Continuous Adverse Weather Removal via Degradation-Aware Distillation

Xin Lu Jie Xiao Yurui Zhu Xueyang Fu†
MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition,
School of Information Science and Technology, University of Science and Technology of China
luxion@mail.ustc.edu.cn, xyfu@ustc.edu.cn

Abstract

All-in-one models for adverse weather removal aim to process various degraded images using a single set of parameters, making them ideal for real-world scenarios. However, they encounter two main challenges: catastrophic forgetting and limited degradation awareness. The former causes the model to lose knowledge of previously learned scenarios, reducing its overall effectiveness. While the later hampers the model's ability to accurately identify and respond to specific types of degradation, limiting its performance across diverse adverse weather conditions. To address these issues, we introduce the Incremental Learning Adverse Weather Removal (ILAWR) framework, which uses a novel degradation-aware distillation strategy for continuous weather removal. Specifically, we first design a degradation-aware module that utilizes Fourier priors to capture a broad range of degradation features, effectively mitigating catastrophic forgetting in low-level visual tasks. Then, we implement multilateral distillation, which combines knowledge from multiple teacher models using an importance-guided aggregation approach. This enables the model to balance adaptation to new degradation types with the preservation of background details. Extensive experiments confirm that ILAWR outperforms existing models across multiple benchmarks, proving its effectiveness in continuous adverse weather removal.

1. Introduction

Adverse weather, including rain, snow, and haze, significantly reduces image quality, hindering critical vision tasks such as object tracking [39], detection [25], semantic segmentation [60], and face recognition [56]. This has driven research into single-image restoration techniques like deraining [12, 46], de-snowing [5, 17], and de-hazing [7, 36], largely using deep neural networks [9, 26, 28, 37, 38, 41, 48, 53, 55] that address specific types of degradation [47]. However, for applications like autonomous driving and surveillance, training dedicated networks for each degradation type is inefficient and impractical.

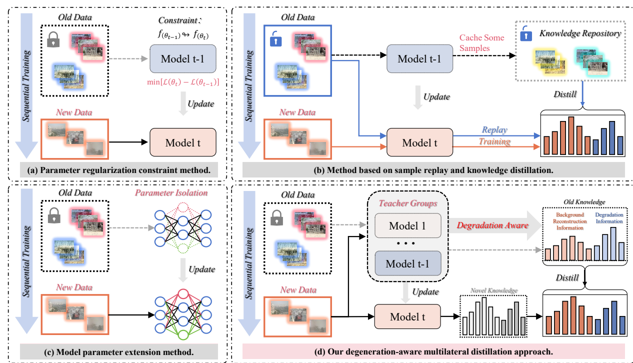


Figure 1. Comparison of our Degradation-Aware Multilateral Distillation (DAMD) approach proposed for low-level vision with previous incremental learning schemes. When compared to knowledge-caching sample replay [8], parameter isolation and expansion [13], and regularization [21, 59] methods, our approach showcases the enhanced restoration performance with decreased memory footprint and improved computational efficiency.

To address this, all-in-one solutions [35, 49, 61] use unified network parameters to handle various degradations simultaneously. These models, however, assume ideal training conditions where all possible degradation-clean mappings are included in the dataset and available at every training iteration. This assumption is unrealistic for real-world applications, where weather variations frequently change the distribution of degraded images, scattering the mapping rules across multiple datasets. This fragmentation complicates dataset assembly for diverse training, and continuous model retraining is computationally costly. On the other hand, training incrementally with new data risks catastrophic forgetting [27], where previously learned information is lost. Incremental learning, or continual learning [54], mitigates catastrophic forgetting and has mostly been used in classification tasks. Approaches to prevent forgetting include regularization [22, 52], replay mechanisms [3, 16], and parameter isolation [1]. The PIGWM model [59] first applied incremental learning to image de-raining, using parameter importance-guided regularization to maintain per-

† : Corresponding author.

formance across updates, allowing effective de-raining despite changing image degradation levels. Other advances, such as parameter expansion and hypergraph convolutional networks [13], have improved generalization for incremental de-raining. The CLAIO model [8] extended incremental weather removal using knowledge replay and distillation techniques on synthetic datasets. Despite these advancements, challenges persist: as illustrated in Figure 1, regularization-based method struggles with growing datasets, parameter expansion significantly increases memory and computation needs, and replay-based CLAIO requires large cache resources. These limitations emphasize the need for more efficient incremental learning methods for image restoration in dynamic conditions.

In this paper, we reassess adverse weather removal algorithms suited to current real-world applications and introduce the ***Incremental Learning Adverse Weather Removal*** (**ILAWR**) framework, which is based on Degradation-Aware Multilateral Distillation (DAMD). ILAWR learns incrementally from streams with varied degradation types, progressively integrating learned mappings without needing specific degradation category information. A comparison between DAMD and prior methods is presented in Figure 1, and implementation details of ILAWR are outlined in Figure 2. Inspired by lifelong learning with diverse teacher models [45], this framework enables adaptive knowledge acquisition from complementary "teachers", effectively balancing stability and adaptability.

Specifically, we developed a Degradation-Aware Module (DAM), based on Fourier priors, which isolates degradation-specific information within the channel dimensions of U-Net's deeper layers [61]. This module divides the model up to $session(t)$ into two teacher groups, as depicted in Figure 4. This structure allows the model to effectively learn diverse degradation cues while maintaining robustness in background reconstruction. Additionally, we introduced an Importance-Guided Aggregation Module (IGAM), which uses an importance-weight matrix to dynamically combine guidance from each teacher group, boosting the model's performance. ILAWR successfully demonstrates incremental learning on four synthetic datasets for rain, snow, haze, and raindrop degradations, using Rain100H [50], RESIDE [20], Snow100K [24], and Raindrop [34] datasets. We also pioneered the first incremental weather removal benchmark using four real-world datasets—SPA+ [61], REVIDE [57], RealSnow [61], and LOL-v2 [51]—enabling continuous removal of rain, snow, fog, and low-light artifacts.

The main contributions of this work include:

- We propose ILAWR, an effective and efficient solution tailored to real-world scenarios, addressing the challenges of incremental learning in adverse weather removal.
- Based on Fourier priors, we design DAM for incremental learning in low-level vision tasks. This module effectively extracts degradation-specific knowledge and background reconstruction knowledge from teacher models.
- We introduce a simple yet effective multilateral distillation approach that employs importance-guided aggregation to combine diverse degradation and reconstruction knowledge from multiple teacher models. This strategy significantly reduces catastrophic forgetting across various degradation types.

Figure 2. The pipeline of the proposed *Incremental Learning Adverse Weather Removal* (ILAWR). Our model learns under the paradigm of incremental learning, gathering data from diverse weather conditions, continuously updating network parameters. The final model $f(\theta_T)$ is capable of restoring various degraded images (gray arrow) to address real environmental variations.

Comprehensive experiments on both synthetic and real-world datasets demonstrate the superior performance of our method in handling incremental adverse weather removal, validating its effectiveness and robustness.

2. Related Work

Adverse Weather Removal. Images captured in nature are subject to degradation due to adverse weather conditions, prompting the development of numerous image restoration algorithms designed to remove effects of adverse weather such as rain, snow, and haze. The cutting-edge image restoration algorithms today are based on deep learning models. [12] first employ the Deep Detail Network (DetanilNet) for deraining. Additionally, advanced network architectures are increasingly applied to single-image deraining [48], desnowing [5], and dehazing tasks [7]. Single-task models are inadequate for real-world applications, leading to the emergence of integrated methods for multi-task image restoration [49]. These approaches utilize a unified set of network parameters for inference across various restoration tasks. As diverse degraded datasets are incrementally gathered from the natural world, models must continually learn new degradation-clean mapping rules. These strategies, bound by the single-shot supervised learning paradigm, face the challenge of costly retraining on all data versus the risk of catastrophic forgetting associated with incremental learning through streaming data input.

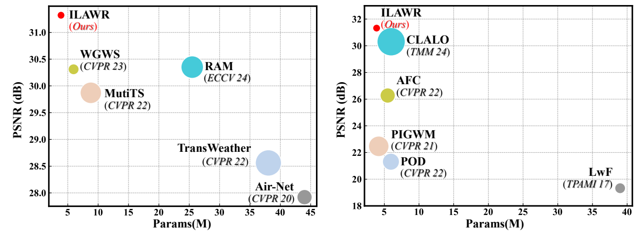
Incremental Learning Schemes. In recent years, an increasing number of studies have turned their attention to the practical deployment of image restoration models on edge devices. PIGWM [59] pioneered the introduction of incremental learning to deraining, using parameter importance in regularization methods to minimize performance gaps between model parameters, facilitating incremental deraining on images with diverse degradation levels. Moreover, techniques utilizing parameter expansion and hypergraph convolution networks [13] are applied for incremental learning in deraining and generalization tasks. AM [15] leverages an associative memory management approach for feature reconstruction in incremental deraining, while DPL [23] introduces feature-level prompt learning through dual hints input for task-invariant and task-specific image knowledge acquisition in the context of incremental deraining. The work closest to ours, CLAIO [8], achieves incremental weather removal on three synthetic datasets using knowledge replay and distillation. PIGWM's reliance on regularization fails to sustain performance as the incremental dataset expands. Techniques involving parameter expansion and hint learning amplify memory and computational overhead, while methods centered on associative memory and replay tend to squander substantial cache. Inspired by humans learning from various teachers throughout their lives, multi-teacher distillation models have been employed in the field of image classification [45, 58] in recent years, yielding superior performance.

3. Motivation

Adverse weather conditions like rain, snow, fog, and low light are primarily generated based on atmospheric optics involving various particles present in the air [30–32], exhibiting high levels of randomness and diversity [29]. Consequently, artificially curated paired datasets always fail to encompass all mapping rules for every degradation type at once. Faced with the continuous acquisition of small batches of data in real-world settings, an ideal restoration model should possess the following capabilities.

Stability and efficiency. Existing all-in-one image restoration models only support one-time training and fine-tuning, and to maintain good performance, they retain a large number of parameters. Due to the imbalance in information across degradation datasets acquired from different batches, this single mixed training approach exacerbates model overfitting on degradation types with larger datasets, while increasing costs, thus hindering the model's generalization to unknown degradation types. As shown in Figure 3, when there is significant variation in data quantity across different weather conditions, our incremental learning method even outperforms the previous state-of-the-art (SOTA) on average metrics, despite their mixed training mode.

Continuous learning capability. In a continuously changing environment, the restoration network should be able to continuously adapt and update in an incremental learning fashion without requiring substantial resources. As illustrated in Figure 1, existing incremental learning approaches rely on regularization constraints [21, 59], sample replay distillation [8], and model parameter expansion [13], lacking an effective incremental learning framework tailored for low-level visual design. Moreover, they also introduce poor memory overhead and computational efficiency.



(a) Comparison with other all-in-one methods. (b) Comparison with other incremental methods.

Figure 3. We simulated biased training data distributions continuously obtained by using training datasets Rain100H (3K), Snow100K (100K), and RESIDE (400K). The circle size indicates the FLOPs of various methods. ILAWR outperformed previous all-in-one and incremental learning models.

Our analysis inspired a novel degradation-aware multilateral distillation technique in Figure 1. This method employs continuous learning guided by knowledge from multiple teacher models, focusing on degradation types and background reconstruction. It optimizes performance without requiring sample replay or parameter expansion.

4. Methodology

As shown in Figure 2, ILAWR sets forth a flow of T incremental sessions as $\mathcal{D} = \{D_1, \cdots, D_T\}$, where each D_t=$\{(x_i^t, y_i^t)\}_{i=1}^{n_t}$ represents a specific adverse weather condition. In $session(t)$, n_t pairs of samples are collected for training, with pair comprising a clean image $y_i^t \in \mathcal{Y}$ and a degraded image $x_i^t \in \mathcal{X}$. $\{\mathcal{X}, \mathcal{Y}\}$ represent the domains of degraded and ground truth image data. Our aim is to train a model $f(\cdot|\theta_T) : \mathcal{X} \mapsto \mathcal{Y}$, with parameters θ_T derived from training in $session(T)$. Notably, during training on $session(t)$, old data $\{D_1, \cdots, D_{t-1}\}$ is unavailable, and final testing occurs across all domains in $\{D_1, \cdots, D_T\}$.

4.1. Overall Pipeline of ILAWR

We devise the Degradation-Aware Multilateral Distillation (DAMD) algorithm to realize ILAWR, training a compact network in an incremental learning paradigm for adverse weather removal tasks, as illustrated in Figure 4. Given an input image $x_i^t \in \mathbb{R}^{C \times H \times W}$, where C is the number of channels, and $H \times W$ is the size of the image. We construct a restoration network $f = \mathcal{P} \circ \psi$ based on the U-Net [61] framework, ψ and $\mathcal{P}$ denotes the feature projection and output projection. We employ Charbonnier loss [4] for content reconstruction, which is mathematically defined as follows:

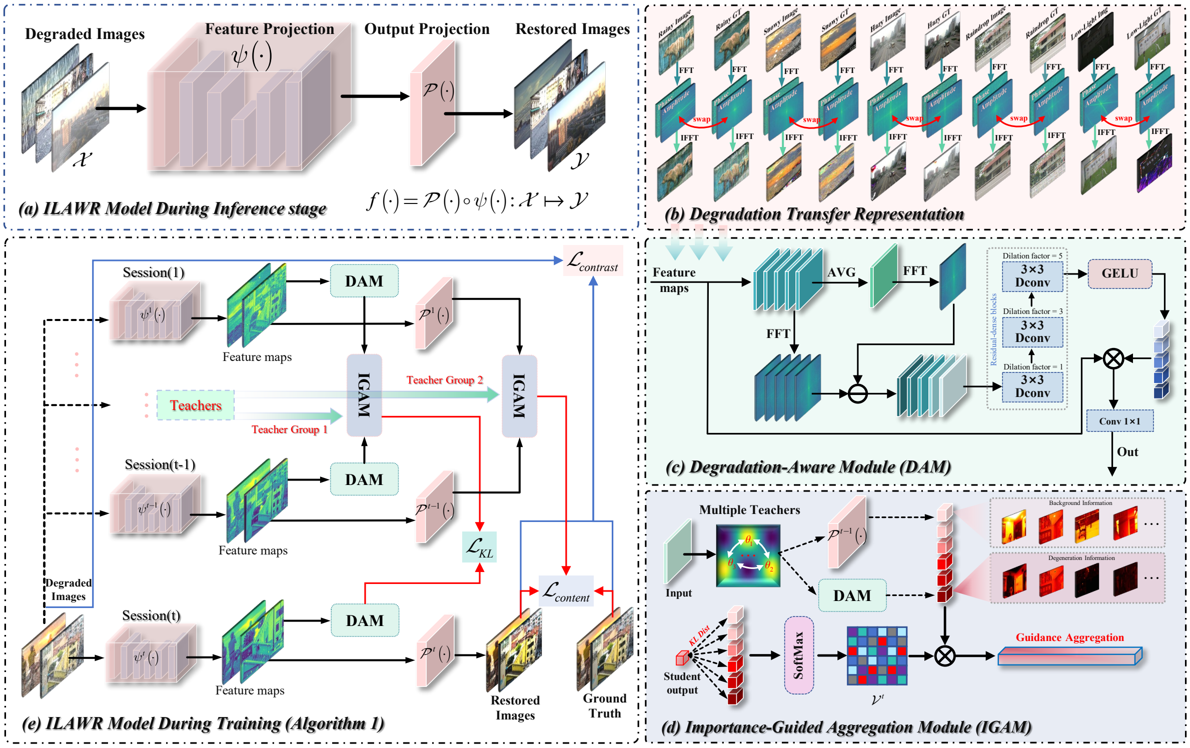


Figure 4. The overall illustration of our method. (a) Illustrates our network during inference stage. For simplicity, the global residual connections are omitted. (b) Depicts the degradation transfer phenomenon observed after exchanging the amplitude spectra of degraded image pairs using Fourier transform, inspiring the design of the Degradation-Aware Module. (c) Degradation-Aware Module (DAM). It modulates feature maps by leveraging the data distribution of the amplitude spectra obtained after Fourier transform to extract degradation information latent in the channel dimension. (d) Visualizes our proposed Importance-Guided Aggregation Module (IGAM) used to obtain the ultimate guidance from each set of teachers. (e) Implements the training process of ILAWR. Refer to Algorithm 1 for specific details.

$$\mathcal{L}_{content} = \frac{1}{n_t}\sum_{i=1}^{n_t}\sqrt{\left\|f^t\left(x_i^t\right)-y_i^t\right\|^2+\epsilon^2}, \tag{1}$$

where ϵ is seen as a tiny constant (*e.g.*, 10^{-5}) for stable and robust convergence. For the perception, we also use the contrast regularization [7, 8], which can be denoted as:

$$\mathcal{L}_{contrast}(f(x_i^t), y_i^t, x_i^t) = \\ -\sum_{i=1}^{n_t}\sum_{l=1}^{L} w_l \log \frac{e^{-\left|E_l(f(x_i^t))-E_l(y_i^t)\right|/\tau}}{e^{-\left|E_l(f(x_i^t))-E_l(y_i^t)\right|/\tau}+e^{-\left|E_l(f(x_i^t))-E_l(x_i^t)\right|/\tau}}, \tag{2}$$

where e denotes the exponential operation, E_l with $l = \{1, 2, \cdots, L\}$ extracts the l-th hidden layer features from the fixed pre-trained model VGG-19 [40], τ (> 0) is the temperature parameter that controls the sharpness of the output, $|\cdot|$ denotes the L_1 distance, and w_l is the weight coefficient for the l-th hidden feature from the fixed VGG-19 network. The value of parameter w_l and τ follow [7].

Therefore, we define the basic loss function for the single weather image restoration task as:

$$\mathcal{L}_{base} = \mathcal{L}_{content} + \alpha\mathcal{L}_{contrast}, \tag{3}$$

α is a hyperparameter used for training the restoration network f in all baseline methods, *i.e.*, for each individual adverse weather removal task and fine-tuning setup.

4.2. Degradation-Aware Distillation.

Given that the Fourier domain can characterize global features of images, an increasing number of recent works integrate frequency domain information to address low-level visual problems [44]. Given an image $x \in \mathbb{R}^{H\times W\times C}$, the Fourier transform $\mathcal{F}$ converts it to Fourier space as the complex component F(x), which is expressed as:

$$\mathcal{F}(x)(u,v) = \frac{1}{\sqrt{HW}}\sum_{h=0}^{H-1}\sum_{w=0}^{W-1} x(h,w)\, e^{-j2\pi\left(\frac{h}{H}u+\frac{w}{W}v\right)}, \tag{4}$$

where u and v indicate the coordinates of the Fourier space. $\mathcal{F}^{-1}(x)$ defines the inverse Fourier transform accordingly. Both the Fourier transform and its inverse procedure can be efficiently implemented using FFT/IFFT algorithms [11]. The amplitude component $\mathbf{A}(x)(u,v)$ and phase component $\mathbf{P}(x)(u,v)$ are expressed as:

$$\begin{aligned} \mathbf{A}(x)(u,v) &= \sqrt{\mathbf{R}^2(x)(u,v)+\mathbf{I}^2(x)(u,v)}, \\ \mathbf{P}(x)(u,v) &= \arctan\left[\frac{\mathbf{I}(x)(u,v)}{\mathbf{R}(x)(u,v)}\right], \end{aligned} \tag{5}$$

Algorithm 1 Degradation-Aware Multilateral Distillation

Input: Flows of T incremental datasets $\mathcal{D}=\{D_1,\cdots,D_T\}$, where $D_t=\left\{\left(x_i^t,y_i^t\right)\right\}_{i=1}^{n_t}$ represents n_t specific adverse weather degraded-clean image pairs. $x_i^t \in \mathcal{X}$ and $y_i^t \in \mathcal{Y}$. $\{\mathcal{X},\mathcal{Y}\}$ represent the domains of degraded and ground truth image data.

Parameter: Restoration network $\left\{f^1,\cdots,f^T\right\}$ parameterized by $\{\theta_1,\cdots,\theta_T\}$, where $f^t=\mathcal{P}^t\circ\psi^t$. $\mathcal{M}^t$ signifies the DAM, employed to extract deep-level degradation information from the network f^t in $Session(t)$, $\mathcal{I}^t$ represents IGAM, merging outputs from multiple teachers into aggregated guidance, E_t is number of training epochs, $\mathcal{L}_{KL}$ refers to the Kullback-Leibler (KL) divergence. α_1,α_2, ϵ, λ, ζ are the hyper-parameters.

Output: All-in-one image restoration model $f(\cdot|\theta):\mathcal{X}\mapsto\mathcal{Y}$

Load pre-trained model (VGG-19) for $\mathcal{L}_{contrast}$.
for $t=1,2,\cdots,T$ **do**
  Random sample a batch $\mathcal{X}_t=\left\{\left(x_i^t,y_i^t\right)\right\}_{i=1}^{R}$.
  for $e=1,2,\cdots,E_t$ **do**
    if $t>1$ **then**
      $Guide_1=\mathcal{I}\left(\sum_{i=1}^{t-1}\mathcal{M}^i\left(\psi^i\left(\mathcal{X}\right)\right)\right)$
      $\mathcal{L}_{teach_1}^t=\mathcal{L}_{KL}^t\left(Guide1,\mathcal{M}^{t-1}\left(\psi^t\left(\mathcal{X}\right)\right)\right)$
      $Guide_2=\mathcal{I}\left(\sum_{i=1}^{t-1}f^i\left(\mathcal{X}\right)\right)$
      $\mathcal{L}_{teach_2}^t=\mathcal{L}_{base_1}^t\left(Guide_2,f^t\left(\mathcal{X}\right),\mathcal{X}\right)$
      Train f: $\mathcal{L}_f^t=\mathcal{L}_{base_1}^t+\lambda\mathcal{L}_{teach_1}^t+\zeta\mathcal{L}_{teach_2}^t$[20]
      Update f^t: $\theta_t\leftarrow\theta_t+\epsilon\nabla\mathcal{L}_f^t$, $f^t.no\_grad()$
    else
      Train f: $\mathcal{L}_f^1=\mathcal{L}_{base_1}^1\left(f^1\left(\mathcal{X}\right),\mathcal{Y},\mathcal{X}\right)$
      Update f^1: $\theta_1\leftarrow\theta_1+\epsilon\nabla\mathcal{L}_f^1$, $f^1.no\_grad()$
    Train $\mathcal{M}$: $\mathcal{L}_{\mathcal{M}}^t=\mathcal{L}_{base_2}^1\left(\mathcal{P}\left(\mathcal{M}^t\left(\mathcal{X}\right)\right),\mathcal{Y},\mathcal{X}\right)$[21]
    Update $\mathcal{M}^t$: $\theta_{\mathcal{M}}\leftarrow\theta_{\mathcal{M}}+\epsilon\nabla\mathcal{L}_{\mathcal{M}}^t$, $\mathcal{M}^t.no\_grad()$
return Parameters θ_T of all-in-one adverse weather removal model $f^T:\mathcal{X}\mapsto\mathcal{Y}$ derived after training in $session(T)$

where $\mathbf{R}(x)(u,v)$ and $\mathbf{I}(x)(u,v)$ represent the real and imaginary parts respectively. The Fourier transform and inverse procedure are applied independently to each channel.

For the image restoration tasks in this study, we conducted Fourier transforms on image pairs corresponding to the 5 degradation types. By observing the characteristics of the spectrum plots as shown in Figure 4(b), through horizontal comparisons, we noted that the disparity in the frequency domain spectra between paired clean and degraded images primarily manifests in the amplitude. Additionally, images under different weather conditions exhibit significant differences in their amplitude spectra. Through vertical comparisons, it is evident that after swapping the amplitude spectra of paired degraded-clean images and subsequently restoring the RGB images via inverse Fourier transforms, corresponding degradation transfers are observed across all 5 weather conditions. Therefore, the amplitude spectra in the Fourier domain notably encapsulate more degradation information within the images.

Therefore, we devised the Degradation-Aware Module (DAM) as illustrated in Figure 4(c), utilizing the Fourier priors of images to aid the model in extracting degradation-specific information. Specifically, for the deep-level feature maps $\psi(x_i^t)\in\mathbb{R}^{C\times H\times W}$ of the network, our aim is to learn the particular degradation information contained within its channel dimensions. Firstly, we compute the channel-wise mean of f to obtain the general features across channels:

$$g=\mathrm{AVG}\left(\psi\left(x_i^t\right)\right)\in\mathbb{R}^{1\times H\times W}, \tag{6}$$

next, we perform Fourier transforms on both $\psi\left(x_i^t\right)$ and g, extracting the amplitude spectra:

$$\begin{aligned}\mathcal{A}_{avg}&\leftarrow\mathcal{F}\left(g\right)\in\mathbb{R}^{1\times H\times W},\\ \mathcal{A}_{initial}&\leftarrow\mathcal{F}\left(\psi\left(x_i^t\right)\right)\in\mathbb{R}^{C\times H\times W},\end{aligned} \tag{7}$$

we isolate the unique information along their channel dimensions as degradation-specific information:

$$\mathcal{A}_{specific}=\mathcal{A}_{initial}-\mathcal{A}_{avg}\left(\in\mathbb{R}^{C\times H\times W}\right). \tag{8}$$

Finally, we obtain the degradation-specific weight tensor:

$$\mathbf{w}=\mathrm{GELU}\left(\mathrm{RDBS}\left(\mathcal{A}_{specific}\right)\right), \tag{9}$$

where $\mathbf{w}\in\mathbb{R}^{C\times1\times1}$, RDBS consists of a series of dense residual connections of dilated convolution blocks, as shown in Figure 4(c). The model $\mathcal{M}$ of Degradation-Aware Module (DAM) is defined as follows:

$$\mathcal{M}\left(\psi\left(x_i^t\right)\right)=\mathrm{Conv}\left(\mathbf{w}\odot\psi\left(x_i^t\right)\right). \tag{10}$$

As depicted in Figure 4(e), the DAM, which extracts degradation-specific information, divides the teachers into two groups, guiding and constraining the student model, enhancing its capability to address different degradation cues while ensuring stability in image background reconstruction. The guidance in Teacher Group 1 is all generated by DAM. To reduce the distribution gap between the student and teacher models in extracting degradation information, we employ Kullback-Leibler (KL) divergence loss to implement distillation:

$$\begin{aligned}\mathcal{L}_{teach_1}^t&=\mathcal{L}_{KL}^t\left(Guide_1,\mathcal{M}_{student}\left(\psi^t\left(\mathcal{X}\right)\right)\right)\\ &=\mathrm{KL}\left[\mathcal{M}_{teacher}||\mathcal{M}_{student}\left(\psi^t\left(\mathcal{X}\right)\right)\right],\end{aligned} \tag{11}$$

where $\mathcal{M}$ refers to the DAM, and details regarding the training updates for model $\mathcal{M}$ can be found in Algorithm 1. The guidance in Teacher Group 2 is all generated by the restoration net f from each session:

$$Guide_2=f\left(\mathcal{X}\right)=\mathcal{P}\left(\psi\left(\mathcal{X}\right)\right), \tag{12}$$

we utilize the restoration results from the previous models as guidance to constrain the stability of the student in background reconstruction. In $session(t)$, the distillation loss employs a basic loss:

$$\begin{aligned}\mathcal{L}_{teach_2}^t&=\mathcal{L}_{base_1}^t\left(Guide_2,f^t\left(x_i^t\right),x_i^t\right)\\ &=\mathcal{L}_{content}^t+\alpha_1\mathcal{L}_{contrast}^t,\end{aligned} \tag{13}$$

where α_1 is a hyperparameter used in Teacher Group 2.

Table 1. **Haze→Rain→Snow** (Setting 1) results on synthetic datasets. The Individual Training result serves as a performance upper bound, and compared to Sequential Fine-tuning , our ILAWR effectively mitigates **catastrophic forgetting**. Compared to state-of-the-art all-in-one and incremental learning methods, our approach demonstrates a superior advantage of up to **0.62dB** (PSNR↑).

Method	Venue	Average		RESIDE [20]		Rain100H [50]		Snow 100K [24]	
		PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Individual Training	-	31.97	0.9357	31.77	0.9412	30.45	0.9345	33.68	0.9314
Sequential Fine-tuning	-	20.63	0.7282	16.20	0.8142	13.47	0.4541	32.21	0.9162
EWC [19]	PNAS 16	22.27	0.7192	28.33	0.9577	13.74	0.4024	24.73	0.7975
LwF [21]	TPAMI 17	19.32	0.6803	20.87	0.8544	13.03	0.3763	24.05	0.8101
MAS [2]	ECCV 18	23.03	0.7489	29.18	0.9661	16.68	0.5010	23.23	0.7797
POD [10]	ECCV 20	21.31	0.6988	20.49	0.8296	15.35	0.3988	28.08	0.8680
PIGWM [59]	CVPR 21	22.46	0.7214	28.93	0.9522	14.01	0.4162	24.45	0.7959
AFC [18]	CVPR 22	26.27	0.8163	26.99	0.9012	22.48	0.6567	29.34	0.8911
TransWeather [42]	CVPR 22	28.56	0.9064	28.52	0.9612	25.20	0.8333	31.95	0.9247
MutiTS [6]	CVPR 22	29.87	0.9237	29.96	0.9692	26.06	0.8606	33.59	0.9412
WGWS [61]	CVPR 23	30.31	0.9385	29.67	0.9724	28.29	0.9014	32.97	0.9417
RAM [35]	ECCV 24	30.35	0.9100	29.75	0.9439	28.31	0.8411	32.99	0.9451
CLAIO [8]	TMM 24	30.70	0.9334	31.03	0.9775	28.52	0.8900	32.54	0.9328
ILAWR	-	**31.32**	**0.9380**	**31.41**	**0.9774**	**28.91**	**0.8960**	**33.64**	**0.9405**

Table 2. **Haze→Rain→Snow→Raindrop** (Setting 2) results on synthetic datasets. Compared to state-of-the-art incremental learning and all-in-one methods, as continuous weather conditions increase to **4**, we demonstrate a more significant advantage (**0.85dB** PSNR↑).

Method	Venue	Average		RESIDE [20]		Rain100H [50]		Snow 100K [24]		Raindrop [34]	
		PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
TransWeather [42]	CVPR 22	27.29	0.8456	27.61	0.8439	23.67	0.8137	29.14	0.8854	28.73	0.8393
MutiTS [6]	CVPR 22	27.19	0.8509	27.84	0.8442	23.71	0.8213	28.69	0.8971	28.51	0.8408
WGWS [61]	CVPR 23	29.09	0.8722	28.57	0.8844	25.64	0.8467	30.83	0.9253	31.31	0.8322
RAM [35]	ECCV 24	29.05	0.8685	28.53	0.8771	24.99	0.8237	31.13	0.9343	31.55	0.8388
PIGWM [59]	CVPR 21	23.25	0.7168	27.82	0.9030	13.81	0.4010	24.13	0.7520	27.25	0.8110
CLAIO [8]	TMM 24	29.28	0.8954	28.74	0.8920	25.81	0.8713	31.59	0.9335	30.96	0.8847
ILAWR	-	**30.13**	**0.9037**	**29.81**	**0.9131**	**27.33**	**0.8862**	**32.21**	**0.9365**	**31.15**	**0.8791**

Figure 5. Setting 1 generalize to the visual effects of the unknown real world, our method achieves superior visual results.

4.3. Importance-Guided Aggregation.

Inheriting diverse knowledge from multiple teachers can improve the generalization of the student model [14, 43]. Given the significant differences among the intermediate features of teachers, employing non-linear transformations to process features is a preferred solution [33]. As illustrated in Figure 4(d), we introduce the Importance-Guided Aggregation Module (IGAM) to consolidate the guidance generated by multiple teachers. During training at $session(t)$, the Guidance Pool generated by the old models is defined as:

$$\mathcal{G} = \{G_1, G_2, \cdots, G_{t-1}\}, \tag{14}$$

the Guidance Pool generated by Teacher Group 1 and Teacher Group 2 is denoted as $\mathcal{G}^1$ and $\mathcal{G}^2$, respectively. The corresponding outputs generated by the student model are denoted as:

$$S^1 = \mathcal{M}\left(\psi\left(x_i^t\right)\right), S^2 = f\left(x_i^t\right) = \mathcal{P}\left(\psi\left(x_i^t\right)\right), \tag{15}$$

when there is a greater discrepancy between the distribution of teacher guidance and student output, it often indicates more significant information. We utilize Kullback-Leibler (KL) divergence along each channel of the feature maps to measure this disparity, resulting in distance vectors:

$$Dist^j = \mathrm{KL}\left(S \| G_j\right), j = 1, \cdots, t-1, \tag{16}$$

we apply SoftMax on $t-1$ distance vectors along the same channel dimensions to generate an importance matrix:

$$\mathcal{V}^t = \mathrm{softmax}\left(\left[Dist^1, \cdots, Dist^{t-1}\right]\right), \tag{17}$$

Table 3. **Haze→Rain→Snow→Low-Light** (Setting 3) results on real datasets. Compared to state-of-the-art incremental learning and all-in-one methods, Our ILAWR achieves superior results on real-world datasets (**1.16dB** PSNR↑).

Method	Venue	Average		REVIDE [57]		SPA+ [61]		RealSnow [61]		LOL-v2 [51]	
		PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
TransWeather [42]	CVPR 22	24.27	0.7963	18.01	0.8054	31.65	0.8266	27.25	0.7325	20.16	0.8207
MutiTS [6]	CVPR 22	24.42	0.8177	17.72	0.8113	33.27	0.8732	26.72	0.7807	19.98	0.8055
WGWS [61]	CVPR 23	25.68	0.8520	18.06	0.8217	34.16	0.9171	28.05	0.8121	22.46	0.8569
RAM [35]	ECCV 24	24.75	0.8291	17.63	0.8059	33.37	0.8979	27.96	0.7963	20.02	0.8164
PIGWM [59]	CVPR 21	22.51	0.7938	16.73	0.7768	30.49	0.8837	22.64	0.6864	20.16	0.8281
CLAIO [8]	TMM 24	25.62	0.8420	17.94	0.8171	33.66	0.9073	27.79	0.8003	23.09	0.8432
ILAWR	-	**26.78**	**0.8566**	**18.76**	**0.8303**	**35.19**	**0.9237**	**28.41**	**0.8109**	**24.74**	**0.8616**

multiplying the importance matrix with the teacher guidance and then merging each feature sub-map yields the final guidance. The IGAM model $\mathcal{I}$ is defined as follows:

$$I\left(\mathcal{G}\right) = \text{Concat}\left(\mathcal{V}^t \odot \mathcal{G}\right). \tag{18}$$

The IGAM model aggregates the diverse guidance from each group of teachers into the final guidance:

$$\begin{aligned} Guide_1 &= \mathcal{I}\left(\sum_{i=1}^{t-1} \mathcal{M}^i\left(\psi^i\left(\mathcal{X}\right)\right)\right), \\ Guide_2 &= \mathcal{I}\left(\sum_{i=1}^{t-1} f^i\left(\mathcal{X}\right)\right). \end{aligned} \tag{19}$$

During the training process in $session(t)$, the models to optimize and update parameters are $\mathcal{M}$ and f. As shown in Algorithm 1, $\mathcal{M}$ and f are individually trained in each iteration, with the other model's parameters frozen. combining Equations 11, 13, and 19 through multilateral distillation aggregation, the total training loss of of model f^t is:

$$\mathcal{L}_f^t = \mathcal{L}_{base_1}^t + \lambda \mathcal{L}_{teach_1}^t + \zeta \mathcal{L}_{teach_2}^t, \tag{20}$$

where λ and ζ are hyperparameters used to balance the guidance from the two groups of teachers. And the total optimization objective model $\mathcal{M}^t$ is defined as follows:

$$\mathcal{L}_{\mathcal{M}}^t = \mathcal{L}_{base_2}^t = \mathcal{L}_{content}^t + \alpha_2 \mathcal{L}_{contrast}^t, \tag{21}$$

where α_1 and α_2 are used to balance fidelity and perceptual aspects. Based on experience, we set $\alpha_1 = 0.3$ and $\alpha_2 = 0.2$. Detailed ablation studies of λ and ζ is in Sec. 5.3.

5. Experiments

In the incremental learning task of adverse weather removal, we conducted experiments on synthetic datasets including RESIDE [20], Rain100H[50], Snow100K [24], Raindrop [34], as well as real datasets REVIDE [57], SPA+ [61], RealSnow [61] and LOL-v2 [51]. During training and testing, we followed the configuration of the current state-of-the-art model CLAIO. We fed the data incrementally into the model for training, and for fair comparison, we also adopted the weather sequence of Haze-Rain-Snow. We quantitatively evaluated the results using peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM).

5.1. Implementation Details

We implemented our method on the PyTorch platform using 8 NVIDIA GTX 1080Ti GPU. The optimizer used is Adam with an exponential decay rate of 0.9. The initial learning rate is set to 5e-5, and a cosine annealing strategy is employed for adjustment. The patch size for images is set to 128, and based on tuning experience, the hyperparameters are set as follows: $\alpha_1 = 0.3$, $\alpha_2 = 0.2$, $\lambda = 0.3$, $\zeta = 0.4$.

5.2. Evaluation and Comparison

Synthetic Datasets. To validate the effectiveness of the proposed incremental learning method, we conducte two sets of experiments using the base restoration network $f\left(\cdot|\theta\right)$: Individual Training (training separately with multiple datasets using different sets of parameters), and Sequential Fine-tuning (sequentially fine-tuning on multiple datasets). As shown in Table 1, when conducting continuous weather removal with three conditions **Haze→Rain→Snow** (Setting 1), we mitigate catastrophic forgetting compared to the sequential fine-tuning method and outperformed previous all-in-one and incremental learning methods in objective metrics. As shown in Figure 5, our ILAWR also exhibits improved generalization performance on real-world datasets. As shown in Table 2, when continuous weather conditions increase to **4**, *i.e.*, **Haze→Rain→Snow→Raindrop** (Setting 2), ILAWR further enhances the performance compared to others.

Real Weather Datasets. As shown in Table 3, We introduce common low-light degradation found in real-life scenarios, providing the first benchmark for continuous weather removal in real-world settings: **Haze→Rain→Snow→Low-Light** (Setting 3). The visual comparison on a real-world dataset in Figure 6 illustrates that ILAWR can better handle continuous adverse weather removal in real-world scenarios. We simultaneously present visualizations of feature maps before and after the DAM module, where degradation-related details are better preserved, extracting more degradation details for teachers.

5.3. Ablation Study

The ablation in this section were all conducted in Setting 1.

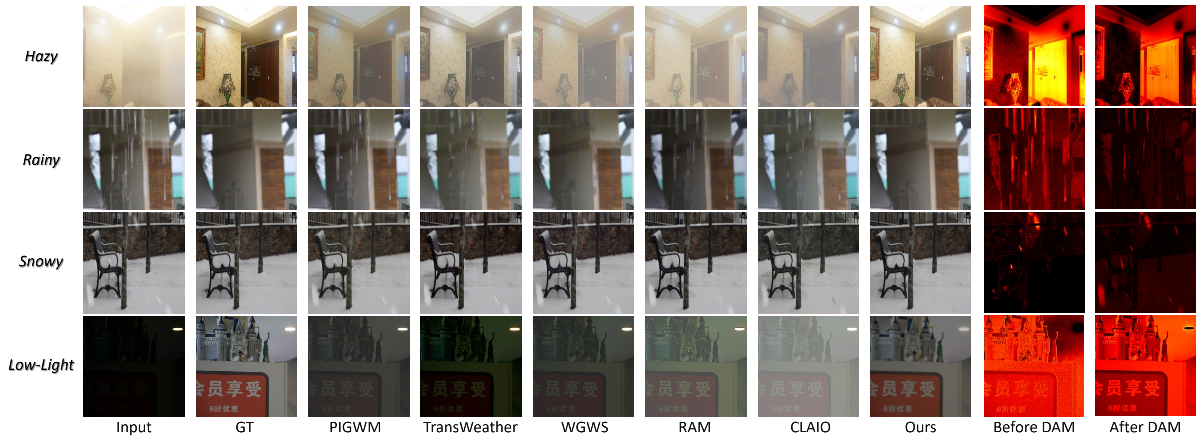


Figure 6. Real-world datasets visualization comparisons of our method with previous approaches for Setting 3 in Table 3. We successfully restored various degraded images, **alleviating catastrophic forgetting** in continuous adverse weather removal, and recovered more background details compared to other methods. The last two columns visualize the feature maps before and after the DAM module, indicating that DAM extracted more degradation-related details while reducing the background information mixed in the degradation features.

Table 4. Ablation experiment of weather sequence.

Task order	Average		RESIDE [20]		Rain100H [50]		Snow100K [24]	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
haze→rain→snow	31.32	0.938	31.41	0.977	28.91	0.896	33.64	0.941
haze→snow→rain	31.31	0.945	31.27	0.965	28.96	0.917	33.72	0.952
rain→haze→snow	31.35	0.943	31.74	0.984	28.86	0.908	33.44	0.937
rain→snow→haze	31.39	0.943	31.57	0.961	28.77	0.922	33.84	0.946
snow→haze→rain	31.28	0.947	31.53	0.974	28.52	0.914	33.78	0.954
snow→rain→haze	31.39	0.940	31.47	0.980	29.12	0.905	33.58	0.934

Table 5. Ablation experiment of DAM and IGAM.

	$f(\cdot\|\theta)$	DAM	IGAM	RESIDE [20] PSNR↑	SSIM↑	Rain 100H [50] PSNR↑	SSIM↑	Snow100K [24] PSNR↑	SSIM↑
①	✓			16.20	0.814	13.47	0.454	32.21	0.916
②	✓	✓		25.14	0.831	22.37	0.852	32.13	0.909
③	✓		✓	26.48	0.953	22.16	0.836	32.16	0.921
④	✓	✓	✓	**31.41**	**0.977**	**28.91**	**0.896**	**33.64**	**0.941**

Table 6. Ablation experiment of $\mathcal{L}_{teach_1}$ and $\mathcal{L}_{teach_2}$.

	$\mathcal{L}_{base}$	$\mathcal{L}_{teach_1}$	$\mathcal{L}_{teach_2}$	RESIDE [20] PSNR↑	SSIM↑	Rain 100H [50] PSNR↑	SSIM↑	Snow100K [24] PSNR↑	SSIM↑
①	✓			16.20	0.814	13.47	0.454	32.21	0.916
②	✓	✓		24.67	0.816	22.45	0.862	32.04	0.897
③	✓		✓	25.88	0.923	22.09	0.827	32.24	0.926
④	✓	✓	✓	**31.41**	**0.977**	**28.91**	**0.896**	**33.64**	**0.941**

Table 7. Model efficiency comparisons on a 256×256 image.

Methods	TransWeather	WGWS	RAM	CLAIO	Ours
#Pram ($M : 10^6$)	38.05	5.97	25.31	5.59	4.95
Inference Time (s)	0.026	0.030	0.067	0.045	0.032

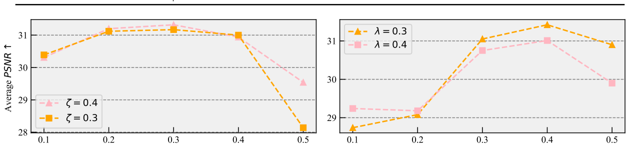


Figure 7. Ablation experiment of λ and ζ.

Weather sequence. By altering the sequence of weather conditions in each session, we investigated the impact of training order on the model. As shown in Table 4, the model's error variation due to changes in training order remained within 1%, indicating stable performance.

DAM and IGAM. A study was conducted to assess the effectiveness of the two proposed modules in this paper, with results presented in Table 5. It was observed that the absence of DAM significantly degraded performance, especially on the Rain100H dataset. The best performance was achieved when both DAM and IGAM were available.

Loss function $\mathcal{L}_{teach_1}$ and $\mathcal{L}_{teach_2}$. As shown in Table 6, $\mathcal{L}_{teach_1}$ was found to significantly enhance performance on the Rain100H dataset. The presence of $\mathcal{L}_{teach_1}$ was beneficial in enhancing performance on Rain100H. Better performance found both $\mathcal{L}_{teach_1}$ and $\mathcal{L}_{teach_2}$ are available, confirming the effectiveness of a multi-teacher model.

The hyperparameters λ and ζ. Based on experience, we set $\alpha_1 = 0.3$ and $\alpha_2 = 0.2$, then conducted ablation studies on λ and ζ, as shown in Figure 7. A high ζ value resulted in insufficient degradation learning, while an excessively high λ led to a sudden performance drop. Balancing the two teachers' knowledge, we adjusted $\lambda = 0.3$ and $\zeta = 0.4$.

Model efficiency. As shown in Table 7, the superior ILAWR achieves faster inference time and lower parameter count compared to competitors.

6. Conclusion

In this paper, we introduce the Incremental Learning for Adverse Weather Removal (ILAWR) framework based on Degradation-Aware Multilateral Distillation (DAM) approach. Leveraging Fourier priors of images, DAM enables the teacher model to extract more degradation information. Through a simple yet effective multilateral distillation approach, our model significantly reduces catastrophic forgetting across various degradation types. Extensive experiments validate that our approach surpasses existing methods on synthetic and real datasets, demonstrating promising prospects for adaptability to actual scenario.

Acknowledgement. This work was supported by the National Natural Science Foundation of China (NSFC) under Grants 62422609 and 62276243.

References

[1] Davide Abati, Jakub Tomczak, Tijmen Blankevoort, Simone Calderara, Rita Cucchiara, and Babak Ehteshami Bejnordi. Conditional channel gated networks for task-aware continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1

[2] Rahaf Aljundi, Francesca Babiloni, Mohamed Elhoseiny, Marcus Rohrbach, and Tinne Tuytelaars. Memory aware synapses: Learning what (not) to forget. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018. 6

[3] Francisco M Castro, Manuel J Marín-Jiménez, Nicolás Guil, Cordelia Schmid, and Karteek Alahari. End-to-end incremental learning. In *Proceedings of the European conference on computer vision (ECCV)*, pages 233–248, 2018. 1

[4] Pierre Charbonnier, Laure Blanc-Feraud, Gilles Aubert, and Michel Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proceedings of 1st international conference on image processing (ICIP)*, 1994. 3

[5] Wei-Ting Chen, Hao-Yu Fang, Cheng-Lin Hsieh, Cheng-Che Tsai, I-Hsiang Chen, Jian-Jiun Ding, and Sy-Yen Kuo. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4196–4205, 2021. 1, 2

[6] Wei-Ting Chen, Zhi-Kai Huang, Cheng-Che Tsai, Hao-Hsiang Yang, Jian-Jiun Ding, and Sy-Yen Kuo. Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17653–17662, 2022. 6, 7

[7] De Cheng, Yan Li, Dingwen Zhang, Nannan Wang, Xinbo Gao, and Jiande Sun. Robust single image dehazing based on consistent and contrast-assisted reconstruction. *arXiv preprint arXiv:2203.15325*, 2022. 1, 2, 4

[8] De Cheng, Yanling Ji, Dong Gong, Yan Li, Nannan Wang, Junwei Han, and Dingwen Zhang. Continual all-in-one adverse weather removal with knowledge replay on a unified network structure. *IEEE Transactions on Multimedia*, 2024. 1, 2, 3, 4, 6, 7

[9] Hang Dong, Jin shan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2154–2164, 2020. 1

[10] Arthur Douillard, Matthieu Cord, Charles Ollion, Thomas Robert, and Eduardo Valle. Podnet: Pooled outputs distillation for small-tasks incremental learning. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 86–102. Springer, 2020. 6

[11] Matteo Frigo and Steven G Johnson. Fftw: An adaptive software architecture for the fft. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)*, pages 1381–1384. IEEE, 1998. 4

[12] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1715–1723, 2017. 1, 2

[13] Xueyang Fu, Jie Xiao, Yurui Zhu, Aiping Liu, Feng Wu, and Zheng-Jun Zha. Continual image deraining with hypergraph convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8):9534–9551, 2023. 1, 2, 3

[14] Jianping Gou, Baosheng Yu, Stephen J. Maybank, and Dacheng Tao. Knowledge distillation: A survey. *Int. J. Comput. Vision*, 129(6):1789–1819, 2021. 6

[15] Yi Gu, Chao Wang, and Jie Li. Incremental image de-raining via associative memory. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*. AAAI Press, 2023. 3

[16] Saihui Hou, Xinyu Pan, Chen Change Loy, Zilei Wang, and Dahua Lin. Learning a unified classifier incrementally via rebalancing. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 831–839, 2019. 1

[17] Da-Wei Jaw, Shih-Chia Huang, and Sy-Yen Kuo. Desnowgan: An efficient single image snow removal framework using cross-resolution lateral connection and gans. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(4):1342–1350, 2020. 1

[18] Minsoo Kang, Jaeyoo Park, and Bohyung Han. Class-incremental learning by knowledge distillation with adaptive feature consolidation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16071–16080, 2022. 6

[19] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017. 6

[20] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2019. 2, 6, 7, 8

[21] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947, 2017. 1, 3, 6

[22] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(12):2935–2947, 2018. 1

[23] Minghao Liu, Wenhan Yang, Yuzhang Hu, and Jiaying Liu. Dual prompt learning for continual rain removal from single images. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, 2023. 3

[24] Yun-Fu Liu, Da-Wei Jaw, Shih-Chia Huang, and Jenq-Neng Hwang. Desnownet: Context-aware deep network for snow

removal. *IEEE Transactions on Image Processing*, 27(6): 3064–3073, 2018. 2, 6, 7, 8

[25] Zhihao Liu, Hui Yin, Xinyi Wu, Zhenyao Wu, Yang Mi, and Song Wang. From shadow generation to shadow removal. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4925–4934, 2021. 1

[26] Xin Lu, Yurui Zhu, Xi Wang, Dong Li, Jie Xiao, Yunpeng Zhang, Xueyang Fu, and Zheng-Jun Zha. Hirformer: Dynamic high resolution transformer for large-scale image shadow removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 6513–6523, 2024. 1

[27] Michael McCloskey and Neal J. Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. *Psychology of Learning and Motivation*, 24:109–165, 1989. 1

[28] Seungjun Nah, Sanghyun Son, Jaerin Lee, and Kyoung Mu Lee. Clean images are hard to reblur: Exploiting the ill-posed inverse task for dynamic scene deblurring. In *International Conference on Learning Representations*, 2021. 1

[29] Srinivasa G. Narasimhan and Shree K. Nayar. Vision and the atmosphere. *International Journal of Computer Vision*, 48: 233–254, 2002. 3

[30] Shree K. Nayar and Srinivasa G. Narasimhan. Vision in bad weather. *Proceedings of the Seventh IEEE International Conference on Computer Vision (ICCV)*, 2:820–827 vol.2, 1999. 3

[31] Shree K. Nayar and Srinivasa G. Narasimhan. Seeing through bad weather. In *International Symposium of Robotics Research (ISRR)*, 2003.

[32] Shree K. Nayar and Srinivasa G. Narasimhan. Models and algorithms for vision through the atmosphere. In *Columbia University*, USA, 2004. AAI3115363. 3

[33] Seonguk Park and Nojun Kwak. Feed: Feature-level ensemble for knowledge distillation. *ArXiv*, abs/1909.10754, 2019. 6

[34] Rui Qian, Robby T. Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for raindrop removal from a single image. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2482–2491, 2017. 2, 6, 7

[35] Chu-Jie Qin, Rui-Qi Wu, Zikun Liu, Xin Lin, Chun-Le Guo, Hyun Hee Park, and Chongyi Li. Restore anything with masks: Leveraging mask image modeling for blind all-in-one image restoration, 2024. 1, 6, 7

[36] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 11908–11915, 2020. 1

[37] Chao Ren, Xiaohai He, Chuncheng Wang, and Zhibo Zhao. Adaptive consistency prior based deep network for image denoising. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8592–8602, 2021. 1

[38] Wenqi Ren, Sibo Liu, Hua Zhang, Jin shan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *European Conference on Computer Vision*, 2016. 1

[39] Andres Sanin, Conrad Sanderson, and Brian C. Lovell. Improved shadow removal for robust person tracking in surveillance scenarios. In *2010 20th International Conference on Pattern Recognition*, pages 141–144, 2010. 1

[40] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations (ICLR)*, 2015. 4

[41] Fu-Jen Tsai, Yan-Tsung Peng, Chung-Chi Tsai, Yen-Yu Lin, and Chia-Wen Lin. Banet: A blur-aware attention network for dynamic scene deblurring. *IEEE Transactions on Image Processing*, 31:6789–6799, 2021. 1

[42] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M. Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2353–2363, 2022. 6, 7

[43] Lin Wang and Kuk-Jin Yoon. Knowledge distillation and student-teacher learning for visual intelligence: A review and new outlooks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6):3048–3068, 2022. 6

[44] Xi Wang, Xueyang Fu, Peng-Tao Jiang, Jie Huang, Mi Zhou, Bo Li, and Zheng-Jun Zha. Decoupling degradation and content processing for adverse weather image restoration. *ArXiv*, abs/2312.05006, 2023. 4

[45] Haitao Wen, Lili Pan, Yu Dai, Heqian Qiu, Lanxiao Wang, Qingbo Wu, and Hongliang Li. Class incremental learning with multi-teacher distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 28443–28452, 2024. 2, 3

[46] Jie Xiao, Xueyang Fu, Aiping Liu, Feng Wu, and Zhengjun Zha. Image de-raining transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45:12978–12995, 2022. 1

[47] Jie Xiao, Xueyang Fu, Feng Wu, and Zhengjun Zha. Stochastic window transformer for image restoration. In *Neural Information Processing Systems*, 2022. 1

[48] Jie Xiao, Xueyang Fu, Man Zhou, HongJiang Liu, and Zhengjun Zha. Random shuffle transformer for image restoration. In *International Conference on Machine Learning*, 2023. 1, 2

[49] Hao Yang, Liyuan Pan, Yan Yang, and Wei Liang. Language-driven all-in-one adverse weather removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 24902–24912, 2024. 1, 2

[50] Wenhan Yang, Robby T. Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2, 6, 7, 8

[51] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, 30:2072–2086, 2021. 2, 7

[52] Lu Yu, Bartlomiej Twardowski, Xialei Liu, Luis Herranz, Kai Wang, Yongmei Cheng, Shangling Jui, and Joost van de Weijer. Semantic drift compensation for class-incremental

learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1

[53] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Cycleisp: Real image restoration via improved data synthesis. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2693–2702, 2020. 1

[54] Friedemann Zenke, Ben Poole, and Surya Ganguli. Continual learning through synaptic intelligence. In *Proceedings of the 34th International Conference on Machine Learning*, pages 3987–3995. PMLR, 2017. 1

[55] Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Björn Stenger, Wei Liu, and Hongdong Li. Deblurring by realistic blurring. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2734–2743, 2020. 1

[56] Wuming Zhang, Xi Zhao, Jean-Marie Morvan, and Liming Chen. Improving shadow suppression for illumination robust face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 41 (3):611–624, 2019. 1

[57] Xinyi Zhang, Hang Dong, Jinshan Pan, Chao Zhu, Ying Tai, Chengjie Wang, Jilin Li, Feiyue Huang, and Fei Wang. Learning to restore hazy video: A new real-world dataset and a new method. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9235–9244, 2021. 2, 7

[58] Linglan Zhao, Jing Lu, Yunlu Xu, Zhanzhan Cheng, Dashan Guo, Yi Niu, and Xiangzhong Fang. Few-shot class-incremental learning via class-aware bilateral distillation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11838–11847, 2023. 3

[59] Man Zhou, Jie Xiao, Yifan Chang, Xueyang Fu, Aiping Liu, Jinshan Pan, and Zheng-Jun Zha. Image de-raining via continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4907–4916, 2021. 1, 3, 6, 7

[60] Yurui Zhu, Xueyang Fu, Chengzhi Cao, Xi Wang, Qibin Sun, and Zheng-Jun Zha. Single image shadow detection via complementary mechanism. In *Proceedings of the 30th ACM International Conference on Multimedia*, page 6717–6726, New York, NY, USA, 2022. Association for Computing Machinery. 1

[61] Yurui Zhu, Tianyu Wang, Xueyang Fu, X. Yang, Xin Guo, Jifeng Dai, Yu Qiao, and Xiao hua Hu. Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21747–21758, 2023. 1, 2, 3, 6, 7