Seeing the Unseen: A Frequency Prompt Guided Transformer for Image Restoration

Shihao Zhou[1,2], Jinshan Pan[3], Jinglei Shi[1⋆], Duosheng Chen[1], Lishen Qu[1], and Jufeng Yang[1,2]

[1] VCIP & TMCC & DISSec, College of Computer Science, Nankai University
[2] Nankai International Advanced Research Institute (SHENZHEN· FUTIAN)
[3] School of Computer Science and Engineering, Nanjing University of Science and Technology

{zhoushihao96,duoshengchen,lishenqu}@mail.nankai.edu.cn,
{jinglei.shi,yangjufeng}@nankai.edu.cn,sdluran@gmail.com

Abstract. How to explore useful features from images as prompts to guide the deep image restoration models is an effective way to solve image restoration. In contrast to mining spatial relations within images as prompt, which leads to characteristics of different frequencies being neglected and further remaining subtle or undetectable artifacts in the restored image, we develop a **F**requency **Pro**mpting image restoration method, dubbed **FPro**, which can effectively provide prompt components from a frequency perspective to guild the restoration model address these differences. Specifically, we first decompose input features into separate frequency parts via dynamically learned filters, where we introduce a gating mechanism for suppressing the less informative elements within the kernels. To propagate useful frequency information as prompt, we then propose a dual prompt block, consisting of a low-frequency prompt modulator (LPM) and a high-frequency prompt modulator (HPM), to handle signals from different bands respectively. Each modulator contains a generation process to incorporate prompting components into the extracted frequency maps, and a modulation part that modifies the prompt feature with the guidance of the decoder features. Experimental results on several popular datasets have demonstrated the favorable performance of our pipeline against SOTA methods on 5 image restoration tasks, including deraining, deraindrop, demoiréing, deblurring, and dehazing. The source code is available at https://github.com/joshyZhou/FPro.

Keywords: Image Restoration · Frequency Prompt · Transformer

1 Introduction

Capturing images in unsatisfactory environments, e.g., rain, haze, usually leads to low-quality ones that accordingly affect the application of downstream tasks in practice. Thus, developing an effective image restoration method to restore clear images from degraded ones is an important task.

⋆ Corresponding author.

Significant progress has been made due to kinds of the deep learning models [4,5,81], and these deep learning-based approaches become predominant ones as they achieve better performance than the conventional hand-crafted prior-based approaches [3,21,31,44,84].

Existing methods, e.g., [33,72,81] achieve promising performances in kinds of image restoration tasks. However, these learning-based methods intend to learn a mapping function between degraded images and clear ones, where the characteristics of the specific degradation are less considered. For example, rain streaks tend to obscure the background partially, whereas raindrops typically result in a more pronounced regional occlusion. Accordingly, these models are hindered from generating better results.

More recently, prompt-learning based methods [50,67,68] serve as an alternative approach to encode useful content of specific degradation for modulating the network, and make a clear performance boost for image restoration. However, we notice that these methods [50,68] pay attention to mining spatial correlations to provide degradation information, whereas the task-specific frequency cues are less studied. Indeed, since various forms of degradation exhibit distinct impacts on image content, they affect information from different frequency bands. Hence, it is crucial to develop an efficient prompt mechanism that explores useful prompts from a frequency perspective for identifying specific characteristics of diverse degradation, which can boost the model to effectively restore images with finer details and non-local structures of the scenes.

This paper proposes a **F**requency **Pro**mpting image restoration method, dubbed **FPro**, to modulate the network by encoding degradation-specific frequency cues as prompts. As mentioned above, existing prompt strategies [50,68] focus on mining spatial relations as useful prompts. In this way, differences between the restored image and the real one within frequency domain [25] are ignored, which remain subtle or undetectable artifacts in the spatial domain. Instead, our FPro aims to enjoy benefits from the capability of prompt learning in different frequency bands at multi-scale resolutions to recover clean images.

We present two designs to enhance FPro for general image restoration: 1). We first decouple input features into separate low-/high-frequency parts using a gated dynamic decoupler, as signals in different frequency bands encode image patterns from distinct views, *i.e.*, local details and global structures. To this end, a gating mechanism is introduced to help learn the enhanced low-pass filters by suppressing the less informative elements within the kernel, which are then employed to generate low-frequency maps. Meanwhile, the corresponding high-pass filter is obtained by subtracting the low-pass filter from the identity kernel, for generating high-frequency maps. 2). We propose a Dual Prompt Block (DPB), which consists of two modulators, *i.e.*, the Low-frequency Prompt Modulator (LPM) and the High-frequency Prompt Modulator (HPM), to handle low- and high-frequency information respectively. Each modulator includes (a) a generation part that incorporates prompting components into the extracted frequency maps, which is supposed to help distinguish various elements within features, such as rain patterns in the context of deraining; and (b) a modulation part that modifies the prompt feature with the guidance of the feature in the restoration process. In terms of functionality, LPM enhances the low-frequency characteristics through a gating mechanism in the Fourier domain before injecting the prompting components, which is proven equivalent to dynamic large-kernel depth-wise convolution in the spatial domain while computationally efficient, and then encodes low-frequency interactions via global cross-attention. As a complement, HPM applies a locally-enhanced gating mechanism to obtain useful high-frequency signals, and then encodes high-frequency interactions via local cross-attention.

We summarize our main contributions as follows:

– We propose FPro, which benefits from prompting learning of frequency components for general image restoration. Instead of mining spatial relations as in previous methods, we explore frequency maps to encode specific degradation information as prompts to guide the image restoration model for restoring finer details and the global structure of the scenes.
– We decouple input features into different frequency bands using learnable low-pass filters, and propose a dual prompt block, which is composed of low-frequency prompt modulator (LPM) and high-frequency prompt modulator (HPM), to explore both details and structures for better restoration.
– Experimental results on 5 restoration tasks: deraining, deraindrop, demoiréing, deblurring and dehazing, show that FPro achieves favorable performance against state-of-the-art methods.

2 Related Work

Image Restoration. The goal of the image restoration task is to recover high-quality images from the degraded ones. Going beyond conventional prior-based solutions [3, 21], this community has witnessed the great success of a body of learning-based approaches [34, 48, 79]. Despite the promising results obtained by various CNN-based architectures [9, 35, 58], the main concern for methods of this kind is that they pose a limited receptive field problem of the basic convolution operation. This means that the feature map contains less global context (corresponding to low-frequency characteristics in an image), and the final prediction can get stuck in this limitation. This drawback has motivated the increased interest in exploring components to capture desired global cues, like attention mechanisms [10, 45, 60], where better restoration performance can be achieved. For instance, MIRNet [82] proposes a dual attention unit to capture contextual information in dual dimensions. NLSN [41] employs a self-attention mechanism to collect global correlation information for super-resolution.

Transformer-based Restoration. In recent years, the idea of using Transformer architecture [64] to address various computer vision-related tasks has been popular. Thanks to their discriminative feature representation capability, they not only earn advantages in solving high-level vision tasks [13, 14, 71], but also are extended to low-level image restoration tasks [27, 85, 88]. Unfortunately, as the complexity of vanilla self-attention is quadratic to the image size, this mechanism suffers from non-trivial computational costs in handling high-resolution input. To address this, some attempts have been made to explore efficient transformer architectures [4, 72, 87]. Specifically, SwinIR [33] introduces a window-based self-attention scheme to improve efficiency. Restormer [81] adopts channel-wise self-attention to reduce the computational costs. The majority of these works have offered reliable solutions to recover clean images, however, some works [11,49] realized that the low-pass filter nature of self-attention, which could lose the high-frequency information, such as textures and edges. Even though these models have achieved superior performance, few high-frequency details can be leveraged to implement image restoration, limiting better recovery as a result.

Visual Prompt Learning. Prompt learning, which emerged recently in the NLP field [1], has resulted in rapid advancements in its adaptation to computer vision tasks [17,24,26]. Contrary to high-level vision problems, motivated by high effectiveness, some works also consider seeking proper prompts for the low-level pipelines [40, 74, 78].

The aim of this paper is not to prompt models for addressing the ALL-in-One problem (in fact, the previous works of [32,40,50] have addressed this nicely by designing various degradation prompt modules). However, our approach is relevant to recent studies [67, 68] exploring degradation-specific information for better image restoration results. In contrast to these attempts that generate raw degradation features with a pre-trained model, we prompt the restoration models from a frequency perspective.

3 Proposed Method

3.1 Overall Pipeline

As depicted in Fig. 1, the overview of our proposed FPro contains the upper restoration branch, like existing works [33, 81], and the bottom prompt branch to extract informative frequency maps and then modulate them as prompts.

Restoration Branch. Given a degraded image $\mathbf{I} \in \mathbb{R}^{H\times W\times 3}$ as input, FPro first employs a convolution layer to obtain shallow feature $\mathbf{F}_s \in \mathbb{R}^{H\times W\times C}$; where $H\times W$ represents the spatial dimension and C is the number of channel. Next, the shallow feature passes through the upper N_1-level encoder-decoder restoration branch to extract deep feature $\mathbf{F}_d \in \mathbb{R}^{H\times W\times C}$. Early layers in Transformer-based models focus on aggregate local patterns [76], whereas the self-attention module acts as a low-pass filter and tends to dilute high-frequency local details [49]. To alleviate the two contradictory factors, we remove the attention mechanism within the encoder of the restoration branch. Specifically, each level of the encoder includes N_2 feed-forward network (FFN) [81] and the paired convolution layer for down-sampling. The encoder features are then fused with the ones in the decoder through skip connections by 1×1 convolution. For the decoder part, each level is composed of N_2 pairs of FFN and multi-head self-attention mechanisms (MSA) [81], along with the convolution layer for up-sampling. Finally, a 3×3 convolution layer is employed to deep feature $\mathbf{F}_d$ for generating residual image $\mathbf{R} \in \mathbb{R}^{H\times W\times 3}$. The restored image $\hat{\mathbf{I}}$ is estimated by: $\hat{\mathbf{I}} = \mathbf{I} + \mathbf{R}$.

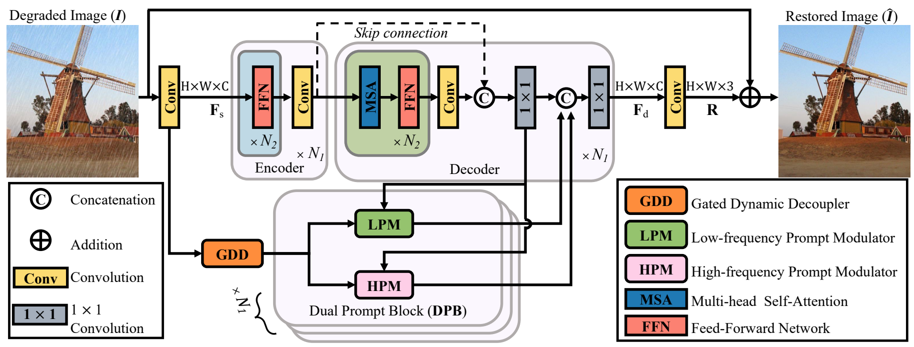

Fig. 1: Overview of the proposed **FPro**. Except for the common upper restoration branch, which is similar to existing methods [33, 81], FPro contains another bottom prompt branch to extract informative features from a frequency perspective. Specifically, the primary components of the prompt branch in this framework are the gated dynamic decoupler (GDD) and dual prompt block (DPB). The GDD is employed to decompose the low-frequency components and corresponding high-frequency characteristics from the input features. Then these frequency-specific features are further processed in DPB, *i.e.*, the high-frequency prompt modulator (HPM) and the low-frequency prompt modulator (LPM), which generates representative frequency prompt to facilitate the clear image reconstruction.

Prompt Branch. In this branch, we take as input the shallow feature $\mathbf{F}_s$ to generate useful frequency prompts, which are further leveraged to facilitate the latent clear image reconstruction. To achieve this goal, we first decompose the input feature into different frequency bands using a gated dynamic decouple (GDD) (see Section 3.2). After that, low-/high-frequency maps are injected with prompt components to distinguish informative elements according to specific tasks, and then modulated as different prompts ($i.e.$, $\mathbf{F}_{hi}^{out}$ and $\mathbf{F}_{low}^{out}$) to interact with the decoder features by 1×1 convolution (see Section 3.3). Next, we present the modules of the prompt branch.

3.2 Gated Dynamic Decoupler

Each type of degradation affects image content in different ways. For instance, rain streaks partially occlude the background while raindrops often cause much greater obstruction, which corresponds to touch high-/low-frequency bands respectively. To handle these differences, as shown in Fig. 2, we decompose the input features into separate frequency parts based on gated and dynamically learned filters. The key ingredient is to introduce a gating mechanism to help generate the gated learnable low-pass filter and the corresponding high-pass filter, which are then employed to obtain low- and high-frequency maps, respectively. These filters are dynamically learned for each spatial location and channel group to balance computation burden and feature diversity. Specifically, given the input shallow feature map $\mathbf{F}_s \in \mathbb{R}^{H\times W\times C}$, we firstly predicts the low-pass

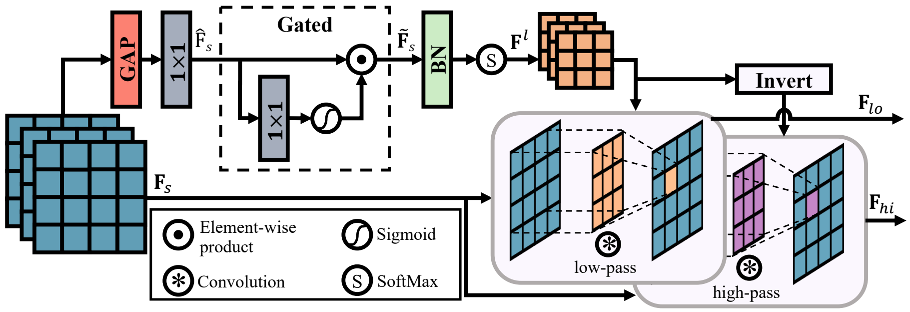

Fig. 2: Illustrations of the Gated Dynamic Decoupler.

filter for each feature channel group, which can be formulated as:

$$\begin{aligned}\hat{\mathbf{F}}_s &= \mathrm{Conv}_{1\times 1}(\mathrm{GAP}(\mathbf{F}_s)),\\ \tilde{\mathbf{F}}_s &= \hat{\mathbf{F}}_s \odot \phi(\mathrm{Conv}_{1\times 1}(\hat{\mathbf{F}}_s)),\\ \mathbf{F}^l &= \mathrm{Softmax}(\mathcal{B}(\tilde{\mathbf{F}}_s))\end{aligned} \tag{1}$$

where $\mathbf{F}^l \in \mathbb{R}^{g\times k^2 \times 1\times 1}$, g is the number of channel groups and k^2 corresponds to the kernel size of the learned filter; $\mathrm{GAP}(\cdot)$ and $\mathrm{Conv}_{1\times 1}(\cdot)$ are global average pooling layer and convolution operation with the filter size of 1×1, respectively; $\phi(\cdot)$ denotes sigmoid activation, $\odot$ refers to the element-wise product, and $\mathcal{B}(\cdot)$ means Batch Normalization. Particularly, $\mathrm{Softmax}(\cdot)$ is a softmax layer, which ensures the generated filters are low-pass [89]. Then, we apply these learned filters to each group input feature $\mathbf{F}_i \in \mathbb{R}^{H\times W\times C_i}$ to obtain low-frequency components:

$$\mathbf{F}^{lo}_{i,c,h,w} = \sum_{p,q} \mathbf{F}^{L}_{i,p,q} \mathbf{F}_{i,c,h+p,w+q}, \tag{2}$$

where $\mathbf{F}^L \in \mathbb{R}^{g\times k\times k}$ is the reshaped filter, i denotes the group index, $C_i = \frac{C}{g}$ refers to number of the group channel, c means the index of a channel, h and w are spatial coordinates, $p, q \in \{-1, 0, 1\}$ point to the surrounding locations.

Meanwhile, we invert this process by subtracting the low-pass filter from the identity kernel to attain the high-pass filter, which is employed to generate the corresponding high-frequency components $\mathbf{F}_{hi}$.

3.3 Dual Prompt Block

Considering that the extracted features, *i.e.*, low-/high-frequency maps, encode image patterns from distinct views (local detail and main structure of the image). We design the Dual Prompt Block that includes two components, *i.e.*, High-frequency Prompt Modulator (HPM) and Low-frequency Prompt Modulator (LPM), to deal with these feature maps, respectively.

High-frequency Prompt Modulator. Given the two input feature maps, including the l-level feature $\mathbf{F}_l \in \mathbb{R}^{\hat{H}\times\hat{W}\times\hat{C}}$ and high-frequency feature $\mathbf{F}_{hi} \in$

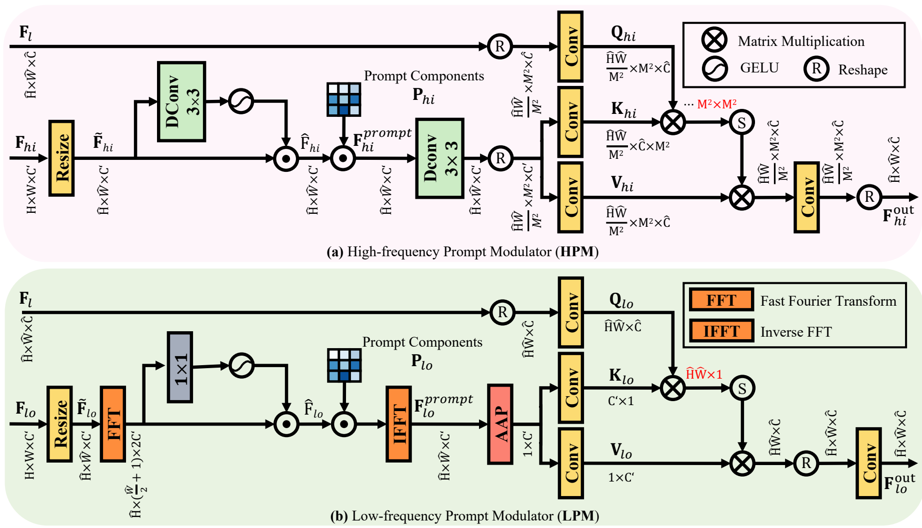

Fig. 3: Illustrations of the proposed components. (a) High-frequency Prompt Modulator (**HPM**); (b) Low-frequency Prompt Modulator (**LPM**).

$\mathbb{R}^{H\times W\times C'}$, we first resize $\mathbf{F}_{hi}$ and obtain $\tilde{\mathbf{F}}_{hi} \in \mathbb{R}^{\hat{H}\times\hat{W}\times C'}$. Towards highlighting high-frequency characteristics, we employ a gating mechanism to adaptively determine the useful frequency information:

$$\hat{\mathbf{F}}_{hi} = \tilde{\mathbf{F}}_{hi} \odot \sigma(\mathrm{DConv}_{3\times3}(\tilde{\mathbf{F}}_{hi})), \tag{3}$$

where $\hat{\mathbf{F}}_{hi} \in \mathbb{R}^{\hat{H}\times\hat{W}\times C'}$ is the processed feature, $\mathrm{DConv}_{3\times3}(\cdot)$ denotes a depth-wise convolution operation with the filter size of 3×3, and $\sigma(\cdot)$ is the GELU activation function [22]. Then, we leverage the learnable high-frequency prompt components $\mathbf{P}_{hi} \in \mathbb{R}^{\hat{H}\times\hat{W}\times C'}$ to make adjustments to the input features, which aims to help distinguish various elements, such as rain patterns and streaks of different orientations and magnitudes in the context of deraining:

$$\mathbf{F}_{hi}^{prompt} = \hat{\mathbf{F}}_{hi} \odot \mathbf{P}_{hi}, \tag{4}$$

where $\mathbf{F}_{hi}^{prompt} \in \mathbb{R}^{\hat{H}\times\hat{W}\times C'}$ is the obtained high-frequency feature prompt.

Next, we modify the high-frequency prompt $\mathbf{F}_{hi}^{prompt}$ according to the input feature $\mathbf{F}_l$. To be specific, we utilize a depth-wise convolution operator, which acts as a high-pass filter [49], to enhance the high-frequency sources in the input $\mathbf{F}_{hi}^{prompt}$. Then, we generate *query* ($\mathbf{Q}_{hi}$) projection from $\mathbf{F}_l$, *key* ($\mathbf{K}_{hi}$) and *value* ($\mathbf{V}_{hi}$) projections from the processed feature map $\hat{\mathbf{F}}_{hi}^{prompt} = \mathrm{DConv}_{3\times3}(\mathbf{F}_{hi}^{prompt})$, respectively. Meanwhile, as the high-frequency information usually corresponds to image details and is a local feature, it could be redundant to calculate global attention. Therefore, before leveraging the linear layer to obtain the matrices of $\mathbf{Q}_{hi}$, $\mathbf{K}_{hi}$, and $\mathbf{V}_{hi}$, the local window self-attention mechanism is adopted to save computational complexity and capture fine-grained high frequencies, which yields $\mathbf{Q}_{hi} = W_p^{Q_{hi}} \cdot R(\mathbf{F}_l)$, $\mathbf{K}_{hi} = W_p^{K_{hi}} \cdot R(\hat{\mathbf{F}}_{hi}^{prompt})$, $\mathbf{V}_{hi} = W_p^{V_{hi}} \cdot R(\hat{\mathbf{F}}_{hi}^{prompt})$. Where $W_p^{(\cdot)}$ represents the projection matrices, and $R(\cdot)$ denotes the window partition strategy [39]. Generally, we have $\mathbf{Q}_{hi} \in \mathbb{R}^{\frac{\hat{H}\hat{W}}{M^2} \times M^2 \times \hat{C}}$, $\mathbf{K}_{hi} \in \mathbb{R}^{\frac{\hat{H}\hat{W}}{M^2} \times \hat{C} \times M^2}$, and $\mathbf{V}_{hi} \in \mathbb{R}^{\frac{\hat{H}\hat{W}}{M^2} \times M^2 \times \hat{C}}$, where M^2 is the size of split windows. The attention matrix is thus calculated to tune the high-frequency prompt:

$$\mathbf{F}_{hi}^{out} = \mathbf{V}_{hi} \cdot \mathrm{Softmax}(\mathbf{K}_{hi} \cdot \mathbf{Q}_{hi} / \sqrt{d}), \tag{5}$$

where $\mathbf{F}_{hi}^{out} \in \mathbb{R}^{\hat{H} \times \hat{W} \times \hat{C}}$ is the output feature map of the high-frequency prompt modulation branch; d is the query/key dimension, following [33].

Low-frequency Prompt Modulator. Given the two input feature maps, including the l-level feature $\mathbf{F}_l \in \mathbb{R}^{\hat{H} \times \hat{W} \times \hat{C}}$ and low-frequency feature $\mathbf{F}_{lo} \in \mathbb{R}^{H \times W \times C'}$, we first resize $\mathbf{F}_{lo}$ and obtain $\tilde{\mathbf{F}}_{lo} \in \mathbb{R}^{\hat{H} \times \hat{W} \times C'}$. Towards handling low-frequency signals effectively, we project $\tilde{\mathbf{F}}_{lo}$ into the frequency domain via the fast Fourier transform (FFT). Then, a gating mechanism is adopted to control the useful low-frequency components flow forward:

$$\hat{\mathbf{F}}_{lo} = \mathcal{F}(\tilde{\mathbf{F}}_{lo}) \odot \sigma(\mathrm{Conv}_{1\times1}(\mathcal{F}(\tilde{\mathbf{F}}_{lo}))), \tag{6}$$

where $\hat{\mathbf{F}}_{lo} \in \mathbb{R}^{\hat{H} \times (\frac{\hat{W}}{2}+1) \times 2C'}$ is the processed feature, $\mathcal{F}(\cdot)$ represents the FFT. Next, we calibrate the input features by injecting learnable low-frequency prompt components $\mathbf{P}_{lo} \in \mathbb{R}^{\hat{H} \times (\frac{\hat{W}}{2}+1) \times 2C'}$, which is then transformed to spatial domain:

$$\mathbf{F}_{lo}^{prompt} = \mathcal{F}^{-1}(\hat{\mathbf{F}}_{lo} \odot \mathbf{P}_{lo}), \tag{7}$$

where $\mathbf{F}_{lo}^{prompt} \in \mathbb{R}^{\hat{H} \times \hat{W} \times C'}$ is the generated low-frequency feature prompt, and $\mathcal{F}^{-1}(\cdot)$ denotes the inverse FFT.

Noted, we perform the feature transformation in the Fourier domain for efficient global information interaction. The convolution theorem [46, 55] indicates the Hadamard product of two signals in the Fourier domain equals to implement the Fourier transform of a convolution of these two signals in the original spatial domain. Base on this insight, we can combine Eq.(6) and Eq.(7):

$$\begin{aligned}
\mathbf{F}_{lo}^{prompt} &= \mathcal{F}^{-1}(\mathcal{F}(\tilde{\mathbf{F}}_{lo}) \odot \sigma(\mathrm{Conv}_{1\times1}(\mathcal{F}(\tilde{\mathbf{F}}_{lo}))) \odot \mathbf{P}_{lo}) \\
&= \mathcal{F}^{-1}(\mathcal{F}(\tilde{\mathbf{F}}_{lo} \circledast \mathcal{F}^{-1}(\sigma(\mathrm{Conv}_{1\times1}(\mathcal{F}(\tilde{\mathbf{F}}_{lo}))) \odot \mathbf{P}_{lo}))) \\
&= \tilde{\mathbf{F}}_{lo} \circledast \mathcal{F}^{-1}(\sigma(\mathrm{Conv}_{1\times1}(\mathcal{F}(\tilde{\mathbf{F}}_{lo}))) \odot \mathbf{P}_{lo})
\end{aligned} \tag{8}$$

where '$\circledast$' is the convolution operation. Since $\mathcal{F}^{-1}(\sigma(\mathrm{Conv}_{1\times1}(\mathcal{F}(\tilde{\mathbf{F}}_{lo}))) \odot \mathbf{P}_{lo})$ is a tensor that shares the same shape with $\tilde{\mathbf{F}}_{lo}$, it can be served as a dynamic depth-wise convolution kernel as large as $\tilde{\mathbf{F}}_{lo}$ in spatial domain while introducing less model complexity.

Subsequently, we further modulate the low-frequency visual prompt $\mathbf{F}_{lo}^{prompt}$ with the guidance of the input feature $\mathbf{F}_l$. Specifically, we adopt an adaptive average pooling operator, which serves as a low-pass filter [65], to enhance the

Table 1: Quantitative comparison on the SPAD dataset [70] for image deraining.

Method	DDN [16]	PReNet [56]	RCDNet [69]	MPRNet [83]	SPAIR [51]	Uformer-S [72]	SCD-Former [18]	IDT [75]	Restormer [81]	DRSformer [7]	FPro (Ours)
PSNR ↑	36.16	40.16	43.36	43.64	44.10	46.13	46.89	47.34	47.98	<u>48.53</u>	**48.99**
SSIM ↑	0.9463	0.9816	0.9831	0.9844	0.9872	0.9913	**0.9941**	0.9929	0.9921	0.9924	<u>0.9936</u>

Table 2: Quantitative comparison on the AGAN-Data [52] for image deraindrop.

Method	Eigen's [15]	Pix2pix [23]	Uformer-S [72]	WeatherDiff$_{128}$ [47]	TransWeather [63]	DuRN [38]	RaindropAttn [54]	AttentiveGAN [52]	IDT [75]	Restormer [81]	FPro (Ours)
PSNR ↑	21.31	28.02	29.42	29.66	30.17	31.24	31.37	31.59	31.63	<u>31.68</u>	**31.96**
SSIM ↑	0.757	0.855	0.906	0.923	0.916	0.926	0.918	0.917	<u>0.936</u>	0.934	**0.937**

low-frequency content in the input $\mathbf{F}_{lo}^{prompt}$. After that, we generate *query* ($\mathbf{Q}_{lo}$) projection from reshaped $\mathbf{F}_l$, *key* ($\mathbf{K}_{lo}$) and *value* ($\mathbf{V}_{lo}$) projections from the average-pooled feature $\hat{\mathbf{F}}_{lo}^{prompt} = \mathrm{AAP}(\mathbf{F}_{lo}^{prompt})$, respectively. Here, $\mathrm{AAP}(\cdot)$ means the adaptive average pooling operation. Then, we employ 1×1 convolutions to yield $\mathbf{Q}_{lo} = W_p^{Q_{lo}}\mathbf{F}_l$, $\mathbf{K}_{lo} = W_p^{K_{lo}}\hat{\mathbf{F}}_{lo}^{\mathrm{prompt}}$, and $\mathbf{V}_{lo} = W_p^{V_{lo}}\hat{\mathbf{F}}_{lo}^{\mathrm{prompt}}$. Here $W_p^{(\cdot)}$ is the 1×1 convolution. Next, we calculate the dot-product of query and key projections, which generates the transposed-attention $\mathbf{A} \in \mathbb{R}^{\hat{H}\hat{W}\times 1}$. Overall, the process of modulating the low-frequency prompt is defined as:

$$\mathbf{F}_{lo}^{out} = \mathbf{V}_{lo} \cdot \mathrm{Softmax}(\mathbf{K}_{lo} \cdot \mathbf{Q}_{lo}/\alpha), \tag{9}$$

where $\mathbf{F}_{lo}^{out} \in \mathbb{R}^{\hat{H}\times\hat{W}\times C'}$ is the output feature map of the low-frequency prompt modulation branch; $\mathbf{Q}_{lo} \in \mathbb{R}^{\hat{H}\hat{W}\times C'}$, $\mathbf{K}_{lo} \in \mathbb{R}^{C'\times 1}$, and $\mathbf{V}_{lo} \in \mathbb{R}^{1\times C'}$ are the input matrices; α is the learnable scaling parameter.

For both low/high-frequency modulators, we perform the attention map calculation several times in parallel, and these results are then concatenated for multi-head self-attention (MSA) [64].

4 Experiments

4.1 Experimental settings

Metrics. We adopt commonly used peak signal-to-noise ratio (PSNR) [73] and structural similarity (SSIM) metrics to evaluate restored images. Meanwhile, perceptual metric NIQE [42] is employed as a non-reference metric. Following previous works [69, 72], PSNR/SSIM computations are implemented on the Y channel in the YCbCr space for the image deraining task, while calculated in the RGB color space for other restoration tasks. In the reported tables, the best and second-best scores are **highlighted** and <u>underlined</u>, respectively.

Implementation Details. FPro contains $N_1 = 3$ levels encoder-decoder, where the encode and decoder share the same N_2=[2,3,6] blocks. We set embedding dimensions C as 48, and the attention heads as [2,4,8]. The expanding channel

Table 3: Quantitative comparison on the TIP-2018 [61] for image demoiré.

Method	AMNet [80]	DMCNN [61]	UNet [59]	WDNet [36]	MopNet [19]	TAPE-Net [37]	FHD^{2}eNet [20]	MBCNN [86]	Uformer-S [72]	Wang *et al.* [66]	FPro (Ours)
PSNR ↑	25.47	26.10	26.49	27.12	27.48	27.52	27.79	28.40	28.63	<u>28.87</u>	**29.25**
SSIM ↑	0.833	0.844	0.864	0.854	0.861	0.866	0.867	0.871	0.872	**0.894**	<u>0.879</u>

Table 4: Quantitative comparison on the SOTS [29] for image dehazing.

Method	AODNet [28]	MSCNN [57]	DehazeNet [2]	EPDN [53]	FDGAN [12]	AirNet [30]	Restormer [81]	PromptIR [50]	FPro (Ours)
PSNR ↑	20.29	22.06	22.46	22.57	23.15	23.18	30.87	<u>31.31</u>	**32.85**
SSIM ↑	0.877	0.908	0.851	0.863	0.921	0.900	0.969	<u>0.973</u>	**0.983**

capacity factor in FFN is 3. The default split window size in HPM is set as $M = 8$. The pixel-unshuffle and pixel-shuffle are employed for downsampling and upsampling. We use the AdamW optimizer with an initial learning rate of 3×10^{-4}, which is gradually reduced to 1×10^{-6} using cosine annealing, to train FPro. Additionally, we adopt the widely used loss function [68] to constrain the network training.

4.2 Main Results

Rain Streak Removal. The proposed FPro is compared with general restoration approches [51,72,81,83] as well as with task-specfic methods [7,16,18,56,69,75]. Tab. 1 shows that FPro makes superior performance over current methods for real image deraining on SPAD [70]. Compared to the previous best approach DRSformer [7], FPro achieves a 0.46 dB performance boost. In addition, FPro obtains 2.1 dB PSNR improvement when compared to the recent model SCD-Former [18]. Fig. 4 provides a visual deraining example, where FPro successfully removes the rain degradation while maintaining the structural information.

Raindrop Removal. For image deraindrop, we compare FPro with some existing deraindrop methods, including Eigen's [15], Pix2pix [23], TransWeather [63], Uformer [72], WeatherDiff$_{128}$ [47], DuRN [38], RaindropAttn [54], AttentiveGAN [52], IDT [75], and Restormer [81]. We report the quantitative results on AGAN-Data [52] in Tab. 2. Our FPro obtains the best performance against all considered methods in terms of both PSNR/SSIM scores. FPro makes a 0.28 dB PSNR performance boost over Restormer [81], and 2.3 dB over the recent

Fig. 4: Qualitative comparisons on the SPAD [70] for real image deraining.

Table 5: Quantitative comparison on the GoPro [43] for image deblurring.

Method	CODE [85]	IPT [4]	MPRNet [83]	MIMO [9]	HINet [6]	MAXIM [62]	Restormer [81]	PromptRestorer [68]	FPro (Ours)
PSNR ↑	31.94	32.58	32.66	32.68	32.71	32.86	32.92	**33.06**	<u>33.05</u>
SSIM ↑	0.928	-	0.959	0.959	0.959	<u>0.961</u>	<u>0.961</u>	**0.962**	<u>0.961</u>

Table 6: Effectiveness of GDD.

Models	PSNR	SSIM
(a) w/o prompt branch	48.40	0.9920
(b) Multi DC [8]	48.52	0.9926
(c) Multi GDD	48.91	0.9934
(d) Single GDD	48.99	0.9936

Table 7: Ablation study of DPB.

Models	PSNR	SSIM
(a) w/o prompt branch	48.40	0.9920
(b) w/o HPM	48.77	0.9931
(c) w/o LPM	48.89	0.9933
(d) Full	48.99	0.9936

method WeatherDiff$_{128}$ [47]. Fig. 5 shows the visual comparisons, where FPro generates a result with finer details.

Moiré pattern Removal. We perform image demoiréing experiments on TIP-2018 [61] dataset, and compare FPro with ten demoiréing methods, including AMNet [80], DMCNN [61], UNet [59], WDNet [36], MopNet [19], TAPE-Net [37], FHD^{2}eNet [20], MBCNN [86], Uformer-S [72], and Wang *et al.* [66]. In Tab. 3, FPro yields a 0.38 performance boost against the previous best method Wang *et al.* [66], and outperforms the recent model TAPE-Net [37] by 1.73 dB.

Haze Removal. We perform image dehazing experiments on SOTS [29] benchmark. We compare FPro with eight representative approaches, including AOD-Net [28], MSCNN [57], DehazeNet [2], EPDN [53], FDGAN [12], AirNet [30], Restormer [81], and PromptIR [50]. As shown in Tab. 4, FPro achieves the best scores among all considered methods. Compared to the recent prompt-based method PromptIR [50], FPro makes a substantial performance gain of 1.54 dB.

Motion Blur Removal. We evaluate image deblurring performance on the GoPro dataset [43]. For synthetic deblurring, we compare FPro with eight representative models: CODE [85], IPT [4], MPRNet [83], MIMO [9], HINet [6], MAXIM [62], Restormer [81] and PromptRestorer [68]. Tab. 5 shows that our FPro achieves competitive performance against recent prompt-based PromptRestorer [68], where we cost half FLOPs. Meanwhile, compared to the recent method CODE [85], FPro obtains a 1.11 dB gain on PSNR while using lower FLOPs.

Fig. 5: Qualitative comparisons on the AGAN-Data [52] for image deraindrop.

4.3 Analysis and Discussion

For the ablation studies, we study different models for rain streak removal on SPAD [70] with 256×256 patches for 300K iterations. Testing is conducted on SPAD testing dataset [70].

Effectiveness of Gated Dynamic Decoupler. To demonstrate the effectiveness of the Gated Dynamic Decoupler, we conduct experiments on different model variants in Tab. 6. Compared to the model equipped with Multiple Dynamic Convolution [8] (DC) for separating different frequency parts (b), directly replacing it with GDD (c) results in a performance gain of 0.39 dB in terms of PSNR. Meanwhile, instead of injecting GDD into each DPB (c) to employ multiple decouplers, we attempt to share one GDD module to divide the low-/high frequency information (d), which slightly reduces the complexity (0.02 M) of the whole framework and brings a 0.08 dB performance boost.

Effectiveness of Dual Propmt Block. To investigate the developed DPB, ablation studies are performed in Tab. 7. Disabling HPM or LPM leads to a clear drop of 0.22 dB and 0.1 dB, respectively. These experimental results indicate that both HPM and LPM play a positive role in restoring high-quality images. Moreover, we present visualizations to better show the effect of DPB. As shown in Fig. 7, we visualize the generated low-/high-frequency feature maps from each branch along with the analysis in the Fourier domain, where the low-frequency prompt feature encodes information such as structures while the high-frequency prompt one focus on information such as edges and texture. Meantime, we provide visual comparisons in Fig. 8 to show the effectiveness of the proposed HPM/LPM.

Perceptual Quality Assessment. To test the perceptual quality of the proposed FPro, following [7], we randomly choose 20 rainy images under real-world scenes from Internet-Data [70] to perform the evaluation. As shown in Tab. 9, compared to other considered methods, FPro achieves a lower NIQE score, which means the generated results contain clearer content and better perceptual quality. Through qualitative comparison in Fig. 6, FPro obtains a visually pleasant result against other models, indicating that it handles unseen degradation well.

Model Efficiency. We provide the comparison of performance (PSNR), complexity (FLOPs and Parameters), and latency (Run-times) for image deraining. FLOPs and Runtimes are measured when input with the size of 256×256, and PSNR scores are tested on SPAD [70]. As shown in Tab. 8, though FPro achieves better performance in terms of PSNR metric, it has less model complexity than

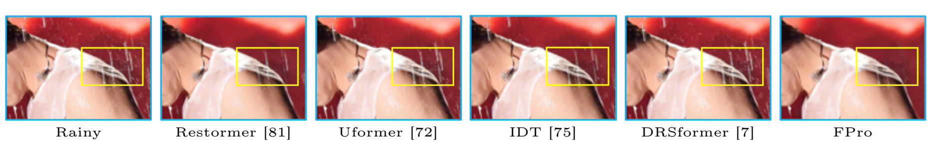

Fig. 6: Qualitative comparisons with state-of-the-art methods on Internet-Data [70] for real rain removal. (Zoom in for a better view.)

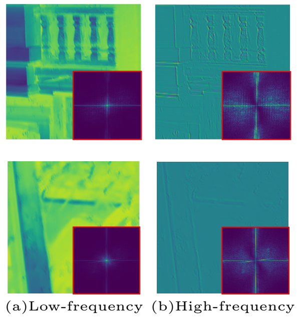

Fig. 7: Feature analysis. we visualize the features from the LPM branch (a), and the HPM one (b). In the right-bottom, we show the results of the average features over the channel dimension in the Fourier domain. (Zoom in for a better view.)

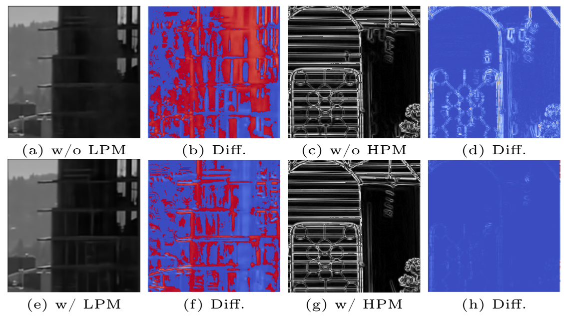

Fig. 8: Effect of DBP. Columns 1 and 3 show low-pass and high-pass filtered results, while columns 2 and 4 show the difference (Diff.) between processed results with corresponding filtered ground-truth. Compared with (a), FPro w/ LPM (e) performs better in capturing information such as structures, resulting in fewer erroneous predictions (f). Compared with (c), FPro w/ HPM (g) restores clear edges and shapes, which indicates it enjoys the benefits from the high-frequency information prompt. (Zoom in for a better view.)

Restormer [81] and DRSformer [7]. Compared to other CNN-/Transformer-based approaches, FPro still has a less or comparable model complexity.

Comparisons with Alternatives to FPro. To further demonstrate the superiority of FPro, we compare it with recent prompt-based methods that mine spatial relations as prompts, including PromptIR [50] and PromptRestorer [68]. As shown in Tab. 10, following PromptIR [50], we train and validate FPro on Rain100L [77]. We achieve a substantial gain of 2.16 dB over PromptIR, and a 0.16 dB performance boost against PromptRestorer.

Table 10: Comparisons with alternatives to FPro.

Models	Params	FLOPs	PSNR
PromptIR [50]	35.6	173	37.04
PromptRestorer [68]	24.4	186	39.04
FPro	22.3	82	39.20

Table 8: Model efficiency analysis on SPAD [70].

Method	MPRNet [83]	SwinIR [33]	Uformer-S [72]	Restormer [81]	IDT [75]	DRSformer [7]	FPro
FLOPs/G	175.8	238.0	**43.9**	174.7	<u>61.9</u>	242.9	81.9
Parameters/M	20.1	**11.5**	20.6	26.1	<u>16.4</u>	33.7	22.3
Run-times/s	**0.03**	1.83	0.12	0.14	0.28	<u>0.08</u>	<u>0.08</u>
PSNR/dB	43.64	44.97	46.13	47.98	47.34	<u>48.53</u>	**48.99**

Table 9: Quantitative results (NIQE) for real-world image deraining.

Methods	Input	Uformer-S [72]	Restormer [81]	IDT [75]	DRSformer [7]	FPro
NIQE ↓	5.8012	5.6971	5.6631	5.6085	5.5942	**5.2999**

(a) Input (b) FPro

Fig. 9: Examples of erroneous reconstruction are shown, where heavy degradation in nighttime real-world scenes leads to typical failures of FPro.

5 Conclusion

In this work, we explore the benefits of prompt learning from a frequency perspective for the task of image restoration. First, when dynamic decoupling the input features with a gating mechanism to select representative elements, we obtain the related frequency components with regard to the specific degradation removal task. Then, we propose modulating the low-/high-frequency signals with separate branches, which concern the intrinsic characteristics of feature maps from different frequency bands. With these modules, our proposed FPro surpasses previous state-of-the-art methods in several image restoration tasks, while performing competitively in terms of computational cost.

Limitations. There remain many avenues for further improvements. For instance, one could achieve better performance by addressing failure cases shown in Fig. 9, where FPro meets challenges in dealing with heavy degradation in the nighttime real-world scene. Intuitively, collecting a large-scale real-world dataset is a potential direction for improvements.

Acknowledgements

This work was supported by Natural Science Foundation of Tianjin, China (No. 20JCJQJC00020), the National Natural Science Foundation of China (Nos. U22B2049, 62302240), Fundamental Research Funds for the Central Universities, and Supercomputing Center of Nankai University (NKSC).

References

1. Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D.: Language models are few-shot learners. In: NeurIPS (2020)
2. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: An end-to-end system for single image haze removal. TIP (2016)
3. Chantas, G., Galatsanos, N.P., Molina, R., Katsaggelos, A.K.: Variational bayesian image restoration with a product of spatially weighted total variation image priors. TIP (2009)
4. Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., Ma, S., Xu, C., Xu, C., Gao, W.: Pre-trained image processing transformer. In: CVPR (2021)
5. Chen, L., Chu, X., Zhang, X., Sun, J.: Simple baselines for image restoration. In: ECCV (2022)
6. Chen, L., Lu, X., Zhang, J., Chu, X., Chen, C.: Hinet: Half instance normalization network for image restoration. In: CVPR Workshops (2021)
7. Chen, X., Li, H., Li, M., Pan, J.: Learning a sparse transformer network for effective image deraining. In: CVPR (2023)
8. Chen, Y., Dai, X., Liu, M., Chen, D., Yuan, L., Liu, Z.: Dynamic convolution: Attention over convolution kernels. In: CVPR (2020)
9. Cho, S.J., Ji, S.W., Hong, J.P., Jung, S.W., Ko, S.J.: Rethinking coarse-to-fine approach in single image deblurring. In: ICCV (2021)
10. Deng, X., Dragotti, P.L.: Deep convolutional neural network for multi-modal image restoration and fusion. TPAMI (2021)
11. Dong, J., Pan, J., Yang, Z., Tang, J.: Multi-scale residual low-pass filter network for image deblurring. In: ICCV (2023)
12. Dong, Y., Liu, Y., Zhang, H., Chen, S., Qiao, Y.: Fd-gan: Generative adversarial networks with fusion-discriminator for single image dehazing. In: AAAI (2020)
13. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: ICLR (2021)
14. d'Ascoli, S., Touvron, H., Leavitt, M.L., Morcos, A.S., Biroli, G., Sagun, L.: Convit: Improving vision transformers with soft convolutional inductive biases. In: ICML (2021)
15. Eigen, D., Krishnan, D., Fergus, R.: Restoring an image taken through a window covered with dirt or rain. In: ICCV (2013)
16. Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., Paisley, J.: Removing rain from single images via a deep detail network. In: CVPR (2017)
17. Gan, Y., Bai, Y., Lou, Y., Ma, X., Zhang, R., Shi, N., Luo, L.: Decorate the newcomers: Visual domain prompt for continual test time adaptation. In: AAAI (2023)
18. Guo, Y., Xiao, X., Chang, Y., Deng, S., Yan, L.: From sky to the ground: A large-scale benchmark and simple baseline towards real rain removal. In: ICCV (2023)
19. He, B., Wang, C., Shi, B., Duan, L.: Mop moiré patterns using mopnet. In: ICCV (2019)

20. He, B., Wang, C., Shi, B., Duan, L.: Fhde^{2}net: Full high definition demoireing network. In: ECCV (2020)
21. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. TPAMI (2010)
22. Hendrycks, D., Gimpel, K.: Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415 (2016)
23. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: CVPR (2017)
24. Jia, M., Tang, L., Chen, B.C., Cardie, C., Belongie, S., Hariharan, B., Lim, S.N.: Visual prompt tuning. In: ECCV (2022)
25. Jiang, L., Dai, B., Wu, W., Loy, C.C.: Focal frequency loss for image reconstruction and synthesis. In: CVPR (2021)
26. Khattak, M.U., Rasheed, H., Maaz, M., Khan, S., Khan, F.S.: Maple: Multi-modal prompt learning. In: CVPR (2023)
27. Kong, L., Dong, J., Ge, J., Li, M., Pan, J.: Efficient frequency domain-based transformers for high-quality image deblurring. In: CVPR (2023)
28. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: Aod-net: All-in-one dehazing network. In: ICCV (2017)
29. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single-image dehazing and beyond. TIP (2018)
30. Li, B., Liu, X., Hu, P., Wu, Z., Lv, J., Peng, X.: All-in-one image restoration for unknown corruption. In: CVPR (2022)
31. Li, Y., Tan, R.T., Guo, X., Lu, J., Brown, M.S.: Rain streak removal using layer priors. In: CVPR (2016)
32. Li, Z., Lei, Y., Ma, C., Zhang, J., Shan, H.: Prompt-in-prompt learning for universal image restoration. arXiv preprint arXiv:2312.05038 (2023)
33. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: ICCV Workshops (2021)
34. Liu, D., Wen, B., Fan, Y., Loy, C.C., Huang, T.S.: Non-local recurrent network for image restoration. In: NeurIPS (2018)
35. Liu, J., Yan, M., Zeng, T.: Surface-aware blind image deblurring. TPAMI (2021)
36. Liu, L., Liu, J., Yuan, S., Slabaugh, G., Leonardis, A., Zhou, W., Tian, Q.: Wavelet-based dual-branch network for image demoiréing. In: ECCV (2020)
37. Liu, L., Xie, L., Zhang, X., Yuan, S., Chen, X., Zhou, W., Li, H., Tian, Q.: Tape: Task-agnostic prior embedding for image restoration. In: ECCV (2022)
38. Liu, X., Suganuma, M., Sun, Z., Okatani, T.: Dual residual networks leveraging the potential of paired operations for image restoration. In: CVPR (2019)
39. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: ICCV (2021)
40. Ma, J., Cheng, T., Wang, G., Zhang, Q., Wang, X., Zhang, L.: Prores: Exploring degradation-aware visual prompt for universal image restoration. arXiv preprint arXiv:2306.13653 (2023)
41. Mei, Y., Fan, Y., Zhou, Y.: Image super-resolution with non-local sparse attention. In: CVPR (2021)
42. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE SPL (2012)
43. Nah, S., Hyun Kim, T., Mu Lee, K.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: CVPR (2017)
44. Narasimhan, S.G., Nayar, S.K.: Contrast restoration of weather degraded images. TPAMI (2003)

45. Niu, B., Wen, W., Ren, W., Zhang, X., Yang, L., Wang, S., Zhang, K., Cao, X., Shen, H.: Single image super-resolution via a holistic attention network. In: ECCV (2020)
46. Oppenheim, A.: Discrete-time signal processing. Prentice-Hall google schola (1999)
47. Özdenizci, O., Legenstein, R.: Restoring vision in adverse weather conditions with patch-based denoising diffusion models. TPAMI (2023)
48. Pan, X., Zhan, X., Dai, B., Lin, D., Loy, C.C., Luo, P.: Exploiting deep generative prior for versatile image restoration and manipulation. TPAMI (2022)
49. Park, N., Kim, S.: How do vision transformers work? In: ICLR (2022)
50. Potlapalli, V., Zamir, S.W., Khan, S., Khan, F.S.: Promptir: Prompting for all-in-one blind image restoration. In: NeurIPS (2023)
51. Purohit, K., Suin, M., Rajagopalan, A., Boddeti, V.N.: Spatially-adaptive image restoration using distortion-guided networks. In: ICCV (2021)
52. Qian, R., Tan, R.T., Yang, W., Su, J., Liu, J.: Attentive generative adversarial network for raindrop removal from a single image. In: CVPR (2018)
53. Qu, Y., Chen, Y., Huang, J., Xie, Y.: Enhanced pix2pix dehazing network. In: CVPR (2019)
54. Quan, Y., Deng, S., Chen, Y., Ji, H.: Deep learning for seeing through window with raindrops. In: ICCV (2019)
55. Rabiner, L.R., Gold, B.: Theory and application of digital signal processing. Englewood Cliffs: Prentice-Hall (1975)
56. Ren, D., Zuo, W., Hu, Q., Zhu, P., Meng, D.: Progressive image deraining networks: A better and simpler baseline. In: CVPR (2019)
57. Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.H.: Single image dehazing via multi-scale convolutional neural networks. In: ECCV (2016)
58. Ren, W., Zhang, J., Pan, J., Liu, S., Ren, J.S., Du, J., Cao, X., Yang, M.H.: Deblurring dynamic scenes via spatially varying recurrent neural networks. TPAMI (2022)
59. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)
60. Song, X., Zhou, D., Li, W., Dai, Y., Shen, Z., Zhang, L., Li, H.: Tusr-net: Triple unfolding single image dehazing with self-regularization and dual feature to pixel attention. TIP (2023)
61. Sun, Y., Yu, Y., Wang, W.: Moiré photo restoration using multiresolution convolutional neural networks. TIP (2018)
62. Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., Li, Y.: Maxim: Multi-axis mlp for image processing. In: CVPR (2022)
63. Valanarasu, J.M.J., Yasarla, R., Patel, V.M.: Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In: CVPR (2022)
64. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: NeurIPS (2017)
65. Voigtman, E., Winefordner, J.D.: Low-pass filters for signal averaging. Review of Scientific Instruments (1986)
66. Wang, C., He, B., Wu, S., Wan, R., Shi, B., Duan, L.Y.: Coarse-to-fine disentangling demoiréing framework for recaptured screen images. TPAMI (2023)
67. Wang, C., Pan, J., Lin, W., Dong, J., Wu, X.M.: Selfpromer: Self-prompt dehazing transformers with depth-consistency. arXiv preprint arXiv:2303.07033 (2023)
68. Wang, C., Pan, J., Wang, W., Dong, J., Wang, M., Ju, Y., Chen, J.: Promptrestorer: A prompting image restoration method with degradation perception. In: NeurIPS (2023)

69. Wang, H., Xie, Q., Zhao, Q., Meng, D.: A model-driven deep neural network for single image rain removal. In: CVPR (2020)
70. Wang, T., Yang, X., Xu, K., Chen, S., Zhang, Q., Lau, R.W.: Spatial attentive single-image deraining with a high quality real rain dataset. In: CVPR (2019)
71. Wang, W., Xie, E., Li, X., Fan, D.P., Song, K., Liang, D., Lu, T., Luo, P., Shao, L.: Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In: ICCV (2021)
72. Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., Li, H.: Uformer: A general u-shaped transformer for image restoration. In: CVPR (2022)
73. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. TIP (2004)
74. Wu, R., Yang, T., Sun, L., Zhang, Z., Li, S., Zhang, L.: Seesr: Towards semantics-aware real-world image super-resolution. arXiv preprint arXiv:2311.16518 (2023)
75. Xiao, J., Fu, X., Liu, A., Wu, F., Zha, Z.J.: Image de-raining transformer. TPAMI (2022)
76. Xiao, T., Singh, M., Mintun, E., Darrell, T., Dollar, P., Girshick, R.: Early convolutions help transformers see better. In: NeurIPS (2021)
77. Yang, W., Tan, R.T., Feng, J., Guo, Z., Yan, S., Liu, J.: Joint rain detection and removal from a single image with contextualized deep networks. TPAMI (2019)
78. Yu, F., Gu, J., Li, Z., Hu, J., Kong, X., Wang, X., He, J., Qiao, Y., Dong, C.: Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild. arXiv preprint arXiv:2401.13627 (2024)
79. Yu, K., Wang, X., Dong, C., Tang, X., Loy, C.C.: Path-restore: Learning network path selection for image restoration. TPAMI (2022)
80. Yue, H., Mao, Y., Liang, L., Xu, H., Hou, C., Yang, J.: Recaptured screen image demoiréing. TCSVT (2021)
81. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: CVPR (2022)
82. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Learning enriched features for real image restoration and enhancement. In: ECCV (2020)
83. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: CVPR (2021)
84. Zhang, K., Li, Y., Zuo, W., Zhang, L., Van Gool, L., Timofte, R.: Plug-and-play image restoration with deep denoiser prior. TPAMI (2021)
85. Zhao, H., Gou, Y., Li, B., Peng, D., Lv, J., Peng, X.: Comprehensive and delicate: An efficient transformer for image restoration. In: CVPR (2023)
86. Zheng, B., Yuan, S., Slabaugh, G., Leonardis, A.: Image demoireing with learnable bandpass filters. In: CVPR (2020)
87. Zheng, C., Zhang, Y., Gu, J., Zhang, Y., Kong, L., Yuan, X.: Cross aggregation transformer for image restoration. In: NeurIPS (2022)
88. Zhou, S., Chen, D., Pan, J., Shi, J., Yang, J.: Adapt or perish: Adaptive sparse transformer with attentive feature refinement for image restoration. In: CVPR (2024)
89. Zou, X., Xiao, F., Yu, Z., Li, Y., Lee, Y.J.: Delving deeper into anti-aliasing in convnets. IJCV (2023)