Learning Dual-Domain Multi-Scale Representations for Single Image Deraining

Shun Zou[1,7*], Yi Zou[2*], Mingya Zhang[3], Shipeng Luo[4], Guangwei Gao[5,7†], Guojun Qi[6]
[1]Nanjing Agricultural University [2]Xiangtan University [3]Nanjing University [4]Northeast Forestry University
[5]Nanjing University of Posts and Telecommunications [6]Westlake University [7]Soochow University
Project page: https://zs1314.github.io/DMSR

2025 IEEE International Conference on Multimedia and Expo (ICME) | 979-8-3315-9495-4/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICME59968.2025.11210243

***Abstract*—Existing image deraining methods typically rely on single-input, single-output, and single-scale architectures, which overlook the joint multi-scale information between external and internal features. Furthermore, single-domain representations are often too restrictive, limiting their ability to handle the complexities of real-world rain scenarios. To address these challenges, we propose a novel Dual-Domain Multi-Scale Representation Network (DMSR). The key idea is to exploit joint multi-scale representations from both external and internal domains in parallel while leveraging the strengths of both spatial and frequency domains to capture more comprehensive properties. Specifically, our method consists of two main components: the Multi-Scale Progressive Spatial Refinement Module (MPSRM) and the Frequency Domain Scale Mixer (FDSM). The MPSRM enables the interaction and coupling of multi-scale expert information within the internal domain using a hierarchical modulation and fusion strategy. The FDSM extracts multi-scale local information in the spatial domain, while also modeling global dependencies in the frequency domain. Extensive experiments show that our model achieves state-of-the-art performance across six benchmark datasets.**

***Index Terms*—Single Image Deraining, Image Restoration, Dual-Domain Paradigm, Multi-Scale Representation**

I. Introduction

Single-image deraining (SID) is a vital task in low-level vision, focused on restoring clear background images from rainy inputs by eliminating or reducing undesirable degradations caused by rain artifacts [1]. The severe interference of rainy images with the performance of downstream tasks has sparked significant research interest in the SID task.

Over the years, numerous methods have been proposed to address the image deraining problem. Early prior-based traditional approaches [2] attempted to remove rain artifacts by analyzing the statistical characteristics of rain streaks and the background. However, in real-world scenarios, rain streaks and droplets are often dense, complex, and diverse, causing these methods to fail in such cases. Recently, many CNN-based methods [3]–[7] have been applied to image deraining, bringing transformative advancements to the field. Compared to traditional algorithms, these methods have significantly improved deraining results. However, convolution, as the core element of CNNs, exhibits spatial invariance and a limited receptive field, making it inadequate for effectively modeling the non-local structures and spatial variations in clear images [8]. To address these limitations, some methods [9]–[13] have adopted Transformers for image deraining, achieving impressive results. Transformers effectively capture global dependencies and model non-local information, enabling high-quality image reconstruction.

However, existing mainstream supervised paradigms still face two major challenges: (1) the inability to explore complementary implicit information across different scales. Most current deep deraining networks rely on single-input single-output architectures. However, spatially varying rain streaks often exhibit diverse scale attributes, which not only overlooks potential explicit information from multiple image scales but also fails to explore complementary implicit information across scales. Generally, multi-scale visual information flow involves representations from external interactions (multi-input multi-output exchanges with the network's external environment) and internal interactions (scale information exchange within internal components of the network). Some researchers have introduced coarse-to-fine mechanisms [1], [14] or multi-patch strategies [15], [16] to leverage multi-scale external rain features. While these methods have achieved impressive performance, they struggle to handle complex and random rain streaks because they rarely consider collaborative multi-scale feature representations from both external and internal domains, neglecting implicit relationships across scales. (2) Rain streaks and droplets consistently exhibit irregular overlaps and highly dynamic geometric patterns, which place higher and more diverse demands on feature representations. Moreover, in real-world rainy images, significant aliasing effects exist between rain residues and the background. Eliminating rain disturbances by inferring pixel residue values inevitably compromises contextual and structural information. When features are derived solely from a single domain or dimension, it becomes challenging to remove all interferences [17]. Therefore, more robust and diverse multi-domain feature representations are critical for achieving high-quality image reconstruction.

To address the abovementioned challenges, we introduce DMSR, a novel dual-domain multi-scale architecture model. It consists of multiple Dual-Domain Scale-Aware Modules (DDSAM), which include two key components: the Multi-Scale Progressive Spatial Refinement Module (MPSRM) and the Frequency Domain Scale Mixer (FDSM). To tackle the first challenge, we propose a collaborative multi-scale paradigm for

*Equal Contribution, †Corresponding author. Email: zs@stu.njau.edu.cn, csgwgao@njupt.edu.cn. This work was supported in part by the Open Fund Project of Provincial Key Laboratory for Computer Information Processing Technology (Soochow University) under Grant KJS2274.

both external and internal domains. We use a coarse-to-fine pyramid input-output flow externally across the entire network while applying various multi-scale architectures within the internal components of DMSR (i.e., MPSRM and FDSM). This facilitates the flow of information within and across scales, enabling the interaction and fusion of complementary implicit information between different scales. Additionally, to address the second challenge, we use MPSRM to extract spatial domain features and introduce Spatial Pixel Guided Attention (SPGA) to aggregate information for each pixel, allowing each pixel to perceive information from an extended region implicitly. Simultaneously, FDSM decouples the features into distinct frequency domain components and modulates them, promoting interactions between different frequencies and refining the spectrum. This adaptive extraction and refinement of rain residuals and background components allows us to fully capture information from different dimensions through the extraction of dual-domain features, facilitating the exchange and complement of expert knowledge. Our contributions are summarized as follows:

- We propose a novel dual-domain multi-scale architecture that rethinks the multi-scale representation paradigm for single-image deraining through a collaborative multi-scale paradigm across internal and external domains, enabling better extraction of rich scale-space features. At the same time, it fully leverages the advantages of the dual-domain paradigm to achieve more diverse, multi-dimensional, and robust high-level feature representations.
- We introduce a Multi-Scale Progressive Spatial Refinement Module for extracting refined spatial pixel features, using gated contextual information to integrate implicit information from the surrounding context and expand the receptive field. Additionally, we develop a Multi-Scale Frequency Domain Mixer that combines global and local characteristics and facilitates information exchange and modulation across different spectral components, thereby generating high-quality deraining results.

II. Method

A. Overall Pipeline

As shown in Fig. 1 (a), DMSR is divided into three scales based on a coarse-to-fine approach. Specifically, given an input rain image, the original image is downsampled by an interpolation operator to 1/2 and 1/4 of the original size, forming a pyramid of rain image inputs, with the coarsest to finest scale inputs referred to as S1, S2, and S3. We first pass S1 through a 3×3 convolutional layer to obtain shallow features $F \in \mathbb{R}^{C \times H \times W}$, where C, H, and W represent the channels, height, and width, respectively. These shallow features are then processed through three DDSAM to obtain multi-scale high-level representation features in both spatial and frequency domains. As shown in Fig. 1 (b), DDSAM consists of multiple residual structures and includes the MPSRM and the FDSM. During this process, the spatial resolution is reduced by half, and the channel number is doubled. Moreover, we incorporate the downsampled rain images S2 and S3 into the main path after passing them through an Embedding Layer and adjusting the channel number with a 3×3 convolution. The dual-domain multi-scale high-level representation features are then passed through another three DDSAMs for gradual restoration into a high-resolution derained image. During this process, features from the encoder and decoder are concatenated to facilitate image restoration. In line with the multi-scale inputs, we also employ multi-scale outputs, generating low-resolution derained images after the first two DDSAMs in the restoration process. Finally, skip connections are applied across the three pyramid input/output image pairs. For simplicity, only the top-level skip connection between the rain image and the derained image is shown in Fig. 1 (a).

B. Multi-Scale Progressive Spatial Refinement Module

The architecture of MPSRM is shown in Fig. 1 (c). To alleviate the knowledge discrepancy across different scales within the same representation range, we perform high-quality restoration in a progressive coarse-to-fine manner within each representation scale, efficiently achieving hierarchical representation. Additionally, MPSRM focuses on modeling spatial context to enhance the spatial pixel representation ability of each feature map. By capturing pixel context relationships, it aids in the precise restoration of background and fine details. Specifically, given the input feature F, global average pooling with different downsampling rates is applied to embed the initial feature F into different scales. For each scale (i.e., each branch), we input the scale features into the Spatial Pixel Guided Attention (SPGA), enhancing the expressiveness of each feature map. The resulting spatially refined features are then integrated into the next branch via addition. This enables MPSRM to progressively reduce intra-scale differences and promote the flow of expert information across different scales. Finally, the progressively fused features from all branches are unified to match the input size of F and summed. Formally, the above process is described as follows:

$$\hat{F}_1 = SPGA(GAP_4(F)), \tag{1}$$

$$\hat{F}_2 = SPGA(GAP_2(F) + \hat{F}_1), \tag{2}$$

$$\hat{F} = f^{3\times3}(F + \hat{F}_1 \uparrow_4 + \hat{F}_2 \uparrow_2), \tag{3}$$

where $GAP_i(\cdot)$ denotes global average pooling with a downsampling rate of i, $\uparrow_i$ represents bilinear upsampling with a scaling factor of i, and $f^{z\times z}(\cdot)$ indicates a convolution with a kernel size of $z \times z$.

C. Spatial Pixel Guided Attention

In traditional spatial self-attention mechanisms [18], spatial context modeling is often achieved by computing spatial feature distribution maps to indicate the importance levels of different regions. However, due to the complexity of real-world rainfall scenarios and the high interweaving of rain streaks with rain-free backgrounds, fine texture details are often disrupted during restoration, resulting in artifacts and noise. To address these issues, we propose a Spatial Pixel Guided Attention (SPGA) mechanism, which shifts from region-based to pixel-level guidance. By gradually generating perception information for

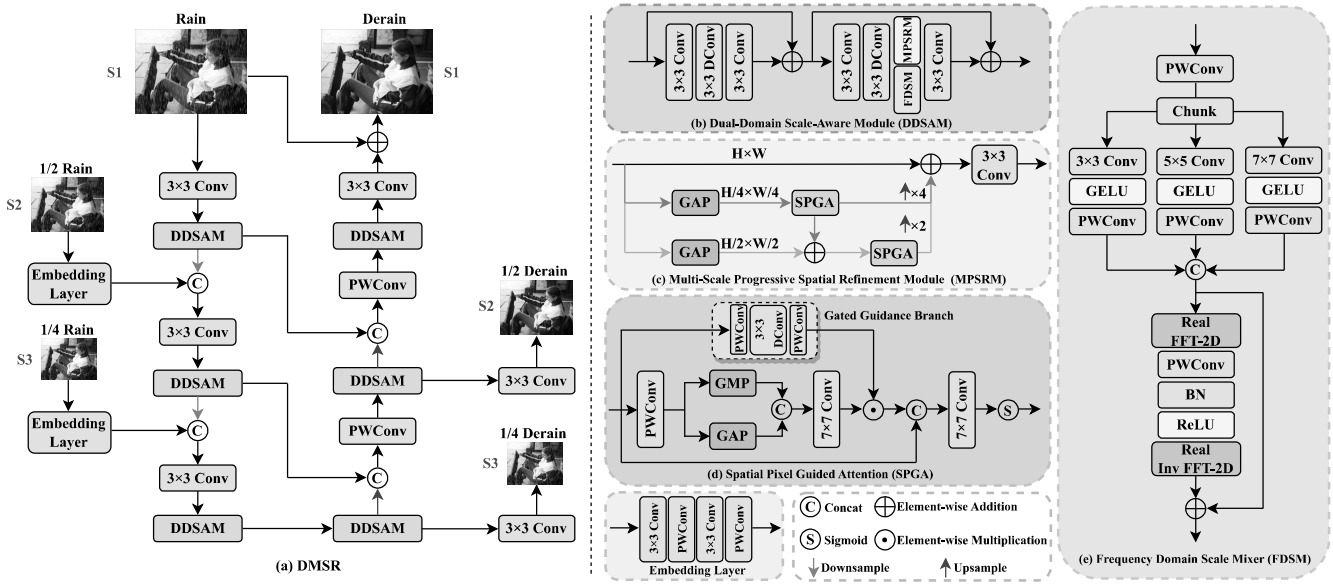

Fig. 1: The overall architecture of the proposed DMSR.

each pixel through gated guidance, SPGA facilitates pixel-wise information interaction, emphasizing the most valuable expert information in feature maps and safeguarding texture details from contamination. The structure of SPGA is illustrated in Fig. 1 (d). Specifically, given the input feature $X \in \mathbb{R}^{C\times H\times W}$, spatial attention is first applied to generate the spatial feature distribution map W_S. Furthermore, we design a gated guidance module to focus features on the most valuable information. The gating mechanism promotes the flow of critical information while preventing excessive redundancy within intermediate layers, thereby mitigating information loss. This process can be mathematically described as:

$$W_S = f^{7\times7}([GAP_2(PW(X)), GMP_2(PW(X))]), \quad (4)$$

$$W_G = PW(DW^{3\times3}(PW(X))), \quad (5)$$

where $PW(\cdot)$ denotes point-wise convolution, $DW^{3\times3}(\cdot)$ represents depth-wise convolution with a kernel size of 3×3, $[\ ,\ ]$ denotes concatenation, and $GMP_i(\cdot)$ represents global max pooling with a downsampling rate of i. Subsequently, we fuse W_S and W_G through an additive operation to obtain the coarse spatial pixel feature map W_M. Then, we use a 7×7 convolution operation to further expand the receptive field, re-weighting the pixel information to enable each pixel to perceive signals from a large region centered around itself, while employing the sigmoid function to introduce non-linearity. Formally, the entire process is expressed as follows:

$$W_M = W_S \odot W_G, \quad (6)$$

$$W = \delta(f^{7\times7}([X, W_M])), \quad (7)$$

where $\delta(\cdot)$ denotes the sigmoid activation function, $\odot$ represents element-wise multiplication.

D. Frequency Domain Scale Mixer

Fourier transform exhibits irreplaceable advantages in single image deraining. First, it can effectively separate image degradation components, as rain streak patterns display prominent and invariant characteristics in the frequency domain, which is crucial for mitigating aliasing effects in rainy images [11], [19]. Second, the transformed frequency components, calculated from all spatial components, serve as global feature filters [20], [21]. The motivation behind the proposed Frequency Domain Scale Mixer (FDSM) is to bridge the spatial and frequency domains, enabling layered modulation and feature extraction of global-local characteristics while facilitating intra-domain multi-scale information interaction and fusion. The structure of FDSM is illustrated in Fig. 1 (e). Specifically, in the spatial-scale mixed domain, we first use PWConv to increase the feature dimensions, which are then divided into three groups to extract multi-scale local mixed features. The mathematical formulation is as follows:

$$X_Z = PW(\vartheta(f^{z\times z}(Chunk(PW(X))))), Z \in [3, 5, 7], \quad (8)$$

$$X_S = [X_3, X_5, X_7]. \quad (9)$$

Here, $\vartheta(\cdot)$ denotes the GELU activation function. In the frequency domain, the spatial-scale mixed features are first transformed using the fast Fourier transform (FFT) to produce real and imaginary components. The combined frequency features are then passed through a PWConv for modulation, emphasizing and interactively merging beneficial frequency components for restoration. After modulation, the features are passed through the inverse fast Fourier transform (IFFT) to convert the frequency domain features back into the spatial domain. The above process is formally expressed as follows:

$$R, I = FFT(X_S), \quad (10)$$

$$\hat{R}, \hat{I} = \phi(BN(PW([\hat{R}, \hat{I}]))), \quad (11)$$

$$X_F = IFFT(\hat{R}, \hat{I}), \quad (12)$$

where $BN(\cdot)$ represents batch normalization, and $\phi(\cdot)$ denotes the ReLU activation function.

III. Experiments

A. Datasets and Metrics

Datasets. We follow the approach of most previous studies to train and validate our model [5], [15], [20], [30], [34].

TABLE I: Comparison of quantitative results on five datasets. Bold and underlined indicate the best and second-best results.

Method	Year	Test100 [22]		Rain100H [23]		Rain100L [23]		Test2800 [24]		Test1200 [25]		Average	
		PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
RESCAN [3]	ECCV2018	21.59	0.726	18.01	0.467	24.15	0.791	24.50	0.765	24.40	0.759	22.53	0.702
PReNet [4]	CVPR2019	23.17	0.752	17.63	0.487	27.76	0.876	27.20	0.825	26.05	0.792	24.36	0.746
SPDNet [26]	ICCV2021	24.25	0.848	25.87	0.809	28.63	0.880	31.05	0.904	30.42	0.893	28.04	0.867
PCNet [27]	TIP2021	23.29	0.762	20.83	0.563	26.64	0.817	27.10	0.818	26.53	0.791	24.88	0.750
MPRNet [15]	CVPR2021	25.66	0.859	28.23	0.850	31.94	0.930	32.14	0.925	31.32	0.901	29.86	0.893
HINet [5]	CVPRW2021	23.21	0.767	20.85	0.598	27.03	0.842	28.36	0.843	27.77	0.821	25.44	0.774
DANet [28]	IJCAI2022	23.96	0.839	23.00	0.791	29.51	0.906	30.32	0.903	29.99	0.888	27.36	0.865
Uformer [29]	CVPR2022	23.87	0.815	22.43	0.700	28.39	0.883	29.71	0.886	28.65	0.856	26.61	0.828
ALformer [30]	ACMMM2022	24.41	0.844	25.10	0.807	29.39	0.903	31.36	0.916	30.40	0.897	28.13	0.874
NAFNet [31]	ECCV2022	25.75	0.845	26.76	0.813	31.27	0.925	31.71	0.918	30.62	0.892	29.22	0.879
MFDNet [32]	TIP2023	25.90	0.870	27.06	0.850	32.76	0.944	31.92	0.925	31.15	0.909	29.76	0.899
HCT-FFN [10]	AAAI2023	24.86	0.847	26.70	0.819	29.94	0.906	31.46	0.915	31.23	0.901	28.84	0.878
DRSformer [9]	CVPR2023	27.86	0.885	28.16	0.864	34.79	0.954	32.80	**0.931**	30.99	0.906	30.92	0.908
ChaIR [6]	KBS2023	28.19	0.879	28.69	0.862	34.52	0.953	**32.85**	**0.931**	31.30	0.903	31.11	0.906
IRNeXT [33]	ICML2023	25.80	0.860	27.22	0.833	31.65	0.931	30.53	0.917	29.02	0.898	28.85	0.888
OKNet [7]	AAAI2024	25.43	0.858	24.01	0.804	31.19	0.928	29.32	0.911	27.56	0.886	27.50	0.877
AST [12]	CVPR2024	26.07	0.859	27.40	0.833	32.03	0.932	31.65	0.921	30.69	0.897	29.57	0.889
SFHformer [11]	ECCV2024	25.67	0.856	27.25	0.832	32.97	0.944	32.27	0.925	**31.50**	0.904	29.94	0.892
Nerd-rain [13]	CVPR2024	27.16	0.869	28.07	0.838	33.72	0.949	32.63	0.927	30.45	0.890	30.41	0.895
FSNet [20]	TPAMI2024	27.95	0.884	28.70	0.860	34.10	0.952	32.68	**0.931**	31.26	**0.910**	30.94	0.908
DMSR (Ours)	–	**28.88**	**0.890**	**29.41**	**0.873**	**35.19**	**0.957**	32.50	**0.931**	31.35	**0.910**	**31.47**	**0.912**

TABLE II: Results of Perceptual Quality Assessment.

Methods	Input	DRSformer [9]	SFHformer [11]	Nerd-rain [13]	FSNet [20]	DMSR
NIQE ↓	5.923	5.814	5.745	5.711	5.667	**5.582**

Specifically, we use 13,712 image pairs collected from multiple datasets for training [22], [23], [35]–[37], and validate on five synthetic datasets (Test100 [22], Rain100H [23], Rain100L [23], Test2800 [24], Test1200 [25]). Additionally, we use real-world rainy images [38] to further validate the effectiveness of the model in real-world scenarios.

Evaluation Metrics. Similar to existing computational methods [1], [20], we use Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) to evaluate the de-rain performance. These metrics are calculated based on the Y channel (luminance) in the YCbCr color space.

B. Implementation Details

During training, we use the Adam optimizer with a patch size of 64×64, a batch size of 12, and a total of 300 epochs. The initial learning rate is set to 2×10^{-4} and follows a cosine annealing scheduler with linear warm-up for the first 3 epochs. For data augmentation, we follow the same strategies as in previous studies [15], [39]. The entire framework is implemented in PyTorch and trained on a single NVIDIA GeForce RTX 4090 GPU. During testing, a sliding window slicing method is applied for image cropping [1].

C. Comparisons with SOTA Methods

We compare DMSR with 20 state-of-the-art image deraining methods: RESCAN [3], PreNet [4], SPDNet [26], PCNet [27], MPRNet [15], HINet [5], DANet [28], Uformer [29], ALformer [30], NAFNet [31], MFDNet [32], HCT-FFN [10], DRSformer [9], ChaIR [6], IRNeXT [33], OKNet [7], AST [12], SFHformer [11], Nerd-rain [13], and FSNet [20]. To ensure fairness, we retrain these methods from scratch in our environment using their official source codes.

Synthetic datasets. Table I presents the comparison results. It is evident that, thanks to the dual-domain multi-scale architecture, DMSR achieves superior performance across five benchmark datasets, notably outperforming the recent state-of-the-art method FSNet [20] by 0.71 dB on Rain100H [23]. Additionally,

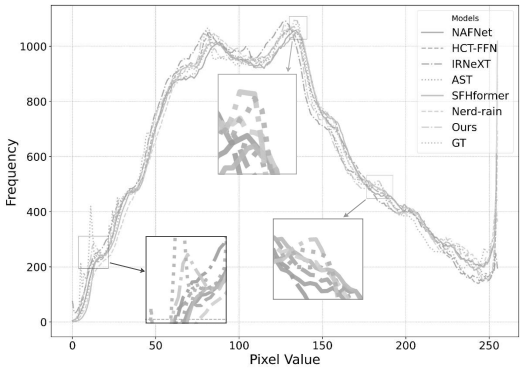

Fig. 2: The average fitting results of the Y channel histogram curve in the YCbCr space on the synthetic dataset.

Fig. 3 illustrates the visual quality comparison of samples generated by recent methods. Leveraging the advantages of our architecture, DMSR effectively removes rain streaks while preserving reliable background textures, delivering clearer deraining results consistent with the quantitative findings. Furthermore, we present the average fitting results of the Y channel histogram in the YCbCr space on the Rain100H dataset [23], showing that our DMSR deraining results closely match the GT image distribution (see Fig. 2).

Real-world datasets. To validate the performance of DMSR in real-world scenarios, we conducted comparisons on real datasets [38]. As shown in Fig. 4, our DMSR outperforms other sota methods in both rain removal and detail restoration, further demonstrating the advantages of our coarse-to-fine multi-scale paradigm across inter-scale and intra-scale dimensions, as well as the dual-domain high-level feature representation.

Perceptual quality assessment. To evaluate the perceptual quality of DMSR, we followed the method in [9], [12] and randomly selected 20 rain images from a real-world internet dataset for assessment [38]. As shown in Table II, indicate that DMSR achieves the lowest NIQE, meaning it produces images with better perceptual quality after rain removal.

More benchmark datasets. The Project page present additional experimental results on other benchmark datasets.

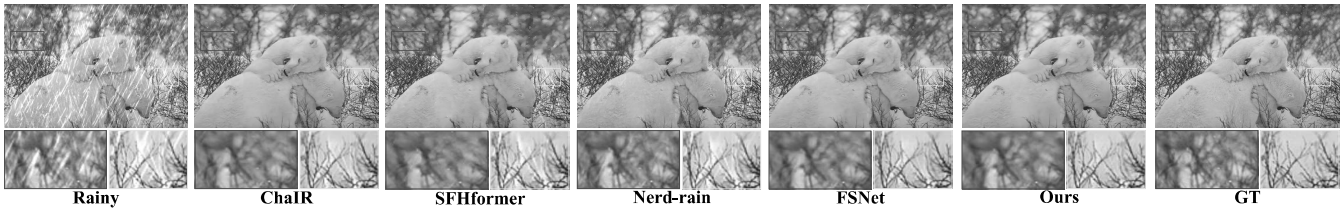

Fig. 3: Visual comparison on the Rain100H dataset [23]. Best viewed by zooming in the figures on high-resolution displays.

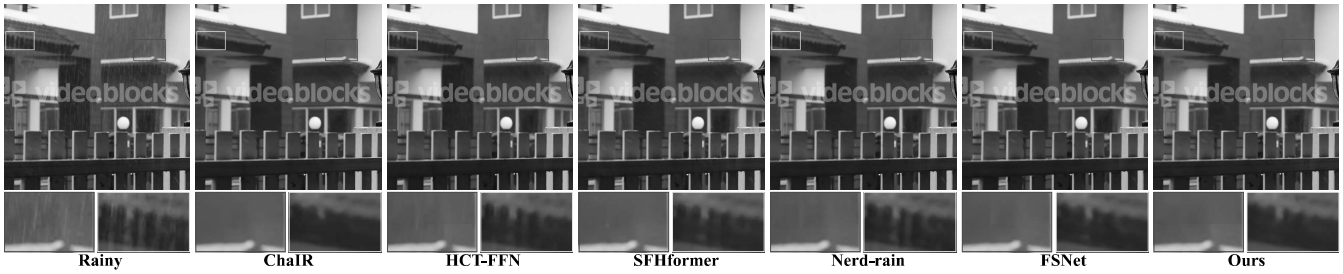

Fig. 4: Visual quality comparison on real-world dataset [38]. Best viewed by zooming in the figures on high-resolution displays.

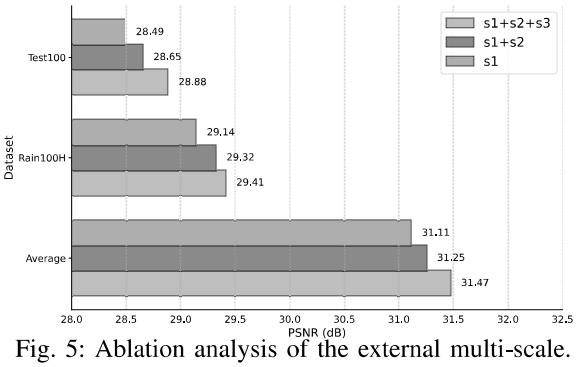

Fig. 5: Ablation analysis of the external multi-scale.

TABLE III: Ablation analysis on MPSRM in the intra-scale representation.

$scale\ branch^4$	$scale\ branch^2$	Test100 PSNR ↑	Test100 SSIM ↑	Rain100H PSNR ↑	Rain100H SSIM ↑	Average PSNR ↑	Average SSIM ↑
✗	✗	28.01	0.872	29.11	0.865	30.81	0.902
✓	✗	28.48	0.884	29.33	0.866	31.09	0.905
✗	✓	28.53	0.887	29.36	0.868	31.14	0.907
✓	✓	28.88	0.890	29.41	0.873	31.47	0.912

TABLE IV: Ablation analysis on FDSM in the intra-scale representation. The numbers in the table represent the convolution operations with corresponding kernel sizes.

method	Test100 PSNR ↑	Test100 SSIM ↑	Rain100H PSNR ↑	Rain100H SSIM ↑	Average PSNR ↑	Average SSIM ↑
3	28.42	0.882	29.21	0.869	31.14	0.906
3+5	28.69	0.886	29.33	0.871	31.38	0.909
3+5+7	28.88	0.890	29.41	0.873	31.47	0.912
3+3+3	28.55	0.884	29.30	0.869	31.29	0.910

D. Ablation Study

To ensure fairness and consistency, all our ablation experiments were conducted on five synthetic datasets under identical environments and training details. Due to space constraints, we only present results on two datasets and the average results across all five datasets.

Effectiveness of external multi-scale. To explore the effectiveness of external multi-scale image representations, we compared models with different scales, as shown in Fig. 5 Compared to single-scale rain removal, richer multi-scale representations provide significant gains to the baseline model. The joint effect of coarse and fine scales often leads to higher-quality results.

Effectiveness of internal multi-scale representation. To further explore the potential of multi-scale architectures, we introduced various forms of internal multi-scale structures in MPSRM and FDSM, creating a synergistic effect with the external multi-scale representations. To validate the effectiveness of internal multi-scale architectures, we progressively removed multi-scale structures within these two components. Specifically, in MPSRM, we denote the branches with $4\times$ downsampling and $2\times$ downsampling as $scale\ branch^4$ and $scale\ branch^2$, respectively, and remove these branches step by step. For FDSM, we replaced the convolution operations across different scales in the spatial local domain with unified 3×3 convolutions or progressively removed these multi-scale convolutions to eliminate its internal multi-scale structure.

The results of these comparisons are shown in Tables III and IV. Removing the internal multi-scale architecture from MPSRM and FDSM consistently led to performance degradation in rain removal. These findings underscore the importance of the dual multi-scale architecture, both external and internal, for achieving high-quality image deraining.

Ablation Study on MPSRM. We conducted a more detailed study of the individual elements within MPSRM. Specifically, we performed ablation experiments on the SPGA, the skip connections within SPGA, and the 3×3 convolution at the end of the module. The results of these experiments are shown in Table V. When all these critical elements are present, the deraining performance is optimal. This demonstrates the effectiveness of our carefully designed micro-level components.

Ablation Study on FDSM. We further analyzed the effectiveness of the FDSM in Table VI. The first row demonstrates the significance of spatial local domain operations, highlighting the necessity of local feature processing before extracting global frequency domain features. The remaining rows emphasize the importance of mixing global features in the frequency domain. The absence of Fourier transform operations significantly weakens FDSM's ability to model global dependencies and facilitate frequency domain interactions, leading to a noticeable

TABLE V: Ablation analysis of the proposed MPSRM.

SPGA	Skip connection	3x3 Conv	Test100 PSNR ↑	Test100 SSIM ↑	Rain100H PSNR ↑	Rain100H SSIM ↑	Average PSNR ↑	Average SSIM ↑
✗	✗	✗	28.34	0.880	28.79	0.862	30.82	0.907
✓	✗	✗	28.71	0.886	29.28	0.868	31.28	0.909
✓	✓	✗	28.82	0.888	29.37	0.871	31.39	0.910
✓	✓	✓	28.88	0.890	29.41	0.873	31.47	0.912

TABLE VI: Ablation analysis of the proposed FDSM.

Multi-Conv	FFT/IFFT	PWConv	Test100 PSNR ↑	Test100 SSIM ↑	Rain100H PSNR ↑	Rain100H SSIM ↑	Average PSNR ↑	Average SSIM ↑
✗	✓	✓	28.29	0.878	28.83	0.867	30.89	0.905
✓	✗	✗	28.41	0.881	29.02	0.869	31.02	0.908
✓	✓	✗	28.76	0.887	29.34	0.870	31.33	0.911
✓	✓	✓	28.88	0.890	29.41	0.873	31.47	0.912

decline in deraining performance. These results underline the critical role of both local spatial operations and global frequency domain interactions in achieving high-quality deraining.

E. Other Applications

To verify whether DMSR can extend its performance to image dehazing, we conducted comparisons on the RESIDE-6K [40] and Haze-4K [41] datasets against other dehazing methods, including DEANet [43], SFHformer [11] and Dehazeformer [42]. As shown in Table VII, our DMSR achieves the best performance. Meanwhile, we further explored the impact of DMSR's restoration performance on downstream visual tasks, with detailed results provided in the Project page.

IV. Conclusion

In this paper, we propose a novel Dual-Domain Multi-Scale Representation Network (DMSR) for single image deraining. Our architecture integrates both inter-scale and intra-scale information through a collaborative scale representation. The MPSRM removes rain streaks or droplets at multiple scales using a progressive coarse-to-fine approach, while SPGA expands the receptive field by enabling pixel-level perception of extended regions. Additionally, the FDSM extracts multi-scale spatial features and global frequency domain features, enhancing spectral representation through frequency modulation. Extensive experiments on six datasets show that DMSR achieves state-of-the-art performance.

TABLE VII: Comparison results of the dehazing experiments. See the Project page for more results.

Method	Year	RESIDE-6K [40] PSNR ↑	RESIDE-6K [40] SSIM ↑	Haze-4K [41] PSNR ↑	Haze-4K [41] SSIM ↑	Average PSNR ↑	Average SSIM ↑
Dehazeformer [42]	TIP2023	26.25	0.931	27.45	0.946	26.85	0.939
DEANet [43]	TIP2024	26.61	0.932	26.94	0.942	26.78	0.937
SFHformer [11]	ECCV2024	27.08	0.940	26.92	0.941	27.00	0.941
DMSR (ours)	–	**28.56**	**0.950**	**27.64**	**0.952**	**28.10**	**0.951**

References

[1] Hongming Chen, Xiang Chen, et al., "Rethinking multi-scale representations in deep deraining transformer," in *AAAI*, 2024.

[2] He Zhang and Vishal M Patel, "Convolutional sparse and low-rank coding-based rain streak removal," in *WACV*, 2017.

[3] Xia Li, Jianlong Wu, et al., "Recurrent squeeze-and-excitation context aggregation net for single image deraining," in *ECCV*, 2018.

[4] Dongwei Ren, Wangmeng Zuo, et al., "Progressive image deraining networks: A better and simpler baseline," in *CVPR*, 2019.

[5] Liangyu Chen, Xin Lu, Jie Zhang, et al., "Hinet: Half instance normalization network for image restoration," in *CVPRW*, 2021.

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

[7] Yuning Cui, Wenqi Ren, and Alois Knoll, "Omni-kernel network for image restoration," in *AAAI*, 2024.

[8] Tianyu Song, Guiyue Jin, Pengpeng Li, Kui Jiang, et al., "Learning a spiking neural network for efficient image deraining," *IJCAI*, 2024.

[9] Xiang Chen, Hao Li, et al., "Learning a sparse transformer network for effective image deraining," in *CVPR*, 2023.

[10] Xiang Chen, Jinshan Pan, et al., "Hybrid cnn-transformer feature fusion for single image deraining," in *AAAI*, 2023.

[11] Xingyu Jiang, Xiuhui Zhang, et al., "When fast fourier transform meets transformer for image restoration," in *ECCV*, 2024.

[12] Shihao Zhou et al., "Adapt or perish: Adaptive sparse transformer with attentive feature refinement for image restoration," in *CVPR*, 2024.

[13] Xiang Chen, Jinshan Pan, and Jiangxin Dong, "Bidirectional multi-scale implicit neural representations for image deraining," in *CVPR*, 2024.

[14] Xintian Mao, Yiming Liu, Fengze Liu, et al., "Intriguing findings of frequency selection for image deblurring," in *AAAI*, 2023.

[15] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao, "Multi-stage progressive image restoration," in *CVPR*, 2021.

[16] Maitreya Suin et al., "Spatially-attentive patch-hierarchical network for adaptive motion deblurring," in *CVPR*, 2020.

[17] Yuning Cui and Alois Knoll, "Dual-domain strip attention for image restoration," *Neural Networks*, 2024.

[18] Alexey Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[19] Ning Gao, Xingyu Jiang, et al., "Efficient frequency-domain image deraining with contrastive regularization," in *ECCV*, 2024.

[20] Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll, "Image restoration via frequency selection," *TPAMI*, 2024.

[21] Yuning Cui, Yi Tao, Luoxi Jing, and Alois Knoll, "Strip attention for image restoration," in *IJCAI*, 2023.

[22] He Zhang, Vishwanath Sindagi, and Vishal M Patel, "Image de-raining using a conditional generative adversarial network," *TCSVT*, 2019.

[23] Wenhan Yang, Robby T Tan, Jiashi Feng, et al., "Deep joint rain detection and removal from a single image," in *CVPR*, 2017.

[24] Xueyang Fu, Jiabin Huang, et al., "Removing rain from single images via a deep detail network," in *CVPR*, 2017.

[25] He Zhang and Vishal M Patel, "Density-aware single image de-raining using a multi-stream dense network," in *CVPR*, 2018.

[26] Qiaosi Yi, Juncheng Li, Qinyan Dai, et al., "Structure-preserving deraining with residue channel prior guidance," in *ICCV*, 2021.

[27] Kui Jiang et al., "Rain-free and residue hand-in-hand: A progressive coupled network for real-time image deraining," *TIP*, 2021.

[28] Kui Jiang, Zhongyuan Wang, Zheng Wang, Peng Yi, et al., "Danet: Image deraining via dynamic association learning," in *IJCAI*, 2022.

[29] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li, "Uformer: A general u-shaped transformer for image restoration," in *CVPR*, 2022.

[30] Kui Jiang, Zhongyuan Wang, Chen Chen, Zheng Wang, Laizhong Cui, and Chia-Wen Lin, "Magic elf: Image deraining meets association learning and transformer," *ACMMM*, 2022.

[31] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun, "Simple baselines for image restoration," in *ECCV*, 2022.

[32] Qiong Wang, Kui Jiang, et al., "Multi-scale fusion and decomposition network for single image deraining," *TIP*, 2024.

[33] Yuning Cui, Wenqi Ren, et al., "Irnext: Rethinking convolutional network design for image restoration," in *ICML*, 2023.

[34] Syed Waqas Zamir, Aditya Arora, et al., "Restormer: Efficient transformer for high-resolution image restoration," in *CVPR*, 2022.

[35] Xueyang Fu, Jiabin Huang, et al., "Removing rain from single images via a deep detail network," in *CVPR*, 2017.

[36] Yu Li, Robby T. Tan, Xiaojie Guo, Jiangbo Lu, and Michael S. Brown, "Rain streak removal using layer priors," in *CVPR*, 2016.

[37] He Zhang and Vishal M Patel, "Density-aware single image de-raining using a multi-stream dense network," in *CVPR*, 2018.

[38] Tianyu Wang, Xin Yang, et al., "Spatial attentive single-image deraining with a high quality real rain dataset," in *CVPR*, 2019.

[39] Kui Jiang, Zhongyuan Wang, et al., "Multi-scale progressive fusion network for single image deraining," in *CVPR*, 2020.

[40] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, et al., "Benchmarking single-image dehazing and beyond," *TIP*, 2018.

[41] Ye Liu, Lei Zhu, Shunda Pei, Huazhu Fu, Jing Qin, Qing Zhang, Liang Wan, and Wei Feng, "From synthetic to real: Image dehazing collaborating with unlabeled real data," in *ACMMM*, 2021.

[42] Yuda Song, Zhuqing He, et al., "Vision transformers for single image dehazing," *IEEE Transactions on Image Processing*, 2023.

[43] Zixuan Chen et al., "Dea-net: Single image dehazing based on detail-enhanced convolution and content-guided attention," *TIP*, 2024.