decoupling contrastive regularization for unpaired image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25479–25489, 2024. 1, 3

[38] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, and L. Ma. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10551–10560, 2021. 9, 11

[39] R.-Q. Wu, Z.-P. Duan, C.-L. Guo, Z. Chai, and C. Li. Ridcp: Revitalizing real image dehazing via high-quality codebook priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22282–22291, 2023. 1, 3, 9, 11

[40] X. Yang, Z. Xu, and J. Luo. Towards perceptual image dehazing by physics-based disentanglement and adversarial training. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018. 1

[41] Y. Yang, C. Wang, R. Liu, L. Zhang, X. Guo, and D. Tao. Self-augmented unpaired image dehazing via density and depth decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2037–2046, 2022. 3, 9, 11

[42] X. Zhang, H. Dong, J. Pan, C. Zhu, Y. Tai, C. Wang, J. Li, F. Huang, and F. Wang. Learning to restore hazy video: A new real-world dataset and a new method. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9239–9248, 2021. 8

[43] S. Zhao, L. Zhang, S. Huang, Y. Shen, and S. Zhao. Dehazing evaluation: Real-world benchmark datasets, criteria, and baselines. *IEEE Transactions on Image Processing*, 29:6947–6962, 2020. 3

[44] Y. Zheng, J. Zhan, S. He, J. Dong, and Y. Du. Curricular contrastive regularization for physics-aware single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5785–5794, 2023. 2

[45] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 4