Teaching Tailored to Talent: Adverse Weather Restoration via Prompt Pool and Depth-Anything Constraint

Sixiang Chen[1], Tian Ye[1], Kai Zhang[1], Zhaohu Xing[1], Yunlong Lin[2], and Lei Zhu[1,3]✉

[1] The Hong Kong University of Science and Technology (Guangzhou), China
[2] Xiamen University, China
[3] The Hong Kong University of Science and Technology, Hong Kong SAR, China
leizhu@ust.hk
Project page: https://ephemeral182.github.io/T3-DiffWeather

Abstract. Recent advancements in adverse weather restoration have shown potential, yet the unpredictable and varied combinations of weather degradations in the real world pose significant challenges. Previous methods typically struggle with dynamically handling intricate degradation combinations and carrying on background reconstruction precisely, leading to performance and generalization limitations. Drawing inspiration from prompt learning and the "**T**eaching **T**ailored to **T**alent" concept, we introduce a novel pipeline, **T^3-DiffWeather**. Specifically, we employ a prompt pool that allows the network to autonomously combine sub-prompts to construct weather-prompts, harnessing the necessary attributes to adaptively tackle unforeseen weather input. Moreover, from a scene modeling perspective, we incorporate general prompts constrained by Depth-Anything feature to provide the scene-specific condition for the diffusion process. Furthermore, by incorporating contrastive prompt loss, we ensures distinctive representations for both types of prompts by a mutual pushing strategy. Experimental results demonstrate that our method achieves state-of-the-art performance across various synthetic and real-world datasets, markedly outperforming existing diffusion techniques in terms of computational efficiency.

Keywords: Adverse weather restoration · Diffusion model · Prompt learning · Teaching tailored to talent

1 Introduction

With the growth of the community, image restoration in adverse weather conditions has become increasingly significant [13, 18, 19, 21, 40, 74, 82]. To meet practical demands effectively, research is increasingly focusing on the all-in-one removal of multiple weather degradations [20, 46, 58, 76, 87] as a primary objective.

Contrasting with restoration tasks targeted at specific weather conditions, multi-weather restoration involves a composition of various weather phenomena.

Early developments employed methods such as neural architecture search [46] or distillation [20] to combine models tailored to individual tasks, but these approaches are complicated and cumbersome. Moreover, several methods [76, 88] have attempted to employ a codebook as a reliable prior for guiding image restoration or utilized shared learnable queries to adapt different weather degradations. However, such paradigms are not aware of the differences and similarities between degradations. Recently, the WGWS framework [102] analyzed the weather-general and weather-specific features and performed targeted parameter learning for such characteristics. Yet, its two-stage design and the need for customized modifications within different architectures remain intricate. In addition, unlike dealing with a single weather, adverse weather will cause unseen and unpredictable degradation combinations in the real world, which poses a challenge to handling degradations adaptively. Hence,

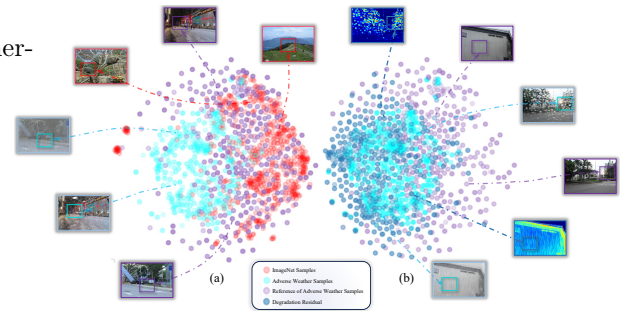

Fig. 1: t-SNE visualization of different feature distributions. (a). Scenes with different contents also have significant commonalities compared to degradations. And there are some differences and commonalities between degradations and degradations. (b). Degradation residuals can represent degradations to a certain extent and be distinguished from the background.

1. How to effectively and flexibly model complicated and unpredictable weather combinations in the real world remains an open question.

Furthermore, benefiting from recent advancements in diffusion models [28], there is the first diffusion-based method—WeatherDiffusion [58]—which has demonstrated the superiority of generative paradigms over regression models in reconstructing clean backgrounds from adverse weather images. However, its shortcomings are evident: the original degraded image as a condition does not adequately guide the reconstruction from noise to clean images, and requires a certain number of steps for sampling. For diffusion models, discovering sufficiently informative conditions is essential for high-quality image reconstruction [27, 42, 48, 81]. Therefore, ***2. Designing a condition that equips rich information on adverse weather samples is critical for diffusion process.***

To address the aforementioned challenges, we introduce a novel paradigm. Inspired by prompt learning [32, 99], we claim that prompts can be employed to craft a comprehensive condition. **i).** Unlike the recent paradigm [59] that utilized a shared set of learnable prompts to adapt to varying degraded images, we design a prompt pool. This pool can fully exploit the differences and similari-

ties among weather degradations, thereby offering the network a wider range of options. Specifically, we enable the network to autonomously construct the necessary weather information based on the input degradation residual and freely assemble a specific set of weather-prompts for unpredictable phenomena. **ii).** We observe that the scene features behind adverse weather often share commonalities (see Fig.1 (a)), inspiring us to devise general prompts specifically for background modeling. Drawing inspiration from the recent breakthroughs in Depth-Anything [83], we note this model exhibits exceptional robustness when handling extreme samples. This observation leads us to extract critical features from Depth-Anything, utilizing them to direct our general prompts. **iii).** Additionally, we introduce a compact contrastive prompt loss to further regulate two types of prompts, and integrate them seamlessly into the diffusion process. At the core of our approach, a "**T**eaching **T**ailored to **T**alent" paradigm is resembled , adeptly guiding distinct weather combinations and their respective scene backgrounds. Building upon this foundation, we introduce a novel diffusion-based architecture: ***T*^{3}*-DiffWeather***. It achieves SOTA performance across various adverse weather benchmarks while requiring only a tenth of the sampling steps compared to the latest WeatherDiffusion [58].

As an overview, the contributions of our work are summarized as follows:

- We introduce a novel prompt pool. Capitalizing on the similarities and differences among various weather conditions, proposed prompt pool empowers the network to autonomously combine sub-prompts, effectively constructing diverse weather-prompts to enhance representation for complicated weather degradations.
- Inspired by the shared attributes within the scenes of degraded samples, we have crafted general prompts specifically tailored for background understanding. For the first time, we propose to utilize the robust features of Depth-Anything [83] as a constraint for general prompts.
- We incorporate a compact contrastive prompt loss to further boost the prompt representation of two designs. Overall, our diffusion-based architecture, T^3-DiffWeather, achieves SOTA performance on multi-weather restoration benchmarks with significantly fewer inference steps.

2 Related Works

2.1 Image Restoration in Adverse Weather Conditions

Dehazing: The field of single-image dehazing has witnessed remarkable advancements in recent years [1, 50, 52, 61, 74, 79, 91]. DehazeFormer [74] adopted a transformer-based approach, tackling the complicated hazy images through distinct window processing. Meanwhile, the MB-TaylorFormer [61] leveraged a linear Transformer architecture grounded in Taylor series expansion to clarify hazy scenes effectively.

Deraining: The progress in single-image rain removal is steadily increasing, including rain streaks [21, 65, 85, 101] and raindrops [13, 60, 62, 80, 82]. IDT [82] developed a transformer-based technique that combines window-based and spatial

transformers to enhance the precision of rain streak modeling. UDR-S2Former [13] leveraged uncertainty to refine the sparse ViT model for improved performance of raindrop removal.

Desnowing: Unlike dehazing and deraining, single-image snow removal presents a greater challenge [9,15–18,23,49,90]. JSTASR [18] introduced a framework capable of addressing both haze and snow removal simultaneously. MSP-Former [16] was the inaugural attempt at a single-image snow removal network utilizing a transformer architecture. Nevertheless, similar to haze and rain removal, these innovative approaches still grapple with limitations when confronted with other variants of extreme weathers.

Multi-Weather Restoration: Adverse weather restoration endeavors to develop a consolidated network to adeptly address weather-induced image degradations [20,46,58,76,86,88,102]. The pioneering work in this domain was the All-in-One [46], the extensive parameterization due to NAS may make it impractical for real-world deployment. TransWeather [76] introduced a weather-type decoder capable of interpreting diverse weather degradations, yet its fixed queries cannot explicitly consider the degradations of different weather and lacks background-level modeling. WeatherDiffusion [58] presented a diffusion-based method that harnessed the capabilities of diffusion models for weather removal, achieving SOTA results across various benchmarks. Nonetheless, the slow inference speed and the absence of precise prompt conditions may hinder its widespread practical application.

2.2 Conditional Diffusion Models

Recent advancements in denoising diffusion probabilistic models (DDPM) [29] have captured intricate distributions with accuracy that exceeds other generative frameworks, including GANs. To further enhance the precision and realism of the generated outputs, diffusion models often incorporated additional conditioning or guidance mechanisms, as evidenced by recent studies [2,3,26,37]. In the field of image restoration, the prevailing approach involves feeding networks with concatenated degraded inputs to yield outputs of high fidelity quality [42,43,58,68,89] compared with traditional regressive models [25,35,38,39,41,66,78]. To further refine the denoising process, networks often incorporated single task-specific prompts such as masks or textual information [27,84] as embedded guidance. However, the recent WeatherDiffusion [58] typically used degraded images as the sole condition, which may result in performance limitations for addressing all-in-one restoration tasks.

2.3 Prompt Learning

Prompt learning has been increasingly applied to computer vision [5,6,57,70,71,96]. This technique involved the insertion of task-specific prompt tokens prior to the input, equipping pre-trained models with the necessary knowledge to perform new tasks without extensive fine-tuning. The approach of Context Optimization

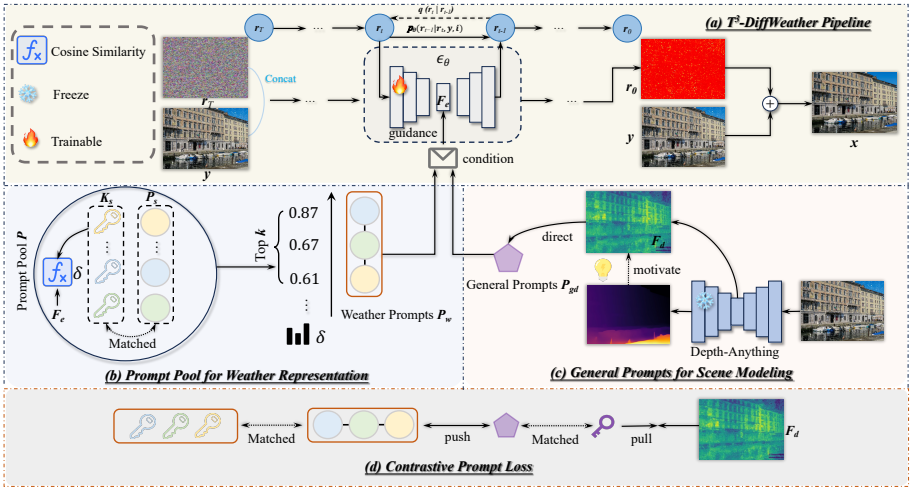

Fig. 2: The overview of proposed method. (a) showcases our pipeline, which adopts an innovative strategy focused on learning degradation residual and employs the information-rich condition to guide the diffusion process. (b) illustrates the utilization of our prompt pool, which empowers the network to autonomously select attributes needed to construct adaptive weather-prompts. (c) depicts the general prompts directed by Depth-Anything constraint to supply scene information that aids in reconstructing residuals. (d) shows the contrastive prompt loss, which exerts constraints on prompts driven by two distinct motivations, enhancing their representations.

(CoOp) [99] leveraged this for the CLIP model [64], refining prompts in a continuous space through backpropagation. Conditional Context Optimization (CoCoOp) [98] innovated further by producing input-dependent prompt residuals to enhance generalization ability. For low-level tasks, PromptIR [59] introduced a learnable prompt module that generates shared prompts responsive to various degradation types. Additionally, recent researches utilized text prompts to guide image restoration networks [22, 48, 75]. However, the use of sparse text embeddings could lead to performance limitations and increase complexity due to the necessity for additional multi-modal models.

3 Methods

Novel Pipeline T^3-DiffWeather. For adverse weather restoration, the intricate blend of degradations in the real world poses significant challenges to obtaining clean backgrounds. Consequently, developing a model that can effectively adapt complicated degradation combinations is crucial. We introduce a novel pipeline, T^3-DiffWeather, whose key principle is "**T**eaching **T**ailored to **T**alent." Inspired by prompt learning [32, 59, 99], our design incorporates instance-wise weather-prompts tailored for specific degradations and general prompts for scene information, efficiently exploiting both the disparate and

shared attributes present in degraded images. We leverage these prompts as the condition to guide the diffusion process with rich information.

Specifically, leveraging insights from Fig.1, we observe that degradations exhibit more distinct features compared to the background (as shown in Fig.1(a)), and degradation residual $\boldsymbol{r}_d$ (subtract the degraded image $\boldsymbol{y}$ from the clean image $\boldsymbol{x}$) provides a clearer representation of the degraded image (as illustrated in Fig.1(b)). We claim that the degradations are a primary factor for the difficulty in restoration. Therefore, we pivot the reconstruction target of the diffusion model toward the degradation residual. The training objective is (see supplementary material):

$$\mathcal{L}_{res} = \mathbb{E}\left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta \left(\sqrt{\bar{\alpha}}(\underbrace{\boldsymbol{x} - \boldsymbol{y}}_{\text{residual } \boldsymbol{r}_d}) + \sqrt{1-\bar{\alpha}}\boldsymbol{\epsilon}, \boldsymbol{y}, \boldsymbol{c} \right) \right\|_2^2 . \tag{1}$$

where c denotes the condition built by weather-prompts $\boldsymbol{\mathcal{P}}_w$ and general prompts $\boldsymbol{\mathcal{P}}_{gd}$. We simply embed them into the latent layer in the diffusion network through cross-attention, similar to the text embedding in SD [67]. This process is naturally efficient and does not take up much computation overhead. Given the feature embedding $\boldsymbol{\mathcal{F}}_e \in \mathbb{R}^{H\times W\times D}$ in the latent layer, The formula can be expressed as follows:

$$\begin{aligned} \boldsymbol{\mathcal{F}}_e^{'} &= \text{softmax}\left(\frac{\boldsymbol{\mathcal{Q}}(\boldsymbol{\mathcal{F}}_e)\boldsymbol{\mathcal{K}}(\boldsymbol{\mathcal{P}}_w)^T}{\sqrt{\boldsymbol{\mathcal{D}}}} \right) \boldsymbol{\mathcal{V}}(\boldsymbol{\mathcal{P}}_w), \\ \hat{\boldsymbol{\mathcal{F}}}_{\boldsymbol{e}} &= \text{softmax}\left(\frac{\boldsymbol{\mathcal{Q}}(\boldsymbol{\mathcal{F}}_e^{'})\boldsymbol{\mathcal{K}}(\boldsymbol{\mathcal{P}}_{gd})^T}{\sqrt{\boldsymbol{\mathcal{D}}}} \right) \boldsymbol{\mathcal{V}}(\boldsymbol{\mathcal{P}}_{gd}), \end{aligned} \tag{2}$$

where $\boldsymbol{\mathcal{Q}}(\cdot), \boldsymbol{\mathcal{K}}(\cdot), \boldsymbol{\mathcal{V}}(\cdot)$ are the query, key, and value functions. The $\hat{\boldsymbol{\mathcal{F}}}_{\boldsymbol{e}}$ is the output feature embedding.

3.1 Prompt Pool for Weather Representation

Motivation I: The restoration from adverse weather is often impeded by the complicated and varied combinations of degradations, which can influence network performance. Unlike the substantial domain gap encountered in general restoration between various types of degradations [7,8,95], weather degradations in the real world exhibit some similar attributes, such as haze veiling and low contrast [102]. Concurrently, degradations specific to distinct conditions manifest unique attributes varying in shape and scale. These differences and similarities inspire us to explicitly leverage degradation characteristics to enhance the representation of degradations.

Leveraging the advancements in prompt learning for image restoration, we posit that the network should adaptively learn the characteristics of degradations and autonomously construct suitable weather-prompts. Consequently, we

introduce the design of a prompt pool. This design triggers the network to selectively choose sub-prompts from the pool, crafting a unique weather-prompt tailored to each sample. Such an autonomous construction explicitly takes into account the similarities (shared sub-prompts) and differences (independent sub-prompts) under varying weather conditions. Specifically, given our prompt pool $\boldsymbol{\mathcal{P}} = \{\boldsymbol{\mathcal{P}}_s^i\}_{i=1}^N$ with each sub-prompts $\boldsymbol{\mathcal{P}}_s^i \in \mathbb{R}^{L_s \times D}$ (L_s denotes the length of tokens), where i represents the index of a specific sub-prompts and N is the prompt pool size. For the input (degradation residual) embedding $\boldsymbol{\mathcal{F}}_e$, the weather-prompt construction function Ψ can be defined as:

$$\boldsymbol{\mathcal{P}}_w = \Psi(\boldsymbol{\mathcal{P}}, \boldsymbol{\mathcal{F}}_e; \Theta), \tag{3}$$

Here, Θ parameterizes the selection process to optimally align the sub-prompts with the embedded feature (related to degradation) $\boldsymbol{\mathcal{F}}_e$. Motivated by the essence of ViT [24], we utilize the query-key mechanism, which enables the network to select the necessary sub-prompts for the input embedding. Specifically, a learnable key $\boldsymbol{\mathcal{K}}_s^i \in \mathbb{R}^{1\times D}$ is designed for each sub-prompts to calculate the correlation with embedding $\boldsymbol{\mathcal{F}}_e$ (query) for choose. The formula can be expressed as follows:

$$\boldsymbol{\mathcal{K}}_s^i \in \mathbb{R}^{1\times D} \overset{\text{match}}{\longleftrightarrow} \boldsymbol{\mathcal{P}}_s^i, \quad \boldsymbol{\mathcal{F}}_e \in \mathbb{R}^{H\times W\times D} \overset{\text{mean}}{\longrightarrow} \boldsymbol{\mathcal{F}}_e^{mean} \in \mathbb{R}^{1\times D}, \quad \delta(\boldsymbol{\mathcal{K}}_s^i, \boldsymbol{\mathcal{F}}_e), \tag{4}$$

where $\delta(\cdot)$ denotes the similarity calculation (we empirically choose cosine similarity). We employ this metric to let the network select the most appropriate sub-prompts from the pool to form effective weather-prompts.

$$\boldsymbol{\mathcal{P}}_w = \bigcup_{i=1}^{k} \boldsymbol{\mathcal{K}}_s^i \quad \text{if} \quad \delta(\boldsymbol{\mathcal{K}}_s^i, \boldsymbol{\mathcal{F}}_e) \geq \delta(\boldsymbol{\mathcal{K}}_s^{i+1}, \boldsymbol{\mathcal{F}}_e), \tag{5}$$

where k is the number of sub-prompts with the top-k similarity we selected. $\bigcup_{i=1}^{k}$ denotes the concatenation of individual sub-prompts to construct the weather-prompts $\boldsymbol{\mathcal{P}}_w$, which embodies the most representative features of the input in relation to the given weather conditions.

Such a manner can be understood as the network using sub-prompts to freely control the weather attributes that need to be learned (see Fig.3), which is novel and efficient compared to previous paradigms. This adaptive combination tailored to individual samples achieves exceptional performance, aligning with the concept of "teaching tailored to talent".

Discussion I: Recently, prompt learning has been used in image restoration [59]. Nonetheless, such an approach often relied on shared parameters to address various degradation scenarios, leading to potential interference among different degradations and overlooking the unique features of instance-wise degradations. We aim to implement a novel strategy "prompt pool". It enables the network to adaptively select appropriate sub-prompts in response to the specific degradation present in the input, thereby concentrating on the inter-attributes and intra-attributes of the weather degradations.

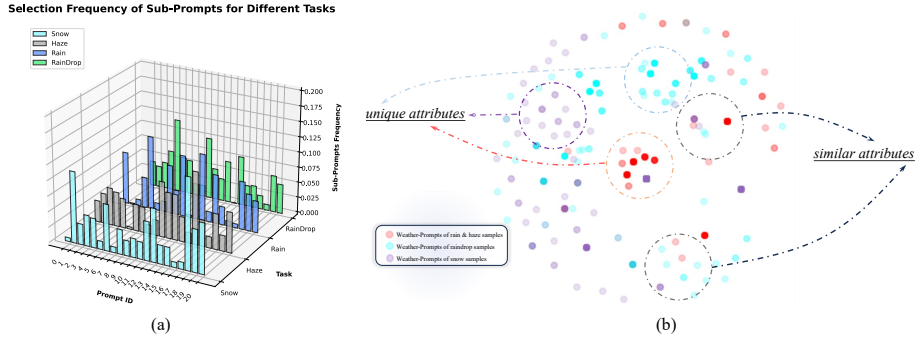

Fig. 3: (a). The selection frequency of sub-prompts. Some similar selection frequencies reflect the network's ability to adaptively exploit common attributes in some similarity between tasks (e.g. rain and raindrop). At the same time, the unique prompt frequencies highlight the flexibility to adapt to the specific characteristics of each weather condition.(b) **t-SNE visualization of weather-prompts for different weather conditions.**

3.2 General Prompts for Scene Modeling

Motivation II: Scene information provides guidance for the reconstruction of degraded residuals. In contrast to previous methods that solely focus on understanding degradations [58, 76, 102], we claim that modeling the scene content is another critical factor in enhancing performance. Inspired by this insight, we contemplate whether clean backgrounds in degraded images possess distinguishable characteristics. Utilizing t-SNE for visualization, we observe the distribution among clean images in Fig.1. There is often a significant distinction between degradations and background, while clean images share commonalities within the latent space. Consequently, we propose general prompts that encourage the network to boost representation with respect to the background.

The proposed general prompts $\boldsymbol{\mathcal{P}}_g \in \mathbb{R}^{L_g \times D}$, unlike the degradation-specific sub-prompts, are designed to encapsulate the common attributes of the scene across various weather distortions. It serves as a versatile anchor

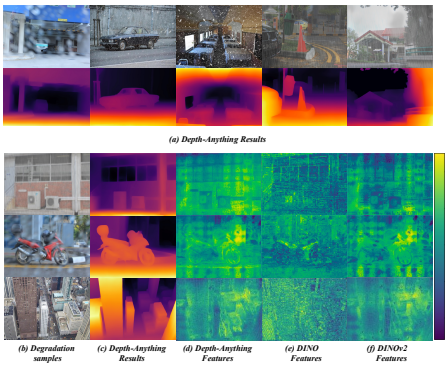

Fig. 4: Motivation of Depth-Anything [83] as a constraint. Depth-Anything has degradation-independent performance, and the intermediate features have better robustness than the previous pre-trained network [4, 56].

within the representational space, fostering a consistent perception of the background. Therefore, for the initialization of prompts, we seek to impose an explicit constraint that directs the learning towards the general attributes of the scenes. It will ensure that the general prompts are not disturbed by the varying degrees of degradations.

Observation: Inspired by scene understanding [12, 34], utilizing depth information has proven to effectively represent clean scenes. Additionally, adverse weather conditions is notably more susceptible to depth-related distortions compared to other degradations. Recently, the Depth-Anything [83] model leverages extensive datasets and the robust representational capabilities of DINOv2 [56] to develop a depth estimation model that applies to any scene. As illustrated in Fig.4 (a), we observe that depth maps estimated by Depth-Anything are almost unaffected by degradations, which to some extent demonstrates the robustness of the intermediate hidden features in scene representation (Fig.4 (b)). Motivated by this discovery, we claim that the latent space features can be used to impose an explicit constraint on the general prompts, thereby better focusing on the background portion.

Depth-Anything Constraint: To direct the general prompts $\boldsymbol{\mathcal{P}}_g \in \mathbb{R}^{L_g \times D}$ towards a more nuanced perceiving of the background, we integrate the latent features from the Depth-Anything model using an attention-based mechanism. Specifically, we define a cross-attention operation where the general prompts form the queries, and the keys and values are derived from the Depth-Anything features. Let $\boldsymbol{\mathcal{F}}_d \in \mathbb{R}^{H \times W \times D}$ represents the depth-aware features, where H and W are the dimensions of the feature map. The cross-attention mechanism is then given by:

$$\boldsymbol{\mathcal{P}}_{gd} = \text{softmax}\left(\frac{\boldsymbol{\mathcal{Q}}_g \boldsymbol{\mathcal{K}}_d^T}{\sqrt{\boldsymbol{\mathcal{D}}}}\right)\boldsymbol{\mathcal{V}}_d, \tag{6}$$

Obtained general prompts $\boldsymbol{\mathcal{P}}_{gd}$ adaptively integrate scene information, providing sufficient prior knowledge for the subsequent perception of the background while mitigating the impact of degradations.

3.3 Contrastive Prompt Loss

The sections above illustrate two types of prompts as the condition for the diffusion model. Additionally, we introduce contrastive prompt loss. It aims to differentially enhance the representations of two uniquely designed prompts. These prompts act as conditions for the diffusion model, with one tailored to model weather degradations and the other, guided by the Depth-Anything model, to perceive the background information. Given the inherently different design objectives of each prompt type, they are hypothesized to act as negative samples for each other. For the positive samples of prompt type, we employ cosine similarity to draw them nearer to the constraint within the latent space. The contrastive

prompt loss $\mathcal{L}_{cp}$ is defined as:

$$\mathcal{L}_{cp} = \frac{1}{b}\frac{1}{k}\sum_{j=1}^{b}\sum_{i=1}^{k}\left[\gamma\left(\boldsymbol{\mathcal{K}}_{gd}, \boldsymbol{\mathcal{F}}_{d}^{mean}\right) - \gamma\left(\boldsymbol{\mathcal{K}}_{s}^{i}, \boldsymbol{\mathcal{K}}_{gd}\right)\right], \tag{7}$$

where b is the batch size. $\boldsymbol{\mathcal{K}}_{gd}$ represents the learnable key matched for general prompts, $\gamma(\cdot)$ denotes the $1-\delta(\cdot)$, which ensures the optimization process.
Discussion II: While our approach draws inspiration from prior contrastive learning paradigms [20, 79, 97], it is distinctly different. 1). We do not require the construction of additional negative samples, as the two types of motivation-driven prompts within our design naturally serve as negatives for each other. 2). Prompts has explicit constraints that draw them closer to feature embeddings, eliminating the need for ground truth images as positive samples. 3). Our prompts can interact with high-dimensional features within the network directly, obviating the process for the traditional contrastive learning step of mapping via a pre-trained network [69] to a feature space.

3.4 Loss Function

To supervise our T^3-DiffWeather model, we employ the noise estimation loss Eq.1 and the contrastive prompt loss Eq.7 during the noise estimation phase. Additionally, our contrastive prompt loss is designed to optimize the prompts adapted to different instance samples. Hence, during the training, we conduct the sampling process for restoring clean images and impose additional supervision on this process through reconstruction loss and contrastive prompt loss. This approach better constrains the optimization trajectory [30, 33] of the diffusion process and releases the potential of prompts. The overall objective function can be expressed as follows:

$$\mathcal{L}_{total} = \lambda_1\mathcal{L}_{res} + \lambda_2\mathcal{L}_{cp} + \lambda_3\left\|(\boldsymbol{r}_{d}^{sample} + \boldsymbol{y}) - \boldsymbol{x}\right\|_{\text{psnr}} + \lambda_4\mathcal{L}_{cp}^{sample} \tag{8}$$

where psnr denotes the PSNR loss [11, 14] we choose empirically. λ_1, λ_2, λ_3 and λ_4 are set to 1 empirically.

4 Experiments

4.1 Implementation

Pipeline implementation. T^3-DiffWeather builds upon the backbone followed by previous diffusion design [54]. We employ uniform initialization techniques to set up the weights for sub-prompts in the prompt pool and general prompts, including their respective keys. Specifically, we designate a total of 20 sub-prompts within the prompt pool, with each sub-prompt comprising 64 tokens (L_s), from which we select the top 5 (k) to form the required weather-prompts. The token number for general prompts (L_g) is set at 256 to ensure the balance between

Table 1: Snow.

Method	Snow100K-S [49] PSNR ↑	Snow100K-S [49] SSIM ↑	Snow100K-L [49] PSNR ↑	Snow100K-L [49] SSIM ↑
SPANet[CVPR'19] [77]	29.92	0.8260	23.70	0.7930
JSTASR[ECCV'20] [18]	31.40	0.9012	25.32	0.8076
RESCAN[ECCV'18] [47]	31.51	0.9032	26.08	0.8108
DesnowNet[TIP'18] [49]	32.33	0.9500	27.17	0.8983
DDMSNet[TIP'21] [94]	34.34	0.9445	28.85	0.8772
MPRNet[CVPR'21] [93]	34.97	0.9457	29.76	0.8949
NAFNet[ECCV'22] [10]	34.79	0.9497	30.06	0.9017
Restormer[CVPR'22] [92]	**35.03**$^{\triangle}$	**0.9487**$^{\triangle}$	**30.52**$^{\triangle}$	**0.9092**$^{\triangle}$
All-in-One[CVPR'20] [46]	-	-	28.33	0.8820
TransWeather[CVPR'22] [76]	32.51	0.9341	29.31	0.8879
TKL&MR[CVPR'22] [20]	34.80	0.9483	30.24	0.9020
WeatherDiff$_{64}$[PAMI'23] [58]	35.83	0.9566	30.09	0.9041
WeatherDiff$_{128}$[PAMI'23] [58]	35.02	0.9516	29.58	0.8941
AWRCP[ICCV'23] [88]	<u>36.92</u>	<u>0.9652</u>	<u>31.92</u>	<u>0.9341</u>
★ T^3-DiffWeather (Ours)	**37.51**	**0.9664**	**32.37**	**0.9355**

Table 2: Rain & Haze.

Method	Outdoor-Rain [44] PSNR ↑	Outdoor-Rain [44] SSIM ↑
CycleGAN[ICCV'17] [100]	17.62	0.6560
pix2pix[ICCV'17] [31]	19.09	0.7100
HRGAN[CVPR'19] [45]	21.56	0.8550
PCNet[TIP'21] [36]	26.19	0.9015
MPRNet[CVPR'21] [93]	28.03	0.9192
NAFNet[ECCV'22] [10]	29.59	0.9027
Restormer[CVPR'22] [92]	**29.97**$^{\triangle}$	**0.9215**$^{\triangle}$
All-in-One[CVPR'20] [46]	24.71	0.8980
TransWeather[CVPR'22] [76]	28.83	0.9000
TKL&MR[CVPR'22] [20]	29.92	0.9167
WeatherDiff$_{64}$[PAMI'23] [58]	29.64	0.9312
WeatherDiff$_{128}$[PAMI'23] [58]	29.72	0.9216
AWRCP[ICCV'23] [88]	<u>31.39</u>	<u>0.9329</u>
★ T^3-DiffWeather (Ours)	**31.99**	**0.9365**

Table 3: Raindrop.

Method	RainDrop [60] PSNR ↑	RainDrop [60] SSIM ↑
pix2pix[ICCV'17] [31]	28.02	0.8547
DuRN[CVPR'19] [51]	31.24	0.9259
RaindropAttn[ICCV'19] [63]	31.44	0.9263
AttentiveGAN[CVPR'18] [60]	31.59	0.9170
CCN[CVPR'21] [62]	31.34	0.9286
IDT[PAMI'22] [82]	31.87	0.9313
UDR-S^{2}Former[ICCV'23] [13]	**32.64**$^{\triangle}$	**0.9427**$^{\triangle}$
All-in-One[CVPR'20] [46]	31.12	0.9268
TransWeather[CVPR'22] [76]	30.17	0.9157
TKL&MR[CVPR'22] [20]	30.99	0.9274
WeatherDiff$_{64}$[PAMI'23] [58]	30.71	0.9312
WeatherDiff$_{128}$[PAMI'23] [58]	29.66	0.9225
AWRCP[ICCV'23] [88]	<u>31.93</u>	<u>0.9314</u>
★ T^3-DiffWeather (Ours)	**32.66**	**0.9411**

Fig. 5: These tables provide quantitative comparisons with state-of-the-art image desnowing, deraining, and adverse weather removal methods, employing PSNR and SSIM as metrics—where higher values signify better restoration. The best and second-best metrics are shown with **bold text** and <u>underlined text</u>, respectively. The triangle $\triangle$ represents the SOTA metric trained on a single dataset. Above half of the tables present comparisons of task-specific methods for a single dataset, while the bottom section showcases the performance of the proposed T^3-DiffWeather method across all four test sets against state-of-the-art adverse weather solutions, including All-in-One [46], TransWeather [76], TKL&MR [20], WeatherDiffusion [58] and AWRCP [88].

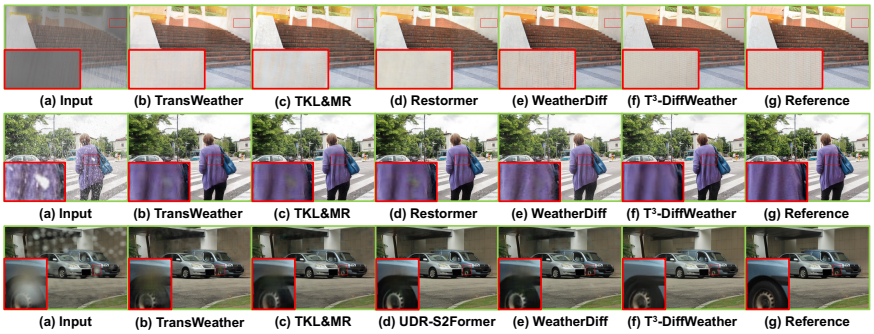

Fig. 6: Visual comparisons in adverse weather conditions on Snow100K [49], Outdoor-Rain [76] and RainDrop [60] datasets.

performance and manageable overhead. When constraining the general prompts using Depth-Anything [83], we utilize the ViT-S architecture, which demands minimal memory usage while maintaining robustness. During the diffusion process, we opt for DDIM [72] sampling. Owing to our focus on reconstructing degradation residuals and the rich representations of the condition, setting the number of sampling steps to merely 2 suffices to achieve impressive performance. Additional architectural details can be found in the supplementary materials.

Training details. To train our T^3-DiffWeather model, we leverage the comprehensive AllWeather dataset referenced in [76], including 18,069 images from the

Snow100K [49], Outdoor-Rain [44], and RainDrop [60] datasets, the same as previous adverse weather restoration methods [46, 58, 76, 88]. Our T^3-DiffWeather pipeline is developed on the PyTorch framework and undergoes training on two NVIDIA A800 GPUs. This pipeline includes 800K training iterations, utilizing the Adam optimizer with momentum parameters set to 0.9 and 0.995. Training commences with an initial learning rate of 1.5×10^{-4}, which is reduced using a cosine annealing schedule. To promote stability during the learning phase, an exponential moving average strategy weighted at 0.995 is employed for parameter updates, as supported by findings in [55] and [73]. The diffusion procedure consists of $1,000$ timesteps, labeled as T, with an incrementally ascending noise schedule β_t ranging from 0.0001 to 0.02. The training employs image patches of 256×256 pixels. Augmentation techniques like horizontal flipping and fixed-angle random rotation are used in the training. Please refer to the supplementary material to view detailed training and testing dataset configurations.

4.2 Quantitative comparison

We perform a comparative analysis of metrics between synthetic and real datasets. Specifically, we compare the model performance for a single task and the performance of a multi-weather image restoration model trained on multiple weather datasets. Our quantitative analysis reveals the competitive advantage of our T^3-DiffWeather pipeline over existing state-of-the-art algorithms in image restoration with various weather impacts. As shown in Tab.1, T^3-DiffWeather achieves excellent performance in image snow removal, as evidenced by the highest PSNR and SSIM metrics on the Snow100K-S [49] and Snow100K-L [49] datasets. It is particularly noteworthy that the PSNR on Snow100K-S is 1.68db higher than the previous best diffusion model, WeatherDiffusion [58], indicating a significant improvement in recovery quality. This is mainly due to our new pipeline design and suitable and effective conditions. In addition, our method ranks first in the deraining and dehazing task (Tab.2) and maintains the leading position in the raindrop removal (Tab.3).

4.3 Visual Comparison

Fig.6 visually compares state-of-the-art image restoration techniques on a synthetic dataset designed to simulate real-world conditions. WeatherDiffusion [58] marginally enhances detail definition but does not remove degradations entirely in some areas. Restormer improves color fidelity but does not entirely eliminate synthetic artifacts. T^3-DiffWeather markedly improves texture and color accuracy, closely matching the reference. It significantly reduces synthetic distortions, maintaining scene authenticity, as seen in the detailed insets.

Furthermore, Fig.7 also shows a visual comparison of restoration methods applied to images of real-world scenarios. Also based on a diffusion model, our method can better remove degradation in the real world and restore complex scene textures than WeatherDiffusion [58]. In addition, UDR-S2Former [13] has deficiencies in handling real rainy scenes. In comparison, our method visually

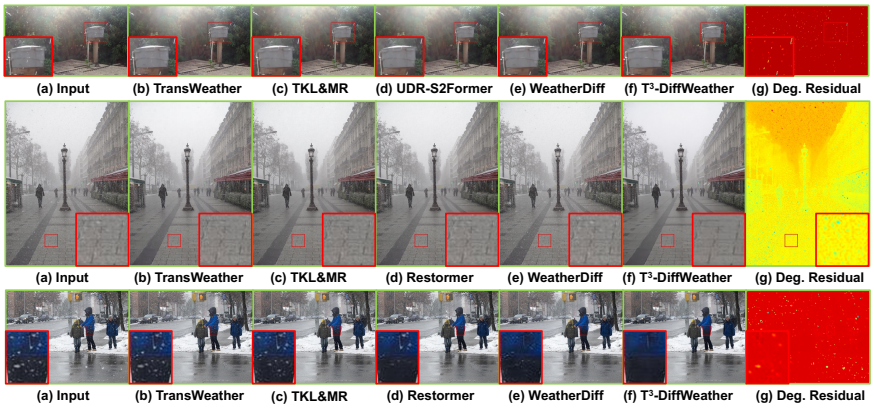

Fig. 7: Visual comparisons of the real-world adverse weather samples.

removed all degradations details, which proves the competitiveness of our method compared to specific methods. We also show the heat map of our degradation residual. We find that our method always focuses on degradations in terms of the restoraiton object, which proves the effectiveness of our pipeline.

4.4 Comparison of Parameters and Computational Complexity

Table 4: Com. of parameters and GFLOPs (256×256 resolution) for diffusion process.

Method	#Params	#GFLOPs
Single Image Restoration		
IR-SDE [53]	135.3M	119.1G×100 steps
Refusion [54]	131.4M	63.4G×50 steps
Adverse Weather Restoration		
WeatherDiffusion [58]	113.68M	248.4G×25 steps
T^3-DiffWeather (ours)	69.38M	59.82G×2 steps

As shown in Tab.4, our pipeline significantly reduces the number of parameters required for diffusion compared to previous designs. Moreover, with only two steps in the sampling process, the computational complexity at a 256×256 resolution is a mere **1/52** of that of the SOTA WeatherDiffusion [58]. Additionally, the computational complexity of the single image restoration diffusion architecture Refusion [54] is nine times of ours, underscoring the efficacy of proposed conditions and constraints and the superiority of our holistic approach aimed at reconstructing degradation residuals.

4.5 Ablation Studies

In order to verify the efficacy of each key component of T^3-DiffWeather, we conduct a series of ablation experiments. All these variants are trained using the same configurations as in the implementation details, and the ablation results are tested on Outdoor-Rain [44].

Effectiveness of Prompt Pool. The Tab.5 highlight the vital role of the proposed prompt pool in adverse weather restoration. Leveraging the advances in prompt learning, our prompt pool method autonomously crafts tailored sub-prompts for each specific degradation scenario. This approach has shown substantial improvements over methods without a prompt pool or with unmatched keys and significantly surpasses the previous design [59]. With our method, the network selects the most representative sub-prompts with 64 token numbers of sub-prompts to balance the complexity and performance, affirming the prompt pool's utility in capturing the attributes of complex weather patterns.

Table 5: Abl. of Prompt Pool.

Method	**PSNR** ↑	**SSIM** ↑
w/o. prompt pool	31.05	0.9325
w/o. matched keys	31.72	0.9349
w. prompt [59]	31.38	0.9330
w. prompt pool (Ours)	**31.99**	**0.9365**
Length L_s	**PSNR** ↑	**SSIM** ↑
32 (Sub-prompts)	31.79	0.9358
64 (Sub-prompts) (Ours)	31.99	0.9365
128 (Sub-prompts)	**32.04**	**0.9366**

Improvements of proposed General Prompts. Tab.6 illustrates the efficacy of incorporating the robust Depth-Anything [83] features as a constraint for general prompts. The depth-aware features intrinsic to Depth-Anything have demonstrated superior performance over other pretrained models (e.g. DINO [4], DINOv2 [56]), particularly regarding scene understanding. Moreover, general prompts without such explicit constraints exhibit limitations in holistically background modeling.

Table 6: Abl. of General Prompts.

Method	**PSNR** ↑	**SSIM** ↑
w/o. General Prompts	31.52	0.9342
w/o. Depth-Anything [83]	31.67	0.9349
w. DINO [4]	31.77	0.9357
w. DINOv2 [56]	31.82	0.9359
w. Depth-Anything [83] (Ours)	**31.99**	**0.9365**
Length L_g	**PSNR** ↑	**SSIM** ↑
128 (General Prompts)	31.73	0.9354
192 (General Prompts)	31.88	0.9360
256 (General Prompts) (Ours)	31.99	**0.9365**
320 (General Prompts)	**32.03**	0.9365

Furthermore, our experiments revealed a performance bottleneck associated with increasing the number of general prompts. To optimize efficiency without compromising gains, we determine that selecting 256 tokens yields the optimal balance, effectively avoiding unnecessary computational overhead. More ablation experiments can be found in the supplementary material.

5 Conclusion

This paper draws inspiration from the prompt learning and the concept of "Teaching Tailored to Talent", proposing a novel T^3-DiffWeather. It utilizes weather-prompts constructed from free combinations of sub-prompts, and general prompts constrained by Depth-Anything to provide rich information for the diffusion process from the degradation and background perspectives. Experimental results demonstrate that our method achieves SOTA performance on various synthetic and real-world data sets, with excellent computational efficiency.

Acknowledgements

This work is supported by the Guangzhou-HKUST(GZ) Joint Funding Program (No. 2023A03J0671), the National Natural Science Foundation of China (Grant No. 61902275), the Guangzhou Industrial Information and Intelligent Key Laboratory Project (No. 2024A03J0628), and Guangzhou-HKUST(GZ) Joint Funding Program (No. 2024A03J0618).

References

1. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: An end-to-end system for single image haze removal. IEEE Transactions on Image Processing **25**(11), 5187–5198 (2016)
2. Cao, S., Chai, W., Hao, S., Wang, G.: Image reference-guided fashion design with structure-aware transfer by diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3524–3528 (2023)
3. Cao, S., Chai, W., Hao, S., Zhang, Y., Chen, H., Wang, G.: Difffashion: Reference-based fashion design with structure-aware transfer by diffusion models. IEEE Transactions on Multimedia (2023)
4. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 9650–9660 (2021)
5. Chai, W., Guo, X., Wang, G., Lu, Y.: Stablevideo: Text-driven consistency-aware diffusion video editing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 23040–23050 (2023)
6. Chai, W., Wang, G.: Deep vision multimodal learning: Methodology, benchmark, and trend. Applied Sciences **12**(13), 6588 (2022)
7. Chen, H., Gu, J., Liu, Y., Magid, S.A., Dong, C., Wang, Q., Pfister, H., Zhu, L.: Masked image training for generalizable deep image denoising. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1692–1703 (2023)
8. Chen, H., Li, W., Gu, J., Ren, J., Sun, H., Zou, X., Zhang, Z., Yan, Y., Zhu, L.: Low-res leads the way: Improving generalization for super-resolution by self-supervised learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 25857–25867 (2024)
9. Chen, H., Ren, J., Gu, J., Wu, H., Lu, X., Cai, H., Zhu, L.: Snow removal in video: A new dataset and a novel method. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 13165–13176. IEEE (2023)
10. Chen, L., Chu, X., Zhang, X., Sun, J.: Simple baselines for image restoration. In: European Conference on Computer Vision. pp. 17–33. Springer (2022)
11. Chen, L., Lu, X., Zhang, J., Chu, X., Chen, C.: Hinet: Half instance normalization network for image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 182–192 (2021)
12. Chen, P.Y., Liu, A.H., Liu, Y.C., Wang, Y.C.F.: Towards scene understanding: Unsupervised monocular depth estimation with semantic-aware representation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)

13. Chen, S., Ye, T., Bai, J., Chen, E., Shi, J., Zhu, L.: Sparse sampling transformer with uncertainty-driven ranking for unified removal of raindrops and rain streaks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13106–13117 (2023)
14. Chen, S., Ye, T., Liu, Y., Chen, E.: Dual-former: Hybrid self-attention transformer for efficient image restoration. arXiv preprint arXiv:2210.01069 (2022)
15. Chen, S., Ye, T., Liu, Y., Chen, E., Shi, J., Zhou, J.: Snowformer: Scale-aware transformer via context interaction for single image desnowing. arXiv preprint arXiv:2208.09703 (2022)
16. Chen, S., Ye, T., Liu, Y., Liao, T., Ye, Y., Chen, E.: Msp-former: Multi-scale projection transformer for single image desnowing. arXiv preprint arXiv:2207.05621 (2022)
17. Chen, S., Ye, T., Xue, C., Chen, H., Liu, Y., Chen, E., Zhu, L.: Uncertainty-driven dynamic degradation perceiving and background modeling for efficient single image desnowing. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 4269–4280 (2023)
18. Chen, W.T., Fang, H.Y., Ding, J.J., Tsai, C.C., Kuo, S.Y.: Jstasr: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16. pp. 754–770. Springer (2020)
19. Chen, W.T., Fang, H.Y., Hsieh, C.L., Tsai, C.C., Chen, I., Ding, J.J., Kuo, S.Y., et al.: All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4196–4205 (2021)
20. Chen, W.T., Huang, Z.K., Tsai, C.C., Yang, H.H., Ding, J.J., Kuo, S.Y.: Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17653–17662 (2022)
21. Chen, X., Li, H., Li, M., Pan, J.: Learning a sparse transformer network for effective image deraining. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5896–5905 (2023)
22. Chen, Z., Zhang, Y., Gu, J., Yuan, X., Kong, L., Chen, G., Yang, X.: Image super-resolution with text prompt diffusion. arXiv preprint arXiv:2311.14282 (2023)
23. Cheng, B., Li, J., Chen, Y., Zhang, S., Zeng, T.: Snow mask guided adaptive residual network for image snow removal. arXiv preprint arXiv:2207.04754 (2022)
24. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
25. Fang, Y., Wang, Z., Zhang, L., Cao, J., Chen, H., Xu, R.: Spiking wavelet transformer. arXiv preprint arXiv:2403.11138 (2024)
26. Guo, J., Chai, W., Deng, J., Huang, H.W., Ye, T., Xu, Y., Zhang, J., Hwang, J.N., Wang, G.: Versat2i: Improving text-to-image models with versatile reward. arXiv preprint arXiv:2403.18493 (2024)
27. Guo, L., Wang, C., Yang, W., Huang, S., Wang, Y., Pfister, H., Wen, B.: Shadowdiffusion: When degradation prior meets diffusion model for shadow removal. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14049–14058 (2023)

28. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
29. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
30. Hou, J., Zhu, Z., Hou, J., Liu, H., Zeng, H., Yuan, H.: Global structure-aware diffusion process for low-light image enhancement. Advances in Neural Information Processing Systems **36** (2024)
31. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017)
32. Jia, M., Tang, L., Chen, B.C., Cardie, C., Belongie, S., Hariharan, B., Lim, S.N.: Visual prompt tuning. In: European Conference on Computer Vision. pp. 709–727. Springer (2022)
33. Jiang, H., Luo, A., Fan, H., Han, S., Liu, S.: Low-light image enhancement with wavelet-based diffusion models. ACM Transactions on Graphics (TOG) **42**(6), 1–14 (2023)
34. Jiang, H., Larsson, G., Shakhnarovich, M.M.G., Learned-Miller, E.: Self-supervised relative depth learning for urban scene understanding. In: Proceedings of the european conference on computer vision (eccv). pp. 19–35 (2018)
35. Jiang, J., Ye, T., Bai, J., Chen, S., Chai, W., Jun, S., Liu, Y., Chen, E.: Five a Θ{+} network: You only need 9k parameters for underwater image enhancement. arXiv preprint arXiv:2305.08824 (2023)
36. Jiang, K., Wang, Z., Yi, P., Chen, C., Wang, Z., Wang, X., Jiang, J., Lin, C.W.: Rain-free and residue hand-in-hand: A progressive coupled network for real-time image deraining. IEEE Transactions on Image Processing **30**, 7404–7418 (2021)
37. Jiang, Z., Zhou, Z., Li, L., Chai, W., Yang, C.Y., Hwang, J.N.: Back to optimization: Diffusion-based zero-shot 3d human pose estimation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 6142–6152 (2024)
38. Jin, Y., Lin, B., Yan, W., Ye, W., Yuan, Y., Tan, R.T.: Enhancing visibility in nighttime haze images using guided apsf and gradient adaptive convolution (2023)
39. Jin, Y., Sharma, A., Tan, R.T.: Dc-shadownet: Single-image hard and soft shadow removal using unsupervised domain-classifier guided network. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5027–5036 (2021)
40. Jin, Y., Yan, W., Yang, W., Tan, R.T.: Structure representation network and uncertainty feedback learning for dense non-uniform fog removal. In: Asian Conference on Computer Vision. pp. 155–172. Springer (2022)
41. Jin, Y., Yang, W., Tan, R.T.: Unsupervised night image enhancement: When layer decomposition meets light-effects suppression. In: European Conference on Computer Vision. pp. 404–421. Springer (2022)
42. Jin, Y., Ye, W., Yang, W., Yuan, Y., Tan, R.T.: Des3: Adaptive attention-driven self and soft shadow removal using vit similarity. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 2634–2642 (2024)
43. Li, H., Yang, Y., Chang, M., Chen, S., Feng, H., Xu, Z., Li, Q., Chen, Y.: Srdiff: Single image super-resolution with diffusion probabilistic models. Neurocomputing **479**, 47–59 (2022)
44. Li, R., Cheong, L.F., Tan, R.T.: Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)

45. Li, R., Cheong, L.F., Tan, R.T.: Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1633–1642 (2019)
46. Li, R., Tan, R.T., Cheong, L.F.: All in one bad weather removal using architectural search. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3175–3185 (2020)
47. Li, X., Wu, J., Lin, Z., Liu, H., Zha, H.: Recurrent squeeze-and-excitation context aggregation net for single image deraining. In: Proceedings of the European conference on computer vision (ECCV). pp. 254–269 (2018)
48. Lin, X., He, J., Chen, Z., Lyu, Z., Fei, B., Dai, B., Ouyang, W., Qiao, Y., Dong, C.: Diffbir: Towards blind image restoration with generative diffusion prior. arXiv preprint arXiv:2308.15070 (2023)
49. Liu: Desnownet: Context-aware deep network for snow removal. IEEE TIP (2018)
50. Liu, X., Ma, Y., Shi, Z., Chen, J.: Griddehazenet: Attention-based multi-scale network for image dehazing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7314–7323 (2019)
51. Liu, X., Suganuma, M., Sun, Z., Okatani, T.: Dual residual networks leveraging the potential of paired operations for image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7007–7016 (2019)
52. Liu, Y., Yan, Z., Chen, S., Ye, T., Ren, W., Chen, E.: Nighthazeformer: Single nighttime haze removal using prior query transformer. arXiv preprint arXiv:2305.09533 (2023)
53. Luo, Z., Gustafsson, F.K., Zhao, Z., Sjölund, J., Schön, T.B.: Image restoration with mean-reverting stochastic differential equations. arXiv preprint arXiv:2301.11699 (2023)
54. Luo, Z., Gustafsson, F.K., Zhao, Z., Sjölund, J., Schön, T.B.: Refusion: Enabling large-size realistic image restoration with latent-space diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1680–1691 (2023)
55. Nichol, A.Q., Dhariwal, P.: Improved denoising diffusion probabilistic models. In: International Conference on Machine Learning. pp. 8162–8171. PMLR (2021)
56. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al.: Dinov2: Learning robust visual features without supervision. arXiv preprint arXiv:2304.07193 (2023)
57. Ouyang, Y., Chai, W., Ye, J., Tao, D., Zhan, Y., Wang, G.: Chasing consistency in text-to-3d generation from a single image. arXiv preprint arXiv:2309.03599 (2023)
58. Özdenizci, O., Legenstein, R.: Restoring vision in adverse weather conditions with patch-based denoising diffusion models. IEEE Transactions on Pattern Analysis and Machine Intelligence (2023)
59. Potlapalli, V., Zamir, S.W., Khan, S., Khan, F.S.: Promptir: Prompting for all-in-one blind image restoration. arXiv preprint arXiv:2306.13090 (2023)
60. Qian, R., Tan, R.T., Yang, W., Su, J., Liu, J.: Attentive generative adversarial network for raindrop removal from a single image. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2482–2491 (2018)
61. Qiu, Y., Zhang, K., Wang, C., Luo, W., Li, H., Jin, Z.: Mb-taylorformer: Multi-branch efficient transformer expanded by taylor formula for image dehazing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12802–12813 (2023)

62. Quan, R., Yu, X., Liang, Y., Yang, Y.: Removing raindrops and rain streaks in one go. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9147–9156 (2021)
63. Quan, Y., Deng, S., Chen, Y., Ji, H.: Deep learning for seeing through window with raindrops. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2463–2471 (2019)
64. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PMLR (2021)
65. Ren, D., Zuo, W., Hu, Q., Zhu, P., Meng, D.: Progressive image deraining networks: A better and simpler baseline. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3937–3946 (2019)
66. Ren, H., Zhou, Y., Zhu, J., Fu, H., Huang, Y., Lin, X., Fang, Y., Ma, F., Yu, H., Cheng, B.: Rethinking efficient and effective point-based networks for event camera classification and regression: Eventmamba. arXiv preprint arXiv:2405.06116 (2024)
67. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
68. Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M.: Image super-resolution via iterative refinement. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(4), 4713–4726 (2022)
69. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
70. Song, E., Chai, W., Wang, G., Zhang, Y., Zhou, H., Wu, F., Chi, H., Guo, X., Ye, T., Zhang, Y., et al.: Moviechat: From dense token to sparse memory for long video understanding. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18221–18232 (2024)
71. Song, E., Chai, W., Ye, T., Hwang, J.N., Li, X., Wang, G.: Moviechat+: Question-aware sparse memory for long video question answering. arXiv preprint arXiv:2404.17176 (2024)
72. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
73. Song, Y., Ermon, S.: Improved techniques for training score-based generative models. Advances in neural information processing systems **33**, 12438–12448 (2020)
74. Song, Y., He, Z., Qian, H., Du, X.: Vision transformers for single image dehazing. arXiv preprint arXiv:2204.03883 (2022)
75. Sun, H., Li, W., Liu, J., Chen, H., Pei, R., Zou, X., Yan, Y., Yang, Y.: Coser: Bridging image and language for cognitive super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 25868–25878 (2024)
76. Valanarasu, J.M.J., Yasarla, R., Patel, V.M.: Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2353–2363 (2022)
77. Wang, T., Yang, X., Xu, K., Chen, S., Zhang, Q., Lau, R.W.: Spatial attentive single-image deraining with a high quality real rain dataset. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12270–12279 (2019)

78. Wang, Z., Fang, Y., Cao, J., Zhang, Q., Wang, Z., Xu, R.: Masked spiking transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1761–1771 (2023)
79. Wu, H., Qu, Y., Lin, S., Zhou, J., Qiao, R., Zhang, Z., Xie, Y., Ma, L.: Contrastive learning for compact single image dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10551–10560 (2021)
80. Wu, H., Yang, Y., Chen, H., Ren, J., Zhu, L.: Mask-guided progressive network for joint raindrop and rain streak removal in videos. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 7216–7225 (2023)
81. Xia, B., Zhang, Y., Wang, S., Wang, Y., Wu, X., Tian, Y., Yang, W., Van Gool, L.: Diffir: Efficient diffusion model for image restoration. arXiv preprint arXiv:2303.09472 (2023)
82. Xiao, J., Fu, X., Liu, A., Wu, F., Zha, Z.J.: Image de-raining transformer. IEEE Transactions on Pattern Analysis and Machine Intelligence (2022)
83. Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., Zhao, H.: Depth anything: Unleashing the power of large-scale unlabeled data. arXiv preprint arXiv:2401.10891 (2024)
84. Yang, T., Ren, P., Xie, X., Zhang, L.: Pixel-aware stable diffusion for realistic image super-resolution and personalized stylization. arXiv preprint arXiv:2308.14469 (2023)
85. Yang, W., Tan, R.T., Feng, J., Liu, J., Guo, Z., Yan, S.: Deep joint rain detection and removal from a single image. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1357–1366 (2017)
86. Yang, Y., Aviles-Rivero, A.I., Fu, H., Liu, Y., Wang, W., Zhu, L.: Video adverse-weather-component suppression network via weather messenger and adversarial backpropagation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13200–13210 (2023)
87. Yang, Y., Wu, H., Aviles-Rivero, A.I., Zhang, Y., Qin, J., Zhu, L.: Genuine knowledge from practice: Diffusion test-time adaptation for video adverse weather removal. arXiv preprint arXiv:2403.07684 (2024)
88. Ye, T., Chen, S., Bai, J., Shi, J., Xue, C., Jiang, J., Yin, J., Chen, E., Liu, Y.: Adverse weather removal with codebook priors. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12653–12664 (2023)
89. Ye, T., Chen, S., Chai, W., Xing, Z., Qin, J., Lin, G., Zhu, L.: Learning diffusion texture priors for image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2524–2534 (2024)
90. Ye, T., Chen, S., Liu, Y., Ye, Y., Bai, J., Chen, E.: Towards real-time high-definition image snow removal: Efficient pyramid network with asymmetrical encoder-decoder architecture. In: Proceedings of the Asian Conference on Computer Vision. pp. 366–381 (2022)
91. Ye, T., Zhang, Y., Jiang, M., Chen, L., Liu, Y., Chen, S., Chen, E.: Perceiving and modeling density for image dehazing. In: European Conference on Computer Vision. pp. 130–145. Springer (2022)
92. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5728–5739 (2022)
93. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14821–14831 (2021)

94. Zhang, K., Li, R., Yu, Y., Luo, W., Li, C.: Deep dense multi-scale network for snow removal using semantic and depth priors. IEEE Transactions on Image Processing **30**, 7419–7431 (2021)
95. Zhang, R., Gu, J., Chen, H., Dong, C., Zhang, Y., Yang, W.: Crafting training degradation distribution for the accuracy-generalization trade-off in real-world super-resolution. In: International conference on machine learning. pp. 41078–41091. PMLR (2023)
96. Zhao, Z., Chai, W., Wang, X., Boyi, L., Hao, S., Cao, S., Ye, T., Hwang, J.N., Wang, G.: See and think: Embodied agent in virtual environment. arXiv preprint arXiv:2311.15209 (2023)
97. Zheng, Y., Zhan, J., He, S., Dong, J., Du, Y.: Curricular contrastive regularization for physics-aware single image dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5785–5794 (June 2023)
98. Zhou, K., Yang, J., Loy, C.C., Liu, Z.: Conditional prompt learning for vision-language models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 16816–16825 (June 2022)
99. Zhou, K., Yang, J., Loy, C.C., Liu, Z.: Learning to prompt for vision-language models. International Journal of Computer Vision **130**(9), 2337–2348 (2022)
100. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)
101. Zhu, L., Deng, Z., Hu, X., Fu, C.W., Xu, X., Qin, J., Heng, P.A.: Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 121–136 (2018)
102. Zhu, Y., Wang, T., Fu, X., Yang, X., Guo, X., Dai, J., Qiao, Y., Hu, X.: Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21747–21758 (2023)