This article has been accepted for publication in IEEE Transactions on Circuits and Systems for Video Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2024.3429557

Omni-Kernel Modulation for Universal Image Restoration

Yuning Cui, Wenqi Ren and Alois Knoll, *Fellow, IEEE*

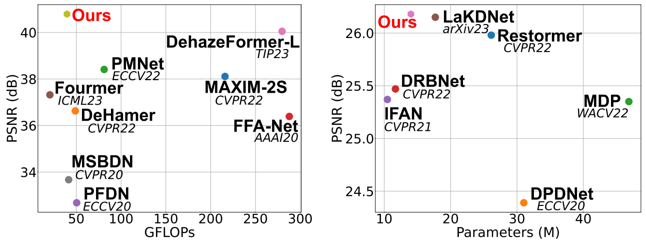
***Abstract*—Image restoration is the process of recovering a clean image from a degraded observation. In order to achieve this, it is essential to refine features at multiple scales. This paper develops an effective omni-kernel modulation module to enhance multi-scale representation learning for image restoration. The module consists of three branches, namely global, large, and local branches, which are designed to learn global-to-local feature representations efficiently. Specifically, the global branch achieves a global perceptive field via the dual-domain channel attention and frequency-gated mechanism. Furthermore, to provide multi-grained receptive fields, the large branch is formulated using different shapes of depth-wise convolutions with unusually large kernel sizes. Moreover, we complement local information with a point-wise depth-wise convolution. Finally, we demonstrate the effectiveness of our omni-kernel modulation module in two cases: general image restoration and all-in-one image restoration tasks. Incorporating our method into a convolutional backbone results in a model that achieves state-of-the-art performance on the 15 datasets for three representative image restoration tasks, including image dehazing, desnowing, and defocus deblurring. Moreover, by integrating our module into a pure Transformer-based backbone, the model demonstrates competitive performance against state-of-the-art algorithms in two all-in-one image restoration settings: the three-task and five-task settings.**

***Index Terms*—Omni-kernel modulation, image restoration, all-in-one image restoration.**

I. INTRODUCTION

IMAGE restoration aims to reconstruct a sharp image from a degraded counterpart, which may suffer from degradations such as haze, noise, and blur [1], [2]. To deal with this longstanding and ill-posed problem, conventional approaches have utilized various hand-crafted features and assumptions to restrict the solution space. However, these approaches are not applicable to more challenging real-world scenarios.

In recent years, convolutional neural networks (CNNs) have demonstrated superior performance over traditional approaches on image restoration tasks by learning generalizable priors from large-scale datasets [5]–[7]. To boost performance, many sophisticated functional units have been developed or borrowed from other domains for image restoration, such as the encoder-decoder architecture [8], residual connection [9], dilated convolution [10], and attention mechanisms [11], [12]. More recently, Transformer models have been introduced into image restoration, significantly advancing the state-of-the-art performance [1], [13], [14]. However, pursuing large-scale receptive fields using self-attention leads to high complexity.



(a) (b)

Fig. 1. (a) FLOPs *vs.* PSNR on the SOTS-Indoor [3] dataset for image dehazing. (b) The number of parameters *vs.* PSNR on the DPDD [4] dataset for image defocus deblurring. Our network achieves a better trade-off between performance and computation overhead over other state-of-the-art algorithms.

In contrast to the convolution operator, which has limited receptive fields, the Transformer can effectively model long-range dependencies. Inspired by this operation, a few recent works on CNNs strike back by designing efficient frameworks with large kernel convolutions, such as 31 × 31 in RepLKNet [15] and 51 × 51 in SLaK [16]. In the context of image restoration, LKDNet [17] decomposes a 21 × 21 convolution into a smaller depth-wise convolution and a depth-wise dilated convolution for image dehazing. LaKDNet [18] leverages large kernel convolutions (*e.g.,* 9 × 9) followed by point-wise convolutions to obtain large effective receptive fields for image deblurring. MAN [19] decomposes a large kernel convolution into a depth-wise convolution, a depth-wise dilated convolution, and a point-wise convolution. Nevertheless, the receptive fields generated by these methods remain constrained, and they do not provide global receptive fields.

This paper examines the potential of unusual large kernel convolutions for image restoration by utilizing a 63 × 63 depth-wise convolution. Furthermore, large strip-based convolutions are used to enhance representation learning for high-fidelity reconstruction. To mitigate the computational overhead associated with these large convolutions, they are deployed only in the bottleneck position. Moreover, we employ dual-domain channel attention and a frequency-gated mechanism to obtain global receptive fields. In addition to pursuing large receptive fields, we leverage a 1×1 depth-wise convolution to complement local information for small degradations. Finally, the omni-kernel modulation (OKM) is formed by organizing the abovementioned designs in parallel, thereby enhancing the ability of the network to handle multi-scale degradations.

The employment of OKM in the bottleneck of a CNN-based backbone results in a model that achieves state-of-the-art performance on 15 datasets for three representative image restora-

---

Y. Cui and A. Knoll are with School of Computation, Information and Technology, Technical University of Munich, Munich, Germany.

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

This article has been accepted for publication in IEEE Transactions on Circuits and Systems for Video Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2024.3429557

tion tasks. In particular, our model demonstrates a notable superiority over the recent Transformer model Fourmer [20] by 3.47 dB PSNR with comparable complexity on the SOTS-Indoor [3] dataset (see Figure 1 (a)). For single-image defocus deblurring, our model achieves a performance gain of 0.2 dB PSNR over the strong Transformer model Restormer [1] in the combined category of the DPDD [4] dataset, while consuming 46% fewer parameters. Moreover, the model demonstrates remarkable proficiency in the image desnowing task, outperforming the recent algorithm IRNeXt [12] by 0.7 dB PSNR on the widely used CSD [21] dataset.

Furthermore, our OKM can effectively serve the all-in-one image restoration. When incorporated into a pure Transformer-based backbone, our model outperforms PromptIR [22] by 0.19 dB PSNR when averaged across three image restoration tasks under the all-in-one setting. Besides, compared to IDR [23], our model obtains an average performance improvement of 1.88 dB under the five-task setting. Overall, the contributions of this study are summarized as follows:

- We present an omni-kernel modulation module that efficiently captures multi-scale receptive fields for image restoration. The large-scale information is modulated via dual-domain processing and large kernel convolutions of different shapes.
- By incorporating the proposed plug-and-play module into a CNN-based backbone, the model achieves state-of-the-art performance on 15 widely used benchmark datasets for three image restoration tasks, including image dehazing, desnowing, and defocus deblurring.
- The Transformer model, equipped with the proposed module, performs favorably against state-of-the-art algorithms under two all-in-one image restoration settings.

This paper is an extension of the conference paper [24]. The improvements made to the previous version are as follows:

- The established CNN-based model is extended to three remote sensing image dehazing datasets and obtains competitive performance compared to general and elaborately designed algorithms. Furthermore, we utilize another nighttime dehazing dataset, GTA5 [25], for evaluation.
- We investigate the effectiveness of our module under two all-in-one image restoration settings. The OKM-based model exhibits superior performance compared to the recent PromptIR [22] under the three-task setting, with a 0.19 dB PSNR improvement. Furthermore, our model demonstrates a notable performance enhancement of 1.88 dB PSNR over the IDR [23] model under the five-task setting. Additionally, we present experimental results for the single-task setting in accordance with PromptIR [22] and for evaluating the generalization ability following [23].
- We conduct more ablation studies in general and all-in-one image restoration to confirm the efficacy of OKM.

II. RELATED WORKS

A. Image Restoration

As a challenging problem, image restoration aims to recover a clean image from its degraded version, playing an important role in many scenarios. Due to its highly ill-posed property, many conventional algorithms have been proposed mainly based on various assumptions and hand-crafted features, which are inapplicable to complicated practical applications.

Recently, deep learning methods have demonstrated superior performance compared to traditional competitors by learning generalizable priors from large-scale datasets [26]. These frameworks can be broadly classified into CNN-based and Transformer-based methods. CNN-based methods have long dominated image restoration by designing advanced functional units [27]. For example, mixed convolution attention is utilized in [28] to reduce feature redundancy based on depth-wise convolutions of different kernel sizes. PGC-Net [29] leverages multi-scale patch-wise features for modulation. SFNet [30] employs a dynamic selective frequency module to identify the most informative frequency for reconstruction. To model long-range dependencies, Transformer [31] has been incorporated into image restoration [14], [32]. To reduce the complexity of self-attention, common practices include restricting self-attention regions [13], [33] and switching from the spatial dimension to the channel dimension [1]. Despite these measures, Transformer models remain expensive for pursuing long-range dependencies. Moreover, the Transformer block is unable to capture multi-scale receptive fields. This paper presents an efficient and effective modulation module that is capable of learning multi-scale representations, including global information.

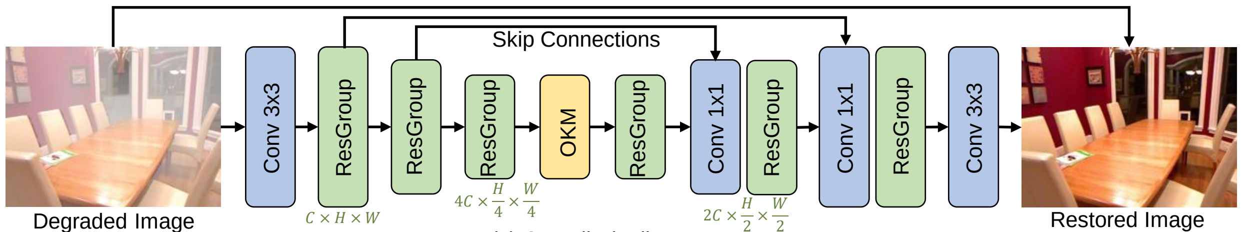
B. Large Kernel Network

In recent years, inspired by the plausible reasons behind the success of Transformer, namely the long-range dependencies modeling ability, CNN-based methods have employed large kernel convolutions in an attempt to rival the effectiveness of Transformer. For example, RepLKNet [15] achieves a kernel size of 31×31 by following several guidelines for designing large convolutions, thereby significantly narrowing the performance gap between CNNs and Transformer models. SLaK [16] employs sparse factorized 51×51 kernels to confront Transformer methods. In the realm of image restoration, LaKDNet [18] employs a combination of large kernel (9×9) depth-wise convolutions and point-wise convolutions to expand the effective receptive field. MAN [19] develops the large kernel attention by decomposing a convolution into three distinct types of convolutions. LKD-Net [17] decomposes a depth-wise convolution into a smaller depth-wise convolution and a depth-wise dilated convolution. Our strategy differs from the aforementioned schemes in four aspects: **(a)** We investigate the potential of unusually large kernel convolutions for image restoration, specifically 63×63; **(b)** In addition to the square depth-wise convolution, we employ strip versions in different directions to provide diverse shapes of receptive fields; **(c)** The dual-domain processing is employed to achieve full-size receptive fields. **(d)** An extremely lightweight 1×1 depth-wise convolution is used to complement local information.

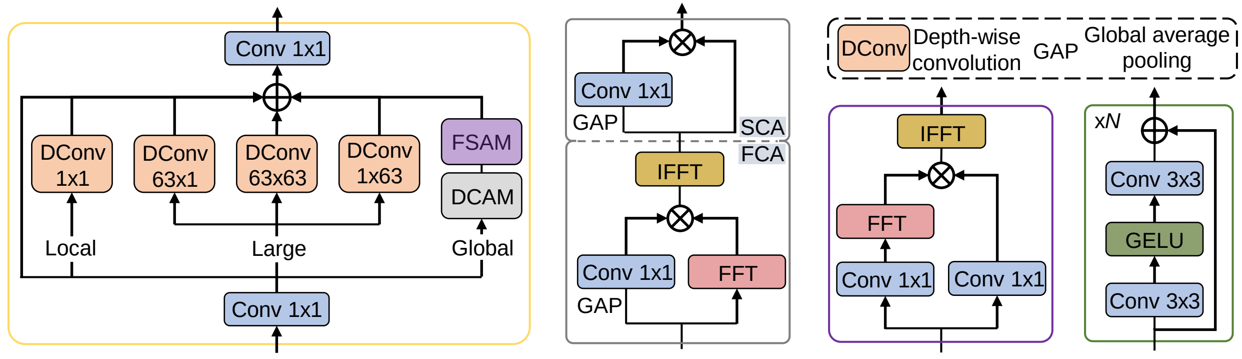
C. Global Modeling Network

Transformer has become renowned for their capacity for modeling long-range dependencies [1], [13], [34], [35]. However, these methods cannot fully capture global information

This article has been accepted for publication in IEEE Transactions on Circuits and Systems for Video Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2024.3429557



(a) Overall Pipeline



(b) OKM (c) DCAM (d) FSAM (e) ResGroup

Fig. 2. The architecture of the CNN-based backbone involving our OKM. FFT and IFFT denote fast Fourier transform and its inverse operation, respectively.

due to the quadratic complexity of self-attention. Recently, researchers have sought to bypass this limitation by leveraging frequency learning techniques. A common practice is to adopt existing tools, such as the fast Fourier transform [36], wavelet transform [10], and global average pooling [37], to generate explicit or implicit frequency features from spatial inputs [38]. These features are then modulated through convolutions, and if necessary, the inverse transform converts the modulated features into the spatial domain for further processing. In contrast to these approaches, our frequency modulation is realized via a frequency-gated mechanism, wherein the attention weights are learned from the inputs and imposed on features via multiplication. This improves the adaptability of the model to manage degradations of different types and degrees. Besides, the dual-domain channel-wise operator also enhances the representation learning ability of the model.

D. *All-in-One Image Restoration*

Although task-specific and general image restoration approaches have produced promising performance on various tasks, users must select the appropriate version for the current task based on prior knowledge. Furthermore, employing different copies for different degradation types and levels is impractical for resource-constrained equipment. To address these issues, all-in-one image restoration has emerged as a promising topic by dealing with multiple degradations using a unified model [39], [40]. For example, AirNet [41] employs a contrastive learning strategy to differentiate diverse degradation types. PromptIR [22] pre-defines a pool of prompts to guide the restoration process. IDR [23] first establishes the task-oriented prior hubs, which are used to perform restoration in the second stage. These methods are designed to distinguish different degradation types. In contrast, our OKM introduces the enhanced multi-scale representation learning ability into the all-in-one model, achieving promising performance.

III. METHODOLOGY

We begin by describing the pipeline of the CNN-based model, which involves our omni-kernel modulation module (OKM). We then delineate the architectural details of OKM.

A. *Overall Pipeline*

As shown in Figure 2 (a), the model adopts the encoder-decoder architecture, consisting of three scales in both the encoder and decoder stages. The ResGroup comprises multiple residual blocks, each with two 3×3 convolutions and the GELU [42] nonlinearity in between. OKM is only inserted into the bottleneck position, where features have the lowest resolution, in order to reduce the computational overhead.

Given an input image $I \in \mathbb{R}^{3 \times H \times W}$, we first leverage a 3×3 convolution to project the image into embedding features of size $C \times H \times W$, where C denotes the number of channels, and $H \times W$ specifies the spatial pixels. Next, the resulting features are fed into the encoder stage to extract in-depth representations. The downsampling operation is implemented by a strided convolution, which expands the number of channels while reducing the spatial dimension. After being processed by the proposed OKM, the features pass through the decoder network to restore high-resolution representations. During this process, the decoder features are concatenated with the encoder features to assist in recovery, followed by a 1×1 convolution to reduce the number of channels by half. The upsampling layer is accomplished by a transposed convolution, which enlarges the spatial dimension and reduces the number of channels. Finally, a 3×3 convolution generates the learned residual image, which is then added to the input image to produce the final restored output.

This article has been accepted for publication in IEEE Transactions on Circuits and Systems for Video Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2024.3429557

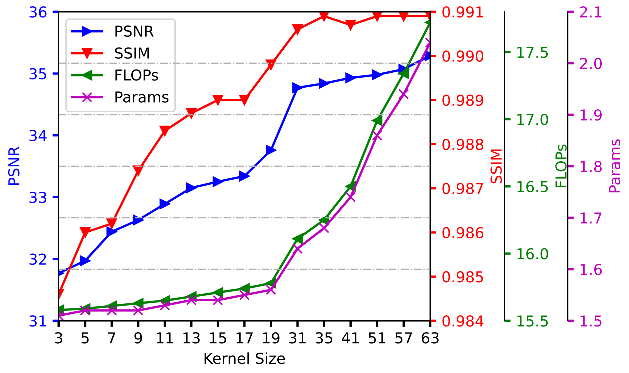


Fig. 3. Experimental results for using depth-wise convolutions of different kernel sizes in the large branch.

B. Omni-Kernel Modulation (OKM)

The architecture of OKM is shown in Figure 2 (b). Given features $X \in \mathbb{R}^{C\times H\times W}$, after being processed by a 1×1 convolution, the features are fed into three branches, namely local, large, and global branches, to enhance multi-scale representation learning. The results of the three branches are fused by addition and modulated by another 1×1 convolution. Next, we introduce the internal components these branches.

1) Large Branch: In this branch, we apply a cost-effective depth-wise convolution of kernel size $K \times K$ in order to achieve a large receptive field. In addition to the regular depth-wise convolution, we also employ $1 \times K$ and $K \times 1$ depth-wise convolutions in parallel to the square one to harvest strip-shaped contextual information, inspired by strip-based self-attention [33], [43]. To circumvent the introduction of a significant computational burden associated with large kernel convolutions, the module is placed in the bottleneck position.

The potential of employing large convolutions for image restoration is investigated by gradually increasing the kernel size K. The experimental results are presented in Figure 3. In general, the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) metrics increase as the kernel size is enlarged from $K = 3$ to $K = 63$. The employment location of the module allows for the adoption of an unusually large kernel size to capture large-scale different shapes of receptive fields with few parameters and low complexity. Ultimately, $K = 63$ is selected in the large branch of OKM for enhanced performance. The cost-effective large branch yields a notable performance enhancement of 3.97 dB PSNR while introducing only additional 0.56M parameters and 2.28 GFLOPs.

2) Global Branch: Image restoration frameworks are primarily trained on cropped image patches, with the bottleneck features typically having a spatial size of 64×64. Consequently, we adopt the largest odd kernel in the large branch. However, the full-size images are considerably larger than the training patches. As a result, a 63×63 kernel cannot cover the global field. To alleviate this issue, we superadd the global modeling capability by employing dual-domain processing in the global branch. The global branch comprises a dual-domain *channel* attention module (DCAM) and a frequency-based *spatial*[1] attention module (FSAM).

Given features $X_{Global} \in \mathbb{R}^{C\times H\times W}$, the DCAM first applies frequency channel attention (FCA) to X_{Global} as follows:

$$X_{\mathrm{FCA}} = \mathcal{IF}(\mathcal{F}(X_{Global}) \otimes W^{\mathrm{FCA}}_{1\times 1}(\mathrm{GAP}(X_{Global}))) \quad (1)$$

where $\mathcal{F}$ and $\mathcal{IF}$ are fast Fourier transform and its inverse operator, respectively; X_{FCA}, $W_{1\times 1}$ and GAP indicate the output of FCA, a 1×1 convolution and global average pooling, respectively; $\otimes$ is element-wise multiplication. The Fourier processing enables the refinement of global features in an efficient manner in accordance with the convolution theorem. The resulting features are subsequently fed into the spatial channel attention module (SCA), which is expressed as:

$$X_{\mathrm{DCAM}} = X_{\mathrm{FCA}} \otimes W^{\mathrm{SCA}}_{1\times 1}(\mathrm{GAP}(X_{\mathrm{FCA}})) \quad (2)$$

where X_{DCAM} is the output of DCAM. DCAM only enhances dual-domain features at the channel-wise coarse granularity. Then, we apply the frequency-based attention module among the spatial dimension to refine the spectrum at a fine-grained level, which is formally expressed as:

$$X_{\mathrm{FSAM}} = \mathcal{IF}(\mathcal{F}(W^{1}_{1\times 1}(X_{\mathrm{DCAM}})) \otimes W^{2}_{1\times 1}(X_{\mathrm{DCAM}})) \quad (3)$$

where X_{FSAM} is the output of FSAM. By doing this, the model attends to informative frequency components for high-fidelity image reconstruction.

3) Local Branch: In light of the crucial role played by local information in image restoration [1], [13], we introduce an extremely lightweight yet effective local branch for local signals modulation using a 1×1 depth-wise convolutional layer, as illustrated in Figure 2 (b). The efficacy of this approach is demonstrated in Table VII.

IV. EXPERIMENTS: GENERAL IMAGE RESTORATION

We apply the previously introduced CNN-based model to general image restoration tasks by training separate copies for different datasets. The experiments are conducted on 15 datasets for three tasks: image dehazing, image defocus deblurring, and image desnowing. In the tables, the best and second best results are marked in **bold** and underlined.

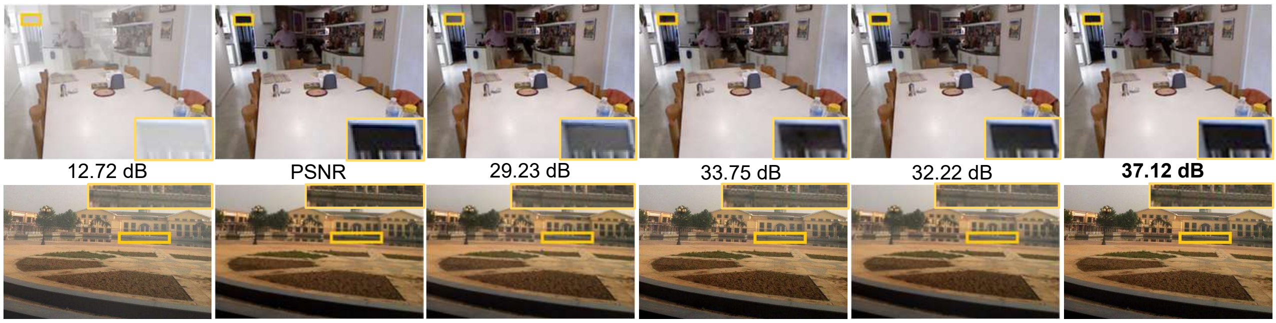
A. Implementation Details

According to the task complexity, we scale the model by setting different numbers of residual blocks in each ResGroup, *i.e.,* $N = 4$ for dehazing and desnowing, and $N = 16$ for deblurring. Unless stated otherwise, the following hyperparameters are adopted. The models are trained using the Adam optimizer with dual-domain L_1 loss functions [24]. The batch size is set to 8. The initial learning rate is set to $2e^{-4}$ and gradually decreased to $1e^{-6}$ using cosine annealing. For data augmentation, the cropped patches of size 256×256 are randomly horizontally flipped with a probability of 0.5. FLOPs are calculated on the 256×256 patch size.

---

[1] This *spatial* indicates the spatial dimension instead of the spatial domain.

This article has been accepted for publication in IEEE Transactions on Circuits and Systems for Video Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2024.3429557


12.72 dB | PSNR | 29.23 dB | 33.75 dB | 32.22 dB | **37.12 dB**

18.26 dB Input | PSNR Target | 29.17 dB GridDehazeNet | 26.98 dB FFA-Net | 24.40 dB MAXIM-2S | **33.18 dB Ours**

Fig. 4. Image dehazing comparisons on the SOTS [3] dataset. The top and bottom images are obtained from SOTS-Indoor and SOTS-Outdoor, respectively.

TABLE I

IMAGE DEHAZING COMPARISONS ON THE DAYTIME SYNTHETIC AND REAL-WORLD DATASETS.

Method	SOTS-Indoor [3]		SOTS-Outdoor [3]		Dense-Haze [44]		NH-HAZE [45]		O-HAZE [46]		I-Haze [47]	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
DCP [48]	16.62	0.818	19.13	0.815	10.06	0.39	10.57	0.52	16.78	0.65	12.38	0.40
AOD-Net [49]	20.51	0.816	24.14	0.920	13.14	0.41	15.40	0.57	15.03	0.54	10.86	0.41
GridDehazeNet [50]	32.16	0.984	30.86	0.982	13.31	0.37	13.80	0.54	23.51	0.83	18.73	0.77
MSBDN [51]	33.67	0.985	33.48	0.982	15.37	0.49	19.23	0.71	24.36	0.75	19.62	0.62
FFA-Net [52]	36.39	0.989	33.57	0.984	14.39	0.45	19.87	0.69	22.12	0.77	19.72	0.73
AECR-Net [53]	37.17	0.990	-		15.80	0.47	19.88	0.72	-		-	
PMNet [54]	38.41	0.990	34.74	0.985	16.79	0.51	20.42	0.73	24.64	0.83	-	
MAXIM-2S [55]	38.11	0.991	34.19	0.985	-		-		-		-	
DeHamer [35]	36.63	0.988	35.18	0.986	16.62	0.56	**20.66**	0.68	-		-	
SDCE [38]	-		-		16.85	0.60	20.42	0.74	24.92	0.84	20.81	0.82
DehazeFormer-L [34]	40.05	**0.996**	-		-		-		-		-	
Fourmer [20]	37.32	0.990	-		15.95	0.49	19.91	0.72	-		-	
Ours	**40.79**	**0.996**	**37.68**	**0.995**	**16.92**	**0.64**	20.48	**0.80**	**25.64**	**0.94**	**21.72**	**0.87**

B. Datasets

1) Image Dehazing: We evaluate the method on four types of dehazing datasets: daytime synthetic/real-world, nighttime, and remote sensing dehazing datasets. We adopt RESIDE [3] for daytime synthetic scenarios. Separate models are trained on its two subsets: the indoor training set (ITS) and the outdoor training set (OTS). Subsequently, the ITS-trained and OTS-trained models are applied to the test sets of RESIDE (SOTS-Indoor and SOTS-Outdoor), respectively. The model is trained on ITS for 1000 epochs. For OTS, the model is trained for 30 epochs with an initial learning rate of $1e^{-4}$.

Additionally, the model is evaluated on four real-world datasets. Following [35], [37], the models are trained for 5000 epochs on a patch size of 800×1200 with a batch size of 2. Furthermore, we evaluate the model on two nighttime datasets, NHR [56] and GTA5 [25], by separately training the model for 300 epochs with an initial learning rate of $1e^{-4}$.

Besides, the model is applied to remote sensing with SateHaze1k [57]. The model is individually trained on its three subsets (Thin, Moderate, and Thick) for 1000 epochs with a batch size of 32 and an initial learning rate of $8e^{-4}$.

2) Image Desnowing: We evaluate our model on three datasets for image desnowing, including Snow100K [6], SRRS [58], and CSD [21]. We adopt the same dataset configuration as previous methods [12], [37], [58] for fairness. The model is trained for 2000 epochs on each dataset.

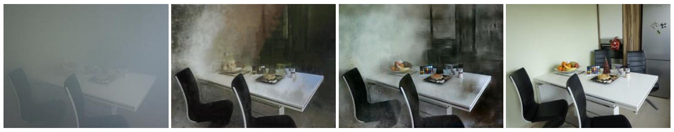

12.24 dB Input | 12.02 dB DeHamer | **13.57 dB Ours** | PSNR Target

Fig. 5. Real-world dehazing comparisons on the Dense-Haze [44] dataset.

3) Image Defocus Deblurring: Consistent with [1], [9], we utilize DPDD [4] for evaluation. This dataset comprises 500 indoor/outdoor scenes captured with a DSLR camera. Each scene contains four images labeled as right view, left view, center view, and all-in-focus ground truth. We train the model by feeding the center view images and reducing the discrepancy between the predicted output and the ground truth.

C. Experimental Results

1) Image Dehazing Results: We conduct dehazing experiments on four kinds of datasets: a daytime synthetic dataset (RESIDE [3]), four daytime real-world datasets (Dense-Haze [44], NH-HAZE [45], O-Haze [46], and I-Haze [47]), two nighttime datasets (NHR [56], GTA5 [25]), and a remote sensing dataset (SateHaze1k [57]). The results for daytime datasets are presented in Table I. Our model achieves the best results on most metrics. Specifically, it outperforms the expen-

This article has been accepted for publication in IEEE Transactions on Circuits and Systems for Video Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2024.3429557

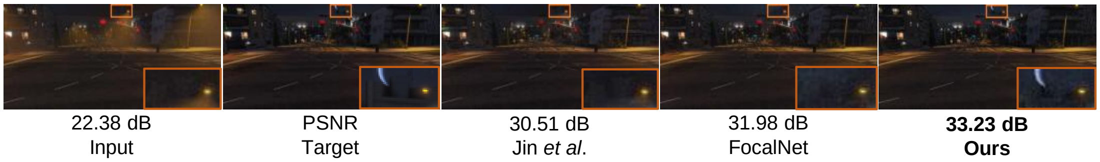


Fig. 6. Nighttime image dehazing comparisons on the GTA5 [25] dataset.

TABLE II
NIGHTTIME IMAGE DEHAZING COMPARISONS ON THE GTA5 [25] DATASET.

Methods	GS [60]	MRP [61]	Ancuti *et al* [62]	Yan *et al* [25]	CycleGAN [63]	Jin *et al* [59]	FocalNet [37]	**Ours**
PSNR↑	21.02	20.92	20.59	27.00	21.75	30.38	30.62	**31.53**
SSIM↑	0.639	0.646	0.623	0.850	0.696	0.904	0.909	**0.920**

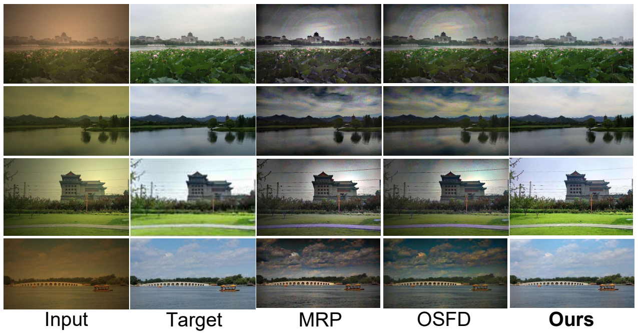


Fig. 7. Nighttime image dehazing comparisons on the NHR [56] dataset.

TABLE III
NIGHTTIME IMAGE DEHAZING COMPARISONS ON THE NHR [56] DATASET.

Method	GS [60]	MRP [61]	OSFD [56]	HCD [64]	FSNet-S [65]	FocalNet [37]	**Ours**
PSNR↑	17.32	19.93	21.32	23.43	24.35	25.35	**27.92**
SSIM↑	0.629	0.777	0.804	0.953	0.965	0.969	**0.979**

TABLE IV
IMAGE DEHAZING RESULTS ON THE REMOTE SENSING SATEHAZE1K [57] DATASET UNDER THREE LEVELS: THIN, MODERATE, AND THICK. MODELS ARE SEPARATELY TRAINED AND TESTED ON EACH SUBSET.

Methods	Thin		Moderate		Thick	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
DCP [48]	13.15	0.725	9.78	0.574	10.25	0.585
AOD-Net [49]	19.54	0.854	20.10	0.885	15.92	0.731
H2RL-Net [66]	20.91	0.880	22.34	0.906	17.41	0.768
FCFT-Net [67]	23.59	0.913	22.88	0.927	20.03	0.816
Uformer [13]	22.82	0.907	24.47	0.939	20.36	0.815
C^{2}PNet [68]	19.62	0.880	24.79	0.940	16.83	0.790
Restormer [1]	23.08	0.912	24.73	0.933	18.58	0.762
Trinity-Net [69]	21.55	0.884	23.35	0.895	20.97	0.823
Ours	**23.80**	**0.973**	**26.77**	**0.978**	**22.36**	**0.953**

TABLE V
IMAGE DESNOWING COMPARISONS ON THE CSD [21], SRRS [58], AND SNOW100K [6] DATASETS.

Method	CSD		SRRS		Snow100K		Params (M)
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	
DesnowNet [6]	20.13	0.81	20.38	0.84	30.50	0.94	15.6
All in One [70]	26.31	0.87	24.98	0.88	26.07	0.88	44
JSTASR [58]	27.96	0.88	25.82	0.89	23.12	0.86	65
HDCW-Net [21]	29.06	0.91	27.78	0.92	31.54	**0.95**	6.99
TransWeather [40]	31.76	0.93	28.29	0.92	31.82	0.93	21.9
FSNet-S [65]	35.33	0.98	31.39	**0.98**	33.36	**0.95**	3.95
FocalNet [37]	37.18	**0.99**	31.34	**0.98**	33.53	**0.95**	3.74
IRNeXt [12]	37.29	**0.99**	**31.91**	**0.98**	33.61	**0.95**	5.46
Ours	**37.99**	**0.99**	31.70	**0.98**	**33.75**	**0.95**	4.72

sive DehazeFormer-L [34] by 0.74 dB PSNR on the SOTS-Indoor [3] dataset with only 14% FLOPs, as shown in Figure 1 (a). Furthermore, our model demonstrates superior performance on all real-world datasets compared to SDCE [38], which is elaborately designed for real-world scenarios. The visual comparisons for SOTS and Dense-Haze are illustrated in Figure 4 and Figure 5, respectively. Our results are more faithful than those of other competitive algorithms.

Additionally, we assess the efficacy of our method on the nighttime dataset, GTA5 [25], wherein the ground truth images are captured in night scenes. Table II shows that our method exhibits a notable enhancement in performance compared to the general method [37] and the specially designed method [59]. Figure 6 shows that our model recovers more details (*e.g.*, the moon) from the nighttime hazy image than the other two algorithms. Furthermore, we present results on the nighttime image dehazing dataset NHR [56], whose ground-truth images are in the daytime scenes. Table III shows that our method outperforms the recent FocalNet [37] by 2.57 dB PSNR and 0.01 SSIM. Figure 7 illustrates that the results yielded by our network are closer to the ground-truth targets.

As image dehazing plays an important role in remote sensing, we apply our model to three subsets of a widely used remote sensing dehazing dataset, SateHaze1k [57]. Table IV shows that our model performs better than other general and task-specific competitors in removing remote sensing degradations of three different degrees. In particular, on SateHaze1k-Thick, our method significantly outperforms the specially devised Trinite-Net [69] by 1.39 dB PSRN and 0.13 SSIM.

2) Image Desnowing Results: We evaluate the proposed model on three widely used datasets for image desnowing, including Snow100K [6], SRRS [58], and CSD [21]. Table V shows that the network exhibits a strong snow removal

This article has been accepted for publication in IEEE Transactions on Circuits and Systems for Video Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2024.3429557

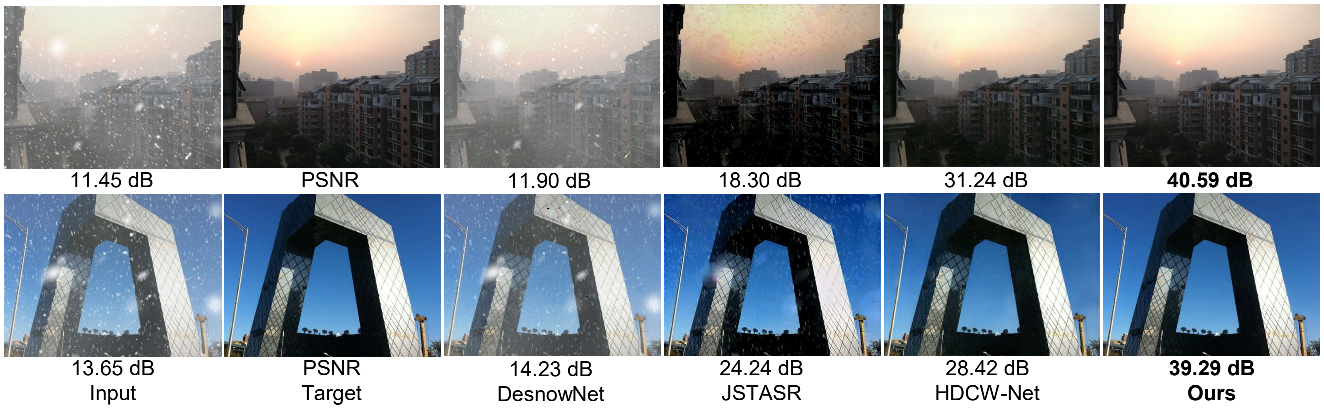


Fig. 8. Image desnowing comparisons on the CSD [21] dataset.

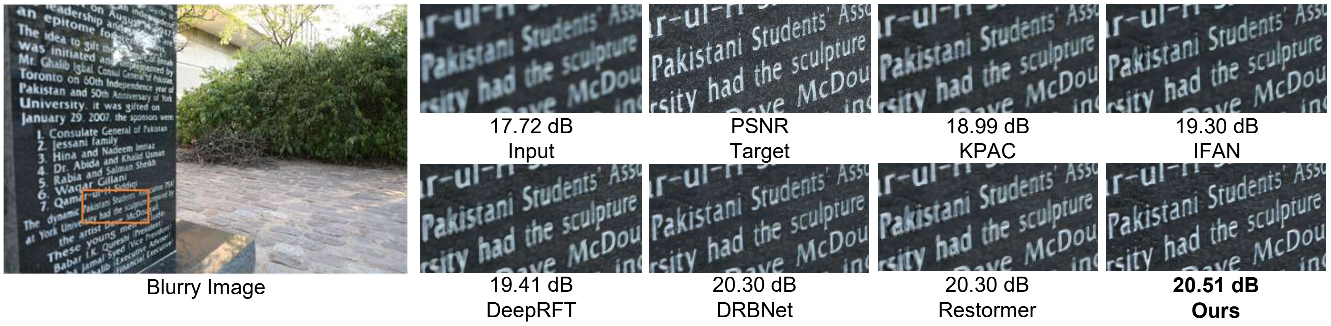


Fig. 9. Image defocus deblurring comparisons on the DPDD [4] dataset.

TABLE VI
SINGLE-IMAGE DEFOCUS DEBLURRING COMPARISONS ON THE DPDD [4] DATASET.

Method	Indoor Scenes				Outdoor Scenes				Combined				Params
	PSNR↑	SSIM↑	MAE↓	LPIPS↓	PSNR↑	SSIM↑	MAE↓	LPIPS↓	PSNR↑	SSIM↑	MAE↓	LPIPS↓	(M)
DPDNet [4]	26.54	0.816	0.031	0.239	22.25	0.682	0.056	0.313	24.34	0.747	0.044	0.277	31.03
KPAC [71]	27.97	0.852	0.026	0.182	22.62	0.701	0.053	0.269	25.22	0.774	0.040	0.227	2.06
DRBNet [9]	-				-				25.73	0.791	-	0.183	11.69
IFAN [72]	28.11	0.861	0.026	0.179	22.76	0.720	0.052	0.254	25.37	0.789	0.039	0.217	10.48
MDP [73]	28.02	0.841	0.027	-	22.82	0.690	0.052	-	25.35	0.763	0.040	0.303	46.86
Restormer [1]	28.87	**0.882**	0.025	**0.145**	23.24	0.743	0.050	**0.209**	25.98	0.811	0.038	0.178	26.16
LaKDNet [18]	-				-				26.15	0.810	-	**0.155**	17.7
Ours	**28.99**	0.877	**0.024**	0.169	**23.51**	**0.751**	**0.049**	0.241	**26.18**	**0.812**	**0.037**	0.206	14.02

capability. Specifically, it outperforms the recent algorithm IRNeXt [12] by 0.14 dB PSNR on the Snow100K dataset while using 14% fewer parameters. On the CSD dataset, which contains more complex snow scenarios, the superiority of our model is more pronounced, indicating the efficacy of our approach in addressing intricate snow degradations. The visual comparisons in Figure 8 illustrate that our results are more visually pleasing than those of other methods.

3) Image Defocus Deblurring Results: We verify the effectiveness of the network for single-image defocus deblurring using the DPDD [4] dataset. Table VI shows that our model achieves superior performance compared to other methods on most metrics. In particular, our model achieves a notable improvement of 0.27 dB PSNR over the strong Transformer model Restormer [1] in outdoor scenes while utilizing only 54% parameters. Moreover, compared to LaKDNet [18], which also employs large kernel convolutions, our model yields performance gains of 0.03 dB PSNR and 0.002 SSIM on the combined category with 21% fewer parameters, as illustrated in Figure 1 (b). Figure 9 shows that our method generates a more visually faithful result than competitors.

D. Ablation Studies

Unless stated otherwise, we perform ablation studies by training the model with one residual block in each ResGroup on ITS [3] for 300 epochs and evaluating on SOTS-Indoor [3]. The baseline is obtained by removing OKM from this model.

1) Effects of Individual Components: We progressively add the designed large branch, small branch, and global branch to the baseline model. Table VII shows that the baseline model achieves 31.32 dB PSNR on SOTS-Indoor. The unusually large regular convolution results in a significant gain of 3.75 dB

This article has been accepted for publication in IEEE Transactions on Circuits and Systems for Video Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2024.3429557

TABLE VII
ABLATION STUDIES FOR THE PROPOSED COMPONENTS IN OKM.

#	Baseline	Large Branch			Small Branch	Global Branch			PSNR↑	SSIM↑	Params/M	FLOPs/G
		63×63	63×1	1×63		DCAM/FCA	DCAM/SCA	FSAM				
1	✓								31.32	0.98357	1.48	15.44
2	✓	✓							35.07	0.99082	2.02	17.65
3	✓	✓	✓						35.15	0.99082	2.03	17.69
4	✓	✓		✓					35.16	0.99093	2.03	17.69
5	✓	✓	✓	✓					35.29	0.99088	2.04	17.72
6	✓	✓	✓	✓	✓				35.48	0.99120	2.04	17.72
7	✓					✓			32.84	0.98748	1.49	15.44
8	✓	✓	✓	✓	✓	✓			35.82	0.99151	2.05	17.72
9	✓						✓		32.35	0.98676	1.49	15.44
10	✓	✓	✓	✓	✓	✓	✓		36.12	0.99188	2.07	17.72
11	✓							✓	33.32	0.98879	1.81	15.57
12	✓	✓	✓	✓	✓	✓	✓	✓	36.48	0.99204	2.40	17.86

TABLE VIII
ABLATION STUDIES FOR THE NUMBER OF RESBLOCKS ON IMAGE DEHAZING AND DESNOWING DATASETS.

ResBlocks	Image Dehazing				Image Desnowing						Params (M)	FLOPs (G)
	SOTS-Indoor [3]		SOTS-Outdoor [3]		CSD [21]		Snow100K [58]		SRRS [6]			
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑		
N=4	40.79	0.996	37.68	0.995	37.99	0.99	33.75	0.95	31.70	0.98	4.72	39.67
N=16	41.82	0.997	39.39	0.996	39.97	0.99	33.95	0.96	32.13	0.98	14.02	126.9

TABLE IX
ABLATION STUDIES FOR FSAM. F AND S DENOTE THE FREQUENCY AND SPATIAL PATHS IN FSAM, RESPECTIVELY. F-F MEANS USING TWO FREQUENCY PATHS.

Method	F-F	S-S	F-S (FSAM)
PSNR↑	23.98	31.35	33.32
SSIM↑	0.925	0.984	0.989

over the baseline. Furthermore, the vertical and horizontal strip-based convolutions produce performance gains of 0.08 dB and 0.09 dB, respectively. The dual-dimensional variant further improves the performance to 35.29 dB, demonstrating the effectiveness of capturing different shapes of receptive fields. The extremely lightweight small branch boosts the accuracy to 35.48 dB by enhancing local information. Finally, we investigate the efficacy of individual components in the global branch. FCA, SCA, and FSAM achieve performance gains of 1.52 dB, 1.03 dB, and 2 dB over the baseline model, respectively. The combination (#10) of FCA and SCA yields a higher score than the use of FCA alone (#8), suggesting the compatibility of the designs. The complete model, which is equipped with FSAM, produces the best performance, with a gain of 5.16 dB over the baseline model.

2) The Number of ResBlocks: We scale the network according to the task complexity. In this part, we apply the large model used for deblurring to other tasks to investigate the impact of the number of ResBlocks. As shown in Table VIII, with the increase in the number of ResBlocks, the results on image dehazing and desnowing improve significantly. Accordingly, the large model requires a high computation overhead. Consequently, in order to achieve a superior balance between accuracy and computation overhead compared to competitors, we adjust the number of Resblocks for each task.

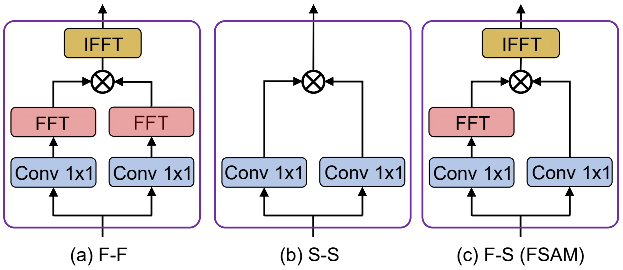


Fig. 10. Architectural comparisons for FSAM.

3) Alternatives to FSAM: We further compare our FSAM with other designs in Table IX. It can be observed that the deployment of isomorphic operations in two paths results in inferior performance compared to our approach, which employs spatial attention weights to modulate spectral features. The architectural comparisons are illustrated in Figure 10.

E. Limitations

As shown in Figure 5, despite the superior visual result over the competitor, there is still a noticeable gap between our result and the ground-truth image. This is due to the limited training samples and the high complexity of real-world scenarios. One promising direction for future research is to explore the utility of synthetic data in addressing real-world degradations using transfer learning or domain adaptation techniques.

V. EXPERIMENTS: ALL-IN-ONE IMAGE RESTORATION

In addition to general image restoration, we apply OKM to all-in-one settings to demonstrate its effectiveness. In this section, we first introduce the datasets and training configurations used in the all-in-one setting. Next, we present the

This article has been accepted for publication in IEEE Transactions on Circuits and Systems for Video Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2024.3429557

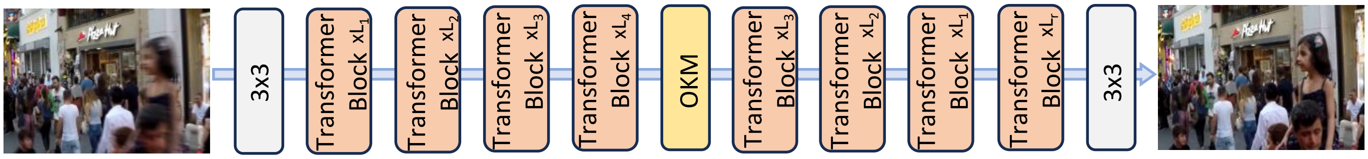

3x3
Transformer Block xL_1
Transformer Block xL_2
Transformer Block xL_3
Transformer Block xL_4
OKM
Transformer Block xL_3
Transformer Block xL_2
Transformer Block xL_1
Transformer Block xL_r
3x3

Fig. 11. The pipeline of the U-shaped Transformer model for all-in-one tasks. The architecture and number of Transformer Block closely follow [1].

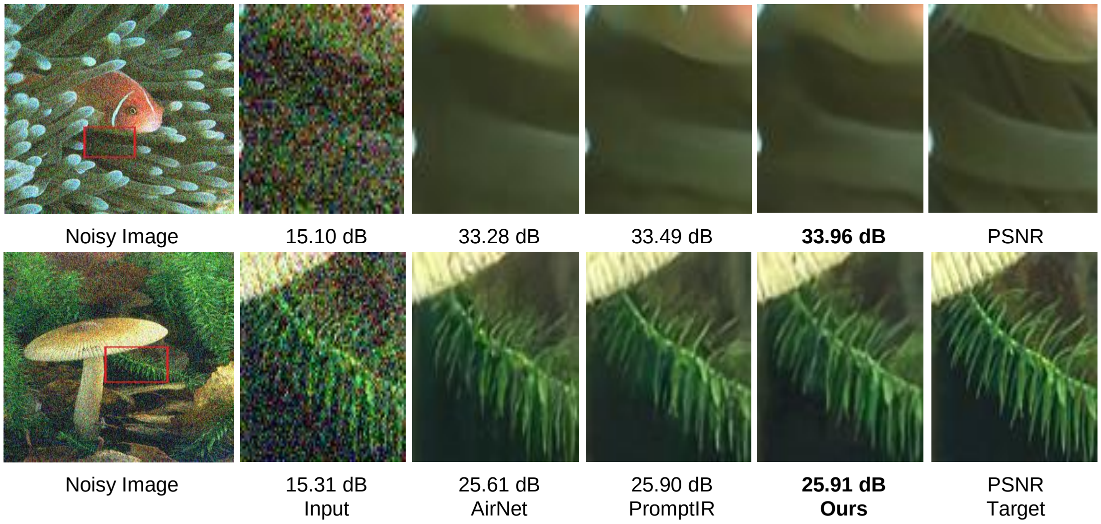

Noisy Image
15.10 dB
33.28 dB
33.49 dB
33.96 dB
PSNR
Noisy Image
15.31 dB
Input
25.61 dB
AirNet
25.90 dB
PromptIR
25.91 dB
Ours
PSNR
Target

Fig. 12. Visual comparisons for image denoising on the BSD68 [74] with $\sigma = 50$ under the three-task setting.

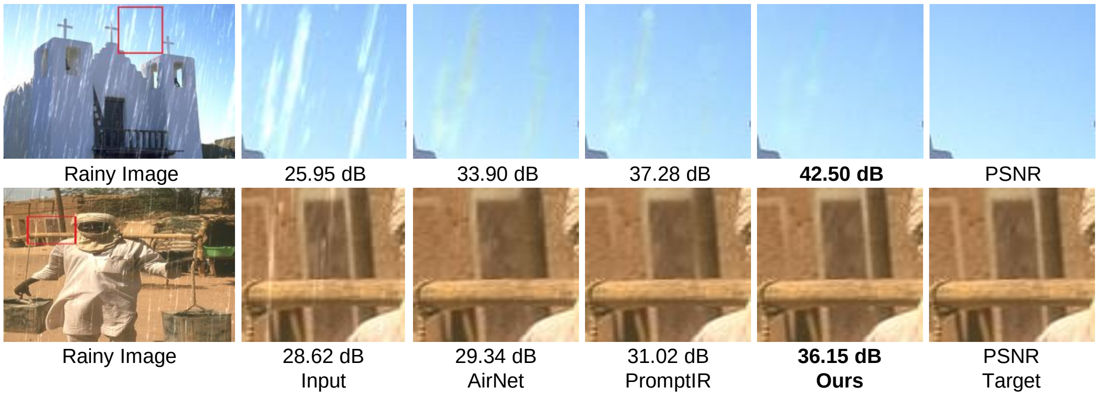

Rainy Image
25.95 dB
33.90 dB
37.28 dB
42.50 dB
PSNR
Rainy Image
28.62 dB
Input
29.34 dB
AirNet
31.02 dB
PromptIR
36.15 dB
Ours
PSNR
Target

Fig. 13. Visual comparisons for image deraining on the Rain100L [75] dataset under the three-task setting.

experimental results under three-task and five-task settings. Finally, we conduct the ablation experiments.

A. Datasets

We choose image denoising, deraining, and dehazing tasks for the three-task all-in-one setting, closely following [22]. Specifically, for image denoising, we employ BSD400 [76] and WED [77] for training. The noise images are generated by adding Gaussian noise with different noise levels of 15, 25, and 50. The trained models are evaluated on BSD68 [74] and Urban100 [78]. For image deraining, the Rain100L [75] dataset is used for training and testing. For image dehazing, RESIDE-β [3] and SOTS-Outdoor [79] are used for training and testing, respectively. The model is trained on a combination of the abovementioned training datasets for the three-task setting. In addition, the GoPro [80] dataset and the LoL-V1 [81] dataset are employed for motion deblurring and low-light image enhancement, respectively, in the five-task setting [23].

B. Implementation Details

The architectural details of our all-in-one model are illustrated in Figure 11. Our OKM is inserted into the bottleneck position of a Transformer [1] network. The details and number of Transformer blocks are identical to [1]. The model is trained using the Adam optimizer with an initial learning rate of $2e^{-4}$ and an L_1 loss function. The patch size is set to 128×128, and random vertical and horizontal flips are adopted for data augmentation. The models are trained for 120 and 150 epochs for the three-task and five-task settings, respectively. The batch

This article has been accepted for publication in IEEE Transactions on Circuits and Systems for Video Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2024.3429557

TABLE X
COMPARISONS UNDER THE THREE-TASK ALL-IN-ONE SETTING.

	Denoising on BSD68 [74]						Deraining on Rain100L [75]		Dehazing on SOTS [3]		Average	
	$\sigma = 15$		$\sigma = 25$		$\sigma = 50$							
Methods	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
BRDNet [82]	32.26	0.898	29.76	0.836	26.34	0.693	27.42	0.895	23.23	0.895	27.80	0.843
LPNet [83]	26.47	0.778	24.77	0.748	21.26	0.552	24.88	0.784	20.84	0.828	23.64	0.738
FDGAN [84]	30.25	0.910	28.81	0.868	26.43	0.776	29.89	0.933	24.71	0.929	28.02	0.883
MPRNet [11]	33.54	0.927	30.89	0.880	27.56	0.779	33.57	0.954	25.28	0.955	30.17	0.899
DL [39]	33.05	0.914	30.41	0.861	26.90	0.740	32.62	0.931	26.92	0.931	29.98	0.876
Restormer [1]	33.86	0.933	31.20	0.888	27.90	0.794	35.56	0.969	29.92	0.970	31.69	0.911
AirNet [41]	33.92	0.933	31.26	0.888	28.00	0.797	34.90	0.968	27.94	0.962	31.20	0.910
PromptIR [22]	33.98	0.933	31.31	0.888	28.06	**0.799**	36.37	0.972	**30.58**	0.974	32.06	0.913
Ours	**34.02**	**0.934**	**31.35**	**0.889**	**28.08**	**0.799**	**38.15**	**0.980**	29.66	**0.975**	**32.25**	**0.915**

TABLE XI
IMAGE DERAINING RESULTS ON THE RAIN100L [75] DATASET UNDER THE SINGLE-TASK SETTING.

Methods	DIDMDN [85]	UMR [86]	SIRR [87]	MSPFN [88]	LPNet [83]	AirNet [41]	Restormer [1]	PromptIR [22]	**Ours**
PSNR↑	23.79	32.39	32.37	33.50	33.61	34.90	36.74	37.04	**39.05**
SSIM↑	0.773	0.921	0.926	0.948	0.958	0.977	0.978	0.979	**0.985**

TABLE XII
IMAGE DENOISING RESULTS ON THE URBAN100 [78] AND BSD68 [74] DATASETS UNDER THE SINGLE-TASK SETTING.

	Urban100 [78]						BSD68 [74]						Average	
	$\sigma = 15$		$\sigma = 25$		$\sigma = 50$		$\sigma = 15$		$\sigma = 25$		$\sigma = 50$			
Methods	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
CBM3D [89]	33.93	0.941	31.36	0.909	27.93	0.840	33.50	0.922	30.69	0.868	27.36	0.763	30.80	0.874
DnCNN [90]	32.98	0.931	30.81	0.902	27.59	0.833	33.89	0.930	31.23	0.883	27.92	0.789	30.74	0.878
IRCNN [91]	27.59	0.833	31.20	0.909	27.70	0.840	33.87	0.929	31.18	0.882	27.88	0.790	29.90	0.864
FFDNet [92]	33.83	0.942	31.40	0.912	28.05	0.848	33.87	0.929	31.21	0.882	27.96	0.789	31.05	0.884
BRDNet [82]	34.42	0.946	31.99	0.919	28.56	0.858	34.10	0.929	31.43	0.885	28.16	0.794	31.44	0.889
AirNet [41]	34.40	0.949	32.10	0.924	28.88	0.871	34.14	0.936	31.48	0.893	28.23	0.806	31.54	0.897
PromptIR [22]	34.77	0.952	32.49	0.929	29.39	0.881	**34.34**	**0.938**	**31.71**	**0.897**	**28.49**	**0.813**	31.87	0.902
Ours	**34.92**	**0.953**	**32.69**	**0.931**	**29.65**	**0.884**	**34.34**	**0.938**	**31.71**	**0.897**	28.48	0.812	**31.97**	**0.903**

TABLE XIII
COMPUTATION OVERHEAD COMPARISONS FOR ALL-IN-ONE ALGORITHMS.

Method	AirNet [41]	PromptIR [22]	Ours
Params/M	8.93	35.59	31.24
FLOPs/G	311	158.4	143.21

sizes are set to 32 and 8, respectively, for the all-in-one and single-task settings.

C. Three-Task Setting Results

The results for the three-task setting are presented in Table X. Our method outperforms the recent state-of-the-art PromptIR [22] on most metrics while using 12% fewer parameters and 10% lower complexity (see Table XIII). Specifically, our model exhibits an improvement of 0.19 dB PSNR when averaged across all degradations and levels. Notably, on the deraining dataset of Rain100L [75], our method produces a significant gain of 1.78 dB PSNR. The visual comparisons are presented in Figure 12 and Figure 13. As illustrated, our model is more effective than competing methods in restoring image details and removing degradations.

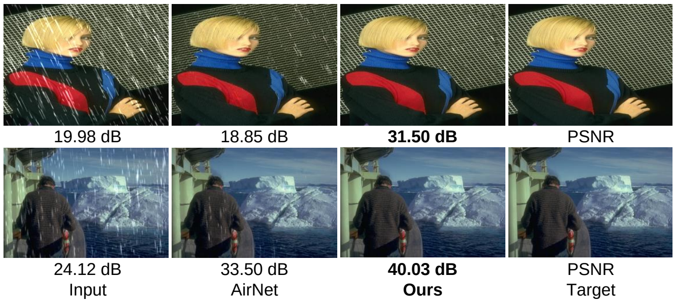


Fig. 14. Visual comparisons for image deraining on the Rain100L [75] dataset under the single-task setting.

Furthermore, following [22], [41], we conduct experiments by training distinct models for each task. The deraining results are presented in Table XI. Our model achieves a notable performance enhancement of 2.01 dB PSNR compared to the all-in-one PromptIR [22] method. Moreover, Table XII shows that our model outperforms PromptIR [22] by 0.1 dB PSNR when averaging across two denoising datasets. In particular, on Urban100 [78], which contains images of higher resolution and quality than the BSD68 [74] dataset, the performance gain

This article has been accepted for publication in IEEE Transactions on Circuits and Systems for Video Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2024.3429557

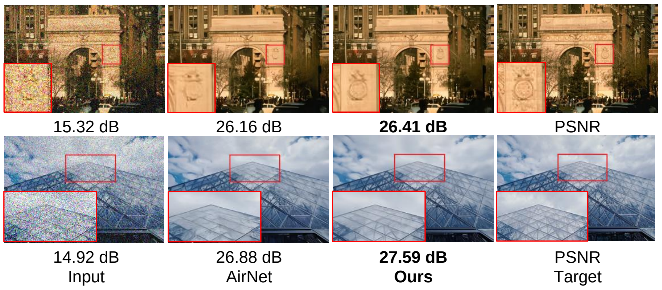


Fig. 15. Visual comparisons for image denoising on the BSD68 [74] dataset with $\sigma = 50$ under the single-task setting.

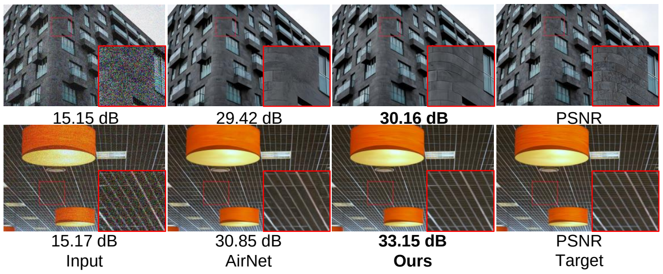


Fig. 16. Visual comparisons for image denoising on the Urban100 [78] dataset with $\sigma = 50$ under the single-task setting.

TABLE XIV

QUANTITATIVE COMPARISONS UNDER THE FIVE-TASK ALL-IN-ONE SETTING. THE FIRST AND SECOND SETS ARE GENERAL AND ALL-IN-ONE IMAGE RESTORATION ALGORITHMS, RESPECTIVELY.

Methods	Dehazing on SOTS		Deraining on Rain100L		Denoising on BSD68		Deblurring on GoPro		Low-Light on LOL		Average	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
NAFNet [93]	25.23	0.939	35.56	0.967	31.02	0.883	26.53	0.808	20.49	0.809	27.76	0.881
HINet [94]	24.74	0.937	35.67	0.969	31.00	0.881	26.12	0.788	19.47	0.800	27.40	0.875
MPRNet [11]	24.27	0.937	**38.16**	**0.981**	31.35	0.889	26.87	0.823	20.84	0.824	28.27	0.890
DGUNet [95]	24.78	0.940	36.62	0.971	31.10	0.883	27.25	0.837	21.87	0.823	28.32	0.891
MIRNetV2 [96]	24.03	0.927	33.89	0.954	30.97	0.881	26.30	0.799	21.52	0.815	27.34	0.875
SwinIR [32]	21.50	0.891	30.78	0.923	30.59	0.868	24.52	0.773	17.81	0.723	25.04	0.835
Restormer [1]	24.09	0.927	34.81	0.962	31.49	0.884	27.22	0.829	20.41	0.806	27.60	0.881
DL [39]	20.54	0.826	21.96	0.762	23.09	0.745	19.86	0.672	19.83	0.712	21.05	0.743
Transweather [40]	21.32	0.885	29.43	0.905	29.00	0.841	25.12	0.757	21.21	0.792	25.22	0.836
TAPE [97]	22.16	0.861	29.67	0.904	30.18	0.855	24.47	0.763	18.97	0.621	25.09	0.801
AirNet [41]	21.04	0.884	32.98	0.951	30.91	0.882	24.35	0.781	18.18	0.735	25.49	0.846
IDR [23]	25.24	0.943	35.63	0.965	**31.60**	0.887	27.87	0.846	21.34	0.826	28.34	0.893
Ours	**30.46**	**0.979**	38.03	**0.981**	31.35	**0.890**	**28.14**	**0.858**	**23.12**	**0.845**	**30.22**	**0.911**

TABLE XV

IMAGE DENOISING COMPARISONS REGARDING PSNR UNDER DIFFERENT NOISE LEVELS (15,25,50). THE RESULTS ARE GENERATED BY APPLYING THE PRE-TRAINED MODEL UNDER THE FIVE-TASK SETTING TO TWO DENOISING DATASETS. '-' MEANS THAT THE DATA IS NOT AVAILABLE.

Methods	Urban100 [78]			Kodak24 [98]			Average	Memory
	15	25	50	15	25	50		
DL [39]	21.10	21.28	20.42	22.63	22.66	21.95	21.67	-
Transweather [40]	29.64	27.97	26.08	31.67	29.64	26.74	28.62	-
TAPE [97]	32.19	29.65	25.87	33.24	30.70	27.19	29.81	-
AirNet [41]	33.16	30.83	27.45	34.14	31.74	28.59	30.99	2.35G
IDR [23]	33.82	31.29	**28.07**	34.78	**32.42**	29.13	31.58	-
Restormer [1]	33.72	31.26	28.03	34.78	32.37	29.08	31.54	9.39G
Ours	**34.01**	**31.35**	**28.07**	**34.88**	32.38	**29.21**	**31.65**	8.45G

can be as large as 0.26 dB for $\sigma = 50$. Figure 14, Figure 15, and Figure 16 illustrate that our model generates high-fidelity results by removing different types of blurs.

D. Five-Task Setting Results

The results for the five-task setting are presented in Table XIV. Our method demonstrates the most favorable performance across the majority of metrics. On average, our model exhibits a 1.88 dB PSNR advantage over the second-best method. For image dehazing, our method achieves a notable 5.22 dB PSNR improvement over the IDR [23] algorithm. Following [23], we evaluate the generalization ability of our model by directly applying the pre-trained model under the five-task setting to Urban100 [78] and Kodak24 [98]. Table XV shows that our model outperforms IDR [23] by 0.07 dB PSNR in terms of average score. Moreover, our method provides an average performance improvement of 0.11 dB over Restormer [1] while utilizing a lower memory footprint.

E. Discussion

The experimental results confirm the efficacy of our OKM in different all-in-one settings. The reasons for this effectiveness are twofold. Firstly, our omni-kernel modulation enables the model to capture perceptive fields of different scales, which are essential for addressing degradations of different sizes within a single task. Additionally, covering a wide range of receptive fields is beneficial for addressing various types of degradations using a unified model, as different types of degradations impact images in different frequencies. Specifically, noise primarily causes high-frequency blurring that corresponds to local receptive fields, whereas low-light image enhancement aims to modulate global illumination, which requires large-scale perceptive fields. Secondly, the attention weights in the global branch are learned from the inputs, enhancing the adaptability of the model to deal with different degradations.

F. Ablation Studies

We perform ablation studies to investigate the impact of different combinations of image restoration tasks on the final

This article has been accepted for publication in IEEE Transactions on Circuits and Systems for Video Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2024.3429557

TABLE XVI
ABLATION STUDIES FOR THE COMBINATIONS OF DIFFERENT DEGRADATION TYPES UNDER THE THREE-TASK SETTING. SCORES ARE PRESENTED AS PSNR/SSIM.

Degradation			Denoising on BSD68			Deraining	Dehazing
Noise	Rain	Haze	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$	Rain100L	SOTS
✓			**34.34/0.938**	**31.71/0.897**	**28.48/0.812**	-	-
	✓		-	-	-	**39.05/0.985**	-
		✓	-	-	-	-	**29.84**/0.974
✓	✓		34.30/0.938	31.66/0.896	28.42/0.810	38.54/0.984	-
✓		✓	34.02/0.934	31.36/0.890	28.09/0.799	-	29.52/0.975
	✓	✓	-	-	-	37.47/0.979	29.39/0.975
✓	✓	✓	34.02/0.934	31.35/0.889	28.08/0.799	38.15/0.980	29.66/**0.975**

performance. As shown in Table XVI, the best results are mostly obtained under the single-task setting. It is worth mentioning that when dehazing is combined with deraining or denoising, dehazing has a more detrimental impact on the combined task than the remaining one. This is possibly because dehazing aims to restore low-frequency signals, while the other two tasks target high frequencies. The model cannot avoid interference between different tasks.

VI. CONCLUSION

This paper proposes an omni-kernel modulation module that can capture global-to-local receptive fields for universal image restoration. The module has three branches, namely local, large, and global branches, to model local, large, and global features, respectively. The large branch is designed by exploring the unusually large regular and strip-based depth-wise convolutions for image restoration. The novel global branch utilizes dual-domain channel attention and frequency-based spatial attention for modulating global representations. Moreover, the lightweight local branch introduces locality to the model. With our designs, the models achieve state-of-the-art performance on 15 different benchmark datasets for general image restoration tasks and perform favorably against state-of-the-art algorithms under two all-in-one settings.

REFERENCES

[1] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5728–5739, 2022.

[2] Yuning Cui and Alois Knoll. Enhancing local–global representation learning for image restoration. *IEEE Transactions on Industrial Informatics*, 20(4):6522–6530, 2024.

[3] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang. Benchmarking single image dehazing and beyond. *IEEE Transactions on Image Processing*, 2018.

[4] Abdullah Abuolaim and Michael S Brown. Defocus deblurring using dual-pixel data. In *Proceedings of the European Conference on Computer Vision*, pages 111–126, 2020.

[5] Dongdong Chen, Mingming He, Qingnan Fan, Jing Liao, Liheng Zhang, Dongdong Hou, Lu Yuan, and Gang Hua. Gated context aggregation network for image dehazing and deraining. In *2019 IEEE Winter Conference on Applications of Computer Vision*, pages 1375–1383, 2019.

[6] Yun-Fu Liu, Da-Wei Jaw, Shih-Chia Huang, and Jenq-Neng Hwang. Desnownet: Context-aware deep network for snow removal. *IEEE Transactions on Image Processing*, 27(6):3064–3073, 2018.

[7] Yuning Cui and Alois Knoll. Exploring the potential of channel interactions for image restoration. *Knowledge-Based Systems*, 282:111156, 2023.

[8] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4641–4650, October 2021.

[9] Lingyan Ruan, Bin Chen, Jizhou Li, and Miuling Lam. Learning to deblur using light field generated and real defocus images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 16304–16313, June 2022.

[10] Wenbin Zou, Mingchao Jiang, Yunchen Zhang, Liang Chen, Zhiyong Lu, and Yi Wu. Sdwnet: A straight dilated network with wavelet transformation for image deblurring. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1895–1904, 2021.

[11] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 14821–14831, 2021.

[12] Yuning Cui, Wenqi Ren, Sining Yang, Xiaochun Cao, and Alois Knoll. Irnext: Rethinking convolutional network design for image restoration. In *International Conference on Machine Learning*, 2023.

[13] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 17683–17693, 2022.

[14] H. Chen, Y. Wang, T. Guo, C. Xu, Y. Deng, Z. Liu, S. Ma, C. Xu, C. Xu, and W. Gao. Pre-trained image processing transformer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021.

[15] Xiaohan Ding, Xiangyu Zhang, Jungong Han, and Guiguang Ding. Scaling up your kernels to 31x31: Revisiting large kernel design in cnns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11963–11975, 2022.

[16] Shiwei Liu, Tianlong Chen, Xiaohan Chen, Xuxi Chen, Qiao Xiao, Boqian Wu, Tommi Kärkkainen, Mykola Pechenizkiy, Decebal C Mocanu, and Zhangyang Wang. More convnets in the 2020s: Scaling up kernels beyond 51x51 using sparsity. In *International Conference on Learning Representations*, 2023.

[17] Pinjun Luo, Guoqiang Xiao, Xinbo Gao, and Song Wu. Lkd-net: Large kernel convolution network for single image dehazing. *arXiv preprint arXiv:2209.01788*, 2022.

[18] Lingyan Ruan, Mojtaba Bemana, Hans-peter Seidel, Karol Myszkowski, and Bin Chen. Revisiting image deblurring with an efficient convnet. *arXiv preprint arXiv:2302.02234*, 2023.

[19] Yan Wang, Yusen Li, Gang Wang, and Xiaoguang Liu. Multi-scale attention network for single image super-resolution. *arXiv preprint arXiv:2209.14145*, 2022.

[20] Man Zhou, Jie Huang, Chun-Le Guo, and Chongyi Li. Fourmer: An efficient global modeling paradigm for image restoration. In *International Conference on Machine Learning*, pages 42589–42601, 2023.

[21] Wei-Ting Chen, Hao-Yu Fang, Cheng-Lin Hsieh, Cheng-Che Tsai, I Chen, Jian-Jiun Ding, Sy-Yen Kuo, et al. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4196–4205, 2021.

[22] Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. *Advances in Neural Information Processing Systems*, 2023.

[23] Jinghao Zhang, Jie Huang, Mingde Yao, Zizheng Yang, Hu Yu, Man Zhou, and Feng Zhao. Ingredient-oriented multi-degradation learning for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2023.

[24] Yuning Cui, Wenqi Ren, and Alois Knoll. Omni-kernel network for image restoration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1426–1434, 2024.

[25] Wending Yan, Robby T Tan, and Dengxin Dai. Nighttime defogging using high-low frequency decomposition and grayscale-color networks. In *European Conference on Computer Vision*, pages 473–488, 2020.

[26] Yuning Cui and Alois Knoll. Dual-domain strip attention for image restoration. *Neural Networks*, 171:429–439, 2024.

[27] Yuning Cui, Yi Tao, Luoxi Jing, and Alois Knoll. Strip attention for image restoration. In *International Joint Conference on Artificial Intelligence*, 2023.

This article has been accepted for publication in IEEE Transactions on Circuits and Systems for Video Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2024.3429557

[28] Xiaoqin Zhang, Jinxin Wang, Tao Wang, and Runhua Jiang. Hierarchical feature fusion with mixed convolution attention for single image dehazing. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(2):510–522, 2021.

[29] Dong Zhao, Long Xu, Lin Ma, Jia Li, and Yihua Yan. Pyramid global context network for image dehazing. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(8):3037–3050, 2020.

[30] Yuning Cui, Yi Tao, Zhenshan Bing, Wenqi Ren, Xinwei Gao, Xiaochun Cao, Kai Huang, and Alois Knoll. Selective frequency network for image restoration. In *International Conference on Learning Representations*, 2023.

[31] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

[32] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1833–1844, 2021.

[33] Yawei Li, Yuchen Fan, Xiaoyu Xiang, Denis Demandolx, Rakesh Ranjan, Radu Timofte, and Luc Van Gool. Efficient and explicit modelling of image hierarchies for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 18278–18289, 2023.

[34] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32:1927–1941, 2023.

[35] Chun-Le Guo, Qixin Yan, Saeed Anwar, Runmin Cong, Wenqi Ren, and Chongyi Li. Image dehazing transformer with transmission-aware 3d position embedding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5812–5820, 2022.

[36] Xintian Mao, Yiming Liu, Wei Shen, Qingli Li, and Yan Wang. Deep residual fourier transformation for single image deblurring. *arXiv preprint arXiv:2111.11745*, 2021.

[37] Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Focal network for image restoration. In *Proceedings of the IEEE International Conference on Computer Vision*, 2023.

[38] Zhanchen Zhu, Daokun Zhang, Zhikang Wang, Siyuan Feng, and Peibo Duan. Spectral dual-channel encoding for image dehazing. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.

[39] Qingnan Fan, Dongdong Chen, Lu Yuan, Gang Hua, Nenghai Yu, and Baoquan Chen. A general decoupled learning framework for parameterized image operators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[40] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M. Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2353–2363, June 2022.

[41] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022.

[42] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016.

[43] Xiaoyi Dong, Jianmin Bao, Dongdong Chen, Weiming Zhang, Nenghai Yu, Lu Yuan, Dong Chen, and Baining Guo. Cswin transformer: A general vision transformer backbone with cross-shaped windows. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 12124–12134, 2022.

[44] Codruta O Ancuti, Cosmin Ancuti, Mateu Sbert, and Radu Timofte. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In *IEEE International Conference on Image Processing*, pages 1014–1018, 2019.

[45] Codruta O. Ancuti, Cosmin Ancuti, and Radu Timofte. Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, June 2020.

[46] Codruta O Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 754–762, 2018.

[47] Cosmin Ancuti, Codruta O Ancuti, Radu Timofte, and Christophe De Vleeschouwer. I-haze: A dehazing benchmark with real hazy and haze-free indoor images. In *Advanced Concepts for Intelligent Vision Systems: 19th International Conference, ACIVS 2018, Poitiers, France, September 24–27, 2018, Proceedings 19*, pages 620–631, 2018.

[48] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2010.

[49] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision*, Oct 2017.

[50] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7314–7323, 2019.

[51] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2020.

[52] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11908–11915, 2020.

[53] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10551–10560, June 2021.

[54] Tian Ye, Yunchen Zhang, Mingchao Jiang, Liang Chen, Yun Liu, Sixiang Chen, and Erkang Chen. Perceiving and modeling density for image dehazing. In *Proceedings of the European Conference on Computer Vision*, pages 130–145, 2022.

[55] Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. Maxim: Multi-axis mlp for image processing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5769–5780, June 2022.

[56] Jing Zhang, Yang Cao, Zheng-Jun Zha, and Dacheng Tao. Nighttime dehazing with a synthetic benchmark. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 2355–2363, 2020.

[57] Binghui Huang, Li Zhi, Chao Yang, Fuchun Sun, and Yixu Song. Single satellite optical imagery dehazing using sar image prior based on conditional generative adversarial networks. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, pages 1806–1813, 2020.

[58] Wei-Ting Chen, Hao-Yu Fang, Jian-Jiun Ding, Cheng-Che Tsai, and Sy-Yen Kuo. Jstasr: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal. In *Proceedings of the European Conference on Computer Vision*, pages 754–770. Springer, 2020.

[59] Yeying Jin, Beibei Lin, Wending Yan, Yuan Yuan, Wei Ye, and Robby T Tan. Enhancing visibility in nighttime haze images using guided apsf and gradient adaptive convolution. In *Proceedings of the ACM International Conference on Multimedia*, pages 2446–2457, 2023.

[60] Yu Li, Robby T Tan, and Michael S Brown. Nighttime haze removal with glow and multiple light colors. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 226–234, 2015.

[61] Jing Zhang, Yang Cao, Shuai Fang, Yu Kang, and Chang Wen Chen. Fast haze removal for nighttime image using maximum reflectance prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7418–7426, 2017.

[62] Cosmin Ancuti, Codruta O Ancuti, Christophe De Vleeschouwer, and Alan C Bovik. Night-time dehazing by fusion. In *IEEE International Conference on Image Processing*, pages 2256–2260, 2016.

[63] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2223–2232, 2017.

[64] Tao Wang, Guangpin Tao, Wanglong Lu, Kaihao Zhang, Wenhan Luo, Xiaoqin Zhang, and Tong Lu. Restoring vision in hazy weather with hierarchical contrastive learning. *Pattern Recognition*, 145:109956, 2024.

[65] Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Image restoration via frequency selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(2):1093–1108, 2024.

[66] Xiang Chen, Yufeng Li, Longgang Dai, and Caihua Kong. Hybrid high-resolution learning for single remote sensing satellite image dehazing. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2021.

[67] Yufeng Li and Xiang Chen. A coarse-to-fine two-stage attentive network for haze removal of remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 18(10):1751–1755, 2020.

This article has been accepted for publication in IEEE Transactions on Circuits and Systems for Video Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2024.3429557

[68] Yu Zheng, Jiahui Zhan, Shengfeng He, Junyu Dong, and Yong Du. Curricular contrastive regularization for physics-aware single image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5785–5794, 2023.

[69] Kaichen Chi, Yuan Yuan, and Qi Wang. Trinity-net: Gradient-guided swin transformer-based remote sensing image dehazing and beyond. *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[70] Ruoteng Li, Robby T. Tan, and Loong-Fah Cheong. All in one bad weather removal using architectural search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2020.

[71] Hyeongseok Son, Junyong Lee, Sunghyun Cho, and Seungyong Lee. Single image defocus deblurring using kernel-sharing parallel atrous convolutions. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2642–2650, 2021.

[72] Junyong Lee, Hyeongseok Son, Jaesung Rim, Sunghyun Cho, and Seungyong Lee. Iterative filter adaptive network for single image defocus deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2034–2042, 2021.

[73] Abdullah Abuolaim, Mahmoud Afifi, and Michael S Brown. Improving single-image defocus deblurring: How dual-pixel images help through multi-task learning. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, pages 1231–1239, 2022.

[74] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the IEEE International Conference on Computer Vision*, 2001.

[75] Wenhan Yang, Robby T Tan, Jiashi Feng, Zongming Guo, Shuicheng Yan, and Jiaying Liu. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[76] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010.

[77] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing*, 2016.

[78] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[79] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *Proceedings of the European Conference on Computer Vision*, September 2018.

[80] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, July 2017.

[81] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv:1808.04560*, 2018.

[82] Chunwei Tian, Yong Xu, and Wangmeng Zuo. Image denoising using deep cnn with batch renormalization. *Neural Networks*, 2020.

[83] Hongyun Gao, Xin Tao, Xiaoyong Shen, and Jiaya Jia. Dynamic scene deblurring with parameter selective sharing and nested skip connections. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[84] Yu Dong, Yihao Liu, He Zhang, Shifeng Chen, and Yu Qiao. Fd-gan: Generative adversarial networks with fusion-discriminator for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020.

[85] He Zhang and Vishal M. Patel. Density-aware single image de-raining using a multi-stream dense network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2018.

[86] Rajeev Yasarla and Vishal M Patel. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[87] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. Semi-supervised transfer learning for image rain removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[88] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2020.

[89] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space. In *IEEE International Conference on Image Processing*, 2007.

[90] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 2017.

[91] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[92] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018.

[93] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *Proceedings of the European Conference on Computer Vision*, 2022.

[94] Liangyu Chen, Xin Lu, Jie Zhang, Xiaojie Chu, and Chengpeng Chen. Hinet: Half instance normalization network for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 182–192, June 2021.

[95] Chong Mou, Qian Wang, and Jian Zhang. Deep generalized unfolding networks for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022.

[96] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for fast image restoration and enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[97] Lin Liu, Lingxi Xie, Xiaopeng Zhang, Shanxin Yuan, Xiangyu Chen, Wengang Zhou, Houqiang Li, and Qi Tian. Tape: Task-agnostic prior embedding for image restoration. In *European Conference on Computer Vision*, pages 447–464, 2022.

[98] Rich Franzen. Kodak lossless true color image suite. http://r0k.us/graphics/kodak/, 1999. Online accessed 24 Oct 2021.

BIOGRAPHY SECTION

Yuning Cui (Student Member, IEEE) received the B.Eng. degree from Central South University, China, in 2016, and the M.Eng. degree from National University of Defense Technology, China, in 2018. He is currently pursuing a Ph.D. degree at the Chair of Robotics, Artificial Intelligence and Real-time Systems within the School of Computation, Information and Technology at the Technical University of Munich. His research interest lies in image restoration.

Wenqi Ren (Member, IEEE) received the Ph.D. degree from Tianjin University, China, in 2017. From 2015 to 2016, he worked with Prof. Ming-Husan Yang as a Joint-Training Ph.D. Student with the Electrical Engineering and Computer Science Department, University of California at Merced. He is currently a Professor at the School of Cyber Science and Technology, Shenzhen Campus, Sun Yat-sen University, Shenzhen, China. His research interests include image processing and related high-level vision problems.

Alois Knoll (Fellow, IEEE) received his diploma (M.Sc.) degree in Electrical/Communications Engineering from the University of Stuttgart, Germany, in 1985 and his Ph.D. (*summa cum laude*) in Computer Science from Technical University of Berlin, Germany, in 1988. Since 2001, he has been a professor at the Department of Informatics, Technical University of Munich, Germany. His research interests include cognitive, medical and sensor-based robotics, multi-agent systems, adaptive systems, multimedia information retrieval, model-driven development of embedded systems with applications to automotive software and electric transportation, as well as simulation systems for robotics and traffic.