ClearAIR: A Human-Visual-Perception-Inspired All-in-One Image Restoration

Xu Zhang[1], Huan Zhang[2], Guoli Wang[3], Qian Zhang[3], Lefei Zhang[1]*

[1]National Engineering Research Center for Multimedia Software, School of Computer Science, Wuhan University
[2]School of Information Engineering, Guangdong University of Technology
[3]Horizon Robotics
{zhangx0802, zhanglefei}@whu.edu.cn, huanzhang2021@gdut.edu.cn, {guoli.wang, qian01.zhang}@horizon.auto

Abstract

Recently, All-in-One image restoration (AiOIR) has advanced significantly, offering promising solutions for complex real-world degradations. However, most existing approaches heavily rely on degradation-specific representation learning, which can lead to oversmoothing and artifacts in the restored images. To address this limitation, we propose ClearAIR, a novel AiOIR framework inspired by human visual perception and designed with a hierarchical restoration strategy in a coarse-to-fine manner. First, leveraging the global priority characteristic of early human visual perception, we employ an image quality assessment model to evaluate the overall image structure and degradation level. Next, we introduce a Semantic Guidance Unit to provide coarse semantic region guidance and a Task Identifier to predict local degradation types, enabling a more informed characterization of local degradation patterns. Finally, aiming at the challenge of local detail restoration, we propose an Internal Clue Reuse Mechanism that deeply mines the internal information of the image in a self-supervised manner to enhance the model's capacity for fine-detail recovery. Experimental results demonstrate that ClearAIR achieves superior restoration performance across diverse synthetic and real-world datasets.

Code — https://github.com/House-yuyu/ClearAIR

Introduction

Image restoration aims to recover a clean image from its degraded version and has made significant progress with deep learning. Early approaches employed task-specific networks for individual degradation types, such as denoising (Zhang et al. 2017; Zhang, Zuo, and Zhang 2018), dehazing (Cai et al. 2016; Song et al. 2023), deraining (Jiang et al. 2020; Chen et al. 2023), deblurring (Nah, Hyun Kim, and Mu Lee 2017), and low-light enhancement (Wei et al. 2018; Cai et al. 2023), achieving strong performance within their intended domains. However, these methods lack generalization across tasks. Although general-purpose restoration models (Chen et al. 2022a; Zamir et al. 2022) have been developed to handle multiple degradations, they often still require separate models for each degradation type, resulting in complex inference pipelines and increased computational costs.

*Corresponding author.

In recent years, All-in-One image restoration (AiOIR) methods (Wu et al. 2024a; Li et al. 2025a) have emerged as promising solutions. These frameworks can simultaneously handle diverse degradation types through various mechanisms. Early efforts, like AirNet (Li et al. 2022), focused on creating specific degradation encoders to capture distinctive feature representations. Subsequent innovations, such as those found in ProRes (Ma et al. 2023) and PromptIR (Potlapalli et al. 2023), enhanced performance by incorporating visual prompts. More recent research (Ai et al. 2024; Zhang et al. 2025c) has harnessed the powerful feature extraction capabilities of large-scale visual models to improve texture reconstruction and ensure structural integrity. However, these AiOIR methods overlook a critical issue: spatially non-uniform degradations can significantly alter the local statistical properties of an image. Most existing AiOIR approaches apply a uniform processing strategy across the entire image, failing to account for variations in degradation distribution and severity across different regions.

To address this limitation, as shown in Fig. 1, we design a progressive restoration pipeline inspired by human visual perception (HVP), which refines image quality hierarchically from global structure to fine local details. First, as in early HVP stages, which emphasize global structure, we integrate an MLLM-based Image Quality Assessment (IQA) model to evaluate the image's overall quality. Second, to better account for spatially varying degradation patterns, we incorporate a Semantic Guidance Unit (SGU) to support region-level segmentation and provide coarse guidance for identifying areas likely affected by degradation. Third, guided by the spatial cues from the SGU, we apply a task identifier to estimate the predominant degradation type in local neighborhoods. This allows ClearAIR to adaptively select region-appropriate restoration strategies, avoiding a uniform one-size-fits-all treatment across the image. Finally, to enhance the recovery of fine-grained local details, we propose an Internal Clue Reuse Mechanism (ICRM) that leverages internal image statistics to refine local structures.

Our main contributions can be summarized as follows:

- We present ClearAIR, a novel AiOIR framework inspired by HVP. By adopting a coarse-to-fine hierarchical restoration process, it gradually improves both structural integrity and perceptual quality.
- We propose an HVP-inspired pipeline integrating global

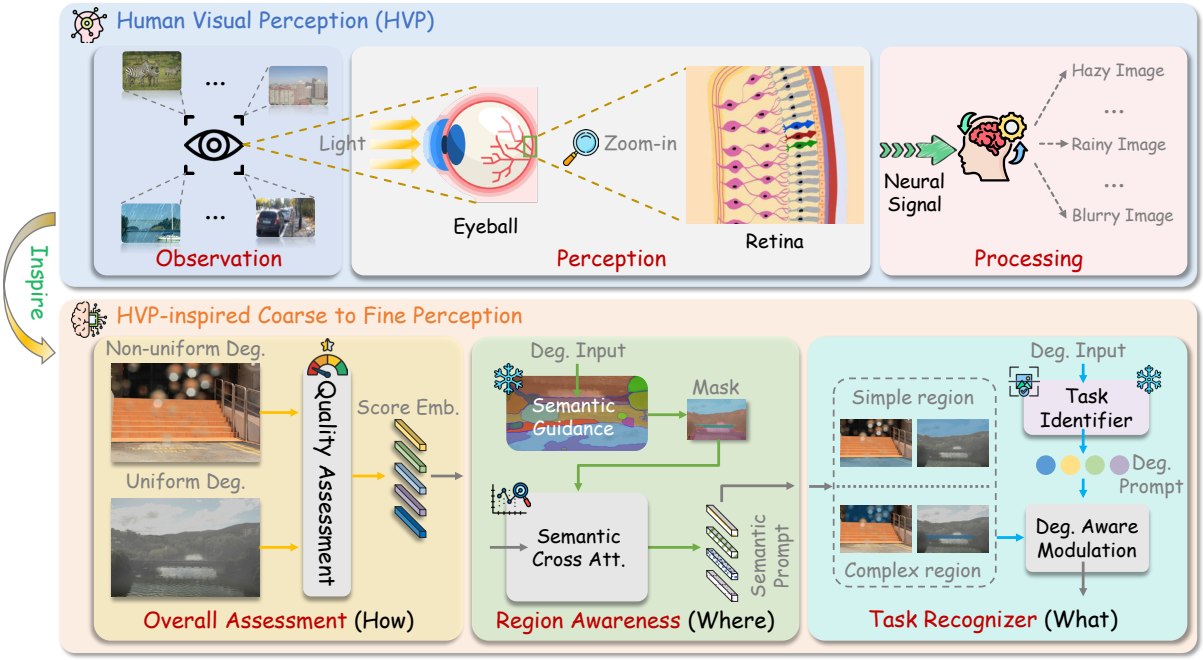

Figure 1: A coarse-to-fine image processing pipeline inspired by human visual perception.

IQA and local attention cues. An MLLM-based IQA evaluates quality, while the SGU and task identifier guide regional analysis and degradation estimation.

- We introduce ICRM that leverages self-supervised learning to exploit the intrinsic structure of the image, thereby enhancing the model's ability to recover fine local details.

Related Work

All-in-One Image Restoration

All-in-One image restoration (Ma et al. 2025; Wu et al. 2025a; Zhang et al. 2025d; Wu et al. 2025b) has emerged as a promising direction in low-level vision, aiming to restore clean images from diverse degradation types using a single unified model. Compared with task-specific (Zhang et al. 2023b; Li et al. 2025b; Cao et al. 2025; Li et al. 2024a; Xiao et al. 2024; Li et al. 2024c; Zhao et al. 2024; Li et al. 2024b; Xiao and Wang 2025; Chang et al. 2025a,b; Xiao et al. 2025b,a; Kong et al. 2023; Zhang et al. 2025b) and general restoration methods (Gou et al. 2020; Zhang et al. 2025a; Gou et al. 2024), all-in-one approaches offer significant advantages in multi-task capability, making it more suitable for practical applications with diverse degradation scenarios. For example, AirNet (Li et al. 2022) introduced a contrastive learning strategy to learn discriminative degradation representations. Prompt-based methods such as PromptIR (Potlapalli et al. 2023) and ProRes (Ma et al. 2023) further improved multi-degradation handling by incorporating vision prompts into the network. More recently, DA-CLIP (Luo et al. 2023) and MPerceiver (Ai et al. 2024) leveraged pre-trained large-scale vision models to boost performance on complex restoration tasks. Perceive-IR (Zhang et al. 2025c) showed that jointly recognizing degradation types and severity improves restoration, underscoring the importance of comprehensive degradation perception in all-in-one frameworks.

Despite these advancements, most all-in-one image restoration methods adopt a uniform processing strategy, failing to account for the spatial variability of degradation. Moreover, even in cases of uniformly distributed degradation, the difficulty of restoration varies significantly depending on the texture complexity of different regions. For example, flat regions are generally easier to restore, while areas with complex textures pose greater challenges.

Human Visual Perception

In visual cognition, humans exhibit specific characteristics. Typically, a visual image is first perceived as a unified whole before being analyzed in terms of its constituent parts. Recently, several studies have leveraged this perceptual mechanism to achieve promising results. For instance, Dream (Xia et al. 2024) reversely models the hierarchical processing of the human visual perception (HVP) into a computable encoding-decoding framework, revealing that both biological and artificial vision systems aim for efficient visual information coding at their core. In this paper, we propose a hierarchical image perception pipeline that mimics human visual processing: global quality assessment to semantic-driven regional localization and degradation identification of distorted regions. By integrating global coarse-grained understanding and local fine-grained perception, our method not only improves the visual naturalness of restored images but also maintains semantic consistency under complex degradation, enabling more human-like AiOIR.

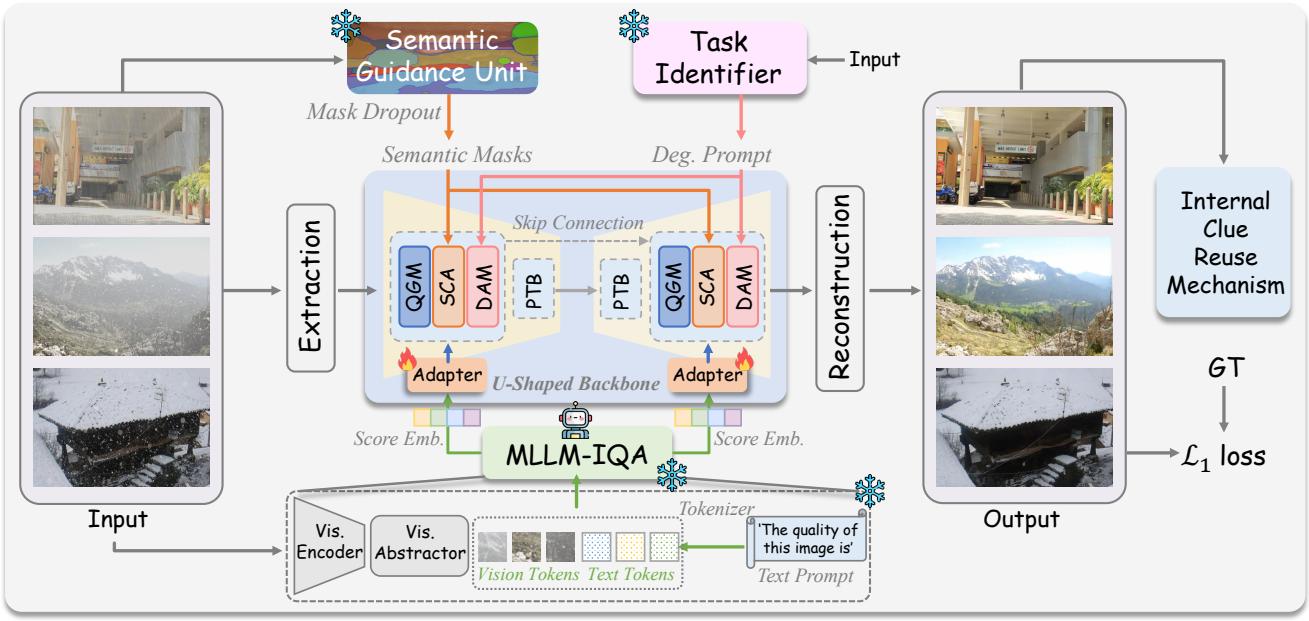

Figure 2: An overview architecture of the proposed ClearAIR.

Image Quality Assessment

Image Quality Assessment (IQA) methods primarily rely on quality scores to evaluate image quality. These methods can be categorized into reference and no-reference approaches, depending on whether they require a high-quality reference image. In the field of image restoration, no-reference IQA is commonly used because it can directly regress a quality score without needing a reference image, aligning well with practical restoration needs. In recent years, Multi-Modal Large Language Models (MLLM)-based IQA methods have leveraged the foundational knowledge of MLLM to achieve superior performance and more detailed assessment results (Wu et al. 2024b,c). Q-Bench (Wu et al. 2024b) demonstrates that general MLLM possess some low-level perception capabilities. Q-Instruct (Wu et al. 2024c) further enhances these capabilities by introducing a large-scale dataset. Recently, DeQA (You et al. 2025) has advanced this area by using MLLM to regress precise quality scores, achieving remarkable performance. In this paper, we aim to harness the powerful ability of MLLMs to mine multi-modal cues, employing them as an initial estimator for overall image quality. This serves as a solid foundation for subsequent human visual perception processes.

Methodology

Overall Pipeline

The overall framework is illustrated in Fig. 2. ClearAIR consists of four components: 1) MLLM-based IQA: It generates a quality score embedding from visual and textual tokens, which guides the restoration backbone via the Quality Guidance Module (QGM). 2) Semantic Guidance Unit (SGU): It provides region-level semantic masks, with features fused through Semantic Cross Attention (SCA). 3) Task Identifier: It estimates the predominant degradation type and encodes this prediction into a degradation prompt used by the Degradation-aware Modulation (DAM). 4) Internal Clue Reuse Mechanism: It exploits self-supervised learning to extract internal image cues, enhancing fine-detail reconstruction. The optimization objective of the overall process can be represented as:

$$\mathcal{L}_{total} = \mathcal{L}_1 + \alpha \cdot \mathcal{L}_{inter}, \tag{1}$$

where α is a hyperparameter set to 0.25. The details of $\mathcal{L}_{inter}$ are provided in the Internal Clue Reuse Mechanism section.

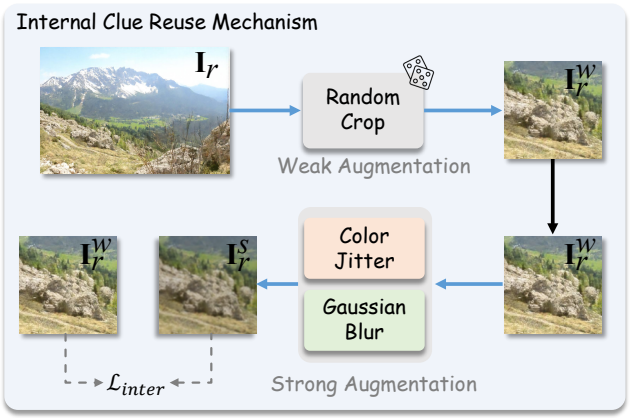

Figure 3: The proposed Internal Clue Reuse Mechanism.

MLLM-based Image Quality Assessment. We employ DeQA (You et al. 2025) as our MLLM-IQA model. As

shown in Fig. 2, a vision encoder is used to encode the input image into visual tokens. In addition, a vision abstractor is utilized as part of the connector module, which further compresses the visual tokens. Finally, the visual and textual tokens are fused and fed into a large MLLM for response prediction. We do not directly use the quality map to guide the restoration process. Instead, we extract the state $\mathcal{Q}$ from the layer preceding the 'quality level' token. This representation more faithfully captures the MLLM-IQA model's underlying reasoning about image quality. Subsequently, $\mathcal{Q}$ is integrated into the Quality Guidance Module (QGM) as score embeddings via an affine transformation. Given a degraded image $\mathbf{I}_d$ along with its corresponding textual description $\mathbf{T}_d$, the above process can be expressed as follows:

$$\mathcal{Q} = \mathcal{M}_{iqa}(\mathbf{I}_d, \mathbf{T}_d), \tag{2}$$

$$\mathbf{F}_q = MLP(\mathcal{Q}), \tag{3}$$

$$\mathbf{X}^{out}_{qgm} = \mathbf{X}^{in}_{qgm} \odot \text{Linear}(\mathbf{F}_q) + \text{Linear}(\mathbf{F}_q), \tag{4}$$

where $\mathcal{M}_{iqa}$ refers MLLM-based IQA model (DeQA). $\mathbf{F}_q$ denotes the features after being transformed by the adapter. $\mathbf{X}^{in}_{qgm}$ and $\mathbf{X}^{out}_{qgm}$ represent the input and output features of the QGM, respectively.

Semantic Guidance Unit. In this part, we introduce the SGU, which leverages a pre-trained Segment Anything Model (Ravi et al. 2024; Rong et al. 2025) to extract high-level semantics. Given a degraded image $\mathbf{I}_d$, SGU generates N_m binary masks:

$$\mathbf{I}_{mask} \in \{0,1\}^{H\times W\times N_m}, \tag{5}$$

where each mask $m_i \in \mathbb{R}^{H\times W\times 1}$ highlights a distinct region. These masks are integrated with shallow features $\mathbf{F}_s \in \mathbb{R}^{H\times W\times C}$ via Mask Average Pooling (MAP). For each mask m_i, we compute the average feature within the masked region and broadcast it back:

$$\bar{\mathbf{f}}_i = \frac{1}{|\Omega_i|}\sum_{(h,w)\in\Omega_i} \mathbf{F}_s(h,w),\ \mathbf{F}_{sem}(h,w) = \bar{\mathbf{f}}_i,\ \forall (h,w)\in\Omega_i, \tag{6}$$

where $\Omega_i = \{(h,w) \mid m_i(h,w) = 1\}$. The output $\mathbf{F}_{sem} \in \mathbb{R}^{H\times W\times C}$ encodes semantic-aware structural priors.

To enhance robustness to fluctuations in mask quality resulting from degradation severity or model scale, we introduce mask dropout during training, removing a random subset of masks and merging their regions into the background. Finally, $\mathbf{F}_{sem}$ interacts with the restoration backbone through Semantic Cross Attention (SCA) enabling region-level semantic guidance in the restoration process. This process can be expressed as follows:

$$\mathbf{Q} = \mathbf{F}^{in}_{sca},\ \mathbf{K} = \mathbf{W}_k\mathbf{F}_{sem},\ \mathbf{V} = \mathbf{W}_v\mathbf{F}_{sem}, \tag{7}$$

$$\mathbf{F}^{out}_{sca} = Softmax\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right)\mathbf{V}, \tag{8}$$

where $\mathbf{F}^{in}_{sca}$ and $\mathbf{F}^{out}_{sca}$ represent the input and output features of the SCA module, respectively.

Task Identifier. We employ DA-CLIP (Luo et al. 2023) as the backbone of the Task Identifier to generate both content embeddings $\mathcal{F}_c$ and degradation embeddings $\mathcal{F}_d$. Given the diversity of degradation types in our experimental setup, we conduct pre-training for each specific degradation pattern. The degradation embedding is then transformed into a degradation prompt $\mathcal{F}_p$, which can be described as:

$$\mathcal{F'}_d = \mathcal{P} \odot Softmax\big(MLP(\mathcal{F}_d)\big),\ \mathcal{F}_p = MLP(\mathcal{F'}_d), \tag{9}$$

where $\mathcal{P}$ denotes a set of learnable prompts. Subsequently, the feature $\mathbf{X}^{in}_{dam}$ is fed into the Degradation-aware Modulation (DAM), which can be formulated as:

$$\mathbf{X'}_{dam} = SimpleGate\Big(Conv\big(Norm(\mathbf{X}^{in}_{dam})\big)\Big), \tag{10}$$

$$\mathbf{X}^{out}_{dam} = Conv\big(SelfAtt.(\mathbf{X'}^{in}_{dam}, \mathcal{F}_c)\big). \tag{11}$$

By integrating degradation-related features into the network, the model can adaptively perceive different types of degradation features, thereby acquiring the ability to perform fine-grained recognition of the degradation.

Internal Clue Reuse Mechanism. As shown in Fig. 3, we propose ICRM to enhance the model's ability to preserve fine details in restored images. Specifically, we apply data augmentation with varying levels of strength to the restored output $\mathbf{I}_r$ from the restoration model. First, weak augmentation is applied to $\mathbf{I}_r$, which can be represented as:

$$\mathbf{I}^w_r = \mathcal{F}_{weak}(\mathbf{I}_r), \tag{12}$$

where $\mathcal{F}_{weak}(\cdot)$ denotes a weak data augmentation operation, and $\mathbf{I}^w_r$ is the output image after applying weak augmentation. Subsequently, strong augmentation is performed on the weakly augmented image $\mathbf{I}^w_r$, which can be expressed

$$\mathbf{I}^s_r = \mathcal{F}_{strong}(\mathbf{I}^w_r), \tag{13}$$

where $\mathcal{F}_{strong}(\cdot)$ represents a more aggressive augmentation transformation, and $\mathbf{I}^s_r$ is the resulting strongly augmented image. In our implementation, weak augmentation consists of random cropping, while strong augmentation includes color jittering and Gaussian blurring. Finally, we compute the L2 distance between the weakly and strongly augmented results to form an internal consistency loss:

$$\mathcal{L}_{inter} = \gamma \cdot \|\mathbf{I}^w_r - \mathbf{I}^s_r\|^2_2, \tag{14}$$

where γ is a hyperparameter that controls the weight of this loss. In our experiments, we set the initial value of γ = 0.05.

Settings	No. Datasets	Degradation Type
Three Deg.	3	Noise, Haze, Rain
Five Deg.	5	Noise, Haze, Rain, Blur, Low-light
All-Weather	3	Haze, Rain, Raindrop, Snow
Composited Deg.	1	Haze, Rain, Low-light, Snow

Table 1: The details of the All-in-One setting.

Method	Source	Params.	*Dehazing*		*Deraining*		*Denoising*						Average	
			SOTS		Rain100L		$BSD68_{\sigma=15}$		$BSD68_{\sigma=25}$		$BSD68_{\sigma=50}$			
DL (Fan et al. 2019)	TPAMI'19	2M	26.92	.931	32.62	.931	33.05	.914	30.41	.861	26.90	.740	29.98	.876
AirNet (Li et al. 2022)	CVPR'22	9M	27.94	.962	34.90	.967	33.92	.933	31.26	.888	28.00	.797	31.20	.910
IDR (Zhang et al. 2023a)	CVPR'23	15M	29.87	.970	36.03	.971	33.89	.931	31.32	.884	28.04	.798	31.83	.911
PromptIR (Potlapalli et al. 2023)	NeurIPS'23	36M	30.58	.974	36.37	.972	33.98	.933	31.31	.888	28.06	.799	32.06	.913
NDR (Yao et al. 2024)	TIP'24	28M	28.64	.962	35.42	.969	34.01	.932	31.36	.887	28.10	.798	31.51	.910
Gridformer (Wang et al. 2024)	IJCV'24	34M	30.37	.970	37.15	.972	33.93	.931	31.37	.887	28.11	.801	32.19	.912
InstructIR (Conde, Geigle, and Timofte 2024)	ECCV'24	16M	30.22	.959	37.98	.978	34.15	.933	31.52	.890	28.30	.803	32.43	.913
Perceive-IR (Zhang et al. 2025c)	TIP'25	42M	30.87	.975	38.29	.980	34.13	.934	**31.53**	.890	**28.31**	**.804**	32.63	.917
AdaIR (Cui et al. 2025)	ICLR'25	29M	31.06	.980	38.64	.983	34.12	.934	31.45	**.892**	28.19	.802	32.69	.918
VLU-Net (Zeng et al. 2025)	CVPR'25	35M	30.71	.980	**38.93**	**.984**	34.13	**.935**	31.48	**.892**	28.23	**.804**	32.70	**.919**
ClearAIR (Ours)	-	31M	**31.08**	**.981**	38.61	**.984**	**34.18**	**.935**	31.50	.891	**28.31**	**.804**	**32.74**	**.919**

Table 2: Comparison to state-of-the-art AiOIR methods on Three Degradations task.

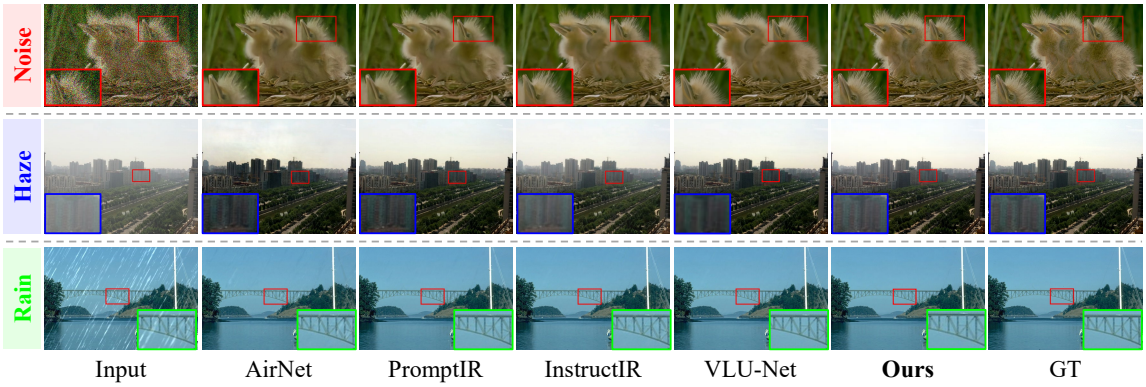

Figure 4: Visual comparisons of ClearAIR with state-of-the-art AiOIR methods on Three Degradations task.

Experiments

Experimental Setup

Datasets. We conduct experiments under two settings: All-in-One and Single-task following the protocols established in prior works (Zhang et al. 2025c). As shown in Table 1, the All-in-One setting entails a comprehensive evaluation across four tasks, with further details on the Single-task setting provided in the **Appendix**.

Implementation Details. We employ DeQA (You et al. 2025) as the MLLM-IQA model and chose Restormer (Zamir et al. 2022) as the restoration backbone. We follow a configuration similar to the original Restormer. Specifically, from level-1 to level-4, the numbers of Prompt Transformer Blocks are set to [3, 5, 6, 8], the attention heads in MDTA modules are [1, 2, 4, 8], and the channel dimensions are [48, 96, 192, 384]. We optimize the network using AdamW ($\beta_1 = 0.9, \beta_2 = 0.999$) with a learning rate of 2×10^{-4} and batch size of 4. Training runs for 300K iterations. The total loss weights are set as $\lambda_1 = 0.1$ and $\lambda_2 = 0.05$. All experiments are conducted on NVIDIA GeForce RTX 4090 GPUs. During training, inputs are randomly cropped to 256×256 patches, and random horizontal and vertical flips are applied for data augmentation.

All-in-One Image Restoration Results

Three Degradations Task. We evaluate our model on three restoration tasks: denoising, dehazing, and deraining. As shown in Table 2, ClearAIR achieves the best average performance, with notable gains in challenging cases like high-noise removal and severe haze. Compared to AdaIR and VLU-Net, ClearAIR delivers competitive or superior results by better modeling human visual perception. For example, it achieves 31.08 dB PSNR on SOTS (vs. 30.71 dB for VLU-Net), demonstrating that perceptual improvements can compensate for the absence of physical priors. Qualitative results in Figure 4 further confirm ClearAIR's effectiveness: it preserves texture under heavy noise, removes rain streaks while maintaining sharpness, and restores both global contrast and fine details in dense haze.

Five Degradations Task. We extend ClearAIR to five degradation tasks, using GoPro for deblurring and LOL for low-light enhancement. As shown in Table 3, ClearAIR achieves superior performance on most tasks, excelling particularly in deblurring. Although slightly behind specialized methods in low-light enhancement and denoising, it remains highly competitive and achieves the highest average PSNR (30.45 dB) and SSIM (0.916), demonstrating strong multi-task capability. Additional visual results are provided in the **Appendix**.

Method	Source	Params.	*Dehazing*		*Deraining*		*Denoising*		*Deblurring*		*Low-Light*		Average	
			SOTS		Rain100L		BSD68$_{\sigma=25}$		GoPro		LOLv1			
DL (Fan et al. 2019)	TPAMI'19	2M	20.54	.826	21.96	.762	23.09	.745	19.86	.672	19.83	.712	21.05	.743
AirNet (Li et al. 2022)	CVPR'22	9M	21.04	.884	32.98	.951	30.91	.882	24.35	.781	18.18	.735	25.49	.847
IDR (Zhang et al. 2023a)	CVPR'23	15M	25.24	.943	35.63	.965	**31.60**	.887	27.87	.846	21.34	.826	28.34	.893
PromptIR (Potlapalli et al. 2023)	NeurIPS'23	33M	26.54	.949	36.37	.970	31.47	.886	28.71	.881	22.68	.832	29.15	.904
Gridformer (Wang et al. 2024)	IJCV'24	34M	26.79	.951	36.61	.971	31.45	.885	29.22	.884	22.59	.831	29.33	.904
InstructIR (Conde, Geigle, and Timofte 2024)	ECCV'24	16M	27.10	.956	36.84	.973	31.40	.887	29.40	.886	**23.00**	.836	29.55	.907
Perceive-IR (Zhang et al. 2025c)	TIP'25	42M	28.19	.964	37.25	.977	31.44	.887	29.46	.886	22.81	.833	29.84	.909
AdaIR (Cui et al. 2025)	ICLR'25	29M	30.53	.978	38.02	.981	31.35	.888	28.12	.858	**23.00**	.845	30.20	.910
VLU-Net (Zeng et al. 2025)	CVPR'25	35M	**30.84**	**.980**	**38.54**	**.982**	31.43	**.891**	27.46	.840	22.29	.833	30.11	.905
ClearAIR (Ours)	-	31M	30.12	.978	38.20	**.982**	31.53	.888	**29.67**	**.887**	22.83	**.846**	**30.45**	**.916**

Table 3: Comparison to state-of-the-art AiOIR methods on Five Degradations task.

Method	Source	Params.	*Snow100K-S*		*Snow100K-L*		*Outdoor-Rain*		*RainDrop*		Average	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
All-in-One (Li, Tan, and Cheong 2020)	CVPR'20	-	-	-	28.33	.882	24.71	.898	31.12	.927	28.05	.902
Transweather (Valanarasu, Yasarla, and Patel 2022)	CVPR'22	38M	32.51	.934	29.31	.888	28.83	.900	30.17	.916	30.20	.909
TKL (Chen et al. 2022b)	CVPR'22	29M	34.42	.947	30.22	.907	29.27	.915	31.81	.931	31.43	.925
WGWSNet (Zhu et al. 2023)	CVPR'23	26M	34.31	.946	30.16	.901	29.32	.921	32.38	.938	31.54	.926
WeatherDiff (Özdenizci and Legenstein 2023)	TPAMI'23	83M	35.83	.957	30.09	.904	29.64	.931	30.71	.931	31.57	.931
AWRCP (Ye et al. 2023)	ICCV'23	-	36.92	.965	31.92	.934	31.39	.933	31.93	.931	33.04	.941
Histoformer (Sun et al. 2024)	ECCV'24	30M	37.41	.966	32.16	.926	32.08	.939	**33.06**	**.944**	33.68	.945
T^3-DiffWeather (Chen et al. 2024)	ECCV'24	69M	37.51	.966	32.37	**.936**	31.09	.937	32.66	.941	33.41	.945
ClearAIR (Ours)	-	31M	**37.79**	**.967**	**32.53**	.932	**32.45**	**.941**	32.82	.942	**33.90**	**.946**

Table 4: Comparison to state-of-the-art All-in-One methods on All-Weather task.

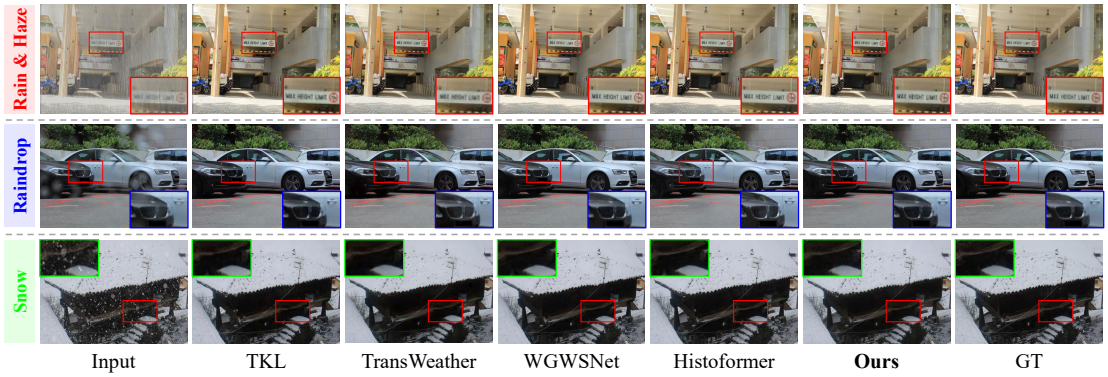

Figure 5: Visual comparisons of ClearAIR with state-of-the-art AiOIR methods on All-Weather task.

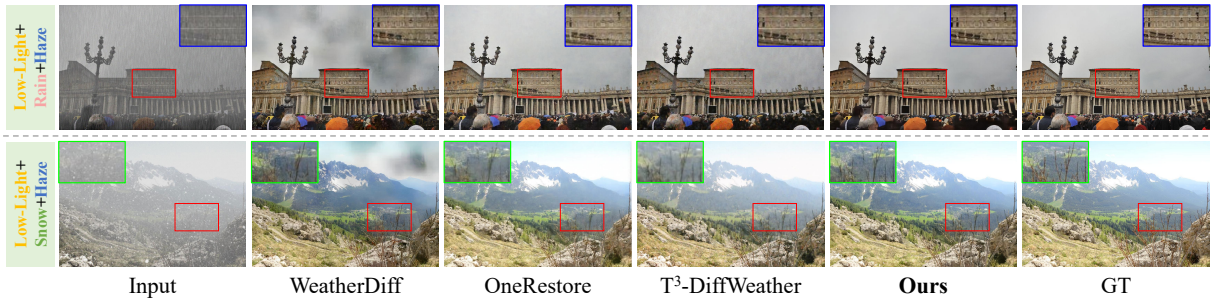

Figure 6: Visual comparisons of ClearAIR with state-of-the-art AiOIR methods on Composited Degradation task.

All-Weather Task. We evaluate ClearAIR on All-Weather task including: snow, rain & haze, and raindrop. As shown in Table 4, ClearAIR achieves the best or second-best performance across all datasets. These consistent gains across diverse weather conditions demonstrate its effectiveness in handling complex adverse weather degradations. Qualitative comparisons in Figure 5 further illustrate this: ClearAIR produces clearer, more natural results, effectively removing weather artifacts while better preserving fine details and textures.

Composited Degradation Task. We evaluate ClearAIR under challenging Composited Degradations task, considering both individual and combined cases. As shown in Table 5, ClearAIR achieves an average gain of 0.62 dB over OneRestore (Guo et al. 2024), outperforming existing All-in-One models, which validates its effectiveness in modeling human visual perception. Qualitative results in Figure 6 demonstrate superior removal of composite degradations while preserving fine details and textures.

Method	Source	Params.	PSNR↑	SSIM↑
AirNet	CVPR'22	9M	23.75	0.814
TransWeather	CVPR'22	38M	23.13	0.781
WeatherDiff	TPAMI'23	83M	22.49	0.799
PromptIR	NeurIPS'23	33M	25.90	0.850
WGWSNet	CVPR'23	26M	26.96	0.863
OneRestore	ECCV'24	6M	28.72	0.882
ClearAIR (Ours)	-	31M	**29.34**	**0.886**

Table 5: Comparison to state-of-the-art All-in-One methods on Composited Degradations task (CDD-11 dataset).

Ablation Study

This section analyzes the impact of different design choices in ClearAIR on model performance. All experiments are conducted on the Rain100L dataset (Yang et al. 2017) using a training of 100K iterations.

Effects of Perception Order. To investigate the impact of perception order, the sequence of coarse-grained quality assessment (How), fine-grained region semantics perception and degradation type recognition (Where and What), we design two additional experimental setups based on the baseline order: Where-What-How and What-How-Where (indexes a and b). As shown in Table 6, the Where-What-How order yields the worst performance. This may be because perceiving region-level semantic information first disrupts the structural integrity that is crucial for coarse quality assessment. Notably, the What-How-Where order achieves the second-best result, which aligns with the workflow of some All-in-One restoration methods that begin with degradation characterization. In contrast, our method, inspired by the human visual perception process, achieves the best overall performance.

Effects of Different Components

As shown in Table 7, we conduct ablation studies to evaluate the contribution of each proposed component. The experimental settings are as follows: (1) w/o MLLM-IQA (IQA): replaces quality guidance with a learnable parameter; (2) w/o SGU: replaces semantic priors with learnable parameters; (3) w/o Task identifier (TI): removes degradation prompts, using a learnable parameter instead; (4) w/o LCRM: removes the Internal Clue Reuse Mechanism. Results show that variants (a–c), which replace structured priors (quality, semantic, or task information) with unstructured learnable parameters, suffer significant performance drops. This confirms that explicit modeling of such priors is crucial for effective restoration guidance and task adaptation. In contrast, variants (d–f), which retain partial structured information (e.g., improved quality or degradation estimation), achieve noticeable gains, highlighting the importance of accurate perceptual and degradation awareness. Finally, removing LCRM (variant g) also degrades performance, though less severely, indicating its role in enhancing restoration by reusing internal image structures and contextual cues.

Index	Order	PSNR↑	SSIM↑
a	Where-What-How	37.89	0.982
b	What-How-Where	38.04	0.983
Ours	How-Where-What	38.21	0.986

Table 6: Effectiveness of perception order.

Index	IQA	SGU	TI	ICRM	PSNR↑	SSIM↑
a	✓	✗	✗	✓	37.57	0.980
b	✗	✓	✗	✓	37.43	0.978
c	✗	✗	✓	✓	37.52	0.980
d	✓	✓	✗	✓	38.05	0.985
e	✓	✗	✓	✓	37.93	0.984
f	✗	✓	✓	✓	37.87	0.984
g	✓	✓	✓	✗	38.03	0.985
Ours	✓	✓	✓	✓	**38.21**	**0.986**

Table 7: Effectiveness of different components.

Conclusion

In this paper, we propose ClearAIR, a novel AiOIR framework inspired by human visual perception and designed with a hierarchical, coarse-to-fine restoration strategy. By mimicking the human visual system's tendency to first perceive an image as a whole before focusing on local details, our method integrates global structure assessment, attention-driven local region analysis, and internal clue reuse for fine-grained restoration. The combination of an image quality assessment model, the semantic guidance unit, and a task recognizer enables accurate localization and understanding of degradation patterns. Furthermore, the proposed internal clue reuse mechanism enhances the model's ability to recover detailed textures in a self-supervised manner. Experimental results demonstrate that ClearAIR achieves state-of-the-art performance on both synthetic and real-world datasets.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 62431020, the National Key Research and Development Program of China under Grant 2024YFE0111800, and the Fundamental Research Funds for the Central Universities under Grant 2042025kf0030.

References

Ai, Y.; Huang, H.; Zhou, X.; Wang, J.; and He, R. 2024. Multimodal prompt perceiver: Empower adaptiveness generalizability and fidelity for all-in-one image restoration. In *CVPR*, 25432–25444.

Cai, B.; Xu, X.; Jia, K.; Qing, C.; and Tao, D. 2016. Dehazenet: An end-to-end system for single image haze removal. *IEEE TIP*, 25(11): 5187–5198.

Cai, Y.; Bian, H.; Lin, J.; Wang, H.; Timofte, R.; and Zhang, Y. 2023. Retinexformer: One-stage Retinex-based Transformer for Low-light Image Enhancement. In *ICCV*, 12504–12513.

Cao, J.; Zeng, Z.; Zhang, X.; Zhang, H.; Fan, C.; Jiang, G.; and Lin, W. 2025. Unveiling the underwater world: CLIP perception model-guided underwater image enhancement. *PR*, 162: 111395.

Chang, L.; Wang, Y.; Deng, L.; Du, B.; and Xu, C. 2025a. WaterDiffusion: Learning a Prior-involved Unrolling Diffusion for Joint Underwater Saliency Detection and Visual Restoration. In *AAAI*, 1998–2006.

Chang, L.; Wang, Y.; Du, B.; and Xu, C. 2025b. Color Correction Meets Cross-Spectral Refinement: A Distribution-Aware Diffusion for Underwater Image Restoration. *arXiv preprint arXiv:2501.04740*.

Chen, L.; Chu, X.; Zhang, X.; and Sun, J. 2022a. Simple baselines for image restoration. In *ECCV*, 17–33.

Chen, S.; Ye, T.; Zhang, K.; Xing, Z.; Lin, Y.; and Zhu, L. 2024. Teaching Tailored to Talent: Adverse Weather Restoration via Prompt Pool and Depth-Anything Constraint. In *ECCV*, 95–115.

Chen, W.-T.; Huang, Z.-K.; Tsai, C.-C.; Yang, H.-H.; Ding, J.-J.; and Kuo, S.-Y. 2022b. Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. In *CVPR*, 17653–17662.

Chen, X.; Li, H.; Li, M.; and Pan, J. 2023. Learning a Sparse Transformer Network for Effective Image Deraining. In *CVPR*, 5896–5905.

Conde, M. V.; Geigle, G.; and Timofte, R. 2024. High-quality image restoration following human instructions. In *ECCV*, 1–21.

Cui, Y.; Zamir, S. W.; Khan, S.; Knoll, A.; Shah, M.; and Khan, F. S. 2025. Adair: Adaptive all-in-one image restoration via frequency mining and modulation. In *ICLR*, 57335–57356.

Fan, Q.; Chen, D.; Yuan, L.; Hua, G.; Yu, N.; and Chen, B. 2019. A general decoupled learning framework for parameterized image operators. *IEEE TPAMI*, 43(1): 33–47.

Gou, Y.; Li, B.; Liu, Z.; Yang, S.; and Peng, X. 2020. Clearer: Multi-scale neural architecture search for image restoration. In *NeurIPS*, 17129–17140.

Gou, Y.; Zhao, H.; Li, B.; Xiao, X.; and Peng, X. 2024. Test-Time Degradation Adaptation for Open-Set Image Restoration. In *ICML*, 16167–16177.

Guo, Y.; Gao, Y.; Lu, Y.; Zhu, H.; Liu, R. W.; and He, S. 2024. Onerestore: A universal restoration framework for composite degradation. In *ECCV*, 255–272.

Jiang, K.; Wang, Z.; Yi, P.; Chen, C.; Huang, B.; Luo, Y.; Ma, J.; and Jiang, J. 2020. Multi-scale progressive fusion network for single image deraining. In *CVPR*, 8346–8355.

Kong, L.; Dong, J.; Ge, J.; Li, M.; and Pan, J. 2023. Efficient frequency domain-based transformers for high-quality image deblurring. In *CVPR*, 5886–5895.

Li, B.; Liu, X.; Hu, P.; Wu, Z.; Lv, J.; and Peng, X. 2022. All-in-one image restoration for unknown corruption. In *CVPR*, 17452–17462.

Li, H.; Chen, X.; Dong, J.; Tang, J.; and Pan, J. 2025a. Foundir: Unleashing million-scale training data to advance foundation models for image restoration. In *ICCV*, 12626–12636.

Li, R.; Tan, R. T.; and Cheong, L.-F. 2020. All in one bad weather removal using architectural search. In *CVPR*, 3175–3185.

Li, W.; Han, W.; Deng, L.-J.; Xiong, R.; and Fan, X. 2025b. Spiking Variational Graph Representation Inference for Video Summarization. *IEEE TIP*, 34: 5697–5709.

Li, W.; Wang, P.; Xiong, R.; and Fan, X. 2024a. Spiking Tucker Fusion Transformer for Audio-Visual Zero-Shot Learning. *IEEE TIP*, 33: 4840–4852.

Li, Z.; Li, J.; Li, Y.; Li, L.; Liu, D.; and Wu, F. 2024b. In-loop filtering via trained look-up tables. In *VCIP*, 1–5.

Li, Z.; Yuan, Z.; Li, L.; Liu, D.; Tang, X.; and Wu, F. 2024c. Object Segmentation-Assisted Inter Prediction for Versatile Video Coding. *IEEE TBC*, 70(4): 1236–1253.

Luo, Z.; Gustafsson, F. K.; Zhao, Z.; Sjölund, J.; and Schön, T. B. 2023. Controlling vision-language models for universal image restoration. In *ICLR*, 16226–16246.

Ma, J.; Cheng, T.; Wang, G.; Zhang, Q.; Wang, X.; and Zhang, L. 2023. ProRes: Exploring Degradation-aware Visual Prompt for Universal Image Restoration. *arXiv preprint arXiv:2306.13653*.

Ma, J.; Hu, S.; Zhang, X.; Wan, J.; Huang, J.; Zhang, L.; and Khan, S. 2025. EvoIR: Towards All-in-One Image Restoration via Evolutionary Frequency Modulation. *arXiv preprint arXiv:2512.05104*.

Nah, S.; Hyun Kim, T.; and Mu Lee, K. 2017. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, 3883–3891.

Özdenizci, O.; and Legenstein, R. 2023. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE TPAMI*, 45(8): 10346–10357.

Potlapalli, V.; Zamir, S. W.; Khan, S.; and Khan, F. S. 2023. PromptIR: Prompting for All-in-One Blind Image Restoration. In *NeurIPS*, 71275–71293.

Ravi, N.; Gabeur, V.; Hu, Y.-T.; Hu, R.; Ryali, C.; Ma, T.; Khedr, H.; Rädle, R.; Rolland, C.; Gustafson, L.; et al. 2024. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*.

Rong, F.; Lan, M.; Zhang, Q.; and Zhang, L. 2025. MPG-SAM 2: Adapting SAM 2 with Mask Priors and Global Context for Referring Video Object Segmentation. *arXiv preprint arXiv:2501.13667*.

Song, Y.; He, Z.; Qian, H.; and Du, X. 2023. Vision Transformers for Single Image Dehazing. *IEEE TIP*, 32: 1927–1941.

Sun, S.; Ren, W.; Gao, X.; Wang, R.; and Cao, X. 2024. Restoring images in adverse weather conditions via histogram transformer. In *ECCV*, 111–129.

Valanarasu, J. M. J.; Yasarla, R.; and Patel, V. M. 2022. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *CVPR*, 2353–2363.

Wang, T.; Zhang, K.; Shao, Z.; Luo, W.; Stenger, B.; Lu, T.; Kim, T.-K.; Liu, W.; and Li, H. 2024. Gridformer: Residual dense transformer with grid structure for image restoration in adverse weather conditions. *IJCV*, 132(10): 4541–4563.

Wei, C.; Wang, W.; Yang, W.; and Liu, J. 2018. Deep retinex decomposition for low-light enhancement. In *BMVC*.

Wu, G.; Jiang, J.; Jiang, K.; and Liu, X. 2024a. Learning from history: Task-agnostic model contrastive learning for image restoration. In *AAAI*, 5976–5984.

Wu, G.; Jiang, J.; Jiang, K.; Liu, X.; and Nie, L. 2025a. Learning Dynamic Prompts for All-in-One Image Restoration. *IEEE TIP*, 34: 3997–4010.

Wu, G.; Jiang, J.; Wang, Y.; Jiang, K.; and Liu, X. 2025b. Debiased All-in-one Image Restoration with Task Uncertainty Regularization. In *AAAI*, 8386–8394.

Wu, H.; Zhang, Z.; Zhang, E.; Chen, C.; Liao, L.; Wang, A.; Li, C.; Sun, W.; Yan, Q.; Zhai, G.; et al. 2024b. Q-bench: A benchmark for general-purpose foundation models on low-level vision. In *ICLR*, 12547–12573.

Wu, H.; Zhang, Z.; Zhang, E.; Chen, C.; Liao, L.; Wang, A.; Xu, K.; Li, C.; Hou, J.; Zhai, G.; et al. 2024c. Q-instruct: Improving low-level visual abilities for multi-modality foundation models. In *CVPR*, 25490–25500.

Xia, W.; De Charette, R.; Oztireli, C.; and Xue, J.-H. 2024. Dream: Visual decoding from reversing human visual system. In *WACV*, 8226–8235.

Xiao, Y.; Yuan, Q.; Jiang, K.; Chen, Y.; Wang, S.; and Lin, C.-W. 2025a. Multi-Axis Feature Diversity Enhancement for Remote Sensing Video Super-Resolution. *IEEE TIP*, 34: 1766–1778.

Xiao, Y.; Yuan, Q.; Jiang, K.; Huang, W.; Zhang, Q.; Zheng, T.; Lin, C.-W.; and Zhang, L. 2025b. Spiking Meets Attention: Efficient Remote Sensing Image Super-Resolution with Attention Spiking Neural Networks. *arXiv preprint arXiv:2503.04223*.

Xiao, Z.; Kai, D.; Zhang, Y.; Zha, Z.-J.; Sun, X.; and Xiong, Z. 2024. Event-adapted video super-resolution. In *ECCV*, 217–235.

Xiao, Z.; and Wang, X. 2025. Event-based Video Super-Resolution via State Space Models. In *CVPR*, 12564–12574.

Yang, W.; Tan, R. T.; Feng, J.; Liu, J.; Guo, Z.; and Yan, S. 2017. Deep joint rain detection and removal from a single image. In *CVPR*, 1357–1366.

Yao, M.; Xu, R.; Guan, Y.; Huang, J.; and Xiong, Z. 2024. Neural Degradation Representation Learning for All-in-One Image Restoration. *IEEE TIP*, 33: 5408–5423.

Ye, T.; Chen, S.; Bai, J.; Shi, J.; Xue, C.; Jiang, J.; Yin, J.; Chen, E.; and Liu, Y. 2023. Adverse weather removal with codebook priors. In *ICCV*, 12653–12664.

You, Z.; Cai, X.; Gu, J.; Xue, T.; and Dong, C. 2025. Teaching large language models to regress accurate image quality scores using score distribution. In *CVPR*, 14483–14494.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M.-H. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, 5728–5739.

Zeng, H.; Wang, X.; Chen, Y.; Su, J.; and Liu, J. 2025. Vision-Language Gradient Descent-driven All-in-One Deep Unfolding Networks. In *CVPR*, 7524–7533.

Zhang, H.; Zhang, X.; Cai, N.; Di, J.; and Zhang, Y. 2025a. Joint multi-dimensional dynamic attention and transformer for general image restoration. *CVIU*, 261: 104491.

Zhang, H.; Zhang, X.; Zhu, L.; Zhang, Y.; Cao, J.; and Ling, W.-K. 2025b. Enhancing 3D video watching experiences: Tackling compression and 3D warping distortions in synthesized view with perceptual guidance. *ESWA*, 264: 125853.

Zhang, J.; Huang, J.; Yao, M.; Yang, Z.; Yu, H.; Zhou, M.; and Zhao, F. 2023a. Ingredient-Oriented Multi-Degradation Learning for Image Restoration. In *CVPR*, 5825–5835.

Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; and Zhang, L. 2017. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE TIP*, 26(7): 3142–3155.

Zhang, K.; Zuo, W.; and Zhang, L. 2018. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE TIP*, 27(9): 4608–4622.

Zhang, X.; Cai, N.; Zhang, H.; Zhang, Y.; Di, J.; and Lin, W. 2023b. AFD-Former: A Hybrid Transformer With Asymmetric Flow Division for Synthesized View Quality Enhancement. *IEEE TCSVT*, 33(8): 3786–3798.

Zhang, X.; Ma, J.; Wang, G.; Zhang, Q.; Zhang, H.; and Zhang, L. 2025c. Perceive-ir: Learning to perceive degradation better for all-in-one image restoration. *IEEE TIP*.

Zhang, X.; Zhang, H.; Wang, G.; Zhang, Q.; Zhang, L.; and Du, B. 2025d. UniUIR: Considering Underwater Image Restoration as an All-in-One Learner. *IEEE TIP*, 34: 6963–6977.

Zhao, R.; Xiong, R.; Zhao, J.; Zhang, J.; Fan, X.; Yu, Z.; and Huang, T. 2024. Boosting spike camera image reconstruction from a perspective of dealing with spike fluctuations. In *CVPR*, 24955–24965.

Zhu, Y.; Wang, T.; Fu, X.; Yang, X.; Guo, X.; Dai, J.; Qiao, Y.; and Hu, X. 2023. Learning Weather-General and Weather-Specific Features for Image Restoration Under Multiple Adverse Weather Conditions. In *CVPR*, 21747–21758.