Fourmer: An Efficient Global Modeling Paradigm for Image Restoration

Man Zhou [*1] **Jie Huang** [*2] **Chunle Guo** [3] **Chongyi Li** [3]

Abstract

Global modeling-based image restoration frameworks have become popular. However, they often require a high memory footprint and do not consider task-specific degradation. Our work presents an alternative approach to global modeling that is more efficient for image restoration. The key insights which motivate our study are two-fold: 1) Fourier transform is capable of disentangling image degradation and content component to a certain extent, serving as the image degradation prior, and 2) Fourier domain innately embraces global properties, where each pixel in the Fourier space is involved with all spatial pixels. While adhering to the "spatial interaction + channel evolution" rule of previous studies, we customize the core designs with Fourier spatial interaction modeling and Fourier channel evolution. Our paradigm, Fourmer, achieves competitive performance on common image restoration tasks such as image de-raining, image enhancement, image dehazing, and guided image super-resolution, while requiring fewer computational resources. The code for Fourmer is publicly available at `https://manman1995.github.io/`.

1. Introduction

Image restoration aims to recover a clear image from its degraded version. It is challenging as there are infinite possible results for a degraded image. Recent research efforts have focused on solving the single image restoration problem, which can be divided into two categories: traditional optimization methods and deep learning-based methods (Zhang et al., 2018; Ren et al., 2018; Fu et al., 2021; Zhang et al., 2020; Liu et al., 2021a; Zamir et al., 2022; Guo et al., 2023).

[*]Equal contribution [1]S-Lab, Nanyang Technological University, Singapore [2] Department of Automation, University of Science and Technology of China, Hefei, China [3]School of Computer Science, Nankai University, Tianjin, China. Correspondence to: Chongyi Li <lichongyi25@gmail.com>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

Traditional methods formulate image restoration as an optimization problem and use various image priors, such as dark channel prior for image dehazing (He et al., 2010), histogram distribution prior for underwater image enhancement (Li et al., 2016), non-local mean prior for image denoising (Dixit & Phadke, 2013), sparse image prior for guided image super-resolution (Kim & Kwon, 2010), to constrain the solution space. However, these methods often have limited versatility and require time-consuming optimization.

Deep learning-based methods, particularly CNNs, have achieved promising results in image restoration tasks compared to traditional methods (Liu et al., 2020; Ma et al., 2021; Zhou et al., 2022b). Recently, transformer and MLPs-based global modeling paradigms (Zamir et al., 2022; Tu et al., 2022) have been used in image restoration, surpassing CNN-based methods. However, these frameworks are often used without considering the intrinsic characteristics of specific image restoration tasks and require significant computational resources. We, therefore, wonder "Can we provide a customized and efficient global modeling-based image restoration paradigm?"

In this work, we present a customized and efficient global modeling paradigm, called Fourmer, for image restoration, motivated by our observations on the capabilities of the Fourier transform in image restoration tasks, as shown in Figure 1. The core insights of our approach include using the Fourier transform to disentangle image degradation and content, serving as a general image restoration prior, and utilizing the global properties of the Fourier domain where each pixel is connected to all spatial pixels.

Our approach, Fourmer, builds on the "spatial interaction + channel evolution" rule of existing global modeling paradigms, such as transformer and MLP-Mixer as shown in Figure 3, but customizes the core designs with Fourier spatial interaction and Fourier channel evolution. These designs provide new insights into global modeling network structures for image restoration. Our approach is described in Figure 3 and has been tested on common image restoration tasks, including image de-raining, image enhancement, image dehazing, and guided image super-resolution. The results suggest that our paradigm achieves competitive performance while requiring fewer computational resources. Our main goal is to provide an alternative,

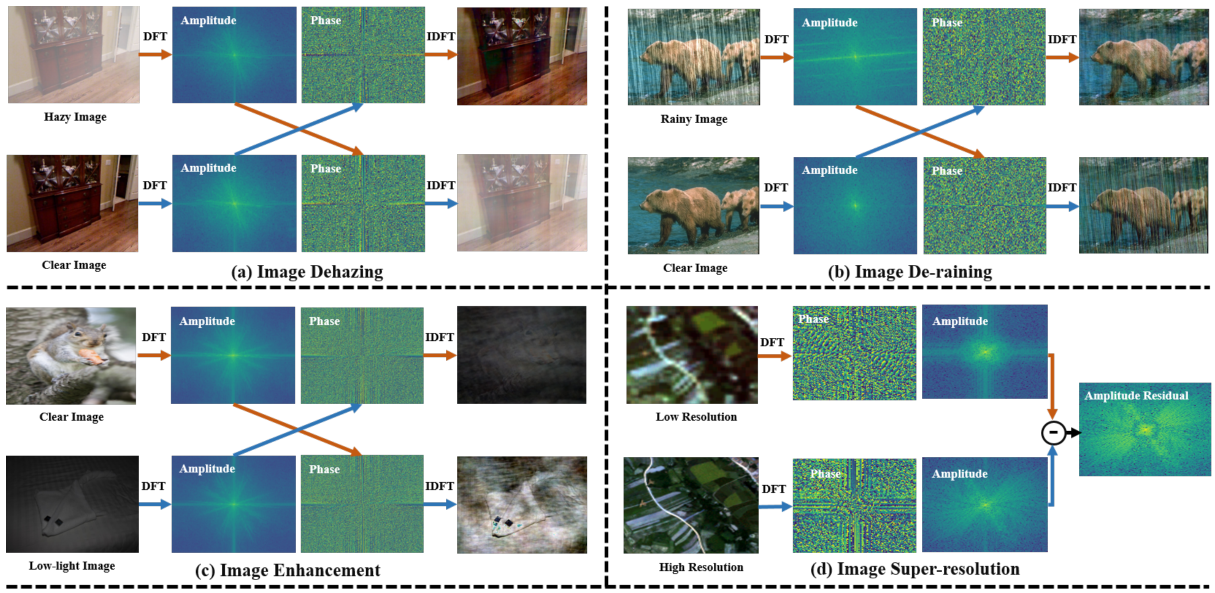

Figure 1. Motivations. Our motivation comes from our observations of the capabilities of the Discrete Fourier Transform (DFT) in common image restoration tasks. We observed that by swapping the amplitude and phase components of a degraded image and its clear version, the degradation (haze, rain) is transferred to the clear version, as shown in (a) and (b). This suggests that the Fourier transform is able to disentangle image degradation and content to a certain extent, and that the degradation is mainly in the amplitude component. To further verify our observation, we also swapped the amplitude and phase components of a degraded image and a clear image with different content, as shown in (c). The degradation is still related to the amplitude component, such as darkness in image enhancement. Similarly, in (d), a low-resolution image and its high-resolution counterpart are different in amplitude component. These observations motivate us to use the Fourier transform as the image degradation prior embedded into image restoration frameworks.

efficient, and customized global modeling-based approach for image restoration, rather than to outperform previous computation-intensive frameworks.

Our contributions are summarized as follows: (1) We propose a global modeling paradigm for image restoration that balances effectiveness and efficiency in comparison to existing global modeling-based frameworks. (2) We incorporate a Fourier-based general image degradation prior into our core structures of Fourier spatial modeling and Fourier channel evolution, providing new insights into the designs of global modeling-based image restoration. (3) Our paradigm Fourmer achieves competitive performance on several mainstream image restoration tasks, such as image de-raining, enhancement, dehazing, and guided super-resolution, while requiring fewer computational resources. Overall, our approach offers a new perspective on global modeling-based image restoration, which can be used as an alternative, efficient, and versatile approach to existing methods.

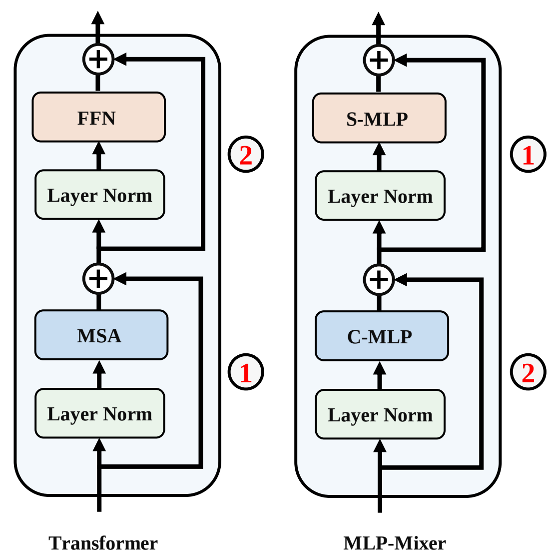

Figure 2. Rule of existing global modeling paradigms: "① spatial interaction + ② channel evolution".

2. Related Work

Image Restoration. Image restoration aims to restore an image degraded by degradation factors (e.g., rain, haze, noise) to a clear counterpart, which has been studied for a long time. Traditional image restoration methods are usually designed as an optimization problem, which incorporates specific priors of the latent clear image to constrain the solution space (He et al., 2009; Li et al., 2016; Dixit & Phadke, 2013; Kim & Kwon, 2010). For example, dark channel prior (He et al., 2009) is proposed for image dehazing and histogram distribution prior (Li et al., 2016) is developed for underwater image enhancement. These methods involve iteration optimization, consuming considerable computational resources and limiting their applications.

Recently, deep learning-based methods have achieved impressive performance in a data-driven manner. Among them, most algorithms are designed with CNN-based architectures. Early works stack deep convolution layers for improving model representation ability, such as VDSR (Kim et al., 2016), DnCNN (Zhang et al., 2017), and ARCNN (Dong et al., 2015). Advanced methods have adopted more powerful architecture designs, such as residual block (Tai et al., 2017; Ehrlich & Davis, 2019) and dense block (Zhang et al., 2020; Dong et al., 2020). Besides, attention mechanism (Zhang et al., 2018; 2021b; zhou et al., 2022) and multi-stage mechanism (Zamir et al., 2021; Chen et al., 2021c) have been brought into image restoration algorithms that improve the performance. However, the locality property of the convolution operation community limits the perception of context-wise global information that is critical for image restoration (Dixit & Phadke, 2013; Berman et al., 2016).

Global Modeling. In recent years, global modeling techniques have gained popularity in the computer vision community. A line of these methods is based on transformer (Vaswani et al., 2017), which has been adapted in numerous vision tasks such as vision recognition (Liu et al., 2021b; Xia et al., 2022) and segmentation (Chen et al., 2021b; Cao et al., 2021). Different from CNN-based architectures, transformer learns long-range dependencies between image patch sequences for global-aware modeling (Dosovitskiy et al., 2020). Various image restoration algorithms based on transformer have been proposed and achieve superior performance in restoration tasks such as image dehazing (Guo et al., 2022a; Yu et al., 2022a), image deraining (Xiao et al., 2022; Zhou et al., 2021b; Xiao et al., 2021; Zhou et al., 2021a; Zhou & Wang, 2021; Fu et al., 2021; Guo et al., 2022b), and low-light image enhancement (Xu et al., 2022; Li et al., 2023; Huang et al., 2022a;b; 2023). Among them, a pioneer work IPT directly applies vanilla transformers to image patches (Chen et al., 2021a), while Uformer (Wang et al., 2022) and SwinIR (Liang et al., 2021) apply window-based local attention models on several image restoration tasks. However, the huge computation cost and parameters of these transformer frameworks limit practical applications.

As another line of global modeling paradigm, multi-layer perceptrons (MLPs)-based methods have attracted attention in vision problems (Tolstikhin et al., 2021). To adapt this architecture for image restoration problems, MAXIM adopts a multi-axis MLP based mechanism to perceive information with global receptive field (Tu et al., 2022). Nevertheless, it still costs enormous computational resources and is thus hard to be applied to compact devices. In summary, all the above architectures are not fully explored priors that are specific for image restoration tasks, which is important for tiqperformance imporvement. Recently, Fourier transformation has presented its effectiveness for global modeling (Chi et al., 2019; 2020). Instead of further exploring the efficacy of Fourier as global modeling in high-level tasks such as image classification, video action classification, human keypoint detection in (Chi et al., 2019), our work is the first to focus on the customized image restoration framework designs. The work proposed in (Chi et al., 2019) pays more attention to the global property while our framework further explores the intrinsic prior tailored for image restoration. In addition, different from existing Fourier techniques (Chi et al., 2020) that emphasize the micro basic operator with the global receptive field, our work focuses on the macro framework design. In our work, we pay more attention to the customized image restoration global modeling framework. We investigate incorporating restoration prior with Fourier transform to conduct effective global modeling, which is efficient for practical application.

Existing transformer-based methods (Wang et al., 2022; Zamir et al., 2022) and MLP-based methods (Tu et al., 2022) do not contain the intrinsic knowledge of the specific image restoration tasks and only roughly focus on the global operator designs. In contrast, our framework is the first to explore the customized image restoration global modeling paradigm. Unlike these methods that only consider global modeling, our work with efficient structure also meets the requirement of image restoration on edge devices with limited computational sources. In general, our proposed framework incorporates both advantages of the global modeling mechanism and general image degradation prior that are introduced by Fourier transform, thus achieving better performance.

3. Method

3.1. Preliminary

Fourier transform is a widely used technique for analyzing the frequency content of an image. For images with multiple color channels, the Fourier transform is applied to each channel separately. Given an image $\boldsymbol{x} \in \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C}}$, the Fourier transform $\mathcal{F}$ converts it to Fourier space as the complex component $\mathcal{F}(\boldsymbol{x})$, which is expressed as:

$$\mathcal{F}(\boldsymbol{x})(\boldsymbol{u},\boldsymbol{v}) = \frac{1}{\sqrt{\mathrm{HW}}}\sum_{\boldsymbol{h}=0}^{\mathrm{H}-1}\sum_{\boldsymbol{w}=0}^{\mathrm{W}-1}\boldsymbol{x}(\boldsymbol{h},\boldsymbol{w})e^{-j2\pi(\frac{\boldsymbol{h}}{\mathrm{H}}\boldsymbol{u}+\frac{\boldsymbol{w}}{\mathrm{W}}\boldsymbol{v})}, \tag{1}$$

where $\boldsymbol{u}$ and $\boldsymbol{v}$ and the coordinates of the Fourier space. $\mathcal{F}^{-1}(\boldsymbol{x})$ defines the inverse Fourier transform. Both the Fourier transform and its inverse procedure can be efficiently

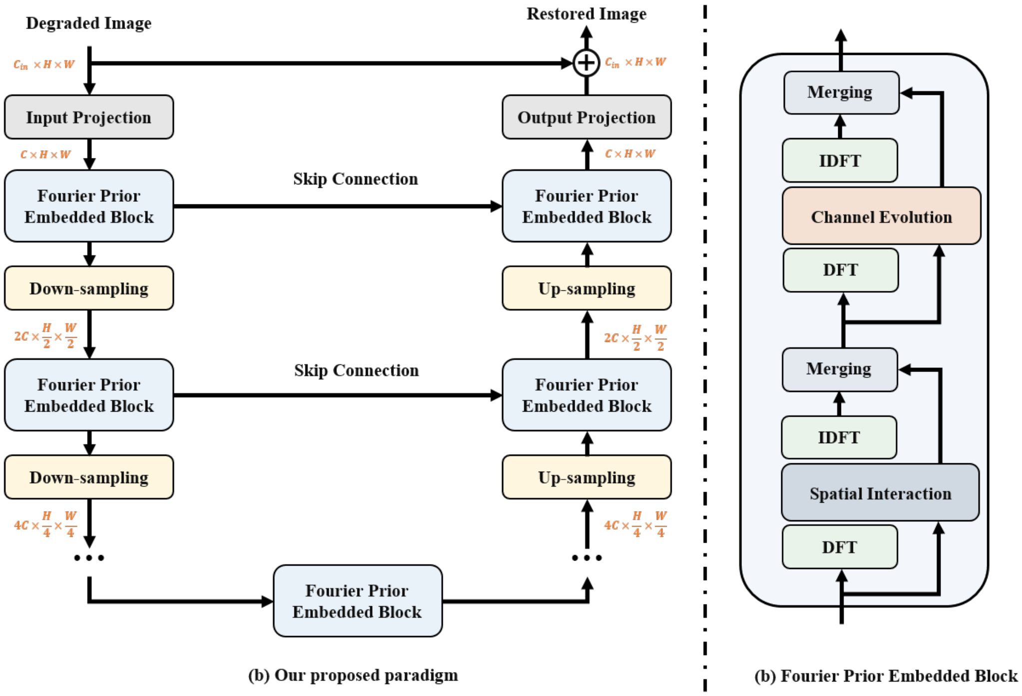

Figure 3. Overview of the proposed efficient and customized global modeling paradigm for image restoration.

implemented using FFT/IFFT algorithms (Frigo & Johnson, 1998). The amplitude component $\mathcal{A}(\boldsymbol{x})(\boldsymbol{u},\boldsymbol{v})$ and phase component $\mathcal{P}(\boldsymbol{x})(\boldsymbol{u},\boldsymbol{v})$ are expressed as:

$$\begin{aligned}\mathcal{A}(\boldsymbol{x})(\boldsymbol{u},\boldsymbol{v})) &= \sqrt{\mathcal{R}^2(\boldsymbol{x})(\boldsymbol{u},\boldsymbol{v})) + \mathcal{I}^2(\boldsymbol{x})(\boldsymbol{u},\boldsymbol{v}))},\\ \mathcal{P}(\boldsymbol{x})(\boldsymbol{u},\boldsymbol{v})) &= \arctan[\frac{\mathcal{I}(\boldsymbol{x})(\boldsymbol{u},\boldsymbol{v}))}{\mathcal{R}(\boldsymbol{x})(\boldsymbol{u},\boldsymbol{v}))}],\end{aligned} \tag{2}$$

where $\mathcal{R}(\boldsymbol{x})(\boldsymbol{u},\boldsymbol{v})$ and $\mathcal{I}(\boldsymbol{x})(\boldsymbol{u},\boldsymbol{v})$ represent the real and imaginary parts respectively. The Fourier transform and its inverse procedure are applied independently to each channel of the image or feature maps.

We use Fourier transform to conduct a detailed frequency analysis for image restoration. By analyzing the properties of phase and amplitude components in the Fourier space, we observed that the degradation effect is mainly presented in the amplitude component. This can be seen by swapping the amplitude and phase components between a degraded image and its clear version, as shown in Figure 1. This phenomenon indicates that the Fourier transform can effectively separate image degradation and content to some extent, and that the degradation primarily occurs in the amplitude component. This motivates us to use Fourier transform as an image degradation prior in image restoration frameworks.

3.2. Overall Framework

Structure Flow. Our main goal is to develop an effective and efficient global modeling paradigm for image restoration, detailed in Figure 3. Given a degraded image $\mathbf{I} \in \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C_{in}}}$, the approach first applies a convolution layer to protect the image into shallow feature embedding $\mathbf{X_0} \in \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C}}$. Following a U-shaped network design, the shallow embedding is passed through N encoder stages. Each stage consists of a stack of the proposed core building module, the Fourier Prior Embedded (FPE) Block, and a downsampling layer. The FPE Block takes advantage of the inborn global modeling properties of Fourier transform and adheres to the underlying global modeling rule "spatial interaction + channel evolution" to customize the Fourier spatial and channel information interaction. The downsampling layer downsamples the 2D spatial feature maps using a 3×3 convolution with stride 2. Similarly, in the decoder stages, we use the stack of the proposed FPE Block and one upsampling layer for feature reconstruction in each stage. To assist the recovery process, each stage takes the high-level decoder features concatenated with the same stage low-level encoder features via skip connections as input. This helps in preserving the fine structural and textural details in the restored images. Finally, a convolution layer is applied to generate a residual image $\mathbf{I} \in \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C_{in}}}$, which is added to the degraded image to obtain the final result $\mathcal{H}_{\mathrm{O}}$.

Optimization Flow. In addition to the novel network designs, we also introduce a new loss function for optimizing the network training for better results in both spatial and frequency domains. The new loss function consists of two

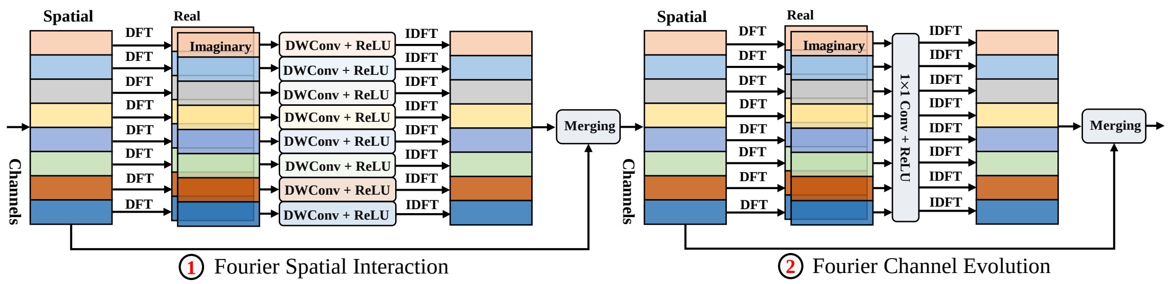

Figure 4. Details of the Fourier Prior Embedded Block. The FPE Block follows the same global modeling rule "spatial interaction + channel evolution" but is with new designs: Fourier spatial interaction modeling and Fourier channel evolution.

parts: a spatial domain loss and a frequency domain loss.

In the spatial domain, we adopt the $\mathcal{L}_1$ loss function, as expressed in Equation (3). This ensures that the network output $\mathcal{H}_{\mathrm{O}}$ is as close as possible to the corresponding ground truth image GT in the pixel-level.

$$\mathcal{L}_{spa} = \|\mathcal{H}_O - \mathrm{GT}\|_1 . \tag{3}$$

To improve the restoration performance and better reconstruct global information, we also add frequency domain supervision via Fourier transform. In the frequency domain, we first use the Discrete Fourier Transform (DFT) to convert $\mathcal{H}_O$ and GT into the Fourier space. Then, the $\mathcal{L}_1$-norm of the amplitude difference and phase difference between $\mathcal{H}_O$ and GT are calculated and summed to produce the total frequency loss as expressed in Equation (4).

$$\mathcal{L}_{fre} = \|\mathcal{A}(\mathcal{H}_O) - \mathcal{A}(\mathrm{GT})\|_1 + \|\mathcal{P}(\mathcal{H}_O) - \mathcal{P}(\mathrm{GT})\|_1 . \tag{4}$$

Finally, the overall loss function is formulated as

$$\mathcal{L} = \mathcal{L}_{spa} + \lambda \mathcal{L}_{fre}, \tag{5}$$

where λ is the weight factor and is set to 0.1. By minimizing this loss function, the network is trained to produce better results in both spatial and frequency domains.

3.3. Fourier Prior Embedded Block

As shown in Figure 4, the Fourier Prior Embedded Block as a core building module contains two fundamental elements: Fourier Spatial Interaction and Fourier Channel Evolution.

Fourier Spatial Interaction. The Fourier Spatial Interaction first takes the feature maps as input and then applies Fourier transform to convert the spatial features into real and imaginary components. Suppose that the features are denote as $\mathbf{X} \in \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{B}}$, the corresponding Fourier transform is expressed as

$$\mathbf{X}_{\mathcal{I}}^{(\boldsymbol{b})}, \mathbf{X}_{\mathcal{R}}^{(\boldsymbol{b})} = \mathcal{F}(\mathbf{X}^{(\boldsymbol{b})}), \tag{6}$$

where $\boldsymbol{b} = 1, \ldots, \mathrm{B}$, $\mathbf{X}_{\mathcal{I}}^{(\boldsymbol{b})}$ and $\mathbf{X}_{\mathcal{R}}^{(\boldsymbol{b})}$ indicate the real and imaginary respectively. Then, we implement the spatial interaction by a stack of depth-wise convolution with the kernel size of 3×3 and the ReLU function. Specifically, $\mathbf{X}_{\mathcal{I}}^{(\boldsymbol{b})}$ and $\mathbf{X}_{\mathcal{R}}^{(\boldsymbol{b})}$ share the common depth-wise operator while different channels are independently performed. The spatial interaction can be written as follows:

$$\mathbf{S}_{\mathcal{I}}^{(\boldsymbol{b})} = \sigma \cdot \mathrm{DW}^{(\boldsymbol{b})}(\mathbf{X}_{\mathcal{I}}^{(\boldsymbol{b})}), \tag{7}$$

$$\mathbf{S}_{\mathcal{R}}^{(\boldsymbol{b})} = \sigma \cdot \mathrm{DW}^{(\boldsymbol{b})}(\mathbf{X}_{\mathcal{R}}^{(\boldsymbol{b})}), \tag{8}$$

where σ and DW indicate the ReLU function and depth-wise convolution respectively. Next, we apply the inverse DFT to transform the filtered frequency components of $\mathbf{S}_{\mathcal{I}}^{(\boldsymbol{b})}$ and $\mathbf{S}_{\mathcal{R}}^{(\boldsymbol{b})}$ back to the spatial domain

$$\mathbf{X}_{\mathbf{S}}^{\boldsymbol{b}} = \mathcal{F}^{-1}(\mathbf{S}_{\mathcal{I}}^{(\boldsymbol{b})}, \mathbf{S}_{\mathcal{R}}^{(\boldsymbol{b})}). \tag{9}$$

The spectral convolution theorem in Fourier theory states that processing information in the frequency domain can reveal the overall frequency composition. We then combine the Fourier-transformed spatial features, $\mathbf{X}_{\mathbf{S}}$, by concatenating each component, $\mathbf{X}_{\mathbf{S}}^{\boldsymbol{b}}$, with the spatial features processed by a half-instance normalization block, resulting in the final output, $\mathbf{S}_{\mathbf{X}}$.

Fourier Channel Evolution. The Fourier Channel Evolution performs point-wise channel interaction by first decomposing the output $\mathbf{S}_{\mathbf{X}}$ from the Fourier Spatial Interaction into real and imaginary components $\mathbf{C}_{\mathcal{R}}$ and $\mathbf{C}_{\mathcal{I}}$. It then applies a stack of convolution operator with a kernel size of 1×1 and the ReLU function to perform the channel interaction, where each position in the frequency space is shared. The Fourier Channel Interaction can be written as:

$$\mathbf{CX}_{\mathcal{I}} = \sigma \cdot \mathbf{conv}(cat[\mathbf{C}_{\mathcal{I}}^1, \ldots, \mathbf{C}_{\mathcal{I}}^B]), \tag{10}$$

$$\mathbf{CX}_{\mathcal{R}} = \sigma \cdot \mathbf{conv}(cat[\mathbf{C}_{\mathcal{R}}^1, \ldots, \mathbf{C}_{\mathcal{R}}^B]), \tag{11}$$

where conv indicates the convolution with a kernel size of 1×1. The filtered frequency components $\mathbf{CX}_{\mathcal{I}}^{(\boldsymbol{b})}$ and $\mathbf{CX}_{\mathcal{R}}^{(\boldsymbol{b})}$ are then transformed back to the spatial domain using the inverse DFT:

$$\mathbf{C}_{\mathbf{S}}^{\boldsymbol{b}} = \mathcal{F}^{-1}(\mathbf{CX}_{\mathcal{I}}^{(\boldsymbol{b})}, \mathbf{CX}_{\mathcal{R}}^{(\boldsymbol{b})}). \tag{12}$$

Table 1. Quantitative comparison for image dehazing. '-' indicate the result is not available.

Method	SOTS		Dense-Haze		NH-HAZE		Param (M)	GFLOPs
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑		
DCP	15.09	0.7649	10.06	0.3856	10.57	0.5196	-	-
DehazeNet	20.64	0.7995	13.84	0.4252	16.62	0.5238	0.01	-
AOD-Net	19.82	0.8178	13.14	0.4144	15.40	0.5693	0.002	0.1
GridDehazeNet	32.16	0.9836	13.31	0.3681	13.80	0.5370	0.96	21.5
FFA-Net	36.39	0.9886	14.39	0.4524	19.87	0.6915	4.68	288.1
MSBDN	33.79	0.9840	15.37	0.4858	19.23	0.7056	31.35	41.5
KDDN	34.72	0.9845	14.28	0.4074	17.39	0.5897	5.99	-
AECR-Net	37.17	0.9901	15.80	0.4660	19.88	0.7173	2.61	43.0
Fourmer (Ours)	**37.32**	**0.9901**	**15.95**	**0.4917**	**19.91**	**0.7214**	1.29	20.6

Finally, we perform a similar merging process with the Fourier Spatial Interaction, achieving the global modeling for both spatial and channel dimensions.

4. Experiment

We conduct extensive experiments on common image restoration tasks, including image de-raining, image enhancement, image dehazing, and guided image super-resolution. These results will provide insight into the performance of the proposed paradigm and how it compares to other existing methods in the field. For these tasks, the only difference in our frameworks is the number of features as some tasks require more features to optimize the network.

4.1. Experimental Settings

Low-light image enhancement. We evaluate our paradigm on two popular low-light image enhancement benchmarks, including LOL (Chen Wei, 2018) and Huawei (Hai et al., 2021). LOL dataset consists of 500 low-/normal-light image pairs and splits 485 for training and 15 for testing. Huawei dataset contains 2,480 paired images and splits 2,200 for training and 280 for testing. We compare our paradigm with 13 state-of-the-art low-light image enhancement methods: SRIE (Fu et al., 2016), RetinexNet (Chen Wei, 2018), MBLLEN (Lv et al., 2018), EnlightenGAN (Jiang et al., 2021), GLADNet (Wang et al., 2018), Xu et al. (Xu et al., 2020), TBEFN (Lu & Zhang, 2020), KinD (Zhang et al., 2019), Zero-DCE++ (Li et al., 2021), DRBN (Yang et al., 2020), RetinexDIP (Zhao et al., 2021), RUAS (Liu et al., 2021a), KinD++ (Zhang et al., 2021a), and URetinex (Wu et al., 2022).

Image De-raining. Following the work (Zamir et al., 2021), our paradigm is evaluated on 13,712 clean-rain image pairs, gathered from multiple synthetic datasets. We perform evaluations on the Rain100H and Rain100L. We compare our paradigm and 9 representative state-of-the-art methods: DerainNet (Yang et al., 2017b), SEMI (Wei et al., 2019), DID-MDN (Zhang & Patel, 2018), UMRL (Yasarla & Patel, 2019), RESCAN (Li et al., 2018b), PReNet (Ren et al., 2019), MSPFN (Jiang et al., 2020), MPRNet (Zamir et al., 2021), and HINet (Chen et al., 2021c).

Image Dehazing. We evaluate our paradigm on synthetic and real-world datasets. For synthetic scenes, we employ RESIDE (Li et al., 2018a) dataset. The subset Indoor Training Set (ITS) of RESIDE contains a total of 13,990 hazy indoor images, generated from 1,399 clear images. The subset Synthetic Objective Testing Set (SOTS) of RESIDE consists of 500 indoor and 500 outdoor hazy images. In addition, we adopt two real-world datasets: Dense-Haze (Ancuti et al., 2019) and NH-HAZE (Ancuti et al., 2020) to evaluate the generalization. Both datasets consist of 55 paired images. We compare our paradigm with 7 representative methods: DCP (He et al., 2010) and DehazeNet (Cai et al., 2016), AOD-Net (Li et al., 2017), GridDehazeNet (Liu et al., 2019), FFA-Net (Qin et al., 2020), MSBDN (Dong et al., 2020), and AECR-Net (Wu et al., 2021).

Guided Image Super-resolution. Following (Zhou et al., 2022a; Yan et al., 2022), we adopt the representative task of guided image super-resolution, pan-sharpening, for evaluations. The WorldView II, WorldView III, and GaoFen2 datasets (Zhou et al., 2022a; Yan et al., 2022) are used. We choose the 11 representative pan-sharpening methods for comparison: 1) 6 state-of-the-art deep-learning-based methods, including PNN (Masi et al., 2016), PANNET (Yang et al., 2017a), MSDCNN (Yuan et al., 2018), SRPPNN (Cai & Huang, 2021), GPPNN (Xu et al., 2021), and INNformer(Zhou et al., 2022a); 2) 5 traditional methods, including SFIM (Liu., 2000), Brovey (Gillespie et al., 1987), GS (Laben & Brower, 2000), IHS (Haydn et al., 1982), and GFPCA (Liao et al., 2017).

Evaluation Metrics. The performance of the proposed paradigm is evaluated using several commonly-used image quality assessment (IQA) metrics. These metrics include relative dimensionless global error in synthesis (ERGAS) (Alparone et al., 2007), peak signal-to-noise ratio (PSNR), Structural Similarity Index (SSIM), and spectral angle mapper (SAM) (Yuhas & Boardman, 1992).

Table 2. Quantitative comparison for image de-raining. '-' indicate the result is not available.

Method	Test100		Rain100H		Rain100L		Test1200		Param (M)	GFLOPs
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑		
DerainNet	22.77	0.810	14.92	0.592	27.03	0.884	23.38	0.835	0.058	1.453
SEMI	22.35	0.788	16.56	0.486	25.03	0.842	26.05	0.822	-	-
DIDMDN	22.56	0.818	17.35	0.524	25.23	0.741	29.65	0.901	0.373	1.686
UMRL	24.41	0.829	26.01	0.832	29.18	0.923	30.55	0.910	0.98	-
RESCAN	25.00	0.835	26.36	0.786	29.80	0.881	30.51	0.882	1.04	20.361
PReNet	24.81	0.851	26.77	0.858	32.44	0.950	31.36	0.911	0.17	73.021
MSPFN	27.50	0.876	28.66	0.860	32.40	0.933	32.39	0.916	13.22	604.70
MPRNet	30.27	0.897	30.41	0.890	36.40	0.965	32.91	0.916	3.64	141.28
HINet	30.29	0.906	30.65	0.894	37.28	0.970	33.05	0.919	3.72	170.71
Fourmer (Ours)	**30.54**	**0.911**	**30.76**	**0.896**	**37.47**	**0.970**	**33.05**	**0.921**	0.4	16.753

Table 3. Quantitative comparison for image enhancement. '-' indicate the result is not available.

Method	LOL		Huawei		Param (M)	GFLOPs
	PSNR↑	SSIM↑	PSNR↑	SSIM↑		
SRIE	12.28	0.596	13.04	0.477	-	-
RobustRetinex	13.88	0.664	14.60	0.559	-	-
RetinexNet	16.77	0.425	16.65	0.485	0.84	148.54
MBLLEN	17.56	0.729	16.63	0.526	0.45	21.37
EnGAN	17.48	0.674	17.03	0.514	8.37	72.61
GLADNet	19.72	0.680	17.76	0.521	1.13	275.32
Xu et al.	16.78	0.766	16.12	0.586	8.62	68.45
TBEFN	17.35	0.781	16.88	0.575	0.49	24.11
KinD	20.86	0.802	16.48	0.540	8.54	36.57
ZeroDCE	15.29	0.518	12.46	0.407	0.08	20.24
DRBN	20.13	0.801	18.46	**0.635**	0.58	42.41
RUAS	16.41	0.500	13.76	0.516	0.003	0.86
KinD++	21.30	0.822	15.78	0.452	8.28	2970.50
URetinex	21.32	**0.835**	18.79	0.607	1.23	68.37
Fourmer (Ours)	**23.57**	0.832	**19.17**	0.621	0.08	5.03

Table 4. Quantitative comparison for guided image super-resolution. '-' indicate the result is not available.

Method	Worldview II				GaoFen2				Worldview III				Param (M)	GFLOPs
	PSNR↑	SSIM↑	SAM↓	ERGAS↓	PSNR↑	SSIM↑	SAM↓	EGAS↓	PSNR↑	SSIM↑	SAM↓	EGAS↓		
SFIM	34.1297	0.8975	0.0439	2.3449	36.9060	0.8882	0.0318	1.7398	21.8212	0.5457	0.1208	8.9730	-	-
Brovey	35.8646	0.9216	0.0403	1.8238	37.7974	0.9026	0.0218	1.372	22.5060	0.5466	0.1159	8.2331	-	-
GS	35.6376	0.9176	0.0423	1.8774	37.2260	0.9034	0.0309	1.6736	22.5608	0.5470	0.1217	8.2433	-	-
IHS	35.2962	0.9027	0.0461	2.0278	38.1754	0.9100	0.0243	1.5336	22.5579	0.5354	0.1266	8.3616	-	-
GFPCA	34.5581	0.9038	0.0488	2.1411	37.9443	0.9204	0.0314	1.5604	22.3344	0.4826	0.1294	8.3964	-	-
PNN	40.7550	0.9624	0.0259	1.0646	43.1208	0.9704	0.0172	0.8528	29.9418	0.9121	0.0824	3.3206	0.689	1.1289
PANNET	40.8176	0.9626	0.0257	1.0557	43.0659	0.9685	0.0178	0.8577	29.6840	0.9072	0.0851	3.4263	0.688	1.1275
MSDCNN	41.3355	0.9664	0.0242	0.9940	45.6874	0.9827	0.0135	0.6389	30.3038	0.9184	0.0782	3.1884	2.39	3.9158
SRPPNN	41.4538	0.9679	0.0233	0.9899	47.1998	0.9877	0.0106	0.5586	30.4346	0.9202	0.0770	3.1553	17.114	21.1059
GPPNN	41.1622	0.9684	0.0244	1.0315	44.2145	0.9815	0.0137	0.7361	30.1785	0.9175	0.0776	3.2593	1.198	1.3967
INNformer	41.6903	0.9704	0.0227	0.9514	47.3528	0.9893	0.0102	0.5479	30.5365	0.9225	0.0747	3.0997	0.706	1.3907
Fourmer (Ours)	**41.8325**	**0.9731**	**0.0219**	**0.9506**	**47.5334**	**0.9912**	**0.0102**	**0.5448**	**30.5987**	**0.9241**	**0.0738**	**3.0763**	0.715	1.386

4.2. Comparisons

The quantitative performance comparison is presented in Tables 1, 2, 3, and 4, where the best results are highlighted in bold. From the results, it can be observed that our paradigm achieves promising performance with fewer computational burdens against the compared methods across all tasks and on all testing datasets. These results indicate that the proposed paradigm is able to achieve high-quality results while being computationally efficient, making it a valuable contribution to the field of image restoration.

In Figures 5, 6, 7, and 8, we show the visual comparison that our method has produced the more pleasing results. Due to to the constraint of limited space, we only present the results of representative methods. As can be seen, our proposed method achieves the best performance against other state-of-the-art algorithms.

4.3. Ablation Studies and Analysis

To understand the impact of each key component on the proposed paradigm, comprehensive ablation studies are conducted. We take the task of guided image super-resolution as an example and perform main experiments on the Worldview II dataset.

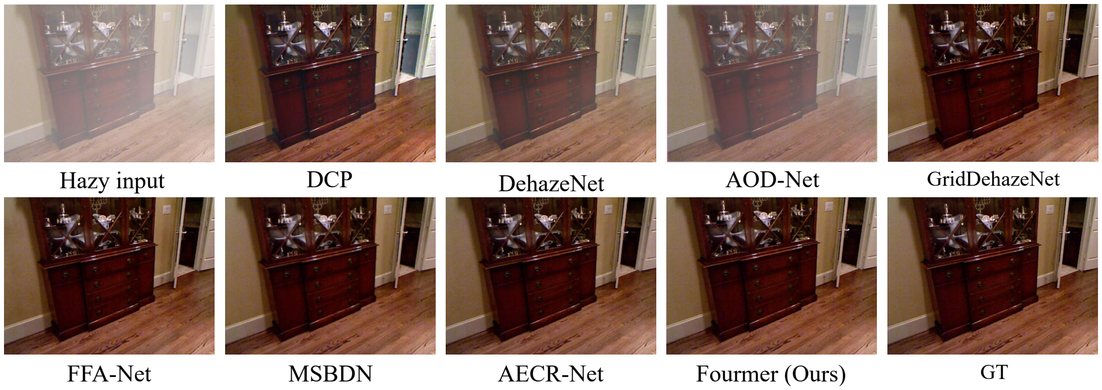

Figure 5. Visual comparison on image dehazing task.

Figure 6. Visual comparison on low-light image enhancement task.

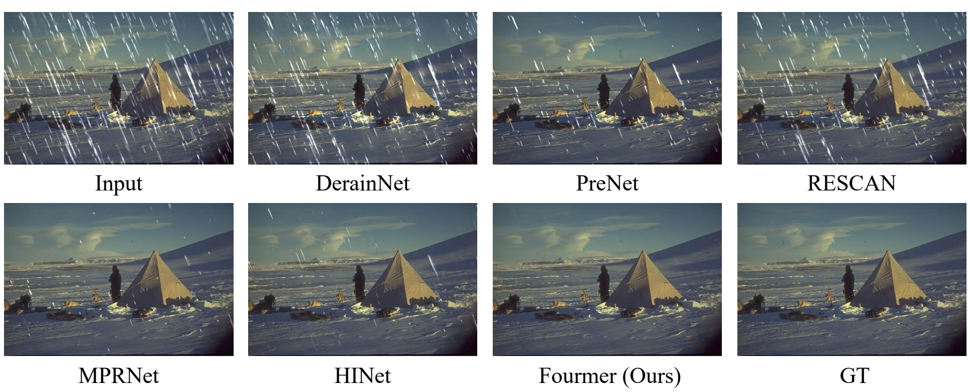

Figure 7. Visual comparison on image de-raining task.

Table 5. Ablation study for the Fourier Prior Embedded Block.

Method	Worldview II				GaoFen2				Worldview III			
	PSNR↑	SSIM↑	SAM↓	ERGAS↓	PSNR↑	SSIM↑	SAM↓	EGAS↓	PSNR↑	SSIM↑	SAM↓	EGAS↓
GPPNN	41.1622	**0.9684**	0.0244	1.0315	44.2145	0.9815	0.0137	0.7361	30.1785	0.9175	0.0776	3.2593
w/FPE	**41.4513**	0.9675	**0.0236**	**1.0001**	**45.5436**	**0.9823**	**0.0135**	**0.6557**	**30.4127**	**0.9201**	**0.0770**	**3.1562**

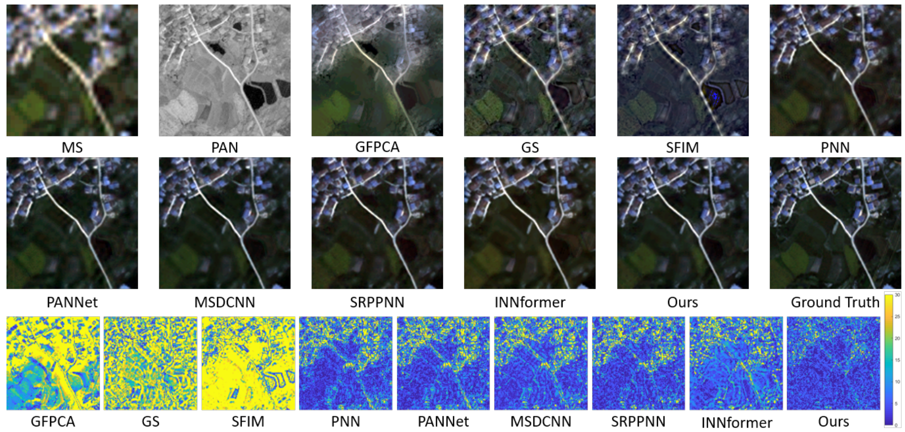

Figure 8. Visual comparison on guided image super-resolution task. The residual maps between the results of different methods and the ground truth are provided at the bottom.

Table 6. Ablation study for the Fourier Spatial Interaction (FSI) and Fourier Channel Evolution (FCE).

Method	PSNR↑	SSIM↑	SAM↓	ERGAS↓
Ours	**41.8325**	**0.9731**	**0.0219**	**0.9506**
w/o FSI	41.4513	0.9675	0.0236	1.0001
w/o FCE	41.5823	0.9705	0.0227	0.9517

Efficacy of Fourier Prior Embedded Block. To verify the efficacy of the Fourier Prior Embedded Block, it is inserted into an existing network, specifically GPPNN (Xu et al., 2021). This is done by replacing the basic blocks of GPPNN with the proposed FPE Blocks. The results presented in Table 5 show the benefit of the introduction of the FPE Block, as it improves network performance.

In addition to the verification of the FPE block, the gain introduced by the Fourier Spatial Interaction and Fourier Channel Evolution is examined separately. As shown in Table 6, both FSI and FCE contribute to the good performance of the FPE block. This indicates that the combination of the Fourier Spatial Interaction and Fourier Channel Evolution with the FPE block leads to even better performance. The results suggest that the proposed FPE block, along with the Fourier Spatial Interaction and Fourier Channel Evolution, form an effective design for image restoration tasks.

Impact of Hierarchical Number. To explore the impact of hierarchical numbers, i.e., the numbers of domsampling stages in our U-shape network, we experiment with the proposed network with varying hierarchical numbers. The corresponding quantitative results for the number K ranging from 1 to 4 are reported in Table 7.

Effectiveness of Frequency Loss. The new frequency loss aims to directly emphasize global frequency information optimization. In Table 8, we remove it to examine its effectiveness. The results in Table 8 demonstrate that removing it severally degrades all metrics, indicating its significance.

Analysis on Our Framework's Effectiveness. The common sense in traditional image restoration algorithms is to explore the intrinsic knowledge and image prior. Besides, the effectiveness of global modeling for image restoration has been demonstrated in previous works. Our work incorporates both advantages of global modeling and general image degradation prior that are introduced by Fourier transform. Specifically, recent works (Dai et al., 2022; Yu et al., 2022b) have demonstrated that "spatial interaction + channel evolution" is the core contribution of the effectiveness within transformer structures. Our work stands on the rule with new designs in Fourier space, achieving better results.

Table 7. Ablation study for the hierarchical number.

K	PSNR↑	SSIM↑	SAM↓	ERGAS↓
1	41.1827	0.9646	0.0255	1.0209
2	41.3324	0.9655	0.0249	1.0125
3	41.5331	0.9682	0.0240	0.9839
4	**41.8325**	**0.9731**	**0.0219**	**0.9506**

Table 8. Ablation study for the frequency loss.

Fre Loss	PSNR↑	SSIM↑	SAM↓	ERGAS↓
	41.7840	0.9725	0.0221	0.9508
✓	**41.8325**	**0.9731**	**0.0219**	**0.9506**

5. Conclusion

In this paper, we presented a novel and efficient global modeling approach for image restoration. We analyzed existing global modeling approaches and identified the key design principles of "spatial interaction + channel evolution". We also examined the properties of the Fourier prior for image restoration, including its decomposition of image degradation and content. Based on these insights, we developed the core designs of Fourier spatial modeling and Fourier channel evolution. Our approach achieves competitive performance while requiring fewer computational resources.

References

Alparone, L., Wald, L., Chanussot, J., Thomas, C., Gamba, P., and Bruce, L. M. Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data fusion contest. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10):3012–3021, 2007.

Ancuti, C. O., Ancuti, C., Sbert, M., and Timofte, R. Dense haze: A benchmark for image dehazing with dense-haze and haze-free images. In *ICIP*, pp. 1014–1018, 2019.

Ancuti, C. O., Ancuti, C., and Timofte, R. NH-HAZE: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In *CVPRW*, pp. 444–445, 2020.

Berman, D., Treibitz, T., , and Avidan, S. Non-local image dehazing. In *CVPR*, pp. 1674–1682, 2016.

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

Cai, J. and Huang, B. Super-resolution-guided progressive pansharpening based on a deep convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 59(6):5206–5220, 2021.

Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., and Wang, M. Swin-unet: Unet-like pure transformer for medical image segmentation. *arXiv:2105.05537*, 2021.

Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., Ma, S., Xu, C., Xu, C., and Gao, W. Pre-trained image processing transformer. In *CVPR*, 2021a.

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., and Zhou, Y. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv:2102.04306*, 2021b.

Chen, L., Lu, X., Zhang, J., Chu, X., and Chen, C. HINet: Half instance normalization network for image restoration. In *CVPR*, pp. 182–192, 2021c.

Chen Wei, Wenjing Wang, W. Y. J. L. Deep retinex decomposition for low-light enhancement. In *BMVC*, 2018.

Chi, L., Tian, G., Mu, Y., Xie, L., and Tian, Q. Fast nonlocal neural networks with spectral residual learning. In *ACMMM*, pp. 2142–2151, 2019.

Chi, L., Jiang, B., and Mu, Y. Fast fourier convolution. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., and Lin, H. (eds.), *NeurIPS*, 2020.

Dai, J., Shi, M., Wang, W., Wu, S., Xing, L., Wang, W., Zhu, X., Lu, L., Zhou, J., Wang, X., et al. Demystify transformers & convolutions in modern image deep networks. *arXiv:2211.05781*, 2022.

Dixit, A. and Phadke, A. Image de-noising by non-local means algorithm. In *CVPR*, pp. 275–277, 2013.

Dong, C., Deng, Y., Loy, C. C., and Tang, X. Compression artifacts reduction by a deep convolutional network. In *ICCV*, pp. 576–584, 2015.

Dong, H., Pan, J., Xiang, L., Hu, Z., Zhang, X., Wang, F., and Yang, M.-H. Multi-scale boosted dehazing network with dense feature fusion. In *CVPR*, pp. 2157–2167, 2020.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Ehrlich, M. and Davis, L. S. Deep residual learning in the jpeg transform domain. In *CVPR*, pp. 3484–3493, 2019.

Frigo, M. and Johnson, S. G. FFTW: An adaptive software architecture for the fft. *ICASSP*, 3, 1998.

Fu, X., Zeng, D., Huang, Y., Zhang, X.-P., and Ding, X. A weighted variational model for simultaneous reflectance and illumination estimation. In *CVPR*, pp. 2782–2790, 2016.

Fu, X., Xiao, Z., Yang, G., Liu, A., Xiong, Z., et al. Unfolding taylor's approximations for image restoration. *NeurIPS*, 34:18997–19009, 2021.

Gillespie, A. R., Kahle, A. B., and Walker, R. E. Color enhancement of highly correlated images. ii. channel ratio and "chromaticity" transformation techniques - sciencedirect. *Remote Sensing of Environment*, 22(3):343–365, 1987.

Guo, C., Wu, R., Jin, X., Han, L., Chai, Z., Zhang, W., and Li, C. Underwater ranker: Learn which is better and how to be better. In *AAAI*, 2023.

Guo, C.-L., Yan, Q., Anwar, S., Cong, R., Ren, W., and Li, C. Image dehazing transformer with transmission-aware 3d position embedding. In *CVPR*, 2022a.

Guo, X., Fu, X., Zhou, M., Huang, Z., Peng, J., and Zha, Z.-J. Exploring fourier prior for single image rain removal. In *Proceedings of the 30th International Joint Conferences on Artificial Intelligence*, pp. 935–941, 2022b.

Hai, J., Xuan, Z., Yang, R., Hao, Y., Zou, F., Lin, F., and Han, S. R2rnet: Low-light image enhancement via real-low to real-normal network. *arXiv:2106.14501*, 2021.

Haydn, R., Dalke, G. W., Henkel, J., and Bare, J. E. Application of the ihs color transform to the processing of multisensor data and image enhancement. *National Academy of Sciences of the United States of America*, 79 (13):571–577, 1982.

He, K., Sun, J., and Tang, X. Single image haze removal using dark channel prior. In *CVPR*, 2009.

He, K., Sun, J., and Tang, X. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2010.

Huang, J., Liu, Y., Zhao, F., Yan, K., Zhang, J., Huang, Y., Zhou, M., and Xiong, Z. Deep fourier-based exposure correction network with spatial-frequency interaction. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIX*, pp. 163–180. Springer, 2022a.

Huang, J., Zhou, M., Liu, Y., Yao, M., Zhao, F., and Xiong, Z. Exposure-consistency representation learning for exposure correction. In *Proceedings of the 30th ACM International Conference on Multimedia*, pp. 6309–6317, 2022b.

Huang, J., Zhao, F., Zhou, M., Xiao, J., Zheng, N., Zheng, K., and Xiong, Z. Learning sample relationship for exposure correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9904–9913, 2023.

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

Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., and Wang, Z. Enlightengan: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, 30:2340–2349, 2021.

Kim, J., Lee, J. K., and Lee, K. M. Accurate image super-resolution using very deep convolutional networks. In *CVPR*, pp. 1646–1654, 2016.

Kim, K. I. and Kwon, Y. Single-image super-resolution using sparse regression and natural image prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1127–1133, 2010.

Laben, C. and Brower, B. Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening. *US Patent 6011875A*, 2000.

Li, B., Peng, X., Wang, Z., Xu, J., and Feng, D. AOD-Net: All-in-one dehazing network. In *ICCV*, pp. 4770–4778, 2017.

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

Li, C., Guo, J.-C., Cong, R.-M., Pang, Y.-W., and Wang, B. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. *IEEE Transactions on Image Processing*, 25(12): 5664–5677, 2016.

Li, C., Guo, C., and Chen, C. L. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

Li, C., Guo, C.-L., Zhou, M., Liang, Z., Zhou, S., Feng, R., and Loy, C. C. Embedding fourier for ultra-high-definition low-light image enhancement. *arXiv preprint arXiv:2302.11831*, 2023.

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

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

Liao, W., Xin, H., Coillie, F. V., Thoonen, G., and Philips, W. Two-stage fusion of thermal hyperspectral and visible rgb image by pca and guided filter. In *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2017.

Liu., J. G. Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details. *International Journal of Remote Sensing*, 21(18):3461–3472, 2000.

Liu, R., Mu, P., Yuan, X., Zeng, S., and Zhang, J. A generic first-order algorithmic framework for bi-level programming beyond lower-level singleton. In *ICML*, pp. 6305–6315, 2020.

Liu, R., Ma, L., Zhang, J., Fan, X., and Luo, Z. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *CVPR*, pp. 10561–10570, 2021a.

Liu, X., Ma, Y., Shi, Z., and Chen, J. GridDehazeNet: Attention-based multi-scale network for image dehazing. In *ICCV*, pp. 7314–7323, 2019.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. CVPR, pp. 10012–10022, 2021b.

Lu, K. and Zhang, L. Tbefn: A two-branch exposure-fusion network for low-light image enhancement. *IEEE Transactions on Multimedia*, 23:4093–4105, 2020.

Lv, F., Lu, F., Wu, J., and Lim, C. Mbllen: Low-light image/video enhancement using cnns. In *BMVC*, volume 220, pp. 4, 2018.

Ma, L., Liu, R., Zhang, J., Fan, X., and Luo, Z. Learning deep context-sensitive decomposition for low-light image enhancement. *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

Masi, G., Cozzolino, D., Verdoliva, L., and Scarpa, G. Pansharpening by convolutional neural networks. *Remote Sensing*, 8(7):594, 2016.

Qin, X., Wang, Z., Bai, Y., Xie, X., and Jia, H. FFA-Net: Feature fusion attention network for single image dehazing. In *AAAI*, volume 34, pp. 11908–11915, 2020.

Ren, D., Zuo, W., Hu, Q., Zhu, P., and Meng, D. Progressive image deraining networks: A better and simpler baseline. In *CVPR*, pp. 3937–3946, 2019.

Ren, W., Ma, L., Zhang, J., Pan, J., Cao, X., Liu, W., and Yang, M.-H. Gated fusion network for single image dehazing. In *CVPR*, pp. 3253–3261, 2018.

Tai, T., Yang, J., and Liu, X. Image super-resolution via deep recursive residual network. In *CVPR*, 2017.

Tolstikhin, I. O., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., Yung, J., Steiner, A., Keysers, D., Uszkoreit, J., et al. Mlp-mixer: An all-mlp architecture for vision. *NeurIPS*, 34:24261–24272, 2021.

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

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *NeurIPS*, 30, 2017.

Wang, W., Wei, C., Yang, W., and Liu, J. Gladnet: Low-light enhancement network with global awareness. In *ICAFGR*, pp. 751–755, 2018.

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

Wei, W., Meng, D., Zhao, Q., Xu, Z., and Wu, Y. Semi-supervised transfer learning for image rain removal. In *CVPR*, pp. 3877–3886, 2019.

Wu, H., Qu, Y., Lin, S., Zhou, J., Qiao, R., Zhang, Z., Xie, Y., and Ma, L. Contrastive learning for compact single image dehazing. In *CVPR*, pp. 10551–10560, 2021.

Wu, W., Weng, J., Zhang, P., Wang, X., Yang, W., and Jiang, J. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *CVPR*, pp. 5901–5910, 2022.

Xia, Z., Pan, X., Song, S., Li, L. E., and Huang, G. Vision transformer with deformable attention. In *CVPR*, pp. 4794–4803, 2022.

Xiao, J., Zhou, M., Fu, X., Liu, A., and Zha, Z.-J. Improving de-raining generalization via neural reorganization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4987–4996, 2021.

Xiao, J., Fu, X., Liu, A., Wu, F., and Zha, Z.-J. Image de-raining transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–18, 2022.

Xu, K., Yang, X., Yin, B., and Lau, R. W. Learning to restore low-light images via decomposition-and-enhancement. In *CVPR*, pp. 2281–2290, 2020.

Xu, S., Zhang, J., Zhao, Z., Sun, K., Liu, J., and Zhang, C. Deep gradient projection networks for pan-sharpening. In *CVPR*, pp. 1366–1375, 2021.

Xu, X., Wang, R., Fu, C.-W., and Jia, J. SNR-Aware low-light image enhancement. In *CVPR*, pp. 17714–17724, June 2022.

Yan, K., Zhou, M., Liu, L., Xie, C., and Hong, D. When pan-sharpening meets graph convolution network and knowledge distillation. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022.

Yang, J., Fu, X., Hu, Y., Huang, Y., Ding, X., and Paisley, J. PanNet: A deep network architecture for pan-sharpening. In *ICCV*, pp. 5449–5457, 2017a.

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

Yang, W., Wang, S., Fang, Y., Wang, Y., and Liu, J. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *CVPR*, pp. 3063–3072, 2020.

Yasarla, R. and Patel, V. M. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In *CVPR*, pp. 8405–8414, 2019.

Yu, H., Zheng, N., Zhou, M., Huang, J., Xiao, Z., and Zhao, F. Frequency and spatial dual guidance for image dehazing. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIX*, pp. 181–198. Springer, 2022a.

Yu, W., Luo, M., Zhou, P., Si, C., Zhou, Y., Wang, X., Feng, J., and Yan, S. Metaformer is actually what you need for vision. In *CVPR*, pp. 10819–10829, 2022b.

Yuan, Q., Wei, Y., Meng, X., Shen, H., and Zhang, L. A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(3):978–989, 2018.

Yuhas, A. F. G. and Boardman, J. M. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm. *Proc. Summaries Annu. JPL Airborne Geosci. Workshop*, pp. 147–149, 1992.

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

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

Zhang, H. and Patel, V. M. Density-aware single image de-raining using a multi-stream dense network. In *CVPR*, pp. 695–704, 2018.

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

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., and Fu, Y. Image super-resolution using very deep residual channel attention networks. In *ECCV*, pp. 286–301, 2018.

Zhang, Y., Zhang, J., and Guo, X. Kindling the darkness: A practical low-light image enhancer. In *ACMMM*, pp. 1632–1640, 2019.

Zhang, Y., Tian, Y., Kong, Y., Zhong, B., and Fu, Y. Residual dense network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43 (7):2480–2495, 2020.

Zhang, Y., Guo, X., Ma, J., Liu, W., and Zhang, J. Beyond brightening low-light images. *International Journal of Computer Vision*, 129(4):1013–1037, 2021a.

Zhang, Y., Li, K., Li, K., Sun, G., Kong, Y., and Fu, Y. Accurate and fast image denoising via attention guided scaling. *IEEE Transactions on Image Processing*, 30: 6255–6265, 2021b.

Zhao, Z., Xiong, B., Wang, L., Ou, Q., Yu, L., and Kuang, F. Retinexdip: A unified deep framework for low-light image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(3):1076–1088, 2021.

Zhou, M. and Wang, R. Control theory-inspired model design for single image de-raining. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 69(2):649–653, 2021.

Zhou, M., Wang, F., Wei, X., Wang, R., and Wang, X. Pid controller-inspired model design for single image deraining. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 69(4):2351–2355, 2021a.

Zhou, M., Xiao, J., Chang, Y., Fu, X., Liu, A., Pan, J., and Zha, Z.-J. Image de-raining via continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4907–4916, 2021b.

Zhou, M., Fu, X., Huang, J., Zhao, F., Liu, A., and Wang, R. Effective pan-sharpening with transformer and invertible neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022a.

Zhou, M., Yan, K., Huang, J., Yang, Z., Fu, X., and Zhao, F. Mutual information-driven pan-sharpening. In *CVPR*, pp. 1798–1808, 2022b.

zhou, m., Yu, H., Huang, J., Zhao, F., Gu, J., Loy, C. C., Meng, D., and Li, C. Deep fourier up-sampling. In *Advances in Neural Information Processing Systems*, volume 35, pp. 22995–23008. Curran Associates, Inc., 2022.