GDANET: FREQUENCY-SPACE COLLABORATIVE LEARNING WITH WAVELET-INSPIRED DIRECTIONAL PERCEPTION FOR SINGLE IMAGE DERAINING

Anonymous authors
Paper under double-blind review

ABSTRACT

Rain streaks are complex degradations in images, exhibiting multi-scale characteristics and diverse directional patterns, often spatially entangled with background textures. Therefore, effectively removing rain while preserving fine-grained image details remains a significant challenge in single image deraining. Traditional methods, whether based on pure spatial filtering or global frequency analysis, often fall short in simultaneously capturing both multi-scale and orientation-selective features crucial for rain removal. Inspired by the inherent ability of wavelet transforms to provide powerful multi-scale and directional feature representations, we propose a novel frequency-space collaborative learning framework, named GDANet. GDANet achieves content-adaptive rain suppression and detail restoration through the synergistic integration of three key components. Wavelet-inspired Directional Feature Module (WDFM) leverages multi-directional Gabor kernels to extract orientation-aware multi-scale rain features, effectively addressing the directional diversity of rain streaks. The Frequency-Spectrum Dynamic Gate Aggregation (FSDGA) dynamically modulates spectral components, reducing phase distortion while faithfully preserving structural details in the frequency domain. The Multi-scale Gated Fine-Grained module (MGFM) employs adaptive spatial attention to precisely separate complex rain patterns from background textures in the spatial domain. Furthermore, we design a multi-scale perception loss that jointly supervises content, edges, and log-frequency spectra across various resolutions, ensuring comprehensive detail preservation. GDANet demonstrates strong parameter efficiency: compared with the DMSR, it introduces only 0.255M additional parameters (from 0.838M to 1.093M). . Extensive experiments on Rain13K show GDANet achieves 37.47 dB PSNR on Rain100L, with clear gains in detail preservation. On the real-world RealRain-1K-L dataset, it achieves 41.02 dB PSNR and 0.986 SSIM, demonstrating strong performance in practical scenarios.

1 INTRODUCTION

Image deraining aims to recover clean scenes from rain-degraded images, benefiting applications like autonomous driving and surveillance (Garg & Nayar, 2007). Rain introduces occlusion, scattering, and blur, impairing both image quality and high-level tasks (Yang et al., 2017). Traditional methods rely on hand-crafted priors such as dictionary learning (Luo et al., 2015) and sparse coding, but struggle with diverse real-world rain patterns.

Deep learning has brought a paradigm shift. CNNs learn complex mappings from large-scale data (Zhang & Patel, 2018; Jiang et al., 2020; Li et al., 2018), but their limited receptive field causes structural distortions on large rain streaks (Ren et al., 2019; Zamir et al., 2021). Vision Transformers model global dependencies (Zamir et al., 2022; Chen et al., 2023c), yet suffer from quadratic complexity (Chen et al., 2023a) and lack explicit rain orientation modeling.

This dual limitation has prompted researchers to explore more efficient alternatives for global modeling and to revisit structural and directional priors in image degradation. Recently, state-space models (SSMs) and diffusion models have been introduced to image restoration. Mamba-based methods

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

like MaIR (Li et al., 2025a) and MambaIRv2 (Guo et al., 2025) balance global receptive fields and efficiency at CVPR 2025. FourierMamba (Li et al., 2025b) integrates Mamba into Fourier space to correlate different frequencies, and has been published at ICML 2025. Diffusion models also show strong generative capabilities for image restoration. Low-Rank Residual Diffusion Models (Tan & Yuan, 2026) achieve state-of-the-art performance on deraining benchmarks at CVPR 2026. However, these methods still underutilize the directional, scale-dependent, and spectral distribution priors of rain streaks in the frequency domain. Specifically, most existing methods operate primarily in the spatial domain, overlooking the unique characteristics of rain streaks in the frequency domain, which often results in poor high-frequency detail recovery. Recent frequency-domain approaches have demonstrated the benefits of spectral processing (Zou et al., 2025b), yet their multi-scale feature fusion and directional awareness remain underdeveloped. Among many frequency-domain methods, SFSNiD (Cong et al., 2024) proposes a Spatial-Frequency domain Information Interaction module that dynamically filters and aggregates multi-channel amplitude and phase spectrums for nighttime dehazing. Its core component, Frequency Spectrum Dynamic Aggregation (FSDA), adaptively modulates frequency components based on local spectral content. However, FSDA is designed for nighttime haze, where degradations are localized and frequency-inconsistent, and does not explicitly model the multi-scale characteristics of rain streaks. To this end, we propose FSDGA based on FSDA by introducing multi-scale spectral decomposition and a dynamic gating mechanism to adapt it for image deraining. We further validate the frequency-domain inductive bias of FSDGA through spectral analysis — experiments show that FSDGA reduces the phase L1 error by approximately 47% compared to FSDA, and its radial-band energy distribution is highly consistent with the ideal deraining operator.

To address these challenges, we propose GDANet (Gabor-guided Dual-domain Attention Network), a frequency-space collaborative learning framework that integrates CNNs, Fourier transform, and wavelet-inspired directional perception. The design of GDANet is motivated by a key observation: rain streaks exhibit complementary characteristics in the spatial, frequency, and directional domains. Spatial multi-scale convolutions capture local density variations and separate rain from background textures; Fourier spectral modulation distinguishes rain energy from background structures while preserving structural details; and Gabor-based directional filtering provides orientation priors that standard convolutions lack. By integrating these three views into a unified framework, GDANet avoids the limitations of single-domain methods and achieves robust deraining across diverse rain patterns. Specifically, GDANet consists of three complementary modules. The Multi-scale Gated Fine-Grained (MGFM) module employs adaptive spatial attention to precisely separate complex rain patterns from background textures in the spatial domain. The Frequency-Spectrum Dynamic Gate Aggregation (FSDGA) module dynamically modulates spectral components, reducing phase distortion while faithfully preserving structural details in the frequency domain. The Wavelet-inspired Directional Feature Module (WDFM) leverages 10 directional Gabor kernels to extract orientation-aware multi-scale rain features, effectively addressing the directional diversity of rain streaks. Furthermore, we design a multi-scale perception loss that jointly supervises content, edges, and log-frequency spectra across various resolutions, ensuring comprehensive detail preservation. Extensive experiments on Rain13K show GDANet achieves 37.47 dB PSNR on Rain100L, with clear gains in detail preservation. On the real-world RealRain-1K-L dataset, it achieves 41.02 dB PSNR and 0.986 SSIM, demonstrating strong performance in practical scenarios.

2 Related Work

2.1 Spatial-Domain Image Deraining

Image deraining has evolved from physics-based priors to data-driven approaches. Early methods used dictionary learning (Luo et al., 2015) and sparse coding, but generalize poorly to complex rain. Deep learning brought a paradigm shift. Early works like DetailNet (Fu et al., 2017) separate high-frequency details, while JORDER (Yang et al., 2017) jointly detects and removes rain. Multi-scale architectures became mainstream, including multi-stream dense networks (Zhang & Patel, 2018), multi-scale progressive fusion (Jiang et al., 2020), and recurrent attention networks (Li et al., 2018). However, CNNs have limited receptive fields, causing structural distortions on large rain streaks (Ren et al., 2019). Unsupervised methods like CSUD (Dong et al., 2025) use channel consistency and self-reconstruction to address lacking paired data.

Vision Transformers overcome this via self-attention. Restormer (Zamir et al., 2022) models global dependencies; DRSformer (Chen et al., 2023c) introduces sparse attention; Sparse Sampling Transformer (Chen et al., 2023a) handles raindrops and streaks. Transformers are also used in adverse weather (Sun et al., 2024). NeRD-Rain (Chen et al., 2024) uses implicit neural representations. Mamba-based methods like MaIR (Li et al., 2025a) and MambaIRv2 (Guo et al., 2025) balance global modeling and efficiency. PRE-Mamba (Ruan et al., 2025) extends to event camera deraining. Despite advances, self-attention remains quadratic and lacks explicit rain orientation priors, motivating frequency-domain representations. In summary, spatial-domain methods have progressed from hand-crafted priors to deep architectures, but they still struggle to explicitly model rain orientation and frequency characteristics. This gap motivates dual-domain and direction-aware designs.

2.2 Frequency-Domain and Multi-Scale Image Restoration

Spatial-domain methods are limited by local receptive fields (Li et al., 2019). Dual-domain learning incorporating frequency representations has emerged. Spectral processing benefits deraining (Zou et al., 2025b). Fourier-based priors exploit distinct rain patterns in frequency domain. However, most treat spectrum as a whole, ignoring amplitude-phase heterogeneity: amplitude reflects degradation intensity, phase governs structure. SFSNiD (Cong et al., 2024) proposes FSDA for nighttime dehazing, but haze degradations are localized and frequency-inconsistent, whereas rain spans multiple frequency bands across size, orientation, and density. Thus, directly applying haze-oriented modulation fails to suppress multi-scale rain while preserving background details.

Recent works decouple phase and amplitude. FourierMamba (Li et al., 2025b) integrates Mamba into Fourier space. FADformer (Gao et al., 2024) explores frequency-domain deraining with contrastive regularization. However, these lack scale-aware dynamic focusing for multi-scale rain. Overall, frequency-domain methods offer complementary cues to spatial processing, yet most existing approaches treat amplitude and phase uniformly and lack multi-scale spectral decomposition. A more fine-grained spectral modulation is needed for dense, multi-scale rain.

2.3 Wavelet Transform and Directional Perception

Fourier methods lack spatial locality. Discrete wavelet transform (DWT) offers spatial-frequency localization but traditional wavelets (Haar, Daubechies) have high translation sensitivity and only horizontal, vertical, diagonal directions, causing aliasing on arbitrary rain streaks. Direction-aware wavelet architectures like DAWN (Jiang et al., 2023) and DAWN+ (Jiang et al., 2025) improve performance but rely on vector decomposition, may not capture continuous orientation variations.

Gabor kernels are orientation-selective, rotatable to arbitrary angles, with elliptical spectral responses fitting elongated rain streaks. We combine wavelet-inspired decomposition with 10 directional Gabor kernels, providing finer directional selectivity while avoiding translation sensitivity and direction sparsity of standard DWT. This design offers a more flexible directional prior than existing wavelet-based deraining methods, which is essential for handling arbitrary rain orientations.

3 Method

Our method is built upon the multi-scale encoder-decoder backbone of DMSR (Figure 1(a)). The input image is downsampled to 1/2 and 1/4 resolutions to construct a multi-scale feature pyramid. Cross-scale feature fusion is performed through cascaded MFDAM (Multi-scale Frequency-Spatial Dynamic Aggregation Module, Figure 1(b)), followed by multi-branch reconstruction.

3.1 Multi-Scale Gated Fine-Grained Module (MGFM)

Rain streaks exhibit significant multi-scale characteristics, and a single-scale convolution cannot capture them simultaneously. Moreover, rain streaks are easily confused with background textures. To address this, we design the Multi-Scale Gated Fine-Grained Module (MGFM, Figure 1(c)), which first applies average pooling, then generates T_1 and T_2 through two 3×3 convolution branches. T_2 is fed into the Spatial Pixel Multi-Scale Gated Attention (SPMGA) to generate T_3, which is added to T_1. The fused result is processed by another SPMGA and added to the original feature map.

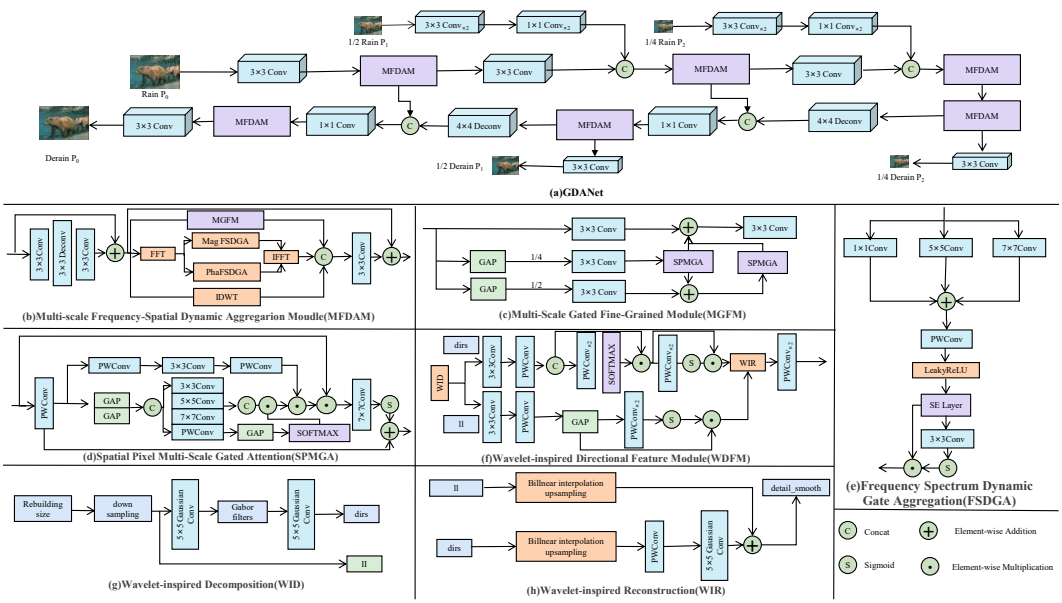

Figure 1: Overall architecture of GDANet. (a) The multi-scale encoder-decoder backbone. (b) The Multi-scale Frequency-Spatial Dynamic Aggregation Module (MFDAM), which integrates MGFM, FSDGA, and WDFM. (c) MGFM. (d) SPMGA. (e) FSDGA. (f) WDFM. (g) Wavelet-inspired Decomposition (WID). (h) Wavelet-inspired Reconstruction (WIR).

The architecture of SPMGA is shown in Figure 1(d), adopting a three-branch design. The upper branch is a residual connection. The lower branch adopts a PWConv-3 $\times$ 3Conv-PWConv bottleneck. The middle branch is the core attention module: average pooling and max pooling are used in parallel, followed by parallel 3×3, 5×5, and 7×7 convolutions to capture rain streaks at different scales. Channel attention weights are generated through PWConv and GAP+Softmax, which modulate the lower-branch features before being output through 7×7 Conv and Sigmoid.

3.2 Frequency-Spectrum Dynamic Gate Aggregation (FSDGA)

Rain streaks also exhibit multi-scale characteristics in the frequency domain, and spatial convolutions alone cannot distinguish whether a high-frequency component belongs to rain or background. The Fourier transform maps rain streaks to spectral energy concentrated along specific orientations and radii, distinguishable from background textures. This motivates us to perform deraining in the frequency domain.

We design the Frequency-Spectrum Dynamic Gate Aggregation (FSDGA) module, as shown in Figure 1(e). Given input $X \in \mathbb{R}^{C\times H\times W}$, FSDGA first applies a 2D DFT:

$$X_{freq} = \mathcal{F}_{2D}(X), \quad Mag = |X_{freq}|, \quad Pha = \angle X_{freq}. \tag{1}$$

Magnitude reflects rain energy intensity, while phase encodes positions and structural orientations; they require separate processing. Multi-scale spectral features are extracted via parallel 3×3, 5×5, and 7×7 convolutions:

$$F_{mag}^{ms} = \sum_{k\in\{3,5,7\}} Conv_k(Mag), \quad F_{pha}^{ms} = \sum_{k\in\{3,5,7\}} Conv_k(Pha). \tag{2}$$

The fused features are adjusted by pointwise convolution and weighted by SE channel attention:

$$F_1 = SE(PWConv(F_{mag}^{ms})), \quad F_1' = SE(PWConv(F_{pha}^{ms})). \tag{3}$$

To avoid uniformly suppressing both rain and background, we introduce a dynamic gating mechanism:

$$G = \sigma(Conv_{3\times3}(F_1)), \quad F_1^{gated} = F_1 \odot G. \tag{4}$$

When the local spectrum exhibits rain characteristics, $G \to 0$, suppressing that frequency; when it exhibits background characteristics, $G \to 1$, preserving it. The spectral components are updated via residual connections:

$$Mag' = Mag + F_1^{gated}, \quad Pha' = Pha + F_1'^{gated}, \tag{5}$$

and the complex spectrum is reconstructed and transformed back:

$$X'_{freq} = Mag' \cdot \cos(Pha') + j \cdot Mag' \cdot \sin(Pha'), \quad X_{out} = \mathcal{F}_{2D}^{-1}(X'_{freq}). \tag{6}$$

3.3 Wavelet-inspired Directional Feature Module (WDFM)

Standard DWT is translation-variant and covers only three directions, while rain streaks exhibit diverse orientations. Motivated by the multi-resolution analysis capability of wavelets and the directional selectivity of Gabor kernels, we design the Wavelet-inspired Directional Feature Module (WDFM, Figure 1(f)) with 10 directional Gabor kernels combined with Gaussian-smoothed LL subbands.

WDFM first performs WID (Wavelet-inspired Decomposition, Figure 1(g)), yielding a low-frequency component ll and high-frequency directional components $dirs = \{d_1, \ldots, d_{10}\}$. The LL subband is smoothed with a 5×5 Gaussian kernel G_5 and enhanced through channel attention:

$$ll_{smooth} = G_5 * ll \ (\text{groups} = C), \quad ll_{enhanced} = P_1 \odot \alpha, \tag{7}$$

where $P_1 = PWConv(DW_{3\times3}(ll))$ and $\alpha = \sigma(PWConv(PWConv(GAP(P_1))))$. We define 10 directional angles $\theta_d \in \{-75°, \ldots, 75°\}$ and construct a Gabor kernel for each direction:

$$g_{\theta_d}(x, y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cdot \cos(2\pi f x' + \psi), \tag{8}$$

where $x' = x\cos\theta_d + y\sin\theta_d$, $y' = -x\sin\theta_d + y\cos\theta_d$, with $\sigma = 2.0$, $\gamma = 0.5$, $f = 0.25$, $\psi = 0$, kernel size 7×7. Each kernel is normalized and stacked as $\mathcal{K}_{dir} \in \mathbb{R}^{10\times1\times7\times7}$. Directional responses are extracted via grouped convolution:

$$dir_{resp} = \text{conv2d}(ll_{smooth}, \mathcal{K}_{dir}, \text{padding} = 3, \text{groups} = C). \tag{9}$$

After concatenation, temperature-scaled Softmax ($\tau = 1.5$) produces adaptive directional weights:

$$w_d = \text{Softmax}\left(\frac{logits_d}{\tau}\right), \quad \sum_{d=1}^{10} w_d = 1. \tag{10}$$

The weighted high-frequency features are enhanced through spatial attention, and low/high-frequency components are reconstructed via WIR (Figure 1(h)). A residual scaling coefficient λ controls the contribution of the reconstructed features:

$$y = x + \lambda \cdot PWConv([x; \text{reconstructed}]), \quad \lambda = 0.2. \tag{11}$$

3.4 Multi-Scale Perception Loss

Our multi-scale perception loss consists of three complementary terms: content loss for pixel-level accuracy, edge loss via Laplacian of Gaussian for boundary preservation, and frequency loss for log-magnitude spectrum consistency. All losses are imposed at multiple scales.

$$\mathcal{L}_{total} = \lambda_c \mathcal{L}_{content} + \mathcal{L}_{edge} + \mathcal{L}_{freq}, \quad \lambda_c = 1.5. \tag{12}$$

We adopt the Charbonnier loss with $\epsilon = 10^{-3}$:

$$\mathcal{L}_{\text{Charbonnier}}(x, y) = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(x_i - y_i)^2 + \epsilon^2}. \tag{13}$$

Multi-scale content, edge, and frequency losses are computed at original, 1/2, and 1/4 resolutions with weights $(0.6, 0.25, 0.15)$ for content, $(0.6, 0.3, 0.1)$ for edge and frequency:

$$\mathcal{L}_{content} = \sum_{s\in\{0,1,2\}} w_s^c \mathcal{L}_C(\hat{I}_s, I_s), \tag{14}$$

$$\mathcal{L}_{edge} = \sum_{s\in\{0,1,2\}} w_s^e \mathcal{L}_C(\mathcal{L}(\hat{I}_s), \mathcal{L}(I_s)), \tag{15}$$

$$\mathcal{L}_{freq} = \sum_{s\in\{0,1,2\}} w_s^f \mathcal{L}_C(\log A_{\hat{I}_s}, \log A_{I_s}). \tag{16}$$

This multi-resolution, multi-domain loss ensures pixel-wise accuracy, perceptual clarity, and enriched high-frequency details.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Our model is implemented using PyTorch and optimized with the Adam optimizer. During training, image patches of size 64×64 are used with a batch size of 120. The model is trained for 3000 epochs, starting with an initial learning rate of 2×10^{-4}, which is then decayed using a cosine annealing strategy. A linear warmup is applied for the first 3 epochs to stabilize training. Data augmentation techniques, including random cropping and flipping, are employed. Experiments are conducted on 4 NVIDIA RTX 5060Ti GPUs. For testing, a sliding window strategy is adopted to handle arbitrary input resolutions.

4.1.1 DATASETS AND EVALUATION METRICS

We evaluate our model on several widely used synthetic benchmark datasets, including Rain100L (Yang et al., 2017), Rain100H (Yang et al., 2017), Test100 (Zhang et al., 2019), Test2800 (Fu et al., 2017), and Test1200 (Zhang & Patel, 2018). These datasets cover a wide range of rain densities and background textures, from light rain streaks in Rain100L to heavy rain in Rain100H, and diverse real-world scenarios in Test100, Test2800, and Test1200.

To further assess performance on real-world data, we evaluate on the Raindrop Clarity (Day) dataset, which contains two subsets: Raindrop-A (raindrop-focused) and Raindrop-B (background-focused), and the RealRain-1k dataset, which provides paired real rain images with two density levels: RealRain-1k-L (light rain) and RealRain-1k-H (heavy rain). The primary evaluation metrics are Peak Signal-to-Noise Ratio (PSNR, in dB) and Structural Similarity Index (SSIM).

4.2 EFFICIENCY ANALYSIS

To evaluate the computational efficiency of our modules, we measure parameter count and inference speed on an NVIDIA RTX 5060Ti GPU with batch size 1 and input resolution 256×256. Each module is warmed up for 10 iterations and averaged over 200 runs.

Table 1: Module-level efficiency comparison. All metrics are measured at 256×256 resolution.

Architecture	Module	Params (M)	FPS
Ours (GDANet)	MGFM	0.763	171.9
	FSDGA	**0.058**	**493.2**
	WDFM	0.272	113.0
DMSR (Baseline)	MPSRM	0.763	181.2
	FDSM	0.075	1059.8

As shown in Table 1, our FSDGA module achieves 493.2 FPS with only 0.058M parameters, outperforming the original FSDM (0.075M) in parameter efficiency while delivering a +0.44 dB PSNR gain. MGFM maintains comparable parameter count to MPSRM (0.763M) with a slight speed reduction (171.9 vs. 181.2 FPS) but substantial accuracy improvement. WDFM, as a newly added directional perception module, runs at 113.0 FPS and provides directional priors absent in the baseline. The complete GDANet model runs at 4.26 FPS, with the main bottleneck lying in the multi-scale encoder-decoder backbone rather than our proposed modules. These results demonstrate that our modules achieve a favorable trade-off between restoration quality and computational efficiency, making the framework practical for real-world deployment.

4.3 Comparison with other Methods

We conduct extensive comparisons between our proposed GDANet and several leading image deraining algorithms. The methods selected for comparison represent diverse architectural advancements in the field, including CNN-based, Transformer-based, and dual-domain approaches.

4.3.1 Results on Synthetic Datasets

Table 2 presents the quantitative comparison of GDANet on various synthetic datasets, covering a wide range of rain densities and background textures. Across all five benchmarks, GDANet achieves the best PSNR on four datasets, demonstrating its effectiveness in handling diverse rain patterns and preserving image details. These consistent gains validate the robustness of our frequency-spatial collaborative design.

Table 2: Quantitative comparison (PSNR/SSIM) on synthetic deraining datasets. Higher is better.

Method	Test100		Rain100H		Rain100L		Test2800		Test1200		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
MSPFN (Jiang et al., 2020)	27.50	0.876	28.66	0.860	32.40	0.933	32.82	0.930	32.39	0.916	30.75	0.903
PCNet (Jiang et al., 2021)	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 (Zamir et al., 2021)	25.66	0.859	28.23	0.850	31.94	0.930	32.14	0.925	31.32	0.901	29.86	0.893
NAFNet (Chen et al., 2022)	25.75	0.845	26.76	0.813	31.27	0.925	31.71	0.918	30.62	0.892	29.22	0.879
Uformer (Wang et al., 2022)	23.87	0.815	22.43	0.700	28.39	0.883	29.71	0.886	28.65	0.856	26.61	0.828
DRSformer (Chen et al., 2023c)	27.86	0.885	28.16	0.864	34.79	0.954	32.80	0.931	30.99	0.906	30.92	0.908
HCT-FFN (Chen et al., 2023b)	24.86	0.847	26.70	0.819	29.94	0.906	31.46	0.915	31.23	0.901	28.84	0.878
AST (Zhou et al., 2024)	26.07	0.859	27.40	0.889	36.88	0.966	33.73	0.939	32.56	0.913	31.33	0.913
SFFormer (Wang et al., 2024)	25.67	0.856	27.25	0.832	32.97	0.944	32.27	0.925	31.50	0.904	29.93	0.892
DMSR (Zou et al., 2025b)	28.88	0.890	29.41	0.873	35.19	0.957	32.50	0.931	31.35	0.910	31.47	0.912
DAIT (Zou et al., 2025a)	29.65	0.895	29.68	0.875	35.98	0.964	33.00	0.933	31.52	0.913	31.97	0.916
SIDC-GAN (Sk et al., 2026)	28.51	0.818	30.70	0.792	27.03	0.884	33.00	0.876	33.20	0.866	30.49	0.847
GDANet (Ours)	**30.04**	**0.899**	**30.14**	**0.888**	**37.47**	**0.970**	**33.59**	**0.938**	29.76	0.888	**32.20**	**0.917**

Our GDANet achieves the best performance on four out of five synthetic datasets. Specifically, on Rain100L, our method reaches 37.47 dB, outperforming the strongest competitor AST by 0.59 dB. On Test100 and Test2800, we also achieve the highest PSNR of 30.04 dB and 33.59 dB, respectively. These improvements can be attributed to the complementary design of our three modules: MGFM effectively distinguishes multi-scale rain streaks from background textures in the spatial domain, while FSDGA performs precise spectral modulation in the frequency domain, and WDFM captures directional priors via multi-directional Gabor kernels. The consistent gains across datasets with varying rain densities demonstrate the robustness of our frequency-spatial collaborative learning framework.

It is worth noting that on Test1200, our method obtains 29.76 dB, which is lower than the baseline DMSR (31.35 dB). We attribute this to the fact that Test1200 contains dense and fragmented rain streaks, and our strong directional constraints and multi-resolution frequency restrictions may over-constrain the reconstruction of fine textures in such scenarios. Nevertheless, the model still maintains good structural similarity (SSIM 0.888) on this dataset, and the consistent performance gains on the other four datasets validate the overall effectiveness of our module design.

Visual Comparison. Figures 2 and 3 illustrate the visual quality comparison on two challenging scenes. Our method effectively removes rain while preserving fine background details, avoiding common artifacts such as over-smoothing or structural distortions. The Gabor filter bank injects directional priors into feature extraction, enabling more accurate separation of rain streaks from background edges.

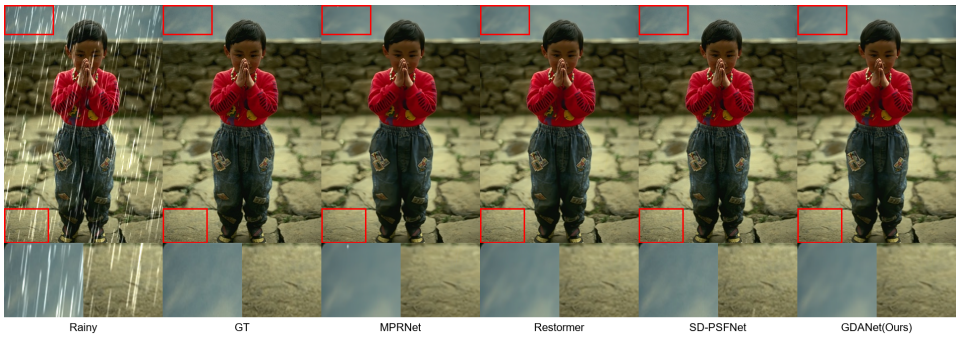

Figure 2: Visual comparison on a rainy image. From left to right: Rainy input, GT, MPRNet, Restormer, SD-PSFNet, and GDANet (Ours). The bottom row shows zoomed-in patches from the red-boxed regions. Our method recovers finer details and better preserves background textures.

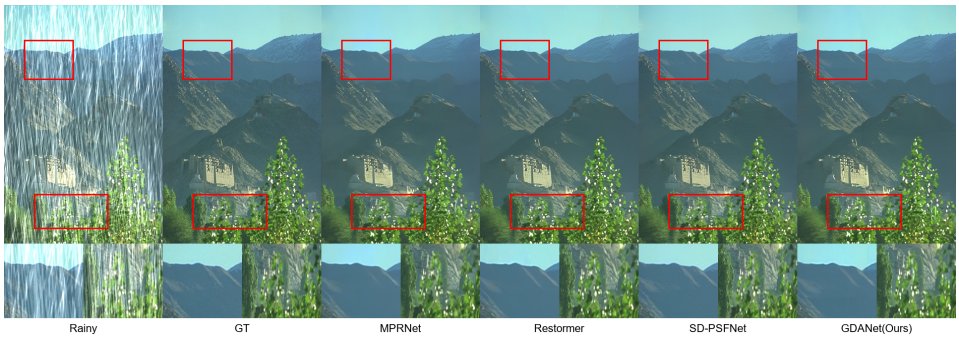

Figure 3: Visual comparison on another challenging scene. Our GDANet effectively removes rain streaks while preserving the structural details of the background, outperforming other methods in both global and local regions.

4.3.2 Results on Raindrop and Real-World Datasets

We further evaluate GDANet on the Raindrop Clarity (Day) dataset and the RealRain-1k dataset. Due to space limits, the detailed quantitative results and analysis are provided in Appendix A and Appendix A. In summary, GDANet achieves 32.15 dB / 0.933 on Raindrop-A, 26.56 dB / 0.827 on Raindrop-B, 41.02 dB / 0.986 on RealRain-1k-L, and 39.33 dB / 0.982 on RealRain-1k-H, demonstrating strong generalization to raindrop and real-world rain scenarios.

4.4 Ablation Study

To validate the effectiveness of each proposed module, we conduct ablation studies on five benchmarks. Starting from the DMSR baseline, we progressively integrate FSDGA, MGFM, WDFM, and the multi-scale perception loss. As shown in Table 3, each module contributes positively to the overall performance.

Table 3: Ablation study of the proposed modules. ✓ indicates the module is enabled.

FSDGA	MGFM	WDFM	Test100		Rain100H		Rain100L		Test2800		Test1200		Average	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
×	×	×	28.88	0.890	29.41	0.873	35.19	0.957	32.50	0.931	31.35	0.910	31.47	0.912
✓	×	×	28.73	0.894	29.58	0.878	36.04	0.964	33.15	0.935	32.06	0.910	31.91	0.916
✓	✓	×	29.32	0.899	30.00	0.885	36.56	0.969	33.31	0.936	31.09	0.903	32.05	0.918
✓	✓	✓	**30.04**	**0.899**	**30.14**	**0.888**	**37.47**	**0.970**	**33.59**	**0.938**	29.76	0.888	**32.20**	**0.917**

Effectiveness of FSDGA. Introducing FSDGA alone improves the average PSNR from 31.47 dB to 31.91 dB (+0.44 dB), with consistent gains on Test2800 and Test1200. This demonstrates that adaptive spectral modulation with dynamic gating effectively separates rain streaks from background structures in the frequency domain. The improvement is particularly notable on Test2800 (+0.65 dB) and Test1200 (+0.71 dB), where rain streaks span multiple scales. This confirms that FSDGA's multi-scale spectral decomposition and dynamic gating mechanism successfully capture both fine and coarse rain components, which a single-scale frequency filter cannot achieve.

Effectiveness of MGFM. Adding MGFM further boosts the average PSNR to 32.05 dB (+0.14 dB over FSDGA alone). On Rain100H and Rain100L, the improvements reach 0.42 dB and 0.52 dB, respectively. The spatial-channel attention mechanism enhances the network's ability to discriminate rain streaks from background textures. Notably, the improvement on Rain100H (0.42 dB) indicates that MGFM's multi-scale spatial filtering is particularly effective when rain streaks are dense and heavily overlapped with background details.

Effectiveness of WDFM. The complete model with WDFM achieves the highest average PSNR of 32.20 dB. Notably, on Test100 and Rain100L, the improvements reach 1.31 dB and 1.43 dB over the baseline, validating the effectiveness of Gabor-based directional filtering in capturing multi-directional rain streak features. The significant gain on Test100 (1.31 dB) suggests that WDFM's 10-directional Gabor kernels provide crucial directional priors that are absent in the baseline, enabling the model to handle complex rain orientations more effectively.

4.5 Frequency-Domain Behavior and Inductive Bias Analysis

To verify whether FSDGA truly provides a frequency-domain inductive bias, we conduct a spectrum analysis on an untrained probe network. The detailed metrics and analysis are provided in Appendix A. In summary, FSDGA reduces the phase L1 error by approximately 47% compared to FSDA, and its radial-band energy distribution closely matches the ideal deraining operator, confirming that the frequency prior of FSDGA is properly learned.

5 Conclusion

In this paper, we proposed GDANet, a frequency-spatial collaborative learning framework for single image deraining. GDANet integrates three complementary modules: MGFM for multi-scale spatial attention, FSDGA for adaptive spectral modulation, and WDFM for wavelet-inspired decomposition and Gabor-based directional perception. Experiments on five synthetic benchmarks and real-world datasets demonstrate that GDANet achieves state-of-the-art performance, with 37.47 dB PSNR on Rain100L and 41.02 dB PSNR on RealRain-1K-L. Ablation and frequency-domain analyses validate the effectiveness of each module.

AI Use Statement

In this work, we used generative AI tools for drafting and polishing sections of the manuscript, as well as for literature retrieval. All AI-assisted content has been reviewed and verified by the authors. We take responsibility for the final content of this work.

Reproducibility Statement

We provide the training details, evaluation metrics, and dataset descriptions in Section 4.1. The code and pre-trained models will be released upon acceptance to facilitate reproducibility.

References

Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European conference on computer vision*, pp. 17–33. Springer, 2022.

S. Chen, T. Ye, J. Bai, E. Chen, J. Shi, and L. Zhu. Sparse sampling transformer with uncertainty-driven ranking for unified removal of raindrops and rain streaks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023a.

X. Chen, J. Pan, J. Lu, Z. Fan, and H. Li. Hybrid cnn-transformer feature fusion for single image deraining. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 378–386, 2023b.

Xiang Chen, Hao Li, Mingqiang Li, and Jinshan Pan. Learning a sparse transformer network for effective image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5896–5905, 2023c.

Xiang Chen, Jinshan Pan, and Jiangxin Dong. Bidirectional multi-scale implicit neural representations for image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25627–25636, 2024.

Xiaofeng Cong, Jie Gui, Jing Zhang, Junming Hou, and Hao Shen. A semi-supervised nighttime dehazing baseline with spatial-frequency aware and realistic brightness constraint. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2631–2640, 2024.

Guanglu Dong, Tianheng Zheng, Yuanzhouhan Cao, Linbo Qing, and Chao Ren. Channel consistency prior and self-reconstruction strategy based unsupervised image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7469–7479, 2025.

Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

Ning Gao, Xingyu Jiang, Xiuhui Zhang, and Yue Deng. Efficient frequency-domain image deraining with contrastive regularization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024.

K. Garg and S. K. Nayar. Vision and rain. *International Journal of Computer Vision*, 75(1):3–27, 2007.

Hang Guo, Yong Guo, Yaohua Zha, Yulun Zhang, Wenbo Li, Tao Dai, Shu-Tao Xia, and Yawei Li. Mambairv2: Attentive state space restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 28124–28133, 2025.

K. Jiang, Z. Wang, P. Yi, C. Chen, B. Huang, Y. Luo, J. Ma, and J. Jiang. Multi-scale progressive fusion network for single image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

K. Jiang, Z. Wang, P. Yi, C. Chen, Z. Wang, X. Wang, J. Jiang, and C.-W. Lin. Rain-free and residue hand-in-hand: A progressive coupled network for real-time image deraining. *IEEE Transactions on Image Processing*, 30:7404–7418, 2021.

Kui Jiang, W. Liu, Z. Wang, X. Zhong, J. Jiang, and C.-W. Lin. DAWN: Direction-aware attention wavelet network for image deraining. In *Proceedings of the 31st ACM International Conference on Multimedia (ACM MM)*, pp. 7065–7074, 2023.

Kui Jiang, Junjun Jiang, Zheng Wang, Zihan Geng, and Xianming Liu. DAWN+: Wavelet-based image deraining meets direction-aware attention and mutual representation. *IEEE Transactions on Neural Networks and Learning Systems*, 36:18244–18258, 2025.

Boyun Li, Haiyu Zhao, Wenxin Wang, Peng Hu, Xi Peng, and Jiancheng Lv. Mair: A locality- and continuity-preserving mamba for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025a.

Dong Li, Y. Liu, X. Fu, J. Zhang, B. Cheng, Z. Wang, X. Chen, H. Zhang, Y. Xu, and R. Li. Fouriermamba: Fourier learning integration with state space models for image deraining. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2025b.

S. Li, I. B. Araujo, W. Ren, Z. Wang, E. K. Tokuda, R. H. Junior, R. Cesar-Junior, J. Zhang, X. Guo, and X. Cao. Single image deraining: A comprehensive benchmark analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3838–3847, 2019.

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 (ECCV)*, 2018.

Yu Luo, Yong Xu, and Hui Ji. Removing rain from a single image via discriminative sparse coding. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015.

Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

Ciyu Ruan, Ruishan Guo, Zihang Gong, Jingao Xu, Wenhan Yang, and Xinlei Chen. Pre-mamba: A 4d state space model for ultra-high-frequent event camera deraining. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9169–9180, 2025.

S. Sk, B. Das, A. S. Al-Shamayleh, A. Ali, S. Acharjee, and S. S. Chaudhuri. Sidc-gan: Single image de-rain cascade unet gan. *IEEE Open Journal of Vehicular Technology*, 2026.

Shangquan Sun, Wenqi Ren, Xinwei Gao, Rui Wang, and Xiaochun Cao. Restoring images in adverse weather conditions via histogram transformer. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024.

Junfu Tan and Jiang Yuan. Low-rank residual diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 35747–35757, 2026.

X. Wang, H. Zhang, K. Cai, D. Miao, Q. Zhang, and M. Li. Sfformer: Adaptive sparse and frequency-guided transformer network for single image derain. In *Pattern Recognition and Computer Vision (PRCV)*, 2024.

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/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17683–17693, 2022.

Wenhan Yang, Robby T. Tan, Jiashi Feng, Ji Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

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/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5728–5739, 2022.

H. Zhang, V. Sindagi, and V. M. Patel. Image de-raining using a conditional generative adversarial network. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(11):3943–3956, 2019.

He Zhang and Vishal M. Patel. Density-aware single image de-raining using a multi-stream dense network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

Shihao Zhou, Duosheng Chen, Jinshan Pan, Jinglei Shi, and Jufeng Yang. Adapt or perish: Adaptive sparse transformer with attentive feature refinement for image restoration. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2952–2963. IEEE, 2024.

Shun Zou, Yi Zou, Juncheng Li, Guangwei Gao, and Guojun Qi. Cross paradigm representation and alignment transformer for image deraining. In *Proceedings of the 33rd ACM International Conference on Multimedia (ACM MM)*, 2025a.

Shun Zou, Yi Zou, Mingya Zhang, Shipeng Luo, Guangwei Gao, and Guojun Qi. Learning dual-domain multi-scale representations for single image deraining. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, 2025b.

A Appendix

Frequency-Domain Behavior and Inductive Bias Analysis

To verify whether FSDGA truly provides a frequency-domain inductive bias for deraining, rather than merely fitting the data in the spatial domain, we conduct a spectrum analysis on an *untrained* probe network. The probe network consists of a commonly initialized $3 \to 32$ convolution followed by the frequency module under comparison. The input is a Rain100H/Rain100L image, and we compute the L1/L2 errors of the magnitude and phase spectra after `rfft2`.

Aligned comparison metrics. To measure whether the module's modification agrees with the ideal deraining operator, we define the ideal and actual spectral modifications as

$$dX_{\text{ideal}} = \mathcal{F}(\text{conv_in}(I_{\text{clean}})) - \mathcal{F}(\text{conv_in}(I_{\text{rainy}})), \tag{17}$$

$$dX_{\text{mod}} = \mathcal{F}(\text{module}(\text{conv_in}(I_{\text{rainy}}))) - \mathcal{F}(\text{conv_in}(I_{\text{rainy}})), \tag{18}$$

where $\mathcal{F}(\cdot)$ denotes the complex spectrum produced by `rfft2` (DC column excluded by default). For complex spectra A, B, we define the real inner product as

$$\langle A, B \rangle = \text{Re}\left(\sum_i A_i B_i^* \right), \tag{19}$$

where B_i^* is the complex conjugate of B_i. The three aligned metrics are then

$$\cos = \frac{\langle dX_{\text{mod}}, dX_{\text{ideal}} \rangle}{\|dX_{\text{mod}}\|_2 \cdot \|dX_{\text{ideal}}\|_2}, \quad \text{proj} = \frac{\langle dX_{\text{mod}}, dX_{\text{ideal}} \rangle}{\|dX_{\text{ideal}}\|_2^2}, \tag{20}$$

$$\text{ratio} = \frac{\sum_i |dX_{\text{mod},i}|}{\sum_i |dX_{\text{ideal},i}|}. \tag{21}$$

Untrained spectral preservation. Table 4 reports the untrained frequency statistics. Even before any training, FSDGA already preserves the input spectrum much better than FSDA: the magnitude L1 error drops from 1.701 to 0.923 (a 46% reduction), and the phase L1 error drops from 0.2205 to 0.1189 (also a 46% reduction). The phase L2 error exhibits an even more dramatic reduction, from 0.1479 to 0.0395 (a 73% reduction). This indicates that the multi-scale spectral decomposition in FSDGA provides a strong architectural prior: the module knows where to modify even without training, which significantly reduces the burden on the subsequent learning process.

Table 4: Untrained spectral statistics on Rain100H/Rain100L. Lower is better.

Setting	Mag L1	Phase L1	Phase L2
FSDA (in→out)	1.701	0.2205	0.1479
FSDGA (in→out)	0.923	0.1189	0.0395

Frequency prior and direction decoupling. Table 5 reports the aligned comparison. Both FSDA and FSDGA obtain cos and proj values close to zero before training, because the randomly initialized dynamic gate has not yet learned the deraining direction. However, the magnitude of the modification $|dX|$ reveals a clear difference: the ideal operator produces $|dX| = 10.3079$, whereas

FSDA produces $|dX| = 3.6575$ and FSDGA produces $|dX| = 2.4092$. More importantly, the phase modification $|d\phi|$ of FSDGA (0.1189) is nearly half that of FSDA (0.2205), indicating that FSDGA makes much smaller and more precise phase adjustments, thereby avoiding unnecessary global phase distortion.

Table 5: Aligned comparison on the same `conv_in` feature domain. The ideal row is the reference operator.

Change vs. rainy	cos	proj	Mag $\|dX\|$	ratio	Phase $\|d\phi\|$
Rain-free − rainy (ideal)	–	–	10.3079	1.0000	0.7505
FSDA (baseline) − rainy	0.0951	0.0334	3.6575	0.3548	0.2205
FSDGA (ours) − rainy	−0.0045	−0.0021	2.4092	0.2337	0.1189

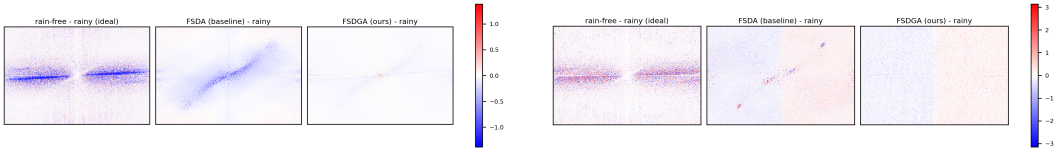

Figure 4: Aligned magnitude (left) and phase (right) delta maps. From left to right in each panel: ideal (rain-free − rainy), FSDA − rainy, and FSDGA − rainy.

Radial-band analysis. Table 6 further breaks down the $|dX|$ energy distribution across radial frequency bands. For bands 2–7, the share difference between FSDGA and the ideal operator is generally less than 1%, whereas FSDA deviates noticeably in low-frequency bands 0–1. This confirms that FSDGA's multi-scale spectral decomposition correctly identifies which frequency bands should be modified, while FSDA lacks this band-selection ability.

Table 6: Radial-band share of $|dX|$ (%) on Rain100H/Rain100L.

Band	Radius $[lo, hi)$	Ideal	FSDA	FSDGA
0	$[0.000, 0.125)$	13.95	12.60	16.13
1	$[0.125, 0.250)$	20.64	18.33	18.89
2	$[0.250, 0.375)$	20.87	20.95	20.02
3	$[0.375, 0.500)$	17.75	19.20	17.81
4	$[0.500, 0.625)$	13.89	15.19	14.14
5	$[0.625, 0.750)$	9.93	10.58	9.93
6	$[0.750, 0.875)$	2.43	2.60	2.55
7	$[0.875, 1.000)$	0.54	0.55	0.52

Summary of frequency-domain findings. FSDGA effectively decouples deraining into two subproblems—"band localization" (which frequency bands should be modified) and "direction alignment" (in which direction to modify them). The multi-scale spectral convolution prior solves the former, while the dynamic gate learns the latter during training. The baseline FSDA lacks this decoupling, causing both aspects to be learned poorly.

Results on Raindrop Datasets

We evaluate GDANet on the Raindrop Clarity (Day) dataset, which contains two subsets: Raindrop-A (raindrop-focused, where raindrops are sharp and the background is blurred) and Raindrop-B (background-focused, where the background is clear but raindrops are blurred). This dataset requires the model to handle both raindrop removal and focus-dependent blur restoration.

Table 7: Quantitative results (PSNR/SSIM) of GDANet on Raindrop-A and Raindrop-B datasets.

Dataset	PSNR	SSIM
Raindrop-A	32.15	0.933
Raindrop-B	26.56	0.827
Average	**29.35**	**0.880**

As shown in Table 7, GDANet achieves 32.15 dB / 0.933 on Raindrop-A and 26.56 dB / 0.827 on Raindrop-B, with an average of 29.35 dB / 0.880. The higher performance on Raindrop-A indicates that our method is particularly effective when raindrops are sharp and well-defined, which aligns with the design of FSDGA: the dynamic gating mechanism can precisely identify and suppress the spectral components corresponding to sharp raindrops while preserving the blurred background structure. On Raindrop-B, where raindrops are blurred, the frequency-domain separation becomes more challenging, yet our method still delivers competitive results.

RESULTS ON REAL-WORLD DATASET

We evaluate GDANet on the RealRain-1k dataset, which provides paired real rain images with light and heavy rain densities. Unlike synthetic datasets, RealRain-1k contains complex rain patterns with varying intensities, orientations, and physical characteristics.

Table 8: Quantitative results (PSNR/SSIM) on the RealRain-1k dataset.

Dataset	PSNR	SSIM
RealRain-1k-L (Light Rain)	41.02	0.986
RealRain-1k-H (Heavy Rain)	39.33	0.982
Average	**40.18**	**0.984**

As shown in Table 8, GDANet achieves 41.02 dB / 0.986 on RealRain-1k-L and 39.33 dB / 0.982 on RealRain-1k-H, with an average of 40.18 dB / 0.984. The slight performance drop from light rain to heavy rain (1.69 dB) is expected, as heavy rain introduces more severe occlusion and scattering. The high SSIM (0.982) on RealRain-1k-H indicates that our model preserves structural information even under dense rain, which can be attributed to the combined effect of MGFM's multi-scale spatial filtering and WDFM's Gabor-based directional filtering.