Complexity Experts are Task-Discriminative Learners for Any Image Restoration

Eduard Zamfir[1] Zongwei Wu[1] Nancy Mehta[1] Yuedong Tan[1]
Danda Pani Paudel[2] Yulun Zhang[3] Radu Timofte[1]

[1]University of Würzburg [2]INSAIT Sofia University [3]Shanghai Jiao Tong University

Abstract

Recent advancements in all-in-one image restoration models have revolutionized the ability to address diverse degradations through a unified framework. However, parameters tied to specific tasks often remain inactive for other tasks, making mixture-of-experts (MoE) architectures a natural extension. Despite this, MoEs often show inconsistent behavior, with some experts unexpectedly generalizing across tasks while others struggle within their intended scope. This hinders leveraging MoEs' computational benefits by bypassing irrelevant experts during inference. We attribute this undesired behavior to the uniform and rigid architecture of traditional MoEs. To address this, we introduce "complexity experts" – flexible expert blocks with varying computational complexity and receptive fields. A key challenge is assigning tasks to each expert, as degradation complexity is unknown in advance. Thus, we execute tasks with a simple bias toward lower complexity. To our surprise, this preference effectively drives task-specific allocation, assigning tasks to experts with the appropriate complexity. Extensive experiments validate our approach, demonstrating the ability to bypass irrelevant experts during inference while maintaining superior performance. The proposed MoCE-IR model outperforms state-of-the-art methods, affirming its efficiency and practical applicability. The source will be publicly made available at eduardzamfir.github.io/MoCE-IR/

1. Introduction

Image restoration [47, 66, 67] is a fundamental problem in computer vision, dealing with reconstructing high-quality images from deteriorated observations. Adverse conditions such as noise [22, 66], haze [5, 38], or rain [7, 58] significantly impact the practical utility of images in downstream applications across various domains, including autonomous navigation [7, 49] or augmented reality [10, 15, 45]. Deep learning-based approaches have remarkably advanced this field, particularly for task-specific image restoration prob-
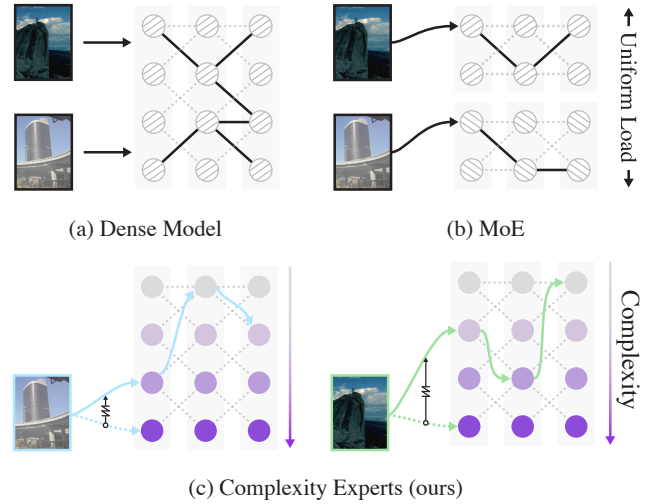

Figure 1. *Motivation.* (a) Dense all-in-one restoration methods [24, 57] often inefficiently allocate parameters when handling multiple degradation types. (b) While recent Mixture-of-Experts (MoE) approaches [60, 61] address this through sparse computation, their rigid routing mechanisms uniformly distribute inputs across experts without considering the natural relationships between degradations. (c) To overcome these limitations, we introduce Complexity Experts - adaptive processing blocks with size-varying computational units. Our framework dynamically allocates model capacity using a spring-inspired force mechanism that continuously guides routing decisions toward simpler experts when possible, with the force proportional to the complexity of the input degradation. While initially designed for computational efficiency, this approach naturally emerges as a task-discriminative learning framework, assigning degradations to the most suitable experts. This makes it particularly effective for all-in-one restoration methods, where both task-specific processing and cross-degradation knowledge sharing are crucial.

lems [6, 8, 16, 28, 52, 64]. Recent all-in-onerestoration models [9, 12, 13, 24, 27, 36, 49, 65], however, demonstrate the possibility of handling multiple degradation types within a single model, offering more practical solutions compared to traditional task-specific approaches without the extensive re-training need for each new degradation type. Noteable

works employ visual [25, 36, 50] or language-based prompting [1, 9, 30], contrastive learning [24, 65], and diffusion-based models [1, 30].

Despite their success, we observe that the aforementioned models often suffer from inefficiencies, as parameters tied to specific degradations often remain inactive or underutilized when addressing unrelated tasks [57]. This naturally suggests the use of mixture-of-experts (MoE) architectures for task-specific processing. Yet, current MoE-based approaches [29, 57, 61] typically incorporate routing mechanisms based on language [57] or degradation priors [29, 61], leading to imbalanced optimization, where some experts generalize well, while others struggle with their intended tasks. This inconsistency limits the potential computational benefits of bypassing task-irrelevant experts during inference.

We attribute this undesired behavior to two limitations in current approaches. First, the uniform and rigid architecture of existing MoE models fails to account for the varying complexity requirements across different restoration tasks. For instance, motion blur demands localized processing with strong spatial awareness, while haze removal requires broader contextual understanding [4, 35], necessitating adaptive processing aligned to the task requirements. Second, the challenge of appropriately routing tasks to experts is complicated by the unknown complexity of each degradation type a priori. Typically, MoE models [37, 39, 43, 69] aim to balance expert utilization, preventing the model from collapsing into a single-expert dependency.

In this work, we introduce a novel mixture-of-complexity-experts (MoCE) framework for all-in-one image restoration that directly addresses these limitations. Our key innovation lies in designing expert blocks with increasing computational complexity and receptive fields, allowing the model to adaptively match processing capacity with task requirements.

To address the routing challenge, we introduce a complexity-aware allocation mechanism that preferentially directs tasks to lower-complexity experts. This approach implements a complexity-proportional bias term, analogous to a mechanical spring force, to guide expert selection. Surprisingly, this straightforward approach results in effective task-specific allocation, naturally directing inputs to experts with appropriate complexity levels. Our MoCE framework offers two significant advantages: it ensures efficient inference by selectively bypassing irrelevant experts, thereby reducing computational overhead, while simultaneously maintaining high-quality restoration performance across diverse tasks. Through extensive experiments across multiple image restoration tasks, we demonstrate that our approach not only preserves restoration quality but surpasses recent state-of-the-art methods. These results validate both the effectiveness of our complexity-based expert design and the practicality of our routing strategy for real-world applications. Our approach establishes a new benchmark for all-in-one image

restoration, enhancing both efficiency and fidelity.

The principal contributions of this work are threefold:
- We propose MoCE-IR, achieving SoTA all-in-one image restoration performance with improved efficiency compared to prior works.
- Our novel MoE layer selectively activates complexity experts based on input requirements, unifying task-specific and holistic learning in a single architecture.
- We develop a complexity-aware routing mechanism that balances restoration quality with computational efficiency by adaptive expert allocation.

2. Related Works

Task Specific Image Restoration. Reconstructing the clean image from its degraded counterpart is a highly ill-posed problem, however, a great body of work have addressed image restoration from a data-driven learning perspective, achieving tremendous results compared to prior hand-crafted methods [22, 26, 47, 52, 64, 66, 67]. Most proposed solutions build on convolutional [6, 47, 66, 67] or Transformer-based architectures [8, 28, 52, 64] addressing single degradation tasks, such as denoising [8, 66, 67], dehazing [41, 42, 55] or deraining [18, 40]. Contrary to CNN-based networks, Transformer offer strong modeling capabilities for capturing global dependencies, which makes them outstanding image restorers [28, 64, 68]. Self-attention's quadratic complexity w.r.t the image size poses a challenge for resource-constrained applications. Conversely, convolutions offer fast and efficient processing with limited global context, but they scale more effectively with increasing input size. Recently, FFTformer [21] introduced an efficient approximation of query and key interactions in the Fourier domain, significantly reducing computational demands compared to window-based attention methods for image deblurring. Building on this, we incorporate specialized experts into a Transformer-based architecture [64] to efficiently capture both shared and distinct contextual information, addressing the diverse demands of image restoration tasks.

All-in-One Image Restoration. Restoring degraded images often requires multiple models, complicating practical implementation, especially when images exhibit a combination of various degradations. An emerging field known as all-in-one image restoration is advancing in low-level computer vision, utilizing a single deep blind restoration model to tackle multiple degradation types simultaneously [7, 19, 36, 49, 63, 65]. The seminal work, AirNet [24] achieves blind all-in-one image restoration using contrastive learning to extract degradation representations, which guide the restoration process. Similarly, IDR [65] employs a meta-learning-based two-stage approach to model degradation through underlying physical principles. Prompt-based learning [25, 36, 50] has also gained traction, with [36] introducing tunable prompts to encode degradation-specific in-
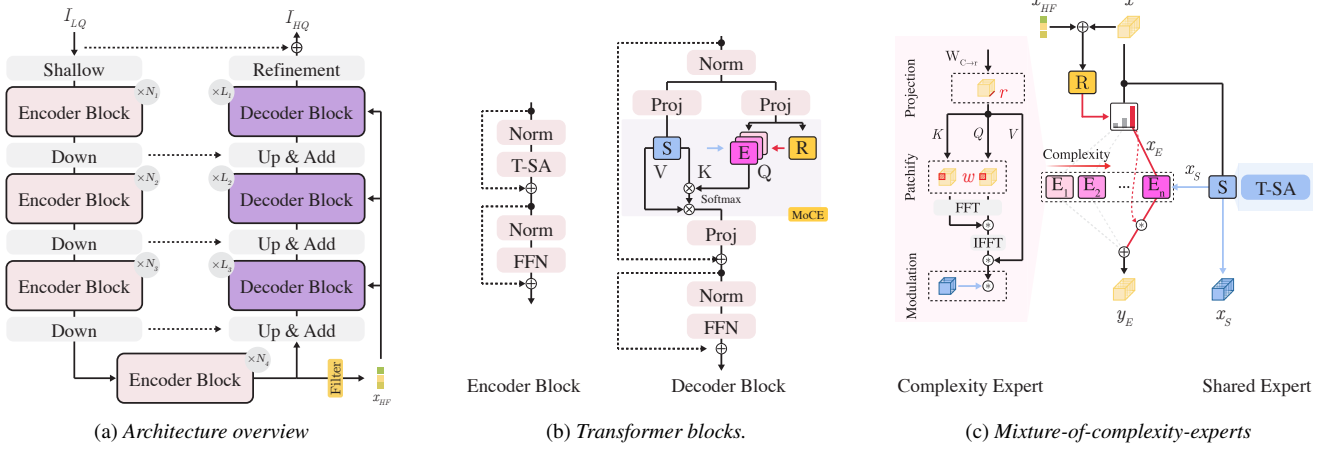
Figure 2. *Proposed MoCE-IR framework.* Despite recent advances in MoE-based image restoration [57, 61, 62], inconsistent expert behavior—where some experts over-generalize while others underperform—limits their computational efficiency. We address this through *complexity experts*: flexible blocks with varying computational capacity and receptive fields. Our MoCE-IR employs an asymmetric encoder-decoder architecture where each decoder block contains a mixture-of-complexity-experts layer for adaptive capacity routing.

formation, albeit with a high parameter count. With recent advancements in multimodal integration, such as combining images and language, there has been growing interest in language-guided image restoration. Notable approaches include DA-CLIP [30], InstructIR [9], and UniProcessor [12]. However, these methods often depend on high-quality text prompts and typically require large language models, limiting their applicability on resource-constrained devices. In contrast, our contribution lies in a novel, parameter-efficient approach that focuses on selective activation, effectively addressing diverse degradations while remaining lightweight and adaptable.

Dynamic Networks. Dynamic networks have evolved from basic conditional computation [3] to sophisticated Mixture-of-Experts (MoE) architectures that expand model capacity while maintaining efficient inference costs [37, 43, 46]. This approach has proven particularly effective in transformer-based NLP and high-level vision tasks. In image restoration, Path-Restore [59] introduced content-aware patch routing with difficulty-regulated rewards, while recent all-in-one models [29, 57, 61] leverage various priors for expert routing. Our work addresses the challenge of adaptively routing corrupted images to appropriate computational units within a generalist restoration framework. By dynamically aligning image characteristics with computational requirements, we achieve a unified model in both parameter and architectural space, advancing beyond previous limitations.

3. Method

In this section, we outline the core principles of our All-in-One image restoration method. As shown in Fig. 2a, our pipeline features a U-shaped architecture [44, 64] with an asymmetric encoder-decoder design. A 3×3 convolution

first extracts shallow features from the degraded input, which then pass through 4 levels of encoding and decoding stages. While both stages utilize Transformer blocks [36, 64], our novel MoCE layers are exclusively integrated into the decoder blocks (Fig. 2b). To further enhance the decoder's feature enrichment capabilities, we introduce high-frequency guidance via Sobel-filtered global feature vector to promote the frequency awareness of the gating functions. Finally, a global residual connection links the shallow features to the output of the refinement stage, refining the crucial high-frequency details before producing the restored image.

3.1. Mixture-of-Complexity-Experts

To address the limitations of uniform MoE architectures in image restoration, we design an efficient input-adaptive model for All-in-One image restoration, dynamically adjusting the necessary processing capacity to the current task requirements. Technically, at the core of our framework is a specialized MoCE layer comprising n complexity experts $\mathbf{E}$ with progressively increasing computational complexity and receptive fields, departing from the uniform architecture of traditional MoE designs [37, 43, 46] shown in Fig. 2b. This is complemented by a shared expert $\mathbf{S}$ that learns task-invariant features. The interaction between these components is orchestrated through a two-level gating mechanism: first projecting routed tokens into expert-specific embeddings, then combining specialized and agnostic features through element-wise multiplication, and finally merging the outputs via cross-attention. This design enables our model to capture both degradation-specific features and inter-degradation relationships, while maintaining a natural bias toward computationally efficient processing paths.

Expert Design. Each complexity expert block as shown

in Fig. 2c is designed based on two fundamental principles: ensuring computational efficiency and capturing hierarchical spatial features, which are critical for restoring fine details and addressing diverse degradation patterns. Given the growing computational demands associated with increasing expert numbers and capacities, it is crucial to prioritize an overall efficient expert design. To this end, we implement a nested expert structure, progressively reducing channel dimensionality r within each subsequent expert to control the computational overhead. Simultaneously, we increase the receptive field, *i.e.* window partition size w, within each expert to adaptively balance the localized and global processing, tailored to the specific requirements of the input degradation. Thus, our complexity experts are constructed by scaling the design along two key axes: window partition size and channel dimension.

More specifically, each expert $\mathbf{E}$ projects its input tokens to the embedding $\mathbf{x}_E \in \mathbb{R}^{H \times W \times r}$, where $r = C/2^i$ for $i \in \{i, ..., n\}$, to highlight the most relevant information along the channel dimension. Following this, a window-based self-attention (WSA) mechanism captures spatial information effectively. To optimize performance, we employ an FFT-based approximation [21] for efficient matrix multiplication between queries Q and keys K in the Fourier domain. The refined features are then transformed back into the spatial domain and re-projected to their original dimensionality. Lastly, the output of the expert block $\hat{\mathbf{x}}_E^i$ is modulated through element-wise multiplication with the holistic features from the shared expert, $\mathbf{S}$. As shared expert, we employ the transposed self-attention (T-SA) module [64] in the channel dimension, while the complexity experts operate spatially. More concretely, the modulated expert features $\mathbf{y_E^i}$ are obtained as following:

$$\mathbf{y_E^i} = \hat{\mathbf{x}}_E^i \odot \mathbf{S}(\mathbf{x}) \tag{1}$$
$$\text{with } \hat{\mathbf{x}}_E^i = \mathbf{E}^i(\mathbf{x}_E^i) = \text{FFT-WSA}_{\mathbf{w}_i}(\mathbf{W}_{C \to r_i}^i \mathbf{x}_E^i) \tag{2}$$
$$\text{and } \mathbf{S}(\mathbf{x}) = \text{T-SA}(\mathbf{x}) \tag{3}$$

where linear layers for projection are denoted as $\mathbf{W}$, $\odot$ denotes element-wise multiplication and i is the index of the current complexity expert $\mathbf{E}^i$ with window size w_i and embedding dimensionality r_i. For example, the most lightweight expert is characterized by the smallest embedding dimension r_1 and window size w_1, while the dimensions and window partitions for the other experts can increase linearly. Throughout our network, the capacity of the shared expert remains constant, with $r = C$.

3.2. Complexity-aware Routing

Drawing inspiration from sparse MoE [43, 46], we integrate linear layers within each decoder block to enable a routing mechanism, associating input features $\mathbf{x} \in \mathbb{R}^{H \times W \times C}$ with their corresponding specialized complexity experts $\mathbf{E}$. However, in the context of image restoration, a key challenge arises in achieving a scale-invariant tokenization of input images, ensuring consistency across varying resolutions and scales. While previous MoE approaches [37, 43, 46] employ a token-based routing, we opt for an image-level routing strategy instead, where we select experts for entire input image. The routing function $g(\mathbf{x})$ orchestrates the allocation of input samples based on the required compute to corresponding complexity experts $\mathbf{E}^i$, indexed by $i \in 1, ..., n$ with n representing the total number of experts. Within $g(\mathbf{x})$, we select the top-k elements of the softmax distribution associating input tokens to experts, setting all other elements to zero. In practice, we use $k = 1$ while ϵ is sampled independently $\epsilon \sim \mathcal{N}(0, 1/n^2)$ entry-wise enabling noisy top-1 routing:

$$g(\mathbf{x}) = \text{top}_k(\text{Softmax}(\mathbf{W}\mathbf{x} + \epsilon)) \tag{4}$$

During training, we incorporate an ==auxiliary loss== $\mathcal{L}_{\text{aux}}$ inspired by the load balancing loss proposed by Riquelme *et al.* [43]. We complement the importance term to favor experts based on their computational complexity, allowing for more effective balancing aligned with expert capacity. As in [43], the balanced expert utilization is encouraged through an importance loss, where the importance of each expert $\mathbf{E}^i$ for a batch of images is defined as the sum of the routing weights associated with the i-th expert across the entire batch. Additionally, we introduce a complexity bias $\mathbf{b}$ by calculating the number of learnable parameters p_i in each expert $\mathbf{E}^i$ and normalizing these values with respect to the largest parameter count of the experts, $p_{\max}$:

$$\text{Imp}_i(\mathbf{x}) = \left(\sum_{x \in \mathbf{x}} \text{Softmax}(Wx)_i \right) * \mathbf{b}, \tag{5}$$
$$\text{with } \mathbf{b} = [p_1/p_{\max}, p_2/p_{\max}, \ldots, p_n/p_{\max}], \tag{6}$$

where W is the layer-specific weight for the router $g(\mathbf{x})$. This approach assigns progressively lower weights to experts with lower complexity. For intuitive understanding, this mechanism parallels a mechanical spring system, where the restoring force is proportional to both displacement and the spring constant. In our context, the expert's parameter count corresponds to displacement, while the normalization factor serves as the material-specific constant, collectively determining the magnitude of complexity bias. Without this, the routers lack a basis for selecting specific experts, leading to randomized routing and thus suboptimal model capacity utilization. Lastly, the complexity-aware importance Imp^* is used to compute the total auxiliary loss $\mathcal{L}_{\text{aux}}$ defined as:

$$\mathcal{L}_{\text{aux}}(\mathbf{x}) = 1/2 \, \text{CV}(\text{Imp}^*(\mathbf{x}))^2 + 1/2 \, \text{CV}(\text{Load}(\mathbf{x}))^2, \tag{7}$$

where CV denotes the coefficient of variation [43]. The impact of the auxiliary loss on the routing is illustrated in Fig. 5 and Tab. 5a.

Table 1. *Comparison to state-of-the-art on three degradations.* PSNR (dB, ↑) and SSIM (↑) are reported on the full RGB images. Our method sets a new state-of-the-art on average across all benchmarks on two model scales, showcasing its scalability while being more efficient than prior work. '-' represents unreported results and '∗' indicates VLM-based approaches. The **best** performances are highlighted.

| | Method | Params. | Dehazing | | Deraining | | Denoising | | | | | | Average | |
			SOTS		Rain100L		BSD68$_{\sigma=15}$		BSD68$_{\sigma=25}$		BSD68$_{\sigma=50}$			
Light	BRDNet [48]	-	23.23	.895	27.42	.895	32.26	.898	29.76	.836	26.34	.693	27.80	.843
	LPNet [14]	-	20.84	828	24.88	.784	26.47	.778	24.77	.748	21.26	.552	23.64	.738
	FDGAN [11]	-	24.71	.929	29.89	.933	30.25	.910	28.81	.868	26.43	.776	28.02	.883
	DL [13]	2M	26.92	.931	32.62	.931	33.05	.914	30.41	.861	26.90	.740	29.98	.876
	AirNet [24]	9M	27.94	.962	34.90	.967	33.92	.933	31.26	.888	28.00	.797	31.20	.910
	MoCE-IR-S (*ours*)	11M	**30.94**	**.979**	**38.22**	**.983**	**34.08**	**.933**	**31.42**	**.888**	**28.16**	**.798**	**32.57**	**.916**
Heavy	MPRNet [63]	16M	25.28	.955	33.57	.954	33.54	.927	30.89	.880	27.56	.779	30.17	.899
	PromptIR [36]	36M	30.58	.974	36.37	.972	33.98	.933	31.31	.888	28.06	.799	32.06	.913
	Gridformer [51]	34M	30.37	.970	37.15	.972	33.93	.931	31.37	.887	28.11	.801	32.19	.912
	Art-PromptIR [54]	33M	30.83	.979	37.94	.982	34.06	.934	31.42	.891	28.14	.801	32.49	.917
	DA-CLIP* [30]	125M	29.46	.963	36.28	.968	30.02	.821	24.86	.585	22.29	.476	-	-
	UniProcessor* [12]	1002M	**31.66**	.979	38.17	.982	34.08	**.935**	31.42	**.891**	28.17	**.803**	32.70	**.918**
	MoCE-IR (*ours*)	25M	31.12	**.979**	**38.72**	**.984**	**34.13**	.932	**31.47**	.888	**28.20**	.800	**32.73**	.917

Table 2. *Comparison to state-of-the-art on five degradations.* PSNR (dB, ↑) and SSIM (↑) are reported on the full RGB images with (∗) denoting general image restorers, others are specialized all-in-one approaches. Under more challenging degradations, our MoCE-IR framework outperforms prior work across two model scales. The **best** performances are highlighted.

| | Method | Params. | Dehazing | | Deraining | | Denoising | | Deblurring | | Low-Light | | Average | |
			SOTS		Rain100L		BSD68$_{\sigma=25}$		GoPro		LOLv1			
Light	SwinIR* [26]	1M	21.50	.891	30.78	.923	30.59	.868	24.52	.773	17.81	.723	25.04	.835
	DL [13]	2M	20.54	.826	21.96	.762	23.09	.745	19.86	.672	19.83	.712	21.05	.743
	TAPE [27]	1M	22.16	.861	29.67	.904	30.18	.855	24.47	.763	18.97	.621	25.09	.801
	AirNet [24]	9M	21.04	.884	32.98	.951	30.91	.882	24.35	.781	18.18	.735	25.49	.847
	MoCE-IR-S (*ours*)	11M	**30.33**	**.974**	**37.21**	**.978**	**31.25**	**.884**	**28.90**	**.877**	**21.68**	**.851**	**29.88**	**.913**
Heavy	NAFNet* [6]	17M	25.23	.939	35.56	.967	31.02	.883	26.53	.808	20.49	.809	27.76	.881
	DGUNet* [33]	17M	24.78	.940	36.62	.971	31.10	.883	27.25	.837	21.87	.823	28.32	.891
	Restormer* [64]	26M	24.09	.927	34.81	.962	31.49	.884	27.22	.829	20.41	.806	27.60	.881
	MambaIR [16]	27M	25.81	.944	36.55	.971	31.41	.884	28.61	.875	22.49	.832	28.97	.901
	Transweather [49]	38M	21.32	.885	29.43	.905	29.00	.841	25.12	.757	21.21	.792	25.22	.836
	IDR [65]	15M	25.24	.943	35.63	.965	**31.60**	.887	27.87	.846	21.34	.826	28.34	.893
	Gridformer [51]	34M	26.79	.951	36.61	.971	31.45	.885	29.22	.884	22.59	.831	29.33	.904
	InstructIR-5D [9]	17M	27.10	.956	36.84	.973	31.40	.873	29.40	.886	23.00	.836	29.55	.908
	MoCE-IR (*ours*)	25M	**30.48**	**.974**	**38.04**	**.982**	31.34	**.887**	**30.05**	**.899**	**23.00**	**.852**	**30.58**	**.919**

4. Experiments

We conduct experiments by strictly following previous works in general image restoration [36, 65] under two different settings: *(i) All-in-One* and *(ii) Composited degradations*. In the all-in-one setting, a unified model is trained across multiple degradation types, where we consider *three* and *five* distinct degradations. In the Composited degradations setting, a single model is trained to handle both isolated degradation occurrences as well as combinations of up to three different types of degradations.

Implementation Details. Our MoCE-IR framework is end-to-end trainable requiring no multi-stage optimization of any components. The architecture features an asymmetrical, four-level encoder-decoder design. The encoder consists of four levels with varying numbers of transformer blocks, arranged as $[4, 6, 6, 8]$ from the top level to the lowest. The decoder includes three levels, containing $[2, 4, 4]$ transformer blocks,

respectively. In each MoCE layer, we use $n = 4$ nested experts, where the dimensionality of the expert embedding is given by $R = C/2^i$ for $i \in \{i, ..., n\}$. The network's initial embedding dimension is set to $C = 32$, doubling at each subsequent level while the spatial resolution is halved. We follow the training configuration of prior work [36], and train our models for 120 epochs with a batch size of 32. We optimize the L_1 loss in the RGB and Fourier domain using the Adam [20] optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) with an initial learning rate of 2×10^{-4} and cosine decay. During training, we utilize crops of size 128^2 with horizontal and vertical flips as augmentations.

Datasets. For all-in-one, we follow existing work [24, 36] and include following datasets: For image denoising, we combine the BSD400 [2] and WED [31] datasets, adding Gaussian noise at levels $\sigma \in [15, 25, 50]$ to create noisy images. Testing is conducted on the BSD68 [32]. For deraining, we use Rain100L [56]. The dehazing task utilises the SOTS

Table 3. *Comparison to state-of-the-art on composited degradations.* PSNR (dB, ↑) and SSIM (↑) are reported on the full RGB images with (*) denoting general image restorers, others are specialized all-in-one approaches. Our MoCE-IR method consistently outperforms even larger models, with favorable results in composited degradation scenarios.

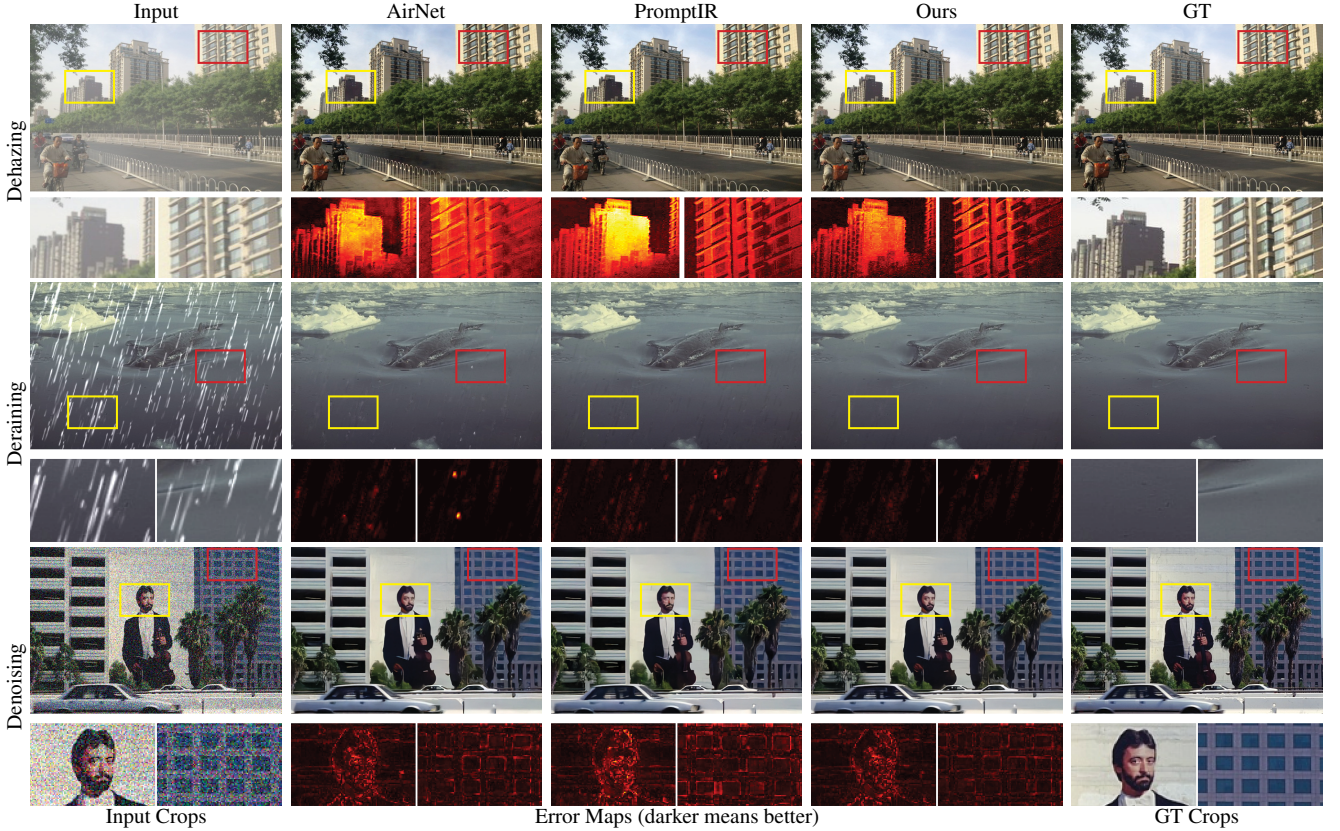| Method	Params.	CDD11-Single				CDD11-Double					CDD11-Triple		Avg.
		Low (L)	Haze (H)	Rain (R)	Snow (S)	L+H	L+R	L+S	H+R	H+S	L+H+R	L+H+S	
AirNet [24]	9M	24.83 .778	24.21 .951	26.55 .891	26.79 .919	23.23 .779	22.82 .710	23.29 .723	22.21 .868	23.29 .901	21.80 .708	22.24 .725	23.75 .814
PromptIR [36]	36M	26.32 .805	26.10 .969	31.56 .946	31.53 .960	24.49 .789	25.05 .771	24.51 .761	24.54 .924	23.70 .925	23.74 .752	23.33 .747	25.90 .850
WGWSNet [70]	26M	24.39 .774	27.90 .982	33.15 .964	34.43 .973	24.27 .800	25.06 .772	24.60 .765	27.23 .955	27.65 .960	23.90 .772	23.97 .771	26.96 .863
WeatherDiff [71]	83M	23.58 .763	21.99 .904	24.85 .885	24.80 .888	21.83 .756	22.69 .730	22.12 .707	21.25 .868	21.99 .868	21.23 .716	21.04 .698	22.49 .799
OneRestore [17]	6M	26.48 **.826**	32.52 **.990**	33.40 .964	34.31 .973	25.79 **.822**	25.58 .799	25.19 .789	**29.99** .957	**30.21** .964	24.78 **.788**	24.90 .791	28.47 .878
MoCE-IR-S (*ours*)	11M	**27.30** .822	**32.77** .989	**34.35** .969	**35.58** .978	**26.29** .821	**26.74** **.808**	**25.90** .789	29.88 **.964**	30.26 **.968**	**25.42** .785	**24.98** .791	**29.11** **.880**

Figure 3. Comparing MoCE-IR-S with AirNet [24] and PromptIR [36] for the all-in-one setting with three degradations. Our method effectively eliminates both haze and rain streaks while preserving image sharpness, delivering pristine restoration quality. We include the error heatmap with color transition from black to white denotes increasing pixel-wise erroneous.

[23] dataset. For deblurring and low-light enhancement, we employ the GoPro [34] and the LOL-v1 [53] dataset, respectively. To develop a unified model for all tasks, we merge these datasets in a *three* (AIO-3) or *five* (AIO-5) degradation setting, and train for 120 epochs and directly evaluate them across different tasks. For the *composite* degradation setting, we use the CDD11 dataset [17] and train our method for 200 epochs, keeping all other settings unchanged.

4.1. Comparison to State-of-the-Art Methods

All-in-One: Three Degradations. We compare our all-in-one restoration approach with specialized restoration methods, including both image-only models, such as MPR-

Net [63], AirNet [24], PromptIR [36], and Gridformer [51], and vision-language models like DA-CLIP [30], Instruc-tIR [9], and UniProcessor [12]. All methods were trained simultaneously on three types of degradation: haze, rain, and Gaussian noise. The results in Tab. 1 show that our proposed model, MoCE-IR, consistently outperforms previous works, both in light and heavy model scales. In the light category, MoCE-IR achieves the highest scores across all benchmarks with an average improvement in PSNR of 1.37 dB, surpassing other models like AirNet [24]. In the heavy scale, MoCE-IR competes closely with UniProcessor [12], achieving an average PSNR of 32.73 dB with 98% fewer parameters, validating our efficiency and effectiveness.

All-in-One: Five Degradations. Building on recent studies [24, 65], we expand the three-degradation setting to include deblurring and low-light image enhancement, further validating our method's effectiveness. As shown in Tab. 2, our approach excels by learning specialized experts with input-adaptive complexity and maximizing synergies across degradations. Our framework outperforms previous methods on the lightest model scale, with an average PSNR improvement of 4.31 dB over AirNet [24]. Even compared to larger networks, our lightest model surpasses the VLM-based InstructIR [9] by 0.33 dB on average, and on a comparable model scale, our framework extends this advantage by an additional 0.50 dB.

All-in-One: Composited Degradations. To simulate more realistic degradation scenarios, recent work [17] has trained all-in-one restorers not only on a mix of different degradations but also on more challenging cases where multiple degradation types are combined within a single image. We extend the previous AIO settings by including rain, haze, snow, and low illumination as individual degradation types, along with various composite degradation scenarios, resulting in a total of eleven distinct restoration settings. As shown in Tab. 3, our method outperforms previous all-in-one restorers by an average improvement of 0.64 dB over OneRestore [17], while being among the smallest models in size. This further validates our effectiveness with adaptive modeling capacity based on input complexity.

Visual Results. We show visual results obtained under the three degradation settings in Fig. 3. In certain demanding hazy scenarios, both AirNet [24] and PromptIR [36] reveal constraints in completely eradicating haziness, resulting in noticeable color intensity discrepancies, whereas our approach ensures precise color reconstruction. Furthermore, in challenging rainy scenes, these popular approaches continue to exhibit notable remnants of rain, which our method adeptly eliminates, distinguishing itself from other approaches. Additionally, our method also generates clear and sharp denoised outputs. We further provide the error maps for visual differences where our method contains less noticeable errors. Coupled with the quantitative comparisons, these findings underscore the effectiveness of our method. *Please refer to the supplemental for more visuals.*

4.2. Ablation Studies

We conduct detailed studies on the components within our method, highlighting the emergence of task-discriminative learners. All experiments are conducted in the all-in-one setting with three degradations. *Please refer to the supplemental for additional architectural ablations.*

Complexity-Efficiency Trade-off. Our MoCE-IR framework addresses the critical challenge of scalability in vision tasks by optimizing both reconstruction quality and compu-
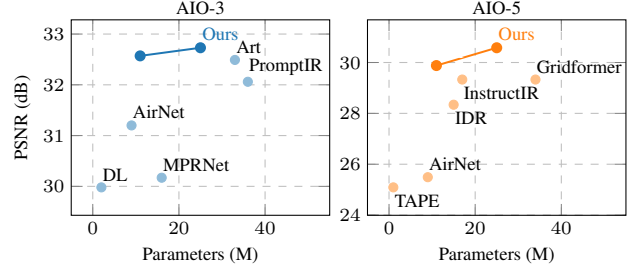

Figure 4. *Complexity-efficiency tradeoff.* Visualization of PSNR and parameter counts of proposed method compared to prior work. Proposed MoCE-IR surpasses prior methods, achieving SoTA results in all-in-one image restoration with enhanced efficiency.

Table 4. *Computational demands.* GFLOPS are computed on an input image of size 224×224 using a NVIDIA RTX 4090 GPU.

Method	Params.	Memory	FLOPS
AirNet [24]	**8.93M**	4829M	238G
PromptIR [36]	35.59M	9830M	132G
IDR [65]	15.34M	4905M	98G
MoCE-IR (*ours*)	25.35M	5887M	80.59G
MoCE-IR (*ours*)	11.47M	**4228M**	**36.93G**

tational efficiency. We visualize in Fig. 4 the progressive improvements in reconstruction fidelity of our framework on two complexity scales. Furthermore, Tab. 4 validates that our approach significantly outperforms SoTA methods in terms of memory utilization. Different from prior MoE architectures that maintain fixed computational demands during inference, our complexity experts enable dynamic FLOPS allocation. For example, our lightweight model operates at an average of 36.9 FLOPS, and the heavyweight operates at an average 80.59 FLOPS. This adaptive computation, combined with lower memory consumption and parameter counts, gives MoCE-IR a substantial advantage over contemporary methods.

Expert Scaling. As shown in Tab. 5a, we systematically evaluate different scaling strategies regarding the embedding dimensionality r of each complexity expert. Our investigation compares exponential and nested parameter scaling, where the window partition is always fixed and follows a $2^{(2+i)}$ progression. Further, we investigate the bias normalization, whether normalizing to the lowest or highest expert capacity. Overall, the nested parameter scaling emerges as the optimal choice, providing sufficient expert channel diversity while maintaining reasonable computational costs. Normalizing router weights by the highest expert capacity, rather than the minimum, proves crucial for performance as minimum capacity normalization forces router weights of higher capacity experts to escalate uncontrollably, thus excluding them from the gating mechanism.

Impact of Complexity-aware Routing. Based on the findings from Tab. 5a, we opt for nested expert scaling. Further, we evaluate our complexity-aware routing strategy against
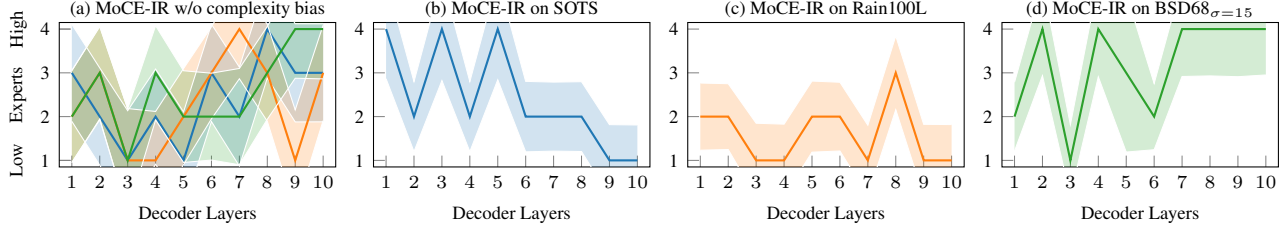
Figure 5. *Routing visualization for the AIO-3 setting.* (a) While standard load balancing [43] achieves uniform utilization across experts, it fails to exploit shared dependencies between tasks and overlooks task-specific characteristics, resulting in reduced restoration quality. (b)-(d) Complexity-aware routing promotes task-discriminative learning by assigning more complex degradations to experts with stronger contextual understanding capabilities and vice versa. This approach allows some experts to generalize across multiple degradation types while others specialize in specific tasks, enhancing the system's adaptivity to diverse degradations. We visualize the average decisions made by each router for dehazing, deraining and denoising. The y-axis indicates increasing expert complexity.

Table 5. *Architecture analysis.* (a) We investigate scaling the expert embedding r and the bias normalization for MoCE-IR-S. (b) We compare our routing against [43]. Additionally, we provide performance results for AirNet [24] as reference. Our proposed loss not only enhances overall performance in the AIO-3 setting, but also improves zero-shot generalization on unseen degradations. PSNR (dB, ↑) and SSIM (↑) are reported on full RGB images.

(a) *Ablation on the expert scaling.*

Embedding r scaling	Bias normalization	Avg. AIO-3	
Nested	p_{Min}	32.44	.913
Exponential	p_{Max}	32.47	.914
Nested	p_{Max}	**32.57**	**.916**

(b) *Ablation on the complexity-aware routing.*

Method	In-Domain		Out-of-Domain					
	Avg. AIO-3		Snow100K		DPDD		GoPro	
AirNet [24]	31.20	.910	22.31	.762	20.17	.654	21.95	.748
MoCE-IR-S								
w/ Load balance [43]	32.30	.913	22.79	.765	21.42	.659	21.87	.747
w/ Complexity bias (*ours*)	**32.57**	**.916**	**22.95**	**.772**	**21.74**	**.669**	**22.45**	**.756**
Improvement	+.27	+.003	+.16	+.007	+.32	+.010	+.58	+.009

the conventional load balancing approach [43], which enforces uniform expert utilization. As shown in Tab. 5b, our MoCE-IR framework demonstrates superior performance by preferentially routing to lower-complexity experts, activating higher-complexity experts only when their additional computational capacity provides clear advantages, emerging to a better model utilization in the context of image restoration. Besides, this allocation of processing resources not only improves efficiency but also enhances out-of-distribution generalization, particularly for degradations that diverge significantly from the training distribution.

Routing Analysis. Fig. 5 shows the average experts chosen in each layer for different degradation types - haze, rain, and noise. Experts with higher complexity are predominantly chosen for degradations requiring broader contextual understanding, such as haze and noise. In contrast, predominantly lightweight experts are selected when the network is tasked with rain streak removal. We observe that some experts handle multiple tasks, while others specialize in specific ones.

Our complexity-conditioned routing enables task-specific learning by directing inputs to the most relevant experts, bypassing irrelevant components. This input adaptivity prevents the MoCE layers from collapsing into a single-expert dependency, ensuring efficient and diverse operation.

In contrast, the baseline MoE approach, lacking complexity-awareness and scaled experts, uniformly distributes the degraded test samples across all the experts, adhering strictly to the original load-balancing strategy. However, our results demonstrate that this uniform allocation is suboptimal for all-in-one restoration, as it fails to leverage task-specific commonalities and distinctions effectively.

5. Conclusion, Limitations and Discussion

This paper introduces MoCE-IR, an efficient all-in-one image restoration model built on the novel concept of complexity experts. By designing experts with varying computational complexity and receptive fields, the approach naturally aligns different degradation tasks with appropriate resources overcoming limitations of prior works. This inherent bias toward simpler paths proves effective at task-discriminative learning, enabling efficient inference by bypassing irrelevant experts without compromising restoration quality. Extensive experiments across diverse degradation settings, including more demanding scenarios with multiple composited degradations, demonstrate our computational efficiency through selective expert utilization and consistent improvement over state-of-the-art methods.

While MoCE-IR shows substantial improvements, opportunities remain. The current image-level routing imposes scalability constraints. Future work could explore Soft-MoE [37] architectures for fine-grained spatial routing or early-exiting approaches [39] to further enhance processing speeds and input adaptability. Extending to synthetic-to-real adaptation, especially for video, represents a crucial next step. Performance optimization through mixed-precision computation across experts could further enhance speed and efficiency for resource-limited deployments.

References

[1] Yuang Ai, Huaibo Huang, Xiaoqiang Zhou, Jiexiang Wang, and Ran He. Multimodal prompt perceiver: Empower adaptiveness generalizability and fidelity for all-in-one image restoration. In *CVPR*, 2024. 2

[2] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *TPAMI*, 33(5):898–916, 2010. 5

[3] Yoshua Bengio, Nicholas Léonard, and Aaron Courville. Estimating or propagating gradients through stochastic neurons for conditional computation. *arXiv preprint arXiv:1308.3432*, 2013. 3

[4] Dana Berman and Shai Avidan. Non-local image dehazing. In *CVPR*, 2016. 2

[5] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *TIP*, 25(11):5187–5198, 2016. 1

[6] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *ECCV*, 2022. 1, 2, 5

[7] Yu-Wei Chen and Soo-Chang Pei. Always clear days: Degradation type and severity aware all-in-one adverse weather removal. *arXiv preprint arXiv:2310.18293*, 2023. 1, 2

[8] Zheng Chen, Yulun Zhang, Jinjin Gu, Linghe Kong, Xin Yuan, et al. Cross aggregation transformer for image restoration. *NeurIPS*, 2022. 1, 2

[9] Marcos V Conde, Gregor Geigle, and Radu Timofte. Instructir: High-quality image restoration following human instructions. In *ECCV*, 2024. 1, 2, 3, 5, 6, 7

[10] Xin Dang, Hai Wang, Jie Ren, and Le Chen. An application performance optimization model of mobile augmented reality based on hd restoration. In *CBD*, pages 201–206. IEEE, 2020. 1

[11] Yu Dong, Yihao Liu, He Zhang, Shifeng Chen, and Yu Qiao. Fd-gan: Generative adversarial networks with fusion-discriminator for single image dehazing. In *AAAI*, 2020. 5

[12] Huiyu Duan, Xiongkuo Min, Sijing Wu, Wei Shen, and Guangtao Zhai. Uniprocessor: A text-induced unified low-level image processor. In *ECCV*, 2024. 1, 3, 5, 6

[13] Qingnan Fan, Dongdong Chen, Lu Yuan, Gang Hua, Nenghai Yu, and Baoquan Chen. A general decoupled learning framework for parameterized image operators. *TPAMI*, 43(1): 33–47, 2019. 1, 5

[14] Hongyun Gao, Xin Tao, Xiaoyong Shen, and Jiaya Jia. Dynamic scene deblurring with parameter selective sharing and nested skip connections. In *CVPR*, 2019. 5

[15] Florin Girbacia, Silviu Butnariu, Alex Petre Orman, and Cristian Cezar Postelnicu. Virtual restoration of deteriorated religious heritage objects using augmented reality technologies. *European Journal of Science and Theology*, 9(2):223–231, 2013. 1

[16] Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. Mambair: A simple baseline for image restoration with state-space model. In *ECCV*, 2024. 1, 5

[17] Yu Guo, Yuan Gao, Yuxu Lu, Ryan Wen Liu, and Shengfeng He. Onerestore: A universal restoration framework for composite degradation. In *ECCV*, 2024. 6, 7

[18] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. In *CVPR*, 2020. 2

[19] Yitong Jiang, Zhaoyang Zhang, Tianfan Xue, and Jinwei Gu. Autodir: Automatic all-in-one image restoration with latent diffusion. *arXiv preprint arXiv:2310.10123*, 2023. 2

[20] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv*, 2017. 5

[21] Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *CVPR*, 2023. 2, 4

[22] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018. 1, 2

[23] 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. 6

[24] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-In-One Image Restoration for Unknown Corruption. In *CVPR*, 2022. 1, 2, 5, 6, 7, 8

[25] Zilong Li, Yiming Lei, Chenglong Ma, Junping Zhang, and Hongming Shan. Prompt-in-prompt learning for universal image restoration. *arXiv preprint arXiv:2312.05038*, 2023. 2

[26] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *ICCV*, 2021. 2, 5

[27] 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 *ECCV*, 2022. 1, 5

[28] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *ICCV*, 2021. 1, 2

[29] 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*, 2023. 2, 3

[30] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Controlling vision-language models for multi-task image restoration. In *ICLR*, 2024. 2, 3, 5, 6

[31] 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. *TIP*, 26(2):1004–1016, 2016. 5

[32] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, 2001. 5

[33] Chong Mou, Qian Wang, and Jian Zhang. Deep generalized unfolding networks for image restoration. In *CVPR*, 2022. 5

[34] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, 2017. 6

[35] Jinshan Pan, Deqing Sun, Hans Pfister, and Ming-Hsuan Yang. Blind image deblurring using dark channel prior. In *CVPR*, 2016. 2

[36] Vaishnav Potlapalli, Syed Waqas Zamir, Salman Khan, and Fahad Khan. Promptir: Prompting for all-in-one image restoration. In *NeurIPS*, 2023. 1, 2, 3, 5, 6, 7

[37] Joan Puigcerver, Carlos Riquelme Ruiz, Basil Mustafa, and Neil Houlsby. From sparse to soft mixtures of experts. In *ICLR*, 2024. 2, 3, 4, 8

[38] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In *CVPR*, 2019. 1

[39] David Raposo, Sam Ritter, Blake Richards, Timothy Lillicrap, Peter Conway Humphreys, and Adam Santoro. Mixture-of-depths: Dynamically allocating compute in transformer-based language models, 2024. 2, 8

[40] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *CVPR*, 2019. 2

[41] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *CVPR*, 2018. 2

[42] Wenqi Ren, Jinshan Pan, Hua Zhang, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks with holistic edges. In *ICCV*, 2020. 2

[43] Carlos Riquelme, Joan Puigcerver, Basil Mustafa, Maxim Neumann, Rodolphe Jenatton, André Susano Pinto, Daniel Keysers, and Neil Houlsby. Scaling vision with sparse mixture of experts. *NeurIPS*, 2021. 2, 3, 4, 8

[44] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241. Springer, 2015. 3

[45] Giovanni Saggio, Davide Borra, et al. Augmented reality for restoration/reconstruction of artefacts with artistic or historical value. In *Augmented reality: some emerging application areas*, pages 59–86. InTech Publications, 2011. 1

[46] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer, 2017. 3, 4

[47] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. Memnet: A persistent memory network for image restoration. In *ICCV*, 2017. 1, 2

[48] Chunwei Tian, Yong Xu, and Wangmeng Zuo. Image denoising using deep cnn with batch renormalization. *NN*, 2020. 5

[49] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *CVPR*, 2022. 1, 2, 5

[50] Cong Wang, Jinshan Pan, Wei Wang, Jiangxin Dong, Mengzhu Wang, Yakun Ju, and Junyang Chen. Promptrestorer: A prompting image restoration method with degradation perception. In *NeurIPS*, 2023. 2

[51] 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. *ICCV*, 2024. 5, 6

[52] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *CVPR*, 2022. 1, 2

[53] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 6

[54] Gang Wu, Junjun Jiang, Kui Jiang, and Xianming Liu. Harmony in diversity: Improving all-in-one image restoration via multi-task collaboration. In *ACMMM*, 2024. 5

[55] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *CVPR*, 2021. 2

[56] Fuzhi Yang, Huan Yang, Jianlong Fu, Hongtao Lu, and Baining Guo. Learning texture transformer network for image super-resolution. In *CVPR*, 2020. 5

[57] Hao Yang, Liyuan Pan, Yan Yang, and Wei Liang. Language-driven all-in-one adverse weather removal. In *CVPR*, 2024. 1, 2, 3

[58] Rajeev Yasarla and Vishal M Patel. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In *CVPR*, 2019. 1

[59] Ke Yu, Xintao Wang, Chao Dong, Xiaoou Tang, and Chen Change Loy. Path-restore: Learning network path selection for image restoration. *TPAMI*, 44(10):7078–7092, 2022. 3

[60] Xiaoyan Yu, Shen Zhou, Huafeng Li, and Liehuang Zhu. Multi-expert adaptive selection: Task-balancing for all-in-one image restoration. *arXiv preprint arXiv:2407.19139*, 2024. 1

[61] Eduard Zamfir, Zongwei Wu, Nancy Mehta, Danda Pani Paudel, Yulun Zhang, and Radu Timofte. Efficient degradation-aware any image restoration, 2024. 1, 2, 3

[62] Eduard Zamfir, Zongwei Wu, Nancy Mehta, Yulun Zhang, and Radu Timofte. See more details: Efficient image super-resolution by experts mining. In *ICML*. PMLR, 2024. 3

[63] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, 2021. 2, 5, 6

[64] 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 *CVPR*, 2022. 1, 2, 3, 4, 5

[65] Jinghao Zhang, Jie Huang, Mingde Yao, Zizheng Yang, Hu Yu, Man Zhou, and Feng Zhao. Ingredient-oriented multi-degradation learning for image restoration. In *CVPR*, 2023. 1, 2, 5, 7

[66] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *CVPR*, 2017. 1, 2

[67] Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. Residual non-local attention networks for image restoration. *arXiv preprint arXiv:1903.10082*, 2019. 1, 2

[68] Haiyu Zhao, Yuanbiao Gou, Boyun Li, Dezhong Peng, Jiancheng Lv, and Xi Peng. Comprehensive and delicate: An efficient transformer for image restoration. In *CVPR*, 2023. 2

[69] Yanqi Zhou, Tao Lei, Hanxiao Liu, Nan Du, Yanping Huang, Vincent Zhao, Andrew M Dai, Quoc V Le, James Laudon, et al. Mixture-of-experts with expert choice routing. *NeurIPS*, 2022. 2

[70] Yurui Zhu, Tianyu Wang, Xueyang Fu, Xuanyu Yang, Xin Guo, Jifeng Dai, Yu Qiao, and Xiaowei Hu. Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. In *CVPR*, 2023. 6

[71] Ozan Özdenizci and Robert Legenstein. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *TPAMI*, 2023. 6