# LEARNING TO RESTORE MORE: CONTINUAL CAPABILITY EXPANSION FOR PRETRAINED IMAGE RESTORATION MODELS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Image restoration models are typically trained with a fixed set of capabilities. When new restoration requirements emerge, existing solutions usually train additional models or jointly retrain the original model with both new and historical data. Instead of designing another restoration backbone, we investigate how a trained restorer can continually acquire new capabilities without forgetting those learned previously. We propose **RestoreMore**, a continual capability-expansion framework that preserves the pretrained restoration model as a frozen capability anchor and learns residual expansion modules for newly arriving degradations. RestoreMore introduces a capability-oriented bi-level routing mechanism at multiple feature stages. The first routing level identifies restoration capabilities relevant to the current input, while the second selects and combines a sparse set of complementary degradation experts. This design enables newly introduced tasks to selectively reuse historical restoration knowledge and progressively enriches the expert bank available for subsequent restoration tasks. Extensive experiments on a wide range of restoration benchmarks demonstrate that RestoreMore consistently acquires new restoration abilities while preserving and improving previously learned capabilities.

## 1 INTRODUCTION

Image restoration aims to recover high-quality images from observations corrupted by noise, rain, and other adverse factors. Most existing methods, however, are developed under a fixed-capability assumption. A degradation-specific model Gao et al. (2024c); Guo et al. (2025b); Peng et al. (2026); Feng et al. (2026) is optimized for a predefined task and usually performs poorly when the degradation changes. All-in-one image restoration alleviates the need to maintain separate networks by learning multiple restoration tasks within a unified model Gao et al. (2024b); Mao et al. (2026); Tang et al. (2026); Gao et al. (2026b). Nevertheless, the capability set of an all-in-one model is still determined by the degradation types available during its original training.

This closed-set assumption limits the long-term utility of trained restoration models. A denoising model may subsequently be required to remove rain or haze, while an all-in-one model originally trained for noise, rain, and blur may later encounter snow or previously unseen imaging artifacts. A common solution is to train an additional task-specific model or jointly retrain the original model using the expanded collection of new and historical datasets. The former continually increases storage and deployment costs, whereas the latter requires access to all previous data and repeatedly optimizes knowledge that has already been learned. Fig. 1 illustrates the difference between these existing strategies and our capability-expansion perspective.

Several studies have explored continual learning for image restoration Cheng et al. (2024); Lu et al. (2025); Gong & Ma (2026), demonstrating the feasibility of incrementally extending restoration capabilities. However, existing approaches are generally built on dedicated incremental architectures or restricted restoration task families, with their primary objective being the mitigation of catastrophic forgetting. They do not explicitly address how to preserve the restoration pathway of an arbitrary pretrained model or how to selectively reuse complementary knowledge accumulated from different degradation types and feature stages.



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

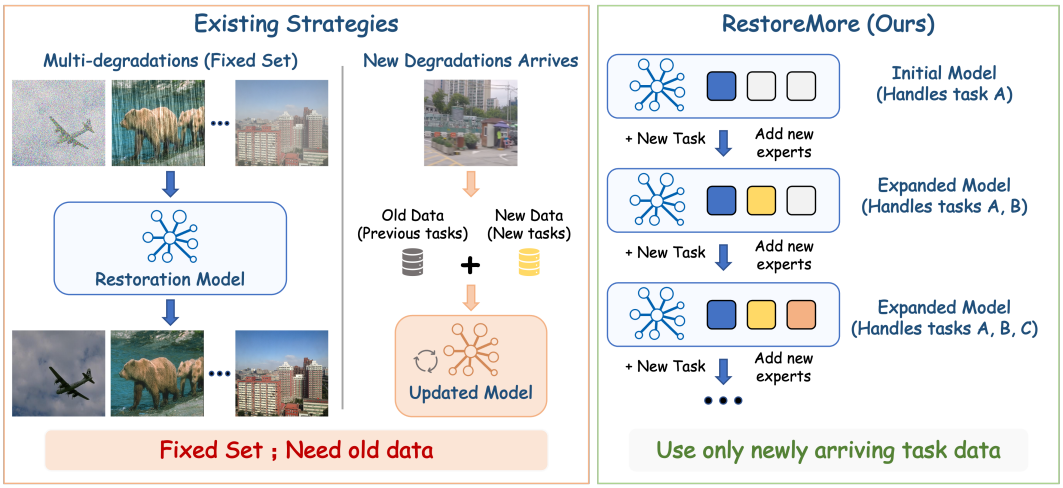


Figure 1: Comparison of existing restoration strategies and RestoreMore. Existing methods are limited to a fixed degradation set and typically require joint retraining with historical and new data when new degradations arrive. In contrast, RestoreMore continually expands a pretrained restorer by adding new experts using only newly arriving task data.

In this paper, we study image restoration from a continual capability-expansion perspective. Rather than constructing another restoration backbone for a predetermined task collection, we ask a different question: given an already trained image restoration model, how can its restoration capability be continually expanded to previously unsupported degradations? After learning a new degradation, the expanded model should acquire the corresponding restoration ability while retaining the capabilities learned before expansion. Meanwhile, the accumulated expert bank should enable knowledge reuse across related degradation types rather than treating each newly arriving task in isolation.

To achieve this objective, we propose RestoreMore, which preserves the pretrained restoration model as a frozen capability anchor and introduces lightweight residual experts for newly arriving degradations. A null expert retains the original restoration pathway when no additional correction is required. RestoreMore further employs capability-oriented bi-level routing at selected feature stages. The first level identifies the Top-$M$ relevant restoration capabilities, while the second selects and combines a sparse set of Top-$K$ complementary experts. This stage-wise design enables each newly arriving task to selectively reuse restoration knowledge accumulated from previous tasks without modifying the pretrained backbone or historical experts.

The contributions of this work are summarized as follows:

- We propose RestoreMore, a framework for continual restoration capability expansion that enables an already trained image restoration model to progressively acquire capabilities for previously unsupported degradations.
- We propose a capability-preserving expansion framework with capability-oriented bi-level routing. The pretrained model is retained as a frozen restoration anchor, while lightweight residual experts are incrementally introduced for new degradations. Stage-wise capability retrieval and sparse expert composition enable newly arriving tasks to selectively reuse complementary restoration knowledge accumulated from previous tasks.
- Extensive experiments demonstrate that RestoreMore consistently acquires new restoration abilities while preserving and improving previously learned capabilities.

# 2 Related Work

## 2.1 Image Restoration

Image restoration aims to recover high-quality visual content from observations. Prior-driven restoration methods formulate image recovery using explicit structural assumptions or degradation models Rong et al. (2024); Feng et al. (2026). Although these methods provide clear interpretations, their manually defined assumptions are often insufficient for complex and spatially varying degradations. The development of deep learning has shifted image restoration toward data-driven representation learning, allowing restoration models to learn degradation characteristics directly from large-scale datasets Guo et al. (2025a); Gao et al. (2025a).

ALGNet Gao et al. (2024a) combines adaptive local and global feature extraction to recover both fine textures and large-scale structures, while XYScanNet Liu et al. (2025) employs alternating spatial scans to improve long-range dependency modeling. EfDeRain+ Guo et al. (2025b) treats deraining as a predictive filtering process rather than explicitly decomposing the rain layer. PGH$^2$Net Su et al. (2025) incorporates bright- and dark-channel priors together with histogram equalization to guide hierarchical dehazing. Diffusion-based methods, including UPID-EDM Wen et al. (2024) and Diff-Unmix Zeng et al. (2024), further exploit generative priors and iterative denoising to improve perceptual fidelity. FSNet Cui et al. (2023) reorganize spatial and frequency information to improve generalization across restoration settings. ACL Gu et al. (2025) introduces efficient long-range modeling through linear attention, while StarIR Cui et al. (2026) and MHNet Gao et al. (2025b) exploit hierarchical or mixed-frequency representations to capture restoration cues shared by different degradations. BaryIR Tang et al. (2026) aligns multiple degraded distributions through a barycentric representation, thereby improving degradation invariance and cross-task generalization. AdaIR Cui et al. (2025) disentangles degradation and content information in spatial and frequency domains, whereas Perceive-IR Zhang et al. (2025) estimates both degradation categories and fine-grained severity to support more adaptive restoration. AllRestorer Mao et al. (2026) combines visual and textual information into a composite scene embedding for describing degradation conditions within a unified model. Large pretrained models and generative priors have further broadened the design space of all-in-one restoration. VLU-Net Zeng et al. (2025) uses vision-language representations to extract degradation-aware cues and integrates them into an unfolding-based restoration process. AutoDIR Jiang et al. (2024) combines vision-language guidance with latent diffusion priors, while Defusion Luo et al. (2025) constructs degradation instructions and performs diffusion in the degradation space for generalized restoration.

Despite these advances, existing restoration methods are still mainly developed under a fixed-capability assumption. When an unsupported degradation appears, the common solution is to train another model or retrain the unified model using both new and historical datasets.

## 2.2 Continual Learning

Continual learning aims to enable a model to acquire knowledge from a sequence of tasks while mitigating catastrophic forgetting. Existing methods are commonly based on regularization, replay, optimization constraints, or architectural expansion. Regularization-based approaches restrict changes to parameters or predictions that are important for previously learned tasks Li & Hoiem (2017), whereas replay-based methods revisit stored or generated historical samples Rebuffi et al. (2017); Lopez-Paz & Ranzato (2017). The development of pretrained models has further motivated parameter-efficient continual learning. MoE-Adapters Yu et al. (2024) dynamically introduces and selects adapter experts for continual vision-language learning. SEMA Wang et al. (2025) evaluates whether existing adapters can represent an incoming data distribution and expands the module set only when necessary. CaRE Lou et al. (2026) further adopts a two-stage routing strategy, in which task-relevant routers are first selected and then used to retrieve and aggregate complementary experts at individual network layers. These methods demonstrate that modular adaptation and sparse knowledge retrieval can improve the stability–plasticity balance in continual learning with pretrained models. Continual learning has also been extended to image restoration. KR Cheng et al. (2024) combines knowledge replay with feature-level distillation for sequential adverse-weather removal. ILAWR Lu et al. (2025) employs degradation-aware distillation to reduce interference between newly introduced and previously learned weather-removal tasks. MINI Gong & Ma (2026)

separates task-specific restoration knowledge through meta-convolutional parameters, thereby reducing destructive updates across incremental tasks.

Nevertheless, existing continual restoration methods mainly focus on mitigating forgetting within restricted restoration task families, without explicitly preserving an arbitrary pretrained restoration pathway or enabling stage-wise reuse of complementary knowledge across degradation types. In contrast, RestoreMore freezes the pretrained restorer as a capability anchor and progressively expands its capability set through lightweight residual experts and capability-oriented bi-level routing.

# 3 METHOD

## 3.1 PROBLEM FORMULATION

We study continual restoration capability expansion, whose goal is to enable an already trained image restoration model to progressively restore degradation types that were not included in its original training set. Let $F_0(\cdot;\theta_0)$ denote a pretrained image restoration model with an initial capability set $\mathcal{S}_0 = \{d_1, d_2, \ldots, d_B\}$, where each $d_i$ represents a degradation type supported by $F_0$. The initial model may be a task-specific restorer with $B = 1$ or an all-in-one restorer with $B > 1$.

After the initial training, previously unsupported degradation types arrive sequentially. At incremental stage $t$, a paired dataset $\mathcal{D}_t = (x_{t,n}, y_{t,n})_{n=1}^{N_t}$ is provided for a new degradation $d_{B+t}$, where $x_{t,n}$ and $y_{t,n}$ denote the degraded image and its clean target, respectively. The capability set is then expanded as $\mathcal{S}_t = \mathcal{S}_{t-1} \cup \{d_{B+t}\}$. Before continual expansion, a one-time capability indexing procedure is performed for $\mathcal{S}_0$. Afterwards, historical training images are neither stored nor revisited, and each incremental stage uses only the newly arriving dataset $\mathcal{D}_t$. Details of the initial capability indexing are provided in Appendix A.1. The expanded model should restore all degradation types observed so far without requiring a degradation label $\hat{y} = F_t(x)$, $d(x) \in \mathcal{S}_t$. The continual expansion process considers two primary objectives:

$$\begin{aligned} \text{Acquisition:} \quad & Q_{t,B+t} \uparrow, \\ \text{Retention:} \quad & Q_{t,i} \geq Q_{t-1,i} - \epsilon, \qquad d_i \in \mathcal{S}_{t-1}, \end{aligned} \tag{1}$$

where $Q_{t,i}$ denotes the restoration quality of $F_t$ on degradation $d_i$. Unlike conventional all-in-one restoration, which jointly optimizes a model over a fixed collection of degradations, our formulation progressively expands the capability boundary of an existing restorer. The pretrained backbone and previously learned expansion components remain frozen, while lightweight components are introduced only for newly arriving degradations.

## 3.2 OVERALL PIPELINE

The overall pipeline of RestoreMore is illustrated in Fig. 2. The pretrained restorer remains frozen and provides the original restoration pathway, while stage-wise residual expansion modules are introduced to acquire new degradation-specific capabilities. Suppose that the pretrained restorer contains $L$ major feature-processing stages. The original feature at stage $l$ is computed as

$$u_l = \mathcal{B}_l\left(h_{l-1}; \theta_{0,l}\right), \qquad l = 1, \ldots, L, \tag{2}$$

where $\mathcal{B}_l$ denotes the original restoration transformation. Its parameters remain frozen throughout continual expansion $\theta_0^{(t)} = \theta_0, \forall t$. At selected stages, RestoreMore introduces a residual capability expansion module $\mathcal{M}_l^{(t)}$:

$$h_l = u_l + \mathcal{M}_l^{(t)}(u_l). \tag{3}$$

Thus, the pretrained feature $u_l$ always remains in the forward pathway, while the expansion branch provides only the additional transformation required by incrementally acquired capabilities. When a new degradation $d_{B+t}$ arrives, a capability prototype and router together with a degradation expert and its routing key are introduced at each selected feature stage. All historical prototypes, routers, experts, and routing keys remain frozen. Only the newly introduced components are optimized using $\mathcal{D}_t$. Once training is completed, they are frozen and incorporated into the capability bank for subsequent tasks. For notational simplicity, the aggregate effect of all stage-wise expansion modules can be abstractly expressed as:

$$F_t(x) = F_0(x) + \Delta F_t\left(x; \mathcal{P}^{(t)}, \mathcal{R}^{(t)}, \mathcal{E}^{(t)}\right), \tag{4}$$

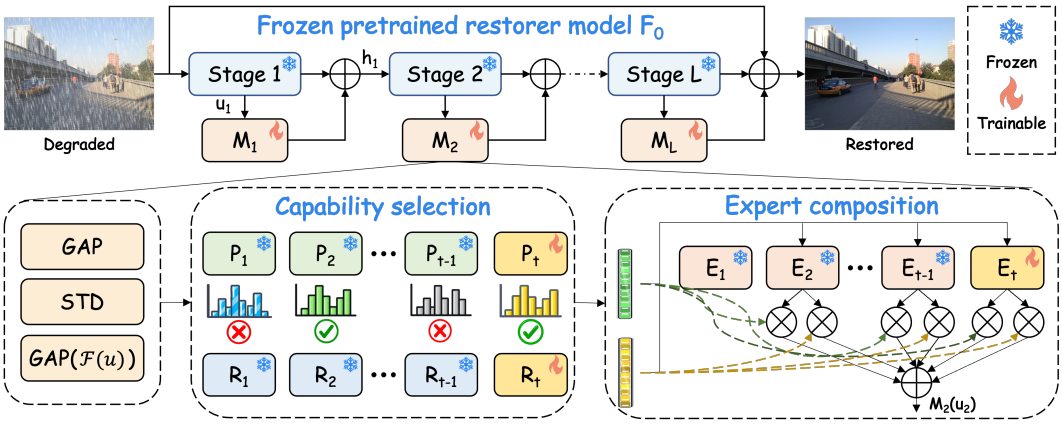


Figure 2: Overview of the proposed RestoreMore framework. A pretrained restoration model is preserved as a frozen capability anchor, and lightweight residual expansion modules are inserted at selected feature stages. Inside each expansion module, a degradation descriptor is first constructed from spatial and frequency statistics. Top-$M$ capability selection is then performed over accumulated capability prototypes and routers. The selected routers further retrieve complementary experts from the expert-key bank through query–key matching. Frozen historical components and trainable current-task components are indicated by snowflake and flame icons, respectively.

where $\mathcal{P}^{(t)}$, $\mathcal{R}^{(t)}$, and $\mathcal{E}^{(t)}$ denote the accumulated capability prototypes, capability routers, and expert bank, respectively. The complete incremental optimization procedure is given in Appendix A.3.

### 3.3 Capability-Oriented Bi-Level Routing

The core of RestoreMore is a capability-oriented bi-level routing mechanism that selectively retrieves and composes restoration knowledge accumulated during continual learning. At each feature stage, the first routing level identifies restoration capabilities relevant to the input, while the second selects complementary degradation experts associated with the retrieved capabilities. At stage $l$, RestoreMore maintains $(p_{i,l}, R_{i,l})_{d_i \in \mathcal{S}_t}$, $p_{i,l}$ and $R_{i,l}$ denote the prototype and router of capability $d_i$, respectively. The incremental expert bank is:

$$\mathcal{E}_l^{(t)} = \{(E_{\varnothing,l}, k_{\varnothing,l}), (E_{1,l}, k_{1,l}), \ldots, (E_{t,l}, k_{t,l})\}, \tag{5}$$

where $k_{j,l} \in \mathbb{R}^{d_r}$ is the routing key associated with expert $E_{j,l}$. The null expert satisfies $E_{\varnothing,l}(u) = 0$, allowing the expansion branch to apply no additional correction when the pretrained pathway is sufficient.

**Capability selection.** Unlike recognition, image restoration does not naturally provide a discrete semantic representation for identifying the underlying restoration requirement. We therefore construct a degradation-sensitive descriptor using spatial statistics and frequency characteristics:

$$q_l = \psi_l\left([\mathrm{GAP}(u_l), \mathrm{STD}(u_l), \mathrm{GAP}\left(|\mathcal{F}(u_l)|\right)]\right), \tag{6}$$

where GAP and STD denote channel-wise spatial average and standard deviation, $\mathcal{F}$ denotes the two-dimensional Fourier transform, and $\psi_l$ is a lightweight projection network. To keep the descriptor space consistent across incremental stages, $\psi_l$ is calibrated during capability indexing and subsequently frozen; details are provided in Appendix A.1. For each learned degradation $d_i$, the capability affinity $a_{i,l} = \frac{q_l^\top p_{i,l}}{\|q_l\|_2 \|p_{i,l}\|_2}$. The normalized capability probability is:

$$\pi_{i,l} = \frac{\exp(a_{i,l}/\tau_c)}{\sum_{d_n \in \mathcal{S}_t} \exp(a_{n,l}/\tau_c)}, \tag{7}$$

where $\tau_c$ is a temperature parameter. The first routing level selects the $M$ most relevant restoration capabilities $\mathcal{I}_l = \mathrm{TopM}(\{\pi_{i,l}\}_{d_i \in \mathcal{S}_t})$. The selected probabilities are renormalized as:

$$\widetilde{\pi}_{i,l} = \frac{\pi_{i,l}}{\sum_{n \in \mathcal{I}_l} \pi_{n,l}}. \tag{8}$$

Table 1: Quantitative comparison of different capability-expansion strategies on CNN-, Transformer-, and Mamba-based image restoration models. $^{O}$, $^{A}$, and $^{R}$ denote the original pretrained model, joint retraining with mixed old and new datasets, and RestoreMore trained only with newly arriving task data, respectively.

| Methods | Deraining | | Desnowing | | Dehazing | | Deblurring | | Denoising | | Average | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PSNR ↑ | SSIM ↑ | PSNR ↑ | SSIM ↑ | PSNR ↑ | SSIM ↑ | PSNR ↑ | SSIM ↑ | PSNR ↑ | SSIM ↑ | PSNR ↑ | SSIM ↑ |
| ALGNet$^{O}$ Gao et al. (2024a) | - | - | - | - | - | - | 33.49 | 0.964 | - | - | - | - |
| ALGNet$^{A}$ Gao et al. (2024a) | 35.89 | 0.969 | 30.89 | 0.924 | 32.76 | 0.968 | 29.41 | 0.882 | 30.97 | 0.881 | 31.98 | 0.925 |
| **ALGNet**$^{R}$ Gao et al. (2024a) | 36.37 | 0.973 | 30.97 | 0.927 | 34.47 | 0.974 | 33.18 | 0.952 | 31.31 | 0.885 | 33.26 | 0.942 |
| Perceive-IR$^{O}$ Zhang et al. (2025) | 38.29 | 0.980 | - | - | 30.87 | 0.975 | 29.46 | 0.886 | 31.53 | 0.890 | - | - |
| Perceive-IR$^{A}$ Zhang et al. (2025) | 38.92 | 0.981 | 32.99 | 0.948 | 34.52 | 0.980 | 29.79 | 0.889 | 31.44 | 0.885 | 33.53 | 0.937 |
| **Perceive-IR**$^{R}$ Zhang et al. (2025) | 39.25 | 0.982 | 33.15 | 0.949 | 34.77 | 0.981 | 31.73 | 0.927 | 31.62 | 0.891 | 34.10 | 0.946 |
| StarIR$^{O}$ Cui et al. (2026) | 38.50 | 0.984 | - | - | 30.89 | 0.979 | - | - | 31.51 | 0.893 | - | - |
| StarIR$^{A}$ Cui et al. (2026) | 38.71 | 0.981 | 32.85 | 0.950 | 34.32 | 0.978 | 28.45 | 0.880 | 31.45 | 0.885 | 33.16 | 0.935 |
| **StarIR**$^{R}$ Cui et al. (2026) | 39.42 | 0.983 | 33.26 | 0.952 | 34.80 | 0.982 | 30.63 | 0.928 | 31.58 | 0.889 | 33.94 | 0.947 |

Selecting multiple capabilities enables the model to exploit complementary restoration knowledge. For example, rain and snow removal may share high-frequency suppression patterns, while denoising and deblurring can provide complementary texture and structural cues. Stage-wise routing further allows different feature depths to retrieve different forms of historical knowledge.

**Expert composition.** For each selected capability, the corresponding router maps $q_l$ to a fixed-dimensional routing query $r_{i,l} = R_{i,l}(q_l) \in \mathbb{R}^{d_r}$. The compatibility between capability router $i$ and expert $j$ is computed through query–key matching:

$$z_{ij,l} = \frac{r_{i,l}^{\top} k_{j,l}}{\|r_{i,l}\|_2 \|k_{j,l}\|_2}, \tag{9}$$

$$g_{ij,l} = \frac{\exp(z_{ij,l}/\tau_e)}{\sum_{n\in\{\varnothing,1,\ldots,t\}} \exp(z_{in,l}/\tau_e)}. \tag{10}$$

Because $r_{i,l}$ has a fixed dimensionality, adding a new expert requires only appending an expert–key pair $(E_{t,l}, k_{t,l})$ and does not modify the structure or parameters of historical routers. A detailed explanation of this property is provided in Appendix A.2. Only the Top-$K$ experts are activated for each selected capability:

$$\mathcal{J}_{i,l} = \mathrm{TopK}\left(\{g_{ij,l}\}_{j\in\{\varnothing,1,\ldots,t\}}\right). \tag{11}$$

Their outputs are combined as:

$$v_{i,l} = \sum_{j\in\mathcal{J}_{i,l}} \widetilde{g}_{ij,l} E_{j,l}(u_l), \tag{12}$$

where $\widetilde{g}_{ij,l}$ denotes the expert weights renormalized over $\mathcal{J}_{i,l}$. The final expansion residual is $\mathcal{M}_l^{(t)}(u_l) = \sum_{i\in\mathcal{I}_l} \widetilde{\pi}_{i,l} v_{i,l}$. Therefore, only a sparse subset of accumulated experts participates in each forward pass. The expert architecture, cold-start strategy, optimization objective, and backbone-specific implementation are deferred to Appendix A.3–A.4, while the computational complexity is discussed in Appendix A.5.

## 4 Experiments

In this section, we first present the qualitative and quantitative comparison results, followed by the ablation studies. The experimental setup and more experiments are provided in Appendix A.6 and A.7.

### 4.1 Results

Table 1 evaluates RestoreMore on representative CNN- (StarIR Cui et al. (2026)), Transformer- (Perceive-IR Zhang et al. (2025)), and Mamba-based (ALGNet Gao et al. (2024a)) restoration models by comparing three settings: the original pretrained model ($^{O}$), joint retraining with mixed old and new datasets ($^{A}$), and continual capability expansion with RestoreMore ($^{R}$). Despite using only the datasets of newly arriving tasks rather than revisiting the original training data, RestoreMore consistently outperforms joint retraining across all three architectures. Compared with the




Figure 3: Qualitative comparison across five restoration tasks.

all-in-one retraining baseline, RestoreMore improves the average PSNR/SSIM from 31.98/0.925 to 33.26/0.942 for ALGNet, from 33.53/0.937 to 34.10/0.946 for Perceive-IR, and from 33.16/0.935 to 33.94/0.947 for StarIR. These consistent gains indicate that the proposed capability-expansion mechanism is not tied to a particular restoration architecture.

The results also show that RestoreMore better preserves the capabilities of the original models. This is particularly evident for ALGNet, whose original deblurring performance is 33.49 dB, but decreases substantially to 29.41 dB after mixed-data retraining. RestoreMore retains 33.18 dB while simultaneously acquiring four additional restoration capabilities, reducing the degradation on the original task from 4.08 dB to only 0.31 dB. For the stronger multi-task pretrained models, RestoreMore not only maintains historical performance but can further improve it. For example, Perceive-IR$^R$ improves the original results by 0.96 dB on deraining, 3.90 dB on dehazing, and 2.27 dB on deblurring. These results show that preserving the original restoration backbone while selectively reusing accumulated experts can effectively retain historical capabilities during continual expansion. Interestingly, several historical tasks even outperform their original models, suggesting that the expanded expert bank can provide complementary restoration knowledge beyond the initial capability set.

Meanwhile, RestoreMore remains effective on the newly introduced tasks. For ALGNet, which initially supports only deblurring, RestoreMore achieves higher performance than joint retraining on all four newly introduced tasks, with particularly clear gains of 1.71 dB on dehazing. Similar improvements are observed for Perceive-IR and StarIR, including gains of 1.94 dB and 2.18 dB on deblurring over their corresponding joint-training baselines.

Fig. 3 provides qualitative comparisons across the five restoration tasks. Consistent with the quantitative results in Table 1, RestoreMore produces visually cleaner results while retaining more faithful structures and local details. For deraining and desnowing, residual rain streaks and snow-like artifacts are more effectively suppressed without noticeably degrading the background content. For dehazing, RestoreMore recovers clearer scene visibility and more natural contrast, particularly around distant buildings and sky regions. In the deblurring examples, object boundaries, fine structures, and text regions are reconstructed more sharply, while competing results tend to retain blur or lose local details. For denoising, RestoreMore suppresses noise while better preserving fine textures instead of introducing excessive smoothing.

Table 2: Zero-shot generalization results on real-world degradation datasets.

| Methods | RealRain-1k-L | | | RTTS | | | SIDD | |
|---|---|---|---|---|---|---|---|---|
| | PSNR ↑ | SSIM ↑ | LPIPS ↓ | FADE ↓ | BRISQUE ↓ | NIMA ↑ | PSNR ↑ | SSIM ↑ |
| ECFNet Gao et al. (2026a) | 23.69 | 0.756 | 0.401 | 1.394 | 29.105 | 4.372 | 24.33 | 0.471 |
| Defusion Luo et al. (2025) | <u>27.37</u> | **0.903** | <u>0.371</u> | 1.277 | <u>23.520</u> | <u>4.615</u> | 24.52 | 0.495 |
| Perceive-IR Zhang et al. (2025) | 27.31 | <u>0.901</u> | 0.372 | <u>1.264</u> | 23.694 | 4.613 | <u>24.53</u> | <u>0.497</u> |
| **Perceive-IR**$^R$ Zhang et al. (2025) | **27.55** | **0.903** | **0.369** | **1.240** | **23.494** | **4.616** | **24.61** | **0.499** |

Table 3: Zero-shot generalization results on an unseen degradation type.

| Methods | UIEB | | | C60 | |
|---|---|---|---|---|---|
| | PSNR ↑ | SSIM ↑ | LPIPS ↓ | UCIQE↑ | UIQM↑ |
| ECFNet Gao et al. (2026a) | 20.49 | 0.856 | 0.223 | 0.539 | 2.517 |
| ACL Gu et al. (2025) | 20.94 | 0.867 | 0.201 | 0.552 | 2.485 |
| Perceive-IR Zhang et al. (2025) | <u>21.77</u> | <u>0.891</u> | <u>0.169</u> | <u>0.557</u> | <u>2.555</u> |
| **Perceive-IR**$^R$ Zhang et al. (2025) | **21.92** | **0.894** | **0.165** | **0.558** | **2.559** |

## 4.2 Generalization

### 4.2.1 Zero-shot Generalization to Real-World Scenes

As shown in Table 2, applying RestoreMore to Perceive-IR consistently improves its real-world generalization across deraining, dehazing, and denoising. On RealRain-1k-L Li et al. (2022), Perceive-IR$^R$ achieves the best PSNR and LPIPS of 27.55 dB and 0.369, respectively, while obtaining the best SSIM of 0.903 jointly with Defusion. Compared with the original Perceive-IR, RestoreMore improves PSNR by 0.24 dB and reduces LPIPS from 0.372 to 0.369. Similar improvements are observed on real-world hazy and noisy images. These consistent improvements indicate that continual capability expansion does not compromise the generalization ability of the pretrained model. Instead, the newly acquired restoration knowledge can improve robustness to real-world degradation distributions.

### 4.2.2 Generalization to an Unseen Degradation Type

We further evaluate whether the expanded model can generalize beyond the degradation types observed during training. As reported in Table 3, Perceive-IR$^R$ achieves the best results across all metrics. Although the improvements are moderate, their consistency across both full-reference and no-reference metrics suggests that continual capability expansion preserves, and can slightly enhance, the pretrained model's ability to generalize to degradation types never encountered during training.

### 4.2.3 Generalization to Unseen Degradation Severity

We additionally examine generalization to degradation severities outside the training distribution. The denoising task is trained with noise levels $\sigma \in \{15, 25, 50\}$, while evaluation is conducted at the unseen and substantially stronger noise levels of $\sigma = 60$ and $100$. As shown in Table 4, Perceive-IR$^R$ achieves the best performance in three of the four settings and remains within 0.01 dB of the best result on Urban100 at $\sigma = 60$. The gains remain consistent as the noise level moves beyond the training range, indicating that the additional capabilities introduced by RestoreMore do not lead to over-specialization to the observed degradation severities.

Table 4: Zero-shot generalization results on unseen noise levels.

| Methods | CBSD68 | | Urban100 | |
|---|---|---|---|---|
| | 60 | 100 | 60 | 100 |
| VLU-Net Zeng et al. (2025) | 27.11 | 20.59 | 27.64 | 21.53 |
| Perceive-IR Zhang et al. (2025) | <u>27.13</u> | 20.65 | 27.65 | <u>21.55</u> |
| Defusion Luo et al. (2025) | 27.11 | <u>20.72</u> | **27.67** | 21.49 |
| **Perceive-IR**$^R$ Zhang et al. (2025) | **27.27** | **20.74** | <u>27.66</u> | **21.58** |

Table 5: Ablation study on individual components of RestoreMore.

| Settings | Frozen | Capability routing | Expert routing | PSNR | △ PSNR |
|---|---|---|---|---|---|
| (a) | | | | 31.98 | - |
| (b) | ✔ | ✔ | | 32.56 | + 0.58 dB |
| (c) | ✔ | | ✔ | 32.69 | + 0.71 dB |
| (d) | | ✔ | ✔ | 32.25 | + 0.27 dB |
| (e) | ✔ | ✔ | ✔ | 33.26 | + 1.28 dB |

Table 6: Effects of the order in which new restoration tasks are learned.

| Method | R → S → H → N | N → R → S → H | H → N → R → S | S → H → R → N |
|---|---|---|---|---|
| PSNR | 33.26 | 33.19 | 33.24 | 33.20 |

### 4.3 Ablation Studies

#### 4.3.1 Effect of Each Component

We investigate the contribution of the main components of RestoreMore using ALGNet as the base restoration model. As shown in Table 5, setting (a) retrains ALGNet on the mixed datasets of all degradation types and serves as the baseline, achieving an average PSNR of 31.98 dB. In contrast, settings (b)–(e) use only the dataset of the newly arriving degradation during continual expansion. Despite the absence of historical raw data, all variants outperform the jointly retrained baseline, demonstrating the effectiveness of expanding restoration capabilities from newly available task data.

Freezing the pretrained restoration model plays an important role in preserving previously learned mappings. When the frozen capability anchor is combined with capability routing or expert routing, settings (b) and (c) improve the baseline by 0.58 dB and 0.71 dB, respectively. Expert routing provides a slightly larger gain, suggesting that selectively reusing complementary historical experts is particularly beneficial for acquiring new restoration capabilities. By comparison, jointly applying capability and expert routing without freezing the original model in setting (d) yields only a 0.27 dB improvement. This result indicates that routing alone cannot fully prevent interference with previously learned restoration knowledge when the backbone remains trainable. The complete RestoreMore in setting (e) combines the frozen capability anchor with both levels of routing and achieves the best average PSNR of 33.26 dB, outperforming the mixed-data retraining baseline by 1.28 dB. It also improves upon the strongest partial variant (c) by 0.57 dB. These results show that the three components are complementary.

#### 4.3.2 Effect of Task Order

We further evaluate the sensitivity of RestoreMore to the order in which new restoration tasks are introduced. As shown in Table 6, the average PSNR remains highly stable across four different task sequences. This small variation indicates that RestoreMore is not strongly dependent on a specific task order. The frozen capability anchor limits interference with previously learned restoration mappings, while capability-oriented routing and expert reuse allow newly arriving tasks to selectively exploit relevant historical knowledge regardless of when they are introduced.

## 5 Conclusion

In this work, we propose RestoreMore, which preserves the pretrained restorer as a frozen capability anchor and incrementally introduces lightweight residual expansion modules for newly arriving degradations. A capability-oriented bi-level routing mechanism was developed to first identify relevant restoration capabilities and then sparsely select complementary degradation experts at multiple feature stages. This design enables effective reuse of accumulated restoration knowledge for new tasks while progressively enriching the expert bank for subsequent capability expansion. Extensive experiments across diverse restoration benchmarks and model architectures demonstrate that RestoreMore consistently acquires new restoration abilities while preserving, and in many cases improving, existing ones.

## AI USE STATEMENT

In this work, we used generative AI tools to assist with manuscript editing and scientific figure preparation. GPT-based tools were used to improve the language, grammar, and readability of the manuscript without changing the underlying scientific content. Generative AI was additionally used to assist in the preparation of Fig. 5, using only the experimental values obtained by the authors as input. The underlying data and numerical results in the figure were not generated or modified by AI.

All AI-assisted outputs were manually reviewed and verified by the authors. The authors take full responsibility for the final manuscript, including all text, figures, experimental results, scientific claims, and conclusions.

## REFERENCES

Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1122–1131, 2017. doi: 10.1109/CVPRW.2017.150.

Pablo Arbeláez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):898–916, 2011. doi: 10.1109/TPAMI.2010.161.

Wei-Ting Chen, Hao-Yu Fang, Jian-Jiun Ding, Cheng-Che Tsai, and Sy-Yen Kuo. Jstasr: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16*, pp. 754–770. Springer, 2020.

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

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*, 26:8184–8196, 2024.

Lark Kwon Choi, Jaehee You, and Alan Conrad Bovik. Referenceless prediction of perceptual fog density and perceptual image defogging. *IEEE Transactions on Image Processing*, 24(11): 3888–3901, 2015.

Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Image restoration via frequency selection. *TPAMI*, pp. 1–16, 2023.

Yuning Cui, Syed Waqas Zamir, Salman Khan, Alois Knoll, Mubarak Shah, and Fahad Shahbaz Khan. AdaIR: Adaptive all-in-one image restoration via frequency mining and modulation. In *The Thirteenth International Conference on Learning Representations*, 2025.

Yuning Cui, Syed Waqas Zamir, Ming-Hsuan Yang, Alois Knoll, Fahad Shahbaz Khan, and Salman Khan. Starir: Convolutional image restoration with spatial-frequency fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–18, 2026.

Hansen Feng, Lizhi Wang, Yiqi Huang, Yuzhi Wang, Lin Zhu, and Hua Huang. Learning physics-informed noise models from dark frames for low-light raw image denoising. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(4):3952–3969, 2026.

Rich Franzen. Kodak lossless true color image suite. *source: http://r0k. us/graphics/kodak*, 4(2):9, 1999.

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)*, pp. 1715–1723, 2017. doi: 10.1109/CVPR.2017.186.

Hu Gao, Bowen Ma, Ying Zhang, Jingfan Yang, Jing Yang, and Depeng Dang. Learning enriched features via selective state spaces model for efficient image deblurring. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 710–718, 2024a.

Hu Gao, Jing Yang, Ying Zhang, Ning Wang, Jingfan Yang, and Depeng Dang. Prompt-based ingredient-oriented all-in-one image restoration. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(10):9458–9471, 2024b.

Hu Gao, Jing Yang, Ying Zhang, Jingfan Yang, Bowen Ma, and Depeng Dang. Learning optimal combination patterns for lightweight stereo image super-resolution. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 5566–5574, 2024c.

Hu Gao, Bowen Ma, Ying Zhang, Jingfan Yang, Jing Yang, and Depeng Dang. Frequency domain task-adaptive network for restoring images with combined degradations. *Pattern Recognition*, 158:111057, 2025a.

Hu Gao, Ying Zhang, Jing Yang, and Depeng Dang. Mixed hierarchy network for image restoration. *Pattern Recognition*, 161:111313, 2025b.

Hu Gao, Bowen Ma, Ying Zhang, Jingfan Yang, Jing Yang, Xingjian Wang, and Depeng Dang. Emphasizing crucial features for efficient image restoration. *Pattern Recognition*, pp. 113575, 2026a.

Hu Gao, Changshuo Wang, Yulong Chen, and Lizhuang Ma. Learning adaptive dynamical features via multi-$\tau$ liquid-mamba for all-in-one image restoration. *arXiv preprint arXiv:2606.22801*, 2026b.

Xiaoxuan Gong and Jie Ma. A minimalistic unified framework for incremental learning across image restoration tasks. volume 38, pp. 34000–34019, 2026.

Yubin Gu, Yuan Meng, Jiayi Ji, and Xiaoshuai Sun. Acl: Activating capability of linear attention for image restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 17913–17923, 2025.

Hang Guo, Yong Guo, Yaohua Zha, Yulun Zhang, Wenbo Li, Tao Dai, Shu-Tao Xia, and Yawei Li. Mambairv2: Attentive state space restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 28124–28133, 2025a.

Qing Guo, Hua Qi, Jingyang Sun, Felix Juefei-Xu, Lei Ma, Di Lin, Wei Feng, and Song Wang. Efficientderain+: Learning uncertainty-aware filtering via rainmix augmentation for high-efficiency deraining. *International Journal of Computer Vision*, 133(4):2111–2135, 2025b.

Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5197–5206, 2015.

Yitong Jiang, Zhaoyang Zhang, Tianfan Xue, and Jinwei Gu. Autodir: Automatic all-in-one image restoration with latent diffusion. In *Computer Vision – ECCV 2024: 18th European Conference, Milan, Italy, September 29–October 4, 2024, Proceedings, Part XL*, pp. 340–359, 2024.

D. Kingma and J. Ba. Adam: A method for stochastic optimization. *Computer Science*, 2014.

Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *TIP*, 28(1):492–505, 2018.

Wei Li, Qiming Zhang, Jing Zhang, Zhen Huang, Xinmei Tian, and Dacheng Tao. Toward real-world single image deraining: A new benchmark and beyond. *arXiv preprint arXiv:2206.05514*, 2022.

Yu Li, Robby T. Tan, Xiaojie Guo, Jiangbo Lu, and Michael S. Brown. Rain streak removal using layer priors. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2736–2744, 2016. doi: 10.1109/CVPR.2016.299.

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

Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 136–144, 2017.

Hanzhou Liu, Chengkai Liu, Jiacong Xu, Peng Jiang, and Mi Lu. Xyscannet: An interpretable state space model for perceptual image deblurring. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pp. 779–789, 2025.

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. doi: 10.1109/TIP.2018.2806202.

David Lopez-Paz and Marc'Aurelio Ranzato. Gradient episodic memory for continual learning. volume 30, 2017.

I. Loshchilov and F. Hutter. Sgdr: Stochastic gradient descent with warm restarts. 2016.

Meng Lou, Yunxiang Fu, and Yizhou Yu. Scaling continual learning to 300+ tasks with bi-level routing mixture-of-experts. In *Forty-third International Conference on Machine Learning*, 2026.

Xin Lu, Jie Xiao, Yurui Zhu, and Xueyang Fu. Continuous adverse weather removal via degradation-aware distillation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 28113–28123, 2025.

Wenyang Luo, Haina Qin, Zewen Chen, Libin Wang, Dandan Zheng, Yuming Li, Yufan Liu, Bing Li, and Weiming Hu. Visual-instructed degradation diffusion for all-in-one image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12764–12777, June 2025.

Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing*, 26(2):1004–1016, 2016.

Jiawei Mao, Yu Yang, Xuesong Yin, Ling Shao, and Hao Tang. All-in-one transformer for image restoration under adverse weather degradations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(6):6628–6641, 2026.

D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pp. 416–423 vol.2, 2001.

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

Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. *CVPR*, pp. 257–265, 2016.

Karen Panetta, Chen Gao, and Sos Agaian. Human-visual-system-inspired underwater image quality measures. *IEEE journal of oceanic engineering*, 41(3):541–551, 2015.

Xiting Peng, Yandi Zhang, Xiaoyu Zhang, Junyi Wang, Shi Bai, Yubo Cao, Hongye Chen, and Tianshu Li. Fcs-ednet: Exploring magnetic particle imaging deblurring with neural network. *IEEE Transactions on Image Processing*, 35:480–494, 2026. doi: 10.1109/TIP.2025.3648497.

Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 2001–2010, 2017.

Jianxiang Rong, Hua Huang, and Jia Li. Imu-assisted accurate blur kernel re-estimation in non-uniform camera shake deblurring. *IEEE Transactions on Image Processing*, 33:3823–3838, 2024.

Xiongfei Su, Siyuan Li, Yuning Cui, Miao Cao, Yulun Zhang, Zheng Chen, Zongliang Wu, Zedong Wang, Yuanlong Zhang, and Xin Yuan. Prior-guided hierarchical harmonization network for efficient image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 7042–7050, 2025.

Hossein Talebi and Peyman Milanfar. Nima: Neural image assessment. *IEEE transactions on image processing*, 27(8):3998–4011, 2018.

Xiaole Tang, Xiaoyi He, Jiayi Xu, Xiang Gu, and Jian Sun. Learning continuous wasserstein barycenter space for generalized all-in-one image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–16, 2026. doi: 10.1109/TPAMI.2026.3669121.

Huiyi Wang, Haodong Lu, Lina Yao, and Dong Gong. Self-expansion of pre-trained models with mixture of adapters for continual learning. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10087–10098, 2025.

Yuanbo Wen, Tao Gao, and Ting Chen. Unpaired photo-realistic image deraining with energy-informed diffusion model. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 360–369, 2024.

Miao Yang and Arcot Sowmya. An underwater color image quality evaluation metric. *IEEE transactions on image processing*, 24(12):6062–6071, 2015.

Wenhan Yang, Robby T. Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. *CVPR*, pp. 1685–1694, 2016.

Jiazuo Yu, Yunzhi Zhuge, Lu Zhang, Ping Hu, Dong Wang, Huchuan Lu, and You He. Boosting continual learning of vision-language models via mixture-of-experts adapters. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23219–23230, 2024.

Haijin Zeng, Jiezhang Cao, Kai Zhang, Yongyong Chen, Hiep Luong, and Wilfried Philips. Unmixing diffusion for self-supervised hyperspectral image denoising. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 27820–27830, 2024.

Haijin Zeng, Xiangming Wang, Yongyong Chen, Jingyong Su, and Jie Liu. Vision-language gradient descent-driven all-in-one deep unfolding networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7524–7533, June 2025.

He Zhang and Vishal M. Patel. Density-aware single image de-raining using a multi-stream dense network. *CVPR*, pp. 695–704, 2018.

He Zhang, Vishwanath A. Sindagi, and Vishal M. Patel. Image de-raining using a conditional generative adversarial network. *TCSVT*, 30:3943–3956, 2017.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 586–595, 2018.

Xu Zhang, Jiaqi Ma, Guoli Wang, Qian Zhang, Huan Zhang, and Lefei Zhang. Perceive-ir: Learning to perceive degradation better for all-in-one image restoration. *IEEE Transactions on Image Processing*, pp. 1–1, 2025.

# A Appendix

## A.1 Initialization of the Capability Bank

RestoreMore starts from a pretrained model $F_0$ with initial capability set $\mathcal{S}_0$. Before continual expansion, we perform a one-time indexing procedure to construct the stage-wise capability representations required by the first routing level. This procedure is performed only for the initial capability set and does not update the pretrained restoration backbone.

For an initial degradation $d_i \in \mathcal{S}_0$, its samples are forwarded through $F_0$ and the degradation descriptor $q_l$ is extracted at each selected feature stage. The projection $\psi_l$ is calibrated during this stage and is subsequently frozen so that descriptors from later tasks remain in the same representation space. The initial capability prototype is computed as:

$$p_{i,l} = \mathrm{Norm}\left[\frac{1}{|\mathcal{D}_i|}\sum_{x\in\mathcal{D}_i} q_l(x)\right]. \tag{13}$$

The capability router $R_{i,l}$ is initialized jointly during this indexing stage. Since the corresponding restoration capability is already represented by $F_0$, the null expert provides the default residual option for these initial capabilities. For a newly arriving degradation $d_{B+t}$, its prototype is initialized from the current training features and updated using exponential moving averaging:

$$p_{B+t,l} \leftarrow \eta p_{B+t,l} + (1-\eta)\frac{1}{|\mathcal{B}_t|}\sum_{x\in\mathcal{B}_t} q_l(x), \tag{14}$$

where $\mathcal{B}_t \subset \mathcal{D}_t$ denotes the current mini-batch. Once stage $t$ is completed, the prototype is frozen. After the initial indexing stage, historical raw training images are discarded. All subsequent incremental stages therefore use only the dataset of the newly arriving degradation.

### A.2 Why Query–Key Routing Supports Continual Expansion

A standard MoE router usually predicts one gating logit for each available expert. Its output dimension therefore grows with the expert bank $R_{i,l}(q_l) \in \mathbb{R}^{|\mathcal{E}_l^{(t)}|}$. Adding a new expert would consequently require expanding the output layer of historical routers, which conflicts with our objective of keeping previously learned components unchanged. RestoreMore instead separates capability routing from expert indexing. Each capability router produces a fixed-dimensional query $r_{i,l} \in \mathbb{R}^{d_r}$, while every expert owns a routing key $k_{j,l} \in \mathbb{R}^{d_r}$. Their compatibility is evaluated through Eq. equation 9. Consequently, adding a new capability only appends $(E_{t,l}, k_{t,l})$ to the expert bank, without changing any historical router.

This design separates knowledge accumulation from historical parameter preservation. New restoration knowledge is accumulated through an expanding expert–key bank, whereas previously learned routers and experts remain structurally unchanged. The same query–key space also allows newly introduced experts to become candidates for future restoration tasks.

### A.3 Optimization and Cold-Start Strategy

At incremental stage $t$, the pretrained backbone and all historical capability components are frozen. The trainable parameters are:

$$\Theta_t = \{R_{B+t,l}, E_{t,l}, k_{t,l}\}_{l=1}^{L}, \tag{15}$$

while the prototype $p_{B+t,l}$ is updated using Eq. equation 14.

For a training pair $(x_t, y_t) \in \mathcal{D}_t$, the restoration objective is:

$$\begin{aligned}\mathcal{L}_{\mathrm{res}} =& \sqrt{\|F_t(x_t) - y_t\|_2^2 + \epsilon^2} \\ &+ \lambda_f \left\|\mathcal{F}(F_t(x_t)) - \mathcal{F}(y_t)\right\|_1.\end{aligned} \tag{16}$$

During training, the current capability is always included in the Top-$M$ capability set:

$$\mathcal{I}_l^{\mathrm{train}} = \{B+t\} \cup \mathrm{Top}_{M-1}\left(\{\pi_{i,l}\}_{i\neq B+t}\right). \tag{17}$$

Similarly, the newly introduced expert is forced into the Top-$K$ set of its own capability router:

$$\mathcal{J}_{B+t,l}^{\mathrm{train}} = \{t\} \cup \mathrm{Top}_{K-1}\left(\{g_{B+t,j,l}\}_{j\neq t}\right). \tag{18}$$

This cold-start strategy guarantees that the newly introduced router, expert, and routing key receive sufficient optimization before their routing scores become reliable. Meanwhile, the remaining $M-1$ capabilities and $K-1$ experts are dynamically selected from previously learned components, allowing the new task to reuse historical restoration knowledge from the beginning of training. No historical image replay or output-level knowledge distillation is used. After optimization, the new prototype, router, expert, and routing key are frozen and added to the capability bank.

### A.4 Expert Architecture and Backbone Instantiation

Each degradation expert is implemented as a lightweight bottleneck spatial adapter:

$$E_{j,l}(u) = \gamma_{j,l} W_{j,l}^{\mathrm{up}} \left[ \mathrm{DWConv}_{3\times 3} \left( \sigma \left( W_{j,l}^{\mathrm{down}} \, \mathrm{Norm}(u) \right) \right) \right], \tag{19}$$

where $W_{j,l}^{\mathrm{down}}$ reduces the channel dimension, $\mathrm{DWConv}_{3\times 3}$ captures local spatial degradation patterns, and $W_{j,l}^{\mathrm{up}}$ restores the original feature dimension. The residual scale $\gamma_{j,l}$ is initialized to zero. RestoreMore is attached to the outputs of major feature-processing stages rather than every individual block. The same residual expansion interface $h_l = u_l + \mathcal{M}_l^{(t)}(u_l)$ is used for CNN-, Transformer-, and Mamba-based restorers. Therefore, RestoreMore does not modify the internal convolution, attention, or state-space operations of the original backbone.

### A.5 Sparse Expert Computation

Assume that $t$ incremental experts have been accumulated at each selected stage and that one expert requires computation $C_E$. A dense expert mixture evaluates all experts and incurs $\mathcal{O}(tC_E)$ expert computation. RestoreMore selects $M$ capability routers, each activating at most $K$ experts. The dominant expert computation is therefore bounded by $\mathcal{O}(MKC_E)$. Because $M$ and $K$ are fixed, the number of evaluated spatial experts does not grow with the task sequence.

Capability-prototype matching and query–key scoring still increase with the size of the capability bank. However, these operations involve only low-dimensional vector similarities and are substantially cheaper than evaluating spatial restoration experts. Thus, RestoreMore maintains sparse expert computation while progressively enlarging its available restoration knowledge.

### A.6 Experimental Setup

#### A.6.1 Datasets

**i) Image deraining.** For image deraining, we use 13,712 paired clean–rain images collected from multiple datasets Yang et al. (2016); Zhang et al. (2017); Fu et al. (2017); Li et al. (2016) for training. Evaluation is conducted on four commonly used benchmarks, including Rain100H Yang et al. (2016), Rain100L Yang et al. (2016), Test100 Zhang et al. (2017), and Test1200 Zhang & Patel (2018).

**ii) Image desnowing.** For image desnowing, we adopt Snow100K Liu et al. (2018), SRRS Chen et al. (2020), and CSD Chen et al. (2021). Following the training protocol of previous work Cui et al. (2023), 2,500 paired images are randomly sampled for training and 2,000 images are used for evaluation.

**iii) Image dehazing.** For image dehazing, we use the daytime synthetic subsets of RESIDE Li et al. (2018), including the Indoor Training Set (ITS), Outdoor Training Set (OTS), and Synthetic Objective Testing Set (SOTS). The model is trained separately on ITS and OTS and evaluated on SOTS-Indoor and SOTS-Outdoor, respectively, each containing 500 paired images.

**iv) Image deblurring.** For image deblurring, we use the GoPro dataset Nah et al. (2016), which contains 2,103 training pairs and 1,111 testing pairs.

**v) Image denoising.** For image denoising, we construct a composite training set containing 800 images from DIV2K Agustsson & Timofte (2017), 2,650 images from Flickr2K Lim et al. (2017), 400 images from BSD500 Arbeláez et al. (2011), and 4,744 images from WED Ma et al. (2016). Additive white Gaussian noise is synthesized with the noise level $\sigma$ randomly selected from $\{15, 25, 50\}$. The trained model is evaluated on CBSD68 Martin et al. (2001), Urban100 Huang et al. (2015), and Kodak24 Franzen (1999).

For the joint-retraining baseline, the training sets of the above restoration tasks are combined into a mixed dataset. For the unified quantitative comparison, we report results on Rain100L Yang et al. (2016) for deraining, Snow100K Liu et al. (2018) for desnowing, SOTS-Outdoor Li et al. (2018) for dehazing, GoPro Nah et al. (2016) for deblurring, and CBSD68 Martin et al. (2001) with $\sigma = 25$ for denoising.

Table 7: Initial and newly acquired restoration capabilities of the pretrained models used in our experiments. R, S, H, B, and N denote deraining, desnowing, dehazing, deblurring, and denoising, respectively.

| Pretrained model | Architecture | Initial capabilities | Newly acquired capabilities |
|---|---|---|---|
| ALGNet Gao et al. (2024a) | Mamba | B | R, S, H, N |
| Perceive-IR Zhang et al. (2025) | Transformer | R, H, B, N | S |
| StarIR Cui et al. (2026) | CNN | R, H, N | S, B |

### A.6.2 Training Details

All models are optimized using Adam Kingma & Ba (2014) with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The initial learning rate is set to $2 \times 10^{-4}$ and gradually decayed to $1 \times 10^{-7}$ using cosine annealing Loshchilov & Hutter (2016). During training, $256 \times 256$ image patches are randomly cropped with a batch size of 32, and each model is optimized for $4 \times 10^5$ iterations. Random horizontal and vertical flipping is adopted for data augmentation. For fair comparison, all deep learning-based baselines are fine-tuned or retrained following the hyperparameter settings reported in their original papers.

### A.6.3 Initial and Incrementally Acquired Restoration Capabilities

We evaluate RestoreMore using pretrained restoration models built on different architectural families. Their original restoration capabilities are directly inherited from the corresponding pretrained models, while RestoreMore is used only to learn degradation types that are not initially supported.

We use the abbreviations R, S, H, B, and N to denote deraining, desnowing, dehazing, deblurring, and denoising, respectively. The capability configurations used in our experiments are summarized in Table 7. ALGNet initially provides only image deblurring capability. We therefore use it to examine the more challenging setting in which a task-specific pretrained model is progressively expanded toward multi-degradation restoration. Starting from B, RestoreMore sequentially learns four additional capabilities, R, S, H, and N, without revisiting the original deblurring training data. Perceive-IR already supports R, H, B, and N in its pretrained form. RestoreMore is applied only to the previously unsupported desnowing task, allowing us to evaluate whether an existing all-in-one model can be extended beyond its original task boundary. StarIR initially supports R, H, and N. RestoreMore further introduces S and B, providing another all-in-one-to-more setting with multiple incremental expansion stages. These configurations together evaluate both task-specific-to-more and all-in-one-to-more capability expansion within the same framework.

For experiments involving multiple newly introduced tasks, the exact incremental orders are specified together with the corresponding experimental results. In particular, ALGNet uses B as its initial capability, while R→S→H→N is used as the default incremental sequence unless otherwise stated.

### A.6.4 Default Hyperparameters

Unless otherwise specified, capability expansion modules are inserted at the outputs of the major feature stages of each pretrained restorer. We use $M = 2$, $K = 3$ for capability and expert selection, respectively. When fewer than $K$ experts are available during early incremental stages, all available experts are considered. The routing temperatures are set to $\tau_c = 0.1$, $\tau_e = 1.0$, and the frequency-domain reconstruction weight is set to $\lambda_f = 0.1$. The same RestoreMore configuration is used across CNN-, Transformer-, and Mamba-based restorers unless otherwise specified.

### A.6.5 Evaluation Metrics

We evaluate performance using both reference-based and no-reference metrics. The reference-based metrics include Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) Zhang et al. (2018). The no-reference metrics include the Underwater Colour Image Quality Evaluation Metric (UCIQE) Yang & Sowmya (2015), Underwater Image Quality Measure (UIQM) Panetta et al. (2015), Fog Aware Density Evaluator (FADE) Choi et al. (2015), Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) Mittal et al. (2011), and Neural Image Assessment (NIMA) Talebi & Milanfar (2018). Among them,

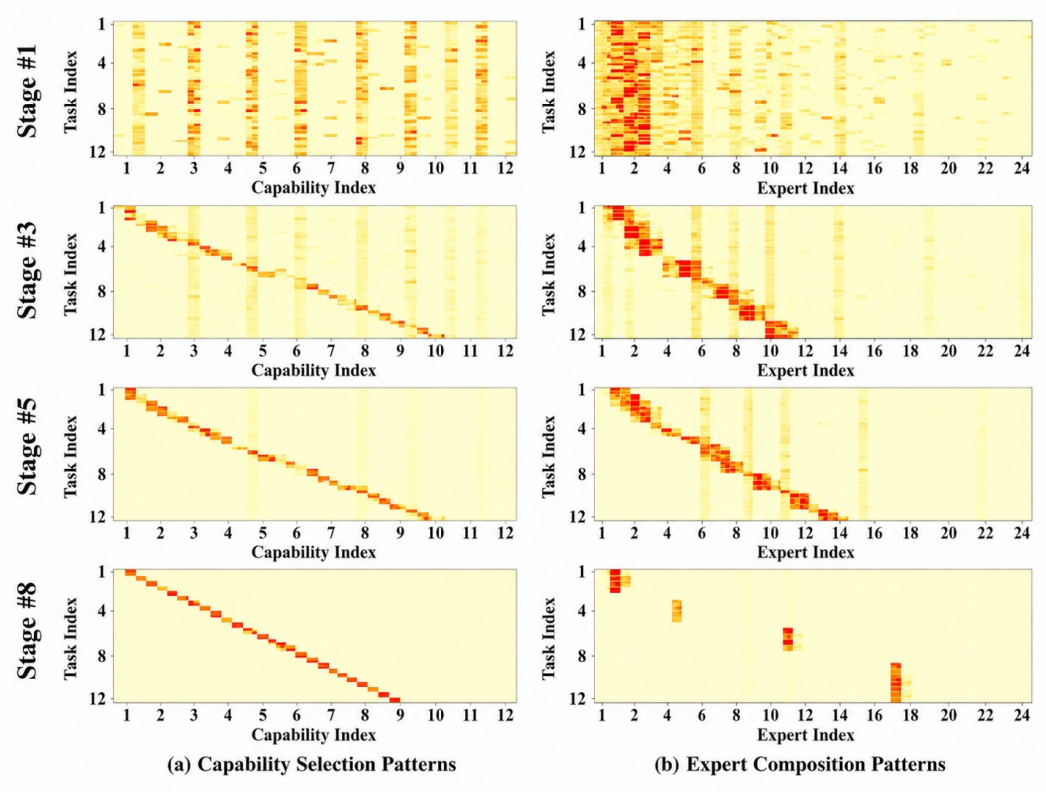


Figure 4: Visualization of the stage-wise bi-level routing patterns in RestoreMore. The left column shows the capability-selection patterns of the first routing level, while the right column shows the expert-composition patterns of the second routing level.

UCIQE and UIQM are specifically designed for underwater image restoration evaluation, while FADE, BRISQUE, and NIMA are commonly used to assess dehazing performance in real-world scenarios. For PSNR, SSIM, UCIQE, UIQM, and NIMA, higher values indicate better performance, whereas lower values are preferred for LPIPS, FADE, and BRISQUE. In the tables, the best and second-best results are highlighted in bold and underlined, respectively.

## A.7 More Experiments

### A.7.1 Analysis of Bi-level Routing Patterns

Fig. 4 visualizes the learned routing behaviors of RestoreMore at different feature stages. The left column shows the capability-selection patterns from the first routing level, and the right column shows the expert-composition patterns from the second routing level. Several observations can be made.

First, the capability-selection patterns become progressively more structured from shallow to deep stages. At Stage 1, the routing responses are relatively dispersed, indicating that shallow features tend to retrieve multiple capabilities and emphasize broadly shared low-level restoration cues. As the stage goes deeper, the activations gradually concentrate around task-aligned regions. In particular, Stage 3, Stage 5, and Stage 8 exhibit increasingly clear diagonal structures, suggesting that deeper representations are more discriminative and can more accurately identify the capability most relevant to each restoration task. This behavior is consistent with our motivation that shallow features mainly capture generic statistics and textures, whereas deeper features encode more task-specific restoration semantics.

Second, the expert-composition patterns reveal a transition from shared knowledge reuse to more selective expert specialization. At Stage 1, a relatively broad set of experts, especially low-index

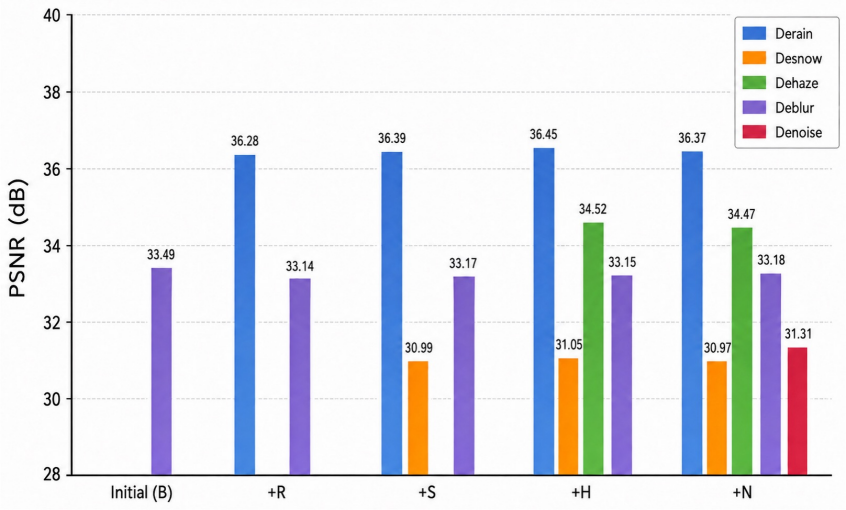


Figure 5: Capability evolution of ALGNet under the incremental sequence R→S→H→N. ALGNet initially supports only deblurring (B), and RestoreMore progressively introduces deraining (R), desnowing (S), dehazing (H), and denoising (N). Bars report the PSNR of all capabilities available after each incremental stage. Previously acquired capabilities remain stable as new restoration tasks are introduced.

experts, is activated across multiple tasks. This indicates that early-stage restoration prefers to reuse a set of shared experts that encode transferable low-level priors. In contrast, the deeper stages show increasingly sparse and localized activations. At Stage 3 and Stage 5, the activated experts align more clearly with task-dependent regions, while Stage 8 exhibits highly concentrated responses on only a few experts. This demonstrates that the second routing level indeed performs selective expert composition rather than dense aggregation, enabling the model to preserve shared knowledge while allocating more specialized experts to different tasks.

Third, the overall sparsity of both columns verifies the effectiveness of the proposed bi-level routing mechanism. Although RestoreMore continually accumulates capability prototypes and experts, each task only activates a small subset of them at each stage. This observation is consistent with the Top-$M$ capability selection and Top-$K$ expert composition design, which prevents unnecessary interference among unrelated tasks and improves scalability as the supported degradation set grows.

### A.7.2 Capability Evolution during Continual Expansion

To further examine how restoration capabilities evolve during continual learning, Fig. 5 reports the performance of ALGNet after each incremental stage under the default R→S→H→N sequence. The original ALGNet model supports only deblurring and achieves 33.49 dB. After deraining is introduced, RestoreMore reaches 36.28 dB on deraining while retaining 33.14 dB on the original deblurring task, corresponding to only a 0.35 dB decrease despite acquiring a previously unsupported capability. As additional tasks are introduced, the previously acquired capabilities remain highly stable. After the final denoising stage, the model simultaneously supports all five restoration tasks, achieving 36.37, 30.97, 34.47, 33.18, and 31.31 dB on deraining, desnowing, dehazing, deblurring, and denoising, respectively. Notably, the original deblurring capability decreases by only 0.31 dB from 33.49 to 33.18 dB over four successive capability-expansion stages. Similarly, once introduced, deraining, desnowing, and dehazing exhibit only minor variations as subsequent tasks are learned. These results provide a stage-by-stage view of the stability–plasticity behavior of RestoreMore. The model continually acquires previously unsupported restoration capabilities while inducing only limited changes to those learned earlier.

Figs. 6 and 7 further visualize how the restoration capability of ALGNet evolves during continual expansion under the R→S→H→N sequence. Starting from the original deblurring model, ALGNet produces a visually sharp result that closely matches the clean target. After the first expansion stage, the resulting model jointly supports deraining and deblurring. The newly introduced deraining ca-

pability effectively suppresses dense rain streaks while recovering the underlying scene content, whereas the previously learned deblurring capability remains visually stable, with object boundaries and local structures still clearly restored. This provides an intuitive example of acquiring a new capability without noticeably sacrificing the original restoration pathway. After desnowing is further introduced, the expanded model successfully removes snow particles while preserving scene structures and natural appearance. Meanwhile, the deraining and deblurring results remain consistent with those obtained at the preceding stage. A similar behavior is observed after adding dehazing: the model substantially improves scene visibility and contrast on hazy inputs, while maintaining strong performance on the previously acquired deraining, desnowing, and deblurring tasks. In particular, fine structures such as object contours, vegetation, buildings, and local textures remain well preserved across the accumulated task set.

The final model, denoted by ALGNet$^R$(R+S+H+B+N), simultaneously supports all five restoration capabilities. As shown in Fig. 7, the newly acquired denoising capability effectively suppresses strong noise while retaining fine object textures. At the same time, the visual quality of deraining, desnowing, dehazing, and deblurring remains comparable to the corresponding earlier checkpoints. For example, rain and snow artifacts are removed without excessive smoothing, hazy scenes recover clear visibility, and blurred regions retain sharp structural details.




Figure 6: Qualitative evolution of ALGNet during the early stages of continual capability expansion. The original ALGNet initially supports only deblurring, and RestoreMore progressively introduces additional restoration capabilities. Representative results from intermediate checkpoints show that newly acquired tasks are effectively learned while previously available restoration capabilities remain visually stable.

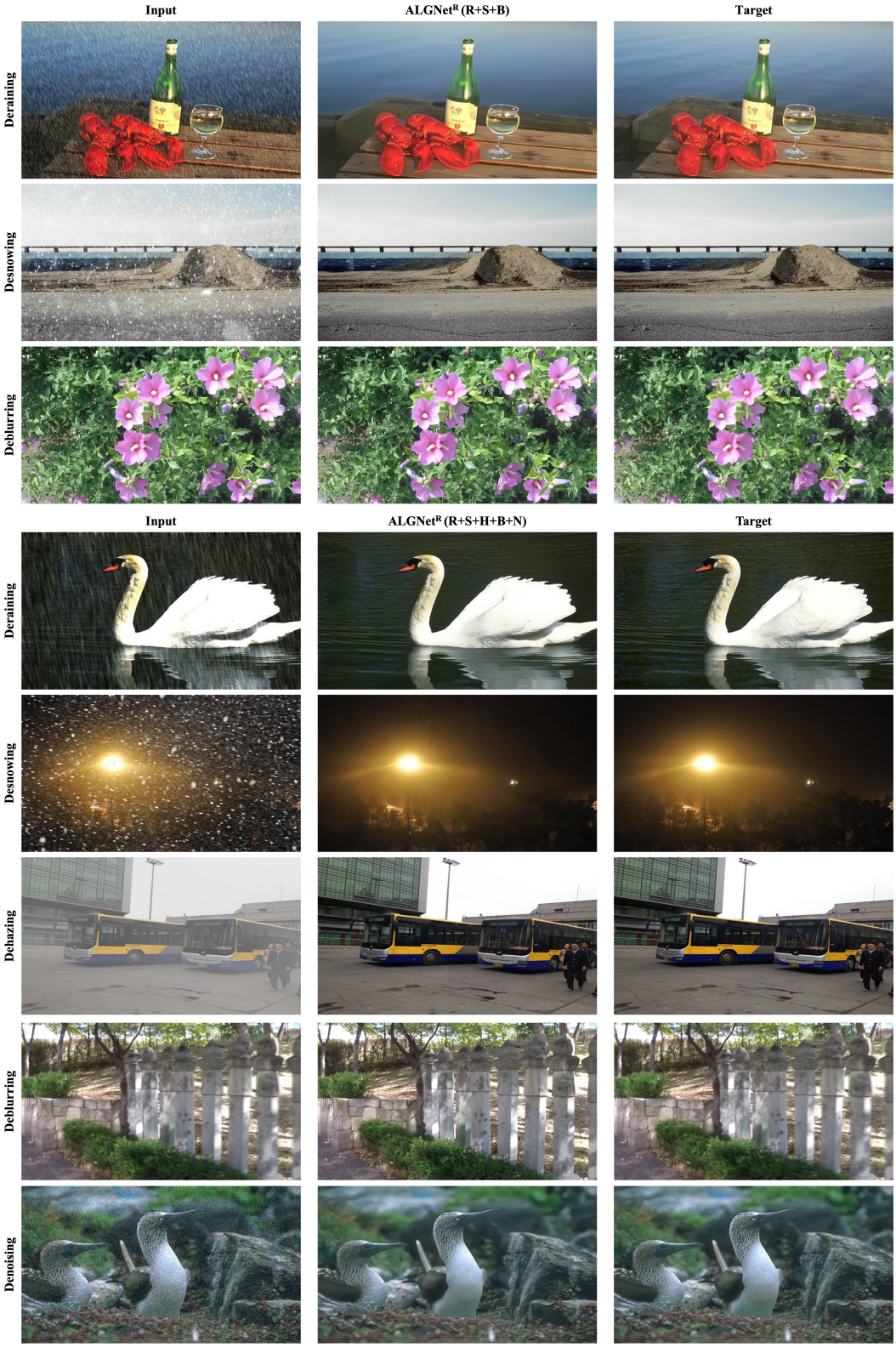


Figure 7: Qualitative results of ALGNet at later stages of continual capability expansion. Restore-More progressively expands the model under the R→S→H→N sequence, ultimately supporting deraining, desnowing, dehazing, deblurring, and denoising within a single expanded model.