Supplementary Materials

Anonymous submission

Introduction

Our supplementary materials give more details and experimental results of our method, which can be summarized as follows:

- We provide detailed information on the training process and how we obtain preliminary dehazed results using physical priors.
- We illustrate more qualitative and quantitative results to demonstrate the superior performance of the proposed method.
- We conduct further ablation studies to validate the effectiveness of our method.

A Details

A.1 Preliminary Dehazed Results

Dark Channel Prior Dark Channel Prior (DCP) is the most famous and effective physical prior for image dehazing. For a real-world hazy image I, We apply 3D Max pooling directly to its additive inverse and use the additive inverse values of the result as the dark channel image I_{dark}. The extraction of the dark channel image can be expressed as:

$$I_{dark} = -Maxpool3D(-I), \tag{1}$$

where $Maxpool3D$ represents the 3D Max pooling operation. The DCP assumes that the dark channel value of a clear image tends towards zero, i.e., $J_{dark} \to 0$. Following this, we can derive the transmission map t_{dcp} from the Atmospheric Scattering Model (ASM), which can be written as:

$$t_{dcp} = 1 - \frac{I_{dark}}{A}, \tag{2}$$

where A denotes the atmospheric light.

To get the atmospheric light A, we first identify the brightest 0.1% pixels in the dark channel image I_{dark}. Subsequently, we can obtain the atmospheric light A by calculating the average pixel values at their corresponding locations in the hazy image I. Utilizing the acquired t_{dcp} and A, we can derive the preliminary image J_{dcp} by reversing the ASM. This process can be formulated as:

$$J_{dcp} = A + \frac{I - A}{t_{dcp}}, \tag{3}$$

Boundary Constraint and Contextual Regularization Boundary Constraint and Contextual Regularization (BCCR) is another physical prior for dehazing that explores the inherent boundary of the transmission map. For a real-world hazy image I, its prior scene radiance is always bounded, that is:

$$C_0 \leq J \leq C_1, \forall x \in \Omega, \tag{4}$$

where C_0 and C_1 are constant vectors that are relevant to the given image I. It is constantly assumed that the extrapolation of a natural image must lie within a radiance cube bounded by the C_0 to C_1. Supposing that atmospheric light A is known, we can get the corresponding boundary J_b. Then we can get the low boundary of the transmission map t_b, leading to the following constraint on t, which can be expressed as:

$$0 \leq t_b \leq t \leq 1, \tag{5}$$

where t_b can be formulated as:

$$t_b = min\left\{\max_{c\in(r,g,b)}\left(\frac{A^c - I^c}{A^c - C_0^c}, \frac{A^c - I^c}{A^c - C_1^c}\right), 1\right\}, \tag{6}$$

According to the assumption that A is brighter than any pixel in the hazy image and $C_0 = 0$, we can get a new transmission map, that is:

$$\hat{t}(x) = \min_{y\in\omega_x}\max_{z\in\omega_y} t_b(z), \tag{7}$$

where x refers to the position of a pixel. Given that the transmission map is invariant to local regions of the image and unsuitable for regions with significant depth variations, a weighted L1 contextual regularization is introduced to optimize the transmission map. With the acquired transmission map t_{bccr} and atmospheric light A, preliminary image J_{bccr} can be derived by reversing the ASM.

A.2 Training Loss

Cycle Consistency Loss Our method encompasses two translation functions, $G(x, C_y) : X \to Y$ and $G(y, C_x) : Y \to X$. Y corresponds to the clear image domain, while X aligns with the hazy image domain. Ideally, if both translation functions are well-learned, an image can be translated to another domain and returned without loss. To ensure this, we utilize the cycle consistency loss to regularize the reconstructed hazy or clear image. Specifically, we employ a combination of pixel-level L1 loss and feature-level LPIPS loss

to minimize the distance between the original input image and its cycle counterpart. This can be expressed as:

$$L_{cyc} = L_{rec}(x, G(G(x, C_y), C_x)) + L_{rec}(y, G(G(y, C_x), C_y)), \tag{8}$$

where x and y are the hazy and clear images. $G(G(x, C_y), C_x)$ and $G(G(y, C_x), C_y)$ represent the cycle hazy and clear images. L_{rec} is the combined distance of L_1 and LPIPS.

GAN Loss In the hazing process, GAN loss is employed to make the generated hazy image indistinguishable from a real-world hazy one. This is accomplished by training a discriminator that uses the CLIP model as a backbone, following the recommendation of Vision-Aided-GAN. Similarly, GAN loss ensures that the dehazed image appears to belong to the clear image domain. It can be expressed as:

$$L_{GAN} = \log D_Y(y)] + \log D_X(x) + \log\left(1 - D_Y(G(x, C_y))\right) + \log\left(1 - D_X(G(y, C_x))\right), \tag{9}$$

By minimizing the GAN loss, hazing and dehazing processes learn to generate images that closely match the target domain.

Identity Loss Identity loss ensures that the generated image retains the details and content of the source domain image, which is crucial for many applications where preserving the input information is vital. By integrating identity loss into the training objective, we enable the network to balance between making the image visually suitable for its domain (hazy or clear) and preserving the original content. The identity loss can be written as:

$$L_{idt} = L_{rec}(G(y, C_y), y) + L_{rec}(G(x, C_x), x), \tag{10}$$

A.3 Implementation Details

We implement our method within the Pytorch framework using Python 3.10, utilizing the AdamW optimizer with a batch size of 1 for network training. We train our framework for 30K iterations, with β_1 set to 0.9, β_2 set to 0.999, and a learning rate l_r of 5e-6. All experiments are conducted on a single 3090 GPU. The training sample is resized to 286×286 and then randomly cropped to 256×256. Additionally, we implement horizontal flipping for data augmentation.

B Experimental Results

We provide more qualitative and quantitative results in this section to further demonstrate the superior performance of the proposed Diff-Dehazer.

B.1 More Results

Table 1 illustrates the quantitative results of URHI and NHAZE. Fig. 1, 2, 3, and 4 present more visual comparisons with several state-of-the-art methods on Fattal's dataset, RTTS, Haze2020, and URHI. As we can see, the proposed Diff-Deahzer achieves satisfactory performance and gets high-quality results with natural color and high contrast.

Moreover, we compare the proposed method with the others in terms of model parameters, FLOPs, (except DCP and BCCR, since they are conventional prior-based methods) in Table 2. Among them, RIDCP and KANet are the extra dehazing methods as suggested by reviewer JLhG. Due to the utilization of pretrained Stable Diffusion, our method inevitably has a large number of parameters and FLOPs. Even so, our method requires less running time than most others. Moreover, our method has significantly fewer FLOPs compared to another diffusion-based method (Diff-Plugin) since the Stable Diffusion Turbo adopted in our method can generate high-quality images in merely a few steps.

We also show the results of RTTS detected by YOLO. Table 2 validates that our dehazed images perform well in downstream tasks, demonstrating the practical benefits of the method in real-world applications. We also highlight the importance of the visual quality of dehazed results, especially in ones closely tied to human decision-making.

B.2 Ablation Study

We employ the SD turbo as the backbone network within our framework. Differently, we design a skipped connection to maintain image fidelity and reduce information loss. Therefore, we conduct an ablation study to validate its effectiveness. As depicted in Fig. 5, we can restore more qualified results with distinguishable textures using the skipped connection.

We illustrate the impact of the guidance scale. We train our method with the guidance as 1, 2.5, 5, 7.5, 9, 10, and 12.5, as depicted in Fig. 6. Our method attains optimal performance across various metrics at a guidance scale of 7.5, except for FID, where it ranks second. Consequently, we adopt a guidance scale of 7.5 for our method.

Additionally, we validate the impact of the weight of physical loss. We train our method with the weight as 0.1, 0.2, 0.5, 0.8, and 1.0, as illustrated in Fig. 7. To achieve a trade-off between the dehazing effect and image over-enhancement, we set the weight as 0.5 in our method.

	URHI				NHAZE			
	FID	NIQE	MUSIQ	CLIPIQA	PSNR	SSIM	LPIPS	VSI
DCP	63.701	4.129	55.966	0.396	13.533	0.599	0.425	0.885
BCCR	65.809	4.184	54.752	0.384	12.939	0.545	0.459	0.885
RefineDNet (TIP2021)	59.701	**3.668**	58.540	0.384	13.676	0.509	0.496	0.874
PSD (CVPR2021)	62.634	4.231	60.417	0.403	11.197	0.520	0.512	0.866
Dehamer (CVPR2022)	56.373	4.260	56.504	0.458	12.170	0.433	0.576	0.838
D4 (CVPR2022)	58.877	4.372	55.958	0.454	12.810	0.433	0.544	0.849
Dehazeformer (TIP2023)	58.154	4.228	57.476	0.455	11.662	0.434	0.570	0.830
MB-Taylorformer (ICCV2023)	55.023	4.174	57.293	0.454	12.092	0.455	0.555	0.837
C2P(CVPR2023)	55.970	4.368	56.542	0.464	12.401	0.442	0.564	0.839
InstructIR (ECCV2024)	55.619	4.460	57.196	**0.465**	12.337	0.442	0.576	0.840
Diffusion-plugin (CVPR2024)	54.669	4.555	53.913	0.442	11.129	0.284	0.601	0.835
Ours	**44.963**	3.730	**62.419**	0.441	**14.159**	**0.656**	**0.391**	**0.901**

Table 1: Quantitative results on URHI and NHAZE.

	Parameters	FLOPs	mAP
DCP	-	-	0.648
BCCR	-	-	0.645
RefineDNet	65.80M	75.41G	0.640
PSD	33.11M	122.7G	0.652
Dehamer	29.44M	25.25G	0.646
D4	10.70M	2.24G	0.639
DehazeFormer	1.28M	13.13G	0.645
C2PNet	7.17M	352.9G	0.646
KANet	55.25M	4.42G	0.644
InstructIR	15.80M	12.39G	0.647
Diff-Plugin	942.61M	1.87T	0.625
Ours	907.89M	173.61G	0.652

Table 2: Quantitative results of model efficiency and detection results.

Figure 1: Visual comparison of samples from Fattal's dataset.

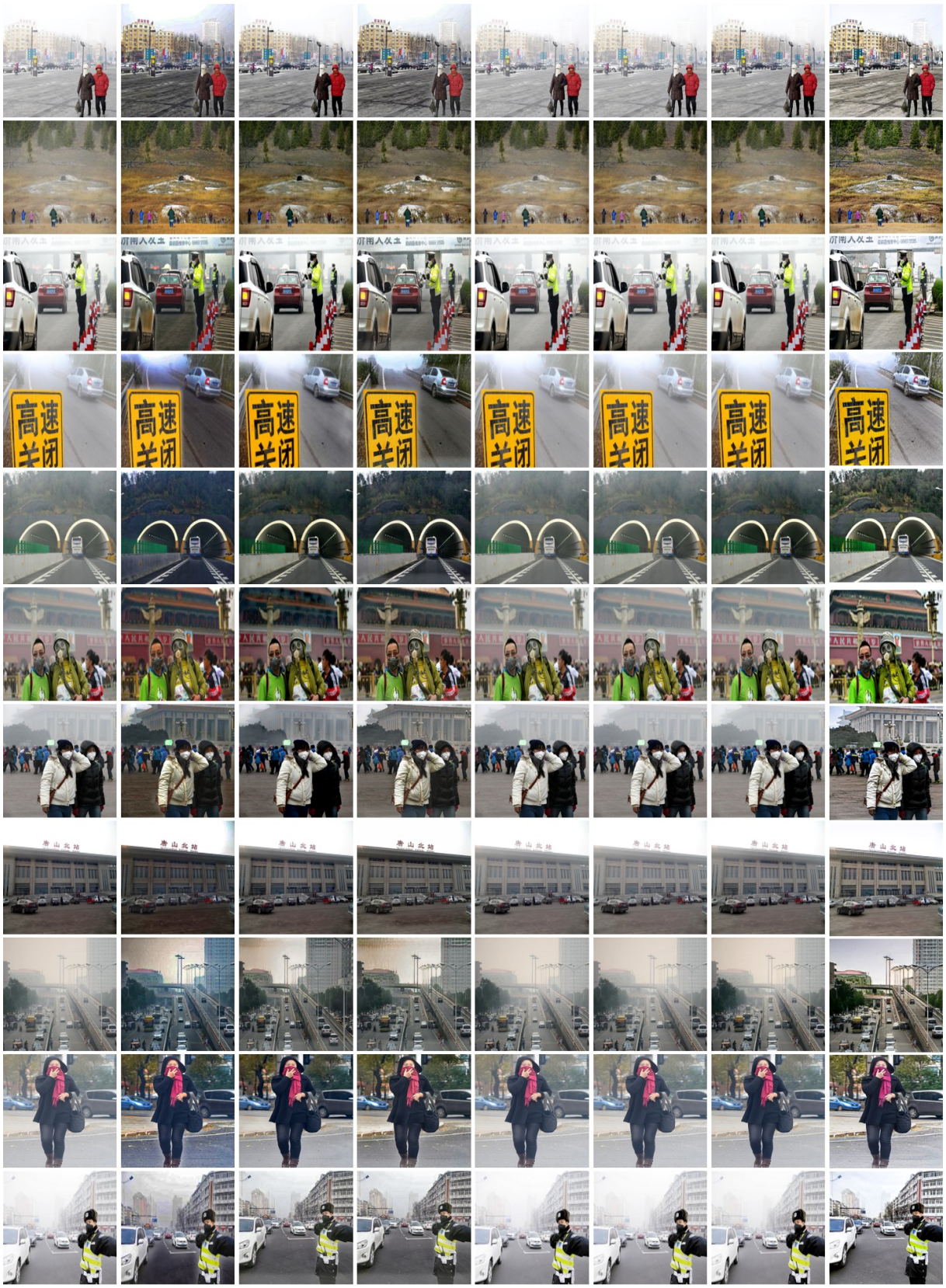


Figure 2: Visual comparison of samples from RTTS.

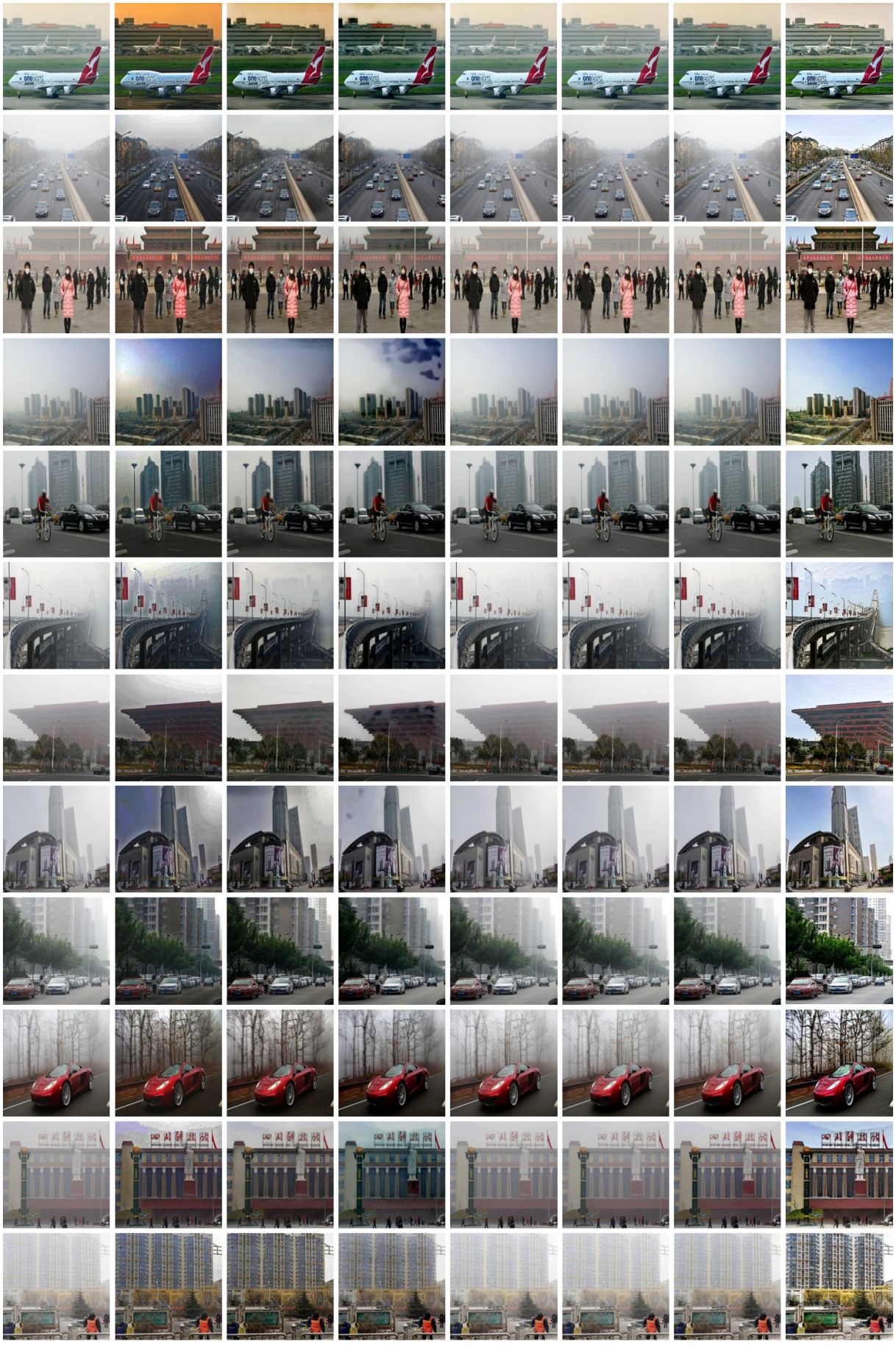


Figure 3: Visual comparison of samples from Haze2020.

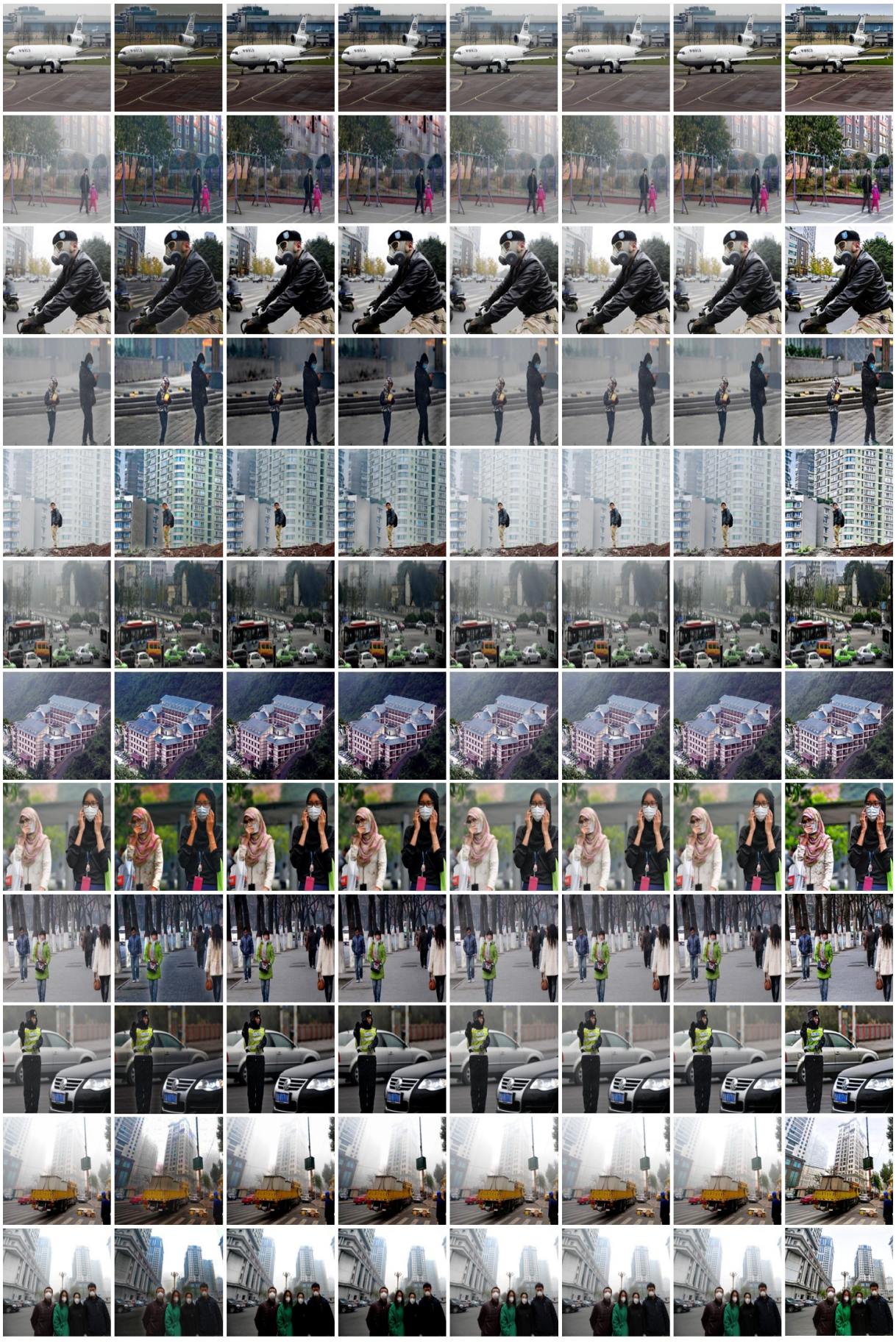


Figure 4: Visual comparison of samples from URHI.

(a) (b) (c)

Figure 5: Ablation study of skipped connection. (a) Hazy image. (b) The result without skipped connection. (c) The result with skipped connection.

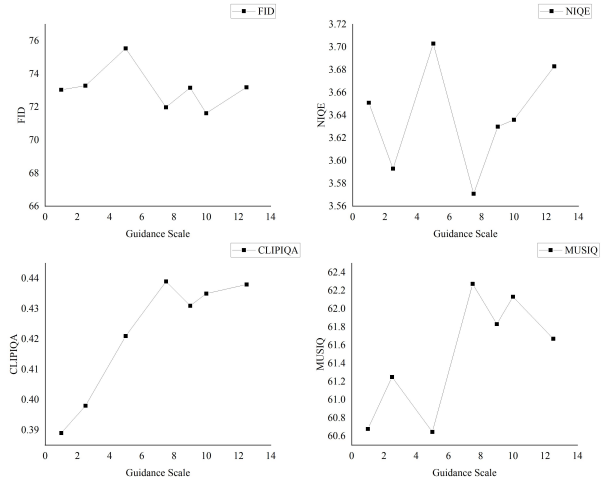
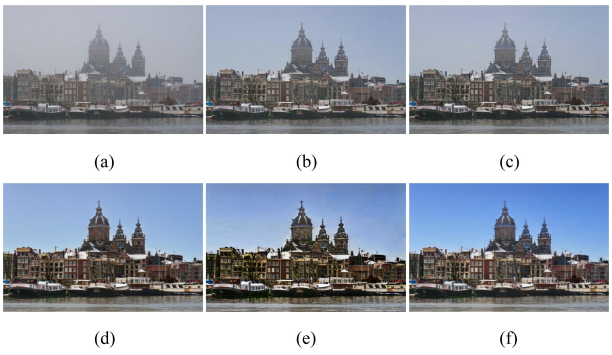


Figure 6: Impact of guidance scale.



(a) (b) (c)

(d) (e) (f)

Figure 7: Ablation study of the weight of physical loss. (a) Hazy image. (b), (c), (d), (e), and (f) are results with the weight, 0.1, 0.2, 0.5, 0.8, and 1.