Gradient as Conditions: Rethinking HOG for All-in-one Image Restoration

Jiawei Wu[1], Zhifei Yang[2], Zhe Wang[1], Zhi Jin[1,3*],

[1]School of Intelligent Systems Engineering, Shenzhen Campus of Sun Yat-sen University, Guangdong, China.
[2]School of Computer Science, Peking University, China.
[3]Guangdong Provincial Key Laboratory of Fire Science and Intelligent Emergency Technology, Shenzhen, China.
wujw97@mail2.sysu.edu.cn, jinzh26@mail.sysu.edu.cn

Abstract

All-in-one image restoration (AIR) aims to address diverse degradations within a unified model by leveraging informative degradation conditions to guide the restoration process. However, existing methods often rely on implicitly learned priors, which may entangle feature representations and hinder performance in complex or unseen scenarios. Histogram of Oriented Gradients (HOG) as a classical gradient representation, we observe that it has strong discriminative capability across diverse degradations, making it a powerful and interpretable prior for AIR. Based on this insight, we propose HOGformer, a Transformer-based model that integrates learnable HOG features for degradation-aware restoration. The core of HOGformer is a Dynamic HOG-aware Self-Attention (DHOGSA) mechanism, which adaptively models long-range spatial dependencies conditioned on degradation-specific cues encoded by HOG descriptors. To further adapt the heterogeneity of degradations in AIR, we propose a Dynamic Interaction Feed-Forward (DIFF) module that facilitates channel–spatial interactions, enabling robust feature transformation under diverse degradations. Besides, we propose a HOG loss to explicitly enhance structural fidelity and edge sharpness. Extensive experiments on a variety of benchmarks, including adverse weather and natural degradations, demonstrate that HOGformer achieves state-of-the-art performance and generalizes well to complex real-world scenarios.

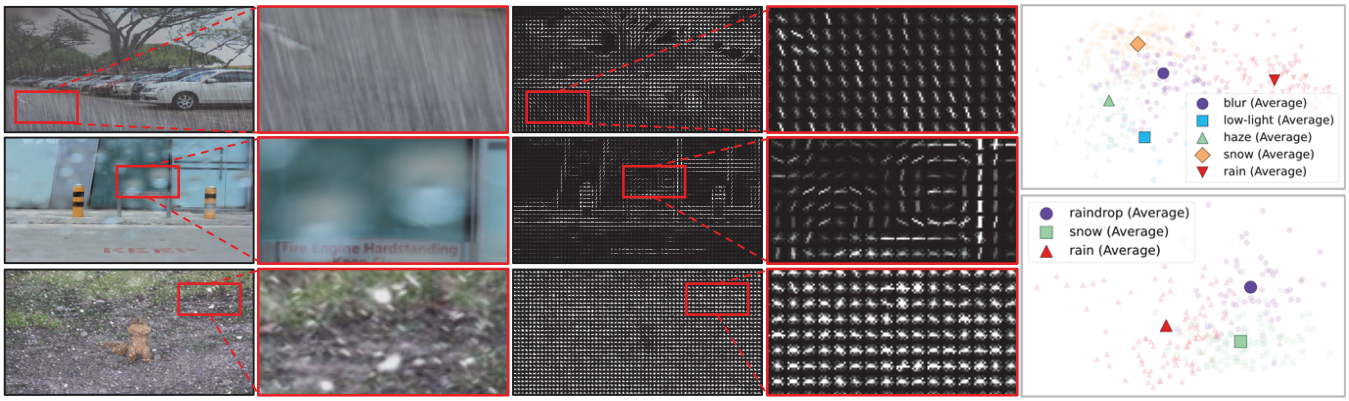
Introduction

Real-world images are frequently affected by diverse degradations such as blur, low-light, and adverse weather. While task-specific deep learning models have achieved remarkable progress in restoring images under individual degradations (He and Jin 2024; Wang, Wu, and Jin 2023; Jin et al. 2025), they often require separate models, leading to high training and deployment overhead. AIR (Jiang et al. 2024a) offers a practical alternative by handling diverse degradations in a single model, which is critical for real-world applications like autonomous driving (Chen et al. 2024).

A key challenge in AIR is designing a model that can adaptively respond to the specific degradation present in each input. To address this, recent works (Li et al. 2022; Potlapalli et al. 2023; Ai et al. 2024; Jiang et al. 2024b) have proposed shared backbone architectures modulated by degradation-aware conditional features, enabling a unified framework to handle diverse tasks. Among them, prompt-based methods such as PromptIR (Potlapalli et al. 2023) inject degradation-specific prompts to guide the network behavior, while others exploit multimodal inputs (Ai et al. 2024) to capture complementary degradation cues. Despite their promise, the success of these methods critically depends on the quality and reliability of the conditioning. In practice, such signals are often derived from implicitly learned priors, which operate as black boxes with limited interpretability. This lack of transparency makes it difficult to disentangle the underlying degradations, especially when they exhibit subtle or overlapping characteristics. Consequently, feature representations optimized for one degradation (e.g., denoising) may conflict with those required for another (e.g., dehazing), leading to feature entanglement and reduced generalization performance across tasks (Li et al. 2020). While alternative methods based on Mixture-of-Experts (MoE) (Zamfir et al. 2025) offer greater flexibility through expert routing, they come with substantial computational overhead, making them less suitable for efficient deployment. These limitations underscore the requirement for a conditional mechanism that is not only explicit and discriminative, but also computationally efficient and robust to diverse degradations.

In this work, we revisit classical feature descriptors and identify the HOG (Dalal and Triggs 2005) features as a surprisingly effective and interpretable prior for AIR. As shown in Figure 1, we observe that the gradient patterns captured by HOG form distinct signatures for different degradations. For example, in the case of rain (Figure 1(a)), falling rain produces vertical streaks that are reflected as strong vertical gradients, while raindrop regions appear as isolated circular patches with low gradient magnitudes. In contrast, snow degradation introduces widespread high-magnitude gradients with low directional variance, forming dense and uniform textures. This discriminative behavior persists across both natural and adverse weather conditions (Figure 1(b)). The effectiveness of HOG stems from its two complementary components, i.e., gradient magnitude and orientation, which explicitly encode local intensity variations and directional structure. These observations raise a central question: *Can the inherent degradation-discriminative property*

*Corresponding author.



(a) (b)

Figure 1: Visualization of HOG feature distributions under various degradations. (a) Example images of different weather conditions with corresponding HOG feature visualizations. (b) HOG features for five natural degradations (Zheng et al. 2024) and three adverse weather degradations (Sun et al. 2024), using 100 randomly selected images for each degradation.

of HOG be leveraged to guide the design of AIR networks, enabling them to explicitly capture degradation-specific gradient patterns for more efficient AIR?

To this end, we propose HOGformer, an AIR framework that integrates HOG as gradient priors in a principled and degradation-aware manner. Instead of incorporating the static handcrafted HOG feature, we reformulate it as a learnable and differentiable module that provides dynamic and context-aware guidance throughout the network. At the core of HOGformer lies the Dynamic HOG-aware Self-Attention (DHOGSA) mechanism, which integrates HOG-derived cues into the self-attention process. This design enables the network to focus on degradation-specific patterns, such as rain streaks or haze-induced texture loss, by explicitly modulating attention maps based on gradient information. To further enhance the adaptability of feature representations, we introduce a Dynamic Interaction Feed-Forward (DIFF) module that facilitates effective channel–spatial interaction, allowing the model to robustly transform features under diverse and heterogeneous degradations. In addition, we propose a dedicated HOG supervision loss that directly constrains the reconstruction of gradient magnitude and orientation. This explicit guidance promotes sharper edges and better preservation of structural details. This unified design enables effective restoration across diverse degradations, achieving state-of-the-art (SOTA) performance on both adverse weather and natural image AIR tasks. Our main contributions are summarized as follows:

- We identify that classical HOG features serve as an explicit and highly discriminative prior for distinguishing diverse degradations in AIR, offering a compelling alternative to implicit conditional mechanisms.
- We propose HOGformer, an AIR network that embeds learnable HOG cues into core network components for degradation-aware adaptation. HOGformer integrates a dynamic HOG-aware self-attention (DHOGSA) mechanism, an efficient dynamic interaction feed-forward (DIFF) module, and a dedicated HOG supervision loss.
- Extensive experiments show that HOGformer achieves state-of-the-art performance across various degradations, including adverse weather and natural conditions.

Related Works

Single-Task Image Restoration

Image restoration is a fundamental problem in computer vision. Traditional methods constrain the solution space using human priors and handcrafted features (Banham and Katsaggelos 1997). With deep learning, various approaches have achieved strong performance across tasks (He and Jin 2024; Wang, Wu, and Jin 2023; Qiu et al. 2023; Zamir et al. 2022a; Guo et al. 2024). Vision Transformers further improve restoration by modeling long-range dependencies (Liang et al. 2021; Zamir et al. 2022a; Wang et al. 2022a). For example, Swin-IR (Liang et al. 2021) adopts a shifted window strategy for local-global context modeling, while Restormer (Zamir et al. 2022a) introduces a multi-stage attention design to balance accuracy and efficiency. Transformer-based models have also been applied to deraining (Chen et al. 2023), desnowing (Chen et al. 2022b), dehazing (Song et al. 2023; Zhang et al. 2024), deblurring (Liang et al. 2024; Kong et al. 2023), and low-light enhancement (Cai et al. 2023). However, these methods require separate training, increasing computational and deployment costs (Sun et al. 2024).

All-in-One Image Restoration

AIR aims to handle diverse degradations using a single model without task-specific retraining (Jiang et al. 2024a; Li et al. 2022; Sun et al. 2024; Zheng et al. 2024). Typically, InstructIR (Conde, Geigle, and Timofte 2024) employs natural language instructions to specify restoration goals but incurs high data preparation costs. Methods such as Painter (Wang et al. 2023) and DA-CLIP (Luo et al. 2024) leverage on-the-fly learning to adapt large models, while DiffUIR (Zheng et al. 2024) builds on residual diffusion to address diverse degradations. PromptIR (Potlapalli

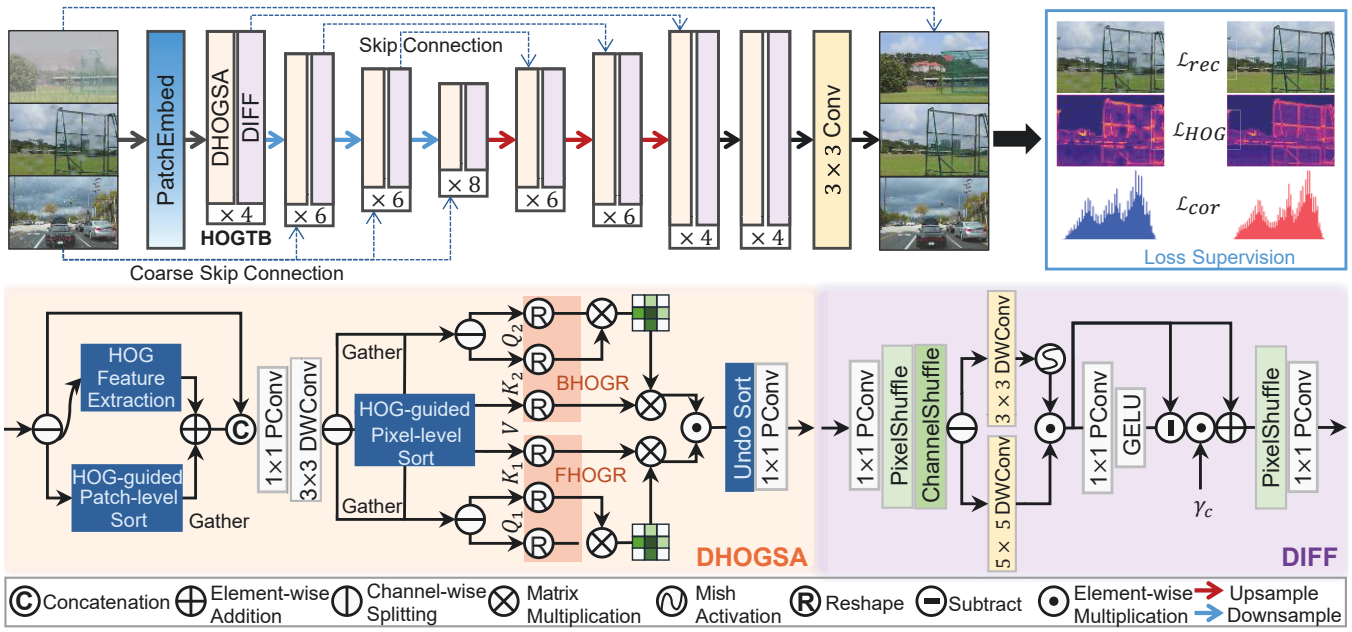


Figure 2: The overall architecture of our HOGformer. It includes the HOG Transformer block with the Dynamic HOG-aware Self-Attention (DHOGSA) module and Dynamic Interaction Feed-Forward (DIFF) module.

et al. 2023) encodes degradation conditions into prompts to guide restoration. These approaches rely on implicitly constructed degradation-specific conditions to enable flexible inference. However, the construction of such conditions is often complex and lacks consistent priors. Recently, Histoformer (Sun et al. 2024) incorporates grayscale histogram priors into Transformer models and shows strong performance on adverse weather restoration. However, histogram-based features offer limited discriminative power; for example, blurry and clean images can share similar histograms. In this work, we observe that HOG offers a clearer way to distinguish between different degradations. Our method utilizes HOG features to provide explicit guidance that is aware of the degradation throughout the restoration process.

Method

Overall Architecture

Motivated by the observation (Figure 1) that the HOG effectively captures degradation-specific patterns, we propose HOGformer, an AIR network that explicitly incorporates HOG-derived cues into the restoration process. As shown in Figure 2, HOGformer employs a multi-stage U-shaped encoder-decoder architecture to progressively model and restore diverse degradations within a single model.

The encoder begins by applying a 3×3 convolution to extract shallow features from the input image I^{lq}. These features are successively downsampled and enriched through stacked HOG Transformer Blocks (HOGTBs), which serve as the core feature extractors. To retain low-frequency priors and facilitate residual learning, coarse skip connections (Sun et al. 2024) are introduced to highlight degradation residuals, incorporating average pooling, pointwise convolution, and depthwise convolution. Downsampling and upsampling are achieved using pixel-unshuffle and pixel-shuffle, respectively, which ensure efficient resolution transitions while preserving structural consistency. Additionally, symmetric skip connections between the encoder and decoder further enhance information flow and detail recovery. The core of the HOGformer lies the HOGTBs, which incorporate dynamic HOG-aware self-attention (**DHOGSA**) and dynamic interaction feed-forward (**DIFF**) as illustrated in Figure 2. DHOGSA utilizes gradient information to adjust attention through HOG-guided sorting at both pixel- and patch-level. This facilitates selective focus on regions sensitive to degradation regions. DIFF facilitates effective channel–spatial interaction to enable robust transformation of features under diverse degradations. Each HOGTB employs a residual paradigm to ensure stable optimization as follows:

$$\begin{aligned} \mathbf{F}'_l &= \mathbf{F}_{l-1} + \text{DHOGSA}(\text{LN}(\mathbf{F}_{l-1})), \\ \mathbf{F}_l &= \mathbf{F}'_l + \text{DIFF}(\text{LN}(\mathbf{F}'_l)), \end{aligned} \tag{1}$$

where F_l denotes the output of the l-th layer, LN is the layer normalization. This structured integration of HOG-guided attention and interaction enables efficient AIR.

Dynamic HOG-Aware Self-Attention (DHOGSA)

The first core module in each HOGTB is DHOGSA, designed to capture long-range and degradation-specific dependencies using gradient-domain priors. Traditional self-attention mechanisms (Zamir et al. 2022a) typically operate within fixed windows or channel-wise structures, making them suboptimal for modeling the non-uniform and spatially varying patterns present in diverse degradations. DHOGSA overcomes this by leveraging differentiable HOG descrip-

tors to explicitly sort features based on their gradient magnitudes and orientations before computing attention. This sorting operation is conducted at both patch and pixel levels, grouping spatial locations with similar degradation patterns to facilitate more effective attention computation.

Local Dynamic-Range Convolution (LDRConv). To facilitate effective long-range modeling, we begin by enhancing local degradation structures through a Local Dynamic-Range Convolution (LDRConv) module. Unlike standard convolutions with fixed receptive fields, LDRConv dynamically reorganizes features based on their gradient distribution. Specifically, LDRConv performs HOG-guided patch-level sorting, followed by modulation using learnable bin-wise HOG priors. This patch-wise operation groups regions with similar gradient properties while preserving global spatial consistency, in contrast to pixel-wise sorting which may disrupt the overall structure. This design ensures that the model retains critical semantic and geometric information, which is essential for image restoration tasks.

Given input features $\mathbf{F} \in \mathbb{R}^{C\times H\times W}$, we first compute the gradient magnitude m and orientation o using Sobel filters:

$$\mathbf{m} = \sqrt{\mathbf{g_x}^2 + \mathbf{g_y}^2}, o = \left\lfloor \frac{\text{atan2}(\mathbf{g_y}, \mathbf{g_x}) + \pi}{2\pi} N_{\text{bin}} \right\rfloor, \quad (2)$$

where $\mathbf{g_x}$ and $\mathbf{g_y}$ are obtained directional gradients using Sobel operation, N_{bin} is the number of HOG bins. To preserving the original representation capacity, we split $\mathbf{F}$ into $\mathbf{F}_1, \mathbf{F}_2$ and only apply HOG-guided patch-wise sorting (Figure 3(a)) to $\mathbf{F}_1$, followed by modulation with learnable HOG priors $\text{HOG}_\theta(\mathbf{F}_1)$ to enhance local gradient awareness under diverse degradations:

$$\begin{aligned} \mathbf{F}_1 &= \text{Sort}_{\text{patch}}(\mathbf{F}_1) + \text{HOG}_\theta(\mathbf{F}_1), \\ \mathbf{F} &= \text{Conv}^{\text{d}}_{3\times3}(\text{Conv}^{\text{p}}_{1\times1}(\text{Concat}(\mathbf{F}_1, \mathbf{F}_2))). \end{aligned} \quad (3)$$

As shown in Figure 3(c), $\text{HOG}_\theta(\mathbf{F}_1)$ is derived by computing gradient-based histograms over local patches, followed by feature projection. LDRConv enhances degradation sensitivity while maintaining structural coherence, providing a strong basis for the subsequent self-attention mechanism.

HOG-Guided Self-Attention. With the enhanced representations from LDRConv, we effectively model the crucial long-range dependencies. Specifically, inspired by the observation that distant pixels affected by the same degradation often exhibit similar HOG responses, we sort pixels (Figure 3(b)) based on HOG descriptors ($\mathbf{m} \cdot \mathbf{o}$). To spatially align features, we generate sorting indices from the information-rich value features $\mathbf{V}$, and then use these indices to consistently sort the query ($\mathbf{Q}$), key ($\mathbf{K}$), and value ($\mathbf{V}$) features. To capture multi-scale dependencies, we introduce two parallel attention branches with distinct histogram reshaping strategies. Bin-wise Histogram Reshaping (BHOGR) groups sorted pixels into fixed-size bins to extract coarse, large-scale features. Frequency-wise Histogram Reshaping (FHOGR) clusters pixels with similar HOG values to capture fine textures and local variations. BHOGR focuses on capturing large-scale structural artifacts, such as the spatially uniform distribution of haze, whereas FHOGR is tailored to detect fine-scale, repetitive degradations like rain streaks. Fusing both branches enables a comprehensive representation $\mathbf{F}$. The entire process is as follows:

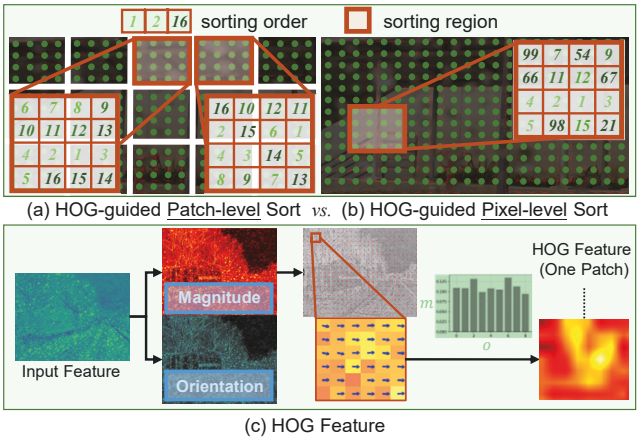


Figure 3: HOG-guided mechanism: (a) Patch-level sorting. (b) Pixel-level sorting. (c) HOG feature extraction process.

$$\begin{aligned} \mathbf{idx} &= \text{Sort}\Big(\mathcal{R}^{C\times HW}_{C\times H\times W}(o(\mathbf{V}) \cdot m(\mathbf{V}))\Big), \\ \mathbf{Q}_{\{\cdot\}\in\{B,F\}} &= \text{Gather}\Big(\mathcal{R}^{C\times HW}_{C\times H\times W}(\mathbf{Q}_{\{\cdot\}}), \mathbf{idx}\Big), \\ \mathbf{K}_{\{\cdot\}\in\{B,F\}} &= \text{Gather}\Big(\mathcal{R}^{C\times HW}_{C\times H\times W}(\mathbf{K}_{\{\cdot\}}), \mathbf{idx}\Big), \\ \mathbf{V} &= \text{Gather}\Big(\mathcal{R}^{C\times HW}_{C\times H\times W}(\mathbf{V}), \mathbf{idx}\Big), \\ \mathbf{A}_{\{\cdot\}\in\{B,F\}} &= \text{Softmax}\left(\frac{\mathcal{R}_{\{\cdot\}}(Q_{\{\cdot\}})\mathcal{R}_{\{\cdot\}}(K_{\{\cdot\}})^\top}{\sqrt{\mathbf{K}}}\right), \\ \mathbf{F} &= \mathbf{A_B}\mathcal{R}_B(\mathbf{V}) \odot \mathbf{A_F}\mathcal{R}_F(\mathbf{V}). \end{aligned} \quad (4)$$

where $\mathbf{K}$ is the number of heads, $\mathcal{R}_{i\in\{B,F\}}$ represents the reshaping operation (BHOGR or FHOGR), and $\odot$ is the Hadamard product.

Dynamic Interaction Feed-Forward (DIFF)

After DHOGSA aggregates spatial features guided by HOG cues, the resulting representations are fed into the proposed Dynamic Interaction Feed-Forward (DIFF) network. Unlike standard, content-agnostic FFNs, DIFF performs dynamic spatial-channel interaction to adaptively refine features based on the content, so that to enhance heavily degraded regions while preserving clean areas. Three complementary designs make it possible: firstly, multi-scale spatial context via parallel branches is captured; secondly, a pixel-wise gating mechanism modulates spatial responses conditioned on input features; thirdly, channel shuffling and aggregation promote cross-channel communication. Therefore, DIFF can achieve content-aware refinement across spatial and channel dimensions.

Loss Supervision

To fully exploit the HOG-guided attention and dynamic feature interaction in HOGformer, we design a tailored op-

Method	Snow100K-S P ↑	Snow100K-S S ↑	Snow100K-L P ↑	Snow100K-L S ↑	Method	Outdoor-Rain P ↑	Outdoor-Rain S ↑	Method	RainDrop P ↑	RainDrop S ↑	Average P ↑	Average S ↑
RESCAN	31.51	0.9032	26.08	0.8108	HRGAN	21.56	0.8550	RaindropAttn	31.44	0.9263	-	-
SnowGAN	32.33	0.9500	27.17	0.8983	PCNet	26.19	0.9015	AttentiveGAN	31.59	0.9274	-	-
DDMSNet	34.34	0.9445	28.85	0.8772	MPRNet	28.03	0.9192	IDT	31.87	0.9313	-	-
DTANet	34.79	0.9497	30.06	0.9017	NAFNet	29.59	0.9027	MAXIM	31.87	0.9352	-	-
Restormer	36.01	0.9579	30.36	0.9068	Restormer	30.03	0.9215	Restormer	32.18	0.9408	-	-
All-in-One	-	-	28.33	0.8820	All-in-One	24.71	0.8980	All-in-One	31.12	0.9268	-	-
TransWeather	32.51	0.9341	29.31	0.8879	TransWeather	28.83	0.9000	TransWeather	30.41	0.9157	30.27	0.9094
Chen et al.	34.42	0.9469	30.22	0.9071	Chen et al.	29.27	0.9147	Chen et al.	31.81	0.9309	31.43	0.9249
WGWSNet	34.31	0.9460	30.16	0.9007	WGWSNet	29.32	0.9207	WGWSNet	32.38	0.9378	31.54	0.9263
WeatherDiff$_{64}$	35.83	0.9566	30.09	0.9041	WeatherDiff$_{64}$	29.64	0.9312	WeatherDiff$_{64}$	30.71	0.9312	31.57	0.9308
WeatherDiff$_{128}$	35.02	0.9516	29.58	0.8941	WeatherDiff$_{128}$	29.72	0.9216	WeatherDiff$_{128}$	29.66	0.9225	30.99	0.9225
AWRCP	36.92	0.9652	31.92	**0.9344**	AWRCP	31.39	0.9329	AWRCP	31.93	0.9314	33.04	0.9409
GridFormer	37.46	0.9640	31.71	0.9231	GridFormer	31.87	0.9335	GridFormer	32.39	0.9362	33.36	0.9392
MoCE-IR	37.10	0.9654	31.88	0.9234	MoCE-IR	31.42	0.9334	MoCE-IR	32.23	0.9386	33.16	0.9400
Histoformer	37.41	0.9656	32.16	0.9261	Histoformer	32.08	0.9389	Histoformer	**33.06**	0.9441	33.67	0.9436
HOGformer-L	**37.93**	**0.9685**	**32.41**	0.9297	**HOGformer-L**	**32.89**	**0.9460**	**HOGformer-L**	32.72	**0.9452**	**33.99**	**0.9474**

Table 1: Quantitative comparisons of adverse weather restoration. The top half of the tables shows results from task-specific methods, while the bottom half displays evaluations of AIR methods. The **best** and second-best results are highlighted.

Figure 4: Visual comparison for deraining (Sun et al. 2024; Özdenizci and Legenstein 2023). Zoom in for the best visualization.

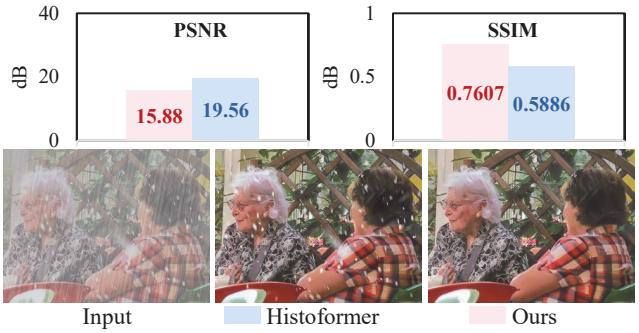


Figure 5: Comparisons on multiple degradations (rain-fog-snow). **Top:** PSNR and SSIM comparisons between Histformer and our method. **Bottom:** Visual example.

timization objective. While pixel-wise losses such as the L_1 loss ensure reconstruction accuracy, they often overlook structural and textural fidelity, resulting in overly smooth outputs. To address this limitation, we formulate a composite loss that aligns with the architectural design:

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \alpha\mathcal{L}_{\text{cor}} + \beta\mathcal{L}_{\text{HOG}}, \tag{5}$$

where α and β are weighting factors. $\mathcal{L}_{\text{rec}}$ denotes the L_1 loss for pixel-level accuracy, $\mathcal{L}_{\text{cor}}$ promotes global photometric consistency via Pearson correlation (Sun et al. 2024), and $\mathcal{L}_{\text{HOG}}$ enforces structural alignment by minimizing the L_2 distance of HOG features between the output and GT.

Experiments

Experimental Setup

We evaluate HOGformer in two standard AIR settings: (I) 3-task adverse weather removal (Sun et al. 2024), including desnowing, draining&dehazing, and raindrop removal; and (II) 5-task general restoration (Zheng et al. 2024), including deraining, low-light enhancement, desnowing, dehazing, and deblurring. One AIR model is trained for each setting.

Datasets and Metrics. For Setting (I), we use the All-Weather dataset (Sun et al. 2024). For Setting (II), we use task-specific datasets: Merged rain dataset (Jiang et al. 2020; Wang et al. 2022b) for deraining, LOL (Wei et al. 2018) for low-light, Snow100K (Liu et al. 2018) for desnowing, RESIDE (Li et al. 2018a) for dehazing, and GoPro (Nah, Hyun Kim, and Mu Lee 2017) for deblurring. To evaluate generalization, we additionally test on real-world datasets: Practical (Yang et al. 2017), MEF (Ma, Zeng, and Wang 2015), NPE (Wang et al. 2013), DICM (Lee, Lee, and Kim 2013), HIDE (Shen et al. 2019), RealBlur (Rim et al. 2020), and T-OLED (Zhou et al. 2021).

For evaluation, we report PSNR (**P**↑) and SSIM (**S**↑) for standard metrics, LPIPS (**L**↓) (Zhang et al. 2018) for perceptual similarity, and NIQE (**N**↓) (Mittal, Soundararajan, and Bovik 2012) for no-reference assessment.

Implementation Details

We implement our model using PyTorch and conduct experiments on NVIDIA Tesla A100 GPUs. We apply random

Method	Rain (5 sets)		Low-light		Snow (2 sets)		Haze		Blur		Complexity	
	P ↑	S ↑	P ↑	S ↑	P ↑	S ↑	P ↑	S ↑	P ↑	S ↑	Params (M)	FLOPs (G)
SwinIR	30.78	0.923	17.81	0.723	-	-	21.5	0.891	24.52	0.773	**0.90**	752.13
MIRNet-v2	33.89	0.924	**24.74**	0.851	-	-	24.03	0.927	26.30	0.799	5.90	140.92
Restormer	**33.96**	**0.935**	20.41	0.806	-	-	30.87	0.969	**32.92**	**0.961**	26.12	141.00
MAXIM	33.24	0.933	23.43	0.863	-	-	34.19	**0.985**	32.86	0.940	14.1	216.00
IR-SDE	-	-	-	-	20.45	0.787	-	-	30.70	0.901	34.20	98.30
WeatherDiff	-	-	-	-	**33.51**	**0.939**	-	-	-	-	82.96	-
RDDM	30.74	0.903	23.22	**0.899**	32.55	0.927	30.78	0.953	29.53	0.876	36.26	**9.88**
AirNet	24.87	0.773	14.83	0.767	27.63	0.860	25.47	0.923	26.92	0.811	8.93	30.13
Painter	29.49	0.868	22.40	0.872	-	-	-	-	-	-	307.00	248.90
ProRes	30.67	0.891	22.73	0.877	-	-	32.02	0.952	27.53	0.851	307.00	248.90
Prompt-IR	29.56	0.888	22.89	0.847	-	-	32.02	0.952	27.21	0.817	35.59	15.81
DA-CLIP	28.96	0.853	24.17	0.882	30.80	0.888	31.39	0.983	25.39	0.805	174.10	118.50
DiffUIR-S	30.25	0.893	23.52	0.895	31.45	0.915	31.83	0.954	27.79	0.830	3.27	**2.40**
DiffUIR-L	31.03	0.904	25.12	0.907	32.65	0.927	32.94	0.956	29.17	0.864	36.26	9.88
HOGformer-S	30.75	0.901	25.36	0.915	32.72	0.929	33.67	0.991	28.37	0.840	**2.91**	20.63
HOGformer-L	**31.63**	**0.914**	**25.57**	**0.917**	**34.08**	**0.941**	**36.60**	**0.994**	**29.95**	**0.884**	16.64	91.77

Table 2: Quantitative results of all-in-one image restoration methods in five tasks. The **best** and second-best results are highlighted. The top is task-specific restoration methods, and the bottom is all-in-one restoration methods.

Method	Rain	Low-light	Snow	Blur	
	N ↓	N ↓	N ↓	P ↑	S ↑
WeatherDiff	-	-	2.96	-	-
CLIP-LIT	-	3.70	-	-	-
Restormer	3.50	3.80	-	**32.12**	**0.926**
RDDM	**3.34**	**3.57**	**2.76**	30.74	0.894
AirNet	3.55	3.45	2.75	16.78	0.628
Prompt-IR	3.52	3.31	2.79	22.48	0.770
DA-CLIP	3.42	3.56	2.72	17.51	0.667
DiffUIR-L	3.38	3.14	2.74	30.63	0.890
HOGformer-L	**3.31**	**3.08**	**2.69**	**30.92**	**0.907**

Table 3: Quantitative results of known task generalization. The **best** and second-best results are highlighted. The top is task-specific methods, and the bottom is AIR methods.

Method	P ↑	S ↑	L ↓
NAFNet	**26.89**	0.774	**0.346**
MPRNet	23.33	**0.807**	0.383
SwinIR	17.72	0.661	0.519
Restormer	20.98	0.632	0.360
RDDM	17.00	0.626	0.545
AirNet	22.73	0.739	0.374
IDR	27.91	0.793	0.346
Prompt-IR	20.47	0.669	0.462
DA-CLIP	15.74	0.606	0.472
DiffUIR-L	**29.55**	0.887	0.281
HOGformer-L	29.33	**0.889**	**0.271**

Table 4: Quantitative results of unknown task on T-OLED. The **best** and second-best results are highlighted. The top is task-specific methods, and the bottom is AIR methods.

horizontal and vertical flips for data augmentation. Following Histoformer (Sun et al. 2024), α is set to 1, and B equals the number of attention heads. N_{bin} is set to 9 according to original HOG setting (Dalal and Triggs 2005).

Comparison with State-of-the-Arts

Setting (I). Table 1 compares HOGformer with task-specific (Li et al. 2018b; Yang et al. 2017; Zhang et al. 2021; Chen et al. 2022a; Zamir et al. 2022a; Li, Cheong, and Tan 2019; Li et al. 2016; Zamir et al. 2021; Chen et al. 2022b; Qian et al. 2018; Quan et al. 2021; Tu et al. 2022) and all-in-one SOTA methods (Li, Tan, and Cheong 2020; Valanarasu, Yasarla, and Patel 2022; Chen et al. 2022c; Zhu et al. 2023; Özdenizci and Legenstein 2023; Ye et al. 2023; Wang et al. 2024; Sun et al. 2024) on synthetic and real-world adverse weather datasets. For fairness, we retrain all-in-one baselines on AllWeather (Li, Tan, and Cheong 2020; Özdenizci and Legenstein 2023). HOGformer consistently outperforms existing methods across all degradations. As shown in Figure 4 and Figure 5, HOGformer also restores clearer details and better handles multiple degradations. This underscores the importance of natural discriminative HOG attributes to model degradations.

Setting (II). Table 2 compares AIR methods under five natural degradations. HOGformer-S achieves competitive results in low-light enhancement, desnowing, and dehazing, outperforming larger models like DA-CLIP (174.1M). Besides, HOGformer-L achieves best results with reasonable parameters than counterparts (Liang et al. 2021; Zamir et al. 2022b,a; Tu et al. 2022; Luo et al. 2023; Özdenizci and Legenstein 2023; Liu et al. 2024; Zamir et al. 2022a; Li et al. 2022; Wang et al. 2023; Ma et al. 2023; Potlapalli et al. 2023; Luo et al. 2024; Zheng et al. 2024). These results highlight its balance between performance and complexity.

Generalization. To further assess generalizability, we evaluate HOGformer on both seen and unseen degradations. As shown in Table 3, HOGformer consistently outperforms previous state-of-the-art methods on seen tasks. In Table 4, HOGformer delivers comparable results on T-OLED with unseen degradations, confirming its robustness and adaptability beyond trained scenarios. This generalization likely stems from the intrinsic properties of HOG features. Unlike learned priors prone to overfitting, HOG captures fundamen-

LDRConv	DHOGSA	DIFF	$\mathcal{L}_{HOG}$	P ↑	S ↑	Params ↓
×	×	×	×	30.14	0.9258	14.65M
✓	×	×	×	30.49	0.9261	14.67M
✓	✓	×	×	31.68	0.9358	16.77M
✓	✓	✓	×	32.40	0.9421	16.64M
✓	✓	✓	✓	**32.89**	**0.9460**	16.64M

Table 5: Ablation study on the proposed core components.

	w/o FHOGR	w/o BHOGR	DHOGSA
P ↑	32.59	31.83	**32.89**
S ↑	0.9432	0.9386	**0.9460**

Table 6: Ablation study on the FHOGR and BHOGR.

tal structural cues common to many degradations. Therefore, the HOG-guided mechanism can highlight such patterns even in real-world degradations.

Ablation Studies

Ablation studies on Outdoor-Rain are conducted to evaluate our methods and to determine optimal hyperparameters.

Effectiveness of Core Components. As shown in Table 5, adding LDRConv significantly improves PSNR and SSIM. This performance gain is attributable to it models local dynamic-range variations commonly found in complex degradations. Introducing DHOGSA further boosts performance by embedding HOG-based priors into the attention mechanism. The reason is that HOG features capture gradient cues that vary across degradations, allowing the model to adjust attention weights adaptively based on perceived degradation as shown in Figure 6. The DIFF module brings additional gains by enabling dynamic feature transformation. The reason is that it enhances the ability of model to process regions with varying degradation levels through spatial-channel interaction. Table 6 shows that both BHOGR and FHOGR are beneficial. We attribute this improvement to BHOGR captures inter-interval dependencies while FHOGR focuses on intra-interval structures. Combining all components yields the best performance, confirming their complementary roles in AIR.

Hyperparameters of HOG Bins. Table 7 analyzes the effect of the number of bins in extracted HOG features. A small number of bins leads to poor performance because it fails to capture essential gradient orientations necessary for identifying degradations. Increasing the number of bins beyond a certain threshold offers only minor improvements while imposing greater computational overhead. To balance accuracy and efficiency, we set the number of bins to 9.

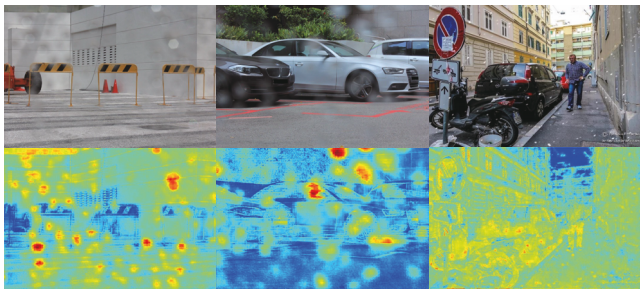


Figure 6: Visualization of the DHOGSA output from the second-to-last HOGTB, highlighting its ability to accurately activate degraded regions.

	13 Bins	9 Bins	5 Bins	1 Bin
P ↑	32.93	32.89	32.64	32.32
S ↑	0.9464	0.9460	0.9433	0.9417

Table 7: Ablation study on the number of bins in extracted HOG features.

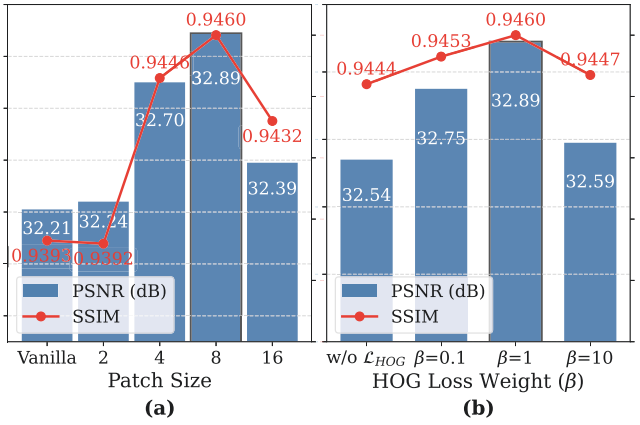


Figure 7: Ablation studies on the patch size of LDRConv (a) and the weight of HOG loss (b).

Hyperparameters of Local Dynamic-Range Convolution. Figure 7 (a) shows that LDRConv with patch size 8 achieves the best performance (32.89dB PSNR, 0.9460 SSIM), outperforming the baseline by +0.68dB and +0.0067 SSIM. Smaller patches offer limited context, while overly large patches (e.g., size 16) degrade structural fidelity.

Hyperparameters of HOG Loss. As illustrated in Figure 7 (b), setting $\beta = 1$ yields the highest performance (32.89dB PSNR, 0.9460 SSIM). A small weight ($\beta = 0.1$) offers marginal gains, while an excessive weight ($\beta = 10$) hampers performance, indicating the need to balance gradient emphasis and overall fidelity.

Conclusion

In this paper, we present HOGformer, an AIR model that leverages the power of HOG features. Through a Dynamic HOG-aware Self-Attention (DHOGSA) mechanism and innovative network components, HOGformer effectively handles diverse image degradations. Our approach demonstrates superior performance while maintaining computational efficiency. Besides, extracting HOG features explicitly helps the model identify various degradations from a gradient perspective and eliminate interfering factors like lighting. This further enhance the generalizability of our method. Future research directions include exploring HOG-based mechanisms in conjunction with emerging architectures (e.g., Mamba) to address the AIR task.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant U24A20251, 62071500, Shenzhen Science and Technology Program under Grant JCYJ20230807111107015.

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.

Banham, M. R.; and Katsaggelos, A. K. 1997. Digital image restoration. *IEEE signal processing magazine*, 14(2): 24–41.

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.

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

Chen, L.; Wu, P.; Chitta, K.; Jaeger, B.; Geiger, A.; and Li, H. 2024. End-to-end autonomous driving: Challenges and frontiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Chen, S.; Ye, T.; Liu, Y.; Chen, E.; Shi, J.; and Zhou, J. 2022b. Snowformer: Scale-aware transformer via context interaction for single image desnowing. *arXiv preprint arXiv:2208.09703*, 2.

Chen, W.-T.; Huang, Z.-K.; Tsai, C.-C.; Yang, H.-H.; Ding, J.-J.; and Kuo, S.-Y. 2022c. 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. Instructir: High-quality image restoration following human instructions. In *ECCV*, 1–21. Springer.

Dalal, N.; and Triggs, B. 2005. Histograms of oriented gradients for human detection. In *CVPR*, volume 1, 886–893. Ieee.

Guo, H.; Li, J.; Dai, T.; Ouyang, Z.; Ren, X.; and Xia, S.-T. 2024. Mambair: A simple baseline for image restoration with state-space model. In *ECCV*, 222–241. Springer.

He, Z.; and Jin, Z. 2024. Latent modulated function for computational optimal continuous image representation. In *CVPR*, 26026–26035.

Jiang, J.; Zuo, Z.; Wu, G.; Jiang, K.; and Liu, X. 2024a. A survey on all-in-one image restoration: Taxonomy, evaluation and future trends. *arXiv preprint arXiv:2410.15067*.

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.

Jiang, Y.; Zhang, Z.; Xue, T.; and Gu, J. 2024b. Autodir: Automatic all-in-one image restoration with latent diffusion. In *ECCV*, 340–359. Springer.

Jin, Z.; Qiu, Y.; Zhang, K.; Li, H.; and Luo, W. 2025. MB-TaylorFormer V2: improved multi-branch linear transformer expanded by Taylor formula for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

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.

Lee, C.; Lee, C.; and Kim, C.-S. 2013. Contrast enhancement based on layered difference representation of 2D histograms. *IEEE transactions on image processing*, 22(12): 5372–5384.

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, B.; Ren, W.; Fu, D.; Tao, D.; Feng, D.; Zeng, W.; and Wang, Z. 2018a. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1): 492–505.

Li, R.; Cheong, L.-F.; and Tan, R. T. 2019. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *CVPR*, 1633–1642.

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

Li, X.; Jin, X.; Lin, J.; Liu, S.; Wu, Y.; Yu, T.; Zhou, W.; and Chen, Z. 2020. Learning disentangled feature representation for hybrid-distorted image restoration. In *ECCV*, 313–329. Springer.

Li, X.; Wu, J.; Lin, Z.; Liu, H.; and Zha, H. 2018b. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *ECCV*, 254–269.

Li, Y.; Tan, R. T.; Guo, X.; Lu, J.; and Brown, M. S. 2016. Rain streak removal using layer priors. In *CVPR*, 2736–2744.

Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; and Timofte, R. 2021. Swinir: Image restoration using swin transformer. In *ICCV*, 1833–1844.

Liang, P.; Jiang, J.; Liu, X.; and Ma, J. 2024. Image deblurring by exploring in-depth properties of transformer. *IEEE Transactions on Neural Networks and Learning Systems*.

Liu, J.; Wang, Q.; Fan, H.; Wang, Y.; Tang, Y.; and Qu, L. 2024. Residual denoising diffusion models. In *CVPR*, 2773–2783.

Liu, Y.-F.; Jaw, D.-W.; Huang, S.-C.; and Hwang, J.-N. 2018. Desnownet: Context-aware deep network for snow removal. *IEEE Transactions on Image Processing*, 27(6): 3064–3073.

Luo, Z.; Gustafsson, F. K.; Zhao, Z.; Sjölund, J.; and Schön, T. B. 2023. Image restoration with mean-reverting stochastic differential equations. *arXiv preprint arXiv:2301.11699*.

Luo, Z.; Gustafsson, F. K.; Zhao, Z.; Sjölund, J.; and Schön, T. B. 2024. Controlling vision-language models for universal image restoration. In *ICLR*.

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, K.; Zeng, K.; and Wang, Z. 2015. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11): 3345–3356.

Mittal, A.; Soundararajan, R.; and Bovik, A. C. 2012. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3): 209–212.

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 Transactions on Pattern Analysis and Machine Intelligence*, 45(8): 10346–10357.

Potlapalli, V.; Zamir, S. W.; Khan, S. H.; and Shahbaz Khan, F. 2023. Promptir: Prompting for all-in-one image restoration. *NeurIPS*, 36: 71275–71293.

Qian, R.; Tan, R. T.; Yang, W.; Su, J.; and Liu, J. 2018. Attentive generative adversarial network for raindrop removal from a single image. In *CVPR*, 2482–2491.

Qiu, Y.; Zhang, K.; Wang, C.; Luo, W.; Li, H.; and Jin, Z. 2023. Mb-taylorformer: Multi-branch efficient transformer expanded by taylor formula for image dehazing. In *ICCV*, 12802–12813.

Quan, R.; Yu, X.; Liang, Y.; and Yang, Y. 2021. Removing raindrops and rain streaks in one go. In *CVPR*, 9147–9156.

Rim, J.; Lee, H.; Won, J.; and Cho, S. 2020. Real-world blur dataset for learning and benchmarking deblurring algorithms. In *ECCV*, 184–201. Springer.

Shen, Z.; Wang, W.; Lu, X.; Shen, J.; Ling, H.; Xu, T.; and Shao, L. 2019. Human-aware motion deblurring. In *ICCV*, 5572–5581.

Song, Y.; He, Z.; Qian, H.; and Du, X. 2023. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 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. Springer.

Tu, Z.; Talebi, H.; Zhang, H.; Yang, F.; Milanfar, P.; Bovik, A.; and Li, Y. 2022. Maxim: Multi-axis mlp for image processing. In *CVPR*, 5769–5780.

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, C.; Wu, H.; and Jin, Z. 2023. Fourllie: Boosting low-light image enhancement by fourier frequency information. In *ACM MM*, 7459–7469.

Wang, S.; Zheng, J.; Hu, H.-M.; and Li, B. 2013. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE transactions on image processing*, 22(9): 3538–3548.

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. *International Journal of Computer Vision*, 132(10): 4541–4563.

Wang, X.; Wang, W.; Cao, Y.; Shen, C.; and Huang, T. 2023. Images speak in images: A generalist painter for in-context visual learning. In *CVPR*, 6830–6839.

Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; and Li, H. 2022a. Uformer: A general u-shaped transformer for image restoration. In *CVPR*, 17683–17693.

Wang, Z.; Zhang, J.; Chen, R.; Wang, W.; and Luo, P. 2022b. Restoreformer: High-quality blind face restoration from undegraded key-value pairs. In *CVPR*, 17512–17521.

Wei, C.; Wang, W.; Yang, W.; and Liu, J. 2018. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*.

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.

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.

Zamfir, E.; Wu, Z.; Mehta, N.; Tan, Y.; Paudel, D. P.; Zhang, Y.; and Timofte, R. 2025. Complexity Experts are Task-Discriminative Learners for Any Image Restoration. In *CVPR*. Springer.

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

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; Yang, M.-H.; and Shao, L. 2021. Multi-stage progressive image restoration. In *CVPR*, 14821–14831.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; Yang, M.-H.; and Shao, L. 2022b. Learning enriched features for fast image restoration and enhancement. *IEEE transactions on pattern analysis and machine intelligence*, 45(2): 1934–1948.

Zhang, K.; Li, R.; Yu, Y.; Luo, W.; and Li, C. 2021. Deep dense multi-scale network for snow removal using semantic and depth priors. *IEEE Transactions on Image Processing*, 30: 7419–7431.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 586–595.

Zhang, X.; Xie, F.; Ding, H.; Yan, S.; and Shi, Z. 2024. Proxy and Cross-Stripes Integration Transformer for Remote Sensing Image Dehazing. *IEEE Transactions on Geoscience and Remote Sensing*.

Zheng, D.; Wu, X.-M.; Yang, S.; Zhang, J.; Hu, J.-F.; and Zheng, W.-S. 2024. Selective hourglass mapping for universal image restoration based on diffusion model. In *CVPR*, 25445–25455.

Zhou, Y.; Ren, D.; Emerton, N.; Lim, S.; and Large, T. 2021. Image restoration for under-display camera. In *CVPR*, 9179–9188.

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.