Prompting Rain Off: Evolving Compact Dual Prompts for Continual De-Raining

Minghao Liu, *Student Member, IEEE*, Wenhan Yang, *Member, IEEE*, and Jiaying Liu, *Fellow, IEEE*

***Abstract*—In recent years, there has been notable progress in single-image rain removal, particularly focusing on static data distributions in these approaches. When dealing with data that constantly changes, the challenge of catastrophic forgetting arises, which is quite common and critical in real-world scenarios. To address this, we propose Evolving COmpact Dual Prompt Learning (EcoDPL), an efficient rehearsal-free continual learning deraining framework designed specifically for low-level vision tasks. Specifically, we design two prompt pools at both image and feature levels and insert these prompts into images and embedding tokens, for better knowledge transfer across tasks. Our adaptive weight generation module, P-Fuser, attaches an attention map to each prompt, to adaptively pay attention to different inputs, and get different weights to fuse prompts, making the inserted prompts more flexible with various inputs. Also, we introduce Grad-Tuner, a dictionary learning strategy, to compress knowledge into fewer prompts. This makes the knowledge more compact and provides more space for new prompts to learn new tasks. Our method stands out by leveraging small, learnable prompts for efficient knowledge retention across tasks, not increasing training time or parameters. Furthermore, we present an augmented method that upgrades the distance function γ from simple cosine distance to a more advanced weight generation network. We also employ a fine-tuned dictionary learning technique, compressing knowledge into a more compact form, and enhancing the ability of prompts to learn new tasks. With our new designs, the model becomes more flexible with various inputs and it compresses knowledge into fewer prompts to free up spaces to learn new tasks. Through extensive experiments on various rain removal datasets, our EcoDPL method consistently outperforms previous continual learning techniques. Notably, although EcoDPL is designed for continual learning with changing data, it also performs well with stationary data, proving its robustness and versatility. Our website is available at: https://starymoon.github.io/Prompting-Rain-Off.**

***Index Terms*—Continual learning, dual prompt learning, rain removal, dictionary learning technique.**

Received 29 August 2025; revised 22 January 2026 and 18 March 2026; accepted 21 April 2026. Date of publication 7 May 2026; date of current version 19 May 2026. This work was supported in part by the National Natural Science Foundation of China under Grant 62332010; in part by the Key Laboratory of Science, Technology and Standard in Press Industry (Key Laboratory of Intelligent Press Media Technology); in part by the Major Key Project of PCL under Grant PCL2025A03; and in part by the Interdisciplinary Frontier Research Project of PCL under Grant PCL2025QYB013. The associate editor coordinating the review of this article and approving it for publication was Dr. Yan-Tsung Peng. *(Corresponding author: Jiaying Liu.)*

Minghao Liu and Jiaying Liu are with the Wangxuan Institute of Computer Technology, Peking University, Beijing 100000, China (e-mail: lmhlmh@stu.pku.edu.cn; liujiaying@pku.edu.cn).

Wenhan Yang is with the Pengcheng Laboratory, Shenzhen, Guangdong 518066, China (e-mail: yangwh@pcl.ac.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/TIP.2026.3689428, provided by the authors.

Digital Object Identifier 10.1109/TIP.2026.3689428

I. Introduction

As ONE of the most prevalent weather degradations, rain significantly reduces visibility and impairs the quality of images. This not only impacts the human visual experience but also affects critical computer vision systems, such as detection [1], segmentation [2], and depth estimation [3]. Such degradation has profound implications for practical applications, including autonomous driving and surveillance systems.

Single-image rain removal has seen considerable progress in recent years. Conventional methods [4], [5], [6], [7], [8], mainly leveraging convolutional networks, have consistently achieved remarkable performance. Yet, the more recent transformer-based route [9] has offered more excellent results. Noted this progress, the issue of rain degradation still poses another challenge. Image deraining is inherently a dynamic and context-sensitive task, as real-world environments exhibit diverse and evolving rain patterns influenced by factors such as geography, season, and climate conditions. To maintain robust performance across these varying scenarios, continual learning is essential, enabling the model to adapt to newly emerging rain types and intensities, while mitigating catastrophic forgetting and sustaining long-term generalization.

Existing continual learning methods can be roughly divided into four categories: parameter isolation-based approaches [10], [11], replay-based mechanisms [12], regularization-based strategies [13], [14], and prompt-based methods [15], [16], [17], [18], [19], [20], [21]. However, these methods come with inherent challenges. For instance, parameter isolation methods have continuously expanding model sizes. Replay-based methods suffer from additional storage and computational costs. Regularization-based methods suffer from inefficiencies when more tasks are involved, while existing prompt-based methods are insufficient for low-level image restoration tasks like deraining. Moreover, these methods struggle to autonomously identify and utilize relevant knowledge for arbitrary samples without explicit task identity, which causes a considerable obstacle in practical use.

Recently, the presence of prompt-based learning has introduced a new continual learning paradigm [22]. By reformulating learning tasks via prompt design [23], there is no need to adjust model weights directly. Prompt-based learning, emphasizing task-specific knowledge encoding via prompts, offers a more resource-efficient way to employ pre-trained models than traditional fine-tuning. As a result, prompting has been identified as a better choice for continual learning [16].

1941-0042 © 2026 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Peking University. Downloaded on July 03,2026 at 08:42:56 UTC from IEEE Xplore. Restrictions apply.

Fig. 1. The comparison between the traditional continual learning method and our EcoDPL method. The first row represents the parameter regularization method, such as PIGWM [18], while the second row illustrates our EcoDPL method. (a) Overview of the respective framework: The EcoDPL framework utilizes a unified model and prompt pools to store task-specific knowledge, employing learnable parameters as a form of memory. EcoDPL selectively updates prompts on an instance-by-instance basis, optimizing them to guide model predictions. It explicitly manages both task-invariant and task-specific knowledge while preserving the plasticity of the model. (b) Visualization of two continual learning methods: Our EcoDPL outperforms traditional methods in rain removal, restoring more details in images.

However, recent prompt-based continual learning methods, such as visual prompt tuning [15], learning to prompt (L2P) [16], dual prompt learning (DPL) [17], parameter importance guided weights modification (PIGWM) [18], CODA-Prompt [19], Hierarchical Prompt Learning [20], and DualPrompt [21], have primarily been developed and optimized for high-level classification tasks. These methods are primarily tailored for high-level classification tasks and rely on an *expansion strategy*. For instance, CODA-Prompt freezes old parameters and incrementally adds new prompt components for new tasks, which inevitably increases model complexity or requires a large pre-allocated buffer. In contrast, our EcoDPL introduces a novel *compression paradigm* via the Grad-Tuner module. As illustrated in Fig. 1, instead of expanding parameters, we employ dictionary learning to actively compress new knowledge into a compact set of basis vectors, maintaining a fixed parameter budget while increasing knowledge density. Furthermore, unlike DualPrompt which divides prompts into general and expert types for semantic distinction, our dual-level design (Image and Feature levels) is specifically engineered for low-level restoration. The image-level prompts and P-Fuser directly modulate the input space to handle spatially non-uniform rain streaks, a capability lacking in existing classification-oriented prompt methods.

Hence, our contributions are:

- We introduce a novel approach to continual learning based on prompt learning, to manage diverse forms of rain streaks using a single model with a fixed set of additional parameters. As far as we know, this is the first attempt to incorporate prompt learning into CL model for low-level vision tasks, achieving superior performance across multiple benchmarks.
- We present an advanced dual prompt learning technique for deraining that employs prompts at both the image and feature levels, which effectively transfers domain-specific knowledge and performs well in various benchmarks.
- We introduce an adaptive weight generation module, P-Fuser, that assigns a learnable attention map to each prompt, which learns to select useful information based on the specific input, enabling our model to adapt to various inputs.
- We design a dynamic compression algorithm, Grad-Tuner, which uses a gradually tuned dictionary learning strategy. This compresses knowledge into a more compact form, effectively capturing prior knowledge with fewer prompts, and reserving more capacity for future tasks.

This paper builds upon our previous work [17], where we have made additional substantial contributions and improvements: 1) In [17], prompts are selected relying on the cosine distance with the input. In this work, we design an Adaptive Weight Generation Module to replace it that assigns a learnable attention map to each prompt, which learns to select useful information considering the context while being robust to noise, enabling our model to adapt to diverse inputs. The details are illustrated in Sec. III-C and Sec. III-D. 2) In our previous work, prompts are as a whole, which lacks flexibility. In our current version, prompts are regarded as basic prompt components, and they are weighted to form a new prompt for each input based on the Adaptive Weight Generation Module, which is demonstrated to be more effective than the previous conference version. 3) We introduce a gradually tuned dictionary learning module, Grad-Tuner. Different from the direct transfer of prompts among tasks in our previous work, it utilizes dictionary learning to compress knowledge into fewer prompts, making the prompt knowledge more compact. This provides more space for new prompts to acquire new knowledge, increasing plasticity while maintaining stability, and offering significant advantages for transferring and accumulating knowledge.

The structure of the paper is as follows: Sec. II offers a concise summary of related works. Sec. III introduces our proposed evolving compact dual prompts learning approach. Sec. IV shows extensive experiments and results. Finally, Sec. V presents the conclusion.

Authorized licensed use limited to: Peking University. Downloaded on July 03,2026 at 08:42:56 UTC from IEEE Xplore. Restrictions apply.

II. Related Work

A. Rain Removal From Single Images

Removing rain from single images has been an important research area, different from video deraining. The main challenge is the limited information in one frame, leading to many different algorithms.

Originally, single-image rain removal methods used filtering methods. Xu et al. [24] used a guided filter for this. A strategy further enhanced by researchers like Zheng et al. [25], who incorporated multiple guided filtering techniques, and Kim et al. [26], who focused on the distinct slender oval shape of rain streaks. You et al. [27] introduce a method based on motion for identifying raindrops.

Later, researchers tried prior-based methods, using image characteristics or learned knowledge. Based on the Maximum a Posteriori (MAP) principle [28], these methods optimize functions to get rain-free images [29]. After that, diverse strategies have been explored, from image decomposition [30] to complex regularization priors [31]. The features of rain streaks, captured through histograms of oriented gradients (HOGs), were utilized by Kang et al. [32] to categorize them into rain and non-rain dictionaries.

With the rise of deep learning, it was applied to rain removal. Eigen et al. [33] first used CNNs for deraining. Qian et al. [34] improved it for larger rain streaks with attentive networks. Fu et al. [35] made progress with DerainNet, deep detail network [4], and later with the residual-guided feature fusion network [36]. Zhang et al. [37] used conditional GANs, and others, like the multi-stream dense network [38], focused on rain density. Recently, understanding the atmospheric process of rain led to new methods. Yang et al. [39] introduced a multi-task RNN architecture for progressive rain removal. Li et al. [40] added a context aggregation network. Ren et al. propose a progressive recurrent deraining network (PReNet) [5] that simplifies this, presenting a straightforward, recurrently unfolding ResNet. Training data is another issue. Many algorithms use synthetic datasets, which might not work well with real images. To address this issue, Wang et al. [41] trained a network on natural rainy scenes. The semi-supervised approach [42] of Wei et al. also addresses sample collection problems. Recently, the field has witnessed a paradigm shift from CNNs to Transformer-based architectures, leveraging their ability to model long-range dependencies. Restormer [43] introduces an efficient transformer model that computes self-attention across channel dimensions, achieving state-of-the-art results on various restoration tasks. Xiao et al. [44] propose an Image De-raining Transformer (IDT) with window-based attention to handle spatial variability. More recently, DRSformer [45] utilizes a sparse transformer to retain top-k relevant features for effective rain removal, while PromptIR [46] employs prompts to adapt to different degradation types within an all-in-one framework. Furthermore, research is beginning to address the challenge of continual adaptation in deraining. A notable work is the Continual Image Deraining with Hypergraph Convolutional Networks [47], which utilizes hypergraph structures to preserve topological information of rain streaks across sequential tasks. Different from these approaches, our work focuses on a prompt-based compression paradigm to efficiently manage knowledge for continual learning.

B. Continual Learning

Continual learning is of great importance due to its ability to continuously learn from ongoing data without forgetting previous knowledge. The key goal is to keep old knowledge while adapting to new data. In the beginning, researchers faced the catastrophic forgetting problem when learning different tasks one after the other. Early efforts included reducing overlap in how different data are represented and replaying old samples [48] or virtual samples from past data [49]. However, limited by computational resources, these methods were confined to small architectures and few samples. The difficulties posed by catastrophic forgetting and the demands of continual learning have recently become a central focus with the improvements of deep neural networks. Continual learning now mainly includes three primary methods: Firstly, there are replay methods [50], where a subset of stored samples of previous task samples are trained again during new task learning to reduce forgetting. Generative models [51] are also used, known for their ability to create high-quality images, paving the way for retraining on generated examples. Secondly, regularization-based methods add additional regularization terms to the loss function, aiding in preserving prior knowledge while learning new data. These methods can be categorized into data-focused and prior-focused approaches. Data-focused techniques [10], [13] rely on knowledge distillation, employing past model predictions as soft labels for earlier tasks. Prior-focused methods [52] estimate distributions over model parameters to guide learning on new data. A notable approach is Elastic Weight Consolidation (EWC) [12], which calculates a Fisher matrix to assess the significance of each parameter and imposes stronger constraints on more crucial parameters. Finally, parameter isolation methods allocate separate model parameters for each task, effectively preventing the forgetting of knowledge. The model either grows new branches for new tasks [11] or uses a static structure [53] with different parts for each task. These often require a task classifier and are mainly used in multi-head structures.

Distinct from these strategies, our innovative approach introduces a fresh continual learning scheme tailored for single-image deraining, diverging from conventional replay-based and regularization-based methods.

C. Prompt for Transfer Learning

Prompt-based learning represents a new development in transfer learning, where pre-trained models are conditioned using specific, handcrafted functions [54], [55]. This technique not only guides models with extra instructions [54], [55] but also simplifies handling various tasks. Designing effective prompting functions is a key challenge, often requiring heuristic methods considering the complexity.

New research promotes adding learnable parameters to prompts [54], [56], enhancing their adaptability and evolution within transfer learning. Innovations like prompt tuning [56]

Authorized licensed use limited to: Peking University. Downloaded on July 03,2026 at 08:42:56 UTC from IEEE Xplore. Restrictions apply.

and prefix tuning [55] highlight this field, boosting model adaptability and performance.

Prompts stand out for their efficiency, capturing complex information with minimal architectural complexity. They are more parameter-efficient compared to other transfer learning techniques like Adapter [57] and LoRA [58], although these methods also perform well in various tasks.

Recently, prompt-based techniques have surpassed traditional transfer learning. As shown by Wang et al. [16], integrating prompt methods into continual learning marks a new phase of innovation and research.

In our research, we introduce prompt learning to develop a continual method for rain removal. Unlike recent prompt-based continual learning methods such as CODA-Prompt [19] and DualPrompt [21] which rely primarily on static prompt selection or simple orthogonality constraints. EcoDPL achieves superior adaptability and reduces forgetting through a tailored dual-level prompting strategy. Specifically, EcoDPL's image-level prompts preserve detailed local rain streak structures directly at the pixel-level, while its feature-level prompts capture higher-level semantic context. Moreover, the proposed P-Fuser module dynamically adjusts prompt contributions via attention maps conditioned explicitly on input features, ensuring prompt fusion is sensitive and adaptive to varying rain patterns. In contrast to prior orthogonality-only dictionary compression methods, EcoDPL's Grad-Tuner progressively refines prompts, actively freeing up model capacity for new tasks. These combined innovations enable EcoDPL to better manage complex, evolving rain scenarios.

III. Evolving Compact Dual Prompt Pools Learning

A. Motivation

In continual learning, traditional methods often struggle with being computationally expensive and complex. Prompt learning, known for its success in transfer learning and adapting models, offers a potentially desirable solution. Using task-specific prompts allows for better continual learning of features. This motivates us to explore further along the route.

We aim to handle various rain streak types by using a single model, augmented by a fixed number of additional parameters as prompts. We regard rain removal as a task that needs a high-low joint optimization, which can be obtained from the modeling of the rain removal problem. So, we propose a dual prompt learning method that works at both image and feature levels. This approach not only addresses various rain streaks but also includes an efficient prompt selection network for determining dynamic conditions.

Unlike traditional prompt learning approaches, which focus too much on task identities and can not distinguish specific and general knowledge, our method bridges this gap with proposed gradually tuned dictionary learning to focus on keeping prior knowledge while reducing the size of learned prompts, ensuring enough capacity for new knowledge.

Finally, to further improve the performance, we combine the model parameters and learnable prompts to regularize as a whole and reduce the learning rate with a pre-determined time schedule.

B. Definition and Configuration

The overall training pipeline of our proposed EcoDPL is illustrated in Fig. 2. We define a series of T tasks for removing rain from images, denoted as $\mathcal{D} = \{D_1, \cdots, D_T\}$. In this series, the t-th task, represented by $D_t = \{(x_i^t, y_i^t)\}_{i=1}^{n_t}$, involves n_t image pairs. Each pair is composed of a rain-affected image $x_i^t \in \mathcal{X}$ and its corresponding clean image $y_i^t \in \mathcal{Y}$. Our aim is to create a model $f(\cdot|\theta_{model}) : \mathcal{X} \rightarrow \mathcal{Y}$, in which θ_{model} denotes parameters of the model. Here, $\mathcal{X}$ symbolizes the domain of images with rain, and $\mathcal{Y}$ symbolizes the domain of clear images. Notably, the data from $D_1, D_2, \cdots, D_n$ are not accessible during the training of D_{n+1}. For any input image $x \in \mathbb{R}^{C\times H\times W}$, where C is the count of channels and $H \times W$ represents the dimensions of the image, our method employs a transformer-based structure for the task of deraining, represented as $f = f_r \circ f_e$. In this model, f_e serves as the preliminary embedding layer for the input images, and f_r consists of a series of self-attention layers, culminating in a specialized network designed for rain removal.

C. Learning With Image-Level Prompts

As shown in the detailed architecture in Fig. 3. To address the challenge of increasing space needs in replay-based techniques and to improve the efficiency of sharing image-level knowledge among different tasks, we incorporate image-level prompt learning into our continual learning framework.

The specific mechanism of prompt selection is depicted in Fig. 4. We define $\mathcal{P} = \{P_1, \cdots, P_M\}$ as an image prompt pool, which includes a fixed number of M prompts $P_i = (K_i, V_i)$. In this scenario, $V_i \in \mathbb{R}^{C\times H\times W}$ signifies a value that matches the dimensions of the input image, and the key $K_i = f_e(V_i)$ is the embedded value corresponding to the key of the input.

As mentioned in Sec. III-A, the input x and its embedded form $x_e = f_e(x)$ indicate, respectively, the original input and its embedding. We deliberately exclude the task index t from x in our representation, highlighting the flexibility of our method to operate without the need for defining task identity.

The input x instance dictates the computation of the ultimate prompt by the weighted summation of the prompt components in the pool. The fused prompt $P_f = (K_f, V_f)$ is obtained by:

$$P_f = \sum_{m=1}^{M} \alpha_m P_m. \tag{1}$$

The weights $\{\alpha_1, \alpha_2, \cdots, \alpha_M\} \in R$ for each prompt component are derived from the following equations:

$$\begin{aligned} &\{\alpha_1, \alpha_2, \cdots, \alpha_M\} \\ &= \gamma(f_e(\boldsymbol{x}) \odot A, K) \\ &= \{\gamma(f_e(\boldsymbol{x}) \odot A_1, K_1), \ldots, \gamma(f_e(x) \odot A_M, K_M)\}, \end{aligned} \tag{2}$$

where $A \in \mathbb{R}^{D\times M}$ contains M attention vectors $A_1, A_2, \cdots, A_M \in \mathbb{R}^D$with learnable parameters corresponding to image prompt components, $\odot$ represents the element-wise product and D represents the embedding dimension, γ represents the cosine distance.

During the model's training for task n, an input x undergoes a process where we first generate a unified prompt as

Authorized licensed use limited to: Peking University. Downloaded on July 03,2026 at 08:42:56 UTC from IEEE Xplore. Restrictions apply.

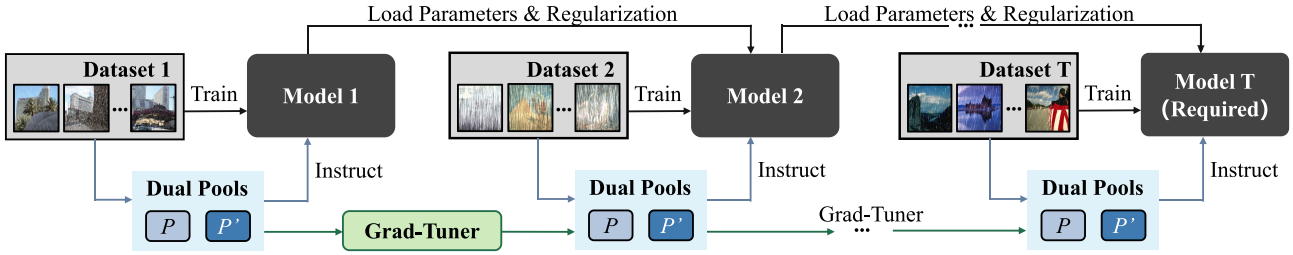


Fig. 2. The training process of our proposed EcoDPL for continual image rain removal. For each task, we train our model with a unique dataset. This dataset acts as a query, instructing two prompt pools to create fusion prompts that instruct the output of the model. We train the model and these dual prompt pools simultaneously. Additionally, the model of each task applies regularization constraints to the next task's model. After training on all the tasks, we get the required model.

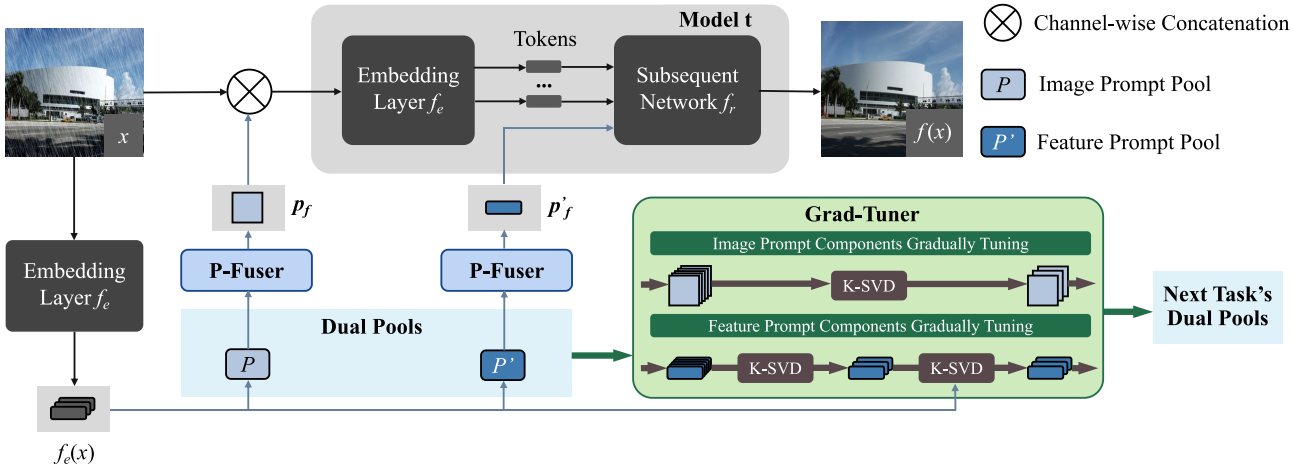


Fig. 3. We use a P-Fuser to combine image prompt pools and feature prompt pools, creating fused prompts. These prompts are then concatenated with input x along the channel dimension, and with embedded $f_e(x)$ along the token length dimension, to instruct the model. After training, we use a Grad-tuner to get dual prompt pools for the next task. For reducing the size of the learned image prompt components, a k-SVD process is used. For feature prompt components, we apply a gradually tuned dictionary learning method. This method compresses the feature prompts first and then fine-tunes them using features gathered during training.

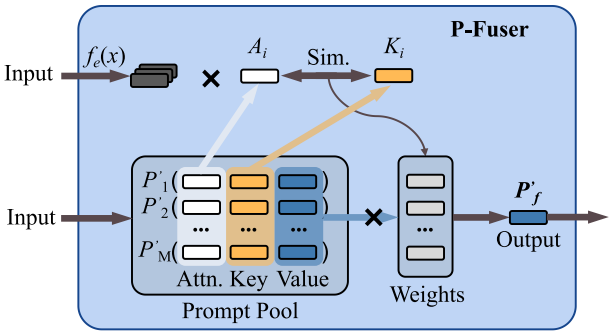


Fig. 4. The detail of prompt pools is shown. First, the input is multiplied by attention maps, and the result is compared with prompt keys to measure similarity. Second, based on these comparisons, weights are calculated. These weights are then used to perform a weighted average of the prompt components, creating a set of combined prompts. The combined prompts from both the input and feature prompt pool are added to the input and embedding tokens, and then processed together in the embedding layer for improved training.

previously described. This prompt is then combined with the input by matching their channel dimensions. Subsequently, this merged data is fed into the embedding layer. These input-sized prompts are designed to seamlessly incorporate knowledge with the input throughout the training process of our model.

$$x_1 = [V_f; x], \tag{3}$$

where [;] indicates combining across the channel dimension.

D. Learning With Feature-Level Prompts

With an input x and our transformer model $f = f_r \circ f_e$, the layer $f_e : \mathbb{R}^{C\times H\times W} \to \mathbb{R}^{L\times D}$ transforms image patches into embedding features $x'_e = f_e(x_1) \in \mathbb{R}^{L\times D}$. Here, L denotes the token length.

At the feature level, we keep a collection of prompt components, denoted by $\mathcal{P}' = \{P'_1, P'_2, \cdots, P'_M\}$. Each component $P'_i = (K'_i, V'_i)$ exists in the space $\mathbb{R}^{L\times D}$ for $i = 1, 2, \cdots, M$. In addition, for task n, we keep a table labeled $Q'_n = [q'^n_1, q'^n_2, \cdots, q'^n_M]$ that tracks the normalized frequencies of each prompt component P'_i.

To compute the ultimate feature-level prompt $P'_f = (K'_f, V'_f)$ for a given input, a weighted sum of these prompt components is performed:

$$P'_f = \sum_{m=1}^{M} \alpha'_m P'_m. \tag{4}$$

Authorized licensed use limited to: Peking University. Downloaded on July 03,2026 at 08:42:56 UTC from IEEE Xplore. Restrictions apply.

The weights $\{\alpha'_1, \alpha'_2, \cdots, \alpha'_M\}$ capture the relationship between the input's feature embedding and each prompt component using the cosine distance function γ:

$$\begin{aligned}&\{\alpha'_1, \alpha'_2, \cdots, \alpha'_M\}\\&= \gamma\left(f_e(x_1) \odot A', K'\right)\\&= \left\{\gamma\left(f_e(x_1) \odot A'_1, K'_1\right), \ldots, \gamma\left(f_e(x_1) \odot A'_M, K'_M\right)\right\}. \end{aligned} \quad (5)$$

$A' \in \mathbb{R}^{D\times M}$ contains M attention vectors $\{A'_1, A'_2, \cdots, A'_M\} \in \mathbb{R}^D$, with learnable parameters that match feature prompt components, $\odot$ represents the element-wise multiplication.

Given this formulation, during task n, for an input x, the final feature-level prompt is computed. Once obtained, this prompt is concatenated with the embedded input along the token length dimension, forming a coherent feature representation:

$$x_2 = [V'_f; f_e(x_1)]. \quad (6)$$

In this equation, x_1 symbolizes the synthesized input, as outlined in Sec. III-B. The combined feature x_2 progresses to the rain-removal network for additional training. As with Sec. III-B, these prompts are kept trainable, enabling the model to enhance its prompt-based representation continuously.

E. Gradually Tuned Dictionary Learning

In our framework, we adopt a progressively refined dictionary learning approach. Initially, we extract the prompts $\mathcal{P}$ using the k-SVD approach, where the dictionary D_1 is constructed to optimally approximate X_1. This is defined as the subsequent optimization:

$$\arg\min_{X_1} \|\mathcal{P} - D_1X_1\|_F^2, \text{subject to} \|x_{i,j}\|_0 \le T', \forall i, j. \quad (7)$$

The term $\|\cdot\|_F$ denotes the Frobenius norm, and T' is a predetermined upper limit. Given D_1, the goal is to identify the optimal coefficients that approximate the prompts $\mathcal{P}$. The coefficients are determined using the following objective:

$$\arg\min_{d_j, x_j} \|\mathcal{P} - \sum_{i\neq j} d_i x_i - d_j x_j\|_2^2, \text{subject to} \|d_j\|_2 = 1. \quad (8)$$

Subsequently, we employ the k-SVD approach for further improvement of the dictionary. By gathering a collection of features, during the training process and revisiting the k-SVD approach, we construct $\mathcal{P}$ to better approximate X_2:

$$\arg\min_{X_2} \|\mathcal{P} - D_2X_2\|_F^2, \text{subject to} \|x_{i,j}\|_0 \le T', \forall i, j. \quad (9)$$

Once D_2 is established, we aim to refine the coefficients to optimally represent $\mathcal{P}$:

$$\arg\min_{d_j, x_j} \|\mathcal{P} - \sum_{i\neq j} d_i x_i - d_j x_j\|_2^2, \text{subject to} \|d_j\|_2 = 1. \quad (10)$$

This gradually tuned dictionary refinement process improves representations of our prompts at each stage and compresses its size by iteratively evolving. The novel two-step dictionary learning approach called Grad-Tuner gradually compresses knowledge into fewer prompts over time. This method differs significantly from prior works, such as CODA-Prompt [19], which rely solely on orthogonality constraints. Grad-Tuner's progressive compression not only helps in reducing model complexity but also ensures high flexibility when learning new tasks, as shown in our experimental results.

F. Parameter-Prompts Joint Regularization

When the neural network undergoes sequential training on Task n and Task $n+1$, our aim is for the network to retain its performance on Task n. To accomplish this, regularization is simultaneously applied to both the model parameters and prompts. For Task n, the parameters of the rain removal model are represented by $\theta^n_{model} = \{\theta^n_1, \theta^n_2, \cdots, \theta^n_r\}$, where r is the network's depth. The set of parameters for all prompts is denoted as $\theta^n_{prompt} = \{\theta'^n_1, \theta'^n_2, \cdots, \theta'^n_s\}$, where s is the total count of parameters associated with these prompts. We define $\theta^n = \theta^n_{model} \cup \theta^n_{prompt}$. For simplicity, θ^n can be expressed as $\{\theta^n_1, \theta^n_2, \cdots, \theta^n_{r+s}\}$. X^n and Y^n represent the sets of rainy images and clean images for Task n, respectively. Assuming (x, y) is a pair of rainy and clean images, when processed by the network, the decline in performance for Task n due to training on Task $n+1$ is quantifiable as:

$$\begin{aligned}&\Delta f(\theta^{n+1}, \theta^n, x, y)\\&= \left|f(x, \theta^n), f\left(x, \theta^{n+1}\right)\right|\\&\triangleq \left|l(f(x, \theta^n), y) - l\left(f\left(x, \theta^{n+1}\right), y\right)\right|, \end{aligned} \quad (11)$$

where $|\cdot|$ is the absolute value operator and l is the total training loss for the deraining network. Regarding the k-th depth parameter for Task n, represented as θ^n_k, we define

$$\delta\theta^n_k = \theta^{n+1}_k - \theta^n_k. \quad (12)$$

To calculate $\Delta f\left(\theta^{n+1}, \theta^n, x, y\right)$, we use a Taylor series expansion of $l(f(x, \theta), y)$ around θ^{n+1}_k. This series is composed of infinitely many terms, each represented as derivatives of the target functions at a single point:

$$\begin{aligned}l\left(f\left(x, \theta^{n+1}_k\right), y\right) = l\left(f\left(x, \theta^n_k\right), y\right) &+ \left(\nabla_{\theta^n_k} l\right)^T \cdot \delta\theta^n_k\\&+ \frac{1}{2}\left(\delta\theta^n_k\right)^T \cdot H \cdot \left(\delta\theta^n_k\right)\\&+ O\left\|\left(\delta\theta^n_k\right)^3\right\|, \end{aligned} \quad (13)$$

where $H = \nabla^2_{\theta^n_k} l\left(\theta^n_k, x\right)$ represents the Hessian matrix for θ^n_k. To reduce the storage, we take an approximation $H \approx 2JJ^T$ for the Hessian matrix, where J is the Jacobian matrix. [59] This approximation, commonly known as the Gauss-Newton approximation, ignores the second-order derivative terms of the loss. It is theoretically justified when the model is near a local minimum (where residuals are small) and is widely adopted to avoid the prohibitive computational cost of calculating the full Hessian. In practice, we omit terms higher than third order. Consequently, we obtain:

$$\begin{aligned}I\left(\theta^{n+1}\right) &= \Delta f\left(\theta^{n+1}, \theta^n, x, y\right)\\&= \sum_{k=1}^{r+s}\left[l\left(f\left(x, \theta^{n+1}_k\right), y\right) - l\left(f\left(x, \theta^n_k\right), y\right)\right]\\&= \sum_{k=1}^{r+s}\left[\left(\nabla_{\theta^n_k} l\right)^T \cdot \delta\theta^n_k + \frac{1}{2}\left(\delta\theta^n_k\right)^T \cdot H \cdot \left(\delta\theta^n_k\right)\right]. \end{aligned} \quad (14)$$

Authorized licensed use limited to: Peking University. Downloaded on July 03,2026 at 08:42:56 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1 EcoDPL

Require: The network, named f, operates with two prompt sets: $\mathcal{P}$ for image-based prompts and $\mathcal{P}'$ for feature-based prompts, universally guided by parameter $\theta = \theta_{model} \cup \theta_{prompt}$. Each set contains M elements, P_i in $\mathcal{P}$ and P'_i in $\mathcal{P}'$, with each prompt defined by two components: K_i, V_i for P_i and K'_i, V'_i for P'_i. For any given task, such as the t-th task, it employs two key frequency sets: Q_t with elements $q_1^t, q_2^t, \ldots, q_M^t$ and F'_t with elements $q_1^{t'}, q_2^{t'}, \ldots, q_M^{t'}$, alongside learnable attention maps A for image prompts and A' for feature prompts. The training for the t-th task spans E_t epochs, across tasks from 1 to T, with a defined learning rate r and a loss function influenced by hyperparameters $\alpha, \beta, \zeta, \eta, \omega$. **Begin**

Initialize the pre-trained VGG-16 model and acquire the datasets $\mathcal{D} = \{D_1, D_2, \ldots, D_T\}$.
for $t = 1, 2, \cdots, T$ **do**
  Set Q_t and Q'_t to the real number 1. Begin with random prompt-sized items for initializing $\mathcal{P}$ and $\mathcal{P}'$. A and A' are initialized as random key-sized items of the respective prompt pools.
  for $e = 1, 2, \cdots, E_t$ **do**
    Extract a mini-batch $B = \left\{\left(x_i^t, y_i^t\right)\right\}_{i=1}^{l}$.
    for (x, y) in B **do**
      Compute the embedding input $x_e = f_e(x)$.
      Lookup the image weights by calculating Eqn. (2).
      Calculate the fused prompts by $P_f = \sum_m \alpha_m P_m$,
      Prepending x_1 with the fused prompts by $x_1 = [V_f; x]$. *// Image-level Augmentation (Sec. III-C)*
      Calculate the embedding input $f_e(x_1)$.
      Lookup the feature weights by calculating Eqn. (5).
      Calculate the fused prompts by $P'_f = \sum_m \alpha'_m P'_m$,
      Prepending x_2 with the fused prompts by $x_2 = [V'_f; f_e(x_1)]$. *// Feature-level Augmentation (Sec. III-D)*
      Calculate the rain-free result $f(x) = f_r(x_2)$.
      Calculate per sample loss by solving Eqn. (15).
      Update Q_t and Q'_t by incrementing the frequency-items corresponding to the chosen prompts by 1.
    end for
    Determine the per batch loss $\mathcal{L}_B$ by summing up $\mathcal{L}_x$.
    Update $\mathcal{P}$ and $\mathcal{P}'$.
  end for
  Update θ.
  Calculate D_1 by solving Eqn. (7).
  Approximate X_1 by minimizing Eqn. (8).
  Initialize D_2 by D_1.
  Calculate D_2 by solving Eqn. (9).
  Approximate X_2 by minimizing Eqn. (10).
end for $= 0$
End

A smaller $I\left(\theta^{n+1}\right)$ results in better retention of Task n's knowledge and serves as an effective regularization during training on Task $n+1$. To fine-tune our model, we use a learning rate decay strategy. Initially, we set a higher learning rate for quick convergence. As training advances, we gradually reduce the learning rate for more precise navigation through the loss landscape. This method helps the optimizer avoid early local minima while enabling detailed exploration of flatter areas in later stages of training.

G. Optimization Objective Function

At each training step, once the fused prompts are obtained as previously described, we input the combined input x_1 into the embedding layer, then select and concatenate N feature prompts. The resulting adapted feature x_2 undergoes processing by the next layers. Additionally, the optimization includes penalties for prompt distance and a combined penalty for both parameters and prompts. Overall, our objective for Task n is to minimize the comprehensive end-to-end training loss:

$$\begin{aligned}\mathcal{L}_x = \min_{\mathcal{P},\mathcal{P}',\theta} & \alpha L_1(f(x), y) + \beta L_p(f(x), y) \\ & + \zeta \sum_{\mathcal{P}_x} \gamma\left(f_e\left(P_{s_i}\right), f_e(x)\right) \cdot q_{s_i}^n \\ & + \eta \sum_{\mathcal{P}'_x} \gamma\left(P'_{s'_i}, f_e(x)\right) \cdot q_{s'_i}^{n'} + \omega I\left(\theta^n\right). \end{aligned} \tag{15}$$

Here, L_1 and L_p stand for the smooth L_1 loss and the perceptual loss, respectively. The third and fourth terms are surrogate losses, aimed at better aligning selected prompts with their corresponding query features. The fifth term refers to the parameter penalty previously mentioned in Eqn. (14). The coefficients $\alpha, \beta, \zeta, \eta, \omega$ adjust the importance of each term in the equation.

IV. Experiment Results

To thoroughly assess our proposed continual learning approach, we combine it with the cutting-edge rain removal baseline, TransWeather [9]. Our EcoDPL method is tested on various popular rain removal datasets, consistently surpassing other current continual learning methods in performance. Specifically, results from all tests indicate that our approach is highly effective in adapting to new tasks and simultaneously maintains strong performance on previous tasks. Qualitative comparisons are provided in Fig. 5, which visually demonstrate that our method restores rain-free images with clearer details and fewer artifacts compared to other baselines.

A. Dataset and Evaluation Metrics

Our continual learning method is evaluated by using three well-known rain removal datasets: Rain100H [39], Rain100L [39], and Rain800 [37]. Initially, our model undergoes training on Rain800 (Task 0) and then proceeds to Rain100H (Task 1), termed as the setting of the Rain800-Rain100H sequence. In addition to this, we also explore another sequence, Rain800-Rain100L. Rain100H and Rain100L datasets each comprise 1,800 pairs of rainy and clean images for training, accompanied by 100 pairs for testing. Rain800 has 600 training images and an additional 200 images for testing. To evaluate the performance of our model, we rely on the metrics of the Peak-Signal-to-Noise Ratio (PSNR) and Structural SIMilarity (SSIM).

B. Training Details

To ensure a fair comparison, we maintain consistency in parameter settings and training methods with the original paper. The distance query function γ is defined as cosine distance.

In each prompt pool, the value of M is established at 100. This setting is empirically determined to balance performance and computational efficiency, as demonstrated by the ablation

Authorized licensed use limited to: Peking University. Downloaded on July 03,2026 at 08:42:56 UTC from IEEE Xplore. Restrictions apply.

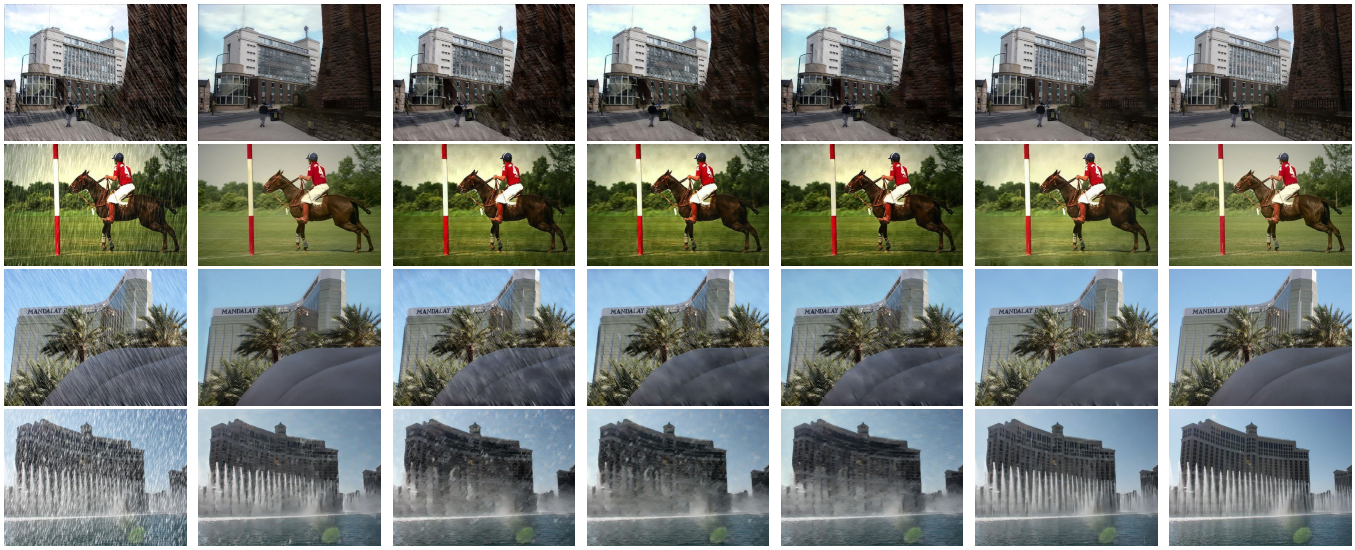


Fig. 5. Visual comparisons of results from rain streak removal using different continual learning methods compared with a baseline. (a) Input: Images with rain from the Rain800 dataset. (b) Task 0: The model is put through both training and evaluation on the Rain800 dataset. (c) Task 1 with SI (Sequential Independent): The model is trained on Rain800 followed by Rain100H, without interlinking the tasks, and then tested on Rain800. (d) Task 1 with replay: Training sequentially on Rain800 and Rain100H with rehearsal techniques, followed by testing on Rain800. (e) Task 1 with PIGWM: The model is trained on both the Rain800 and Rain100H datasets, employing the parameter regularization-based method, PIGWM. (f) Task 1 with EcoDPL: Training proceeds sequentially on Rain800 and Rain100H, utilizing EcoDPL, and then the model is tested on Rain800. (g) GT: This represents the clean, non-rainy image, serving as a ground truth for comparison.

TABLE I
COMPARING THE PERFORMANCE AND THE ADDED COMPLEXITY OF PARAMETERS IN CONTINUAL LEARNING METHODS

	Replay	Parameter Regularization (PR)	Dual Prompt (DP)	DP + PR
Performance (PSNR/SSIM)	22.76/0.8136	22.48/0.8058	22.96/0.8198	25.18/0.8441
Increased Para. Ratio	100% + 4%× tasks	100%	5.2%	105.2%

study on prompt numbers in Section IV-H (Table XI), where performance peaks at this value. For hyperparameters, we set α at 1, β at 0.04, ω at 0.95, and both ζ and η are fixed at $1e-5$. A training period of 50 epochs is applied to each task.

C. Results on Benchmark Datasets

We demonstrate the efficiency of our proposed continual learning algorithm through a range of qualitative and quantitative evaluations, on the previously mentioned datasets and performance measures. Table II and Table V present a detailed comparison among baseline models, classic continual learning approaches, and our method, revealing that our approach more effectively prevents catastrophic forgetting across various datasets. As illustrated in Table II, EcoDPL not only performs well in continual learning scenarios but also single dataset applications compared to the baseline (PSNR/SSIM: 27.52 dB/0.8667), showing the effectiveness of our design strategy. To formalize the analysis of plasticity, we introduce the Normalized Plasticity Score (S_p), calculated as the ratio of the model's performance on the new task to that of an Upper Bound model trained exclusively on it. As detailed in Table VI, the Upper Bound model trained solely on Rain100H achieves a PSNR of 27.62 dB and an SSIM of 0.8595. While previous methods like PIGWM and DPL show limited plasticity with scores around 0.88–0.91, EcoDPL attains remarkable scores of 0.98 for PSNR and 0.97 for SSIM. This confirms that EcoDPL can adapt to novel tasks almost as effectively as a dedicated model, exhibiting superior stability-plasticity trade-off. As indicated in Table I, the introduction of dual prompt pools significantly reduces parameter complexity compared to traditional continual learning methods, adding only an additional 5.2% in parameters. Through the implementation of parameter regularization to prompts, our approach offers more solution choices across various contexts. Utilizing only a dual prompt pool results in performance metrics comparable to those achieved through parameter regularization, although with a decrease in computational complexity. Moreover, the integration of parameter regularization can yield further performance improvements when higher computational complexity is permitted.

D. Comparison With Deraining Methods Under Continual Learning

To demonstrate the robustness and general applicability of our continual learning approach, we extend our evaluation to several state-of-the-art deraining backbones, namely TransWeather, Restormer and PromptIR. We compare various continual learning methods including Sequential

Authorized licensed use limited to: Peking University. Downloaded on July 03,2026 at 08:42:56 UTC from IEEE Xplore. Restrictions apply.

TABLE II

THE COMPARISON EMPHASIZES PSNR AND SSIM METRICS.MODELS ARE TRAINED SEQUENTIALLY ON THE RAIN800-RAIN100H TASK SEQUENCE WITH VARIOUS CONTINUAL LEARNING TECHNIQUES. IN CONTRAST, THE BASELINE MODEL RECEIVES TRAINING SOLELY ON RAIN800. TO ENSURE CONSISTENCY IN EVALUATION, ALL TESTS ARE CONDUCTED ON RAIN800

Methods	Training on Rain800-Rain100H		
	PSNR	SSIM	Degradation on Rain800
Baseline (only Rain800)	26.63	0.8583	0, 0
Ours (only Rain800)	**27.52**	**0.8667**	-0.89, -0.0084
SI	19.87	0.6451	6.76, 0.2132
EWC	21.64	0.7962	4.99, 0.0621
Replay	22.76	0.8136	3.87, 0.0447
Deep generative	22.51	0.8162	4.12, 0.0421
PIGWM	22.48	0.8058	4.15, 0.0525
L2P	22.54	0.7818	4.09, 0.0765
CODA-Prompt	24.54	0.8251	2.09, 0.0332
DualPrompt	22.28	0.7806	4.35, 0.0777
HiDe-Prompt	21.33	0.7731	5.30, 0.0852
DPL	24.39	0.8365	2.24, 0.0218
EcoDPL	**25.18**	**0.8441**	**1.45, 0.0142**

Independent (SI), EWC, PIGWM, DPL, and our proposed EcoDPL across these backbones, trained sequentially on the Rain800-Rain100H dataset sequence. The baseline denotes models trained only on Rain800.

Both Restormer and PromptIR are robust and powerful deraining backbones, though they involve substantially larger model sizes. Nevertheless, employing EcoDPL consistently yields the best performance in mitigating catastrophic forgetting, demonstrating superior PSNR and SSIM values on the initial training dataset (Rain800) across all backbones. The detailed results can be found in Table III.

Comparison with the Conference Version: As an extension of our preliminary conference work DPL [17], EcoDPL demonstrates a significant performance leap. Taking the Rain800-Rain100H task as an example (Table II), EcoDPL surpasses DPL by **0.79 dB** in PSNR (25.18 dB vs. 24.39 dB) on the old task. This enhancement stems from two core innovations: 1) **Grad-Tuner**: unlike DPL which relies on direct prompt accumulation, EcoDPL employs dictionary learning to compress knowledge. This yields a more compact representation, reducing redundancy and reserving more capacity to mitigate catastrophic forgetting. 2) **P-Fuser**: while DPL utilizes simple cosine distance for prompt selection, EcoDPL introduces learnable attention maps to generate adaptive weights. This mechanism allows the model to select prompts more precisely based on local context, enhancing robustness against rain streak perturbations.

E. Visual Results on the Realistic Rain Scene

We show more visual results of different methods whose models are trained sequentially on task sequence Rain800-Rain100H and tested on RID&RIS. RID (Rain in Driving) is applied to automatic driving and RIS (Rain in Surveillance) is applied to video surveillance. Images from both datasets are collected from real scenes, in line with the real application scenario. As illustrated in Fig. (6), our method exhibits significant advantages. We further evaluate EcoDPL on real-world IVIPC_DQA [60] and DNN-SIRR [42] datasets, comparing it with recent prompt-based methods (Fig. 7). While classification-oriented methods (e.g., L2P, CODA-Prompt) struggle with complex rain patterns, EcoDPL effectively restores clean images with clear structures, verifying its robust generalization across diverse domains. In addition to visual comparisons, we introduce non-reference image quality assessment metrics to quantitatively evaluate performance on real-world data, including NIQE [61], NIMA [62], and MANIQA [63], as reported in Table IV. EcoDPL achieves best results.

This superior generalization to real-world scenarios is largely attributed to our dual-prompt design, which effectively bridges the domain gap between synthetic training data and real-world inputs. Specifically, the *feature-level prompts* capture high-level semantic abstractions that are robust to variations in rain appearance, preventing the model from overfitting to specific synthetic rain patterns. Simultaneously, the *image-level prompts*, modulated by the adaptive P-Fuser, operate directly in the pixel space to accommodate the complex and non-uniform spatial distributions typical of real rain. Together, this dual mechanism enables the model to access a more diverse and adaptable feature space than methods relying on static parameters, thereby significantly enhancing practical utility.

F. The Effectiveness of Gradually Tuned Dictionary Learning

We present the experimental results demonstrating the effectiveness of our Grad-Tuner on prompts. To provide a comprehensive understanding of the capability of our Grad-Tuner, we visualize both the original and the learned prompt components. The visual comparison is displayed in the supplementary. The gray region represents the original prompts. The red region represents the prompts after the first step of our dictionary learning method. They demonstrate a degree of separability, with the blue region indicating the prompts after undergoing our dictionary learning process. A clear observation from the visualization is the superior aggregation in the learned prompts compared to their original form, which means that we have obtained a more compact and centralized representation, resulting in fewer prompts required during the reconstruction process.

We then analyze the sparsity and reconstruction fidelity of prompts about the k-SVD dictionary learning method. We begin by computing the original prompt components and subsequently subject them to k-SVD for dictionary learning. A further refinement is then conducted using a second iteration of k-SVD feature adjustment. Sparsity is evaluated by counting non-zero elements in these components. For reconstruction accuracy, we utilize the prompt components derived post-k-SVD to reconstruct the original ones and compute the related reconstruction loss.

Our experiments span different dictionary sizes, specifically 5, 25, 50, 75, and 100, as shown in Fig. 8. The results reveal

Authorized licensed use limited to: Peking University. Downloaded on July 03,2026 at 08:42:56 UTC from IEEE Xplore. Restrictions apply.

TABLE III

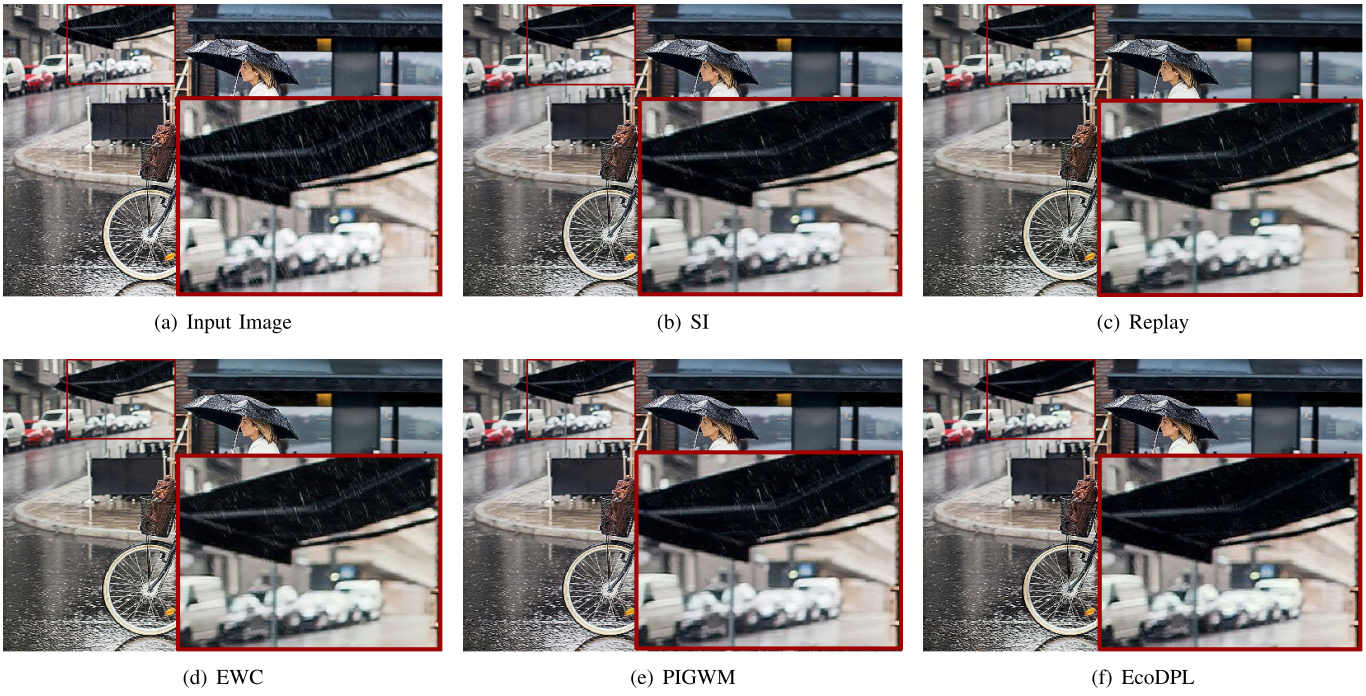
COMPARISON OF PSNR AND SSIM METRICS ACROSS DIFFERENT CONTINUAL LEARNING STRATEGIES AND STATE-OF-THE-ART DERAINING BACKBONES. MODELS ARE TRAINED SEQUENTIALLY ON THE RAIN800-RAIN100H TASK SEQUENCE AND EVALUATED ON THE INITIAL RAIN800 DATASET. THE BASELINE COLUMN REPRESENTS THE PERFORMANCE OF THE MODEL TRAINED SOLELY ON RAIN800, SERVING AS THE UPPER BOUND

Deraining Backbones	Baseline	Continual Learning Methods				
		SI	EWC	PIGWM	DPL	**EcoDPL**
TransWeather [9]	26.63 / 0.8583	19.87 / 0.6451	21.64 / 0.7962	22.48 / 0.8058	24.39 / 0.8365	**25.18 / 0.8441**
Restormer [43]	32.12 / 0.9006	24.97 / 0.8058	30.15 / 0.8824	30.90 / 0.8817	31.04 / 0.8835	**31.17 / 0.8842**
PromptIR [46]	30.95 / 0.9052	31.17 / 0.8842	29.59/ 0.8980	29.19 / 0.8966	28.27 / 0.8913	**30.11 / 0.9077**

TABLE IV

QUANTITATIVE COMPARISON ON REAL-WORLD DATASETS USING NON-REFERENCE METRICS. ↓ INDICATES LOWER IS BETTER (NIQE), WHILE ↑ INDICATES HIGHER IS BETTER (NIMA, MANIQA). BOLD INDICATES THE BEST PERFORMANCE

Methods	**IVIPC_DQA** [61]			**DNN-SIRR** [42]		
	NIQE ↓	NIMA ↑	MANIQA ↑	NIQE ↓	NIMA ↑	MANIQA ↑
L2P	4.38	5.06	0.50	4.54	4.85	0.54
HierPrompt	4.17	5.04	0.46	4.59	4.76	0.52
DualPrompt	4.21	5.06	0.50	4.35	4.84	0.54
CODA-Prompt	4.36	**5.14**	0.54	4.57	4.84	0.57
EcoDPL (Ours)	**3.98**	5.13	**0.57**	**4.18**	**4.92**	**0.60**



(a) Input Image (b) SI (c) Replay

(d) EWC (e) PIGWM (f) EcoDPL

Fig. 6. Visual results of different methods trained sequentially on Rain800 and Rain100H, and tested on a dataset of real scenarios, RID&RIS, of which the rainy images do not have corresponding backgrounds. (a) Input: rainy images from RID&RIS; (b) SI: rain removal model trained with rehearsal; (d) EWC: rain removal model trained with EWC; (e) PIGWM: rain removal model trained with parameter regularization; (f) EcoDPL: rain removal model trained with evolving compact dual prompt learning;.

a consistent trend: as the number of atoms increases, there is a reduction in sparsity, evidenced by an increase in non-zero elements. Correspondingly, there is a noticeable decrease in reconstruction loss, indicating that a denser dictionary yields more accurate prompt reconstructions. The results emphasize the inherent trade-off between sparsity and accuracy in dictionary learning.

G. Ablation Study

1) Ablation Study of Each Component in the Loss Function: In this section, ablation analyses are conducted to thoroughly evaluate the impact of each component in Equation 15. The results, presented in Table VII, highlight that using two sets of prompts significantly helps in addressing catastrophic forgetting. The visual results are displayed in the supplementary.

Authorized licensed use limited to: Peking University. Downloaded on July 03,2026 at 08:42:56 UTC from IEEE Xplore. Restrictions apply.

TABLE V

QUANTITATIVE EVALUATION RESULTS IN PSNR AND SSIM. MODELS ARE TRAINED SEQUENTIALLY ON THE RAIN800-RAIN100L SEQUENCE THROUGH CONTINUAL LEARNING TECHNIQUES, FOLLOWED BY TESTING ON RAIN800

Methods	Training on Rain800-Rain100L		
	PSNR	SSIM	Degradation on Rain800
Baseline	26.63	0.8583	0, 0
SI	20.42	0.5823	6.21, 0.2760
EWC	23.11	0.7840	3.52, 0.0743
Replay	23.20	0.7758	3.43, 0.0825
Deep generative	22.25	0.7462	4.38, 0.1121
PIGWM	23.98	0.8049	2.65, 0.0534
DPL	24.79	**0.8382**	1.84, **0.0201**
EcoDPL	**25.44**	0.8366	**1.19**, 0.0217

TABLE VI

QUANTITATIVE EVALUATION OF PLASTICITY ON THE NEW TASK (RAIN100H). MODELS ARE TRAINED SEQUENTIALLY ON RAIN800-RAIN100H AND EVALUATED ON RAIN100H. THE NORMALIZED PLASTICITY SCORE (S_p) IS THE RATIO OF THE METHOD'S PERFORMANCE TO THE UPPER BOUND (TRAINED EXCLUSIVELY ON RAIN100H).

Methods	Performance		Norm. Plasticity Score (S_p)	
	PSNR	SSIM	S_p (PSNR)	S_p (SSIM)
Baseline (Rain800)	17.93	0.4868	0.65	0.57
PIGWM	24.13	0.7736	0.87	0.90
DPL	24.38	0.7847	0.88	0.91
EcoDPL	**26.94**	**0.8347**	**0.98**	**0.97**
Upper Bound (Rain100H)	*27.62*	*0.8595*	*1.00*	*1.00*

TABLE VII

ABLATION STUDY FOR OPTIMIZATION OBJECTIVE FUNCTION

Training on Rain800-Rain100H, Testing on Rain800					
Image Prompts	Feature Prompts	Parameter Reg.	Prompts Reg.	PSNR	SSIM
✓	✓	✓	✓	25.18	0.8441
×	✓	✓	✓	24.29	0.8313
✓	×	✓	✓	23.10	0.8219
✓	✓	×	✓	23.77	0.8360
✓	✓	✓	×	23.86	0.8333

TABLE VIII

ABLATION STUDY FOR MODULES IN ECODPL

Training on Rain800-Rain100H, Testing on Rain800					
Prompt Components	Attention Vectors	k-SVD	Learning Rate Decay	PSNR	SSIM
✓	✓	✓	✓	25.18	0.8441
×	✓	✓	✓	23.17	0.8176
✓	×	✓	✓	24.89	0.8371
✓	✓	×	✓	24.74	0.8394
✓	✓	✓	×	25.00	0.8424

Additionally, feature-level prompts emerge as particular critical elements, while the application of parameter regularization techniques offers supplementary enhancements to our model's overall performance. Furthermore, the features of the images with the prompt pools are notably cleaner compared to those of the images without the addition of the prompt pools. The visual results are also shown in the supplementary.

TABLE IX

THE QUANTITATIVE PERFORMANCE FOR THE EFFECTIVENESS OF THE FREQUENCY TABLES

Training on Rain800-Rain100H, Testing on Rain800			
Image Prompts Frequency Table	Feature Prompts Frequency Table	PSNR	SSIM
✓	✓	**25.18**	**0.8441**
×	✓	23.47	0.8180
✓	×	24.25	0.8333
×	×	24.58	0.8309

TABLE X

THE QUANTITATIVE PERFORMANCE FOR DIFFERENT WEIGHTS OF COSINE DISTANCE

Training on Rain800-Rain100H, Testing on Rain800			
Image \ Feature	10^{-5}	10^{-4}	10^{-3}
10^{-5}	24.53/0.8365	23.75/0.8279	24.37/0.8374
10^{-4}	25.00/0.8423	**25.18/0.8441**	24.03/0.8322
10^{-3}	24.48/0.8368	22.93/0.8300	22.68/0.8225

TABLE XI

THE QUANTITATIVE PERFORMANCE FOR DIFFERENT NUMBERS OF THE PROMPTS

Training on Rain800-Rain100H, Testing on Rain800				
The number of prompts in a pool	PSNR in Task 0	SSIM in Task 0	PSNR in Task 1	SSIM in Task 1
50	26.68	0.8622	23.25	0.8278
75	26.80	0.8659	24.06	0.8350
100	**27.82**	**0.8696**	**25.18**	**0.8441**
125	27.71	0.8680	24.86	0.8407
150	27.64	0.8680	24.11	0.8330
175	27.25	0.8651	23.88	0.8286
200	27.06	0.8638	23.92	0.8300

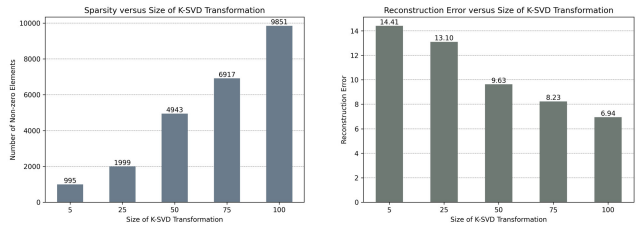
2) Ablation Study of Modules in EcoDPL: We conduct an ablation study to meticulously evaluate the impact of various components within the EcoDPL framework. The study, as detailed in Table VIII, focuses on the model trained on Rain800-Rain100H and tested on Rain800. The results highlight the significance of each module in our method. The integration of all modules, including Prompt Components, Attention Vectors, k-SVD, and Learning Rate Decay, yields the highest PSNR and SSIM scores, highlighting their importance in optimizing the model's effectiveness.

3) Ablation Study of the Frequency Table: We evaluate the quantitative performance of our model using dual prompts for the sequential Rain800-Rain100H datasets, with and without the frequency tables. As depicted in Table IX, it is apparent that the utilization of frequency tables results in improved performance in both scenarios. In particular, the frequency table of the image prompts is particularly important.

4) Ablation Study of the Weight of Cosine Distance: In Table X, upon evaluating the weight of cosine distance in the

Authorized licensed use limited to: Peking University. Downloaded on July 03,2026 at 08:42:56 UTC from IEEE Xplore. Restrictions apply.

Fig. 7. Visual comparisons on real-world scenes from IVIPC_DQA [60] (top two rows) and DNN-SIRR [42] (bottom two rows). Compared to recent methods (b-e) [16], [19], [20], [21], our EcoDPL (f) achieves superior rain removal and detail preservation.



(a) (b)

Fig. 8. Sparsity versus size of k-SVD transformation. Reconstruction error versus size of k-SVD transformation.

optimization target, it is determined that a value of 10^{-4} is a suitable choice. Deviating from this value, either by increasing or decreasing it, results in a degradation of performance.

5) Ablation Study of the Number of the Prompts: We evaluate the quantitative performance of our model using dual prompts for the sequential Rain800-Rain100H datasets, with different numbers of prompt components in a prompt pool. As depicted in Table XI, with the increase in the number of prompts, the performance of the model first increases and then decreases, reaching a peak at 100. Subsequently, the performance begins to decline when the number of prompts increases to 125.

6) Ablation Study on k-SVD Transformation With Different Sizes: We evaluate the quantitative performance of our model using dual prompts for the sequential Rain800-Rain100H datasets, with different sizes of k-SVD transformation. As shown in Table XII, with the increase of the size of k-SVD transformation, the performance of the model first increases and then decreases, reaching a peak at about 25.

TABLE XII
THE QUANTITATIVE PERFORMANCE FOR DIFFERENT SIZES OF k-SVD TRANSFORMATION

Training on Rain800-Rain100H, Testing on Rain800		
The size of k-SVD transformation	PSNR in Task 1	SSIM in Task 1
5	24.22	0.8340
25	**25.18**	**0.8441**
50	24.31	0.8356
75	24.28	0.8355
100	24.79	0.8372

TABLE XIII
PSNR/SSIM RESULTS TRAINED ON SEQUENTIAL TASKS RAIN800-RAIN100H-RAIN100L

Method	Testing Set		
	Rain800	Rain100H	Rain100L
SI	21.30/0.7152	16.53/0.5946	36.96/0.9800
Replay	22.77/0.7758	18.78/0.6861	32.12/0.9541
PIGWM	22.80/0.7564	17.44/0.6331	32.58/0.9561
DPL	24.44/0.8163	19.63/0.7282	31.93/0.9589
EcoDPL	**24.79/0.8201**	**20.26/0.7400**	**33.47/0.9619**
Reference	26.63/0.8583	28.49/0.8802	37.97/0.9825

TABLE XIV
DETAILED PERFORMANCE EVOLUTION (PSNR/SSIM) AT THE END OF EACH TRAINING STAGE FOR THE THREE-TASK SEQUENCE (RAIN800 → RAIN100H → RAIN100L). BOLD INDICATES THE PERFORMANCE ON THE TASK CURRENTLY BEING LEARNED (PLASTICITY). COLUMNS REPRESENT THE TEST DATASETS, AND ROWS REPRESENT THE MODEL STATE AFTER TRAINING ON A SPECIFIC STAGE

Model State (After Training on)	Test Datasets		
	Rain800 (Task 1)	Rain100H (Task 2)	Rain100L (Task 3)
Stage 1 (Rain800)	**27.99 / 0.8736**	17.48 / 0.5005	27.85 / 0.8756
Stage 2 (Rain100H)	22.67 / 0.7971	**31.42 / 0.8997**	32.87 / 0.9445
Stage 3 (Rain100L)	24.79 / 0.8201	20.26 / 0.7400	**33.47 / 0.9619**

H. Generalization Across Multiple Datasets

In this section, we demonstrate that our approach is not restricted to two task scenarios but can be seamlessly extended to handle multiple tasks. Specifically, given a sequence of n tasks, our model undergoes continual training, transitioning from the first task to the n-th task sequentially.

We consider a scenario involving three tasks. Initially, the model is trained on the first two tasks utilizing our proposed methodology. Subsequently, the same pre-trained model serves as the foundation for the training on Task 3. Table XIV presents the results of an experiment involving sequential training on Rain800, Rain100H, and Rain100L datasets. The results confirm the superiority of our approach.

Table XIV details the step-by-step performance evolution across the three stages. After Stage 1, the model performs well on Rain800 but poorly on the unseen Rain100H. Upon

Authorized licensed use limited to: Peking University. Downloaded on July 03,2026 at 08:42:56 UTC from IEEE Xplore. Restrictions apply.

TABLE XV

QUANTITATIVE COMPARISON (PSNR/SSIM) ON THE IMAGE DENOISING TASK SEQUENCE (GAUSSIAN → POISSON). MODELS ARE TRAINED SEQUENTIALLY AND EVALUATED ON THE FIRST TASK (GAUSSIAN NOISE)

Method	PSNR	SSIM
SI	25.59	0.7540
PIGWM	26.03	0.7644
EcoDPL (Ours)	**29.91**	**0.9339**

TABLE XVI

QUANTITATIVE RESULTS OF THE TASK-FREE EXTENSION OF ECODPL, TRAINED SEQUENTIALLY ON RAIN800 →RAINTRAINH→RAINTRAINL WITHOUT TASK IDENTITY.

Task-Free Continual Training, No Task Identity Provided		
Test Set	PSNR	SSIM
Rain800	23.49	0.8007
RainTestH	21.56	0.6987
RainTestL	31.24	0.9414

completing Stage 2 (Rain100H), the model achieves high plasticity on the new task (31.42 dB) while experiencing expected forgetting on Rain800. However, interestingly, after Stage 3 (Rain100L), we observe a phenomenon of *positive backward transfer*: the performance on the oldest task (Rain800) recovers from 22.67 dB to 24.79 dB. This suggests that the knowledge learned from Rain100L shares commonalities with Rain800, which our compact prompts successfully leveraged to repair previous forgetting.

I. Extension to Other Low-Level Tasks

To further validate the claim that our EcoDPL framework is effective for general low-level vision tasks beyond deraining, we conducted an additional experiment on image denoising. We set up a continual learning sequence involving two distinct noise types: Gaussian Noise followed by Poisson Noise. The model was trained on the ImageNet dataset with synthetic noise and evaluated on the Gaussian Noise subset of the Common528 dataset [64] (containing 11 degradation types). As shown in Table XV, while methods like SI and PIGWM suffer from forgetting (PSNR ≈ 25-26 dB), EcoDPL maintains superior performance on the initial task with a PSNR of **29.91 dB** and SSIM of **0.9339**. This result confirms that our dual-prompt design, which handles spatially variant degradations, generalizes well to other restoration tasks, effectively mitigating catastrophic forgetting in broader low-level vision scenarios.

Ablation Study on Task-Free Extension: While EcoDPL relies on explicit task boundaries by default, it can be extended to task-free settings by replacing the boundary trigger with an EWMA-based loss monitoring mechanism, which automatically detects pseudo-boundaries when the loss exceeds $\mu + 3\sigma$ without any task identity. As shown in Table XVI, this task-free variant achieves competitive performance across all three datasets.

V. CONCLUSION

In this work, we introduce EcoDPL, a novel continual learning framework designed specifically for single-image deraining applications. This method uses compact prompts, which are small, trainable parameters, to direct a transformer-based model in removing rain at both image and feature levels, allowing it to transfer knowledge among various tasks. We design a P-Fuser to assign a learnable attention map to each prompt, enabling our model to select information according to the inputs. Additionally, we apply a dictionary learning method, Grad-Tuner, to compress prompts over time, providing more space for prompts to learn new knowledge. Extensive experiments on various rain datasets confirm the effectiveness of our proposed EcoDPL. Furthermore, this approach is adaptable enough to integrate into training processes for lower-level tasks, enhancing its usefulness and adaptability in a range of challenging situations.

VI. LIMITATIONS AND FUTURE WORKS

While EcoDPL demonstrates robust performance, it is worth noting a characteristic of its design: as a plugin-style framework, its absolute upper bound for restoration quality is inherently tied to the representational capacity of the underlying frozen backbone (e.g., TransWeather). Although this dependency limits the model's ability to exceed the backbone's original theoretical maximum, it simultaneously highlights a significant advantage: EcoDPL is highly modular and model-agnostic. This implies that our framework can be seamlessly integrated with stronger, future restoration architectures to unlock even greater potential without redesigning the continual learning mechanism.

REFERENCES

[1] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis.*, Jun. 2020, pp. 213–229.

[2] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with Atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2018, pp. 801–818.

[3] L. Wang, J. Zhang, O. Wang, Z. Lin, and H. Lu, "SDC-Depth: Semantic divide-and-conquer network for monocular depth estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 538–547.

[4] 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.

[5] D. Ren, W. Zuo, Q. Hu, P. Zhu, and D. Meng, "Progressive image deraining networks: A better and simpler baseline," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3932–3941.

[6] D. Ren, W. Shang, P. Zhu, Q. Hu, D. Meng, and W. Zuo, "Single image deraining using bilateral recurrent network," *IEEE Trans. Image Process.*, vol. 29, pp. 6852–6863, 2020.

[7] Y. Wei et al., "DerainCycleGAN: Rain attentive CycleGAN for single image deraining and rainmaking," *IEEE Trans. Image Process.*, vol. 30, pp. 4788–4801, 2021.

[8] K. Jiang et al., "Rain-free and residue hand-in-hand: A progressive coupled network for real-time image deraining," *IEEE Trans. Image Process.*, vol. 30, pp. 7404–7418, 2021.

[9] 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. 2353–2363.

Authorized licensed use limited to: Peking University. Downloaded on July 03,2026 at 08:42:56 UTC from IEEE Xplore. Restrictions apply.

[10] J. Zhang et al., "Class-incremental learning via deep model consolidation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 1120–1129.

[11] J. Xu and Z. Zhu, "Reinforced continual learning," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 899–908.

[12] J. Kirkpatrick et al., "Overcoming catastrophic forgetting in neural networks," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 13, pp. 3521–3526, 2017.

[13] Z. Li and D. Hoiem, "Learning without forgetting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 2935–2947, Dec. 2018.

[14] F. Zenke, B. Poole, and S. Ganguli, "Continual learning through synaptic intelligence," in *Proc. Int. Conf. Machine Learn.*, 2017, pp. 3987–3995.

[15] M. Jia et al., "Visual prompt tuning," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2022, pp. 709–727.

[16] Z. Wang et al., "Learning to prompt for continual learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 139–149.

[17] M. Liu, W. Yang, Y. Hu, and J. Liu, "Dual prompt learning for continual rain removal from single images," in *Proc. 32nd Int. Joint Conf. Artif. Intell.*, 2023, pp. 7215–7223.

[18] M. Zhou et al., "Image de-raining via continual learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 4905–4914.

[19] S. J. Seale et al., "CODA-Prompt: Continual decomposed attention-based prompting for rehearsal-free continual learning," in *Proc. IEEE Int. Conf. Computer Vis. Pattern Recognit.*, Jun. 2022, pp. 11909–11919.

[20] L. Wang, J. Xie, X. Zhang, M. Huang, H. Su, and J. Zhu, "Hierarchical decomposition of prompt-based continual learning: Rethinking obscured sub-optimality," in *Proc. Adv. Neural Inf. Process. Syst.*, 2023, pp. 69054–69076.

[21] Z. Wang et al., "DualPrompt: Complementary prompting for rehearsal-free continual learning," in *Proc. IEEE Eur. Conf. Computer Vis.*, Jun. 2022, pp. 631–648.

[22] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing," *ACM Comput. Surveys*, vol. 55, no. 9, pp. 1–35, 2022.

[23] C. Raffel et al., "Exploring the limits of transfer learning with a unified text-to-text transformer," *J. Mach. Learn. Res.*, vol. 21, no. 140, pp. 1–67, 2019.

[24] J. Xu, W. Zhao, P. Liu, and X. Tang, "Removing rain and snow in a single image using guided filter," in *Proc. IEEE Int. Conf. Comput. Sci. Autom. Eng.*, Jun. 2012, pp. 304–307.

[25] X. Zheng, Y. Liao, W. Guo, X. Fu, and X. Ding, "Single-image-based rain and snow removal using multi-guided filter," in *Proc. Int. Conf. Neural Info. Process.*, Berlin, Germany, 2013, pp. 258–265.

[26] J.-H. Kim, C. Lee, J.-Y. Sim, and C.-S. Kim, "Single-image deraining using an adaptive nonlocal means filter," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 914–917.

[27] 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.

[28] P. Mu, J. Chen, R. Liu, X. Fan, and Z. Luo, "Learning bilevel layer priors for single image rain streaks removal," *IEEE Signal Process. Lett.*, vol. 26, no. 2, pp. 307–311, Feb. 2019.

[29] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2808–2817.

[30] Y.-H. Fu, L.-W. Kang, C.-W. Lin, and C.-T. Hsu, "Single-frame-based rain removal via image decomposition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 1453–1456.

[31] D.-Y. Chen, C.-C. Chen, and L.-W. Kang, "Visual depth guided color image rain streaks removal using sparse coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 8, pp. 1430–1455, Aug. 2014.

[32] L.-W. Kang, C.-W. Lin, and Y.-H. Fu, "Automatic single-image-based rain streaks removal via image decomposition," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1742–1755, Apr. 2012.

[33] D. Eigen, D. Krishnan, and R. Fergus, "Restoring an image taken through a window covered with dirt or rain," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 633–640.

[34] 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.

[35] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley, "Clearing the skies: A deep network architecture for single-image rain removal," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2944–2956, Jun. 2017.

[36] Z. Fan, H. Wu, X. Fu, Y. Hunag, and X. Ding, "Residual-guide feature fusion network for single image deraining," 2018, *arXiv:1804.07493*.

[37] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3943–3956, Nov. 2020.

[38] 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.

[39] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan, "Deep joint rain detection and removal from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1685–1694.

[40] G. Li, X. He, W. Zhang, H. Chang, L. Dong, and L. Lin, "Non-locally enhanced encoder–decoder network for single image de-raining," in *Proc. 26th ACM Int. Conf. Multimedia*, Oct. 2018, pp. 1056–1064.

[41] T. Wang, X. Yang, K. Xu, S. Chen, Q. Zhang, and R. W. H. Lau, "Spatial attentive single-image deraining with a high quality real rain dataset," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12270–12279.

[42] W. Wei, D. Meng, Q. Zhao, Z. Xu, and Y. Wu, "Semi-supervised transfer learning for image rain removal," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3872–3881.

[43] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5718–5729.

[44] J. Xiao, X. Fu, A. Liu, F. Wu, and Z.-J. Zha, "Image de-raining transformer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 11, pp. 12978–12995, Nov. 2023.

[45] 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.*, Jun. 2023, pp. 5896–5905.

[46] S. Khan, V. Potlapalli, F. Shahbaz Khan, and S. W. Zamir, "PromptIR: Prompting for all-in-one image restoration," in *Proc. Adv. Neural Inf. Process. Syst.*, 2023, pp. 71275–71293.

[47] X. Fu, J. Xiao, Y. Zhu, A. Liu, F. Wu, and Z.-J. Zha, "Continual image deraining with hypergraph convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 8, pp. 9534–9551, Aug. 2023.

[48] A. Robins, "Catastrophic forgetting, rehearsal and pseudorehearsal," *Connection Sci.*, vol. 7, no. 2, pp. 123–146, 1995.

[49] R. M. French, "Pseudo-recurrent connectionist networks: An approach to the'sensitivity-stability'dilemma," *Connection Sci.*, vol. 9, no. 4, pp. 353–380, 1997.

[50] A. Chaudhry et al., "Continual learning with tiny episodic memories," in *Proc. Workshop Multi-Task Lifelong Reinforcement Learn.*, 2019, pp. 1–13.

[51] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–18.

[52] C. Zeno, I. Golan, E. Hoffer, and D. Soudry, "Task agnostic continual learning using online variational Bayes," 2018, *arXiv:1803.10123*.

[53] J. Serra, D. Suris, M. Miron, and A. Karatzoglou, "Overcoming catastrophic forgetting with hard attention to the task," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 4548–4557.

[54] B. Lester, R. Al-Rfou, and N. Constant, "The power of scale for parameter-efficient prompt tuning," 2021, *arXiv:2104.08691*.

[55] X. L. Li and P. Liang, "Prefix-tuning: Optimizing continuous prompts for generation," in *Proc. 59th Annu. Meeting Assoc. Comput. Linguistics, 11th Int. Joint Conf. Natural Lang. Process.*, 2021, pp. 4582–4597.

[56] G. Qin and J. Eisner, "Learning how to ask: Querying lms with mixtures of soft prompts," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Human Lang. Technol.*, 2021, pp. 1–16.

[57] R. Wang et al., "K-adapter: Infusing knowledge into pre-trained models with adapters," in *Proc. Findings Assoc. Comput. Linguistics*, 2021, pp. 1405–1418.

[58] J. E. Hu et al., "LoRA: Low-rank adaptation of large language models," in *Proc. Int. Conf. Learning Represent.*, 2021, pp. 1–13.

[59] J. Nocedal and S. J. Wright, *Numerical Optimization*. Cham, Switzerland: Springer, 2006.

[60] Q. Wu, L. Wang, K. N. Ngan, H. Li, F. Meng, and L. Xu, "Subjective and objective de-raining quality assessment towards authentic rain image," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3883–3897, Nov. 2020.

[61] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.

Authorized licensed use limited to: Peking University. Downloaded on July 03,2026 at 08:42:56 UTC from IEEE Xplore. Restrictions apply.

[62] H. Talebi and P. Milanfar, "NIMA: Neural image assessment," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3998–4011, Aug. 2018.
[63] S. Yang et al., "MANIQA: Multi-dimension attention network for no-reference image quality assessment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 1191–1200.
[64] Y. Liu et al., "Unifying image processing as visual prompting question answering," 2023, *arXiv:2310.10513*.

Minghao Liu (Student Member, IEEE) is currently pursuing the B.S. degree in computer science with Peking University, Beijing, China. His current research interests include low-level computer vision and deep-learning-based image processing.

Wenhan Yang (Member, IEEE) received the B.S. degree and Ph.D. degree (Hons.) in computer science from Peking University, Beijing, China, in 2012 and 2018. He is currently an Associate Researcher with Pengcheng Laboratory, Shenzhen, Guangdong, China. His current research interests include image/video processing/restoration, bad weather restoration, human-machine collaborative coding. He has authored over 50 technical articles in refereed journals and proceedings, and holds 9 granted patents. He received the 2023 IEEE Multimedia Rising Star Runner-Up Award, the IEEE ICME-2020 Best Paper Award, the IFTC 2017 Best Paper Award, the IEEE CVPR-2018 UG2 Challenge First Runner-up Award, and the MSA-TC Best Paper Award of ISCAS 2022. He was the Candidate of CSIG Best Doctoral Dissertation Award in 2019. He served as the Area Chair of IEEE ICME-2021/2022/2023/2024, the Session Chair of IEEE ICME-2021, and the Organizer of IEEE CVPR-2019/2020/2021 UG2+Challenge and Workshop.

Jiaying Liu (Fellow, IEEE) received the Ph.D. degree (Hons.) in computer science from Peking University, Beijing China, 2010.

She is currently a Boya Distinguished Professor with the Wangxuan Institute of Computer Technology, Peking University, China. She has authored more than 100 technical articles in refereed journals and proceedings, and holds 70 granted patents. Her current research interests include multimedia signal processing, compression, and computer vision. She is a distinguished member of CCF and a senior member of CSIG. She was a visiting scholar with the University of Southern California, Los Angeles, CA, USA, from 2007 to 2008. She was a Visiting Researcher with Microsoft Research Asia, in 2015 supported by the Star Track Young Faculties Award. She has been a member of Multimedia Systems and Applications Technical Committee (MSA TC), and Visual Signal Processing and Communications Technical Committee (VSPC TC) in IEEE Circuits and Systems Society. She has served as ACM ICMR Steering Committee member and the CAS Representative at the IEEE ICME Steering Committee. She has also served as the Deputy Editor in Chief of APSIPA Trans. on Signal and Information Processing, the Associate Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON CIRCUITS SYSTEMS FOR VIDEO TECHNOLOGY, IEEE \textit{*Multimedia Magazine*}, and \textit{*Journal of Visual Communication and Image Representation*}, the General Chair of ACM MM Asia 2024, the Technical Program Chair of ACM MM Asia 2025/2023/IEEE ICME 2021/ACM ICMR 2021/IEEE VCIP 2019. She was the APSIPA Distinguished Lecturer (2016-2017).

Authorized licensed use limited to: Peking University. Downloaded on July 03,2026 at 08:42:56 UTC from IEEE Xplore. Restrictions apply.