PromptHaze: Prompting Real-world Dehazing via Depth Anything Model

Tian Ye[1], Sixiang Chen[1], Haoyu Chen[1], Wenhao Chai[2], Jingjing Ren[1], Zhaohu Xing[1], Wenxue Li[1], Lei Zhu[1,3*]

[1]The Hong Kong University of Science and Technology (Guangzhou)
[2]University of Washington
[3]The Hong Kong University of Science and Technology
{tye610, schen691, hchen794, jren044, zxing565}@connect.hkust-gz.edu.cn,
wchai@uw.edu, wxli408@gmail.com,leizhu@ust.hk

Abstract

Real-world image dehazing remains a challenging task due to the diverse nature of haze degradation and the lack of large-scale paired datasets. Existing methods based on hand-crafted priors or generative priors struggle to recover accurate backgrounds and fine details from dense haze regions. In this work, we propose a novel paradigm, PromptHaze, for real-world image dehazing via the depth prompt from the Depth Anything model. By employing a prompt-by-prompt strategy, our method iteratively updates the depth prompt and progressively restores the background through a dehazing network with controllable dehazing strength. Extensive experiments on widely-used real-world dehazing benchmarks demonstrate the superiority of PromptHaze in recovering authentic backgrounds and fine details from various haze scenes, outperforming state-of-the-art methods across multiple quality metrics.

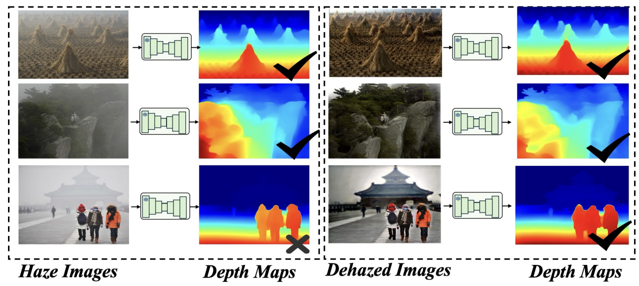

Figure 1: We feed both real hazy images and their corresponding dehazed images into the Depth Anything Model, and could observe that for more common haze scenarios, the Depth Anything Model can provide fairly accurate scene depth information. However, when faced with challenging haze scenarios (the bottom sample), it struggles to provide precise background depth information. The dehazed images are generated by our PromptHaze.

Introduction

The objective of image dehazing (Cai et al. 2016; Liu et al. 2019; Qin et al. 2020) is to eliminate haze in images and restore the fine details of a background accurately, which is essential for numerous advanced computer vision tasks (Xie et al. 2021; Zhang et al. 2020). The formation of a hazy image can be represented by a physical atmospheric scattering model:

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

where $I(x)$ signifies the hazy image, and $J(x)$ represents the corresponding clear image. Here, A refers to the global atmospheric light value, and $t(x)$ is the transmission map. The equation for the transmission map, $t(x) = e^{\beta d(x)}$, is influenced by the scene's depth $d(x)$ and the haze density coefficient β.

Image dehazing, as a classic ill-posed problem, presents multiple possible solutions $J(x)$ for a hazy image $I(x)$. There have been lots of approaches based on priors, such as the dark channel prior (He, Sun, and Tang 2010), the color attenuation prior (Katiyar and Verma 2016) and Non-local prior (Berman, Avidan et al. 2016), attempting to recover a clean, ideal $J(x)$. However, conventional prior-based methods (He, Sun, and Tang 2010; Berman, Avidan et al. 2016) relying on hand-craft statistical priors often struggle to accommodate the wide variety of degradation scenarios encountered in the real world, resulting in subpar outcomes in certain situations.

With the advancement of deep learning technologies, a variety of powerful dehazing networks have emerged, such as AECR-Net (Wu et al. 2021), FFA-Net (Qin et al. 2020), PMNet (Ye et al. 2021), and Dehamer (Chun-Le Guo 2022). However, these methods often achieve high PSNR/SSIM metrics on synthetic dehazing benchmarks but do not generalize well to real-world hazy images. *This discrepancy is due to the diverse nature of real-world haze degradation, which the limited scale of synthetic datasets struggles to fully represent.* Additionally, *the high cost of collecting authentic paired haze datasets makes it challenging to gather a reliable real-world paired haze dataset for model training*. Therefore, there have been efforts to apply semi-supervised (Chen et al. 2021), unsupervised (Li et al. 2021), or Image2Image Translation techniques (Shao et al. 2020) for real-world domain adaptation dehazing. Nonetheless, the actual effectiveness of these methods remains limited and could be further improved. With the development of Gen-

[*]Lei Zhu (leizhu@ust.hk) is the corresponding author.

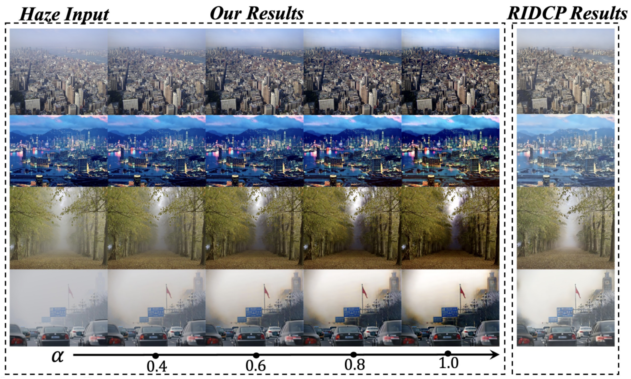

Figure 2: **Left**: Real-world hazy images and corresponding results of the iterative dehazing process using our PromptHaze, which incorporates a controllable dehazing strength factor α. This factor enables adjustable dehazing strength, altering clarity in the dehazed output, as illustrated across the images. **Right**: Comparative results obtained by RIDCP (Wu et al. 2023), the state-of-the-art method for real-world image dehazing. Clearly, our PromptHaze shows superior generalization across various real-world haze scenes, resulting in clearer details and fewer residual haze. The real-world hazy samples from widely-used Fattal's data (Fattal 2014) and RTTS (Li et al. 2018).

erative models (**??????????**), some image restoration efforts (Wu et al. 2023; Chen et al. 2022a; Ye et al. 2023) have explored the use of generative codebook priors for image restoration, such as RIDCP (Wu et al. 2023) for real-world dehazing. This effort attempts to employ generative codebook priors instead of physical-based priors to support real-world image dehazing. RIDCP performs well in many real-world haze scenarios. However, *it struggles to recover clear details and backgrounds from dense haze regions, and its training process is quite cumbersome.*

In this work, we present a fresh paradigm ***PromptHaze*** for real-world image dehazing motivated by utilizing the stable depth prompt from Depth Anything model (Yang et al. 2024). Depth Anything model is a latest, power solution for depth estimation, which significantly improves model generalization ability by utilizing large-scale unlabeled data, achieving state-of-the-art depth results across various scenarios. ***To our surprise, Depth Anything model is capable of delivering stable and reliable depth estimation results under common hazy conditions, as illustrated in the results presented in Figure 1.*** It showcases an impressive zero-shot depth estimation ability, particularly in dealing with typical haze scenarios. Given that the depth information in hazy images is closely linked to the extent of hazy degradation, employing visual prompts from the Depth Anything model to facilitate real-world dehazing emerges as a logical and straightforward approach. However, we found that directly utilizing the output of the depth anything model, such as simply introducing the depth map as an additional condition into the dehazing network, does not yield ideal results. ***The underlying reason is that various thick haze layers could lead to inevitable errors in the estimated depth map. The errors makes it difficult for dehazing model to effectively restore accurate background from challenging haze images.*** This problem motivates us to design a *prompt-by-prompt* dehazing strategy: With the support of the Depth Anything model, our prompt-by-prompt dehazing strategy employs a controllable haze removal network to progressively restore the background. It updates the depth prompt using the results from the previous network inference, then uses this depth prompt to perform the dehazing process again, thereby achieving *stable and reliable iterative cycles*. Additionally, to enhance the dehazing model's capability to model haze degradation, we introduced the prediction of haze-related parameters as a multi-task learning objective. This encourages the encoder of dehazing network to learn *haze-relevant representations*, and concurrently, we introduce haze-related parameter projection onto the decoder through Adaptive Layer Normalization layers. This facilitates the dehazing model's decoder in effectively leveraging the predicted parameters linked with the hazy input.

Compared to state-of-the-art real-world image dehazing methods like DAD, PSD, and RIDCP, our approach, PromptHaze, demonstrates improved real-world applicability, enabling the recovery of stable, authentic backgrounds and details from various types of haze scenes. The contributions of this work can be summarized as follows.

- We demonstrate that the Depth Anything model, serving as a foundational model, can provide strong support for image dehazing methods. Leveraging its robust generalization capability, utilizing stable and reliable depth information to restore degraded images has emerged as a promising direction.
- We are the pioneers in utilizing depth information for real-world image dehazing. The proposed PromptHaze method iteratively updates the depth prompt and background, achieving high-fidelity background recovery.
- We evaluate our PromptHaze on two commonly used real-world dehazing benchmarks and achieve leading performance results across multiple IQA metrics.

Related Work

Single Image Dehazing

Image dehazing, as a classic visual task, has long been widely recognized by the computer vision research community and industry. Over the past decade, deep learning-based dehazing paradigms (Qin et al. 2020; Wu et al. 2021; Cai et al. 2016) have gradually replaced traditional hand-crafted prior-based methods (He, Sun, and Tang 2010; Katiyar and Verma 2016). In the early stages of deep learning-based method development, many approaches (Cai et al. 2016; Li et al. 2021; Zhang, Sindagi, and Patel 2019) often explicitly estimate the transmission map $t(x)$ and global atmospheric light value A in Eq.(1). However, such methods often struggle to cover diverse natural haze scenes. For example, when dealing with heavily foggy or non-uniformly hazy degraded images, these methods (Cai et al. 2016; Li et al. 2017) either fail to generalize or may leave unpleasant artifacts of haze patches. As deep learning technology advances, the computer vision community discovers that attention mechanisms (Qin et al. 2020; Wu et al. 2021) can achieve excellent results in dehazing tasks. Consequently, a series of attention-based models, such as AECR-Net (Wu et al. 2021), FFA-Net (Qin et al. 2020), KDDN (Hong et al. 2020) are proposed. These works make significant progress on synthetic datasets while laying the foundation for subsequent dehazing research.

Some research endeavors attempt to integrate the atmospheric scattering model with unsupervised or semi-supervised methods in order to reduce reliance on large-scale paired data. Several successful examples show promising performance results. However, such methods heavily rely on customized loss constraints or image translation frameworks, making it still challenging to cover the complex and varied haze images found in the real world. RIDCP is a classic groundbreaking work that introduces the use of high-quality codebook prior instead of traditional handcrafted priors into the real-world image dehazing domain. It achieves good generalization effects and stands as one of the pioneering works in the field. Some previous works (Wang et al. 2023; Yang et al. 2022; Yang and Zhang 2022; Fan, Hua, and Li 2021) have attempted to explore integrating or utilizing depth information to enhance image dehazing. *However, these efforts often focus on improving performance on synthetic dehazing benchmarks and do not delve into exploring the potential and impact of foundational models like the depth anything model (Yang et al. 2024) on real-world image dehazing.*

Monocular Depth Estimation

Many early learning-based monocular depth estimation (Ming et al. 2021; Godard et al. 2019; Wofk et al. 2019) relied on complex architectures or prior constraints. Recently, some works (Chen et al. 2016; Xian et al. 2018; Yang et al. 2024) start to gather more training data to achieve good zero-shot depth estimation effects. Among them, Depth Anything Model (Yang et al. 2024) undoubtedly stands as a successful paradigm. It utilizes large-scale unlabeled data and novel supervision methods to enhance the model's zero-shot capability, achieving stable depth estimation results across various scenarios.

Prompt Learning for Vision Model

Prompt learning (Schick and Schütze 2020; Reynolds and McDonell 2021), pioneered in Nature Language Processing (NLP), has recently been applied to computer vision tasks (Khattak et al. 2023; Jia et al. 2022; Zhou et al. 2022). Due to its flexibility and efficiency, it has been used in multimodal learning (Khattak et al. 2023), vision model tuning (Jia et al. 2022), and domain generalization (Zhou et al. 2022). To our knowledge, there is no previous art to utilize depth prompts from foundational models for real-world dehazing. *Our work aims to be a pioneer in this area, drawing the attention of the community to the use of foundational models in the field of real-world image restoration.*

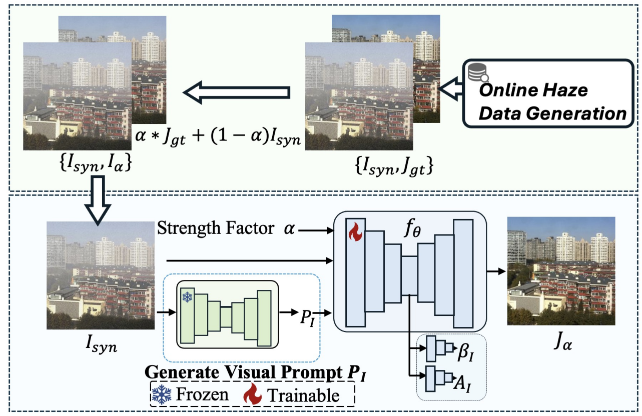

Figure 3: **Overview of the training stage of our PromptHaze.** During each training step, we first randomly sample clean image J_{gt} from the dataset and synthesize hazy image I_{syn} using the offline estimated depth map and the Online Haze Generation Pipeline. Subsequently, using this paired data $\{I_{syn}, J_{gt}\}$ and a random strength factor α , we synthesize a new image I_α. We take the I_α as the network prediction target, with the original synthesized hazy image I_{syn} and the α as inputs to the network. With the aid of the visual prompt feature from the depth anything model, we achieve a flexible dehazing framework with ***controllable dehazing strength.***

Methods

In this section, we will first introduce the motivation behind our proposed method, the PromptHaze. Then, we will discuss our Online Haze Data Generation Pipeline, followed by the training stage and prompt-by-prompt dehazing strategy.

Our Motivation. Although the Depth Anything model provides relatively stable depth information in common haze scenes, challenging haze images contain various levels of image degradation, such as uneven haze layers, low-lighting, color distortion, and noise. These degradation factors could lead to *cumulative errors* in the depth information, resulting in instability and unreliable performance of the dehazing model based on the depth map. To address this, we aim to *achieve a controllable dehazing process, gradually updating the depth prompt through a step-by-step background recovering process, thereby alleviating the impact of cumulative errors.*

Online Haze Data Generation

Online Haze Data Generation Pipeline. Due to the limitations of existing synthetic datasets (Li et al. 2018; Liu et al. 2021) in covering complex and varied haze scenes, utilizing the online haze data pipeline has become the optimal choice for us. Using online data pipeline has been proven feasible by RIDCP (Wu et al. 2023) to generate hazy data containing various complex degradations. We refer to its implementation in RIDCP (Wu et al. 2023) and make slight modifications to meet our needs. The first generation process of the hazy image I_{syn} can be writtern as:

$$I_{syn}(x) = \textbf{JPEG}\left(\mathcal{P}\left(J_{gt}(x)^{\gamma} + \mathcal{N}, e^{\beta d(x)}, A + \Delta A\right)\right). \tag{2}$$

In Eq. (2), γ represents a brightness adjustment factor, $\mathcal{N}$ indicates the Gaussian noise distribution, $d(x)$ denotes the estimated depth map, and β is the factor used to control haze density. To accommodate diverse colors of haze data, we utilize the vector ΔA to manage the color bias of atmospheric light and A to establish the basic global light value. The $\textbf{JPEG}(\cdot)$ denotes the JPEG compression process, which we employ to mimic JPEG artifacts. We utilize the same clean image dataset as RIDCP (Wu et al. 2023), which comprises 500 clean images paired with depth maps.

Synthesize New Target for Controllable Dehazing Model. As mentioned earlier, we aim to achieve a controllable dehazing process that gradually removes haze, progressively restores the background, and updates the depth prompt. In fact, achieving this process is not straightforward. Past dehazing methods (Wu et al. 2023; Chun-Le Guo 2022; Qin et al. 2020; Wu et al. 2021) typically yield only a constant result with each inference from a haze sample, and they cannot control the strength of haze removal. To achieve our goal, we redefine the learning objective of the dehazing network. We introduce a dehazing strength factor α and synthesize the new target I_α using paired synthetic hazy data $\{I_{syn}, J_{gt}\}$:

$$I_\alpha = \alpha * J_{gt} + (1-\alpha) I_{syn}, \tag{3}$$

the range of α is between 0.4 and 1.0. To simplify the learning process, we set the interval to 0.1. During training, a random value within this range is selected as the value of α each sample.

Learning a Model with Controllable Dehazing Strength

Training Step of PromptHaze. With the introduction of I_α and α, achieving the controllable dehazing process has become a feasible approach for PromptHaze. As shown in Figure 3, during each training step, we utilize the Depth Anything model (Yang et al. 2024) D to obtain the depth prompt feature P_I. Subsequently, we input P_I, I_{syn}, I_α into the network, while also introducing the crucial coefficient α into the network. This enables the network to explicitly learn to control the dehazing strength based on the coefficient α, aiming to make the dehazing model output closer to I_{alpha} rather than J_{gt}. We extract the last layer feature before the Depth Anything Model decoder head as the depth prompt feature, which allows for more comprehensive preservation of depth characteristic information. Due to limited space, the design of our model's basic block is detailed in the supplementary materials.

Regression Learning of Haze-Relevant Parameters. Considering that we can obtain additional information about haze degradation from the synthesis pipeline, such as the global atmospheric light value A and the haze intensity value β, we introduce the regression learning of haze-relevant parameters to encourage the encoder of the model to learn haze-relevant representations.As depicted in Figure 3, we employ pooling layers and MLPs to form two output heads. These heads utilize latent features extracted by the encoder to acquire haze-related parameters.

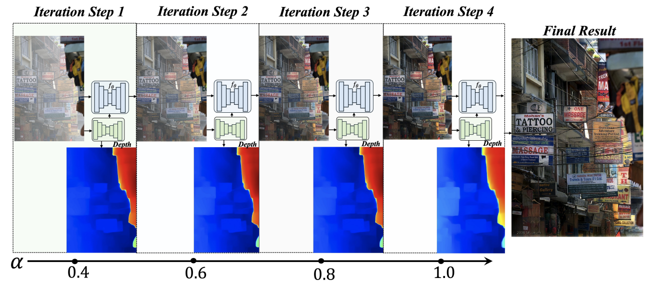

Figure 4: **Overview of the Iterative Dehazing Process with our Prompt-by-Prompt Dehazing Strategy.** The images in the top row demonstrate consecutive iteration steps, showcasing the gradual removal of haze from the input image by applying the dehazing model at each step. The corresponding depth maps are visualized in the bottom row, with the dehazing strength factor, α, increasing from left to right. The depth maps use a color gradient where cooler colors (blue) represent farther distances and warmer colors (red) indicate closer objects. **The iterative process refines the visibility of the scene with each step, as indicated by the clearer images and more defined depth cues in the final results**.

Haze-Relevant Parameters Projection. We empirically find that solely employing regression learning of haze-relevant parameters does not significantly improve the performance. Hence, we propose Haze-related Parameters Projection. We utilize **adaptive layer normalization (AdaLN)** to re-utilize crucial representations before MLPs heads and after the pooling layer, predicting them into each block of the decoder. In formal terms, $AdaLN(x, e) = e_s \textbf{LayerNorm}(x) + e_b$, where x represents an input feature within a decoder block, and e_s, e_b are derived from a linear projection of the embedding of haze-related parameters.

Prompt-by-Prompt Dehazing Strategy

With the trained controllable dehazing model, we could perform the Prompt-by-Prompt Dehazing Strategy. The proposed novel strategy leverages the previous inference results of dehazing to *obtain more accurate depth cues as new prompts*, aiming to achieve better dehazing results. As shown in Figure 4, by utilizing a progressive factor α, we mitigate potential errors in depth prompts caused by haze with multiple inference iterations. Through controllable hierarchical inference, our approach effectively reduces the interference of haze degradation on the Depth Anything Model, gradually refining depth prompts and restoring clean backgrounds. Through our Prompt-by-Prompt Dehazing Strategy, *a complementary cyclic interaction* between the two models is achieved.

Training Objective of PromptHaze

The training objective of PromptHaze consists of three components, which we will introduce one by one.

Basic Reconstruction Loss. We adpot the Charbonnier loss (Charbonnier et al. 1994) as our basic reconstruction objective function. It could be represented by:

$$\mathcal{L}_{\text{rec}} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\left\|J_{\alpha}^{i} - I_{\alpha}^{i}\right\|^{2} + \epsilon^{2}}, \tag{4}$$

within constant ϵ emipirically set to $1e^{-3}$ for our all experiments.

Contrastive Regularization Loss. Contrastive Regularization Learning (Wu et al. 2021) has been proven effective in image dehazing, enhancing the model's generalization capabilities. Therefore, we also introduce Contrastive Regularization Loss (Wu et al. 2021) to aid in improving the generalization ability of our approach:

$$\mathcal{L}_{cr} = CR(J_{\alpha}, I_{\alpha}, I_{syn}), \tag{5}$$

where the anchor is J_{α}, positive sample is I_{α} and negative samples is I_{syn}.

Haze-Relevant Parameters Regression Loss. We directly use the L_1 loss to boost the learning of haze-relevant parameters regression. The loss could be represented by:

$$\mathcal{L}_{reg} = \left\|\beta_I - \beta_{gt}\right\|_1 + \left\|A_I - A_{gt}\right\|_1, \tag{6}$$

where the β_I and A_I denotes the estimated hazy density parameter β and the global light value A, respectively. And the β_{gt} and $_{gt}$ represent the corresponding ground-truth values.

The total training objective is the combination of the above losses:

$$\mathcal{L}_{total} = \mathcal{L}_{rec} + \lambda_{cr}\mathcal{L}_{cr} + \lambda_{reg}\mathcal{L}_{reg} \tag{7}$$

where the λ_{cr} and λ_{reg} denote the scale factors of corresponding loss, respectively.

Experiments

Datasets

Training Datasets. For a fair and convincing comparison, we employ the identical set of 500 clean images utilized in RIDCP (Wu et al. 2023) within our Online Haze Data Generation Pipeline.Please note that the hazy data is generated during the training phase by the data generation pipeline.

Training Datasets. We qualitatively and quantitatively evaluate our PromptHaze method on the RTTS dataset (Li et al. 2018), which comprises over 4,000 real hazy images with diverse scenes, times, resolutions, and degradations. Additionally, we also conduct some visual comparisons using Fattal's dataset (Fattal 2014), consisting of 31 classic real hazy cases.

Training Details. We implement our PromptHaze using the PyTorch framework, harnessing four NVIDIA RTX 4090 GPUs. We utilize the AdamW optimizer with beta values set to 0.9 and 0.999. The batch size is set to 7, and the initial learning rate is set at 2×10^{-4}, employing a cosine annealing strategy for gradual learning rate reduction. Data augmentation techniques, including horizontal flipping, random resizing and cropping, and random image rotation at $45°$ and $90°$, are applied during training. Each paired data is cropped to a size of 256x256. We set the λ_{cr} and the λ_{reg} to 0.5 and 0.2, respectively. The proposed PromptHaze is trained with our Online Haze Data Generation Pipeline for 15K iterations. We make slight modifications to the efficient image restoration model NAFNet (Chen et al. 2022b) to meet our requirements. The parameter size of our dehazing model is 64.7 M, while the Depth Anything Model is the small version, with a parameter size of 24.8 M. For detailed model implementation, please refer to our supplementary material.

Method	FADE↓	BRISQUE↓	NIMA↑	US↑
Hazy image	2.484	37.011	4.3250	0.024
MSBDN (Dong et al. 2020)	1.363	28.743	4.1401	0.037
FFA-Net (Qin et al. 2020)	1.421	35.417	3.9701	0.026
Dehamer (Guo et al. 2022)	1.895	33.866	3.8663	0.051
DAD (Shao et al. 2020)	1.130	32.727	4.0055	0.103
PSD (Chen et al. 2021)	0.920	25.239	4.3459	0.087
RIDCP (Wu et al. 2023)	0.944	18.782	4.4267	0.162
PromptHaze	0.929	14.624	5.1141	0.512

Table 1: Quantitative comparison on RTTS dataset. Red indicates the best result.

Comparison with State-of-the-Art Real-world Dehazing Methods

We compare the performance of our proposed PromptHaze against several state-of-the-art dehazing methods, including classic image dehazing networks (Qin et al. 2020; Dong et al. 2020; Chun-Le Guo 2022), the domain adaptation dehazing method (Shao et al. 2020) based on Image Translation , and the semi-supervised approach (Chen et al. 2021). Our experimental setup is designed from both quantitative and qualitative perspectives. Furthermore, we also conduct a user study (US) to validate the subjective performance of our method.

Quantitative Comparison. We perform a quantitative comparison on publicly available real-world dehazing datasets using non-reference metrics. We initially utilize the Fog Aware Density Evaluator (FADE) (Choi, You, and Bovik 2015) for the esitmation of haze density. Additionally, we incorporate two widely used image quality assessment metrics: BRISQUE (Mittal, Moorthy, and Bovik 2011) and NIMA (Talebi and Milanfar 2018). Quantitative comparisons are conducted on the RTTS dataset with three dehazing methods (MSBDN (Dong et al. 2020), Dehamer (Chun-Le Guo 2022), and FFA-Net (Qin et al. 2020)) showing outstanding performance on synthetic hazy image datasets, along with three real-world dehazing methods (DAD (Shao et al. 2020), PSD (Chen et al. 2021), and RIDCP (Wu et al. 2023)). The results are presented in Table 1. Our proposed PromptHaze achieves the best results in terms of BRISQUE and NIMA. For FADE, our method ranks second, slightly below PSD. However, PSD tends to produce over-enhanced or oversaturated results and often leaves residual haze (See Figure 5), leading to inaccurate assessments. Overall, PromptHaze secures the best outcomes on quantitative metrics.

User Study. We conduct a user study to subjectively evaluate and compare the proposed method against other techniques. For this comparison, we select 80 images from the RTTS dataset and invite eight volunteers. Before the user study commences, we provide three guidelines: 1) The pri-

Figure 5: Visual comparison on RTTS dataset.

mary evaluation criterion is the extent of residual haze, especially the dense fog in distant areas. 2) The restoration of natural color tones. 3) The level of noise and the presence of artifacts in the results. Subsequently, images are grouped and presented to the observers. Each group consists of the input image alongside results generated through different methods. Observers are required to view each group for at least five seconds before selecting the best one. We tally the percentage of times each method is chosen as the best, with the final scores listed in Table 1. The proposed PromptHaze method achieves the highest score, significantly outperforming the runner-up, further demonstrating our method's superior dehazing capability.

Qualitative Comparison. We conduct a qualitative comparison on the RTTS and Fattal datasets, as illustrated in Figures 5 and 6. We could bserve that Dehamer, FFA-Net, and MSBDN largely fail to handle real hazy images effectively. PSD tends to produce overly bright and enhanced results, yet its dehazing capability is quite limited, leaving much haze residue. DAD can effectively remove haze in some cases, while RIDCP also leaves haze residue in some challenging scenes. Compared to other methods, our approach generates the best results in terms of haze residue removal and background detail restoration.

Ablation Study

In the ablation study Table 2, we demonstrate the impact of different configurations on image quality assessment metrics on the RTTS dataset. The two metrics in question are BRISQUE and NIMA, where a lower BRISQUE value indicates higher image quality, and a higher NIMA value signifies better image quality. Due to the page length limitation, we provide further ablation studies and related analyses in our supplementary materials.

Depth Prompt (w/o Depth Prompt). This configuration yields a BRISQUE score of 19.712 and a NIMA score of 4.1041. Compared to PromptHaze, the BRISQUE score is significantly higher, and the NIMA score is lower. This indicates that utilizing the depth prompt in PromptHaze significantly improves the quality of the results, reducing haze residue and improving the aesthetics of the images.

Settings	BRISQUE ↓	NIMA ↑
w/o Depth Prompt	19.712	4.1041
w/o Contrastive Regularization	16.824	4.5190
w/o Prompt-by-Prompt Dehazing Strategy	16.791	4.8437
w/o Haze-Relevant Parameters Projection	15.944	4.8141
Ours(PromptHaze)	**14.624**	**5.1141**

Table 2: The ablation studies on RTTS dataset (Li et al. 2018).

Contrastive Regularization (w/o Contrastive Regularization). With a BRISQUE score of 16.824 and a NIMA score of 4.5190, it is evident that contrastive regularization plays a crucial role in enhancing the visual quality of images, particularly in improving image contrast and clarity.

Prompt-by-Prompt Dehazing Strategy (w/o Prompt-by-Prompt Dehazing Strategy). This configuration results in a BRISQUE score of 16.791 and a NIMA score of 4.8437. Compared to PromptHaze, there is a noticeable decline in image quality. This outcome suggests that employing a Prompt-by-Prompt Dehazing Strategy effectively enhances the overall quality and visual appeal of the images.

Haze-Relevant Parameters Projection (w/o Haze-Relevant Parameters Projection). With a BRISQUE score of 15.944 and a NIMA score of 4.8141, this configuration demonstrates that projecting haze-relevant parameters into the dehazing process can further improve image quality, especially in terms of restoring the true visual effects of the images.

By comparing the BRISQUE and NIMA scores under

Figure 6: Visual comparison on Fattal's data (Fattal 2014)

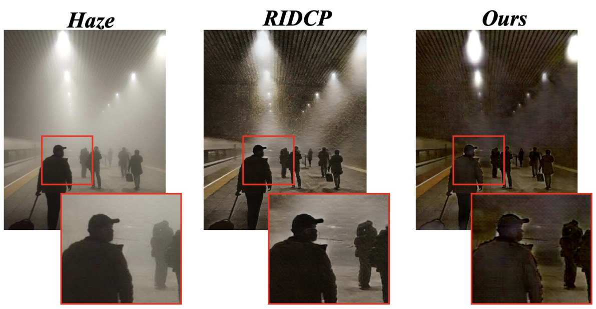

Figure 7: A challenging case of uneven haze for real-world dehazing methods. We can observe that RIDCP (Wu et al. 2023) struggles to remove dense haze in such sample, whereas our proposed method, PromptHaze, can more adeptly eliminate dense haze. However, it may result in unpleasant artifacts and distortions in the color of the foreground. This haze sample from RTTS dataset (Li et al. 2018).

different configurations, we observe that depth prompt, contrastive regularization, the Prompt-by-Prompt Dehazing Strategy, and the projection of haze-relevant parameters all play critical roles in enhancing the quality of dehazed images.

Limitations and Future Works

Through in-depth evaluation and analysis of the proposed method, PromptHaze, we have identified several challenges that urgently need to be addressed. We detail the limitations of our method here, hoping that future work can overcome these issues. Firstly, there is the problem of artifacts, such as those demonstrated in Figure 7. We find that removing haze from some challenging scenes can cause artifacts and color distortions in the foreground, which are undesirable outcomes. Additionally, due to the need for multiple inferences, the speed of the PromptHaze method is not conducive to real-time applications. We look forward to future efforts addressing these challenges.

Conclusion

In this paper, we introduce PromptHaze, a novel approach for real-world image dehazing that leverages the stable and reliable depth information provided by the Depth Anything model. This innovative method marks a significant advancement in the field of image dehazing, offering a new perspective on utilizing depth information for enhancing image clarity in hazy conditions. Our approach shows superiority over previous dehazing methods by employing a unique prompt-by-prompt dehazing strategy, which iteratively refines the depth information and dehazed results to achieve high-fidelity background recovery. Our findings underscore the importance of reliable depth information in tackling the challenging task of real-world image dehazing. Through comprehensive evaluations on widely recognized real-world dehazing benchmarks, PromptHaze has set new standards, outperforming existing state-of-the-art methods across various image quality assessment metrics.

Acknowledgments

This work is supported by the Guangzhou-HKUST(GZ) Joint Funding Program (No. 2023A03J0671), the Guangzhou Municipal Science and Technology Project (Grant No. 2024A04J4230), the Guangzhou Industrial Information and Intelligent Key Laboratory Project (No. 2024A03J0628), the Nansha Key Area Science and Technology Project (No. 2023ZD003), Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things(No.2023B1212010007).

References

Berman, D.; Avidan, S.; et al. 2016. Non-local image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1674–1682.

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

Charbonnier, P.; Blanc-Feraud, L.; Aubert, G.; and Barlaud, M. 1994. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proceedings of 1st International Conference on Image Processing*, volume 2, 168–172. IEEE.

Chen, C.; Shi, X.; Qin, Y.; Li, X.; Han, X.; Yang, T.; and Guo, S. 2022a. Real-World Blind Super-Resolution via Feature Matching with Implicit High-Resolution Priors.

Chen, L.; Chu, X.; Zhang, X.; and Sun, J. 2022b. Simple baselines for image restoration. *arXiv preprint arXiv:2204.04676*.

Chen, W.; Fu, Z.; Yang, D.; and Deng, J. 2016. Single-image depth perception in the wild. *Advances in neural information processing systems*, 29.

Chen, Z.; Wang, Y.; Yang, Y.; and Liu, D. 2021. PSD: Principled synthetic-to-real dehazing guided by physical priors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7180–7189.

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

Chun-Le Guo, S. A. R. C. W. R. C. L., Qixin Yan. 2022. Image Dehazing Transformer with Transmission-Aware 3D Position Embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Dong, H.; Pan, J.; Xiang, L.; Hu, Z.; Zhang, X.; Wang, F.; and Yang, M.-H. 2020. Multi-scale boosted dehazing network with dense feature fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2157–2167.

Fan, G.; Hua, Z.; and Li, J. 2021. Multi-scale depth information fusion network for image dehazing. *Applied Intelligence*, 51(10): 7262–7280.

Fattal, R. 2014. Dehazing using color-lines. *ACM transactions on graphics (TOG)*, 34(1): 1–14.

Godard, C.; Mac Aodha, O.; Firman, M.; and Brostow, G. J. 2019. Digging into self-supervised monocular depth estimation. In *Proceedings of the IEEE/CVF international conference on computer vision*, 3828–3838.

Guo, C.-L.; Yan, Q.; Anwar, S.; Cong, R.; Ren, W.; and Li, C. 2022. Image dehazing transformer with transmission-aware 3d position embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5812–5820.

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

Hong, M.; Xie, Y.; Li, C.; and Qu, Y. 2020. Distilling image dehazing with heterogeneous task imitation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3462–3471.

Jia, M.; Tang, L.; Chen, B.-C.; Cardie, C.; Belongie, S.; Hariharan, B.; and Lim, S.-N. 2022. Visual prompt tuning. In *European Conference on Computer Vision*, 709–727. Springer.

Katiyar, K.; and Verma, N. 2016. Single Image Haze Removal Algorithm using Color Attenuation Prior and Multi-Scale Fusion. *International Journal of Computer Applications*, 141(10).

Khattak, M. U.; Rasheed, H.; Maaz, M.; Khan, S.; and Khan, F. S. 2023. Maple: Multi-modal prompt learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19113–19122.

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

Li, B.; Peng, X.; Wang, Z.; Xu, J.; and Feng, D. 2017. Aodnet: All-in-one dehazing network. In *Proceedings of the IEEE international conference on computer vision*, 4770–4778.

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

Liu, X.; Ma, Y.; Shi, Z.; and Chen, J. 2019. Griddehazenet: Attention-based multi-scale network for image dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 7314–7323.

Liu, Y.; Zhu, L.; Pei, S.; Fu, H.; Qin, J.; Zhang, Q.; Wan, L.; and Feng, W. 2021. From Synthetic to Real: Image Dehazing Collaborating with Unlabeled Real Data. *arXiv preprint arXiv:2108.02934*.

Ming, Y.; Meng, X.; Fan, C.; and Yu, H. 2021. Deep learning for monocular depth estimation: A review. *Neurocomputing*, 438: 14–33.

Mittal, A.; Moorthy, A. K.; and Bovik, A. C. 2011. Blind/referenceless image spatial quality evaluator. In *2011 conference record of the forty fifth asilomar conference on signals, systems and computers (ASILOMAR)*, 723–727. IEEE.

Qin, X.; Wang, Z.; Bai, Y.; Xie, X.; and Jia, H. 2020. FFA-Net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 11908–11915.

Reynolds, L.; and McDonell, K. 2021. Prompt programming for large language models: Beyond the few-shot paradigm. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–7.

Schick, T.; and Schütze, H. 2020. Exploiting cloze questions for few shot text classification and natural language inference. *arXiv preprint arXiv:2001.07676*.

Shao, Y.; Li, L.; Ren, W.; Gao, C.; and Sang, N. 2020. Domain Adaptation for Image Dehazing. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Talebi, H.; and Milanfar, P. 2018. NIMA: Neural image assessment. *IEEE transactions on image processing*, 27(8): 3998–4011.

Wang, C.; Pan, J.; Lin, W.; Dong, J.; and Wu, X.-M. 2023. SelfPromer: Self-Prompt Dehazing Transformers with Depth-Consistency. *arXiv preprint arXiv:2303.07033*.

Wofk, D.; Ma, F.; Yang, T.-J.; Karaman, S.; and Sze, V. 2019. Fastdepth: Fast monocular depth estimation on embedded systems. In *2019 International Conference on Robotics and Automation (ICRA)*, 6101–6108. IEEE.

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

Wu, R.; Duan, Z.; Guo, C.; Chai, Z.; and Li, C. 2023. RIDCP: Revitalizing Real Image Dehazing via High-Quality Codebook Priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Xian, K.; Shen, C.; Cao, Z.; Lu, H.; Xiao, Y.; Li, R.; and Luo, Z. 2018. Monocular relative depth perception with web stereo data supervision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 311–320.

Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J. M.; and Luo, P. 2021. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 34.

Yang, F.; and Zhang, Q. 2022. Depth aware image dehazing. *The Visual Computer*, 38(5): 1579–1587.

Yang, L.; Kang, B.; Huang, Z.; Xu, X.; Feng, J.; and Zhao, H. 2024. Depth anything: Unleashing the power of large-scale unlabeled data. *arXiv preprint arXiv:2401.10891*.

Yang, Y.; Wang, C.; Liu, R.; Zhang, L.; Guo, X.; and Tao, D. 2022. Self-augmented unpaired image dehazing via density and depth decomposition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2037–2046.

Ye, T.; Chen, S.; Bai, J.; Shi, J.; Xue, C.; Jiang, J.; Yin, J.; Chen, E.; and Liu, Y. 2023. Adverse Weather Removal with Codebook Priors. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12653–12664.

Ye, T.; Jiang, M.; Zhang, Y.; Chen, L.; Chen, E.; Chen, P.; and Lu, Z. 2021. Perceiving and Modeling Density is All You Need for Image Dehazing. *arXiv preprint arXiv:2111.09733*.

Zhang, H.; Sindagi, V.; and Patel, V. M. 2019. Joint transmission map estimation and dehazing using deep networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(7): 1975–1986.

Zhang, Z.; Zhao, L.; Liu, Y.; Zhang, S.; and Yang, J. 2020. Unified density-aware image dehazing and object detection in real-world hazy scenes. In *Proceedings of the Asian Conference on Computer Vision*.

Zhou, K.; Liu, Z.; Qiao, Y.; Xiang, T.; and Loy, C. C. 2022. Domain generalization: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.