IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 21 November 2024, accepted 11 December 2024, date of publication 20 December 2024,
date of current version 31 December 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3520484

RESEARCH ARTICLE

Multi-Task Image Restoration Algorithm Under Different Weather Influence Factors

LINGWEN MENG, MINGYONG XIN, JUNWEI ZHANG, FU ZOU, AND QINGQING ZHAO
Electric Power Research Institute of Guizhou Power Grid Co. Ltd., Guiyang 550002, China
Corresponding author: Lingwen Meng (menglingwen_gpg@163.com)

This work was supported by Guizhou Power Grid Co. Ltd., under Grant GZKJXM20222320.

ABSTRACT Bad weather, such as rain, snow, and fog, will reduce the quality of the image acquired, and it will affect the performance of many related visual fields. The existing researches on image restoration under severe weather either focus on the restoration under a certain kind of weather, which cannot be generalized under different weather, or need to introduce additional model structures, which increases the burden of practical applications. Therefore, we explores an integrated image restoration framework, which can restore images under different adverse weather effects. Specifically, the model is trained in two stages. In the pre-training stage, a basic general network that can deal with multiple weather is trained by supervised learning. In the fine-tuning stage, soft prompts are introduced to stimulate the model's ability to cope with different weather, so as to enhance the generalization ability. Specifically, the hidden information of the prompts is explored by low-rank decomposition, and the contras loss is added to prompt the prompts to converge on similar tasks. Furthermore, we address the distribution shift problem by aligning out-of-distribution (OOD) test sample statistics with those of the source data using test-time prompt tuning. Finally, we use the evaluation metrics, PSNR and SSIM, to evaluate the proposed method under four tasks: rain removal, snow removal, fog removal, and raindrop removal. The results demonstrate that the proposed method achieves superior performance compared to the existing state-of-the-art.

INDEX TERMS Transformer architecture, pretrain-finetuning, soft prompts, contrast learning, distribution alignment strategy.

I. INTRODUCTION

With the continuous development of society, image acquisition has become integral to various fields, such as satellite remote sensing and autonomous driving. One of the key challenges in advancing these fields lies in ensuring the quality of collected images. Severe weather conditions, including rain, snow, and fog, often adversely affect the image acquisition process, resulting in degraded image quality and impairing the performance of many vision-based applications. The degradation effects caused by different weather conditions are distinct. For example, raindrops create visible streaks and spots on images, obscuring the line of sight and interfering with image recognition tasks. Snow typically appears as bright spots or occlusions, reducing image clarity and contrast. Fog, on the other hand, scatters light, decreasing visibility and causing the loss of fine details.To address these issues, researchers have conducted extensive studies to mitigate the impact of adverse weather conditions on degraded images. These include rain removal methods [1], fog removal methods [2], snow removal methods [3], and raindrop removal methods [4], [5].

Although these methods have demonstrated promising results, they are typically designed to target specific weather conditions and lack the generalizability needed to handle multiple types of weather degradation. In practical applications, it is often necessary to address diverse weather conditions simultaneously. To meet this demand, researchers have proposed unified approaches that utilize a single set of network parameters to manage different types of degradation

The associate editor coordinating the review of this manuscript and approving it for publication was Andrea Bottino.

[6], [7], [8].However, existing unified methods face notable limitations. For instance, [6], [9] fail to account for the unique characteristics of each weather condition, resulting in suboptimal performance when addressing certain types of degradation. Additionally, methods such as [10] require customized model components, such as encoders, for each weather condition, which imposes a significant burden in practical implementations. Furthermore, approaches like [6], [8] involve optimizing task-specific model parameters by tuning task-general parameters during training. However, training from scratch in this manner may lead to conflicts between parameters, ultimately degrading performance.

Real-world image degradation often changes dynamically, with additional factors such as device variations contributing to substantial differences between training and test data distributions. Such unseen distribution changes during inference render the model sensitive to these shifts, leading to performance degradation [11], [12].

To effectively address the aforementioned challenges, we propose a two-stage image restoration training strategy combined with distribution alignment at test time to mitigate distribution shift issues. Specifically, in the pre-training stage, we leverage a diverse dataset of weather-degraded images for multi-task supervised learning, thereby training a general-purpose image restoration model. However, multi-task training can suffer from task interference, where conflicting objectives across tasks hinder the overall learning capacity of the model [13]. To overcome this limitation, we enhance the model's capability to handle specific weather degradation by introducing task-specific soft prompts during the second stage of training. These soft prompts encode discriminative information unique to each degradation type, enabling the model to adjust its attention weights based on the target task.While the visual manifestations of different weather conditions vary, there exist shared patterns and characteristics underlying the degradation processes, alongside task-specific features [7]. For instance, [14] achieved multi-task processing by exploring intrinsic connections among different image types, optimizing the network's feature extraction and association allocation to enhance task differentiation. Similarly, [15] proposed a Lifelong CycleGAN model for continual multi-task image restoration, enabling continuous learning of new tasks without compromising previously acquired knowledge. Building upon these approaches, we explore both the implicit and explicit interactions between soft prompts during fine-tuning, thereby enhancing task-specific model performance.To capture the implicit interactions among various weather conditions, the prompts undergo low-rank decomposition, yielding two distinct components: generic task prompts and task-specific prompts. Generic task prompts are shared across all tasks and represent common features, while task-specific prompts capture unique features tailored to each type of weather degradation. To further explore explicit interactions, we employ t-SNE analysis [16] to visualize relationships among different weather-degraded images. This insight guides the integration of contrastive learning to improve the efficiency of soft prompts.Finally, we address distribution alignment explicitly during test time. For each test input, we generate randomly augmented views that are passed through the model to compute token embedding statistics. These embeddings' mean and variance are then aligned between the source dataset and the test sample by updating learnable prompts. This alignment ensures robust performance in the presence of distribution shifts.

In summary, the main contributions of this paper are as follows:

- We propose a two-stage image restoration framework capable of effectively removing different types of adverse weather (*i.e.*, rain, snow, fog, raindrops) from degraded images.
- We explore the implicit and explicit interactions between soft prompts during the fine-tuning stage to better train models on specific tasks.
- We align the test data with the source data distribution by adjusting the learnable prompts at test time, solving the problem of effectively adapting the model when the distribution shift occurs.
- We verified the effectiveness of the proposed method. At the same time, our method has achieved relatively advanced results in a variety of image restoration tasks.

II. RELATED WORK

A. SINGLE-WEATHER DEGRADED IMAGE RESTORATION

The restoration of images degraded by specific weather conditions has been extensively studied. For instance, in rain removal research, the GAN-based model proposed in [17] effectively removes rain streaks from images while preserving image details and quality by learning the mapping relationship between rain-degraded and rain-free images. Similarly, in snow removal studies, Deep Convolutional Neural Networks (DCNNs) have demonstrated the ability to accurately distinguish snowflakes from background information, enabling precise snowflake removal. In dehazing research, the dark channel prior method introduced in [18] effectively eliminates haze by removing low-intensity information elements and mitigating block artifacts.For raindrop removal, the study in [19] proposed a network that learns optimized filters to minimize the impact of raindrops on image details, achieving the restoration of clear images. Additionally, other approaches have achieved remarkable results, including the rain removal method in [1], the snow removal method in [3], the fog removal method in [2], and the raindrop removal method in [4].Although these methods exhibit excellent performance, they are not universal solutions capable of addressing all weather-induced degradations. This limitation arises because these approaches are task-specific, with separate models trained for each weather condition. Furthermore, for unknown weather degradation

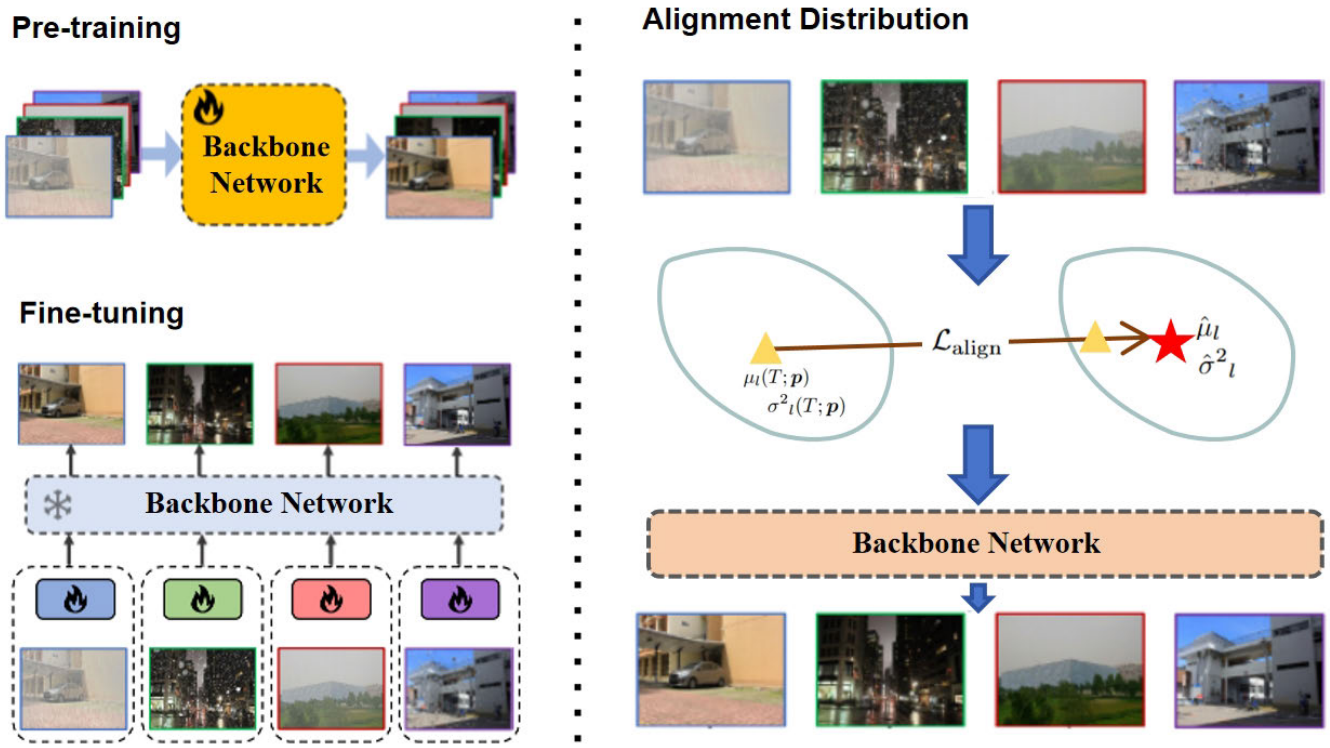

FIGURE 1. The overall of our approach. Our approach is divided into three parts: two-stage training with pre-training and fine-tuning and distribution alignment performed at test time. We pre-train the model in a supervised manner and then fine-tune the model using interactive augmentation prompts while the model parameters are frozen. Finally, the distribution of test data is aligned according to the distribution of training data at test time.

types, these methods often require additional fine-tuning or multiple sets of weather-specific model parameters, which can limit their practical applicability.

B. MULTI-WEATHER DEGRADED IMAGE RESTORATION

Unlike single-weather degraded image restoration tasks, multi-weather degraded image restoration aims to recover clear images under various weather conditions using a unified set of network parameters. Current research methods for all-in-one weather image restoration have shown promising results. These approaches typically involve either designing distinct networks for each weather removal task or employing frameworks where each task has a dedicated encoder while sharing a common decoder across all tasks. For example, in [10], an all-in-one network with multiple encoders was proposed to address severe weather degradations through network architecture search. TransWeather [6] employs a transformer-based end-to-end model that incorporates learnable weather-type embeddings to decode weather degradations. Additionally, [9] introduced a two-stage knowledge learning mechanism to handle multiple severe weather removal tasks. PromptIR [8] proposed an integrated prompt-based learning framework for image restoration, utilizing prompts to encode weather-specific degradation information.

However, the above methods exhibit notable limitations. Approaches such as [10] and [7] rely on task-specific model structures for specific weather scenarios, which can increase computational and practical burdens. On the other hand, [9] and [8] train a general image restoration model while incorporating additional learnable embeddings for task-specific learning. Nevertheless, these methods jointly optimize weather-general and weather-specific parameters from scratch, which often results in parameter conflicts that degrade overall performance.To address these issues, we build upon the methodologies of [20] and [2] by adopting a two-stage training strategy. In the first stage, a general image restoration model is learned to capture shared features across weather conditions. In the second stage, weather-specific prompts are trained while freezing the parameters of the general model. This approach effectively alleviates parameter conflicts, enhancing the model's ability to handle specific weather degradations.

C. PROMPT LEARNING

Prompt learning has emerged as a prominent technique in the field of natural language processing (NLP). According to [21], instead of relying solely on task-specific parameters, prompt learning enhances model performance by utilizing prompts to provide contextual information, thereby fine-tuning the model for a given target task. Rather than retraining the model with traditional parameters for each task, predefined prompts guide the model's behavior, enabling efficient task adaptation. Prompts have

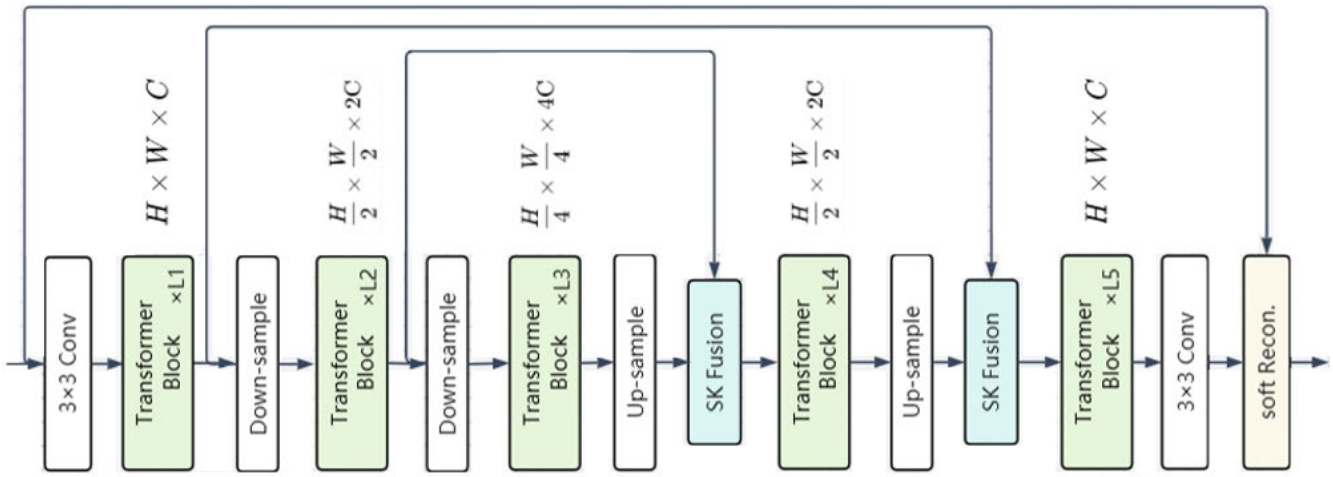

FIGURE 2. The main architecture of the image inpainting model. We applied a 5-stage U-Net, replacing its convolutional block with a Transformerblock.

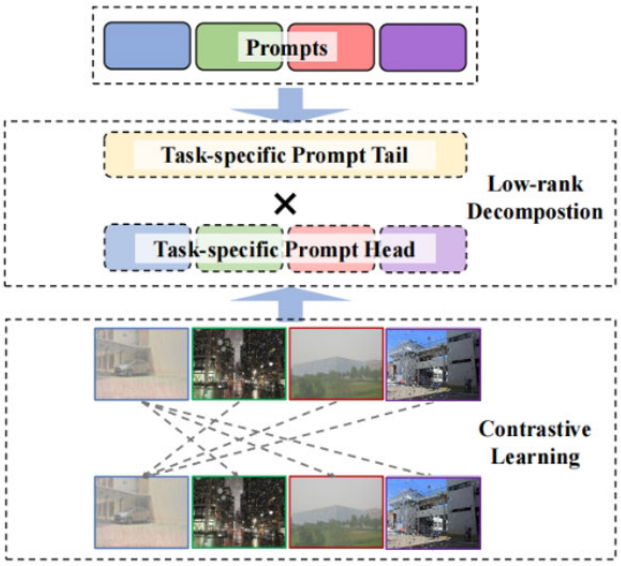

FIGURE 3. Interaction Enchanced prompts. Prompts can be decomposed into a prompt tail shared by all tasks and a prompt head specific to each task, labeled as implicit interaction enhancement. In addition, we constrain prompt heads by contrastive learning based on inter-task relations, which is labeled as explicit interaction enhancement.

been successfully applied to fine-tuning vision tasks [22], natural language generation [23], and speech synthesis [24], demonstrating their ability to facilitate efficient parameter adaptation.

To address the limitations of manually crafted prompts in early prompt learning, researchers introduced soft prompts [25], which are optimized in a continuous space. Compared to hard prompts, soft prompts offer greater flexibility and adaptability, allowing models to better align with the requirements of specific tasks. In the domain of computer vision, high-level tasks such as object recognition and scene parsing have begun leveraging prompt learning to enhance model interpretability and accuracy. More recently, researchers have proposed utilizing prompts for low-level vision tasks. For instance, PromptIR [8] employs prompts to encode degradation-specific information for unseen image restoration.Building upon these advancements, this paper explores the use of soft prompts for task-specific learning. To prevent parameter conflicts, the base model parameters are frozen during the training process, and only the soft prompts are optimized. This avoids the need to train prompts and model parameters from scratch, ensuring efficient and conflict-free adaptation for specific tasks.

D. ALIGNMENT DISTRIBUTION

Deep learning has achieved significant success across numerous fields, largely predicated on the assumption that a large amount of labeled data is available and that the training and test sets are independent and identically distributed (IID) [26]. However, in practice, models are often trained on synthetic data while being tested on real-world data. This discrepancy between the training and testing domains leads to a violation of the IID assumption, a phenomenon commonly referred to as domain shift. Domain Adaptation (DA) has been proposed as a strategy to alleviate the challenges posed by domain shift.Despite its effectiveness, existing DA methods have notable limitations. Some approaches require access to both the source and target domain data during training, while others necessitate training across multiple domains simultaneously. The former assumes that source data remains accessible throughout adaptation, which is not always feasible, while the latter increases computational complexity and resource overhead.

In the field of computer vision, researchers have explored Test Time Adaptation (TTA) [16], [27] as a method to mitigate domain shift by updating pre-trained models online using test data. However, most TTA techniques have been applied to image classification tasks. For instance, [26] introduced a novel approach that constructs secondary degraded images and applies TTA to super-resolution image

inpainting tasks. Additionally, [28] proposed a test time prompt tuning (TPT) approach to extend TTA by introducing distributional alignment within the CLIP framework.Building on this concept, we propose leveraging token distribution statistics from the training data to explicitly align training and test sample distributions. By integrating test time prompt tuning, our method effectively mitigates domain shift and enhances model performance in real-world scenarios.

III. METHOD

In this section, we introduce the overall flow of our proposed method, as shown in Figure 1.

A. UNIFIED FORMULA

The all-in-one severe weather image restoration task aims to restore a clean image from a weather-degraded image. However, since the physical models under different weather phenomena are not unified, we try to construct a unified restoration formula. Let the weather-damaged image I and the original clean image B be used.

According to [29], the degradation of rain can be expressed as:

$$I = T \odot \left(B + \sum_{i}^{n} S_i \right) + (1 - T) \odot A \tag{1}$$

where $\odot$ represents element-wise multiplication, T is the transmission map generated by scattering effects, A is the atmospheric light in the scene, S_i is the rain layer, and n is the total number of raindrop layers.

According to [30], the degradation of snow can be expressed as:

$$I = a \odot z + B \odot (1 - z) \tag{2}$$

where a is the color aberration map and z is the independent snow mask.

According to [2], the degradation of fog can be expressed as:

$$I = T \odot B + (1 - T) \odot A \tag{3}$$

According to [4], the degradation of raindrop can be expressed as:

$$I = (1 - M) \odot B + R \tag{4}$$

where M is the mask and R is the raindrop residue.

In order to achieve a unified image restoration method, the above image degradation formulation can be integrated:

$$B = K \odot I + R + I \tag{5}$$

where $K = \frac{1}{T} - 1, R = \left(1 - \frac{1}{T}\right) A + r$, where r denotes the residual term.

Because the actual degradation process may be more complex than the theoretical model or contain some unknown factors, the model usually only captures the core or most dominant part of the degradation phenomenon. Therefore, the unified formula acts as an abstract model and provides a framework to integrate the effects of different weather conditions. Although it may not perfectly fit the actual situation of every weather, it can capture the main characteristics of multiple degradation types, and adopting prior knowledge without strong constraints can help the trained model better understand and reverse the degradation process, so as to achieve better restoration results.

B. MODEL ARCHITECTURE

Our network architecture is based on SwinTransformer in [20] and Dehazeformer in [2]. SwinTransformer has demonstrated its effects on high-level computer vision tasks. Given an input feature map $X \in \mathbb{R}^{b \times h \times w \times c}$, we apply MHSA, a multi-head self-attention mechanism with occlusion, using a linear layer to project X to queries, keys, and values, and use Windows to group into tokens. Dehazeformer adds an additional convolution operation on V to supplement the possible limitations of MHSA in spatial information aggregation and enhance the ability of the model to process spatial information. Specifically, self-attention is computed in the following way:

$$\begin{aligned} \text{Aggregation}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}} + B \right) V \\ + \text{Conv}\left(\hat{V}\right) \end{aligned} \tag{6}$$

where $Q, K, V \in \mathbb{R}^{b' \times l \times d}$ represents the query, key, and value, l is the number of tokens in the window, d is the dimension, $b' = b \times \left(h \times \frac{w}{l}\right)$, and $B \in \mathbb{R}^{l \times l}$ is the relative position bias term. $\hat{V} \in \mathbb{R}^{b \times h \times w \times c}$ denotes V when no window partition is performed. Therefore, the attention mechanism is used to aggregate the information within the window, and the convolution is also used to aggregate the information in the domain, so as to improve the spatial understanding ability of the model. Besides that, not all transformerblocks adopt MHSA, such as in the shallow layer of the encoder and the decoder stage.

Inspired by the above,as shown in Figure 2,We applied a 5-stage U-Net, replacing its convolutional block with a Transformerblock. In addition, SK fusion and soft reconstruction layers are used to replace the original cascade fusion and global residual.Our network takes an degraded image $I \in \mathbb{R}^{H \times W \times 3}$ as input and outputs a features map $O \in \mathbb{R}^{H \times W \times 4}$. We then split O into $K \in \mathbb{R}^{H \times W \times 1}$ and $R \in \mathbb{R}^{H \times W \times 3}$ in Eq(5).Finally, they are put into the formula to obtain the restored image.

In the fine-tuning stage, we freeze the baseline model backbone parameters and train a set of task-specific prompts to enhance the model's ability to handle specific tasks. The attention level prompt is introduced at the input of the MSHA layer, assuming that the prompt length is m, then Eq(6) can be improved as:

$$\begin{aligned} &\text{Aggregation}(Q, K, V) \\ &= \text{Softmax}\left(\frac{Q[\mathcal{P}k, K]^T}{\sqrt{d}} + B \right) [\mathcal{P}v, V] + \text{Conv}\left(\hat{V}\right) \end{aligned} \tag{7}$$

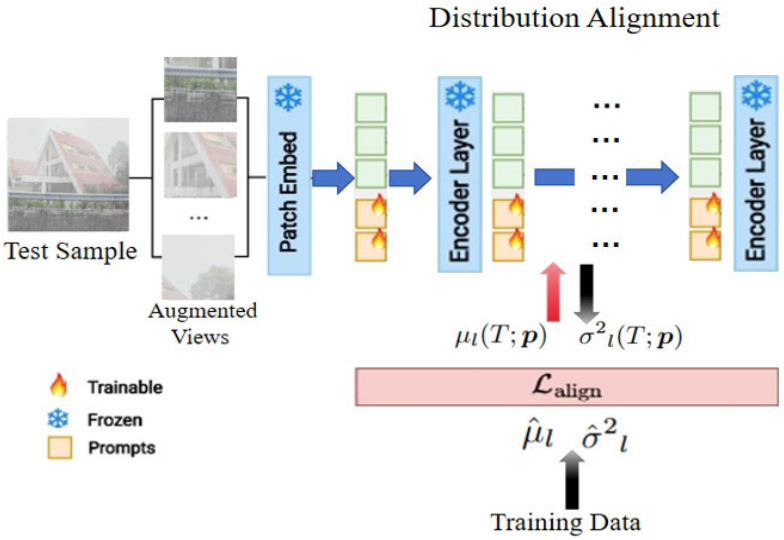

FIGURE 4. Overview of the methodology for distribution alignment. At test time, a single test sample and its augmented view are passed through the encoder and learnable prompts are updated using a distribution alignment loss to align the distribution statistics (mean and variance) of the test sample with the training set statistics.

where $\mathcal{P}k, \mathcal{P}v \in \mathbb{R}^{m\times d}$ denotes the prompt vectors inserted into K and V. Task-specific soft prompts are used to encode task-specific semantic knowledge, which is not shared between tasks. In this paper, we use this method to alleviate the problem of task interference that may occur in multi-task learning. At the same time, the baseline model backbone parameters are frozen to avoid conflicts between parameters.

C. INTERACTION ENCHANCED PROMPTS

Considering that although the effects of different weather conditions are visually different, there are some common patterns and characteristics in the process of image degradation, as well as task-specific characteristics [7]. Therefore, as shown in Figure 3, we further explores the implicit and explicit interactions between prompts in the fine-tuning stage. For implicit interaction, based on the above theory, we uses low-rank decomposition to divide the prompt into two parts, which can be expressed as follows:

$$\mathcal{P}_i = \mathcal{P}_i^s \cdot \mathcal{P}^g \tag{8}$$

where $\cdot$ denotes matrix multiplication, $\mathcal{P}_i^s \in \mathbb{R}^{m\times lr}$ denotes the task-specific prompt for the ith weather removal task, $\mathcal{P}^g \in \mathbb{R}^{lr\times d}$ is the task-general prompt, and lr is the rank of the parameter matrix. Task-specific prompts are tailored to individual tasks, while generic prompts are shared across all tasks. This design ensures that task-specific features are captured within task-specific prompts, while shared knowledge across tasks is embedded in generic prompts. By incorporating these distinctions, the hidden structure within the prompts can be better exploited. Moreover, applying low-rank decomposition to the prompt vectors, splitting them into generic and task-specific components, enhances the efficiency of parameter usage.Generic prompts, being shared across all tasks, reduce the overall parameter count for the model. In contrast, task-specific prompts facilitate specialized processing for individual tasks, enabling the model to internally distinguish between tasks and thereby reducing interference. Furthermore, this decomposition allows the adjustment of only the task-specific portions of the prompts when new tasks are introduced, avoiding the need for extensive modifications to the entire model and significantly enhancing its scalability.

In terms of explicit interactions, certain weather degradations exhibit similarities. For instance, snow and raindrop degradations are closely related, while rain and haze also share similarities—both featuring droplet-like occlusions, with rain additionally introducing streaking artifacts and opacity changes akin to haze. By leveraging these explicit relationships, we apply contrastive learning to task-specific prompts. This approach encourages prompts corresponding to similar degradation tasks to align more closely, enhancing their similarity through explicit interactions. Conversely, for unrelated degradation tasks with less clear interaction relationships, the model aims to achieve lower similarity between their corresponding prompts. To this end, we follow the method of [31] and construct the contras loss. Specifically, suppose that there are N weather tasks, T is the set of all tasks, and $i \in I \equiv \{1\cdots N\}$ is the index of any task. The contras loss can be expressed as follows:

$$\begin{aligned}\mathcal{L}_{contrastive} &= \sum_{i\in I} \mathcal{L}^i_{contrastive} \\ &= \sum_{i\in I} -\frac{1}{|T(i)|} \sum_{p\in T(i)} \log \frac{\frac{\exp\sin(\mathcal{P}_i,\mathcal{P}_p)}{\tau}}{\frac{\sum_{k\in I} \mathbf{1}_{i\neq k} \exp\sin(\mathcal{P}_i,\mathcal{P}_k)}{\tau}}\end{aligned} \tag{9}$$

where $T(i)$ is the index set of all tasks that have similar relationship with the ith task, $|T(i)|$ is its cardinality, τ is the temperature coefficient, $\mathbf{1}_{i\neq k}$ is an indicator function, which takes the value 1 if $i \neq k$ and 0 otherwise, $sim(x, y)$ represents the similarity between two vectors x and y. We enable the prompts to more accurately encode task-specific features by introducing a contrastive loss.

D. TWO-STAGE TRAINING STRATEGY

In order to avoid the conflict between task-generic parameters and task-specific parameters caused by training from scratch, we adopts a two-stage training strategy. In the pre-training stage, we employ a shared backbone network to learn weather general features in a supervised manner using a mixture of datasets from various weather conditions. The loss function in the pre-training stage can be formulated as follows:

$$\mathcal{L}_{pretrain} = \sum_{\rho} \left(\| \hat{X}_\rho, X_\rho \|_1 \right) \tag{10}$$

where $\hat{X}_\rho, X_\rho$ denote the input degraded image and the target clean image for weather ρ, respectively.

TABLE 1. Details of the data sets corresponding to different weather degradation types.

Dataset	Degradation	Train samples	Test samples
Outdoor Rain	Rain	4,500	750
Snow100K	Snow	50,000	50,000
RESIDE-outdoor	Haze	13,099	500
RainDrop	Raindrop	861	58

In the second fine-tuning stage, we freeze the parameters of the backbone and introduce a contras loss to train the prompts. The loss function in the fine-tuning stage can be expressed as follows:

$$\mathcal{L}_{finetune} = \sum_{\rho} \left(\| \hat{X}_\rho, X_\rho \|_1 \right) + \lambda_{cont} \mathcal{L}_{contrastive} \tag{11}$$

where $L_{contrastive}$ is the contrastive loss and λ_{cont} is the hyperparameter for the contrastive loss regularization. The value of λ_{cont} is set to 0.1 in the experiments in this paper.

E. ALIGNMENT DISTRIBUTION

Given a test sample X_{test}, we take multiple augmented views on it and pass them to the encoder with deep learnable prompts, as shown in Figure 4. At each layer of the encoder, we compute the alignment loss between the mean and variance of the test sample and the mean and variance of the training dataset. Our ultimate goal is to align the distribution of the test data with the training data by updating the prompts based on the alignment loss.

We use a random set of augmentations $Aug(\cdot)$ on the test data, resulting in a set of augmented views N_k. The mean and variance statistics of the tokens embedding for the test sample are computed at the output of each transformer layer of the model encoder. For the mean and variance of the distribution of the test samples:

$$\mu_l(T; \boldsymbol{p}) = \frac{1}{N_k} \sum_{\mathrm{x} \in Aug(X)} \widetilde{\boldsymbol{X}}^p{}_{l,\mathrm{x}},$$
$$\sigma^2{}_l(T; \boldsymbol{p}) = \frac{1}{N_k} \sum_{\mathrm{x} \in Aug(X)} \left(\widetilde{\boldsymbol{X}}^p{}_{l,\mathrm{x}} - \mu_l(T; \boldsymbol{p}) \right)^2 \tag{12}$$

where $\mu_l(T; \boldsymbol{p})$ and $\sigma^2{}_l(T; \boldsymbol{p})$ is the mean and variance of the test sample in the Lth layer in the encoder. Similarly, for each layer of the encoder, we calculate the statistics of the training data:

$$\hat{\mu}_l = \mu_l(S, \theta_v) \quad \text{and} \quad \hat{\sigma}^2{}_l = \sigma^2_l(S, \theta_v) \tag{13}$$

θ_v represents the encoder parameters from the pre-trained model. We compute the alignment distribution loss between the mean and variance of the test sample and the training dataset statistics as follows:

$$\mathcal{L}_{\text{align}} = \frac{1}{L} \sum_{l=1}^{L} \left(\|\mu_l(T; \boldsymbol{p}) - \hat{\mu}_l\|_1 + \|\sigma^2{}_l(T; \boldsymbol{p}) - \hat{\sigma}^2{}_l\|_1 \right) \tag{14}$$

With the alignment loss $\mathcal{L}_{\text{align}}$, we align the test data with the training data distribution.

IV. EXPERIMENTS

A. IMPLEMENT DETAILS

1) DATASETS

For four different severe weather degraded images and restoration tasks, we prepare four data sets respectively. For the rain removal task, the OutdoorRain data set is selected in this paper. In the OutdoorRain data set, there are 4500 pairs of images for training samples and 750 pairs of images for testing samples, and each pair of images includes degraded images and clean images. For the snow removal task, we use the Snow100K dataset, which has 50000 pairs of images for training samples and 50000 pairs of images for testing samples. Each pair of images includes degraded images and clean images. For the dehazing task, we choose the RESIDE dataset, which has two parts: indoor images and outdoor images. Considering the actual needs, we choose outdoor images. In the RESIDE_outdoor dataset, there are 13099 pairs of images for training samples and 500 pairs of images for testing samples. Each pair of images includes both degraded and clean images. For the RainDrop removal task, we choose the RainDrop dataset. There are 861 pairs of images in the training sample and 58 pairs of images in the test sample in the Raindrop dataset, and each pair of images includes degraded images and clean images. The specific information of the four datasets is summarized in Table 1.

At the same time, a key point to consider is that the gap in the number of samples between the four datasets is too large, which may lead to suboptimal performance of our model on specific tasks, resulting in poor multi-task learning

TABLE 2. Quantitative comparison on the outdoor rain dataset.

Type	method	PSNR↑	SSIM↑
Typical-weather-remoal	Pix2pix	19.09	0.71
	HRGAN	21.56	0.86
	MPRNet	21.90	0.85
All-weather-removal	All-in-one	24.71	0.90
	TransWeater	23.18	0.84
	Chen et al.	23.94	0.85
	WGWS-net	25.31	0.90
	Ours	**28.61**	**0.90**

TABLE 3. Quantitative comparison on the Snow100K dataset.

Type	method	PSNR↑	SSIM↑
Typical-weather-remoal	DetailsNet	19.18	0.75
	DesnowNet	27.17	0.90
	JSTASR	25.32	0.81
	DDMSNET	28.85	0.88
All-weather-removal	All-in-one	28.33	0.88
	TransWeater	27.80	0.85
	Chen et al.	29.27	0.88
	WGWS-net	29.71	0.89
	Ours	**29.53**	**0.88**

effect, so that the model may show a tendency to prioritize degradation types with a larger number of samples and ignore degradation types with a smaller number of samples. So we need to sample the training data of different datasets and roughly align the number of samples of different datasets. Specifically, we take all samples (4500 pairs) from the OutdoorRain training set, 7500 pairs of images randomly sampled from the Snow100K training set, 5000 pairs of images randomly sampled from the RESIDE dataset, and 1500 pairs of images cyclically sampled from the RainDrop training set. For fair comparison, we keep the original number of test image pairs constant for all datasets during evaluation

2) TRAINING DETAILS

We tune the parameters and learning rate of the neural network using Adam optimizer combined with cosine annealing scheme. In the first pre-training stage, the model is trained for 200 epochs with a batch size of 4, using an initial learning rate of 0.0002. In the second fine-tuning stage, the model is trained for 250 cycles with a batch size of 8 and an initial learning rate of 0.00002 used. In this experiment, random flipping was applied to enhance the image and the image was randomly cropped into blocks of 256 × 256 size.For distribution alignment, we train for 1 cycle based on the alignment loss with a batch size of 64 and an initial learning rate of 0.04 used.We generate 63 augmentation attempts, plus a single test image, for a total of 64 in a batch.

3) EVALUATION METRICS

We used the most commonly used evaluation metrics, PSNR and SSIM, which can be respectively formulated as follows:

$$PSNR = 10\log_{10}(\frac{MAX_I^2}{MSE}) \tag{15}$$

TABLE 4. Quantitative comparison on the RESIZE_outdoor dataset.

Type	method	PSNR↑	SSIM↑
Typical-weather-remoal	EPDN	23.82	0.89
	PFDN	31.45	0.97
	KDDN	33.49	0.97
	MPRNET	31.31	0.97
All-weather-removal	All-in-one	30.49	0.95
	TransWeater	27.66	0.95
	Chen et al.	30.76	0.97
	WGWS-net	30.85	0.98
	Ours	**31.89**	**0.97**

TABLE 5. Quantitative comparison on the RainDrop dataset.

Type	method	PSNR↑	SSIM↑
Typical-weather-remoal	Pix2pix	28.02	0.85
	Attn.GAN	30.55	0.90
	Quan et al.	31.44	0.93
	CCN	31.34	0.95
All-weather-removal	All-in-one	31.32	0.93
	TransWeater	28.98	0.90
	Chen et al.	30.75	0.91
	WGWS-net	31.31	0.93
	Ours	**30.07**	**0.91**

where MAX_I denotes is the maximum possible pixel value of the image, MSE is the mean squared error.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{16}$$

where x, y are the two images being compared, μ denotes the mean value of pixels, σ^2 denotes the variances, σ_{xy} denotes the covariance between x and y and $c1, c2$ are constants used to stabilize the division with weak denominators.

B. COMPARISON WITH STATE-OF-THE-ART METHODS

1) COMPARISON OF RAIN REMOVAL METHODS

As in Table 2, we show the single rain removal method and the multi-weather recovery method. Among them, for a single rain removal method, we show quantitative results for Pix2pix in [32], HRGAN in [29], and MPRNet in [33]. For multi-weather recovery methods, we show quantitative results for All-in-one in [10], TransWeather in [6], and Chen et al. in [9] and WGWS-Net in [7].

2) COMPARISON OF SNOW REMOVAL METHODS

As in Table 3, we show the single snow removal method and the multi-weather recovery method. Among them, for single snow removal methods, we show quantitative results for DetailsNet in [19], DesnowNet in [30], JSTASR in [34], and DDMSNET in [35]. For multi-weather recovery methods, we show quantitative results for All-in-one, TransWeather, Chen et al. and WGWS-Net.

3) COMPARISON OF DEHAZING METHODS

As in Table 4, we shows the single dehazing method and the multi-weather recovery method. Among them, for single

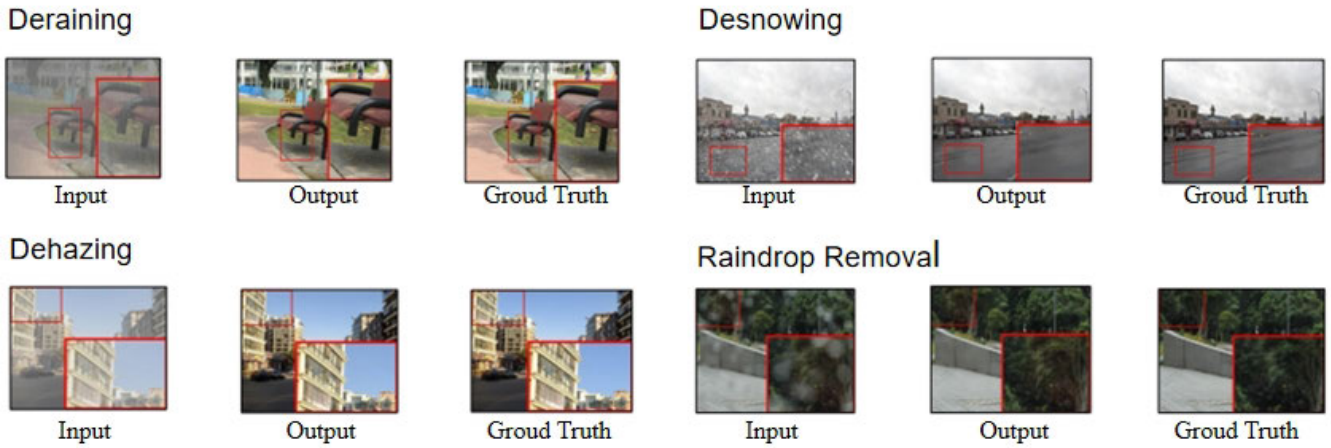

FIGURE 5. Visual analysis of the results. It can be seen that our method achieves good visual results in four kinds of weather.

dehazing methods, we show quantitative results for EPDN in [36], PFDN in [37], KDDN in [38] and MPRNet in [33]. For multi-weather recovery methods, we show quantitative results for All-in-one, TransWeather, Chen et al. and WGWS-Net.

4) COMPARISON OF RAINDROP REMOVAL METHODS

As in Table 5, we show the single raindrop removal method and the multi-weather recovery method. Among them, for single raindrop removal methods, we show quantitative results for Pix2pix in [32], Attn.GAN in [4], Quan et al. in [39] and CCN Quan et al. in [40]. For multi-weather recovery methods, we show quantitative results for All-in-one, TransWeather, Chen et al. and WGWS-Net.

Analyzing the results, our method performs well in rain removal, snow removal, and fog removal, Raindrop removal tasks; although its performance in raindrop removal task is not among the top, the results are still competitive, which may be due to the limited number of raindrop samples. In addition, we analyzed the complexity of different methods by comparing the number of model parameters for various multi-weather recovery methods. As shown in Table 6, our model is simpler in complexity compared to other multi-weather recovery methods.

After that, four images are randomly selected as input, as shown in Figure 5, and the four degraded images are processed by the model and all perform well visually.

TABLE 6. Complexity comparison of different multi-weather recovery methods.

	Method	Params
All-weather-removal	All-in-one	–
	TransWeater	38.05M
	Chen et al.	28.72M
	WGWS-net	9.35M
	Ours	**6.62M**

TABLE 7. Ablation study on different prompt settings.

DRank	Length	Rain	Snow	Haze	Raindrop
0	8	27.63	28.27	30.78	28.72
4	8	**27.67**	**28.31**	**30.84**	**28.76**
4	16	27.61	28.25	30.73	28.66
8	8	27.25	28.13	30.63	28.48

V. ABLATION STUDY

A. IMPLICIT INTERACTION ENHANCED PROMPTS

To explore the implicit interaction between prompts, we study the effect of different values of the rank of the prompt parameter matrix on the model performance. The quantitative results, shown in Table 7, show that the model performs best when the rank is 4 and the prompt length is 8. It is worth mentioning that a rank of 0 means that all tasks share the same prompt parameters, while a rank equal to the prompt length means that prompt parameters are not shared at all between tasks. According to the results, using the same prompt or completely different prompts for all tasks did not produce the best results, which is due to the fact that different weather conditions have both general and specific characteristics. The ablation experiments under different prompting Settings in this experiment verify the effectiveness of using low-rank decomposition prompting.

B. AUGMENTED VIEWS AT TEST TIME

From the Table 8, it can be known that the performance of the model improves as the number of augmented views increases at test time. This is because quantity enhancement makes the effect of distribution alignment increase. As you can see, going from 64 to 128 doesn't make a huge difference, and the cost of doing more computation is much more noticeable. Therefore, we choose 64 augmented views.

C. MODEL COMPONENTS

As shown in Table 9, we start with a Base image restoration model pre-trained in a supervised manner, denoted as "base"

TABLE 8. Ablation Study on different number of augmented views.

Type	Number			
	16	32	64	128
Rain	28.55	28.59	**28.61**	28.63
Snow	29.41	29.48	**29.53**	29.54
Haze	31.74	31.82	**31.89**	31.89
Raindrop	30.03	30.05	**30.07**	30.07

TABLE 9. Ablation study on different model settings.

Types	Base	+ Prompts	+ Contrastive Learning	+ Align
Rain	26.63	27.67	28.43	28.61
Snow	27.57	28.31	29.26	29.53
Haze	30.16	30.84	31.64	31.89
Raindrop	28.37	28.76	30.02	30.07
Avg.inference time	0.246s	0.334s	0.376s	0.427s

in Table 9. The performance is then evaluated by fine-tuning the pre-trained model with soft Prompts, denoted as "+ Prompts" in Table 9. The added prompts are then augmented by Contrastive Learning, denoted as "+ Contrastive Learning" in Table 9. Finally, distribution alignment is performed, denoted as "+ Align" in Table 9, and concreteness is shown by PSNR. As you can see, each part of our approach contributes to the performance improvement. From the analysis of "The average inference time per sample," the computational overhead introduced by each component of the model does not have a significant impact compared to the improvement in model performance.In the raindrop removal task, adding align didn't improve much. The analysis should be because the training pictures of raindrops themselves are not as different from the real pictures as the other three types of weather. For augmented views, not all test images yield good results. For images with distinct local features under certain weather conditions, the results may not be satisfactory even after adjustments using augmented views.

VI. CONCLUSION

In this study, we propose a novel framework for multi-weather image restoration that leverages interactively augmented prompts and distribution alignment strategies. The framework adopts a two-stage training strategy: a pre-training stage to learn task-generic features and a fine-tuning stage to learn task-specific prompts. During the fine-tuning stage, certain parameters are frozen to prevent conflicts between task-generic and task-specific parameters, which can arise when training from scratch.Furthermore, we investigate both the explicit and implicit interactions among the prompts to enhance the model's ability to recover images degraded by various weather conditions. To address the domain shift problem, we implement distribution alignment at test time, bridging the significant gap between the training data and real-world test data. This approach effectively mitigates domain shift, improving model performance in practical scenarios.We validate the effectiveness of the proposed framework through extensive experiments, demonstrating its superior performance compared to state-of-the-art methods.

REFERENCES

[1] X. Chen, H. Li, M. Li, and J. Pan, "Learning a sparse transformer network for effective image deraining," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 5896–5905.
[2] Y. Song, Z. He, H. Qian, and X. Du, "Vision transformers for single image dehazing," *IEEE Trans. Image Process.*, vol. 32, pp. 1927–1941, 2023.
[3] S. Chen, T. Ye, Y. Liu, and E. Chen, "SnowFormer: Context interaction transformer with scale-awareness for single image desnowing," 2022, *arXiv:2208.09703*.
[4] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, "Attentive generative adversarial network for raindrop removal from a single image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2482–2491.
[5] S. You, R. T. Tan, R. Kawakami, Y. Mukaigawa, and K. Ikeuchi, "Adherent raindrop modeling, detection and removal in video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 9, pp. 1721–1733, Sep. 2016.
[6] J. M. Jose Valanarasu, R. Yasarla, and V. M. Patel, "TransWeather: Transformer-based restoration of images degraded by adverse weather conditions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 2343–2353.
[7] Y. Zhu, T. Wang, X. Fu, X. Yang, X. Guo, J. Dai, Y. Qiao, and X. Hu, "Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 21747–21758.
[8] V. Potlapalli, S. W. Zamir, S. Khan, and F. S. Khan, "PromptIR: Prompting for all-in-one blind image restoration," 2023, *arXiv:2306.13090*.
[9] W.-T. Chen, Z.-K. Huang, C.-C. Tsai, H.-H. Yang, J.-J. Ding, and S.-Y. Kuo, "Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17632–17641.
[10] R. Li, R. T. Tan, and L.-F. Cheong, "All in one bad weather removal using architectural search," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3172–3182.
[11] D. Hendrycks and T. Dietterich, "Benchmarking neural network robustness to common corruptions and perturbations," 2019, *arXiv:1903.12261*.
[12] B. Recht, R. Roelofs, L. Schmidt, and V. Shankar, "Do ImageNet classifiers generalize to ImageNet," in *Proc. Int. Conf. Mach. Learn.*, May 2019, pp. 5389–5400.
[13] H. He and J. D. Choi, "The stem cell hypothesis: Dilemma behind multi-task learning with transformer encoders," 2021, *arXiv:2109.06939*.
[14] Y. Ding and K. Wu, "A multi-task learning and knowledge selection strategy for environment-induced color-distorted image restoration," *Appl. Sci.*, vol. 14, no. 5, p. 1836, Feb. 2024.
[15] Y. Li, X. Nie, W. Diao, and S. Zheng, "Lifelong CycleGAN for continual multi-task image restoration," *Pattern Recognit. Lett.*, vol. 153, pp. 183–189, Jan. 2022.
[16] G. Hinton and L. Van Der Maaten, "Visualizing data using t-SNE journal of machine learning research," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[17] H. Zhang and V. M. Patel, "Density-aware single image de-raining using a multi-stream dense network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 695–704.
[18] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.
[19] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1715–1723.
[20] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.
[21] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, and S. Agarwal, "Language models are few-shot learners," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, Jan. 2020, pp. 1877–1901.

[22] K. Zhou, J. Yang, C. C. Loy, and Z. Liu, "Learning to prompt for vision-language models," *Int. J. Comput. Vis.*, vol. 130, no. 9, pp. 2337–2348, Sep. 2022.
[23] T. Tang, J. Li, W. Xin Zhao, and J.-R. Wen, "MVP: Multi-task supervised pre-training for natural language generation," 2022, *arXiv:2206.12131*.
[24] S. Ji, J. Zuo, M. Fang, Z. Jiang, F. Chen, X. Duan, B. Huai, and Z. Zhao, "TextrolSpeech: A text style control speech corpus with codec language text-to-speech models," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2024, pp. 10301–10305.
[25] B. Lester, R. Al-Rfou, and N. Constant, "The power of scale for parameter-efficient prompt tuning," 2021, *arXiv:2104.08691*.
[26] Z. Deng, Z. Chen, S. Niu, T. H. Li, B. Zhuang, and M. Tan, "Efficient test-time adaptation for super-resolution with second-order degradation and reconstruction," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, Jan. 2023, pp. 74671–74701.
[27] Y. Sun, X. Wang, Z. Liu, J. J. Miller, A. A. Efros, and M. Hardt, "Test-time training with self-supervision for generalization under distribution shifts," in *Proc. Int. Conf. Mach. Learn.*, Jan. 2019, pp. 9229–9248.
[28] J. Hassan, H. Gani, N. Hussein, M. U. Khattak, M. Naseer, F. S. Khan, and S. Khan, "Align your prompts: Test-time prompting with distribution alignment for zero-shot generalization," 2023, *arXiv:2311.01459*.
[29] R. Li, L.-F. Cheong, and R. T. Tan, "Heavy rain image restoration: Integrating physics model and conditional adversarial learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1633–1642.
[30] Y.-F. Liu, D.-W. Jaw, S.-C. Huang, and J.-N. Hwang, "DesnowNet: Context-aware deep network for snow removal," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 3064–3073, Jun. 2018.
[31] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2020, pp. 18661–18673.
[32] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5967–5976.
[33] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Multi-stage progressive image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14816–14826.
[34] W. Chen, H. Fang, J. Ding, C.-C. Tsai, and S. Kuo, "JSTASR: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal," in *Proc. 16th Eur. Conf. Comput. Vis. (ECCV)*, Glasgow, U.K., Jan. 2020, pp. 754–770.
[35] K. Zhang, R. Li, Y. Yu, W. Luo, and C. Li, "Deep dense multi-scale network for snow removal using semantic and depth priors," *IEEE Trans. Image Process.*, vol. 30, pp. 7419–7431, 2021.
[36] Y. Qu, Y. Chen, J. Huang, and Y. Xie, "Enhanced Pix2pix dehazing network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8152–8160.
[37] J. Dong and J. Pan, "Physics-based feature dehazing networks," in *Proc. 16th Eur. Conf. Comput. Vis. (ECCV)*, Glasgow, U.K. Cham, Switzerland: Springer, Aug. 2020, pp. 188–204.
[38] M. Hong, Y. Xie, C. Li, and Y. Qu, "Distilling image dehazing with heterogeneous task imitation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3459–3468.
[39] Y. Quan, S. Deng, Y. Chen, and H. Ji, "Deep learning for seeing through window with raindrops," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2463–2471.
[40] R. Quan, X. Yu, Y. Liang, and Y. Yang, "Removing raindrops and rain streaks in one go," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9143–9152.

LINGWEN MENG received the master's degree from Shandong University, in 2012.

She mainly researches risk control and safety technology of power grid operation. She has published more than ten articles and authorized more than 50 invention patents.

MINGYONG XIN received the master's degree from the University of Electronic Science and Technology of China, in 2014.

He started work in August 2014, presided more than two key projects of South Grid, participated in two national key research and development program, four provincial and ministerial projects, and more than ten other South Grid projects, mainly engaged in research work in the field of intelligent sensor and power chip and equipment technology. He has published more than 30 articles and authorized more than 20 invention patents.

JUNWEI ZHANG received the master's degree from Guizhou University, in 2013.

She mainly engaged in scientific and technological achievements and awards, and scientific and technological project management. She has published more than 30 articles, six articles in SCI/EI, and more than 30 authorized invention patents.

FU ZOU received the bachelor's degree from North China Electric Power University, in 2016.

In August 2016, he joined the Tongren Management Office of Power Transmission Operation and Maintenance Branch, Guizhou Grid Company Ltd., and worked as a crew member of power transmission operation and maintenance. In 2019, he was with the Safety Supervision Department, Transmission Operation and Maintenance Branch, Guizhou Power Grid Company Ltd., as the assistant responsible for the safety risk management system. In 2021, he was with the Safety Supervision Department, Power Transmission Operation and Maintenance Branch, Guizhou Grid Company Ltd., as the assistant responsible for emergency power supply protection. In 2023, he was with the Safety Science Research Center, Power Science Research Institute, Guizhou Grid Company Ltd., he will be responsible for comprehensive supervision and assistance. His main research interests include power safety and emergency power supply. He participated in the application of the invention patent of "A dc ice melting device LCC and MMC parallel operation control method and System" and participated in the publication of the article "Emergency Capacity Assessment of Power Grid Enterprises Based on Analytic Hierarchy Process" which was selected in "Power Big Data."

QINGQING ZHAO received the bachelor's degree in automation from Guizhou University, in 2021.

She works with Liupan Water Power Supply Bureau, Guizhou Power Grid. She has been granted four patents as the first person in charge. Her current research interests include security technology and intelligent power transformation.