CDIR: LoRA-Inspired Attention for Efficient Composite Degradation Image Restoration

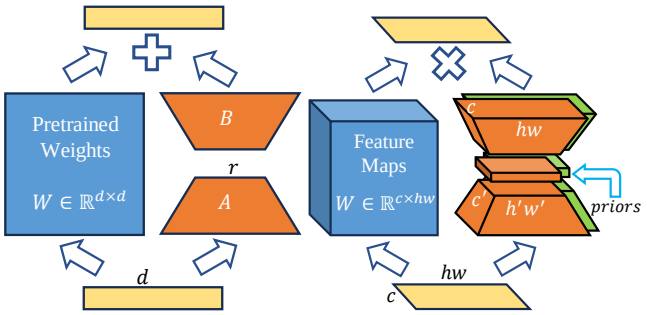
Yuning Cui, *Graduate Student Member, IEEE*, Wenqi Ren, *Senior Member, IEEE*, Boxin Shi, *Senior Member, IEEE*, Jianhou Gan, and Alois Knoll, *Fellow, IEEE*

***Abstract*—Specialized image restoration methods have been extensively explored, each targeting a specific type of degradation. However, real-world images often suffer from composite degradations, prompting growing interest in unified restoration approaches. While recent unified models have shown promising results, many are hindered by high computational complexity, limiting their deployment in resource-constrained settings. Motivated by the parameter-efficient design of Low-Rank Adaptation (LoRA), we propose an efficient attention module specifically designed for composite degradation image restoration. The proposed method adopts a dual-branch architecture, where one branch processes features at full resolution, and the other operates with reduced spatial and channel dimensions to improve efficiency. To better adapt to diverse degradation patterns, the latter branch is further divided into two sub-branches, each incorporating dynamic operations guided by local and contextual priors. These context priors are iteratively updated within each module, drawing inspiration from feedback mechanisms in reinforcement learning, thereby enabling the model to effectively perceive and handle multiple degradation types within a unified structure. Additionally, we introduce a multi-scale feed-forward network to further enhance both performance and computational efficiency. Extensive experiments on two composite degradation benchmarks demonstrate that our proposed network, CDIR, achieves state-of-the-art performance with significantly reduced complexity and fast inference speed. In addition, CDIR shows strong adaptability to various task-specific image restoration scenarios, such as dehazing, desnowing, and deraining. It also performs robustly on domain-specific applications such as ultra-high-definition (UHD), remote sensing, and medical image restoration, highlighting its versatility and practical applicability.**

***Index Terms*—Image restoration, composite degradation, efficient networks, ultra-high-definition, medical image processing, remote sensing.**

I. Introduction

IMAGE restoration seeks to recover high-quality images from their degraded observations. Adverse environmental conditions (*e.g.*, haze, rain, snow) and dynamic factors such as motion blur pose significant challenges to image visibility and compromise the reliability of downstream applications. Over the past decade, numerous task-specific image restoration methods have been developed to address individual degradation types, such as dehazing [1]–[4], deraining [5]–[7], desnowing [8]–[10], and deblurring [11]–[14]. However, these approaches are typically designed to handle a single type of degradation and often struggle in real-world scenarios, where multiple distortions, also known as composite or mixed degradations, frequently co-occur.



(a) LoRA (b) Our Attention Module

Fig. 1: Illustration of LoRA [17] and the proposed multi-branch attention module.

More recently, composite degradation approaches have attracted increasing attention in image restoration. For example, DarkIR [15] addresses both low-light enhancement and deblurring within a single framework by employing a specially designed encoder for enhancement and a decoder for deblurring. MoCE-IR [16] adopts a mixture-of-experts strategy to selectively activate specialized experts based on the input degradation characteristics. Despite their effectiveness, they often incur substantial computational overhead, limiting their practicality in resource-constrained environments.

To address this issue, we propose an efficient image restoration network for composite degradations. LoRA [17] has recently shown promising potential for parameter-efficient fine-tuning through low-rank decomposition, as illustrated in Figure 1(a). Inspired by LoRA's rank *reduction*-based design, we introduce an efficient attention module specifically tailored to handle composite degradations. As depicted in Figure 1(b), the proposed module adopts a two-branch strategy: one branch operates on the full-resolution feature map, while the other processes features at *reduced* spatial and channel dimensions to improve efficiency.

To further enhance adaptability to diverse degradation patterns, the second branch is subdivided into multiple sub-branches, each potentially targeting different types of degradation. These sub-branches incorporate dynamic operations

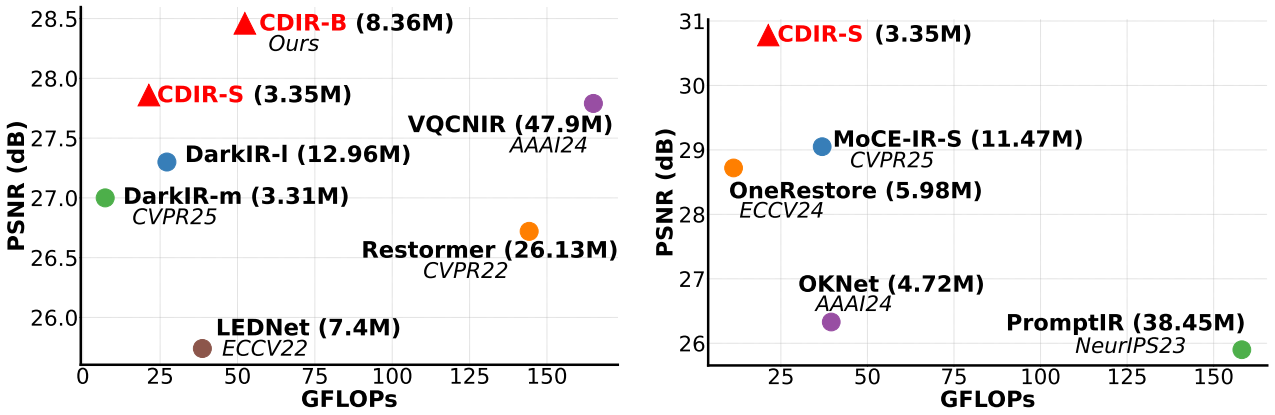
Yuning Cui and Alois Knoll are with the School of Computation, Information and Technology, Technical University of Munich, Munich, Germany.

Wenqi Ren is with the School of Cyber Science and Technology, Shenzhen Campus of Sun Yat-sen University, Shenzhen, China.

Boxin Shi is with the State Key Laboratory of Multimedia Information Processing and National Engineering Research Center of Visual Technology, School of Computer Science, Peking University, Beijing, China.

Jianhou Gan is with the Key Laboratory of Education Informatization for Nationalities, Yunnan Normal University, Kunming, China.

Corresponding author: Jianhou Gan (ganjh@ynnu.edu.cn)



(a) LOLBlur [18] (Table I) (b) CDD-11 [19] (Table IV)

Fig. 2: Computational comparison on two composite degradation datasets. The proposed CDIR achieves state-of-the-art performance with significantly improved computational efficiency.

guided by both local and contextual priors[1]. Drawing inspiration from feedback mechanisms in reinforcement learning, the context priors interact bi-directionally with the dynamic operations and are iteratively updated within each module. This design enables the model to better perceive and adapt to complex degradation patterns, thereby facilitating the recovery of high-quality images from inputs affected by multiple degradations.

Furthermore, a multi-scale feed-forward network (MFFN) is introduced to improve efficiency and maintain competitive performance. By integrating the proposed attention module and MFFN into a Transformer-style architecture, we develop a unified network, dubbed CDIR, specifically tailored for composite degradation image restoration. As shown in Figure 2(a), the compact variant CDIR-S significantly outperforms the recent DarkIR-l [15] by 0.56 dB in PSNR on the LOLBlur [18] dataset for low-light image deblurring, while reducing the number of parameters by **74%** and FLOPs by **21%**. Furthermore, on the more challenging CDD-11 [19] dataset, where images are simultaneously affected by up to three degradation types, our network achieves a substantial PSNR improvement of 1.73 dB over the recent MoCE-IR [16], using only **29%** of the parameters (see Figure 2(b)). The main contributions of this paper are summarized as follows:

- We develop an efficient image restoration network for composite degradations by integrating a LoRA-inspired attention module and a multi-scale feed-forward network.
- We propose a novel LoRA-inspired attention module that operates on reduced spatial and channel dimensions using a multi-branch architecture to address different degradation types. It further incorporates dynamic operations guided by both local and bidirectional context priors to more effectively handle composite degradations.
- Extensive experimental results demonstrate that the proposed model achieves state-of-the-art performance on two composite degradation benchmarks, with significantly fewer parameters, lower computational complexity, and faster inference speed. In addition, it performs competitively on a range of single-degradation tasks, highlighting its strong adaptability across diverse image restoration scenarios. The model also transfers well to domain-specific applications such as UHD, remote sensing, and medical image restoration, further showcasing its robustness and versatility.

[1]In this work, *context priors* refer to degradation-related information extracted from the input image and propagated across blocks, while *local priors* are computed independently within each block.

II. Related Work

A. Image Restoration

Image restoration, a highly ill-posed problem, aims to reconstruct the clean image from its degraded counterpart [20]–[25]. Compared to prior hand-crafted methods [26], [27], recent data-driven approaches have achieved superior results by learning powerful features from large-scale curated data. These approaches are mostly built on convolutional or Transformer-based architectures [1], [11], [28]–[31], addressing various single-degradation tasks, such as dehazing, deraining, desnowing, and deblurring. With advanced functional units, convolutional image restoration networks have achieved promising performance while offering efficient and fast processing. However, these frameworks suffer from limited receptive fields, which hinders their ability to handle large-scale degradations. In contrast, thanks to self-attention [29], [30], [32], Transformer-based networks possess strong modeling capabilities for long-range dependencies. Although they have significantly improved performance across various image restoration tasks, this comes at the cost of quadratic complexity with respect to input size, posing challenges for practical deployment. Recently, Mamba-based models have brought renewed momentum and innovation to the field of image restoration [33]–[35]. For example, PH-Mamba [36] introduces a position-guided hybrid Mamba module to adaptively encode spatial coordinates and discriminative patterns for precise degradation representation. In addition, diffusion-based models have demonstrated remarkable advancements in image restoration [37]. For instance, Yan *et al.* [38] develop a frequency prior-guided network to leverage the diffusion

models to produce frequency priors for generating authentic details. A comprehensive survey of diffusion-based methods can be found in [39].

However, most of the aforementioned methods perform well on a specific degradation type or require separate model instances for different types. This paper introduces an efficient network capable of handling composite degradations using a single model.

On the other hand, LoRA has been widely used in low-level vision tasks as a parameter-efficient fine-tuning method [40]. For example, PiSA-SR disentangles semantic-level and pixel-level objectives by learning two LoRA modules on top of pre-trained models [41]. Our model is orthogonal to these approaches, as it adopts the *reduction* mechanism of LoRA to guide the design of an efficient architecture.

B. Multi-Task Image Restoration

Most recently, multi-task image restoration methods have garnered increasing attention and can be broadly categorized into all-in-one [42]–[45] and composite degradation schemes [15], [16], [18], [19], [46], [47]. The former is trained on compound datasets assembled from several single-degradation tasks and evaluated on their respective test sets. As one of the pioneering solutions, AirNet [48] learns degradation priors from inputs via contrastive learning and leverages them to guide the restoration process. PromptIR [43] encodes degradation information into learnable prompts, while InstructIR [49] utilizes human-written instructions to tackle all-in-one image restoration. The recent AdaIR [50] addresses this task via frequency mining and modulation. Although these methods can handle multiple degradation types within a unified model, they assume that each input image is affected by only one type of degradation. This assumption limits their practicality in real-world applications, where images are often simultaneously impacted by multiple degradations.

To meet this requirement, several composite degradation image restoration methods have recently been proposed. For instance, DarkIR [15] combines spatial and frequency information via frequency and large receptive field attention to address low-light image deblurring. OneRestore [19] employs a cross-attention mechanism that integrates textual and visual descriptors to handle composite degradations. An extensive survey of multi-task image restoration methods is presented in [42]. Despite the promising performance of these methods, a key question remains: *Can the efficiency of composite degradation image restoration approaches be further improved to enhance their practicality in real-world applications*? To this end, we propose an efficient dynamic attention module for composite degradation restoration, drawing inspiration from the LoRA design. In parallel, a variety of priors have been explored in multi-task image restoration to encode degradation-specific information that can guide the restoration process. However, most of these priors are unidirectional, extracted from the input image and passed into the network without further refinement or interaction. In contrast, our method introduces a bidirectional prior mechanism, in which priors are dynamically updated through feedback from intermediate network features.

III. Methodology

This section first outlines the overall pipeline of the proposed network, followed by detailed descriptions of its key components, including the Dual-LoRA Attention (DLA) module and the MFFN.

A. Overall Pipeline

Figure 3(a) illustrates the pipeline of our network, CDIR, which is built upon a widely adopted hierarchical architecture. Shallow features are first extracted by applying a 3×3 convolution to the degraded input image. These features are then passed through an encoder-decoder structure to remove degradation artifacts and recover texture details. During this process, the encoder progressively downsamples the spatial resolution while increasing the channel capacity, transforming the feature shape from $H \times W \times C$ to $\frac{H}{4} \times \frac{W}{4} \times 4C$. Conversely, the decoder restores the feature maps to their original resolution, employing residual connections to facilitate effective reconstruction. Each stage in both the encoder and decoder consists of multiple basic blocks, with the number of blocks per stage ranging from L_1 to L_4. Context priors, generated by the Prior Generation (PG) unit, are injected into each basic block to guide the restoration process and are iteratively updated to better capture degradation characteristics. The computation within a basic block is defined as follows:

$$x_{i+1} = \mathrm{MFFN}(\mathrm{Norm}(x'_i)) + x'_i, \tag{1}$$

$$x'_i = \mathrm{DLA}(\mathrm{Norm}(x_i), P_i) + x_i, \tag{2}$$

where x_i and x_{i+1} denote the input and output feature maps of the basic block, respectively. P_i indicates the input context priors, while P_{i+1} is produced as one of the outputs of DLA. *Norm* denotes the normalization layer. Finally, the residual image is generated via a 3×3 convolution and added to the input degraded image to produce the final output of the model.

B. DLA

1) Motivation: LoRA [17] has emerged as a prominent and widely adopted method for parameter-efficient fine-tuning, based on the premise that the updated matrices in model adaptation have a low intrinsic rank. It introduces low-rank decomposition matrices into pre-trained weights as $W_0 + \Delta W = W_0 + BA$, where W_0 represents pre-trained weight matrix and $B \in \mathbb{R}^{d \times r}$, $A \in \mathbb{R}^{r \times k}$ are low-rank matrices. The low-rank decomposition ($r \ll min(d, k)$) greatly reduces the trainable parameters for downstream tasks. Inspired by its *reduction* mechanism, we develop a DLA module that operates through two super-branch transformations, defined as follows:

$$\begin{aligned}\hat{X} =& \mathcal{F}_1^{H\times W\times C\to H\times W\times C}(X) \\ &\otimes \mathcal{F}_2^{H\times W\times C\to H'\times W'\times C'\to H\times W\times C}(X),\end{aligned} \tag{3}$$

where X and $\hat{X} \in \mathbb{R}^{H \times C \times W}$ denote the input and output feature maps, respectively. $\mathcal{F}^{S_1 \to S_2}(\cdot)$ represents the network processing that involves changing the shape of the feature map from S_1 to S_2. In our case, *reducing* both the channel and spatial dimensions in the second branch ($H' < H$, $W' < W$,

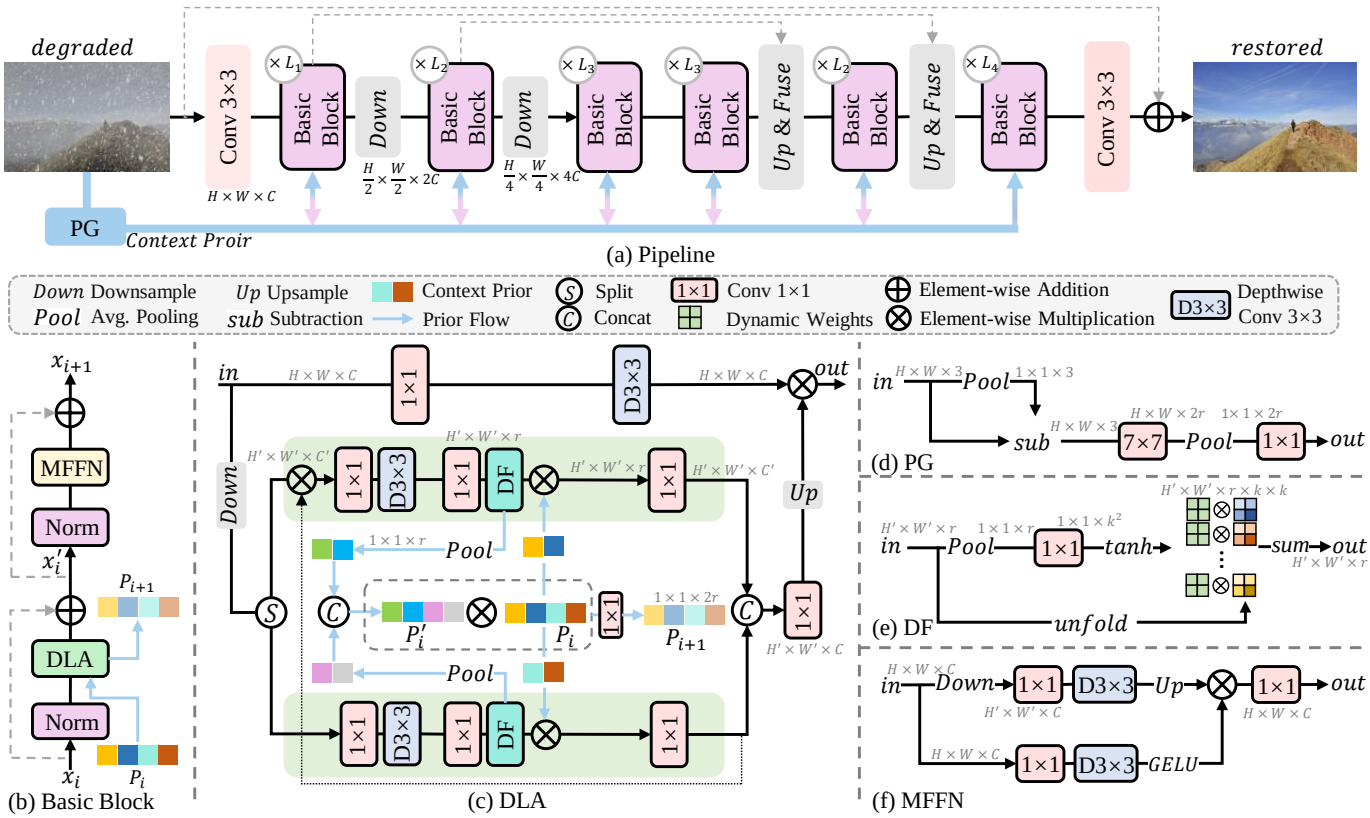


Fig. 3: (a) Overall pipeline of the proposed network for efficient composite degradation image restoration, which adopts a U-shaped architecture. Each stage comprises multiple basic blocks guided by bidirectional context priors. (b) Structure of the basic block, which integrates the Multi-Scale Feedforward Network (MFFN) and the Dual-LoRA Attention (DLA) module. Within DLA, context priors are utilized to guide restoration and are iteratively updated to better capture degradation characteristics. (c) The DLA module reduces both spatial ($HW \rightarrow H'W'$) and channel ($C \rightarrow r$) dimensions to enable the application of Dynamic Filters (DF) for handling diverse degradation types. (d) The Prior Generation (PG) unit extracts context priors from the high-frequency components of the degraded input image. (e) DF dynamically performs degradation removal based on learned kernels. (f) MFFN enhances multi-scale representation learning within the gated feedforward network.

$C' < C$) enables higher efficiency and facilitates multi-scale representation learning. We next describe the details of DLA, followed by the design of the prior-guided flow.

2) Mesoscale Architecture Design: Eq. (3) illustrates that our attention mechanism adopts a multi-branch paradigm, where one branch operates on full-size features and the other follows a LoRA-inspired *reduction* strategy. Notably, the LoRA-inspired super branch can be naturally extended to incorporate multiple branches, allowing the model to better handle composite degradations. An example of such an extension, using a two-branch configuration, is illustrated in Figure 1(b), with architectural details provided in Figure 3(c). In this case, the two super branches in Eq. (3) can be reformulated to reflect our specific design as:

Super-Branch 1: (4)

$$\mathcal{F}_1(X) = W^D_{3\times3}(W_{1\times1}(X)), \tag{5}$$

Super-Branch 2: (6)

$$\mathcal{F}_2(X) = \mathcal{R}^{up}\Big(W_{1\times1}[\mathcal{B}_1(X_1), \mathcal{B}_2\big(\mathcal{B}_1(X_1) \otimes X_2\big)]\Big), \tag{7}$$

$$X_1, X_2 = \text{Split}(\mathcal{R}^{down}(X)), \tag{8}$$

where $W_{1\times1}$ and $W^D_{3\times3}$ denote the 1×1 convolution and 3×3 depthwise version, respectively. $\mathcal{R}^{up}$ and $\mathcal{R}^{down}$ represent upsampling and downsampling operations, respectively, while $[\cdot,\cdot]$ and *Split* denote channel-wise concatenation and splitting. $\mathcal{B}$ denotes the operation of a sub-branch in the super-branch 2. For simplicity, the superscripts of $\mathcal{F}$ and the context priors are omitted. Furthermore, the input to the top sub-branch is first modulated by the output of the bottom sub-branch (see Figure 3(c)), enabling high-order spatial interactions [51] and facilitating multi-scale feature learning [52]. We next detail the internal components of each sub-branch.

3) Microscale Architecture Design: Each sub-branch of Super-Branch 2 adopts a squeeze-expand strategy along the channel dimension, promoting computational efficiency and enabling the use of dynamic filters for handling composite degradations. Taking the bottom sub-branch as an example, let $X_1 \in \mathbb{R}^{H'\times W'\times C'}$ denote the input feature map. The input is first processed using a combination of a 1×1 convolution and a 3×3 depthwise convolution. The resulting features are then compressed along the channel dimension to a reduced shape of $H'\times W'\times r$, where $r \ll C'$. These compressed features are subsequently refined by a Dynamic Filter (DF), guided

by both local and contextual priors. Finally, the features are projected back to the original channel dimension via a 1×1 convolution. The overall computation of this sub-branch can be formally expressed as:

$$\hat{X}_1 = W_{1\times1}^{r\to C}(\mathrm{DF}(X_1') \otimes P_i^1), \tag{9}$$

$$X_1' = W_{1\times1}^{C\to r}(W_{3\times3}^{D} W_{1\times1}(X_1)), \tag{10}$$

where $W_{1\times1}^{C\to r}$ and $W_{1\times1}^{r\to C}$ denote 1×1 convolutions used for channel squeezing and expansion, with reduction and expansion ratios of $\frac{r}{C}$ and $\frac{C}{r}$, respectively. These operations form the basis of our channel-wise *reduction* mechanism. P_i represents the context prior, which guides the integration of the two sub-branches to address different types of degradations.

DF. The incorporation of local-prior-guided dynamic filters enables adaptive restoration and facilitates the handling of multiple degradation types through a multi-branch design. As illustrated in Figure 3(e), given an input $X_1' \in \mathbb{R}^{H'\times W'\times r}$, DF first applies global average pooling, followed by a 1×1 convolution and a *Tanh* nonlinearity to generate adaptive kernels of size $k \times k$. The input is then unfolded into the shape $H' \times W' \times r \times k \times k$. Subsequently, each $k \times k$ patch is element-wise multiplied by the corresponding learned kernel and aggregated to produce the output of DF. The adaptive kernels are shared across spatial and channel dimensions to ensure high efficiency.

4) Context Priors: Existing methods mostly adopt a unidirectional strategy to generate priors for enhancing image restoration performance. For example, PromptIR [43] leverages preset in-situ parameters as prompts, which are independent and lack interactions across different blocks. MoCE-IR [16] employs a Sobel-filtered global feature vector to enhance the frequency awareness of the gating functions in mixture-of-experts architectures. In contrast, inspired by the feedback mechanism in reinforcement learning, our model introduces bidirectional priors that are iteratively updated within each block, thereby improving both degradation perception and the guidance provided to the DF.

Generation. The PG unit produces context priors based on the high-frequency components of the degraded input image (see Figure 3(d)). Given an input image $I \in \mathbb{R}^{H\times W\times 3}$, PG first extracts high-frequency subbands by subtracting the low-frequency component, obtained via global average pooling [53]. This is followed by a 7×7 convolution to adjust the number of channels and enlarge the receptive field. Subsequently, channel-wise context priors are generated through pooling operations and a 1×1 convolution. The complete process of PG can be expressed as:

$$P = W_{1\times1}(\mathrm{Pool}(W_{7\times7}^{3\to 2r}(I_h))), \tag{11}$$

$$I_h = I - \mathrm{Pool}(I) \tag{12}$$

where Pool denotes global average pooling and $P \in \mathbb{R}^{1\times1\times2r}$ represents the generated context priors, which guide the integration of features enhanced by the DF.

Update. The priors derived from Eq. (11) are integrated into each DLA to guide the restoration process and are iteratively updated to enhance the modeling of degradation information. Specifically, let X_d^1 and X_d^2 denote the outputs of the DF in the bottom and top sub-branches of Figure 3(c), respectively. These outputs are globally averaged across the spatial dimensions to produce 1-D channel descriptors. The resulting vectors are then concatenated and used to modulate the input priors via element-wise multiplication. Finally, a pointwise convolution is applied to refine the modulated priors. The entire update process can be formally expressed as:

$$P_{i+1} = W_{1\times1}(P_i \otimes P_i'), \tag{13}$$

$$P_i' = [\mathrm{Pool}(X_d^1), \mathrm{Pool}(X_d^2)], \tag{14}$$

where P_i and P_{i+1} denote the input and updated context priors, respectively.

C. MFFN

To enhance the model's capacity for multi-scale representation learning, we introduce an MFFN, extending the gated feedforward structure proposed in [29]. As illustrated in Figure 3(f), given an input feature map $X \in \mathbb{R}^{H\times W\times C}$, MFFN first downsamples it to a lower spatial resolution of $\mathbb{R}^{H'\times W'\times C}$, followed by a 1×1 convolution and a 3×3 depthwise convolution for refinement. The upsampled features are then modulated by a gating signal generated from a parallel branch, which operates directly on the original-resolution features using convolutional layers and GELU activation. Finally, a 1×1 convolution produces the output feature map $\hat{X}$ of MFFN. The entire process can be formulated as:

$$\hat{X} = W_{1\times1}(\tilde{X}_1 \otimes \tilde{X}_2) \tag{15}$$

$$\tilde{X}_1 = R^{up}(W_{3\times3}^{D} W_{1\times1}(R^{down}(X))), \tag{16}$$

$$\tilde{X}_2 = \mathrm{GELU}(W_{3\times3}^{D} W_{1\times1}(X)). \tag{17}$$

IV. Experiments and Analysis

We evaluate the effectiveness of the proposed CDIR under two-degradation and three-degradation[2] settings for composite degradation image restoration, using the LOLBlur [18] and CDD-11 [19] datasets, respectively. Images in LOLBlur are simultaneously degraded by *low-light* conditions and *blur*, while CDD-11 contains images affected by a total of 11 degradation types, including *haze*, *snow*, *low light*, *rain*, and their combinations: *low+haze*, *low+rain*, *low+snow*, *haze+rain*, *haze+snow*, *low+haze+rain*, and *low+haze+snow*. In addition, we assess the adaptability of our model on standard real-world and synthetic single-degradation datasets across multiple image restoration tasks, including dehazing, desnowing, and deraining. Furthermore, our model is evaluated on domain-specific tasks, such as UHD, remote sensing, and medical image restoration.

Implementation details. Our model is trained using the Adam optimizer and dual-domain l_1 loss function [53], [54]. Regarding the network hyperparameters in Figure 3, we set $H' = \frac{H}{2}$, $W' = \frac{W}{2}$, $r = 8$, and $k = 3$. Our model is evaluated on multiple datasets. Owing to the varying complexity of these datasets, we scale the model by adjusting the channel width and the number of blocks to achieve computational costs

[2]In our study, the three-degradation setting refers to scenarios in which images are simultaneously affected by up to three types of degradations.

TABLE I: Quantitative results on the LOLBlur [18] dataset for composite degradations (*i.e.*, *low light* and *blur*). The proposed lightweight model outperforms the specialized method DarkIR-l [15] by 0.56 dB in PSNR, while reducing the number of parameters by 74% and computational cost by 21%.

Method	NAFNet [55]	LEDNet [18]	RetinexFormer [56]	Restormer [29]	VQCNIR [46]	DarkIR-m [15]	DarkIR-l [15]	CDIR-S	CDIR-B
PSNR	25.36	25.74	26.02	26.72	27.79	27.00	27.30	27.86	28.46
SSIM	0.882	0.850	0.887	0.902	0.875	0.883	0.898	0.919	0.926
Params	12.05M	7.4M	1.61M	26.13M	47.9M	3.31M	12.96M	3.35M	8.36M
FLOPs	12.3G	38.65G	15.57G	144.25G	165G	7.25G	27.19G	21.36G	52.38G

TABLE II: Three model variants of the proposed method.

Variants	Blocks ($L_1 \sim L_4$)	Channels	Params	FLOPs
CDIR-T	[1, 1, 2, 5]	[32, 64, 128, 32]	1.31M	11.92G
CDIR-S	[2, 3, 6, 6]	[32, 64, 128, 32]	3.35M	21.36G
CDIR-B	[3, 3, 7, 7]	[48, 96, 192, 48]	8.36M	52.38G

TABLE III: Training configurations for various image restoration tasks, including dataset, model variant, learning rate (LR), patch size, batch size, and training iterations.

Scenario/Task	Dataset	Model	LR	Patch	Batch	Iteration
Composite degradation						
Two-degradation	LOLBlur	CDIR-S/-B	1e-3	256	16	300K
Three-degradation	CDD-11	CDIR-T/-S	1e-3	256	16	300K
Single-degradation						
Desnowing	CSD	CDIR-B	1e-3	256	12	300K
	SRRS	CDIR-B	1e-3	256	12	300K
	Snow100K	CDIR-B	1e-3	256	12	300K
Dehazing	Dense-Haze	CDIR-S	1e-3	512	4	50K
	NH-HAZE	CDIR-S	1e-3	512	4	50K
	Haze4K	CDIR-T/-B	1e-3	256	12	300K
Deraining	AGAN	CDIR-B	1e-4	256	12	50K
UHD image restoration						
Deblurring	UHD-Blur	CDIR-T	1e-3	768	6	100K
Enhancement	UHD-LL	CDIR-T	1e-3	768	6	100K
Deraining	4K-Rain13k	CDIR-T	1e-3	768	6	100K
Medical image restoration						
Super-resolution	IXI	CDIR-T	2e-4	128	4	300K
Image synthesis	PolarStar M660	CDIR-T	2e-4	128	4	300K
Image denoising	AAPM	CDIR-T	2e-4	128	4	300K
Remote sensing						
Dehazing	SateHaze1k	CDIR-T	1e-3	256	16	50K

comparable to representative methods for each task, thereby ensuring fair comparisons. In particular, CDIR-T (Tiny) is configured with $C = 32$ and layer depths $[L_1, L_2, L_3, L_4] = [1, 1, 2, 5]$; CDIR-S (Small) uses $C = 32$ with $[2, 3, 6, 6]$; while CDIR-B (Base) adopts $C = 48$ and $[3, 3, 7, 7]$. Table II summarizes the configurations and computational overhead of different variants. A summary of the training configurations for various image restoration scenarios is presented in Table III. Notably, our training configurations largely follow previous methods within each domain, without leveraging any task-specific tricks to improve performance. FLOPs are measured using a patch of size $3 \times 256 \times 256$. In the tables, the top-performing and second-best quality scores are highlighted in red and blue, respectively.

TABLE IV: Comparison on the CDD-11 [19] dataset under composite degradations in terms of PSNR/SSIM. CDD-11 comprises 11 degradation combinations categorized into three scenarios: *single*, *double*, and *triple*. The *average* performance is computed by averaging results across all combinations. The two super-row methods correspond to single-task and multi-task image restoration approaches, respectively.

Method	Single	Double	Triple	Average	Params
MIRNetv2 [57]	28.46/.911	24.11/.817	22.35/.719	25.37/.834	5.86M
Restormer [29]	30.29/.931	25.60/.851	23.86/.767	26.99/.865	26.13M
SRUDC [58]	31.18/.930	26.27/.845	23.98/.758	27.64/.860	6.80M
OKNet [10]	29.31/.926	25.16/.846	23.30/.765	26.33/.861	4.72M
AirNet [48]	25.59/.885	22.97/.796	22.02/.717	23.75/.814	8.93M
WeatherDiff [59]	23.81/.860	21.97/.786	21.13/.707	22.49/.799	82.96M
PromptIR [43]	28.88/.920	24.46/.834	23.54/.750	25.90/.850	38.45M
WGWSNet [60]	29.97/.923	25.77/.850	23.94/.772	26.96/.863	25.76M
OneRestore [19]	31.81/.939	27.65/.871	25.23/.796	28.72/.882	5.98M
MoCE-IR-S [16]	32.54/.941	27.73/.869	25.40/.790	29.05/.881	11.47M
CDIR-T	31.79/.942	28.93/.875	25.86/.800	29.67/.886	1.31M
CDIR-S	34.50/.946	29.56/.881	26.39/.808	30.78/.892	3.35M

A. Composite Degradation Image Restoration

Two-degradation setting. Table I presents quantitative comparisons on the LOLBlur [18] dataset for low-light image deblurring. VQCNIR [46] achieves a PSNR of 27.79 dB but requires a large number of parameters and high computational complexity. While the two variants of DarkIR [15] significantly reduce computational overhead compared to VQCNIR, they do so at the cost of degraded performance. In contrast, our model delivers superior accuracy while maintaining high efficiency. Notably, CDIR-S surpasses VQCNIR by 0.07 dB in PSNR while reducing parameter count by **93**% and FLOPs by **87**%. Compared to DarkIR-l, CDIR-S achieves a substantial PSNR improvement of 0.56 dB with only **26**% of the parameters. The base version of our model further improves restoration quality while still maintaining a lightweight design. Visual comparisons are provided in Figure 4. Compared to VQCNIR, our model more effectively restores structural details (*e.g.*, the text on the billboard in the first example) and demonstrates superior brightness perception in the second example.

Three-degradation setting. We further evaluate the effectiveness of our model on the more challenging CDD-11 [19] dataset. The 11 degradation types in CDD-11 are categorized into three scenarios: *single*, *double*, and *triple* degradations, where *triple* refers to images simultaneously affected by three types of degradations. We compare our model against both single-task and multi-task image restoration methods. As shown in Table IV, our CDIR-S consistently achieves

Fig. 4: Qualitative comparison on the LOLBlur [18] dataset for composite degradations (*low light* and *blur*). Compared to competing methods, the proposed model better restores structural details in the first example and demonstrates improved brightness perception in the second example.

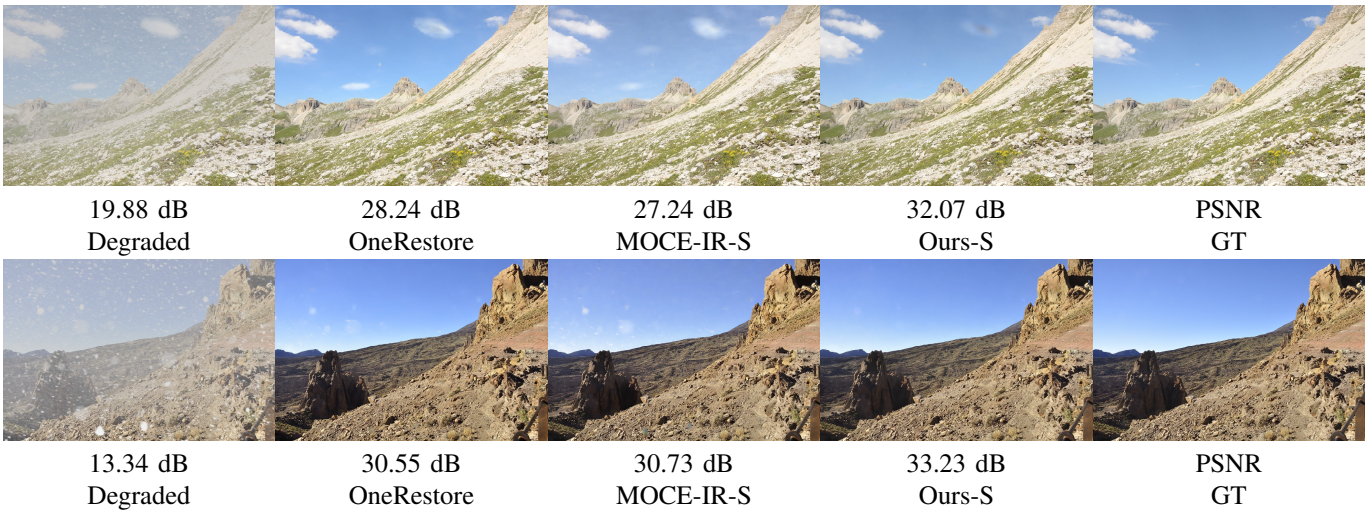


Fig. 5: Qualitative results on the CDD-11 [19] dataset with composite degradations (*haze* and *snow*). The proposed model yields results visually closer to the ground truth than competing methods.

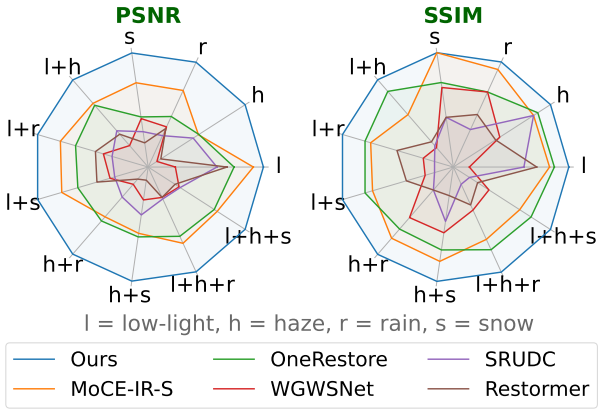
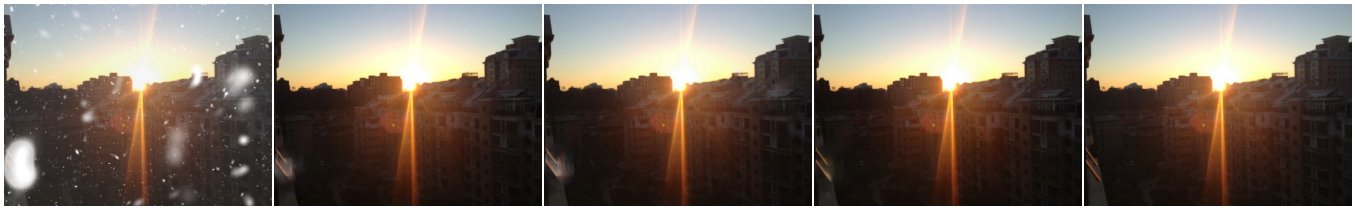


Fig. 6: Quantitative comparison across the 11 degradation categories of CDD-11 [19]. "Ours" refers to our small model.

the best performance across all scenarios while using fewer parameters than multi-task baselines. Specifically, OneRestore [19], a method tailored for composite degradation image restoration, achieves an average PSNR of 28.72 dB across the three scenarios. The more recent MoCE-IR-S [16] offers improved performance but incurs higher parameter complexity. In contrast, our model not only surpasses both methods in accuracy, outperforming MoCE-IR-S by a significant margin of 1.73 dB, but also maintains fewer parameters. Notably, our tiny model achieves the second-best performance on most metrics. A more fine-grained comparison with leading algorithms across individual degradation types is shown in Figure 6. Our model consistently achieves higher PSNR and SSIM scores than competitors, demonstrating its effectiveness in addressing complex composite degradations. Qualitative results in Figure 5 further confirm that our model produces outputs significantly closer to reference images, effectively removing challenging degradations from the input.



15.12 dB	32.81 dB	33.16 dB	34.17 dB	PSNR
Degraded	FocalNet	MBT-L V2	Ours-B	GT

Fig. 7: A qualitative comparison on the SRRS [61] dataset for image desnowing demonstrates that our model more effectively removes snow-related degradations, particularly in challenging regions.

TABLE V: Image desnowing comparisons on three widely used datasets: CSD [8], SRRS [61], and Snow100K [9]. "–" indicates that results are not available. Our model improves PSNR by 0.32 dB over the recent Transformer-based MBT-L V2 [54] on Snow100K, while reducing FLOPs by 39%.

Dataset	Metric	DesnowNet [9]	JSTASR [61]	HDCW-Net [8]	SMGARN [62]	FocalNet [63]	IRNeXt [6]	ConvIR-B [53]	MBT-L V2 [54]	CDIR-B
CSD	PSNR	20.13	27.96	29.06	31.93	37.18	37.29	39.10	-	39.94
	SSIM	0.81	0.88	0.91	0.95	0.99	0.99	0.99	-	0.99
SRRS	PSNR	20.38	25.82	27.78	29.14	31.34	31.91	32.39	32.55	32.86
	SSIM	0.84	0.89	0.92	0.94	0.98	0.98	0.98	0.98	0.98
Snow100K	PSNR	30.50	23.12	31.54	31.92	33.53	33.61	33.92	34.01	34.33
	SSIM	0.94	0.86	0.95	0.93	0.95	0.95	0.96	0.96	0.96
Overhead	Params	15.6M	65M	6.99M	6.83M	3.74M	5.46M	8.63M	7.29M	8.36M
	FLOPs	1.7K	-	9.8G	450.3G	30.6G	42.1G	71.2G	86.0G	52.38G

TABLE VI: Image dehazing comparisons on two widely used real-world datasets: Dense-Haze [64] and NH-HAZE [65]. The proposed model outperforms the recent MBT-L V2 [54] on both benchmarks while using less than half the parameters and only a quarter of the FLOPs.

Dataset	Metric	GDN [66]	MSBDN [67]	FFA-Net [68]	DeHamer [1]	C^2PNet [69]	MBT-L V1 [70]	ConvIR-B [53]	MBT-L V2 [54]	CDIR-S
Dense-Haze	PSNR	13.31	15.13	15.70	16.62	16.88	16.44	16.86	16.90	17.10
	SSIM	0.37	0.56	0.55	0.56	0.57	0.57	0.60	0.61	0.62
NH-HAZE	PSNR	13.80	17.97	18.13	20.66	20.24	20.49	20.66	20.77	20.87
	SSIM	0.54	0.66	0.65	0.68	0.69	0.69	0.69	0.71	0.71
Overhead	Params	0.956M	31.35M	4.46M	132.5M	7.17M	7.43M	8.63M	7.29M	3.35M
	FLOPs	21.5G	41.5G	287.8G	60.3G	461G	88.1G	71.2G	86G	21.36G

TABLE VII: Image dehazing results on the synthetic Haze4k [71] dataset. The proposed model achieves a significant PSNR gain of 0.79 dB over MBT-L V2 [54], with lower computational complexity.

Method	MSBDN [67]	FFA-Net [68]	PMNet [72]	FSNet [73]	ConvIR-S [53]	ConvIR-B [53]	ConvIR-L [53]	MBT-L V1 [70]	MBT-L V2 [54]	CDIR-T Ours	CDIR-B Ours
PSNR	22.99	26.96	33.49	34.12	33.36	34.15	34.50	34.47	34.92	33.39	35.71
SSIM	0.85	0.95	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
Params	31.35M	4.46M	18.90M	13.28M	5.53M	8.63M	14.83M	7.43M	7.29M	1.31M	8.36M
FLOPs	41.5G	287.8G	81.1G	110.5G	42.1G	71.2G	129.9G	88.1G	86G	11.92G	52.38G

B. Single-Degradation Image Restoration

We evaluate the adaptability of our model on several single-degradation image restoration tasks, including dehazing, desnowing, and deraining. In this setting, the model is trained and evaluated independently for each task.

Image desnowing. Table V presents desnowing results on three widely used datasets: CSD [8], SRRS [61], and Snow100K [9]. Our model significantly outperforms the recent Transformer-based MBT-L V2 [54] by 0.31 dB and 0.32 dB in PSNR on SRRS and Snow100K, respectively, while reducing computational complexity by 39%. As shown in Figure 7, our model effectively addresses heavy snow degradation, achieving visually superior results compared to existing methods.

Image dehazing. We also assess our model on widely adopted real-world dehazing datasets (Dense-Haze [64], NH-HAZE [65]) and a synthetic dataset (Haze4k [71]), with results shown in Tables VI and VII, respectively. In particular, our model achieves PSNR gains of 0.2 dB and 0.1 dB over MBT-L V2 on Dense-Haze and NH-HAZE, respectively, while requiring less than half the parameters and only a quarter of the FLOPs. In addition, our base model outperforms MBT-

TABLE VIII: UHD image deblurring results on the UHD-Blur [74] dataset.

Method	PSNR	SSIM	Params
Stripformer† [12]	25.05	0.725	19.7M
FFTformer† [75]	25.41	0.725	16.6M
UHDFormer [74]	28.82	0.844	0.339M
UHDDIP [76]	29.51	0.858	0.81M
ERR [77]	29.72	0.861	1.131M
CDIR-T	31.01	0.886	1.31M

TABLE IX: UHD low-light image enhancement results on UHD-LL [78].

Method	PSNR	SSIM	Params
LLFormer [79]	22.79	0.853	13.15M
UHDFour [78]	26.22	0.900	17.54M
Wave-Mamba [80]	27.35	0.913	1.258M
UHDFormer [74]	27.11	0.927	0.339M
ERR [77]	27.57	0.932	1.131M
CDIR-T	28.04	0.935	1.31M

TABLE X: UHD image deraining results on the 4K-Rain13k [81] dataset.

Method	PSNR	SSIM	Params
Restormer† [29]	33.02	0.933	26.12M
DRSformer† [5]	32.94	0.933	33.65M
UDR-S2Former† [82]	33.36	0.946	8.53M
UDR-Mixer [81]	34.28	0.951	4.90M
ERR [77]	34.48	0.952	1.131M
CDIR-T	34.97	0.956	1.31M

Fig. 8: Image dehazing comparison on the real-world Dense-Haze [64] dataset.

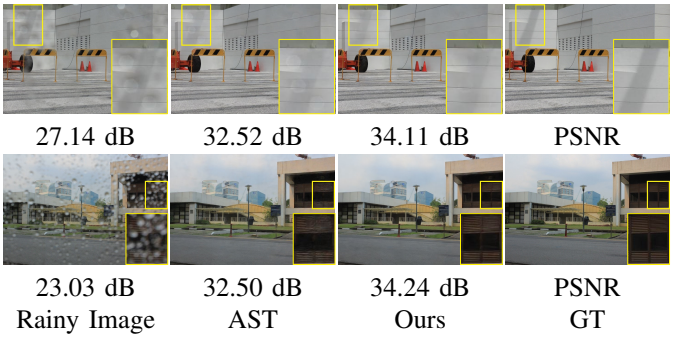


Fig. 9: Visual results on AGAN [83] for image deraining.

TABLE XI: Image deraining results on AGAN [83]. CDIR achieves notable improvements compared to the frequency-based FPro [85] and Transformer-based AST-B [84].

Method	MAXIM [86]	AWRCP [87]	FPro [85]	AST-B [84]	CDIR-B
PSNR	31.87	31.93	31.96	32.32	32.62
SSIM	0.935	0.931	0.937	0.935	0.943

L V2 on the Haze4K [71] dataset while using a comparable number of parameters and lower FLOPs. We also evaluate our tiny variant on this dataset. Due to its reduced computational budget and limited representational capacity, the tiny model underperforms the base model. Notably, compared with ConvIR-S [53], our tiny model delivers a 0.03 dB improvement in PSNR while using less than 30% of the parameters and FLOPs. Figure 8 provides a visual comparison that highlights the superiority of our model in handling real-world hazy scenes.

Image deraining. Table XI presents quantitative results on the AGAN [83] dataset for the raindrop removal task. Following prior work [84], [85], quality metrics are evaluated in the YCbCr color space for fair comparison. The proposed model achieves 32.62 dB PSNR, surpassing the frequency-based FPro [85] by a notable margin of 0.66 dB. Compared to the recent Transformer-based AST-B [84], it further improves PSNR by 0.3 dB and SSIM by 0.008. A visual comparison with AST-B, the second-best method, is shown in Figure 9. As illustrated, our model more effectively removes raindrop artifacts and reconstructs fine structural details, particularly evident in the top-left corner of the first image and the brown fence in the second, as highlighted by the yellow boxes.

C. Domain-Specific Image Restoration

This section presents results on UHD, medical, and remote sensing image restoration tasks. In the tables, methods marked with † denote general approaches, while the others are task-specific.

UHD image restoration. We evaluate the proposed model on three UHD image restoration tasks at a resolution of 3840×2160: deblurring, low-light image enhancement, and deraining, and compare our results with state-of-the-art algorithms in this domain. Tables VIII, IX, and X show that our model outperforms recent methods on these three tasks. In particular, our approach achieves performance gains of 1.29 dB, 0.47 dB, and 0.49 dB in PSNR over ERR [77] on the deblurring, enhancement, and deraining tasks, respectively, while maintaining comparable parameter efficiency. In addition, the proposed method surpasses the Transformer-based [74] and Mamba-based [80] baselines on the enhancement dataset.

Medical image restoration. In addition to natural images, we evaluate the proposed model on three medical image restoration tasks, including MPR super-resolution, PET image synthesis, and CT image denoising, following the protocol in [88]. Table XII presents the comparison results on the IXI [89] dataset for MRI image super-resolution. Our model achieves superior performance compared to the recent Restore-RWKV [88] across all three evaluation metrics. The advantages of our model are also demonstrated in Table XIII and Table XIV, on the PET image synthesis and CT image denoising tasks, respectively. Although not specifically designed for this task, our model demonstrates strong effectiveness and adaptability in medical image processing.

Remote sensing image restoration. Considering the important role of image restoration in remote sensing, we evaluate the proposed model on the remote sensing image dehazing task

TABLE XII: MRI image super-resolution results on IXI [89].

Method	PSNR↑	SSIM↑	RMSE↓	Params
FAWDN [90]	30.04	0.914	35.956	5.07M
SwinMR [91]	30.93	0.925	32.734	11.40M
SDAUT [92]	30.96	0.926	32.593	67.23M
F-UNet [93]	31.26	0.931	31.568	32.12M
Restore-RWKV [88]	31.36	0.931	31.256	1.16M
CDIR-T	31.57	0.934	30.541	1.31M

TABLE XIII: PET image synthesis results on PolarStar M660 [88].

Method	PSNR	SSIM	RMSE	Params
CycleWGAN [94]	36.62	0.929	0.091	1.00M
DCITN [95]	36.09	0.929	0.097	0.08M
DRMC [96]	36.00	0.935	0.100	0.62M
ARGAN [97]	36.73	0.941	0.090	31.14M
Restore-RWKV [88]	36.96	0.943	0.089	1.16M
CDIR-T	37.22	0.947	0.087	1.31M

TABLE XIV: CT image denoising results on AAPM [98].

Method	PSNR	SSIM	RMSE	Params
TransCT [99]	32.62	0.908	9.533	13.23M
Eformer [100]	33.35	0.918	8.803	0.34M
CTformer [101]	33.25	0.913	8.897	1.45M
DenoMamba [102]	33.53	0.915	8.612	112.62M
Restore-RWKV [88]	33.64	0.918	8.514	1.16M
CDIR-T	33.77	0.919	8.386	1.31M

TABLE XV: Numerical results on the SateHaze1k [103] dataset for remote sensing image dehazing, comprising three haze levels: thin, moderate, and thick.

Method	Thin		Moderate		Thick	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Uformer [30]	22.00	0.887	25.15	0.933	20.61	0.820
DehazeFormer [28]	24.26	0.909	25.69	0.938	22.26	0.835
EMPF [104]	22.69	0.896	25.17	0.932	20.23	0.822
Trinity [105]	22.65	0.896	24.73	0.934	20.57	0.824
FMambaIR [106]	24.58	0.912	25.83	0.939	22.65	0.850
CDIR-T	24.84	0.923	27.20	0.939	23.05	0.860

TABLE XVI: Inference speed comparison with leading methods on both composite and single-degradation image restoration tasks. All speeds are measured on an NVIDIA Tesla A100 GPU with 40 GB of memory. The red font highlights the improvements our model achieves, offering superior performance while significantly reducing inference time.

Task	Dataset	Method	GFLOPs	PSNR		Time/s	
Composite (Three-degradation)	CDD-11	MoCE-IR-S	36.93	29.05		0.878	
		Ours-S	21.36	30.78	+1.73	0.143	×6.14
Composite (Two-degradation)	LOLBlur	VQCNIR	165	27.79		0.406	
		Ours-S	21.36	27.86	+0.07	0.141	×2.88
Single-degradation (Dehazing)	Haze4K	MBT-L V2	86	34.92		1.747	
		Ours-B	52.38	35.71	+0.79	0.092	×18.99
Single-degradation (Deraining)	AGAN	AST-B	80.72	32.32		0.533	
		Ours-B	52.38	32.62	+0.30	0.133	×4.01

using the SateHaze1k [103] dataset, which includes three haze levels: thin, moderate, and thick. We train separate models for each haze level. As shown in Table XV, our model achieves the best performance across all three levels and significantly outperforms the Mamba-based method [106].

D. Inference Speed Comparison

This section compares the inference speed of our model with state-of-the-art methods on both composite and single-degradation image restoration tasks. Table XVI shows that the model achieves higher performance while offering faster runtime. Specifically, on the CDD-11 [19] dataset for composite degradation restoration, it outperforms the recent MoCE-IR-S [16] by 1.73 dB in PSNR and delivers a **6.14**× acceleration in inference speed. Additionally, on the Haze4K [71] dataset for image dehazing, our model achieves a remarkable **18.99**× speedup over the Transformer-based MBT-L V2 [54].

E. Ablation Study

TABLE XVII: Ablation studies for the proposed components.

Net	MFFN	DLA									PSNR	FLOPs	Params
		Channel			Branch			Context prior					
		2	8	16	1	2	4	Imgage	Gauss	Pool			
a											27.64	28.30	3.35
b	✓										27.81	23.91	3.36
c	✓	✓			✓						29.29	21.33	3.42
d	✓		✓		✓						29.48	21.42	3.45
e	✓			✓	✓						29.43	21.54	3.49
f	✓		✓			✓					29.54	21.21	3.34
g	✓		✓				✓				29.49	21.10	3.29
h	✓		✓			✓		✓			29.66	21.36	3.35
i	✓		✓			✓			✓		29.70	21.36	3.35
j	✓		✓			✓				✓	29.74	21.36	3.35

We conduct ablation studies to evaluate the effectiveness of our design by training CDIR-S on the CDD-11 [19] dataset. To reduce computational cost, the model is trained for 100K iterations. Table XVII summarizes the ablation results.

The baseline model (Table XVII(a)) employs only Super-Branch 1 (Eq. (4)) in the attention block, which operates on full-resolution feature maps with doubled channel width for fairness, and adopts a standard gated feedforward network [29], achieving a PSNR of 27.64 dB on the CDD-11 dataset. Incorporating our efficient MFFN yields a 0.17 dB improvement while significantly reducing computational complexity (Table XVII(b)).

Channel dimension compression ratio. In our design, r in Eq. (9) is introduced to improve computational efficiency. To this end, we conduct experiments starting with a small value inspired by channel redundancy [109] and gradually increase it to identify a suitable r that achieves a better trade-off between performance and efficiency. Specifically, we train the one-branch model with different values of r. As shown in Table XVII(c-e), setting $r = 8$ achieves the best performance among the evaluated configurations. Based on this observation, we adopt $r = 8$ in our final model to balance performance and computational cost, while effectively leveraging channel redundancy.

The number of branches. In the proposed DLA module, we adopt a two-branch design within Super-Branch 2 (Eq. (6)) to better handle composite degradations. To assess the effect of the branch number on performance, we conduct experiments with varying branch counts while keeping the total number of channels constant. As shown in Table XVII(d,f,g), increasing the number of branches, particularly to four, slightly

Fig. 10: Detection results on degraded and restored images using YOLOv7 [107]. The first two examples are taken from composite degradation datasets, while the last two are from CSD [8], which targets single-degradation (desnowing) tasks.

Fig. 11: Qualitative evaluation on real-world night blurry images [18] using models pre-trained on LOLBlur [18]. The top example is taken from RealBlur-J [108], while the bottom example was captured in the wild using a Sony RX10 IV camera.

reduces performance. Although the four-branch design slightly reduces computational overhead through a mechanism similar to group convolution when the total channel count is fixed, the two-branch version achieves higher restoration quality by better coordinating different degradation characteristics and preserving the expressiveness of each branch with a larger number of channels. In addition, using fewer branches is more favorable for exploiting the parallelism provided by modern GPUs. Based on these results, we adopt the two-branch design in the final model.

Extracting high-frequency components in PG. We further study the effectiveness of extracting high-frequency components from input images for constructing context priors. Table XVII(h-j) shows that generating priors directly from degraded images yields 29.66 dB PSNR. Incorporating Gaussian operators to extract high-frequency components leads to a 0.04 dB improvement. Our proposed strategy achieves the best performance, with comparable computational overhead.

F. Downstream Applications

To evaluate the effectiveness of our model in supporting downstream tasks, we apply the YOLOv7 [107] detector to both the degraded inputs and the restored outputs. The detection results are presented in Figure 10. As shown, our restoration enhances object detection by enabling the identification of more instances and increasing detection confidence. Notably, in the second example, the detector successfully identifies the *bird* near the third person, although with an imprecise outline. In contrast, our restored result provides a more accurate object boundary but leads to misclassification, suggesting that a more advanced detection model may be required to fully exploit the benefits of image restoration.

Fig. 12: More visual results in real-world scenarios [64], [65].

G. Generalization Evaluation

To evaluate the generalization capability of our model, we apply the model pre-trained on the LOLBlur [18] dataset to real-world night blurry images. Qualitative results are presented in Figure 11. The first example, taken from the RealBlur-J [108] dataset, shows that our model more effectively restores sharper text within the red boxes. The second example, captured in the wild, further demonstrates the model's ability to recover clearer text on the billboard and better reconstruct structural details such as the window and road, as highlighted by the white boxes. These results highlight the strong generalization ability of our model and its practical potential in real-world applications.

V. Conclusion

This paper presents an efficient network for composite degradation image restoration. Specifically, we propose a LoRA-inspired attention block that operates in reduced channel and spatial dimensions to improve efficiency. Additionally, a multi-branch structure incorporating dynamic filters guided by local and bidirectional context priors is introduced to effectively handle complex degradations. To further enhance both restoration quality and computational efficiency, we design a multi-scale feedforward network. Extensive experiments on two composite degradation benchmarks validate the effectiveness of the proposed model. Moreover, it performs well on various single-degradation image restoration tasks, including dehazing, desnowing, and deraining. In addition, the model demonstrates strong adaptability to domain-specific scenarios such as medical, remote sensing, and UHD image restoration.

Despite the overall effectiveness of the proposed method, certain limitations remain in challenging degradation scenarios. Figure 12 presents additional visual results on real-world dehazing datasets [64], [65]. As illustrated, when haze is extremely dense or spatially non-uniform, the model may fail to fully recover fine-grained textures and subtle structural details. Potential directions to address these issues include collecting additional real-world data or leveraging the strong generalization capability of large-scale models.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant No. 62136001, U24B20175), Beijing Natural Science Foundation (Grant No. L233024), Beijing Municipal Science & Technology Commission, Administrative Commission of Zhongguancun Science Park (Grant No. Z241100003524012), and GBA Ascend Application Innovation Institute, Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ) (Grant No. GML-ST-2026-04).

References

[1] C.-L. Guo, Q. Yan, S. Anwar, R. Cong, W. Ren, and C. Li, "Image dehazing transformer with transmission-aware 3d position embedding," in *CVPR*, 2022.
[2] Z. Chen, Z. He, and Z.-M. Lu, "Dea-net: Single image dehazing based on detail-enhanced convolution and content-guided attention," *IEEE TIP*, 2024.
[3] R. Wang, Y. Zheng, Z. Zhang, C. Li, S. Liu, G. Zhai, and X. Liu, "Learning hazing to dehazing: Towards realistic haze generation for real-world image dehazing," in *CVPR*, 2025.
[4] X. Liu, Z. Shi, Z. Wu, J. Chen, and G. Zhai, "Griddehazenet+: An enhanced multi-scale network with intra-task knowledge transfer for single image dehazing," *IEEE TITS*, 2022.
[5] X. Chen, H. Li, M. Li, and J. Pan, "Learning a sparse transformer network for effective image deraining," in *CVPR*, 2023.
[6] Y. Cui, W. Ren, S. Yang, X. Cao, and A. Knoll, "Irnext: Rethinking convolutional network design for image restoration," in *ICML*, 2023.
[7] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," in *CVPR*, 2017.
[8] W.-T. Chen, H.-Y. Fang, C.-L. Hsieh, C.-C. Tsai, I. Chen, J.-J. Ding, S.-Y. Kuo *et al.*, "All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss," in *ICCV*, 2021.
[9] Y.-F. Liu, D.-W. Jaw, S.-C. Huang, and J.-N. Hwang, "Desnownet: Context-aware deep network for snow removal," *IEEE TIP*, 2018.
[10] Y. Cui, W. Ren, and A. Knoll, "Omni-kernel network for image restoration," in *AAAI*, 2024.
[11] S.-J. Cho, S.-W. Ji, J.-P. Hong, S.-W. Jung, and S.-J. Ko, "Rethinking coarse-to-fine approach in single image deblurring," in *ICCV*, 2021.
[12] F.-J. Tsai, Y.-T. Peng, Y.-Y. Lin, C.-C. Tsai, and C.-W. Lin, "Stripformer: Strip transformer for fast image deblurring," in *ECCV*, 2022.
[13] K. Zhang, W. Luo, Y. Zhong, L. Ma, B. Stenger, W. Liu, and H. Li, "Deblurring by realistic blurring," in *CVPR*, 2020.
[14] Y. Cui, Y. Tao, Z. Bing, W. Ren, X. Gao, X. Cao, K. Huang, and A. Knoll, "Selective frequency network for image restoration," in *ICLR*, 2023.
[15] D. Feijoo, J. C. Benito, A. Garcia, and M. V. Conde, "Darkir: Robust low-light image restoration," in *CVPR*, 2025.
[16] E. Zamfir, Z. Wu, N. Mehta, Y. Tan, D. P. Paudel, Y. Zhang, and R. Timofte, "Complexity experts are task-discriminative learners for any image restoration," in *CVPR*, 2025.
[17] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen *et al.*, "Lora: Low-rank adaptation of large language models." *ICLR*, 2022.
[18] S. Zhou, C. Li, and C. Change Loy, "Lednet: Joint low-light enhancement and deblurring in the dark," in *ECCV*, 2022.
[19] Y. Guo, Y. Gao, Y. Lu, H. Zhu, R. W. Liu, and S. He, "Onerestore: A universal restoration framework for composite degradation," in *ECCV*, 2024.
[20] W. Dong, H. Zhou, Y. Zhang, X. Liu, and J. Chen, "Ecmamba: Consolidating selective state space model with retinex guidance for efficient multiple exposure correction," *NeurIPS*, 2024.
[21] H. Zhou, W. Dong, X. Liu, Y. Zhang, G. Zhai, and J. Chen, "Low-light image enhancement via generative perceptual priors," in *AAAI*, 2025.
[22] Q. Zhang, W. Xin, S. Wu, Q. Zhu, Q. Song, and S. Zhang, "A cluster tree network for image super-resolution," *IEEE Transactions on Consumer Electronics*, 2025.
[23] H. Zhou, W. Dong, X. Liu, S. Liu, X. Min, G. Zhai, and J. Chen, "Glare: Low light image enhancement via generative latent feature based codebook retrieval," in *ECCV*, 2024.
[24] R. Wang, W. Li, X. Liu, C. Li, Z. Zhang, X. Min, and G. Zhai, "Hazeclip: Towards language guided real-world image dehazing," in *ICASSP*, 2025.
[25] T. Hu, L. Wu, W. Dong, P. Wu, J. Sun, X. Xu, Q. Yan, and Y. Zhang, "Boosting hdr image reconstruction via semantic knowledge transfer," *IEEE TIP*, 2026.

[26] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE TPAMI*, 2010.

[27] K. I. Kim and Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior," *TPAMI*, 2010.

[28] Y. Song, Z. He, H. Qian, and X. Du, "Vision transformers for single image dehazing," *IEEE TIP*, 2023.

[29] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *CVPR*, 2022.

[30] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, "Uformer: A general u-shaped transformer for image restoration," in *CVPR*, 2022.

[31] X. Qin, Y. Quan, Z. Chen, and H. Ji, "Robust unsupervised deep learning for nonblind image deconvolution with inaccurate kernels," *IEEE TNNLS*, 2025.

[32] Q. Yan, L. Zhang, Y. Liu, Y. Zhu, J. Sun, Q. Shi, and Y. Zhang, "Deep hdr imaging via a non-local network," *IEEE TIP*, 2020.

[33] H. Guo, J. Li, T. Dai, Z. Ouyang, X. Ren, and S.-T. Xia, "Mambair: A simple baseline for image restoration with state-space model," in *ECCV*, 2024.

[34] D. Li, Y. Liu, X. Fu, S. Xu, and Z.-J. Zha, "Fouriermamba: Fourier learning integration with state space models for image deraining," *arXiv preprint arXiv:2405.19450*, 2024.

[35] J. Weng, Z. Yan, Y. Tai, J. Qian, J. Yang, and J. Li, "MambaLLIE: Implicit retinex-aware low light enhancement with global-then-local state space," in *NeurIPS*, 2024.

[36] K. Jiang, J. Jiang, X. Liu, H. Yao, and C.-W. Lin, "Ph-mamba: Enhancing mamba with position encoding and harmonized attention for image deraining and beyond," *IEEE TIP*, 2025.

[37] Z. Chen, Y. Zhang, D. Liu, J. Gu, L. Kong, and X. Yuan, "Hierarchical integration diffusion model for realistic image deblurring," in *NeurIPS*, 2023.

[38] Q. Yan, T. Hu, P. Wu, D. Dai, S. Gu, W. Dong, and Y. Zhang, "Efficient image enhancement with a diffusion-based frequency prior," *IEEE TCSVT*, 2025.

[39] X. Li, Y. Ren, X. Jin, C. Lan, X. Wang, W. Zeng, X. Wang, and Z. Chen, "Diffusion models for image restoration and enhancement: A comprehensive survey," *IJCV*, 2025.

[40] C. Korkmaz, N. Mehta, and R. Timofte, "Adaptsr: Low-rank adaptation for efficient and scalable real-world super-resolution," *arXiv preprint arXiv:2503.07748*, 2025.

[41] L. Sun, R. Wu, Z. Ma, S. Liu, Q. Yi, and L. Zhang, "Pixel-level and semantic-level adjustable super-resolution: A dual-lora approach," in *CVPR*, 2025.

[42] J. Jiang, Z. Zuo, G. Wu, K. Jiang, and X. Liu, "A survey on all-in-one image restoration: Taxonomy, evaluation and future trends," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

[43] V. Potlapalli, S. W. Zamir, S. H. Khan, and F. Shahbaz Khan, "Promptir: Prompting for all-in-one image restoration," in *NeurIPS*, 2023.

[44] Y. Ai, H. Huang, X. Zhou, J. Wang, and R. He, "Multimodal prompt perceiver: Empower adaptiveness generalizability and fidelity for all-in-one image restoration," in *CVPR*, 2024.

[45] D. Park, B. H. Lee, and S. Y. Chun, "All-in-one image restoration for unknown degradations using adaptive discriminative filters for specific degradations," in *CVPR*, 2023.

[46] W. Zou, H. Gao, T. Ye, L. Chen, W. Yang, S. Huang, H. Chen, and S. Chen, "Vqcnir: Clearer night image restoration with vector-quantized codebook," in *AAAI*, 2024.

[47] X. Ma, K. Shi, D. Chen, Y. Cao, Q. Yan, and Y. Zhang, "A unified framework for deblurring and hdr imaging in dynamic scenes," *IEEE Journal of Selected Topics in Signal Processing*, 2025.

[48] B. Li, X. Liu, P. Hu, Z. Wu, J. Lv, and X. Peng, "All-in-one image restoration for unknown corruption," in *CVPR*, 2022.

[49] M. V. Conde, G. Geigle, and R. Timofte, "High-quality image restoration following human instructions," in *ECCV*, 2024.

[50] Y. Cui, S. W. Zamir, S. Khan, A. Knoll, M. Shah, and F. S. Khan, "AdaIR: Adaptive all-in-one image restoration via frequency mining and modulation," in *ICLR*, 2025.

[51] Y. Rao, W. Zhao, Y. Tang, J. Zhou, S. N. Lim, and J. Lu, "Hornet: Efficient high-order spatial interactions with recursive gated convolutions," in *NeurIPS*, 2022.

[52] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2net: A new multi-scale backbone architecture," *IEEE TPAMI*, 2019.

[53] Y. Cui, W. Ren, X. Cao, and A. Knoll, "Revitalizing convolutional network for image restoration," *IEEE TPAMI*, 2024.

[54] Z. Jin, Y. Qiu, K. Zhang, H. Li, and W. Luo, "Mb-taylorformer v2: Improved multi-branch linear transformer expanded by taylor formula for image restoration," *IEEE TPAMI*, 2025.

[55] L. Chen, X. Chu, X. Zhang, and J. Sun, "Simple baselines for image restoration," in *ECCV*, 2022.

[56] Y. Cai, H. Bian, J. Lin, H. Wang, R. Timofte, and Y. Zhang, "Retinexformer: One-stage retinex-based transformer for low-light image enhancement," in *ICCV*, 2023.

[57] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Learning enriched features for fast image restoration and enhancement," *IEEE TPAMI*, 2022.

[58] B. Song, X. Chen, S. Xu, and J. Zhou, "Under-display camera image restoration with scattering effect," in *ICCV*, 2023.

[59] O. Özdenizci and R. Legenstein, "Restoring vision in adverse weather conditions with patch-based denoising diffusion models," *IEEE TPAMI*, 2023.

[60] Y. Zhu, T. Wang, X. Fu, X. Yang, X. Guo, J. Dai, Y. Qiao, and X. Hu, "Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions," in *CVPR*, 2023.

[61] W.-T. Chen, H.-Y. Fang, J.-J. Ding, C.-C. Tsai, and S.-Y. Kuo, "Jstasr: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal," in *ECCV*, 2020.

[62] B. Cheng, J. Li, Y. Chen, and T. Zeng, "Snow mask guided adaptive residual network for image snow removal," *CVIU*, 2023.

[63] Y. Cui, W. Ren, X. Cao, and A. Knoll, "Focal network for image restoration," in *ICCV*, 2023.

[64] C. O. Ancuti, C. Ancuti, M. Sbert, and R. Timofte, "Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images," in *ICIP*, 2019.

[65] C. O. Ancuti, C. Ancuti, and R. Timofte, "Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images," in *CVPRW*, 2020.

[66] X. Liu, Y. Ma, Z. Shi, and J. Chen, "Griddehazenet: Attention-based multi-scale network for image dehazing," in *ICCV*, 2019.

[67] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, and M.-H. Yang, "Multi-scale boosted dehazing network with dense feature fusion," in *CVPR*, 2020.

[68] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "Ffa-net: Feature fusion attention network for single image dehazing," in *AAAI*, 2020.

[69] Y. Zheng, J. Zhan, S. He, J. Dong, and Y. Du, "Curricular contrastive regularization for physics-aware single image dehazing," in *CVPR*, 2023.

[70] Y. Qiu, K. Zhang, C. Wang, W. Luo, H. Li, and Z. Jin, "Mb-taylorformer: Multi-branch efficient transformer expanded by taylor formula for image dehazing," in *ICCV*, 2023.

[71] Y. Liu, L. Zhu, S. Pei, H. Fu, J. Qin, Q. Zhang, L. Wan, and W. Feng, "From synthetic to real: Image dehazing collaborating with unlabeled real data," in *ACM MM*, 2021.

[72] T. Ye, Y. Zhang, M. Jiang, L. Chen, Y. Liu, S. Chen, and E. Chen, "Perceiving and modeling density for image dehazing," in *ECCV*, 2022.

[73] Y. Cui, W. Ren, X. Cao, and A. Knoll, "Image restoration via frequency selection," *IEEE TPAMI*, 2023.

[74] C. Wang, J. Pan, W. Wang, G. Fu, S. Liang, M. Wang, X.-M. Wu, and J. Liu, "Correlation matching transformation transformers for uhd image restoration," in *AAAI*, 2024.

[75] L. Kong, J. Dong, J. Ge, M. Li, and J. Pan, "Efficient frequency domain-based transformers for high-quality image deblurring," in *CVPR*, 2023.

[76] L. Wang, C. Wang, J. Pan, W. Zhou, X. Sun, W. Wang, and Z. Su, "Ultra-high-definition restoration: New benchmarks and a dual interaction prior-driven solution," *arXiv e-prints*, pp. arXiv–2406, 2024.

[77] C. Zhao, Z. Chen, Y. Xu, E. Gu, J. Li, Z. Yi, Q. Wang, J. Yang, and Y. Tai, "From zero to detail: Deconstructing ultra-high-definition image restoration from progressive spectral perspective," in *CVPR*, 2025.

[78] C. Li, C.-L. Guo, man zhou, Z. Liang, S. Zhou, R. Feng, and C. C. Loy, "Embedding fourier for ultra-high-definition low-light image enhancement," in *ICLR*, 2023.

[79] T. Wang, K. Zhang, T. Shen, W. Luo, B. Stenger, and T. Lu, "Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method," in *AAAI*, 2023.

[80] W. Zou, H. Gao, W. Yang, and T. Liu, "Wave-mamba: Wavelet state space model for ultra-high-definition low-light image enhancement," in *ACM MM*, 2024.

[81] H. Chen, X. Chen, C. Wu, Z. Zheng, J. Pan, and X. Fu, "Towards ultra-high-definition image deraining: A benchmark and an efficient method," *arXiv preprint arXiv:2405.17074*, 2024.
[82] S. Chen, T. Ye, J. Bai, E. Chen, J. Shi, and L. Zhu, "Sparse sampling transformer with uncertainty-driven ranking for unified removal of raindrops and rain streaks," in *ICCV*, 2023.
[83] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, "Attentive generative adversarial network for raindrop removal from a single image," in *CVPR*, 2018.
[84] S. Zhou, D. Chen, J. Pan, J. Shi, and J. Yang, "Adapt or perish: Adaptive sparse transformer with attentive feature refinement for image restoration," in *CVPR*, 2024.
[85] S. Zhou, J. Pan, J. Shi, D. Chen, L. Qu, and J. Yang, "Seeing the unseen: A frequency prompt guided transformer for image restoration," in *ECCV*, 2024.
[86] Z. Tu, H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. Bovik, and Y. Li, "Maxim: Multi-axis mlp for image processing," in *CVPR*, 2022.
[87] T. Ye, S. Chen, J. Bai, J. Shi, C. Xue, J. Jiang, J. Yin, E. Chen, and Y. Liu, "Adverse weather removal with codebook priors," in *ICCV*, 2023.
[88] Z. Yang, J. Li, H. Zhang, D. Zhao, B. Wei, and Y. Xu, "Restore-rwkv: Efficient and effective medical image restoration with rwkv," *IEEE Journal of Biomedical and Health Informatics*, 2025.
[89] "Ixi dataset," 2023, accessed: 2025-08-02. [Online]. Available: http://braindevelopment.org/ixi-dataset/
[90] L. Chen, X. Yang, G. Jeon, M. Anisetti, and K. Liu, "A trusted medical image super-resolution method based on feedback adaptive weighted dense network," *Artificial Intelligence in Medicine*, 2020.
[91] J. Huang, Y. Fang, Y. Wu, H. Wu, Z. Gao, Y. Li, J. Del Ser, J. Xia, and G. Yang, "Swin transformer for fast mri," *Neurocomputing*, 2022.
[92] J. Huang, X. Xing, Z. Gao, and G. Yang, "Swin deformable attention u-net transformer (sdaut) for explainable fast mri," in *MICCAI*, 2022.
[93] H. Sun, Y. Li, Z. Li, R. Yang, Z. Xu, J. Dou, H. Qi, and H. Chen, "Fourier convolution block with global receptive field for mri reconstruction," *Medical Image Analysis*, 2025.
[94] L. Zhou, J. D. Schaefferkoetter, I. W. Tham, G. Huang, and J. Yan, "Supervised learning with cyclegan for low-dose fdg pet image denoising," *Medical image analysis*, 2020.
[95] Y. Zhou, Z. Yang, H. Zhang, E. I.-C. Chang, Y. Fan, and Y. Xu, "3d segmentation guided style-based generative adversarial networks for pet synthesis," *IEEE Transactions on Medical Imaging*, 2022.
[96] Z. Yang, Y. Zhou, H. Zhang, B. Wei, Y. Fan, and Y. Xu, "Drmc: A generalist model with dynamic routing for multi-center pet image synthesis," in *MICCAI*, 2023.
[97] Y. Luo, L. Zhou, B. Zhan, Y. Fei, J. Zhou, Y. Wang, and D. Shen, "Adaptive rectification based adversarial network with spectrum constraint for high-quality pet image synthesis," *Medical image analysis*, 2022.
[98] C. H. McCollough, A. C. Bartley, R. E. Carter, B. Chen, T. A. Drees, P. Edwards, D. R. Holmes III, A. E. Huang, F. Khan, S. Leng *et al.*, "Low-dose ct for the detection and classification of metastatic liver lesions: results of the 2016 low dose ct grand challenge," *Medical physics*, 2017.
[99] Z. Zhang, L. Yu, X. Liang, W. Zhao, and L. Xing, "Transct: dual-path transformer for low dose computed tomography," in *MICCAI*, 2021.
[100] A. Luthra, H. Sulakhe, T. Mittal, A. Iyer, and S. Yadav, "Eformer: Edge enhancement based transformer for medical image denoising," *arXiv preprint arXiv:2109.08044*, 2021.
[101] D. Wang, F. Fan, Z. Wu, R. Liu, F. Wang, and H. Yu, "Ctformer: convolution-free token2token dilated vision transformer for low-dose ct denoising," *Physics in Medicine & Biology*, 2023.
[102] Ş. Öztürk, O. C. Duran, and T. Çukur, "Denomamba: A fused state-space model for low-dose ct denoising," *arXiv preprint arXiv:2409.13094*, 2024.
[103] B. Huang, L. Zhi, C. Yang, F. Sun, and Y. Song, "Single satellite optical imagery dehazing using sar image prior based on conditional generative adversarial networks," in *WACV*, 2020.
[104] Y. Wen, T. Gao, J. Zhang, Z. Li, and T. Chen, "Encoder-free multiaxis physics-aware fusion network for remote sensing image dehazing," *IEEE TGRS*, 2023.
[105] K. Chi, Y. Yuan, and Q. Wang, "Trinity-net: Gradient-guided swin transformer-based remote sensing image dehazing and beyond," *IEEE TGRS*, 2023.
[106] X. Luan, H. Fan, Q. Wang, N. Yang, S. Liu, X. Li, and Y. Tang, "Fmambair: A hybrid state space model and frequency domain for image restoration," *IEEE TGRS*, 2025.
[107] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *CVPR*, 2023.
[108] J. Rim, H. Lee, J. Won, and S. Cho, "Real-world blur dataset for learning and benchmarking deblurring algorithms," in *ECCV*, 2020.
[109] Z. Chen, Y. Zhang, J. Gu, L. Kong, X. Yang, and F. Yu, "Dual aggregation transformer for image super-resolution," in *ICCV*, 2023.