TOWARDS ALL-IN-ONE IMAGE RESTORATION: A SPARSE MIXTURE-OF-EXPERTS FRAMEWORK WITH TASK-ADAPTIVE ROUTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Image restoration aims to reconstruct high-quality visual content from degraded observations caused by blur, noise, haze, or resolution loss. While recent deep learning approaches have achieved remarkable performance on individual restoration tasks, most existing methods still follow a task-specific paradigm that trains separate models for different degradations. This design limits their practicality in real-world scenarios where degradations are often unknown, mixed or spatially varying. In this paper, we propose MoERestore, a unified image restoration framework that integrates task-adaptive sparse mixture-of-experts routing with degradation-aware conditioning. The proposed architecture dynamically routes each input to a small subset of lightweight experts according to degradation characteristics, enabling task-adaptive capacity allocation while alleviating cross-task interference. To provide explicit guidance for restoration under ambiguous or mixed degradations, we introduce a degradation-aware module that leverages a vision-language model to generate semantic descriptions of corruption patterns. These descriptions are encoded as conditioning signals to guide the restoration process. In addition, a high-frequency feature extraction branch is incorporated to enhance fine-grained texture reconstruction. Extensive experiments on super resolution, deblurring, dehazing, and denoising benchmarks demonstrate that MoERestore achieves strong and consistent performance across diverse restoration scenarios and generalizes well to challenging real-world degradations.

1 INTRODUCTION

Image restoration aims to reconstruct high quality visual content from degraded observations caused by blur, noise, haze, and resolution loss. With the rapid progress of deep learning, significant advances have been achieved in task-specific restoration problems, such as super resolution Huang et al. (2020); Wang et al. (2021); Chen et al. (2025b), deblurring Kupyn et al. (2018); Nah et al. (2017); Tsai et al. (2022), dehazing Cai et al. (2016); Dong et al. (2020); Song et al. (2023), and denoising Zhang et al. (2017; 2018a); Shen et al. (2023). Despite these successes, most existing methods are trained for a single degradation type under simplified assumptions. This task-specific design limits their applicability in real-world scenarios, where degradations are often unknown, mixed, or spatially varying.

To address this limitation, recent studies have begun to explore all-in-one image restoration Zamir et al. (2021; 2022b); Chen et al. (2022); Cui et al. (2023); Xia et al. (2023), where a unified model is expected to address multiple restoration tasks within a single framework. Despite promising progress, building an effective unified restoration model remains challenging for two main reasons. First, different restoration tasks often impose conflicting optimization objectives. For example, denoising tends to suppress high-frequency noise while super resolution requires the reconstruction of high-frequency details. Training a single model with largely shared parameters can therefore introduce cross-task interference, where improvements on one task may degrade performance on another. Second, many unified restoration models Liu et al. (2022); Li et al. (2022); Tan et al. (2024); Zhang et al. (2023); Ma et al. (2023); Yao et al. (2024); Lin et al. (2023); Chen & Pei (2025); Luo et al. (2023b); Wang et al. (2024b); Conde et al. (2024) rely primarily on implicit degradation perception

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.

from visual inputs alone. When degradations are ambiguous or intertwined, the lack of effective degradation-aware guidance can lead to unstable or inconsistent restoration results.

In this work, we argue that effective unified restoration requires two key capabilities: task-adaptive capacity allocation and degradation-aware guidance. Task-adaptive capacity allocation allows the model to dynamically specialize different parts of the network for different restoration behaviors, reducing cross-task interference. Degradation-aware guidance, on the other hand, provides additional semantic cues about corruption patterns when visual evidence alone is insufficient.

Motivated by these observations, we propose MoERestore, a unified image restoration framework that integrates sparse mixture-of-experts routing with degradation-aware conditioning. Specifically, we introduce a task-adaptive sparse mixture-of-experts (Sparse-MoE) routing architecture that dynamically routes each input image to a small subset of lightweight restoration experts according to its degradation characteristics. This routing mechanism enables task-specialized processing while maintaining parameter efficiency. To provide explicit degradation guidance, we further incorporate a degradation-aware module based on a vision-language model that generates semantic descriptions of corruption patterns. These descriptions are encoded as conditioning signals to guide the restoration process. In addition, we introduce a high-frequency feature extraction branch that captures fine-grained texture information directly from the pixel space, complementing the latent diffusion representation. To evaluate unified restoration models under realistic conditions, we collect a real-world test set consisting of 60 in-the-wild images and adopt a vision-language model based perceptual evaluation protocol that enables reference-free assessment of restoration quality. Extensive experiments across super-resolution, deblurring, dehazing, and denoising benchmarks demonstrate that the proposed framework achieves strong and consistent performance across diverse restoration scenarios.Our main contributions are summarized as follows.

- We propose MoERestore, a unified image restoration framework that integrates sparse mixture-of-experts routing with degradation-aware conditioning, enabling a single model to handle diverse restoration tasks within one architecture.
- We design a task-adaptive sparse-MoE routing mechanism that dynamically allocates lightweight restoration experts to reduce cross-task interference.
- We introduce a degradation-aware module that leverages semantic degradation descriptions to guide restoration under unknown or mixed degradations.
- Extensive experiments on benchmarks and real-world images demonstrate that MoERestore achieves strong and consistent performance across multiple restoration scenarios.

2 Related Work

2.1 Task-Specific Image Restoration

Image restoration, a cornerstone of low-level vision, has undergone a paradigm shift from analytical modeling to deep learning. Early classical methods Chen et al. (2023); Li et al. (2022); Zhang et al. (2017) predominantly relied on handcrafted priors, such as Total Variation (TV) regularization and Non-Local Self-Similarity (e.g., BM3D Dabov et al. (2007)). The advent of Convolutional Neural Networks (CNNs), exemplified by DnCNN Zhang et al. (2017), and subsequent Transformer-based architectures like SwinIR Liang et al. (2021), have achieved breakthroughs in specific degradation tasks, including denoising, super-resolution (SR), and deraining. Recently, Transformers have further pushed the performance boundaries owing to their superior long-range dependency modeling. SwinIR leverages a shifted-window mechanism to achieve an SOTA balance across multiple tasks, while Restormer Zamir et al. (2022a) mitigates the computational overhead of high-resolution processing via an efficient transposed attention mechanism. Furthermore, HAT Chen et al. (2025c) integrates hybrid attention with an expanded receptive field to optimize reconstruction fidelity. Despite their success, these methods adhere to a ”one-model-per-task” design philosophy, which limits their scalability and deployment in practical, unpredictable environments.

2.2 All-in-One Image Restoration

To transcend the constraints of task-specific models, All-in-One Image Restoration (AiOIR) seeks to handle diverse degradations within a singular, unified framework. Various strategies have been explored to enhance multi-task learning efficiency: AirNet Li et al. (2022) employs contrastive learning to extract discriminative representations across degradations, while PromptIR Potlapalli et al. and ProRes Ma et al. (2023) introduce learnable prompts to explicitly guide the network in perceiving degradation types. FoundIR Li et al. (2025) further enhances model capacity through large-scale real-world data augmentation. To alleviate parameter competition and inter-task interference during joint training, the Mixture-of-Experts (MoE) architecture has been integrated into restoration pipelines Yang et al. (2024a); Zhang et al. (2024); Luo et al. (2023a); Zamfir et al. (2025); Lin et al. (2024a); Ai et al. (2024a). However, most existing AiOIR methods rely on the network's implicit learning of degradation-aware information, lacking explicit modeling of the restoration logic. Consequently, when confronted with complex joint degradations, these models often suffer from semantic misalignment, leading to physically inconsistent artifacts due to the absence of high-level guidance.

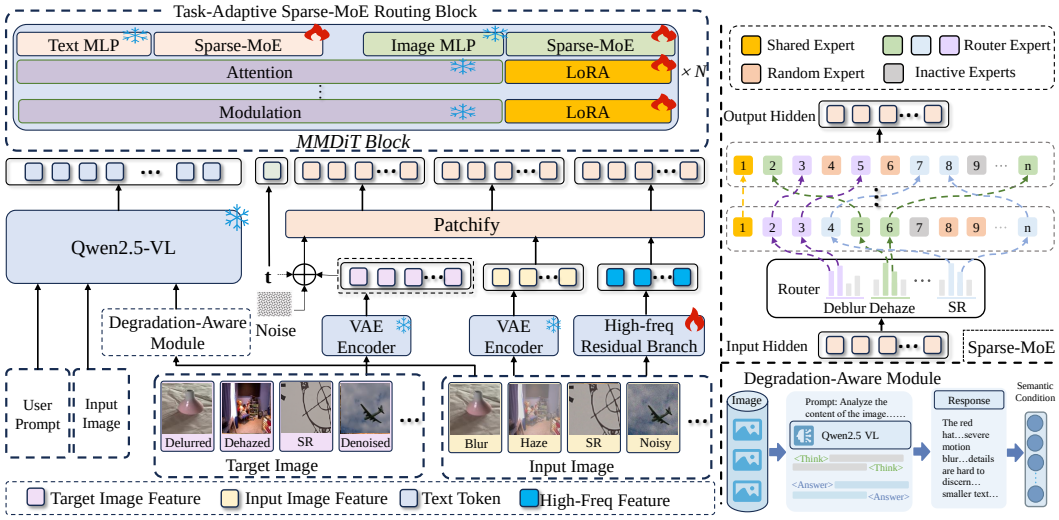

Figure 1: Overview of MoERestore.The input image and user prompt are encoded into text-conditional tokens using Qwen2.5-VL and high-frequency branches. Then, the input and target images are mapped to latent variables via a VAE encoder and also encoded into tokens. Training is then performed using a Task-Adaptive Sparse-MoE Routing Block. The top right side shows a schematic diagram of the Spare-MoE. The bottom right side shows the flow of the Degradation-Aware Module.

3 Method

3.1 Overview

In this section, we introduce MoERestore, a unified image restoration framework that integrates task-adaptive expert routing with degradation-aware conditioning, enabling a single model to handle diverse restoration tasks within one architecture. Given a degraded image x_{deg}, the goal of our model is to predict a restored image $\hat{x}$ that is visually faithful and semantically consistent with the underlying scene content. As illustrated in Figure 1, the framework consists of three key components. The first is a task-adaptive expert routing block, which dynamically allocates restoration capacity through a sparse mixture-of-experts routing mechanism. The second is a degradation-aware module, which provides semantic guidance about the degradation pattern using a vision-language model. The third is a high-frequency feature extraction branch, which introduces additional texture information to enhance fine-detail reconstruction. Formally, the overall restoration process can be written as

$$\hat{x} = \mathcal{F}(x_{deg}, p(x_{deg})), \tag{1}$$

where $\mathcal{F}$ denotes the proposed restoration model, x_{deg} is the degraded input, p is the semantic degradation description, and $\hat{x}$ is the restored output.

3.2 Task-Adaptive Sparse Mixture-of-Experts Routing

A major challenge in unified restoration is that different tasks often induce conflicting optimization behaviors. For example, denoising generally favors smooth and stable structures, whereas super resolution requires recovering sharp edges and high frequency details. When multiple restoration tasks are jointly optimized with largely shared parameters, the model often suffers from cross task interference. To alleviate this issue, we introduce a sparse mixture-of-experts design into the restoration backbone. As illustrated the task-adaptive sparse expert block in Figure 1, we implement a hybrid adaptation strategy that distinguishes. Instead of applying the same adaptation to all inputs, our model dynamically routes each sample to a small subset of lightweight experts according to its degradation characteristics.

Specifically, within each MMDiT block, we keep the core parameters of the attention and modulation layers frozen to preserve pre-trained structural priors, while augmenting them with trainable LoRA adapters to capture task-specific nuances. In contrast, for the Text MLP and Image MLP layers which serve as the primary knowledge repositories of the model, we replace the traditional feed-forward networks with our sparse-MoE architecture. This allows the model to allocate capacity in a task-adaptive manner while preserving efficient parameter sharing. For a selected transformation layer with input feature $\mathbf{x} \in \mathbb{R}^{D_{in}}$ and weight matrix $\mathbf{W} \in \mathbb{R}^{D_{out} \times D_{in}}$, the original transformation is augmented by three expert branches, including shared experts, routed experts, and randomly activated experts during training. The shared experts capture task agnostic restoration priors and are always active. Their output is defined as

$$\mathbf{h}_{shared} = \sum_{i=1}^{N_s} \alpha_i \, \mathrm{Expert}_i^{shared}(\mathbf{x}), \tag{2}$$

where N_s is the number of shared experts and α_i denotes the scaling coefficient of the ith shared expert.

To model task-specific restoration patterns, we further introduce routed experts. A gating network predicts an input dependent routing distribution

$$\mathbf{g}(\mathbf{x}) = \mathrm{Softmax}\big(\mathrm{Gate}(\mathbf{x}) + \mathbf{b}_{router}\big), \tag{3}$$

where $\mathbf{b}_{router}$ is a learnable bias term to balance expert load. Based on the routing scores, we select the top K experts and aggregate their outputs as

$$\mathbf{h}_{routed} = \sum_{j \in \mathcal{T}(\mathbf{x})} g_j(\mathbf{x}) \, \mathrm{Expert}_j^{routed}(\mathbf{x}), \tag{4}$$

where $\mathcal{T}(\mathbf{x})$ denotes the selected top K routed experts for input $\mathbf{x}$, and $g_j(\mathbf{x})$ is the corresponding routing weight.

To encourage more diverse expert utilization during training, we additionally activate a small number of random experts sampled from the remaining pool. Their output is written as

$$\mathbf{h}_{rand} = \sum_{k \in \mathcal{R}(\mathbf{x})} \beta_k \, \mathrm{Expert}_k^{rand}(\mathbf{x}), \tag{5}$$

where $\mathcal{R}(\mathbf{x})$ denotes the randomly selected expert set and β_k is the corresponding weight.

The final output of the adapted layer is obtained by combining the original backbone transformation and the expert responses

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{h}_{shared} + \mathbf{h}_{routed} + \mathbf{h}_{rand}. \tag{6}$$

This design balances shared restoration knowledge and task specific specialization within a unified framework. Shared experts capture common priors across restoration tasks, routed experts focus on degradation dependent processing, and random activation improves expert diversity during training.

3.3 SEMANTIC DEGRADATION-AWARE CONDITIONING

Visual evidence alone may be insufficient for reliable restoration when degradations are ambiguous, intertwined, or spatially non-uniform. To provide more informative guidance, we augment each training sample with a semantic degradation description that summarizes the corruption pattern and restoration intent. Compared with generic task labels, such semantic descriptions provide richer cues about both image content and degradation characteristics.

Given a degraded image x_{deg}, we utilize a Vision-Language Model (VLM) to generate a semantic prompt p that captures both the observed degradation and the desired restoration target. This prompt is then encoded into a conditioning representation $\mathbf{c} = \mathrm{Enc}_{text}(p)$, where Enc_{text} denotes the text encoding module. By leveraging this semantic guidance, the restored image is generated as:

$$\mathbf{c} = \mathrm{Enc}_{text}(p), \hat{x} = \mathcal{F}(x_{deg}, \mathbf{c}). \tag{7}$$

This formulation enables the model to align its restoration behavior with both scene structure and specific corruption patterns. Particularly in real-world scenarios characterized by unknown or complex degradations, such semantic conditioning provides critical degradation awareness, significantly enhancing the robustness and consistency of the restoration process.

3.4 HIGH-FREQUENCY FEATURE EXTRACTION BRANCH

To enhance the model's capability to capture fine-grained textures and local structures of the target image, we design an auxiliary high-frequency feature extraction branch. This branch adopts a lightweight Vision Transformer (ViT-S) architecture. Unlike the primary DiT, which operates within a low-dimensional latent space, this branch extracts high-frequency, patch-level features directly from the pixel space. This effectively captures the high-frequency details that are frequently lost during the latent encoding process. During the feature integration stage, the extracted ViT features are converted into high-frequency tokens via a patching operation. Rather than employing traditional global weighting methods, we adopt a sequence concatenation strategy, where these high-frequency tokens are directly appended to the DiT latent sequence. This design enables the Transformer's self-attention mechanism to explicitly retrieve fine-grained structural information from the reference image in a spatially aligned manner throughout the denoising process.

3.5 TRAINING OBJECTIVE

We optimize the proposed framework with a joint objective that balances reconstruction fidelity and perceptual quality. The primary supervision is the diffusion training loss, which governs the optimization of the generative restoration backbone. To further improve perceptual realism, we employ a perceptual loss that enforces feature consistency between the restored image and the ground truth image. We also introduce a high-frequency fidelity loss to better preserve fine details that are often difficult to recover under severe degradations.

The high-frequency loss is defined in the Fourier domain as

$$\mathcal{L}_{hf} = \left\| \mathcal{M}_{high} \odot \left(\mathcal{F}_{fft}(\hat{x}) - \mathcal{F}_{fft}(x_{gt}) \right) \right\|_1, \tag{8}$$

where $\mathcal{F}_{fft}$ denotes the Fourier transform, $\mathcal{M}_{high}$ is a mask that selects high frequency components, and $\odot$ denotes element wise multiplication.

The overall training objective is formulated as

$$\mathcal{L}_{total} = \lambda_d \mathcal{L}_{diff} + \lambda_p \mathcal{L}_{per} + \lambda_h \mathcal{L}_{hf} + \lambda_c \mathcal{L}_{color}, \tag{9}$$

where $\mathcal{L}_{diff}$ denotes the diffusion training loss, $\mathcal{L}_{per}$ denotes the perceptual loss, $\mathcal{L}_{color}$ denotes color-consistency loss to prevent chromatic shifts. By jointly optimizing these objectives, the model is encouraged to produce restored images that are not only visually plausible but also structurally faithful and rich in fine details.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETTINGS

Training Dataset. We combine publicly available datasets from diverse image restoration tasks to form our training set, including GoPro Nah et al. (2017), FFHQ Karras et al. (2019), LSDIR Li et al. (2023), ITS Li et al. (2018), RSBlur Rim et al. (2022). After filtering and preprocessing, the final training set contains approximately 1000K image pairs covering diverse degradation types such as blur, noise, haze, and resolution loss. To introduce degradation-aware semantic guidance during training, we further generate textual descriptions for each degraded image using the Qwen2.5-VL model. These image descriptions not only include content information about the images but also contain image restoration instructions (e.g., "make the image clearer"), without explicitly specifying the task type. These triplets serve as our training data. Detailed implementation and training configurations are provided in the Appendix.

Evaluation Datasets. To comprehensively evaluate the generalization capability of our model in all-in-one image restoration, we test the model on a wide range of benchmark datasets(detailed in the Appendix) across different tasks, including synthetic benchmark datasets, real-world benchmark datasets, and a more challenging real-world test set that we propose.

Evaluation Protocol. To comprehensively assess the performance of all-in-one image restoration methods, we adopt a multi-dimensional evaluation protocol that combines conventional fidelity metrics, perceptual quality metrics, and VLM-based perceptual assessment. For fidelity evaluation, we report PSNR and SSIM to measure reconstruction accuracy with respect to ground-truth images. For perceptual quality metrics, we use LPIPS Zhang et al. (2018b), MUSIQ Ke et al. (2021), CLIPIQA+ Wang et al. (2023), PIQE Venkatanath et al. (2015), MANIQA Yang et al. (2022), and PaQ-2-PiQ Ying et al. (2020) to evaluate restoration fidelity. Furthermore, we incorporate a VLM-based perceptual evaluation using Gemini to provide a complementary human-aligned assessment. The VLM takes both the degraded input image and the corresponding restored output as context for evaluation, enabling perceptual assessment without requiring ground-truth references.

4.2 QUANTITATIVE AND QUALITATIVE RESULTS

We evaluate our MoERestore on public benchmark datasets for different restoration tasks and report quantitative results in Table 1. We compare our method with state-of-the-art restoration methods, including all-in-one image restoration methods (i.e., PromptIR Potlapalli et al., TransWeather Valanarasu et al. (2022), DACLIP Luo et al. (2023b), DiffUIR Zheng et al. (2024), AutoDIR Jiang et al. (2024), InstructIR Conde et al. (2024), X-Restormer Chen et al. (2024), FoundIR Li et al. (2025), LucidFlux Fei et al. (2025)) and single-task restoration methods (i.e., Real-ESRGAN Wang et al. (2021), BSRGAN Zhang et al. (2021), StableSR Wang et al. (2024a), PASD Yang et al. (2024b), SeeSR Wu et al. (2024b), DreamClear Ai et al. (2024b), OSEDiff Wu et al. (2024a), SUPIR Yu et al. (2024), FaithDiff Chen et al. (2025b), DIP Zheng et al. (2022), NAFNet Chu et al. (2022), DiffBIR Lin et al. (2024b), UniRestore Chen et al. (2025a)). Based on the statistics of multiple evaluation metrics on these test sets, our method demonstrates optimal or second-best performance on most metrics.

As shown in Table 1, our method achieves significant improvements on multiple perceptual quality metrics in real-world super-resolution tasks. On real-world super-resolution datasets (RealPhoto60, RealLQ250), it attains the best scores on key perceptual metrics (like MUSIQ, CLIPIQA+, and MANIQA), with CLIPIQA+ improving by 10.7% on RealPhoto60. It also sets a new state-of-the-art on the synthetic DIV2K-Val dataset, boosting PSNR by 34.7% and SSIM by 50.8%, compared to the second best method LucidFlux. The results demonstrate that our method can effectively restore more natural details while maintaining high fidelity. Our method also demonstrates excellent performance on dehazing, denoising and deblur tasks. Our approach delivers superior performance across the entire range of metrics on the Dense-Haze dataset, and also best PSNR and SSIM on RESIDE. For denoising on PolyU, it also ranks first or second in no-reference metrics (PIQE/MANIQA), confirming superior perceptual quality with fewer artifacts. For deblurring, it achieves the best SSIM (0.8341) on the GoPro dataset. On datasets such as FoundIR, which contain various types of degradations, we also consistently achieve the best PSNR and SSIM. To evaluate the capability of our model in handling mixed degradations, we conducted extensive experiments on the FoundIR

Table 1: Quantitative comparison on various image restoration benchmarks. All methods and metrics are explicitly listed. Bold and underline indicate the best and second-best results, respectively. (B+N denotes the Blur + Noisy degradation and R+H denotes the Rain +Haze degradation in the FoundIR dataset)

Method	PSNR ↑	SSIM ↑	LPIPS ↓	MUSIQ ↑	CLIPIQA+ ↑	CLIPIQA ↑	Q-Align ↑	NIQE ↓	MANIQA ↑	NIMA ↑
DIV2K-Val (Synthetic Super-Resolution)										
ResShift	18.3802	0.4394	0.3738	60.5932	0.5583	0.6017	3.5761	6.1976	0.3421	5.0430
StableSR	18.3269	0.4819	0.3933	57.4246	0.5760	0.5002	3.4226	4.9810	0.2902	5.0341
SinSR	18.0956	0.4259	0.3919	66.0906	0.6128	**0.7166**	3.7336	5.3679	0.4341	5.1810
SeeSR	18.2529	0.4684	0.3497	71.4947	0.7116	0.7149	4.1167	4.2823	0.5104	5.2709
DreamClear	17.5701	0.4291	0.3621	65.8187	0.6585	0.6663	3.9323	4.1634	0.4369	5.1663
SUPIR	17.7567	0.4482	0.3785	67.9074	0.6719	0.5848	4.1659	**3.7701**	0.4148	5.1516
LucidFlux	15.4393	0.3837	0.4312	**73.9045**	**0.7492**	0.7034	**4.5311**	3.7283	**0.5819**	**5.4884**
Ours	**20.7900**	**0.5788**	**0.4933**	64.7600	0.6618	0.6658	4.1203	4.2616	0.5394	5.1015

Method	**RealPhoto60** (Real-world SR)					**RealLQ250** (Real-world SR)				
	MUSIQ ↑	CLIPIQA+ ↑	MANIQA ↑	PaQ-2-PiQ ↑	PIQE ↓	MUSIQ ↑	CLIPIQA+ ↑	Q-Align ↑	NIQE ↓	MANIQA ↑
Real-ESRGAN / ResShift	59.29	0.4398	0.5046	69.0438	25.0258	59.50	0.5529	3.6318	6.6326	0.3397
BSRGAN / StableSR	45.46	0.3397	0.3759	63.3800	-	57.25	0.5804	3.5586	4.6236	0.2937
SinSR	-	-	-	-	-	65.45	0.6054	3.7451	5.4425	0.4230
PASD / SeeSR	63.53	0.4787	0.5194	69.9744	29.2662	70.38	0.7034	4.1423	4.4383	0.4895
DreamClear	70.46	0.5273	0.6080	**75.8422**	26.0232	67.08	0.6810	4.0640	**3.8700**	0.4400
SeeSR / SUPIR	71.74	0.5956	0.6077	73.8402	26.7787	65.81	0.6532	4.1347	3.6591	0.3826
Ours	**71.85**	**0.6592**	**0.6211**	75.0652	**24.1500**	**68.73**	**0.7216**	**4.2715**	4.4842	**0.5552**

Methods	**Dense-HAZE (Dehazing)**					**PolyU (Denoising)**				
	PSNR ↑	SSIM ↑	LPIPS ↓	MUSIQ ↑	CLIPIQA+ ↑	MUSIQ ↑	CLIPIQA+ ↑	PIQE ↓	MANIQA ↑	PaQ-2-PiQ ↑
PromptIR	9.57	0.4333	0.7889	26.16	0.2406	51.80	0.3873	60.3508	0.5182	68.7471
TransWeather	10.51	0.4523	0.7900	27.85	0.2060	32.99	0.3111	80.8918	0.4030	59.8569
DA-CLIP	10.94	0.4590	0.7604	30.07	0.2451	42.13	0.3547	74.2320	0.4774	65.3832
DiffUIR	9.59	0.4326	0.7994	26.42	0.2680	45.53	0.3561	73.2232	0.4611	66.6793
AutoDIR	12.33	0.4862	0.7364	31.97	0.2167	37.67	0.3538	80.3762	0.4496	63.2148
InstructIR	11.03	0.4649	0.7456	32.96	0.1615	**52.14**	**0.4445**	40.8439	0.5254	70.7628
X-Restormer	9.57	0.4295	0.8049	24.09	0.2486	32.99	0.3111	80.8918	0.4030	59.8569
AgenticIR	10.11	0.3884	0.7960	27.73	0.2595	35.98	0.3186	79.2196	0.4580	62.2334
FoundIR	9.29	0.4307	0.7951	27.80	0.2522	49.92	0.4244	42.0997	**0.5293**	**70.9032**
Ours	**14.83**	**0.5089**	**0.6405**	**55.18**	**0.4526**	46.28	0.4344	**34.5812**	0.5292	68.2896

Methods	**HQ-NightRain (Deraining)**					**RS-Cloud (Declouding)**				
	PSNR ↑	SSIM ↑	LPIPS ↓	MUSIQ ↑	CLIPIQA+ ↑	PSNR ↑	SSIM ↑	LPIPS ↓	MUSIQ ↑	CLIPIQA+ ↑
PromptIR	10.43	0.4485	0.5684	38.70	0.3727	11.35	0.7716	0.2713	41.83	0.4739
TransWeather	12.88	0.4774	0.6642	28.58	0.2855	12.93	0.7081	0.3808	30.02	0.3390
DA-CLIP	14.78	0.5398	0.5717	32.11	0.3500	16.43	0.8028	0.2383	35.73	**0.4839**
DiffUIR	11.58	0.4827	0.5687	34.57	0.3277	13.82	0.8140	0.2602	38.44	0.4210
AutoDIR	11.67	0.4697	0.6114	31.65	**0.3766**	18.39	0.8093	0.2519	31.44	0.4390
InstructIR	10.92	0.3972	0.4802	**46.46**	0.3475	14.46	0.8647	0.1624	**47.00**	0.3538
X-Restormer	10.33	0.4376	0.6093	36.12	0.3420	11.48	0.7506	0.3252	35.99	0.3641
AgenticIR	16.29	0.5681	0.5797	32.16	0.3762	17.80	0.7992	0.2572	34.90	0.4297
FoundIR	11.57	0.5220	0.4708	41.68	0.3570	11.71	0.7527	0.3219	35.40	0.4295
Ours	**17.08**	**0.5848**	**0.3849**	38.58	0.3763	**20.11**	**0.8689**	**0.1438**	41.67	0.4291

Dataset	Metric	NAFNet	PromptIR	DiffUIR	AutoDIR	SUPIR	InstructIR	X-Restormer	FoundIR	Ours
GoPro	PSNR ↑	22.18	23.93	23.93	-	-	22.55	23.54	23.55	**23.94**
	SSIM ↑	0.8042	0.8221	0.8241	-	-	0.8184	0.8260	0.8263	**0.8341**
RESIDE	PSNR ↑	25.13	27.26	27.12	-	-	26.17	27.91	27.66	**28.01**
	SSIM ↑	0.8632	0.8957	0.8820	-	-	0.8721	0.9043	0.8971	**0.9122**
FoundIR(B+N)	PSNR ↑	-	22.93	-	21.27	21.44	21.03	22.90	23.87	**24.15**
	SSIM ↑	-	0.7587	-	0.7193	0.6268	0.7003	0.7544	0.7737	**0.7850**
FoundIR(H+R)	PSNR ↑	-	19.63	20.24	14.90	-	13.49	15.17	21.80	**22.14**
	SSIM ↑	-	0.6275	0.7268	0.4213	-	0.5535	0.4354	0.7153	**0.7482**

(Blur+Noisy) composite benchmark. Our method consistently achieves the best performance, surpassing all competing models in terms of both PSNR and SSIM metrics.

In addition, we provide qualitative restoration results to further validate the effectiveness of our method. As shown in Figure 2, we present detailed visual comparisons across multiple restoration tasks, including super-resolution, deblurring, dehazing, and denoising. To evaluate the capability of our method in handling complex scenarios, we conducted experiments on the FoundIR test set featuring mixed degradations (blur and noise). It can be seen that our method skips ahead of other unified restoration models by a significant margin. These results demonstrate that our approach effectively handles diverse tasks without observable cross-task interference (e.g., between denoising and super-resolution), confirming its strong task independence and robust restoration capability.

4.3 Comparison with Commercial Solutions

To further validate the practical effectiveness of our method, we compare MoERestore with several widely-used commercial image restoration solutions, including Seedream 4.0 (ByteDance Seed Vision Team, 2025), Gemini-NanoBanana (DeepMind, 2025), and MeiTu SR (MeiTu, 2025). As shown in Table 2, our method achieves superior performance across all evaluation metrics on the RealLQ250 dataset. Figure 3 further shows the superiority of our method over other methods. These results demonstrate that the proposed unified restoration framework is competitive not only with research-oriented methods but also with advanced commercial image restoration systems.

Table 2: Performance compared to commercial models on the RealLQ250 dataset.

Method	CLIP-IQA+ ↑	Q-Align ↑	MUSIQ ↑	MANIQA ↑	NIMA ↑	CLIP-IQA ↑	NIQE ↓
LQ Input	0.6218	2.1693	44.1541	0.3718	3.8664	0.6079	6.0790
Seedream 4.0	0.5002	3.6931	52.3771	0.2794	4.7024	0.4124	4.9393
Gemini-NanoBanana	0.3780	3.3114	44.6310	0.2548	4.6571	0.4434	6.0865
MeiTu SR	0.6653	4.1464	66.5936	0.4498	5.2103	0.6663	5.4125
Ours	**0.7216**	**4.2715**	**68.7300**	**0.5552**	**5.2815**	**0.7144**	**4.4842**

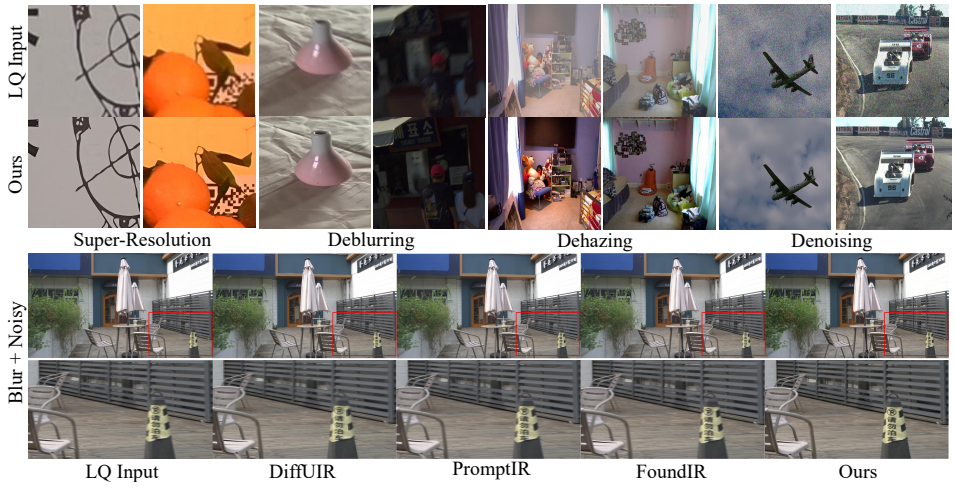

Figure 2: Visualization results of multi-task image restoration. The examples cover both single degradations and complex mixed degradations.

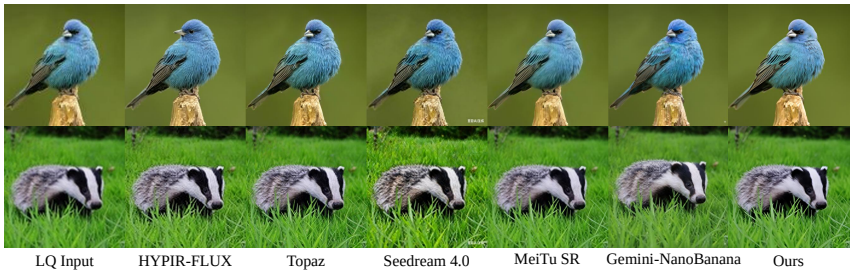

Figure 3: Qualitative comparison with commercial models on RealLQ250.

4.4 Task-aware Expert Activation Analysis

To investigate the behavior of the proposed routing mechanism, we visualize the expert activation patterns across four representative restoration tasks: super-resolution, deblurring, dehazing, and denoising. As shown in Figure 4, different tasks exhibit distinct expert activation distributions. Such task-aware activation suggests that different experts specialize in handling different restoration properties, enabling the unified model to effectively adapt to diverse degradation types while mitigating cross-task interference.

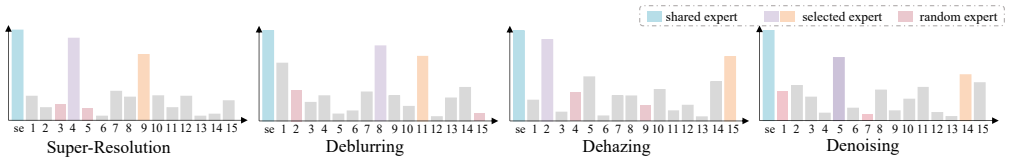

Figure 4: Visualization of task-aware expert activation across four restoration tasks.

4.5 Real-World Evaluation with VLM-based Evaluation

Existing restoration benchmarks are mostly constructed under synthetic or controlled settings, where degradation types are predefined and ground-truth references are available. In real-world scenarios, however, degradations are often unknown, mixed, and spatially varying, making full-reference evaluation inapplicable and substantially increasing the difficulty of unified restoration. To better assess real-world generalization, we collect an extremely challenging test set containing 60 real-world images. We evaluate all methods on this test set using the VLM-based perceptual assessment protocol defined in Appendix D.

Table. 3 summarizes the VLM-based evaluation results on the proposed real-world test set. MoERestore dominates across all dimensions, indicating stronger robustness under complex real-world degradations. Figure 5 further presents representative qualitative comparisons. MoERestore yields superior restoration results under unknown degradation conditions. These qualitative observations are consistent with the VLM-based scores and together demonstrate strong real-world generalization capability of the proposed unified restoration framework.

Table 3: VLM-based evaluation.

Method	ES	TF	GP	SP	TP	DP
PromptIR	6.0	6.5	7.0	6.5	6.5	7.0
TransWeather	6.5	7.0	7.0	6.5	7.0	6.5
DiffUIR	6.0	7.5	7.5	7.0	7.5	7.0
InstructIR	7.0	7.0	7.0	7.0	7.5	6.5
FoundIR	8.0	7.5	8.0	7.0	7.5	7.0
Ours	**9.0**	**9.0**	**9.0**	**9.0**	**8.5**	**9.0**

**ES: Edge Sharpness, TF: Texture Fidelity, GP: Geometric Pres., SP: Spatial Pres., TP: Text Pres., DP: Detail Pres.*

4.6 Ablation Study

As shown in Table 4, we conduct ablation experiments to evaluate the contribution of each component in MoERestore on the GoPro dataset. Removing the Sparse MoE architecture (w/o MoE) causes the most significant drop, with PSNR decreasing by 3.59 dB. This result highlights the importance of task-adaptive expert routing in alleviating cross-task interference within the unified restoration framework. Specifically, the shared experts and random exploration are crucial for universal representation and robust optimization. The SDA module's removal confirms that semantic guidance is key to resolving complex degradation ambiguities. Notably, excluding the High-Frequency branch results in the lowest SSIM (0.6425), highlighting its indispensable role in detail reconstruction. Overall, the synergy of these modules ensures superior multi-task restoration. Overall, these results demonstrate that each component contributes to the final performance of MoERestore, and their combination enables robust multi-task image restoration.

Table 4: Ablation study of our proposed MoERestore on GoPro dataset. Semantic Degradation-Aware (SDA), MoE configurations, and the High-Frequency branch.

Metrics	**Ours**	w/o SDA	w/o MoE Rand	w/o MoE Shared	w/o MoE	w/o High-Freq
SSIM	**0.8341**	0.8216	0.7881	0.7264	0.7009	0.6425
PSNR	**23.94**	23.54	23.05	22.16	20.35	20.05

5 Conclusion

In this paper, we presented MoERestore, a unified framework for multi-task image restoration that integrates sparse mixture-of-experts routing with degradation-aware conditioning. The proposed design enables a single model to adaptively handle diverse restoration tasks while alleviating cross-task interference. Extensive experiments on standard benchmarks and real-world images demonstrate that MoERestore delivers consistent performance across multiple restoration scenarios. These results highlight the effectiveness of combining adaptive expert routing with semantic degradation guidance for unified image restoration.

REFERENCES

Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 126–135, 2017.

Yuang Ai, Huaibo Huang, and Ran He. Lora-ir: Taming low-rank experts for efficient all-in-one image restoration. *arXiv preprint arXiv:2410.15385*, 2024a.

Yuang Ai, Xiaoqiang Zhou, Huaibo Huang, Xiaotian Han, Zhengyu Chen, Quanzeng You, and Hongxia Yang. Dreamclear: High-capacity real-world image restoration with privacy-safe dataset curation. *Advances in Neural Information Processing Systems*, 37:55443–55469, 2024b.

Codruta O Ancuti, Cosmin Ancuti, Mateu Sbert, and Radu Timofte. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In *2019 IEEE international conference on image processing (ICIP)*, pp. 1014–1018. IEEE, 2019.

Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE transactions on image processing*, 25(11):5187–5198, 2016.

I Chen, Wei-Ting Chen, Yu-Wei Liu, Yuan-Chun Chiang, Sy-Yen Kuo, Ming-Hsuan Yang, et al. Unirestore: Unified perceptual and task-oriented image restoration model using diffusion prior. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 17969–17979, 2025a.

Junyang Chen, Jinshan Pan, and Jiangxin Dong. Faithdiff: Unleashing diffusion priors for faithful image super-resolution. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 28188–28197, 2025b.

Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European conference on computer vision*, pp. 17–33. Springer, 2022.

Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 22367–22377, 2023.

Xiangyu Chen, Zheyuan Li, Yuandong Pu, Yihao Liu, Jiantao Zhou, Yu Qiao, and Chao Dong. A comparative study of image restoration networks for general backbone network design. In *European Conference on Computer Vision*, pp. 74–91. Springer, 2024.

Xiangyu Chen, Xintao Wang, Wenlong Zhang, Xiangtao Kong, Yu Qiao, Jiantao Zhou, and Chao Dong. Hat: Hybrid attention transformer for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025c.

Yu-Wei Chen and Soo-Chang Pei. Always clear days: Degradation type and severity aware all-in-one adverse weather removal. *IEEE Access*, 13:7650–7662, 2025.

Xiaojie Chu, Liangyu Chen, and Wenqing Yu. Nafssr: Stereo image super-resolution using nafnet. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1239–1248, 2022.

Marcos V Conde, Gregor Geigle, and Radu Timofte. Instructir: High-quality image restoration following human instructions. In *European Conference on Computer Vision*, pp. 1–21. Springer, 2024.

Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Image restoration via frequency selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(2):1093–1108, 2023.

Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16 (8):2080–2095, 2007.

Yu Dong, Yihao Liu, He Zhang, Shifeng Chen, and Yu Qiao. Fd-gan: Generative adversarial networks with fusion-discriminator for single image dehazing. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pp. 10729–10736, 2020.

Song Fei, Tian Ye, Lujia Wang, and Lei Zhu. Lucidflux: Caption-free universal image restoration via a large-scale diffusion transformer. *arXiv preprint arXiv:2509.22414*, 2025.

Yan Huang, Shang Li, Liang Wang, Tieniu Tan, et al. Unfolding the alternating optimization for blind super resolution. *Advances in Neural Information Processing Systems*, 33:5632–5643, 2020.

Yitong Jiang, Zhaoyang Zhang, Tianfan Xue, and Jinwei Gu. Autodir: Automatic all-in-one image restoration with latent diffusion. In *European Conference on Computer Vision*, pp. 340–359. Springer, 2024.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4401–4410, 2019.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 5148–5157, 2021.

Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8183–8192, 2018.

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, 2018.

Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 17452–17462, 2022.

Hao Li, Xiang Chen, Jiangxin Dong, Jinhui Tang, and Jinshan Pan. Foundir: Unleashing million-scale training data to advance foundation models for image restoration. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 12626–12636, 2025.

Yawei Li, Kai Zhang, Jingyun Liang, Jiezhang Cao, Ce Liu, Rui Gong, Yulun Zhang, Hao Tang, Yun Liu, Denis Demandolx, et al. Lsdir: A large scale dataset for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1775–1787, 2023.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1833–1844, 2021.

Jingbo Lin, Zhilu Zhang, Wenbo Li, Renjing Pei, Hang Xu, Hongzhi Zhang, and Wangmeng Zuo. Unirestorer: Universal image restoration via adaptively estimating image degradation at proper granularity. *arXiv preprint arXiv:2412.20157*, 2024a.

Xin Lin, Jingtong Yue, Kelvin CK Chan, Lu Qi, Chao Ren, Jinshan Pan, and Ming-Hsuan Yang. Multi-task image restoration guided by robust dino features. *arXiv preprint arXiv:2312.01677*, 2023.

Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Yu Qiao, Wanli Ouyang, and Chao Dong. Diffbir: Toward blind image restoration with generative diffusion prior. In *European conference on computer vision*, pp. 430–448. Springer, 2024b.

Lin Liu, Lingxi Xie, Xiaopeng Zhang, Shanxin Yuan, Xiangyu Chen, Wengang Zhou, Houqiang Li, and Qi Tian. Tape: Task-agnostic prior embedding for image restoration. In *European conference on computer vision*, pp. 447–464. Springer, 2022.

Yulin Luo, Rui Zhao, Xiaobao Wei, Jinwei Chen, Yijie Lu, Shenghao Xie, Tianyu Wang, Ruiqin Xiong, Ming Lu, and Shanghang Zhang. Wm-moe: Weather-aware multi-scale mixture-of-experts for blind adverse weather removal. *arXiv preprint arXiv:2303.13739*, 2023a.

Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Controlling vision-language models for universal image restoration. *arXiv preprint arXiv:2310.01018*, 3(8), 2023b.

Jiaqi Ma, Tianheng Cheng, Guoli Wang, Qian Zhang, Xinggang Wang, and Lefei Zhang. Prores: Exploring degradation-aware visual prompt for universal image restoration. *arXiv preprint arXiv:2306.13653*, 2023.

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 *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pp. 416–423, July 2001.

Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3883–3891, 2017.

V Potlapalli, SW Zamir, S Khan, and FS Khan. Promptir: Prompting for all-in-one blind image restoration. arxiv 2023. *arXiv preprint arXiv:2306.13090*, 6.

Jaesung Rim, Haeyun Lee, Jucheol Won, and Sunghyun Cho. Real-world blur dataset for learning and benchmarking deblurring algorithms. In *European conference on computer vision*, pp. 184–201. Springer, 2020.

Jaesung Rim, Geonung Kim, Jungeon Kim, Junyong Lee, Seungyong Lee, and Sunghyun Cho. Realistic blur synthesis for learning image deblurring. In *European conference on computer vision*, pp. 487–503. Springer, 2022.

Stefan Roth and Michael J Black. Fields of experts. *International Journal of Computer Vision*, 82 (2):205–229, 2009.

Kyle Sargent, Ruiqi Gao, Philipp Henzler, Charles Herrmann, Aleksander Holynski, Li Fei-Fei, Jiajun Wu, and Jason Zhang. Vlic: Vision-language models as perceptual judges for human-aligned image compression. *arXiv preprint arXiv:2512.15701*, 2025.

Hao Shen, Zhong-Qiu Zhao, and Wandi Zhang. Adaptive dynamic filtering network for image denoising. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 2227–2235, 2023.

Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32:1927–1941, 2023.

Zhentao Tan, Yue Wu, Qiankun Liu, Qi Chu, Le Lu, Jieping Ye, and Nenghai Yu. Exploring the application of large-scale pre-trained models on adverse weather removal. *IEEE transactions on image processing*, 33:1683–1698, 2024.

Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, Chung-Chi Tsai, and Chia-Wen Lin. Stripformer: Strip transformer for fast image deblurring. In *European conference on computer vision*, pp. 146–162. Springer, 2022.

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*, pp. 2353–2363, 2022.

Narasimhan Venkatanath, D Praneeth, S Channappayya Sumohana, S Medasani Swarup, et al. Blind image quality evaluation using perception based features. In *2015 twenty first national conference on communications (NCC)*, pp. 1–6. IEEE, 2015.

Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 2555–2563, 2023.

Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin CK Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *International Journal of Computer Vision*, 132(12):5929–5949, 2024a.

Tao Wang, Kaihao Zhang, Ziqian Shao, Wenhan Luo, Bjorn Stenger, Tong Lu, Tae-Kyun Kim, Wei Liu, and Hongdong Li. Gridformer: Residual dense transformer with grid structure for image restoration in adverse weather conditions. *International journal of computer vision*, 132(10): 4541–4563, 2024b.

Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1905–1914, 2021.

Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-step effective diffusion network for real-world image super-resolution. *Advances in Neural Information Processing Systems*, 37: 92529–92553, 2024a.

Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, and Lei Zhang. Seesr: Towards semantics-aware real-world image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 25456–25467, 2024b.

Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. Diffir: Efficient diffusion model for image restoration. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 13095–13105, 2023.

Jun Xu, Hui Li, Zhetong Liang, David Zhang, and Lei Zhang. Real-world noisy image denoising: A new benchmark. *arXiv preprint arXiv:1804.02603*, 2018.

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*, pp. 24902–24912, 2024a.

Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. Maniqa: Multi-dimension attention network for no-reference image quality assessment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1191–1200, 2022.

Tao Yang, Rongyuan Wu, Peiran Ren, Xuansong Xie, and Lei Zhang. Pixel-aware stable diffusion for realistic image super-resolution and personalized stylization. In *European conference on computer vision*, pp. 74–91. Springer, 2024b.

Mingde Yao, Ruikang Xu, Yuanshen Guan, Jie Huang, and Zhiwei Xiong. Neural degradation representation learning for all-in-one image restoration. *IEEE transactions on image processing*, 33:5408–5423, 2024.

Zhenqiang Ying, Haoran Niu, Praful Gupta, Dhruv Mahajan, Deepti Ghadiyaram, and Alan Bovik. From patches to pictures (paq-2-piq): Mapping the perceptual space of picture quality. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3575–3585, 2020.

Fanghua Yu, Jinjin Gu, Zheyuan Li, Jinfan Hu, Xiangtao Kong, Xintao Wang, Jingwen He, Yu Qiao, and Chao Dong. Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 25669–25680, 2024.

Eduard Zamfir, Zongwei Wu, Nancy Mehta, Yuedong Tan, Danda Pani Paudel, Yulun Zhang, and Radu Timofte. Complexity experts are task-discriminative learners for any image restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 12753–12763, 2025.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14821–14831, 2021.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5728–5739, 2022a.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for fast image restoration and enhancement. *IEEE transactions on pattern analysis and machine intelligence*, 45(2):1934–1948, 2022b.

Jinghao Zhang, Jie Huang, Mingde Yao, Zizheng Yang, Hu Yu, Man Zhou, and Feng Zhao. Ingredient-oriented multi-degradation learning for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5825–5835, 2023.

Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26 (7):3142–3155, 2017.

Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018a.

Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4791–4800, 2021.

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, 2018b.

Rongyu Zhang, Yulin Luo, Jiaming Liu, Huanrui Yang, Zhen Dong, Denis Gudovskiy, Tomoyuki Okuno, Yohei Nakata, Kurt Keutzer, Yuan Du, et al. Efficient deweahter mixture-of-experts with uncertainty-aware feature-wise linear modulation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 16812–16820, 2024.

Dian Zheng, Xiao-Ming Wu, Shuzhou Yang, Jian Zhang, Jian-Fang Hu, and Wei-Shi Zheng. Selective hourglass mapping for universal image restoration based on diffusion model. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 25445–25455, 2024.

Zihua Zheng, Ni Nie, Zhi Ling, Pengfei Xiong, Jiangyu Liu, Hao Wang, and Jiankun Li. Dip: Deep inverse patchmatch for high-resolution optical flow. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8925–8934, 2022.