

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

Anonymous ICCV submission

Paper ID 8542

## Abstract

**001** *Dehazing involves removing haze or fog from images to*  
**002** *restore clarity and improve visibility by estimating atmo-*  
**003** *spheric scattering effects. While deep learning methods*  
**004** *show promise, the lack of paired real-world training data*  
**005** *and the resulting domain gap hinder generalization to real-*  
**006** *world scenarios. In this context, physics-grounded learn-*  
**007** *ing becomes crucial; however, traditional methods based*  
**008** *on the Atmospheric Scattering Model (ASM) often fall short*  
**009** *in handling real-world complexities and diverse haze pat-*  
**010** *terns. To solve this problem, we propose HazeFlow, a novel*  
**011** *ODE-based framework that reformulates ASM as an ordi-*  
**012** *nary differential equation (ODE). Inspired by Rectified*  
**013** *Flow (RF), HazeFlow learns an optimal ODE trajectory to*  
**014** *map hazy images to clean ones, enhancing real-world de-*  
**015** *hazing performance with only a single inference step. Addi-*  
**016** *tionally, we introduce a non-homogeneous haze generation*  
**017** *method using Markov Chain Brownian Motion (MCBM) to*  
**018** *address the scarcity of paired real-world data. By simula-*  
**019** *ting realistic haze patterns through MCBM, we enhance the*  
**020** *adaptability of HazeFlow to diverse real-world scenarios.*  
**021** *Through extensive experiments, we demonstrate that Haze-*  
**022** *Flow achieves state-of-the-art performance across various*  
**023** *real-world dehazing benchmark datasets.*

## 1. Introduction

Haze considerably diminishes visual quality, thereby posing challenges for diverse real-world applications, including autonomous driving, aerial surveillance, and the analysis of outdoor scenes [25, 32]. The non-homogeneous, spatially varying nature of real-world haze makes dehazing even more complex. Despite extensive research, there remains a substantial performance gap between synthetic and real-world scenarios.

Hazy images are typically represented by the Atmospheric Scattering Model (ASM) [41, 42], which captures the interaction of light with haze particles as it travels through the atmosphere. ASM is formulated as:

$$I(x) = T(x)J(x) + (1 - T(x))A, \quad (1)$$

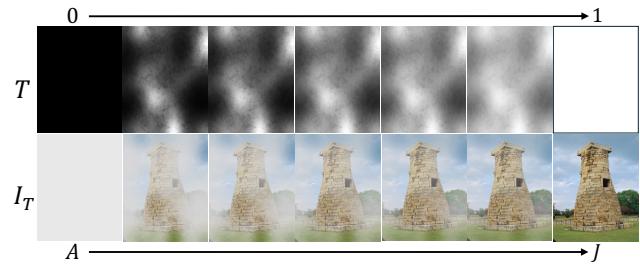


Figure 1. Illustration of dehazing trajectory of hazy image  $I_T$  as transmission map  $T$  gradually increases from 0 to 1.

where  $x$  represents a pixel location,  $I$  denotes the observed hazy image,  $J$  is the clear scene radiance,  $A$  is the global atmospheric light, and  $T$  is the transmission map. The transmission map quantifies the degree of haze at the pixel location  $x$ , and is defined as:

$$T(x) = e^{-\beta d(x)}, \quad (2)$$

where  $\beta$  denotes the haze density coefficient and  $d(x)$  represents the scene depth. For example, as shown in Fig. 1,  $T = 0$  results in complete opacity with  $I$  converge to  $A$ , while  $T = 1$  reveals a clear scene with  $I$  matches  $J$ .

Traditional physical-based haze removal methods [20, 22, 65], often rely on ASM, provide a principled understanding of how haze affects images, but physical models alone struggle to capture the complexity of real-world haze. On the other hand, deep learning approaches [8, 21, 26, 30, 43] trained on large-scale synthetic datasets [31] achieve impressive results on curated benchmarks, but fall short in real-world settings due to domain gap. Techniques such as data augmentation-based approaches [9], CycleGAN-based methods [11, 64], and recent diffusion-based models [38, 63] have indeed improved robustness. However, handling real-world haze artifacts and distribution shifts remains a major challenge.

To fully leverage both physical models and deep learning methodologies, we introduce HazeFlow, a novel ODE-based dehazing framework that redefines dehazing as an optimal transport process between haze and clean image distributions. Instead of directly solving an inverse prob-

066 lem of ASM, we reformulate dehazing as an ordinary dif-  
067 ferential equation (ODE), where the velocity field guides  
068 hazy images along a physically grounded path to their clean  
069 counterparts. This reformulation enables the utilization of  
070 contemporary neural ODE methodologies to facilitate the  
071 training of a dehazing network across the complete dehaz-  
072 ing trajectory, thereby enhancing the stability, accuracy, and  
073 efficiency of haze removal. Motivated by Rectified Flow  
074 (RF) [35] which learns straight, fast sampling trajectories,  
075 we train a neural network to approximate the optimal ODE  
076 trajectory, allowing fast and accurate removal of the haze in  
077 a single inference step.

078 Furthermore, HazeFlow enhances real-world perfor-  
079 mance with a three-stage learning process: Pretrain, Re-  
080 flow, and Distillation. Initially, the Pretrain phase on large-  
081 scale synthesized hazy-clean image pairs constructs a map-  
082 ping from hazy images to their clean counterparts. The Re-  
083 flow stage fine-tunes the model using real hazy and pseudo  
084 clean image pairs, capturing complex real-world haze pat-  
085 terns. Finally, the Distillation stage further improves robust-  
086 ness to real-world artifacts and enhances perceptual qual-  
087 ity. Through three-stage learning, HazeFlow develops a dy-  
088 namic, transmission-aware mapping, enhancing its under-  
089 standing of real-world haze. HazeFlow demonstrates supe-  
090 rior adaptability to real-world scenarios, achieving state-of-  
091 the-art results in both synthetic and real-world benchmark  
092 datasets.

093 In addition, we tackle the issue of the limited avail-  
094 ability of paired real-world datasets. Datasets containing  
095 non-homogeneous haze (e.g., smoke-induced haze) are  
096 scarce, and most synthetic datasets assume homogeneous  
097 haze with a constant scattering coefficient  $\beta$  across the  
098 scene [31, 45, 56]. To mitigate this limitation, we intro-  
099 duce a non-homogeneous haze generation method that uti-  
100 lizes Markov Chain Brownian Motion (MCBM) which ef-  
101 fectively models the random movements of airborne parti-  
102 cles [27, 52]. Specifically, we simulate a spatially varying  
103 scattering coefficient for transmission map using MCBM.  
104 This provides a more realistic and diverse dataset, allowing  
105 the model to generalize better to real-world haze.

106 The contributions of this paper are as follows.

- 107 • For the first time, we reformulate the dehazing process  
108 as solving the ODE of ASM, enabling the learning of a  
109 physics-guided flow. Our three-stage training pipeline en-  
110 hances real-world generalization and achieves stable one-  
111 step estimation.
- 112 • We propose a non-homogeneous haze image generation  
113 technique with MCBM modeling to mitigate the limited  
114 availability of paired real-world hazy datasets.
- 115 • Our method achieves state-of-the-art performance in real-  
116 world dehazing tasks across various benchmark datasets,  
117 demonstrating exceptional efficacy in practical applica-  
118 tions.

## 2. Related Works

### 2.1. Single Image Dehazing

121 Early methods relied on assumptions or priors to estimate  
122 the haze. DCP [22] assumes that haze-free images have at  
123 least one low intensity channel. Other priors, such as the  
124 color attenuation prior [65], color-lines [20], Haze-lines [7]  
125 used color and intensity gradients to predict transmission  
126 map. However, these methods often struggle with complex  
127 scenes, sky regions, and non-uniform haze, producing un-  
128 desired artifacts.

129 Recently, deep learning approaches have significantly  
130 improved dehazing performance by directly learning to re-  
131 move haze from data. Early methods relied heavily on Con-  
132 volutional Neural Networks (CNNs), which are effective  
133 in capturing local characteristics and hierarchical informa-  
134 tion [8, 16, 30, 34, 43, 44, 55]. Recent research has explored  
135 Transformer-based architectures, which excel in modeling  
136 global relationships and capturing long-range dependencies  
137 within the image [21, 50]. However, these approaches face  
138 challenges when applied to real-world hazy images, as they  
139 often overlook the underlying physics model of haze forma-  
140 tion. Without explicitly modeling the atmospheric scatter-  
141 ing process and spatially varying transmission, these meth-  
142 ods can struggle to generalize across diverse haze patterns,  
143 leading to incomplete dehazing or distortion of fine details.

### 2.2. Real-World Dehazing

144 To mitigate the domain gap between synthetic and real-  
145 world hazy images, various approaches have been proposed.  
146 Methods based on CycleGAN [11, 33, 46, 58] utilize cycle  
147 consistency loss [64] to tailor models for real-world scenar-  
148 os by employing unpaired datasets of hazy and clean im-  
149 ages. On the other hand, RIDCP [56] tackled the scarcity  
150 of real-world haze data by generating a dataset with syn-  
151 thesized haze applied to hundreds of clean images. Despite  
152 these efforts, challenges persist in effectively modeling non-  
153 homogeneous real-world haze. PANet [9] attempts to ad-  
154 dress this issue by augmenting non-homogeneous haze.  
155 Specifically, it modifies the values of  $\beta$  or  $A$  to augment  
156 real-world haze. However, this method is limited to aug-  
157 mentations within limited datasets. Furthermore, it strug-  
158 gles to fully encompass the diverse haze conditions encoun-  
159 tered in real-world scenarios such as in RTTS [31]. In this  
160 work, we aim to address these limitations by providing an  
161 enhanced dataset with more diverse haze characteristics, en-  
162 abling more reliable training and generalization.

164 **Diffusion-based Image Restoration** Recently, diffusion  
165 models have shown remarkable success in image restora-  
166 tion tasks by iteratively refining images through learned  
167 denoising processes. To address real-world dehazing, De-  
168 hazeDDPM [60] and Wang *et al.* [54] leverage the diffusion  
169 model [24] to improve dehazing performance by applying

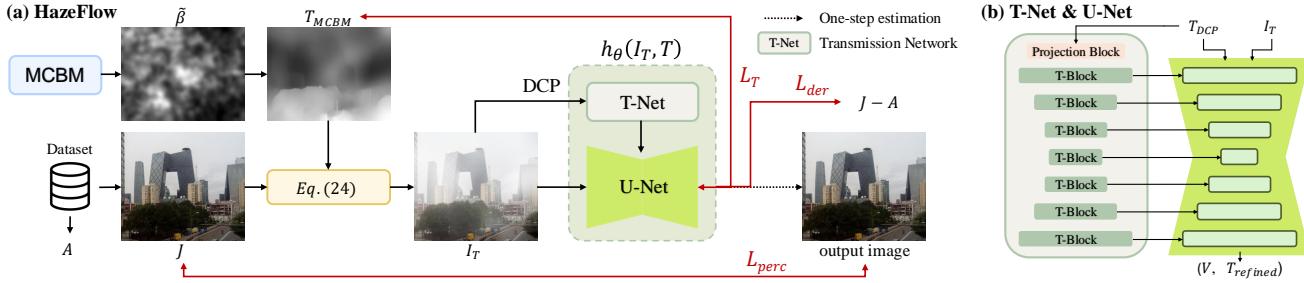


Figure 2. Overview of our proposed method, illustrating (a) HazeFlow and (b) Transmission Network.

conditional diffusion models along with additional refinement techniques. DACLIP [38] and DiffUIR [63] proposed unified frameworks capable of handling multiple restoration tasks, including dehazing through a shared diffusion process. Despite their impressive results, these diffusion-based methods face limitations in inference speed and computational efficiency. In contrast, our methodology employs a flow-based strategy that uses a well-defined ODE formulation, significantly decreasing the number of sampling steps required during the inference process while preserving superior restoration quality.

### 3. ODE-based Dehazing Model

This section begins with a concise summary of RF, which serves as an inspiration for our study, and is followed by an in-depth description of our proposed method, HazeFlow.

#### 3.1. Preliminary

RF solves the optimal transport problem between two data distributions  $\pi_0$  and  $\pi_1$  by transporting data from an initial state  $X_0 \sim \pi_0$  to a target state  $X_1 \sim \pi_1$  over a continuous time interval  $t \in [0, 1]$ . Given observations of  $X_0$  and  $X_1$ , RF learns the velocity field by training a neural network  $r_\theta$  to follow the linear interpolation of  $X_t = tX_1 + (1-t)X_0$ . This is achieved by solving the following least squares regression problem:

$$L(\theta) = \min_{\theta} \int_0^1 \mathbb{E}[||(X_1 - X_0) - r_\theta(X_t, t)||^2] dt. \quad (3)$$

After training the velocity function  $r_\theta$ , target  $X_1$  can be inferred by a conventional ODE-solver (e.g. Euler solver):

$$X_{k+\frac{1}{N}} = X_k + \frac{1}{N}r_\theta(X_k, k), \text{ s.t. } k \in \left\{0, \frac{1}{N}, \dots, \frac{N-1}{N}\right\}, \quad (4)$$

where  $N$  denotes number of steps from  $X_0$  to  $X_1$ , and  $k$  is the discrete step index.

**Reflow.** To straighten the trajectory and yield fast simulation, RF introduces Reflow mechanism. Using  $r_\theta$  trained on pairs of  $(X_1, X_0)$ , new synthetic pairs  $(\hat{X}_1, X_0)$  can be generated, where  $\hat{X}_1$  is obtained from  $X_0$  using a conventional ODE solver (i.e.,  $ODE[r_\theta](X_0)$ ). The newly generated pair is used for training to refine the velocity field through

$r_\phi$ . This process is repeated iteratively to streamline the path. The aim of the objective function is to ensure that the predictions of  $r_\phi$  correspond with the velocity field from  $X_0$  to  $\hat{X}_1$ , resulting in more direct trajectories, as follows:

$$L(\phi) = \min_{\phi} \int_0^1 \mathbb{E}[||(\hat{X}_1 - X_0) - r_\phi(X_t, t)||^2] dt, \quad (5)$$

where  $X_t = t\hat{X}_1 + (1-t)X_0$ . By refining the path using data pairs derived from solving the ODE, the newly trained  $r_\phi$  achieves accurate results with fewer inference steps.

**Distillation.** Using  $r_\phi$  from Reflow, a network  $r_{\phi'}$  requiring one-step inference can be obtained via knowledge distillation, as follows:

$$L(\phi') = \mathbb{E}[\mathbb{D}(ODE[r_\phi](X_0), X_0 + r_{\phi'}(X_0, 0))], \quad (6)$$

where the loss function  $\mathbb{D}$  uses LPIPS [62], which enhances the perceptual quality of images produced. Through these Reflow and Distillation processes, RF enables high-quality one-step generation and has been successfully applied in various tasks [36, 53, 66].

#### 3.2. HazeFlow

**Revisit ASM as an ODE.** In this work, our key observation is that the conventional ASM in Eq. 1 can be reformulated as an ODE:

$$dI_T = (J - A)dT, \quad (7)$$

where  $I_T$  is the hazy image corresponding to the transmission map  $T$ . This reformulation reveals that dehazing fundamentally involves modeling incremental changes in  $T$  and proportional adjustments in  $I_T$ . In this view, the dehazing process becomes a velocity estimation problem, where the goal is to learn a vector field that guides the evolution of the hazy image towards the clean image along the  $T$  path. We achieve this by parameterizing the derivative function using RF [35] through a learnable neural network  $h_\theta$ , allowing the model to approximate the optimal transport trajectory from atmospheric light to clean image as in Fig.1.

Using the RF framework, our method produces diverse and high-quality results and drastically reduces computational costs compared to conventional iterative models (e.g., diffusion models). Once  $h_\theta$  is trained, clean image  $J$  can be

243 directly predicted from the hazy image  $I_T$  with an off-the-  
244 shelf Euler ODE-solver as follows:

$$\begin{aligned} I_{\tau+\frac{1}{M}} &= I_\tau + \frac{1}{M} h_\theta(I_\tau, \tau), \\ \text{s.t. } \tau &\in \left\{ T, T + \frac{1}{M}, \dots, T + \frac{N-1}{M} \right\}, \end{aligned} \quad (8)$$

245 where  $\frac{1}{M}$  denotes the adaptive step size, defined as  $\frac{1-T}{N}$ ,  
246 where  $N$  is the number of sampling steps from  $I_T$  to  $I_1 (= J)$ . Unlike conventional RF-based models [36, 53, 66],  
247 which apply a time variable  $t$  uniformly across the entire  
248 image, our method assigns a locally varying transmission  
249 map  $T$  across images. Since the degree of attenuation varies  
250 for each pixel in a hazy image, the transmission map  $T$  ex-  
251 hibits spatial variability in the image. This spatial varia-  
252 tion is managed through the adaptive step size: areas with  
253 thicker haze (*i.e.*, lower  $T(x)$ ) undergo more transforma-  
254 tions, whereas regions with lighter haze (*i.e.*, higher  $T(x)$ )  
255 require fewer updates, thereby preserving finer details.  
256

### 258 3.3. Network Architecture

259 Our HazeFlow consists of a U-Net and a T-Net, as shown in  
260 Fig. 2 (b). The U-Net takes the hazy image and transmission  
261 map derived from the hazy image as inputs, while the T-Net  
262 processes the transmission map. Input transmission map  
263  $T_{DCP}$  is estimated using DCP [22]. T-Net then embeds the  
264 transmission feature, which is combined with the U-Net to  
265 predict the dehazed image.

266 **Transmission Network (T-Net).** We propose T-Net to ef-  
267 fectively handle the spatially varying input transmission  
268 map  $T$  in a high-dimensional feature space as shown in  
269 Fig. 2 (b). To achieve this, input  $T_{DCP}$  is projected into  
270 the feature space by a Projection Block comprising two  
271  $1 \times 1$  convolution layers, which generates the representa-  
272 tion in the feature space  $T^{l=0}$ . Subsequently, the Trans-  
273 mission Block (T-Block), which consists of two convolution  
274 layers and SiLU [17] activation functions, operates at multi-  
275 ple scales  $l \in \{0, 1, 2, 3\}$ . It takes  $T^l$  as input and generates  
276  $T^{l+1}$  at scale level  $(l+1)$  as:

$$277 T^{l+1} = Conv_{1 \times 1}(SiLU(Conv_{3 \times 3}(SiLU(T^l)))). \quad (9)$$

278 The embedding  $T^l$  is integrated with U-Net features to ef-  
279 fectively handle spatially varying transmissions. Each  $T^l$  is  
280 utilized in both the encoder and decoder of the U-Net.

281 **Transmission Refinement (T-Refinement).** The transmis-  
282 sion map  $T_{DCP}$  derived from DCP can be erroneous in  
283 real-world hazy conditions, especially in regions like sky  
284 where its underlying assumptions break down. To address  
285 this problem, we refine  $T_{DCP}$  in our HazeFlow by jointly  
286 predicting velocity and improved transmission map as fol-  
287 lows:  
288

$$(V, T_{refined}) = h_\theta(I_T, T_{dcp}), \quad (10)$$

289 where  $V$  denotes velocity and  $T_{refined}$  represents the im-  
290 proved transmission map. Specifically, we employ the  
291 ground-truth transmission map  $T_{gt}$  to train the network and  
292 obtain the refined transmission map by minimizing the loss  
293 as:

$$294 L_T(\theta) = \mathbb{E} \|T_{gt} - T_{refined}\|_2^2. \quad (11)$$

295 This approach ensures that the refinement aligns closely  
296 with the realistic haze characteristics defined by  $T_{gt}$ . During  
297 the inference phase, the refined transmission map is  
298 used as initial  $\tau$  in Eq. 8, resulting in more accurate and vi-  
299 sually coherent haze removal. Refer to the Supplementary  
300 Material Sec. S2 for additional details.

301 In the remainder of the study, when the output of  $h_\theta$  is  
302 not explicitly stated, it refers to  $V$  for simplicity.

### 303 3.4. Three-Stage Learning

304 We introduce a three-stage learning scheme including Pre-  
305 train, Reflow, and Distillation. Through this approach, we  
306 progressively adapt the model to real-world scenarios, en-  
307 abling real-world dehazing with fewer sampling steps.

308 **Pretrain** For the Pretrain phase, we first introduce a loss  
309 function  $L_{der}$  to train the neural network  $h_\theta$  on large-scale  
310 synthesized hazy-clean pairs, learning the derivative func-  
311 tion:

$$312 L_{der}(\theta) = \min_{\theta} \int_0^1 \mathbb{E} [\|(J - A) - h_\theta(I_T, T)\|^2] dT. \quad (12)$$

313 Moreover, to enhance perceptual quality of the dehazed im-  
314 age, we incorporate the LPIPS-based loss  $L_{perc}$  using a  
315 one-step estimation approach as:

$$316 L_{perc}(\theta) = \mathbb{E}[\mathbb{D}(J, I_T + (1 - T)h_\theta(I_T, T))]. \quad (13)$$

317 We observe that this accelerates training by yielding percep-  
318 tually satisfactory images [29]. Finally, by combining the  
319 derivative, perceptual, and transmission refinement losses,  
320 the total loss function for training  $h_\theta$  in the Pretrain stage is  
321 defined as:

$$322 L(\theta) = L_{der}(\theta) + L_{perc}(\theta) + w * L_T(\theta), \quad (14)$$

323 where  $w$  is a users-defined parameter to control the weight  
324 of the transmission refinement loss  $L_T(\theta)$ .

325 **Reflow.** The Reflow stage trains a new RF model  $h_\phi$  (shar-  
326 ing the same architecture as  $h_\theta$ ) to adapt to realistic haze  
327 distributions. In particular, we tackle two major challenges:  
328 (1) the lack of clean ground-truth counterparts correspond-  
329 ing to real-world hazy images, and (2) the complexity of ac-  
330 curately modeling the diverse distributions of atmospheric  
331 light  $A$  [9]. Specifically, when a real-world hazy image  
332  $I_T^{real}$  without corresponding ground truth counterpart is  
333 given, the pretrained  $h_\theta$  can generate a pseudo clean im-  
334 age  $\hat{J}$  and spatially-varying atmospheric light  $\hat{A}$  through

335 one step estimation as follows:

$$\hat{J} = I_T^{real} + (1 - T) \cdot h_\theta(I_T^{real}, T), \quad (15)$$

$$\hat{A} = I_T^{real} - T \cdot h_\theta(I_T^{real}, T). \quad (16)$$

338 While  $h_\theta$  from the Pretrain stage assumes a uniform atmospheric light  $A \in \mathbb{R}^3$  over the entire image, Reflow relaxes this assumption by estimating a spatially varying  $\hat{A}$  in  $\mathbb{R}^{H \times W \times 3}$  from real-world images. This allows the model to adapt to a much broader spectrum of real-world haze distributions.

344 Then the pseudo pair  $(\hat{J}, \hat{A})$  serves as a proxy for ground  
345 truth, enabling further training of  $h_\phi$  with the loss given by,  
346

$$L_{der}^{real}(\phi) = \min_{\phi} \int_0^1 \mathbb{E}[||(\hat{J} - \hat{A}) - h_\phi(\gamma(I_T^{real}), T)||^2] dT, \quad (17)$$

348 where  $\gamma$  represents random gamma correction [56] to predict appropriate brightness. To further enhance perceptual  
349 quality, we incorporate the LPIPS loss, yielding the final  
350 objective to train  $h_\phi$  as:  
351

$$L_{Reflow}(\phi) = L_{der}^{real}(\phi) + L_{perc}(\phi). \quad (18)$$

353 Notably, acquiring a ground truth transmission map for the  
354 real-world  $I_T^{real}$  is impractical, hence transmission refinement  
355 is not performed at this stage.

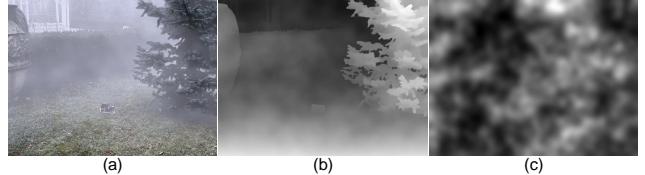
356 **Distillation.** In Distillation stage, we further improve the  
357 robustness of the model against real-world artifacts by distilling  
358  $h_\phi$  to a new model  $h'_\phi$  (with the same architecture as  
359  $h_\phi$ ). Specifically, we improve performance by suppressing  
360 unwanted degradations and enhancing output quality. To  
361 achieve this, the LPIPS loss is employed, as follows:

$$L_{Distill}(\phi') = \mathbb{E}[\mathbb{D}(I_T + (1 - T)h_\phi(I_T, T), \\ G(I_T) + (1 - T)h'_{\phi'}(G(I_T), T))], \quad (19)$$

363 where  $G$  represents real-world data augmentation, consisting  
364 of various degradations [18]. This augmentation forces  
365  $h'_\phi$  to learn how to remove various real-world artifacts, making  
366 the dehazing process more robust to unseen degradation  
367 patterns. Moreover, the Distillation can be iterated multiple  
368 times to progressively improve performance, leading to  
369 cleaner and more visually appealing results.

## 370 4. Non-Homogeneous Haze Generation

371 **Non-Homogeneous Haze.** In ASM, the thickness of the  
372 haze is determined by the value of the transmission map  
373  $T(x)$ , which depends on scene depth  $d(x)$  and the haze  
374 density coefficient  $\beta$ . Previous methods generate haze by sam-  
375 pling a scalar  $\beta$  for the entire image [31, 56], which fails to  
376 capture real-world non-homogeneous characteristics, such  
377 as smoke [2]. Vinay et al. [47] aimed to create realistic  
378 haze using Photoshop; however, producing various non-  
379 homogeneous hazy images demands substantial resources.



380 Figure 3. Example of non-homogeneous haze synthesized via  
381 MCBM. (a) Generated hazy image. (b) Transmission map  
382  $T_{MCBM}$ . (c) Spatially varying density coefficient map  $\beta$ .

383 To address this issue, we draw inspiration from the concept  
384 that air particle movement can be modeled using Brownian  
385 motion [27, 52] and utilize Markov Chain Brownian Motion  
386 (MCBM) to generate haze with non-homogeneous charac-  
387 teristics by spatially varying  $\beta$  and  $d(x)$ .

388 **Haze Synthesis using MCBM.** To simulate non-  
389 homogeneous real-world haze, we employ MCBM to  
390 generate a spatially varying haze density map  $\tilde{\beta}$ . First, a  
391 2D array is initialized to represent the haze density map on  
392 an image grid. We begin with a randomly selected pixel  
393 location  $(i, j)$  within the haze density map. This initial  
394 coordinate  $(i, j)$  is subsequently updated as follows:

$$(i', j') = (i + a, j + b), \quad (20)$$

395 where  $(a, b) \in \{(0, 1), (0, -1), (1, 0), (-1, 0)\}$  are selected  
396 randomly to update the coordinates. To enhance the non-  
397 uniformity of the haze density, we employ the Brownian  
398 motion. Then, the non-uniform haze density map  $\tilde{\beta}$  is up-  
399 dated accordingly as follows:

$$\tilde{\beta}(i' + \Delta i, j' + \Delta j) = \tilde{\beta}(i' + \Delta i, j' + \Delta j) + 1. \quad (21)$$

400 where  $\Delta i, \Delta j \sim \mathcal{N}(0, \sigma^2)$  are the random displace-  
401 ments for each pixel from Brownian motion, rounded to the  
402 nearest integer, with  $\sigma$  representing the standard deviation. In  
403 this way, we update the haze density map  $\tilde{\beta}$  progressively  
404 for  $n$  iterations where  $n$  denotes the number of updates in  
405 Eq. 20. This process models the random movement of haze  
406 particles, emphasizing their non-homogeneous characteris-  
407 tics. Finally, the haze density map,  $\tilde{\beta}$ , is smoothed with a  
408 Gaussian filter to produce a more realistic haze density map  
409 and then normalized to the range of 0 to 1. Additional de-  
410 tails about MCBM are provided in Supplementary Material  
411 Sec. S3.

412 **Haze Dataset Generation.** Based on MCBM, we gener-  
413 ate non-homogeneous transmission map by combining the  
414 uniform  $\beta$  and non-uniform  $\tilde{\beta}$  as:

$$T_{MCBM} = e^{-(\beta + \alpha \cdot \tilde{\beta})d}, \quad (22)$$

415 where the random parameter  $\alpha \in [0.5, 1]$  controls the de-  
416 gree of non-homogeneity. Subsequently, we generate non-  
417 homogeneous haze image using ASM:

$$I = \mathcal{D}(T_{MCBM} \cdot J + (1 - T_{MCBM}) \cdot A), \quad (23)$$

Method	RTTS [31]				Fattal's [19]			
	FADE↓	BRISQUE↓	NIMA↑	PAQ2PIQ↑	FADE↓	BRISQUE↓	NIMA↑	PAQ2PIQ↑
Hazy input	2.484	36.64	4.48	66.05	1.061	21.08	5.38	71.54
PDN [57]	0.876	30.81	4.46	-	-	-	-	-
MBSDN [16]	1.363	27.67	4.53	66.85	0.579	14.15	5.43	72.84
Dehamer [21]	1.895	33.24	4.52	66.70	0.698	15.53	5.16	71.43
DAD [46]	1.130	32.24	4.31	66.79	0.484	29.64	<b>5.46</b>	71.56
PSD [12]	0.920	27.71	4.60	70.43	0.416	23.61	4.99	76.02
D4 [58]	1.358	33.21	4.48	66.84	0.411	20.33	5.44	73.13
RIDCP [56]	0.944	17.29	4.97	70.82	0.408	20.05	5.43	74.64
CORUN [18]	0.824	11.96	<b>5.34</b>	72.56	0.338	14.82	5.39	76.12
HazeFlow	<b>0.583</b>	<b>5.01</b>	5.30	<b>72.97</b>	<b>0.264</b>	<b>13.36</b>	5.40	<b>76.44</b>

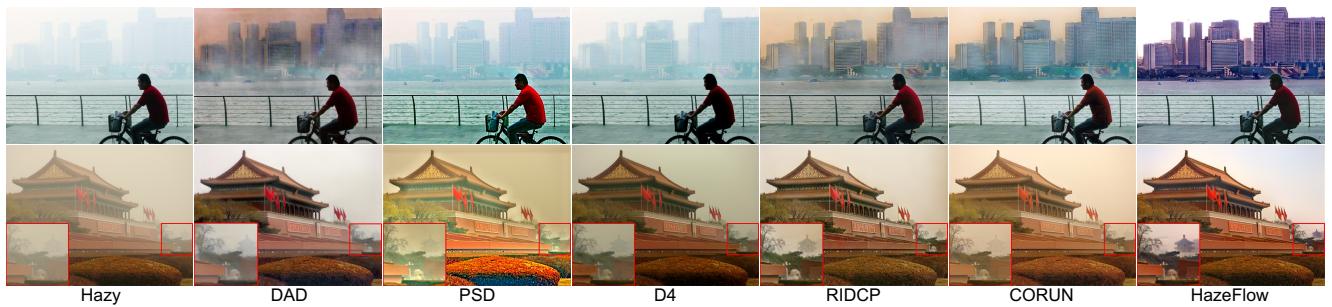
Table 1. Quantitative results on unpaired dataset (RTTS [31], Fattal [19]). Best results are **bolded**.

Figure 4. Qualitative unpaired dataset comparison on RTTS [31] (top row) and Fattal [19] (bottom row).

where  $\mathcal{D}$  represents the real-world degradation process, such as Gamma correction, additive Gaussian noise, and JPEG compression, following [56].

In practice, to generate a large number of non-homogeneous haze images, we use RIDCP500 [56] dataset for the clean image  $J$ , and RA-depth [23] is used to estimate the corresponding scene depth  $d$ . The global atmospheric light  $A$  is uniformly sampled in the range  $[0.25, 1.8]$  accounting for the color bias of atmosphere light, and  $\beta$  is uniformly sampled from the range  $[0.2, 2.8]$ . In Fig. 3, (a) shows an example of our synthetic hazy image, (b) illustrates the  $T_{MCBM}$ , and (c) represents the spatially varying  $\tilde{\beta}$  generated by MCBM. This process produces realistic non-homogeneous haze effects, and the resulting dataset is used during the Pretrain stage of HazeFlow.

## 5. Experiments

Our source code will be available upon acceptance.

### 5.1. Experimental Setting

**Datasets.** In the Pretrain stage, we use a synthetic non-homogeneous dataset generated by our MCBM method with the RIDCP500 [56], consisting of 500 clear images paired with depth maps estimated using [23]. In experiments utilizing unpaired datasets ( $I_T^{real}$  without the corresponding ground truth  $J$ ), we use the URHI dataset [31], which comprises 4,807 real hazy images, for Reflow and Distillation stages. In experiments utilizing paired datasets

of clean and hazy images, we use paired training datasets ( $I_T^{real}$  with the corresponding ground truth  $J$ ) in the Reflow and Distillation phases to maintain a fair comparison with alternative approaches. Additional experiments in which the model is trained on the URHI dataset instead of the training data from the paired dataset can be found in the Supplementary Material Sec. S8.

To evaluate real-world dehazing performance, we use unpaired real-world haze datasets, including RTTS [31], which consists of 4,322 hazy images, and Fattal's dataset [19] containing 31 classic hazy images. Moreover, for paired real-world haze datasets, we use the NH-HAZE [4] and Dense-Haze [3] datasets, both of which contain 50 training data pairs and 5 test data pairs.

**Implementation Details.** The baseline U-Net architecture of HazeFlow follows NCSN++ [49]. In the Pretrain stage, we train HazeFlow for 100K iterations using hazy images generated by our MCBM from the RIDCP500 dataset. We use Adam [28] optimizer and Cosine annealing [37] with an initial learning rate of  $2 \times 10^{-4}$  to  $1 \times 10^{-6}$  for optimization. The weight of the transmission refinement loss is set to  $w = 0.5$ . To focus on learning the derivative function of the ODE, we fix  $A$  to  $[1, 1, 1]$ .

Next, in the Reflow and Distillation phases, the network is trained for only 10K iterations with Adam [28] optimizer.

For inference, we perform single-step generation with  $N = 1$ .

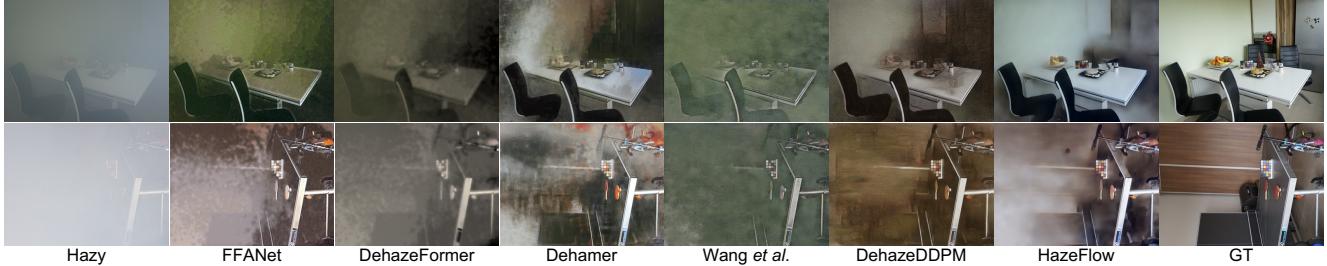


Figure 5. Qualitative paired dataset comparison on Dense-HAZE [3].

**Evaluation Metrics.** For evaluation on unpaired datasets, we use the Fog-Aware Density Evaluator (FADE) [13] to measure the haze density. Furthermore, to evaluate the natural appearance and ensure our results lack any artifacts, we utilize BRISQUE [40]. Meanwhile, NIMA [51] is employed for a thorough assessment of both image quality and aesthetics. Additionally, we incorporate PAQ2PIQ [59] for a further perceptual quality comparison. For evaluation on paired datasets, we measure the Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM) and Learned perceptual image patch similarity (LPIPS) in the RGB color space.

## 5.2. Comparative Evaluation

**Quantitative Results.** First, for evaluation on the unpaired real-world RTTS and Fattal’s datasets, we compare our method with the SOTA approaches, including PDN [57], MSBDN [16] and Dehamer [21], and real-world dehazing methods such as DAD [46], PSD [12], D4 [58], RIDCP [56], and CORUN [18]. In Tab. 1, our approach surpasses other methods in dehazing performance across the majority of evaluation metrics for both datasets. In particular, on RTTS, our method surpasses the second-leading approach by a significant margin, improving FADE by 0.241 and BRISQUE by 6.95. On Fattal’s dataset, our HazeFlow outperforms the next leading method by a margin of 0.074 in FADE score, while also obtaining best results in BRISQUE and PAQ2PIQ.

Next, for evaluation on the paired real-world Dense-HAZE and NH-HAZE datasets, we compare our method against SOTA approaches, including the prior-based method DCP [22], methods trained on the RESIDE [31] dataset (AOD-Net [30], MSBDN [16], GCANet [15], FFANet [43], DehazeFormer [50], Dehamer [21]), and diffusion-based methods (Wang *et al.* [54], DehazeDDPM [60]). For DehazeDDPM, we use the official pre-trained weights for evaluation, while for Wang *et al.*, we fine-tune the model on Dense-HAZE and NH-HAZE using the same number of iterations as HazeFlow to ensure a fair comparison. In Tab. 2, HazeFlow outperforms all competing methods in Dense-HAZE, achieving the best scores across all metrics. Specifically, HazeFlow outperforms the second-best method by a substantial margin, with a 0.0556 improvement in SSIM and 0.1530 in LPIPS. On

Method	Dense-HAZE [3]			NH-HAZE [4]		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
DCP [22]	11.01	0.4165	0.8405	12.72	0.4419	0.5168
AOD-Net [30]	12.88	0.5045	0.7612	15.31	0.4584	0.5121
MSBDN [16]	13.44	0.4301	0.9501	17.34	0.5566	0.5026
GCANet [15]	12.46	0.4774	0.7313	16.64	0.5583	0.4356
FFANet [43]	15.18	0.5757	0.6072	18.48	0.6186	0.3694
DehazeFormer [50]	14.95	0.4979	0.7770	18.15	0.6070	0.4192
Dehamer [21]	16.62	<u>0.5602</u>	0.6543	<b>20.66</b>	0.6844	0.3122
Wang <i>et al.</i> [54]	14.83	0.4064	0.7508	16.11	0.5591	0.4349
DehazeDDPM [60]	15.45	0.4559	<u>0.5808</u>	19.44	0.6278	<u>0.2984</u>
HazeFlow	<b>17.14</b>	<b>0.6158</b>	<b>0.4278</b>	<u>20.06</u>	<b>0.6847</b>	<u>0.2606</u>

Table 2. Quantitative results on paired real-world datasets: Dense-HAZE [3] and NH-HAZE [4]. Best results are **bolded** and second best results are underlined. The last three rows correspond to generative model-based approaches.

the NH-HAZE dataset, HazeFlow achieves the best perceptual quality, ranking first in LPIPS and SSIM while securing the second-highest PSNR. Notably, our method produces perceptually superior outcomes relative to SOTA diffusion-based methods (DehazeDDPM), while delivering competitive objective performance against the regression-based SOTA approach Dehamer. These findings demonstrate the effectiveness of our approach in addressing complex real-world haze, efficiently achieving a balance between restoration quality and structural integrity.

**Qualitative Results.** We provide qualitative comparisons on both unpaired (Fig. 4) and paired datasets (Fig. 5). As shown in Fig. 4, our method excels in restoring background regions with significant depth. The first row shows no residual haze, while the second row highlights effective haze removal from distant objects, as marked. Similarly, Fig. 5 shows that HazeFlow achieves more precise haze removal and color prediction results, closely matching ground truth images. Notably, our model accurately captures the spatial variant in haze and dynamically adjusts its dehazing strength, leading to sharper, artifact-free results across varying haze densities. More results can be found in the Supplementary Material Sec. S9.

## 5.3. Comparison with Diffusion-based Methods

To demonstrate the superior efficiency and effectiveness of our generative approach, we compare it with recent diffusion-based methods [39, 54], including all-in-one models such as DiffUIR [63] and DACLIP [38], which are

Method	DehazeFormer [50]				NAFNet [10]				FocalNet [14]			
	FADE↓	BRISQUE↓	NIMA↑	PAQ2PIQ↑	FADE↓	BRISQUE↓	NIMA↑	PAQ2PIQ↑	FADE↓	BRISQUE↓	NIMA↑	PAQ2PIQ↑
DAD* [46]	6.16	45.45	4.93	64.74	2.82	36.88	4.80	66.52	2.69	33.68	4.82	66.77
D4* [58]	1.78	29.60	4.85	66.71	1.35	28.64	4.81	67.07	1.16	28.78	4.86	67.21
RIDCP [56]	1.44	28.43	5.17	67.96	1.17	<b>21.01</b>	<b>5.00</b>	67.48	1.36	<b>28.07</b>	4.99	65.47
MCBM (Ours)	<b>1.05</b>	<b>23.22</b>	<b>5.25</b>	<b>68.40</b>	<b>1.03</b>	23.68	4.94	<b>67.63</b>	<b>0.80</b>	29.90	<b>5.06</b>	<b>68.44</b>

Table 3. Comparison of different haze augmentation methods applied to baseline models. Methods marked with \* (DAD and D4) require training, while RIDCP and MCBM (Ours) are training-free approaches. Best results are highlighted in **bold**.

trained on large-scale datasets incorporating various types of degradation. Additionally, we compare against Refusion [39], which is trained on non-homogeneous real pairs [5], as well as the dehazing model by Wang *et al.* [54], which is trained on the RESIDE dataset. As shown in Tab. 4, HazeFlow achieves significantly better performance with substantially fewer steps, leading to much faster runtime. This highlights that solving the ODE of ASM with a flow-based approach enhances both speed and restoration quality.

#### 5.4. MCBM Performance Across Different Models

To illustrate the broad applicability of MCBM, we assess its performance with various baselines, such as NAFNet [10], DehazeFormer [50], and FocalNet [14], as shown in Tab. 3. The results show that MCBM consistently achieves the best performance in FADE and PAQ2PIQ across all baselines, confirming its effectiveness in training various models. BRISQUE show best performance with DehazeFormer, and NIMA show better performance with FocalNet. Notably, our MCBM is the only approach that is both training-free and capable of generating non-homogeneous haze, unlike methods such as DAD [46], D4 [58], and RIDCP [56]. This indicates that MCBM haze generation process closely aligns with real-world haze characteristics, making it adaptable to various models.

#### 5.5. Ablations

To evaluate HazeFlow, we conduct experiments on the three-stage learning framework, MCBM and T-Refinement. All ablation studies are performed on the RTTS dataset [31].

**Effectiveness of Three-Stage Learning.** Tab. 5 presents an ablation study on Reflow and Distillation, highlighting their effectiveness and showing significant performance improvements. Specifically, Reflow achieves substantial gains in BRISQUE and PAQ2PIQ, demonstrating its ability to adapt to real-world datasets and accurately learn complex haze distributions. Finally, Distillation further enhances performance for real-world degradation in terms of BRISQUE, PAQ2PIQ, and NIMA. Supplementary Material Sec. S1.2 provide additional visual results of comparison between Reflow and Distillation.

**Effectiveness of MCBM and T-Refinement.** Tab. 6 shows that HazeFlow trained with our MCBM outperforms models trained with previous haze synthesis approach [56], improv-

Method	FADE↓	BRISQUE↓	NIMA↑	PAQ2PIQ↑	Run time (s)	# of steps
DACLIP [38]	1.922	34.78	4.98	66.87	38.09	100
Wang <i>et al.</i> [54]	0.929	27.66	5.00	67.51	2.50	20
DiffUIR [63]	2.019	34.88	4.54	66.81	0.30	3
Refusion [39]	2.047	42.59	4.49	64.67	2.76	100
HazeFlow	<b>0.583</b>	<b>5.01</b>	<b>5.30</b>	<b>72.97</b>	<b>0.21</b>	<b>1</b>

Table 4. Comparison with other diffusion-based models.

Experiment	FADE↓	BRISQUE↓	NIMA↑	PAQ2PIQ↑
Pretrained	$h_\theta$	0.561	18.96	5.26
+ Reflow	$h_\phi$	0.598	8.14	5.26
+ Distillation	$h_{\phi'}$	0.583	5.01	5.30

Table 5. Ablation study on 3-stage framework.

MCBM	T-Refinement	FADE↓	BRISQUE↓	NIMA↑	PAQ2PIQ↑
-	-	0.950	14.63	5.31	70.78
✓	-	0.817	12.67	5.35	71.38
✓	✓	0.583	5.01	5.30	72.97

Table 6. Ablation study for MCBM and Transmission refinement.

ing the FADE score by approximately 0.133. This demonstrate that MCBM significantly enhances real-world haze removal. Additionally, T-Refinement corrects inaccuracies in the transmission map predicted by DCP, enabling HazeFlow to estimate transmission more accurately. This refinement improves the FADE score by 0.234 and further enhances haze removal. Adding T-Refinement also enhances the recovery of color and brightness restoration during inference. Visual results for T-Refinement are provided in the Supplementary Material Sec. S2.

## 6. Conclusion

In this work, we present HazeFlow, a novel framework that unifies physics-guided modeling and neural ODEs to redefine single image dehazing. By reformulating the ASM as an ODE, we transform dehazing into a trajectory optimization problem, where a neural network learns a straight ODE path from atmospheric light to the clean image, guided by the transmission map. This design not only enhances dehazing accuracy but also enables fast inference, achieving high-quality results even with a single step while drastically reducing computational costs compared to iterative methods. Furthermore, by modeling non-homogeneous haze synthesis with MCBM, we enhance our framework’s capacity to handle complex and diverse real-world hazy scenes. This comprehensive approach aims to bridge the gap between synthetic and real-world data, improving generalization and robustness while accelerating inference and advancing dehazing performance.

613 **References**

- [1] Cosmin Ancuti, Codruta O Ancuti, Radu Timofte, and Christophe De Vleeschouwer. I-haze: A dehazing benchmark with real hazy and haze-free indoor images. In *ACIVS*. Springer, 2018. 4
- [2] Codruta O Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In *CVPRW*, 2018. 5, 4
- [3] Codruta O Ancuti, Cosmin Ancuti, Mateu Sbert, and Radu Timofte. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In *ICIP*, pages 1014–1018. IEEE, 2019. 6, 7, 4, 8
- [4] Codruta O Ancuti, Cosmin Ancuti, and Radu Timofte. Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In *CVPRW*, 2020. 6, 7, 2, 4
- [5] Codruta O Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluiu, and Radu Timofte. Ntire 2023 challenge on non-homogeneous dehazing. In *CVPRW*, 2023. 8
- [6] Dana Berman, Shai Avidan, et al. Non-local image dehazing. In *CVPR*, 2016. 3
- [7] Dana Berman, Tali Treibitz, and Shai Avidan. Single image dehazing using haze-lines. *IEEE TPAMI*, 42(3):720–734, 2018. 2
- [8] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE TIP*, 25(11):5187–5198, 2016. 1, 2
- [9] Chih-Ling Chang, Fu-Jen Tsai, Zi-Ling Huang, Lin Gu, and Chia-Wen Lin. Panet: A physics-guided parametric augmentation net for image dehazing by hazing. *arXiv preprint arXiv:2404.09269*, 2024. 1, 2, 4
- [10] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *ECCV*, 2022. 8
- [11] Xiang Chen, Zhentao Fan, Pengpeng Li, Longgang Dai, Caihua Kong, Zhuoran Zheng, Yufeng Huang, and Yufeng Li. Unpaired deep image dehazing using contrastive disentanglement learning. In *ECCV*, 2022. 1, 2
- [12] Zeyuan Chen, Yangchao Wang, Yang Yang, and Dong Liu. Psd: Principled synthetic-to-real dehazing guided by physical priors. In *CVPR*, 2021. 6, 7, 2, 4
- [13] Lark Kwon Choi, Jaehee You, and Alan Conrad Bovik. Referenceless prediction of perceptual fog density and perceptual image defogging. *IEEE TIP*, 24(11):3888–3901, 2015. 7, 2
- [14] Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Focal network for image restoration. In *ICCV*, 2023. 8
- [15] Sowmen Das, Md Saiful Islam, and Md Ruhul Amin. Gca-net: utilizing gated context attention for improving image forgery localization and detection. In *CVPR*, 2022. 7
- [16] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In *CVPR*, 2020. 2, 6, 7
- [17] Stefan Elfwing, Eiji Uchibe, and Kenji Doya. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural networks*, 107:3–11, 2018. 4
- [18] Chengyu Fang, Chunming He, Fengyang Xiao, Yulun Zhang, Longxiang Tang, Yuelin Zhang, Kai Li, and Xiu Li. Real-world image dehazing with coherence-based label generator and cooperative unfolding network. *arXiv preprint arXiv:2406.07966*, 2024. 5, 6, 7, 2, 4
- [19] Raanan Fattal. Single image dehazing. *ACM TOG*, 27(3):1–9, 2008. 6, 3, 7
- [20] Raanan Fattal. Dehazing using color-lines. *ACM TOG*, 34(1):1–14, 2014. 1, 2
- [21] Chun-Le Guo, Qixin Yan, Saeed Anwar, Runmin Cong, Wenqi Ren, and Chongyi Li. Image dehazing transformer with transmission-aware 3d position embedding. In *CVPR*, 2022. 1, 2, 6, 7
- [22] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE TPAMI*, 33(12):2341–2353, 2010. 1, 2, 4, 7, 3
- [23] Mu He, Le Hui, Yikai Bian, Jian Ren, Jin Xie, and Jian Yang. Ra-depth: Resolution adaptive self-supervised monocular depth estimation. In *ECCV*, 2022. 6
- [24] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 2020. 2
- [25] Shih-Chia Huang, Trung-Hieu Le, and Da-Wei Jaw. Dsnet: Joint semantic learning for object detection in inclement weather conditions. *IEEE TPAMI*, 43(8):2623–2633, 2020. 1
- [26] Eun Woo Im, Junsung Shin, Sungyong Baik, and Tae Hyun Kim. Deep variational bayesian modeling of haze degradation process. In *CIKM*, 2023. 1
- [27] Kiyosi Itô, P Henry Jr, et al. *Diffusion processes and their sample paths*. Springer Science & Business Media, 2012. 2, 5
- [28] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6, 1
- [29] Sangyun Lee, Zinan Lin, and Giulia Fanti. Improving the training of rectified flows. *NeurIPS*, 2024. 4
- [30] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *ICCV*, 2017. 1, 2, 7
- [31] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE TIP*, 28(1):492–505, 2018. 1, 2, 5, 6, 7, 8, 3, 4
- [32] Jinlong Li, Runsheng Xu, Jin Ma, Qin Zou, Jiaqi Ma, and Hongkai Yu. Domain adaptive object detection for autonomous driving under foggy weather. In *WACV*, 2023. 1
- [33] Yi Li, Yi Chang, Yan Gao, Changfeng Yu, and Luxin Yan. Physically disentangled intra-and inter-domain adaptation for varicolored haze removal. In *CVPR*, 2022. 2
- [34] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Grid-dehazenet: Attention-based multi-scale network for image dehazing. In *ICCV*, 2019. 2

- 724 [35] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow  
725 straight and fast: Learning to generate and transfer data with  
726 rectified flow. *arXiv preprint arXiv:2209.03003*, 2022. 2, 3,  
727 6
- 728 [36] Xingchao Liu, Xiwen Zhang, Jianzhu Ma, Jian Peng, et al.  
729 Instaflow: One step is enough for high-quality diffusion-  
730 based text-to-image generation. In *ICLR*, 2023. 3, 4
- 731 [37] Ilya Loshchilov and Frank Hutter. Sgdr: Stochas-  
732 tic gradient descent with warm restarts. *arXiv preprint*  
733 *arXiv:1608.03983*, 2016. 6
- 734 [38] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens  
735 Sjölund, and Thomas B Schön. Controlling vision-language  
736 models for multi-task image restoration. *arXiv preprint*  
737 *arXiv:2310.01018*, 2023. 1, 3, 7, 8
- 738 [39] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund,  
739 and Thomas B Schön. Refusion: Enabling large-size realis-  
740 tic image restoration with latent-space diffusion models. In  
741 *CVPR*, 2023. 7, 8
- 742 [40] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad  
743 Bovik. No-reference image quality assessment in the spa-  
744 tial domain. *IEEE TIP*, 21(12):4695–4708, 2012. 7, 2
- 745 [41] Srinivasa G Narasimhan and Shree K Nayar. Vision and the  
746 atmosphere. *IJCV*, 48:233–254, 2002. 1
- 747 [42] Srinivasa G. Narasimhan and Shree K. Nayar. Contrast  
748 restoration of weather degraded images. *IEEE TPAMI*, 25  
749 (6):713–724, 2003. 1
- 750 [43] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and  
751 Huizhu Jia. Ffa-net: Feature fusion attention network for  
752 single image dehazing. In *AAAI*, 2020. 1, 2, 7
- 753 [44] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao,  
754 and Ming-Hsuan Yang. Single image dehazing via multi-  
755 scale convolutional neural networks. In *ECCV*, 2016. 2
- 756 [45] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Se-  
757 mantic foggy scene understanding with synthetic data. *IJCV*,  
758 126:973–992, 2018. 2
- 759 [46] Yuanjie Shao, Lerenhan Li, Wenqi Ren, Changxin Gao, and  
760 Nong Sang. Domain adaptation for image dehazing. In  
761 *CVPR*, 2020. 2, 6, 7, 8, 4
- 762 [47] Lithesh Shetty et al. Non homogeneous realistic single image  
763 dehazing. 2023. 5
- 764 [48] Karen Simonyan and Andrew Zisserman. Very deep convo-  
765 lutional networks for large-scale image recognition. *arXiv*  
766 *preprint arXiv:1409.1556*, 2014. 2
- 767 [49] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Ab-  
768 hishek Kumar, Stefano Ermon, and Ben Poole. Score-based  
769 generative modeling through stochastic differential equa-  
770 tions. In *ICLR*, 2021. 6
- 771 [50] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision  
772 transformers for single image dehazing. *IEEE TIP*, 32:1927–  
773 1941, 2023. 2, 7, 8
- 774 [51] Hossein Talebi and Peyman Milanfar. Nima: Neural image  
775 assessment. *IEEE TIP*, 27(8):3998–4011, 2018. 7
- 776 [52] Nicolaas Godfried Van Kampen. *Stochastic processes in*  
777 *physics and chemistry*. Elsevier, 1992. 2, 5
- 778 [53] Chaoyang Wang, Xiangtai Li, Lu Qi, Henghui Ding, Yunhai  
779 Tong, and Ming-Hsuan Yang. Semflow: Binding semantic  
780 segmentation and image synthesis via rectified flow. *arXiv*  
781 *preprint arXiv:2405.20282*, 2024. 3, 4, 2
- [54] Jing Wang, Songtao Wu, Zhiqiang Yuan, Qiang Tong, and  
Kuanhong Xu. Frequency compensated diffusion model for  
real-scene dehazing. *Neural Networks*, 175:106281, 2024. 2,  
7, 8 782
- [55] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi  
Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Con-  
trastive learning for compact single image dehazing. In  
*CVPR*, 2021. 2 783
- [56] Rui-Qi Wu, Zheng-Peng Duan, Chun-Le Guo, Zhi Chai, and  
Chongyi Li. Ridcp: Revitalizing real image dehazing via  
high-quality codebook priors. In *CVPR*, 2023. 2, 5, 6, 7, 8,  
4 784
- [57] Dong Yang and Jian Sun. Proximal dehaze-net: A prior  
learning-based deep network for single image dehazing. In  
*ECCV*, 2018. 6, 7 785
- [58] Yang Yang, Chaoyue Wang, Risheng Liu, Lin Zhang, Xiao-  
jie Guo, and Dacheng Tao. Self-augmented unpaired image  
dehazing via density and depth decomposition. In *CVPR*,  
2022. 2, 6, 7, 8, 4 786
- [59] Zhenqiang Ying, Haoran Niu, Praful Gupta, Dhruv Mahajan,  
Deepti Ghadiyaram, and Alan Bovik. From patches to pic-  
tures (paq-2-piq): Mapping the perceptual space of picture  
quality. In *CVPR*, 2020. 7, 2 787
- [60] Hu Yu, Jie Huang, Kaiwen Zheng, and Feng Zhao. High-  
quality image dehazing with diffusion model. *arXiv preprint*  
*arXiv:2308.11949*, 2023. 2, 7 788
- [61] He Zhang and Vishal M Patel. Densely connected pyramid  
dehazing network. In *CVPR*, 2018. 3 789
- [62] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman,  
and Oliver Wang. The unreasonable effectiveness of deep  
features as a perceptual metric. In *CVPR*, 2018. 3 790
- [63] Dian Zheng, Xiao-Ming Wu, Shuzhou Yang, Jian Zhang,  
Jian-Fang Hu, and Wei-Shi Zheng. Selective hourglass  
mapping for universal image restoration based on diffusion  
model. In *CVPR*, 2024. 1, 3, 7, 8 791
- [64] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A  
Efros. Unpaired image-to-image translation using cycle-  
consistent adversarial networks. In *ICCV*, 2017. 1, 2 792
- [65] Qingsong Zhu, Jiaming Mai, and Ling Shao. Single im-  
age dehazing using color attenuation prior. In *BMVC*, pages  
1674–1682. Citeseer, 2014. 1, 2 793
- [66] Yixuan Zhu, Wenliang Zhao, Ao Li, Yansong Tang, Jie  
Zhou, and Jiwen Lu. Flowie: Efficient image enhancement  
via rectified flow. In *CVPR*, 2024. 3, 4, 2 794

# 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.

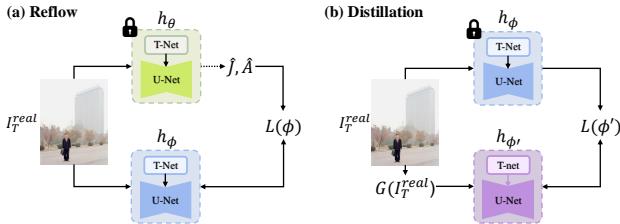


Figure 6. Illustration of (a) Reflow and (b) Distillation in our 3-stage learning framework.

#### 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  $1 \times 10^{-6}$  for Reflow and  $5 \times 10^{-5}$  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:

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

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

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

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:

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

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

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

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

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

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  $\sigma$  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  $\sigma$  is randomly selected from [15, 25, 35]. Finally, the MCBM haze density is generated through normalization.

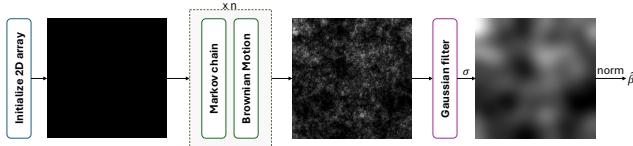


Figure 7. Overall pipeline of MCBM.

892

### S3.2. Additional Visual Results

893

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.

901

### S3.3. t-SNE Visualization of MCBM Haze

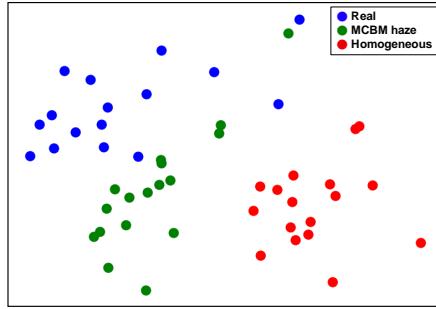


Figure 8. Visualization of feature distributions of pretrained network [48] using t-SNE, comparing real hazy images (blue) obtained from NH-HAZE [4], synthesized non-homogeneous hazy images (green) generated by our MCBM haze density, and synthesized homogeneous hazy images (red).

902

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.

911

## 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	<b>0.6267</b>

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

912

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.

913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925

## S5. Discussion on Estimation Steps

926

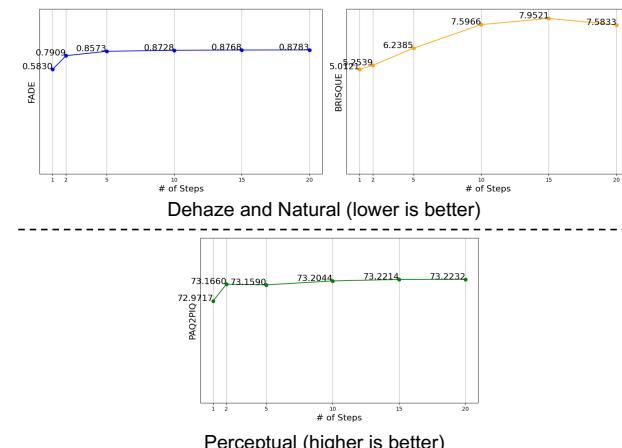


Figure 9. Variation in metrics based on the number of estimation steps.

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.

927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944



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

945

## 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, Re-flow 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	<b>0.583</b>	<b>5.01</b>	<b>72.97</b>	<b>63.94</b>

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

955

956

## 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.

968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996

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	<b>13.51</b>	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	<b>14.49</b>	<b>0.61</b>	11.39	<b>0.56</b>	<b>18.37</b>	<b>0.83</b>	<b>18.79</b>	<b>0.84</b>

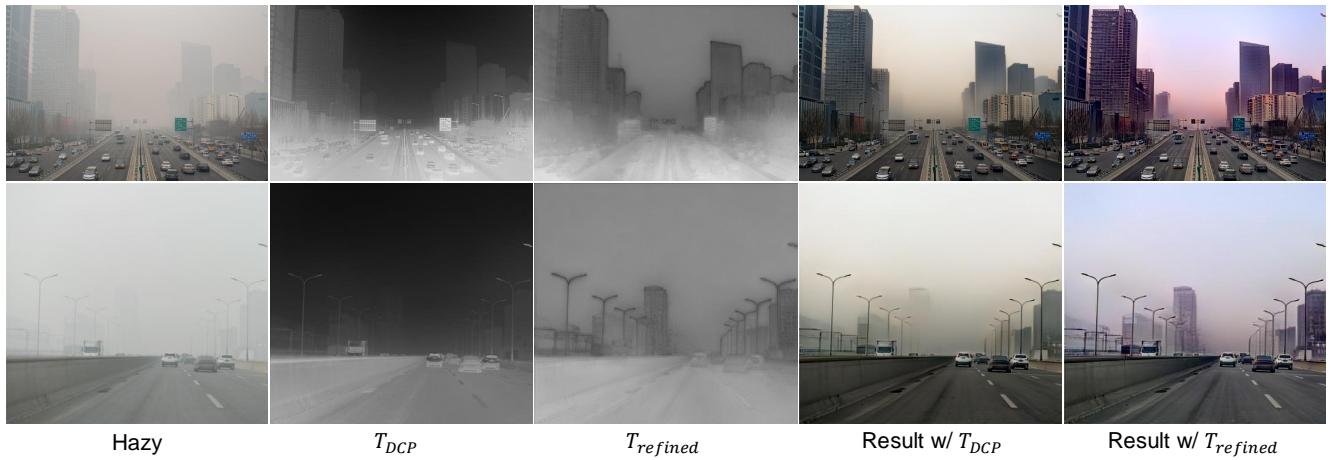
Table 10. Quantitative results on paired dataset (NH-HAZE [4], Dense-HAZE [3], I-HAZE [1], O-HAZE [2]). Best results are **bolded**.Figure 11. Visual comparison between results with  $T_{DCP}$  and with  $T_{refined}$  on the RTTS dataset [31].

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

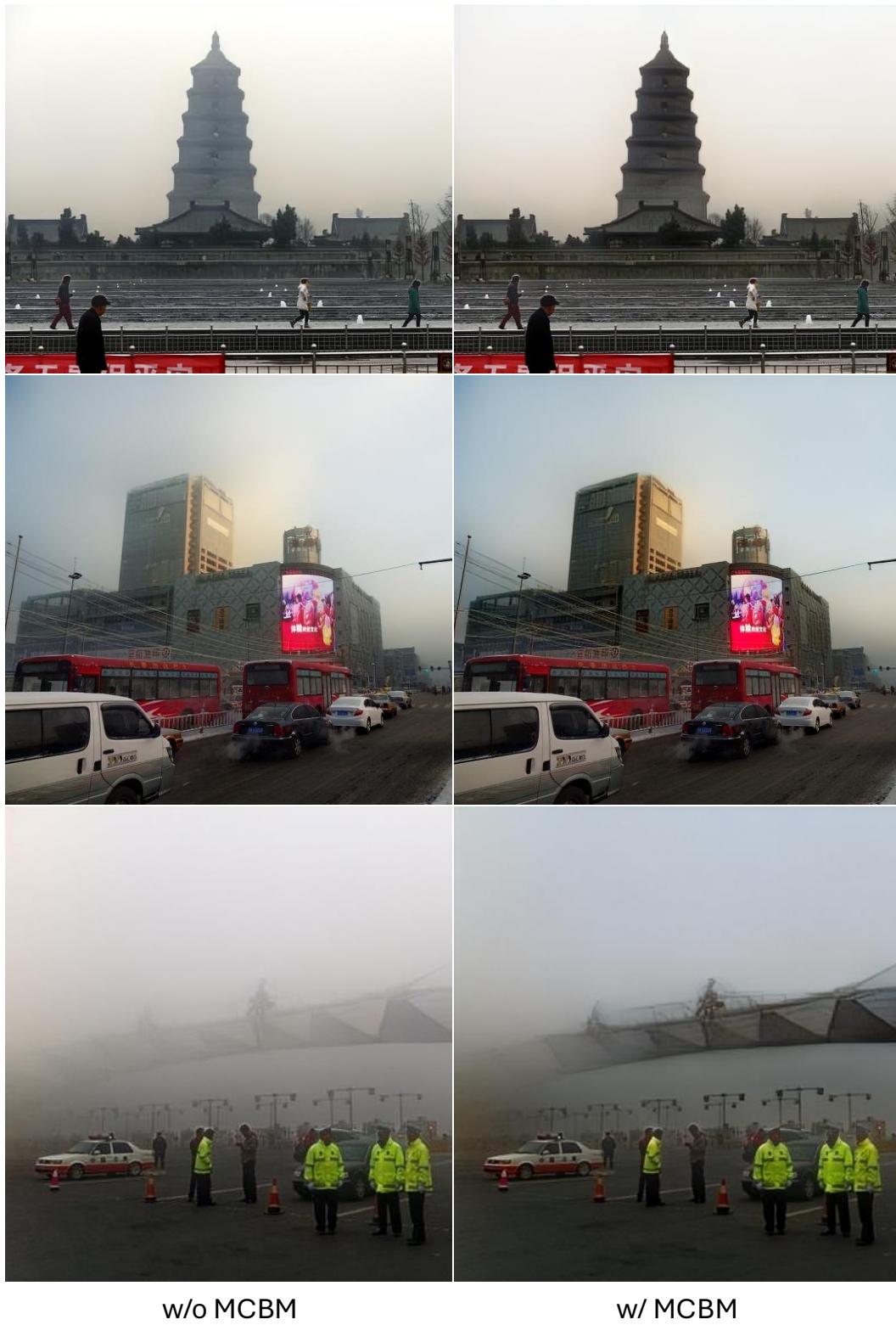


Figure 13. Visual comparison of the results from networks trained with and without MCBM haze on the RTTS dataset [31].



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



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

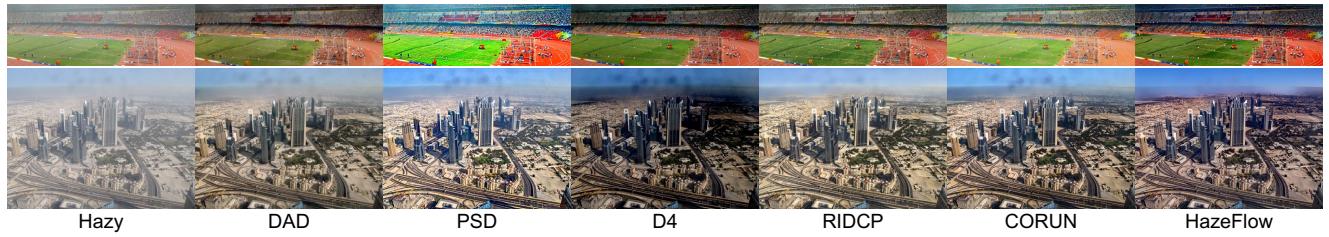


Figure 16. Additional visual comparison on Fattal's dataset [19]



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

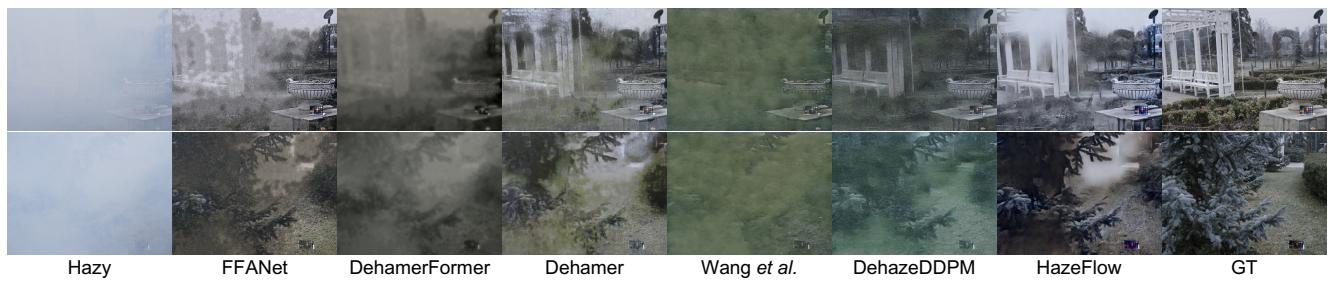


Figure 18. Additional visual comparison on Dense-HAZE dataset [3].