HazeFlow: Revisit Haze Physical Model as ODE and Non-Homogeneous Haze Generation for Real-World Dehazing

Supplementary Material

S1. Reflow and Distillation

For a more detailed understanding, we provide illustrations depicting the Reflow and Distillation processes Fig 6.

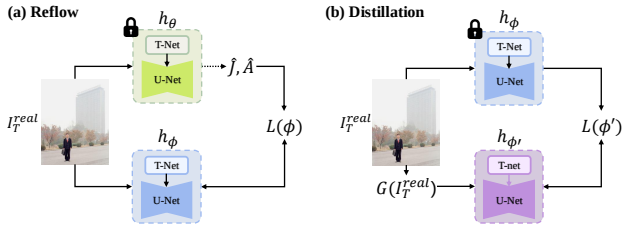

Figure 6. Illustration of (a) Reflow and (b) Distillation in our 3-stage learning framework.

S1.1. Implementation Details

In Reflow and Distillation phase, We train HazeFlow for 10K iterations on the URHI [31], using Adam [28] optimizer. We use learning rate of 1×10^{-6} for Reflow and 5×10^{-5} for Distillation. Other details are consistent with those described in the pretraining phase.

S1.2. Additional Visual Results

For a phase-by-phase comparison, we present visual results on the RTTS dataset [31] in Fig. 12. During the pretraining phase, the network focuses primarily on learning to remove haze effectively. In the Reflow phase, the network is trained with pseudo atmospheric light $\hat{A}$, enabling it to restore appropriate lighting in hazy regions effectively. In the Distillation phase, as shown in Fig. 12, this stage removes artifacts and enhances the natural appearance of the results.

S2. Transmission Refinement

We utilize the refined transmission map $T_{refined}$ across multiple stages, including perceptual loss calculation and sampling during inference, Reflow and Distillation. Specifically, the perceptual loss described in Eq. 13, utilizing the refined transmission map, can be written as follows:

$$L_{perc}(\theta) = \mathbb{E}[\mathbb{D}(J,\ I_T + (1 - T_{refined})h_\theta(I_T, T_{DCP}))]. \tag{24}$$

Similarly, the inference process, as described in Eq. 8, can be expressed as:

$$I_{T+(1-T)\frac{1}{N}} = I_T + (1 - T_{refined})\frac{1}{N}h_\theta(I_T, T_{DCP}). \tag{25}$$

The obtained transmission maps $T_{refined}$ are also utilized for sampling pseudo images during the Reflow phase. The sampling process for the pseudo clean image $\hat{J}$ and pseudo atmospheric light $\hat{A}$ can be expressed as:

$$\hat{J} = I_T^{real} + (1 - T_{refined}) \cdot h_\theta(I_T^{real}, T_{DCP}), \tag{26}$$

$$\hat{A} = I_T^{real} - T_{refined} \cdot h_\theta(I_T^{real}, T_{DCP}). \tag{27}$$

Also, in Distillation phase, we use refined transmission map $T_{refined}$ instead of transmission map obtained by DCP.

$$L_{Distill}(\phi') = \mathbb{E}\big[\mathbb{D}(I_T + (1 - T_{refined})h_\phi(I_T, T_{DCP}), \tag{28}$$

$$\mathcal{G}(I_T) + (1 - T_{refined})h_{\phi'}(\mathcal{G}(I_T), T_{DCP}))\big]. \tag{29}$$

The refined transmission map $T_{refined}$ provides a more accurate estimation of haze density compared to T_{DCP}. As a result, training with $T_{refined}$ enables the model to remove haze more effectively and accurately.

As shown in Fig. 11, our refined transmission map captures haze more effectively. DCP struggles to capture haze in regions with high depth, leaving residual haze in such areas, and it also fails to effectively remove haze in sky regions due to its limitations. In contrast, our method accurately captures haze density, enabling effective haze removal not only in high-depth regions but also in sky areas.

S3. MCBM

S3.1. Implementation Details

For a more detailed explanation of MCBM haze generation process, the overall pipeline is illustrated in Fig. 7. To obtain diverse shapes of non-homogeneous haze density, we treat the number of iterations n and the strength of the Gaussian filter σ as hyperparameters. Specifically, the iteration n represents the number of times the Markov chain and Brownian motion are performed, and a higher n leads to reduced non-homogeneity. We randomly select n by multiplying the total number of pixels by factors of $[4, 5, 6]$. Additionally, to create realistic haze density, it is necessary to smooth to the 2D array generated by the Markov chain and Brownian motion. We apply smoothing using a Gaussian filter, where the strength σ is randomly selected from $[15, 25, 35]$. Finally, the MCBM haze density is generated through normalization.

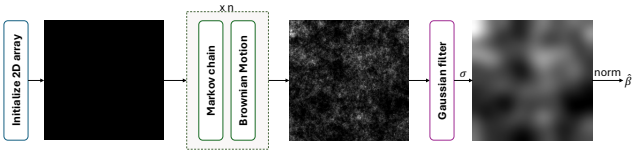

Figure 7. Overall pipeline of MCBM.

S3.2. Additional Visual Results

To demonstrate the effectiveness of our MCBM haze synthesis, we provide additional visual results. For a fair comparison, we evaluate both networks without applying our transmission refinement process on the RTTS [31]. As shown in Fig. 13, the network trained with our MCBM haze synthesis produces significantly clearer results. By learning to capture non-homogeneous haze, the network effectively removes haze more comprehensively.

S3.3. t-SNE Visualization of MCBM Haze

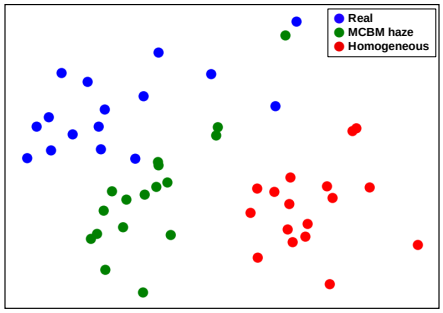

Figure 8. Visualization of feature distributions of pretrained network [48] using t-SNE, comparing real hazy images (blue) obtained from NH-HAZE [4], synthesized non-homogeneous hazy images (green) generated by our MCBM haze density, and synthesized homogeneous hazy images (red).

To verify whether MCBM can closely approximate real-world hazy images, we provide a visualization using t-SNE. As shown in Fig. 8, non-homogeneous hazy images generated by MCBM haze density is much closer to real-world hazy image in the feature space of VGG networks [48], compared to homogeneous hazy image. This approach enables a closer approximation to real haze conditions and the resulting synthetic images facilitate the effective learning of our dehazing networks for real-world scenarios.

S4. User Study

Method	DAD [46]	PSD [12]	D4 [58]	RIDCP [56]	CORUN [18]	HazeFlow
Score↑	0.0273	0.0076	0.0440	0.1214	0.1730	**0.6267**

Table 7. User study result on RTTS and Fattal's datasets.

To provide a more thorough comparison, we conduct a user study. In this study, we randomly select 4 images from Fattal's dataset and 26 images from the RTTS dataset. Participants are asked to evaluate the results based on three criteria: (1) the completeness of haze removal, (2) the absence of artifacts, and (3) the quality of color restoration. They then chose best performing image from the results of our method and those of the state-of-the-art models. The study involved 5 image processing experts and 17 non-expert participants. As shown in Tab. 7, our method received the majority of votes, with a significant margin over the second-ranked model. This result demonstrates that our HazeFlow effectively removes haze, even according to human perception.

S5. Discussion on Estimation Steps

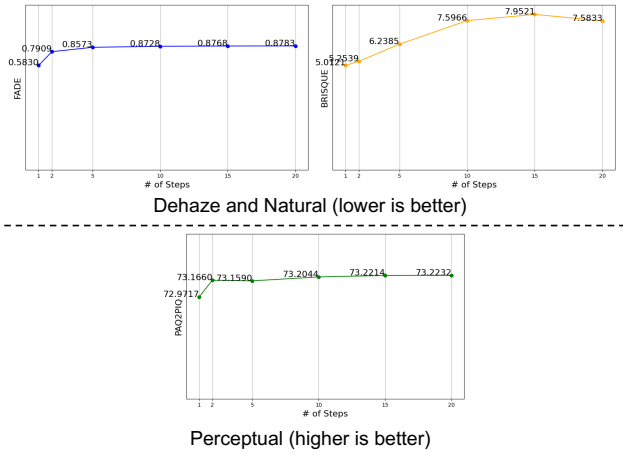

Figure 9. Variation in metrics based on the number of estimation steps.

HazeFlow can predict results using multiple estimation steps, similar to other RF-based models [35, 53, 66]. This section explains the impact of the number of estimation steps on the results. For thorough evaluation, we use four metrics: FADE [13] and BRISQUE [40] assess how well the predicted image has been dehazed and how close it is to the natural image, while PAQ2PIQ [59] evaluates the perceptual quality of the predicted images. As shown in Fig. 9, as the number of steps increases, FADE and BRISQUE scores degrade, while PAQ2PIQ scores improve. This indicates that there is a trade-off between the degree of dehazing and the perceptual quality in the predicted image. Although increasing the number of estimation steps may enhance perceptual quality, this study focuses on achieving effective dehazing while minimizing computational cost. Consequently, one-step estimation is chosen as the final approach. Predicted images for different estimation steps can be seen in Fig. 10.

Figure 10. Visualization of results with varying numbers of estimation steps.

S6. Comparison with Rectified Flow

In this section, we demonstrate that the ASM-based ODE outperforms the naive ODE derived from linear interpolation by comparing it with the baseline, RF [35]. To train RF, we assume the hazy distribution as X_0 and the clean distribution as X_1 in Sec. 3.1. For a fair comparison, Reflow and Distillation are conducted as Eq. (17) and Eq. (19) using one-step estimation. We provide a quantitative comparison in Tab. 8, which shows that HazeFlow outperforms RF in both dehazing capability and perceptual quality. Additionally, a visual comparison is provided in Fig. 14.

Method	FADE↓	BRISQUE↓	PAQ2PIQ↑	MUSIQ↑
Rectified Flow	1.059	32.21	69.81	54.16
HazeFlow	**0.583**	**5.01**	**72.97**	**63.94**

Table 8. Quantitative comparison between RF [35] and our HazeFlow.

S7. Other Transmission Estimation Methods

Although DCP [22] is used to approximate transmission map in our method, various alternative transmission estimation methods can also be employed. Tab. 9 compares different transmission estimation methods including DCP [22], the prior-based method Non-Local [6] and the model-based method DCPDN [61]. We selected DCP due to its high FADE score to enhance dehazing quality, but other methods can be chosen depending on the objective. Note that selecting the initial transmission map offers flexibility and the potential for further performance improvements through alternative approaches.

Method	FADE↓	BRISQUE↓	NIMA↑	PAQ2PIQ↑
DCP [22]	0.583	5.01	5.30	72.97
DCPDN [61]	0.713	4.94	5.19	71.95
Non-local [6]	0.734	1.38	5.16	73.11

Table 9. Ablation on different transmission estimation methods.

S8. Additional Quantitative Results in paired datasets.

We also conduct a comparison without additional training on paired datasets. As shown in Tab. 10, HazeFlow achieves the best performance in both PSNR and SSIM across all datasets except Dense-HAZE. HazeFlow achieves improvements of 0.43 dB in PSNR on O-HAZE and 0.05 in SSIM on NH-HAZE compared to the second-best methods, indicating a significant boost in reconstruction fidelity and structural accuracy. Also, on Dense-HAZE, our method achieves the highest SSIM score while maintaining a comparable PSNR, whereas DAD falls short in SSIM. These results indicates that our approach not only outperforms existing methods but also produces superior perceptual quality with fewer artifacts and sharper details.

S9. Additional Visual Results with SOTA

For a comprehensive comparison with other models, we provide additional visual comparison on RTTS [31] in Fig. 15 and Fattal's dataset [19] in Fig. 16. These figures demonstrate that HazeFlow removes haze more effectively and reduces artifacts better compared to other models. Notably, HazeFlow uniquely preserves structural details in heavily obscured distant regions (e.g., buildings, trees), where competing methods fail to recover fine edges. Furthermore, we provide additional comparisons for paired datasets. Visual results for NH-HAZE can be found in Fig.17, and Dense-HAZE in Fig.18. These results highlight that HazeFlow removes deeper haze and significantly reduces artifacts compared to other models.

Method	NH-HAZE [4]		Dense-HAZE [3]		I-HAZE [1]		O-HAZE [2]	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
DAD [46]	14.34	0.56	**13.51**	0.46	18.02	0.80	18.36	0.75
PSD [12]	10.62	0.52	9.74	0.43	13.79	0.74	11.66	0.68
D4 [58]	12.67	0.50	11.50	0.45	15.64	0.73	16.96	0.72
RIDCP [56]	12.32	0.53	9.85	0.45	16.88	0.78	16.52	0.72
CORUN [18]	11.87	0.56	9.47	0.52	17.14	0.83	18.20	0.83
HazeFlow	**14.49**	**0.61**	11.39	**0.56**	**18.37**	**0.83**	**18.79**	**0.84**

Table 10. Quantitative results on paired dataset (NH-HAZE [4], Dense-HAZE [3], I-HAZE [1],O-HAZE [2]). Best results are **bolded**.

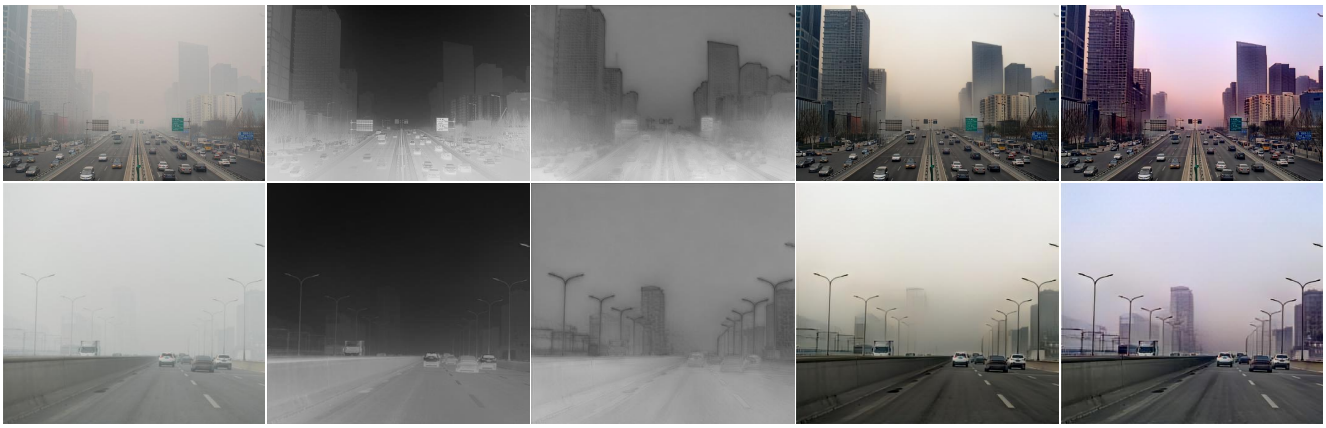

Figure 11. Visual comparison between results with T_{DCP} and with $T_{refined}$ on the RTTS dataset [31].

Figure 12. Visual comparison between the three phases of our method on the RTTS dataset [31].

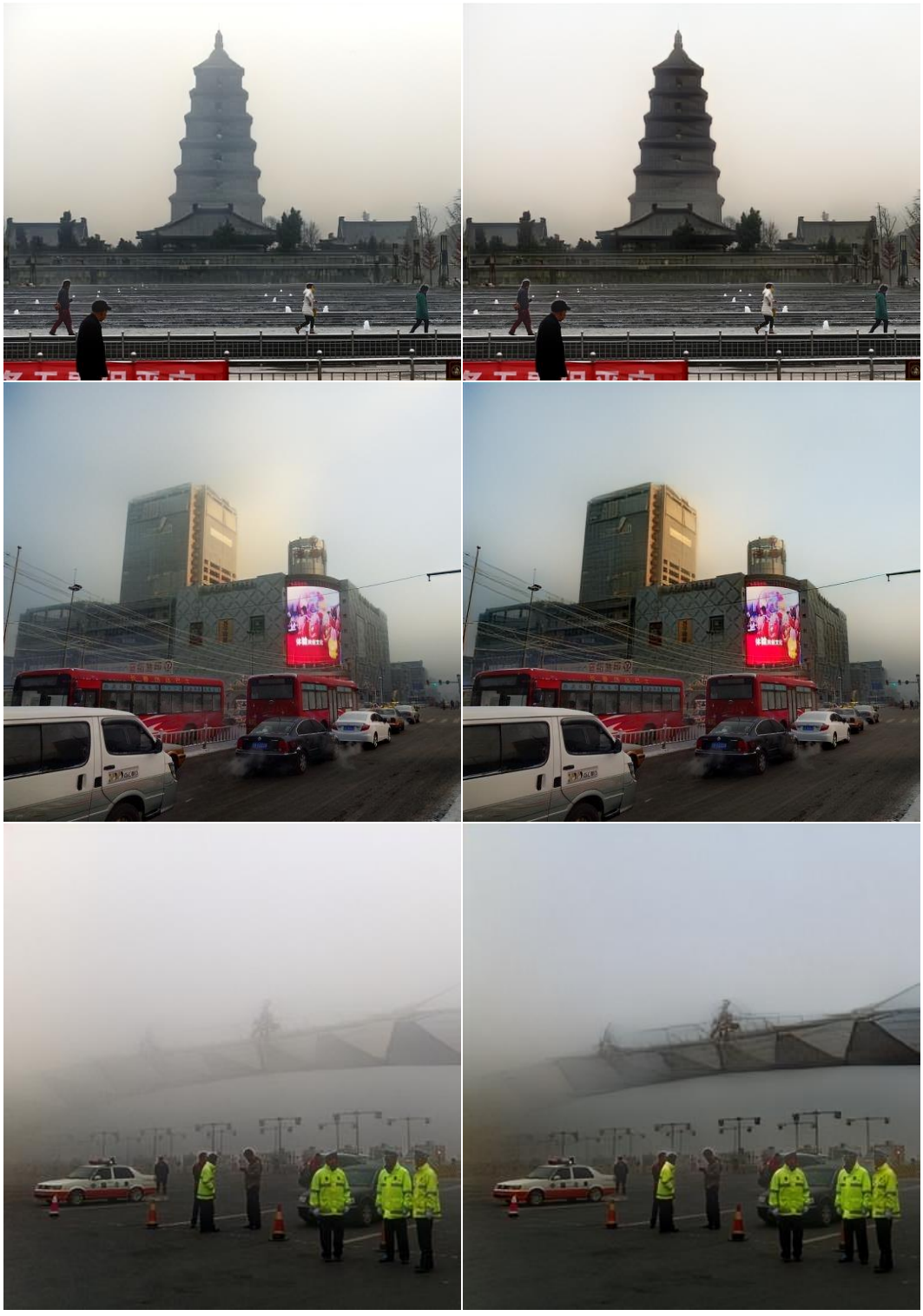

Figure 13. Visual comparison of the results from networks trained with and without MCBM haze on the RTTS dataset [31].

Figure 14. Visual comparison between Rectified Flow [35] and HazeFlow (Ours) on RTTS dataset [31].

Figure 15. Additional visual comparison on RTTS dataset [31]

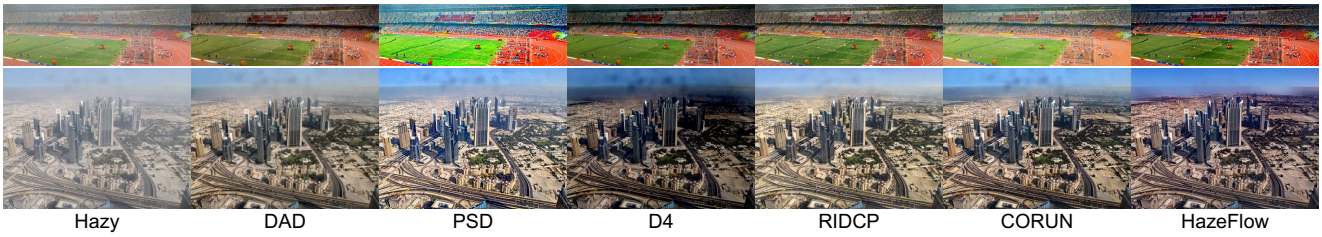

Figure 16. Additional visual comparison on Fattal's dataset [19]

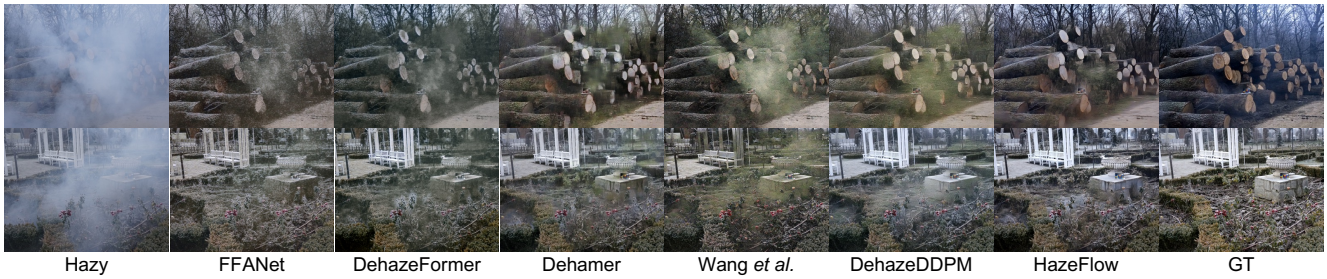

Figure 17. Additional visual comparison on NH-HAZE dataset [4].

Figure 18. Additional visual comparison on Dense-HAZE dataset [3].