SD-PSFNet: Sequential and Dynamic Point Spread Function Network for Image Deraining

Jiayu Wang[1]*, Haoyu Bian[2,3]*, Haoran Sun[1], Shaoning Zeng[2,3]†

[1]School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, China
[2]Yangtze Delta Region Institute (Huzhou), University of Electronic Science and Technology of China, Huzhou, China
[3]School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China
aster20060626@gmail.com, haoyubian04@gmail.com, 2022090916002@std.uestc.edu.cn, zsn@outlook.com

Abstract

Image deraining is crucial for vision applications but is challenged by the complex multi-scale physics of rain and its coupling with scenes. To address this challenge, a novel approach inspired by multi-stage image restoration is proposed, incorporating Point Spread Function (PSF) mechanisms to reveal the image degradation process while combining dynamic physical modeling with sequential feature fusion transfer, named SD-PSFNet. Specifically, SD-PSFNet employs a sequential restoration architecture with three cascaded stages, allowing multiple dynamic evaluations and refinements of the degradation process estimation. The network utilizes components with learned PSF mechanisms to dynamically simulate rain streak optics, enabling effective rain-background separation while progressively enhancing outputs through novel PSF components at each stage. Additionally, SD-PSFNet incorporates adaptive gated fusion for optimal cross-stage feature integration, enabling sequential refinement from coarse rain removal to fine detail restoration. Our model achieves state-of-the-art PSNR/SSIM metrics on Rain100H (33.12dB/0.9371), RealRain-1k-L (42.28dB/0.9872), and RealRain-1k-H (41.08dB/0.9838). In summary, SD-PSFNet demonstrates excellent capability in complex scenes and dense rainfall conditions, providing a new physics-aware approach to image deraining.

Code — https://github.com/Aster-1024/SD-PSFNet

Introduction

Image deraining is a fundamental restoration task aiming to recover rain-free images from rainy ones. This task is crucial for downstream applications (Su et al. 2023) such as object detection (Wang et al. 2022a), autonomous driving (Sun, Ang, and Rus 2019), and surveillance systems (Dey and Bhattacharjee 2020), as rain-degraded images suffer from reduced visibility, decreased contrast, and detail loss that significantly impact visual processing performance. In recent years, deep learning-based deraining methods have made significant progress (Wang et al. 2020a). From early CNN-based models to approaches incorporating Transformer architectures such as Restormer (Zamir et al. 2022) and Uformer (Wang et al. 2022b), as well as GAN-based techniques like SSCGAN (Shao et al. 2021) and LFI-based (Ding et al. 2021), these approaches have demonstrated remarkable capabilities in image deraining tasks.

The primary challenges in image deraining arise from the complexity, multi-scale nature, and high coupling of raindrop distributions with background scenes. Most existing methods primarily focus on learning mapping relationships from data, overlooking the physical and optical properties of raindrop formation (Yang et al. 2020). This neglect leads to an inability to dynamically adjust to different rain patterns, resulting in insufficient deraining and lack of explainability. Although Transformer-based architectures, GAN-based models, and diffusion models demonstrate superior performance in image restoration tasks, their massive parameter counts and computational overhead hinder deployment in real-time applications (Chen et al. 2025b). Additionally, existing methods attempt to incorporate domain knowledge through physics-aware approaches, such as leveraging rain streaks' frequency domain characteristics as inductive bias. However, traditional explicit physical modeling relies on fixed prior assumptions, such as static PSF templates that struggle to capture the diversity of rain degradation, leading to overlooking the multi-scale distribution of rain streaks (Zhu et al. 2020) and optical differences of raindrops.

To address these issues, we refocus our attention on CNN architecture and propose a sequential modeling-based deraining network featuring a novel physics-aware mechanism that relies on multi-scale, data-driven dynamic Point Spread Function (PSF) (Rossmann 1969) prediction. Based on this mechanism, we propose SD-PSFNet, constructing multiple innovative PSF-related modules integrated into a powerful multi-stage architecture, inspired by successful progressive restoration paradigms like MPRNet (Zamir et al. 2021). Additionally, we implement an adaptive gating fusion mechanism to achieve sequential modeling, facilitating the transmission and optimization of cross-stage information within the multi-stage architecture.

Experiments show SD-PSFNet achieves state-of-the-art deraining performance across public datasets while maintaining reasonable computational efficiency. Compared to the baseline MPRNet architecture, our novel physics-aware design delivers a significant 5.04dB PSNR improvement (13.5% gain) on RealRain-1k-L, with particular effective-

*These authors contributed equally.
†Corresponding Author

ness in complex scenes and dense rain conditions. The key contributions are:

- **A Novel PSF-Integrated Deraining Approach**: Multiple integrated PSF mechanism modules dynamically simulate raindrops' optical effects, enhancing representation of complex rain patterns.
- **Sequential Model Design**: Multi-level gated fusion strategies enable full transmission of physical prior information across multiple stages, repeatedly and dynamically refining the estimation of image degradation.
- **Extensive Experimental Evaluation**: Multiple benchmark evaluations, including comparative tests and ablation studies, confirm our method's superior performance.

Related Work

Image Deraining

Image deraining aims to recover rain-free images from rainy ones, representing a specific instance of reversing image degradation. Traditional methods enhanced interpretability by modeling rain's optical properties through image decomposition (Kang, Lin, and Fu 2011) and frequency domain analysis (Zheng et al. 2013). However, these approaches struggle with complex scenes, causing over-smoothing or residual rain streaks.

Deep learning has emerged as the primary approach for image deraining, evolving from CNN-based local feature extraction to global modeling and generative techniques. Early CNNs like DANet (Jiang et al. 2022) and DerainRLNet (Chen and Li 2021) processed rain hierarchically but suffered from limited receptive fields. Later, multi-scale networks such as RESCAN (Li et al. 2018) and DID-MDN (Zhang and Patel 2018) along with physically-guided networks like KGCNN (Wang et al. 2020b) enhanced adaptability through multi-resolution feature fusion and motion blur modeling. Restormer (Zamir et al. 2022) applied Transformer (Vaswani et al. 2017) to low-level vision, achieving linear-complexity long-range dependency modeling via cross-channel self-attention. Generative approaches including Attentive GAN (Qian et al. 2018) produced realistic rain-free images through adversarial training and attention mechanisms, though facing mode collapse issues. Semi-supervised and unsupervised methods such as UD-GAN (Jin et al. 2019) and DerainCycleGAN (Wei et al. 2021) addressed the scarcity of real rain annotations by leveraging unpaired data, while DCDGAN (Chen et al. 2022b) utilized diffusion models for enhanced realism. However, these methods still lack multi-scale physics-aware modeling and struggle to balance computational efficiency with deraining performance.

Physics-Aware Image Restoration

Physics-aware approaches are widely used in image restoration, integrating optical imaging's physical mechanisms into models (Xu et al. 2023). Fusion strategies combining explicit physical modeling with data-driven deep learning approaches have become mainstream methods. For example, ReDNet (Wu et al. 2025) constructed an optically-guided Transformer network that precisely separates stains from background information. In applications focusing on rainy scenes, the ReDT-Det (Chen et al. 2025a) network employs Illumination Fusion Differential Transformer Blocks (IFDTB) to suppress noise and restore details in dark regions, significantly enhancing vehicle detection capabilities in nighttime rainy and foggy conditions. Similarly, researchers have proposed a Dual-Scale Prior (DSP) (Yuan, Meng, and Bai 2025) model that combines physical priors such as non-local self-similarity with the powerful representational capabilities of pre-trained deep denoising networks, achieving robust removal of degradation factors like rain streaks. However, multi-scale rain streak removal still requires further exploration and enhancement.

Method

This section details SD-PSFNet's implementation, including theoretical validation of PSF and the innovative module and framework design.

Dynamic PSF Mechanism

Image degradation processes can typically be modeled as linear space-invariant systems using Point Spread Function (PSF), but real rainy image degradation is more complex, with traditional models struggling to accurately characterize spatially varying degradation fields $K(x, y)$. We transform PSF representation from static presets to data-driven learnable forms. Specifically, we propose that any spatially variable degradation function $K(x, y)$ can be approximated through content-adaptive linear combinations of a learnable degradation pattern dictionary $\{k_j\}_{j=1}^{K_c}$:

$$K(x,y) \approx \sum_{j=1}^{K_c} w_j(x,y) \cdot k_j$$

where k_j represents learnable 2D convolution kernels capturing basic degradation patterns, collectively forming a dictionary representing various rain streak degradation types. Since rain degradation is dominated by finite dimensions like direction and density, a comprehensive pattern dictionary can effectively cover its variation range. In our implementation, neural networks achieve adaptive mapping from image features to combination weights with added physical constraints.

Dynamic Physics-Aware Modules

Point Spread Function (PSF) characterizes imaging systems' response to point light sources, critical for understanding image degradation. Image deraining essentially reverses complex, spatially varying degradation processes. Unlike traditional approaches using static PSFs, SD-PSFNet dynamically adapts to each image's unique degradation patterns, effectively overcoming conventional model limitations.

Multi-Scale PSF Head synthesizes degradation information from multi-level encoder features, dynamically performing adaptive PSF modeling for various degradation patterns across different image regions, overcoming the limitations of traditional methods with single-scale or predefined degradation models.

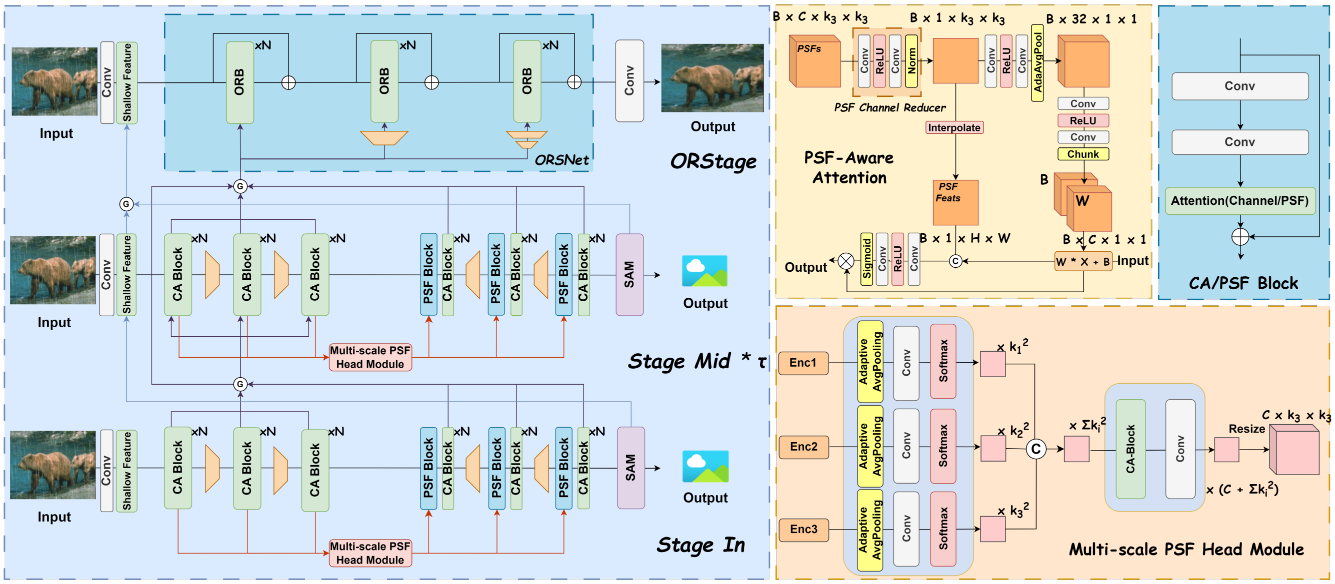

Figure 1: Overview of the SD-PSFNet. The left panel shows the serialized overall restoration framework of SD-PSFNet, including three stages: input Stage In, multiple Stage Mid, and final restoration output ORStage. The right panels detail our key components: PSF Channel Reducer, PSF-Aware Attention, CA/PSF Block, and Multi-Scale PSF Head Module. The network integrates multi-scale dynamic PSF mechanisms for modeling, effectively handling multi-scale rain removal. Input images undergo serialized restoration through three stages with specialized PSF-aware processing.

For a single-scale PSF prediction head, we process encoder features $F_i \in \mathbb{R}^{B\times C_i\times H_i\times W_i}$ to generate the PSF representation:

$$P_i^{raw} = \mathcal{W}_i(\mathcal{AP}(F_i)) \in \mathbb{R}^{B\times k_i^2\times 1\times 1} \quad (1)$$

Here, $\mathcal{AP}$ is adaptive average pooling, B is batch size, and $\mathcal{W}_i$ is a 1×1 convolutional layer mapping C_i-dimensional features to k_i^2 dimensions, functionally equivalent to a multilayer perceptron (MLP). Subsequently, Softmax is applied along the second dimension (kernel dimension) to normalize P_i^{raw} into a valid probability distribution, yielding the flattened weight vector $P_i^{\text{flat}} \in \mathbb{R}^{B\times k_i^2\times 1\times 1}$.

In the multi-scale scenario, we employ multiple prediction heads with specific sizes (3×3, 5×5, and 7×7) to extract complementary degradation information from encoder features $\{F_1, F_2, ..., F_n\}$. These designed heads focus on high-frequency detail degradation, balanced local and global information, and macroscopic low-frequency patterns respectively. The flattened PSFs are concatenated along the channel dimension to form P_{concat}^{flat}. A channel attention block (CAB) (Zhang et al. 2018) then adaptively fuses these multi-scale PSF representations into P_{att}^{flat}. This fused representation is processed by a 1×1 convolutional layer that projects it into $P_{refined}^{flat} \in \mathbb{R}^{B\times(K_c\cdot K\cdot K)\times 1\times 1}$, where K_c is the number of channels and K is the spatial size of the output PSF. Finally, $P_{refined}^{flat}$ is reshaped into the multi-channel PSF representation $P_{multi} \in \mathbb{R}^{B\times K_c\times K\times K}$.

To ensure each channel of P_{multi} forms a valid probability distribution kernel, preventing unrealistic amplification or attenuation of background pixels and simulating energy conservation, we enforce spatial normalization:

$$\sum_{x=1}^{K}\sum_{y=1}^{K} P_{multi}(b, c, x, y) = 1, \quad \forall b \in \{1, ..., B\}, \forall c \in \{1, ..., K_c\} \quad (2)$$

Each channel of P_{multi} represents a principal component in the degradation space. Through channel attention fusion, the model learns to integrate complementary multi-scale information from individual heads, collectively forming a complete representation of the degradation patterns.

PSF Block (PSFB) guides image deraining using PSF features from the **Multi-Scale PSF Head**. It adaptively adjusts input features through dynamic modulation and incorporates physical degradation models, outperforming traditional attention methods in restoring spatially varying defects. Its refinement process is driven by the **PSF Channel Reducer** and **PSF-Aware Attention** components.

PSF Channel Reducer enables information compression by transforming multi-channel PSF from the Multi-Scale PSF Head into a single-channel representation. This module preserves key degradation information through adaptive channel attention. When output channels are fewer than the input, it applies direct mapping using 1×1 convolution. Conversely, it first compresses to $\frac{1}{4}$ of the original channels before mapping to the target count. The compressed PSF then undergoes spatial normalization to ensure its elements sum to 1. This approach maintains the physical properties of the PSF while reducing computational complexity.

PSF-Aware Attention adaptively adjusts feature representations using estimated PSF degradation information, achieving physics-guided feature enhancement. This module integrates PSF encoding with input features through a dual-path mechanism of channel modulation and spatial attention. As shown in Figure 1, for multi-channel PSF ($K_c > 1$), channels are first compressed via the PSF Channel Reducer. A dedicated PSF Generator extracts features at multiple stages to predict degradation kernels at different scales. These predictions are normalized via softmax and fused through channel attention mechanisms to produce a multi-channel PSF representation. The PSF Encoder, integrated within the attention module, processes this information through convolutional layers and adaptive pooling to extract a compact feature representation P_{feat}. This generates channel modulation parameters γ and β:

$$[\gamma, \beta] = \mathcal{F}_{channel}(P_{feat}) \in \mathbb{R}^{B \times C \times 2} \tag{3}$$

where $\mathcal{F}_{channel}$ is a fully-connected network that maps PSF features to modulation parameters matching the input feature channels. Channel modulation enhances features:

$$x_{mod} = x \odot \gamma + \beta \in \mathbb{R}^{B \times C \times H \times W} \tag{4}$$

where $\odot$ denotes element-wise multiplication, with modulation parameters broadcast across all spatial positions. The spatial attention branch upsamples the single-channel PSF to match the input feature's dimensions and combines it with the modulated features to generate spatial attention weights, highlighting regions affected by different degradation levels. Finally, channel-modulated features are multiplied by these spatial attention weights to produce output features.

Sequential Design and Enhanced Feature Fusion

In traditional cascade networks, shallow detail loss and deep semantic gaps often occur. This paper employs a cross-stage gating mechanism to achieve feature transfer through adaptive weights, expressed as:

$$\begin{aligned}\mathbf{F}^{(t)} = G_\theta\left(\mathbf{F}_{\text{current}}, \mathbf{F}_{\text{prev}}\right) \odot \mathbf{F}_{\text{current}} + \\ \left(1 - G_\theta\left(\mathbf{F}_{\text{current}}, \mathbf{F}_{\text{prev}}\right)\right) \odot \mathbf{F}_{\text{prev}}\end{aligned} \tag{5}$$

Here, the gating weight $G_\theta \in [0, 1]$ is learned from the current feature $\mathbf{F}_{\text{current}}$ and the historical feature $\mathbf{F}_{\text{prev}}$. The gating mechanism dynamically fuses features from different sources through an adaptive weighting process. It concatenates the input features and applies global average pooling to capture their overall characteristics. These pooled features pass through convolutional layers to generate fusion weights, determining the contribution of each input feature to the final output, effectively emphasizing more informative features while suppressing less useful ones.

This mechanism operates at three key points: at stage inputs where adaptive gating fuses the current input with previous stage features after shallow feature extraction, at encoder levels through integration of current outputs with historical features, and via enhanced Cross-Stage Feature Fusion (CSFF). This improved CSFF efficiently routes information between network stages through dual adaptive gating units. Here, τ represents the number of middle stages, I_{in} is the degraded input, and $\hat{I}$ is the restored output. Each stage outputs an intermediate restoration result I_i, cross-scale features O_i via CSFF, and feature mappings H_i. The feature extraction module incorporates CSFF through:

$$\mathcal{F}_{\text{shallow}}(I_{\text{in}}, H) = \mathcal{F}_{\text{gate}}(\mathcal{F}_{\text{conv+CAB}}(I_{\text{in}}), H) \tag{6}$$

In the CSFF process, encoder and decoder features from the current stage first pass through an initial gating unit to determine their relative importance. After convolution, these processed features are fed into a second gating unit where they're adaptively fused with features from the previous stage. This fusion undergoes another convolution to generate output features for guiding the next stage. This dual-gating mechanism provides precise control over cross-stage information flow, preserving both low-level details and high-level semantic information throughout the network.

This approach ensures that texture details and semantic information from earlier stages contribute to the refinement in subsequent stages, effectively addressing attenuation and gaps in feature transfer within multi-stage networks.

In the implementation within the original-resolution subnetwork (ORSNet) (Zamir et al. 2021), unlike the original method, the SD-PSFNet design incorporates multi-scale information at specific processing depths, rather than just at the input and output boundaries. Each original resolution block (ORB) receives enriched features from both the main path and transformed side information, enabling fine-grained texture reconstruction at the original resolution.

This approach ensures that texture details and semantic information from earlier stages contribute to the refinement in subsequent stages through gating weights, effectively addressing attenuation and gaps in feature transfer within multi-stage networks.

Overview of SD-PSFNet

SD-PSFNet employs a multi-stage sequential restoration strategy inspired by sequential modeling approaches like LSTM (Hochreiter and Schmidhuber 1997) and multi-stage image restoration models such as MPRNet. Stage-wise image restoration, as a common approach, allows models to repeatedly re-evaluate and refine their estimates of the degradation process while being conducive to module embedding. Notably, although we use designs similar to MPRNet and related module references, such as Supervised Attention Module (SAM) and ORSNet, our approach differs from MPRNet's spatial division strategy (e.g., partitioning images into quarters, half-size, etc.). Instead, we prefer to view SD-PSFNet as a serialized restoration strategy where each stage processes the complete image and dynamically predicts and utilizes PSF features at different resolutions through the multi-scale downsampling structure within UNet.

Our architecture consists of Stage In, Stage Mid, and Original Resolution Stage (ORStage) with enhanced physical information transmission between stages. The overall network is expressed as:

$$\hat{I} = \mathcal{F}_{\text{ORS}}(I_{\text{in}}, H_{\text{mid}}, O_{\text{mid}}) \tag{7}$$

Datasets		Rain100L		Rain100H		Realrain-1k-L		Realrain-1k-H	
Method	Params(M)	PSNR(dB)	SSIM	PSNR(dB)	SSIM	PSNR(dB)	SSIM	PSNR(dB)	SSIM
CNN-based Methods									
DerainNet	0.75	27.03	0.8841	14.92	0.5923	27.09	0.925	22.88	0.8886
DDN	0.06	32.38	0.9259	24.64	0.849	31.18	0.9172	29.17	0.8783
PreNet	0.28	37.48	0.979	29.46	0.899	33.01	0.944	30.88	0.9108
SPANet	0.28	35.33	0.969	25.11	0.827	30.43	0.947	25.76	0.9095
MPRNet	3.64	36.4	0.965	30.41	0.89	36.29	0.972	34.74	0.964
NAFNet	17.11	37	0.978	29.66	0.9	38.8	0.986	36.11	0.976
HINet	88.67	37.28	0.97	30.65	0.894	$41.98^{\dagger}$	$0.9869^{\dagger}$	$40.82^{\dagger}$	$0.983^{\dagger}$
M3SNet	16.70	40.04	0.985	30.64	0.982	41.55	0.9852	40.01	0.979
Ours	**9.63**	**41.47**†	**0.9896**†	**33.12***	**0.9371***	**42.28***	**0.9872***	**41.08***	**0.9838***
Transformer-based Methods									
Restormer	26.10	38.99	0.978	31.46	0.904	40.9	0.9849	39.57	0.9812
PromptIR	32.11	38.34	0.983	28.69	0.877	36.99	0.973	33.61	0.953
DRSFormer	33.66	41.32	0.9887	32.07	0.9316	41.52	0.9812	40.21	0.9824
NeRD-Rain-S	10.53	42^{*}	0.99^{*}	$32.86^{\dagger}$	$0.932^{\dagger}$	38.64	0.979	36.69	0.97

Table 1: Performance comparison (PSNR/SSIM) of deraining methods on synthetic and real-world datasets. $*$ and † represent the methods ranked first and second in performance metrics on the corresponding dataset. The suffixes L and H represent "Light" and "Heavy" rainfall.

Intermediate feature representation H_{i} and cross-stage features O_{i} are generated through:

$$\{I_1, H_1, O_1\} = \mathcal{F}_{\text{StageIn}}(I_{\text{in}}) \tag{8}$$

$$\{I_i, H_i, O_i\} = \mathcal{F}_{\text{StageMid}_i}(I_{\text{in}}, H_{i-1}, O_{i-1}) \tag{9}$$

where i ranges from 2 to τ+1, τ represents the number of middle stages, I_{in} is the degraded input, and $\hat{I}$ is the restored output. Each stage employs an encoder-decoder structure with PSF estimation based on encoder-extracted multi-scale features, upgrading traditional Channel Attention Blocks (CAB) to PSF Blocks (PSFB):

$$X_{\text{out}}^{\text{PSFB}} = X_{\text{in}} + \mathcal{F}_{\text{PSFA}}(\mathcal{F}_{\text{body}}(X_{\text{in}}), \text{PSF}) \tag{10}$$

Where $\mathcal{F}_{\text{body}}$ consists of two convolutional layers with activation functions, and $\mathcal{F}_{\text{PSFA}}$ represents the PSF-Aware Attention that adapts the decoding process based on degradation characteristics. SD-PSFNet includes components from MPRNet (Zamir et al. 2021), including the SAM, which won't be elaborated on here.

Loss Function

The model employs a hybrid loss function that jointly optimizes PSF estimation and physics-guided restoration:

$$\mathcal{L}_{total} = \sum_{s=1}^{\tau+1}(\mathcal{L}_{char}(\mathbf{I}_s, \mathbf{T}) + \alpha_1 \mathcal{L}_{edge}(\mathbf{I}_s, \mathbf{T}) + \alpha_2 \mathcal{L}_{freq}(\mathbf{I}_s, \mathbf{T})) \tag{11}$$

The Charbonnier loss ($\mathcal{L}_{char}$) provides pixel-wise supervision, edge-aware loss ($\mathcal{L}_{edge}$) preserves high-frequency details, and frequency domain loss ($\mathcal{L}_{freq}$) aligns with PSF characteristics. With $\alpha_1 = 0.05$, $\alpha_2 = 0.01$, and $\mathbf{I}_s$ representing stage outputs against ground truth $\mathbf{T}$, this approach maintains pixel accuracy while respecting physical degradation constraints for effective PSF-guided restoration.

Experiments

Datasets and Benchmarks

Four benchmark datasets are introduced for the main comparative experiments, including synthetic datasets: Rain100L (Yang et al. 2019) and Rain100H (Yang et al. 2019), and real-world datasets: RealRain-1K-L (Li et al. 2022), RealRain-1K-H (Li et al. 2022). What's more, SPA-data (Wang et al. 2019), Rain13K (Zamir et al. 2021), and Rain200L (Yang et al. 2017) which are only incorporated in the Generalization Results section. To evaluate the proposed method, several state-of-the-art methods are selected. We compare our method with CNN-based methods (DerainNet (Fu et al. 2017a), DDN (Fu et al. 2017b), PreNet (Ren et al. 2019), SPANet (Wang et al. 2019), MPRNet (Zamir et al. 2021), NAFNet (Chen et al. 2022a), HINet (Chen et al. 2021), M3SNet(Gao et al. 2023)), Transformer-based methods (Restormer (Zamir et al. 2022), PromptIR (Potlapalli et al. 2023), DRSFormer (Chen et al. 2023), NeRD-Rain-S (Chen, Pan, and Dong 2024)). The method employs supervised training, thus excluding comparisons with unsupervised GAN-based deraining approaches.

Implementation Details and Metrics

Our model is implemented using the PyTorch framework with τ=3 stages and PSF channel K_c=40 for main results and trained for 2000 epochs in Table 1. We employ AdamW optimizer with an initial learning rate of 1e-4, incorporating a 3-epoch linear warmup followed by cosine annealing to 1e-6. Mixed-precision training (FP16) accelerates computation while gradient clipping (max norm 2.0) stabilizes convergence. During training, rain-free images and ground truth are aligned and cropped to 128×128 patches, normalized using ImageNet mean-variance statistics, and augmented with random flipping and geometric transformations. For test-

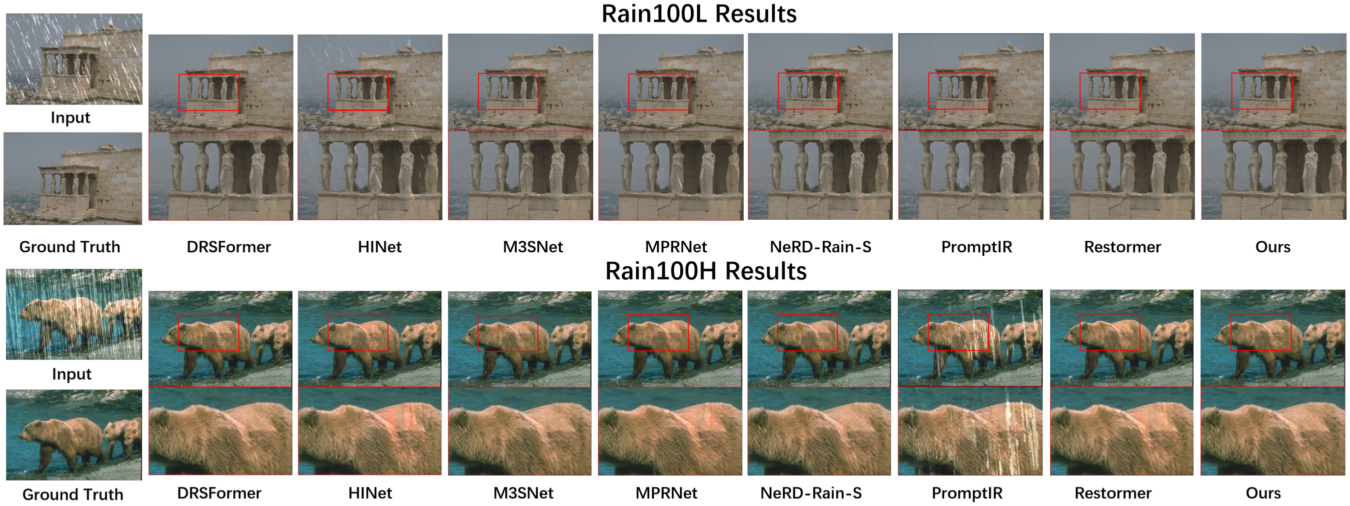

Figure 2: Qualitative deraining performance comparisons on Rain100L (Yang et al. 2019)and Rain100H (Yang et al. 2019)datasets. Our SD-PSFNet achieves competitive visual results comparable to the other SOTA method.

Train ↓ Test →	Realrain-1k-L		Realrain-1k-H		Rain100L		Rain100H		SPA-Data	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
RealRain-1k-L	**42.28**	**0.9872**	40.70	0.9829	26.93	0.8377	14.34	0.3907	**34.89**	**0.9617**
RealRain-1k-H	41.14	0.9855	**41.08**	**0.9838**	23.61	0.7949	16.79	0.4719	33.23	0.9552
Rain100L	27.54	0.8831	23.82	0.7963	**41.47**	**0.9896**	18.79	0.5721	32.97	0.9206
Rain100H	26.98	0.8733	24.12	0.8106	38.07	0.9797	**33.12**	**0.9371**	31.93	0.9121

Table 2: Cross-dataset evaluation results on synthetic and real rain datasets. Each row represents a model trained on a specific dataset, while columns show testing performance on different datasets.

ing, we only apply minimal reflection padding to maintain whole-image inference. For quality evaluation of derained images, we employ PSNR and SSIM metrics. During training and validation, we calculate PSNR only on the Y channel (luminance component in YCrCb), while during testing, we compute both PSNR and SSIM in RGB space. All experiments are conducted from scratch on a machine with one NVIDIA GeForce RTX 4090 GPU.

Main Deraining Results

As shown in Table 1, our proposed CNN-based method achieves impressive performance across most evaluation metrics and datasets. For synthetic benchmarks, we obtain 41.47 dB PSNR/0.9896 SSIM on Rain100L and 33.12 dB PSNR/0.9371 SSIM on Rain100H, demonstrating that our CNN architecture can match or even surpass many Transformer-based methods like Restormer and PromptIR. Although NeRD-Rain-S shows slightly better results on Rain100L, our method demonstrates extreme performance on real-world datasets, achieving 42.28 dB PSNR/0.9872 SSIM on RealRain-1k-L and 41.08 dB PSNR/0.9838 SSIM on RealRain-1k-H. Our CNN-based method bridges the performance gap with state-of-the-art Transformer architectures while maintaining CNN efficiency advantages, handling diverse rain patterns in both synthetic and real-world scenarios. Visualizations are shown in Figures 2 and 3.

Generalization Results

To validate the generalization capability of SD-PSFNet and the significant domain gap between artificial and real rainfall scenarios, as shown in Table 2, models trained on the RealRain-1k dataset demonstrate poor transferability to the Rain100 dataset, and vice versa, with severe performance degradation when crossing domain boundaries. Notably, since the Ground Truth in SPA-Data is derived from video-based deraining methods, our method demonstrates good and more stable performance on SPA-Data. This gap represents a fundamental limitation of data-driven approaches, as models learn dataset-specific features rather than universal deraining principles. Table 3 demonstrates generalization comparison from synthetic to real datasets, showcasing the generalization capability of various deraining methods when applied to real-world rain scenarios without specific fine-tuning. SD-PSFNet achieves the highest PSNR of 26.98dB and maintains consistent superior performance when trained on the Rain100H, demonstrating that our method effectively captures the essential characteristics of rain streaks and shows high generalization capability compared to other models trained on the larger datasets.

Ablation Studies

Ablation studies were conducted to evaluate the effectiveness of each proposed component in SD-PSFNet and reveal

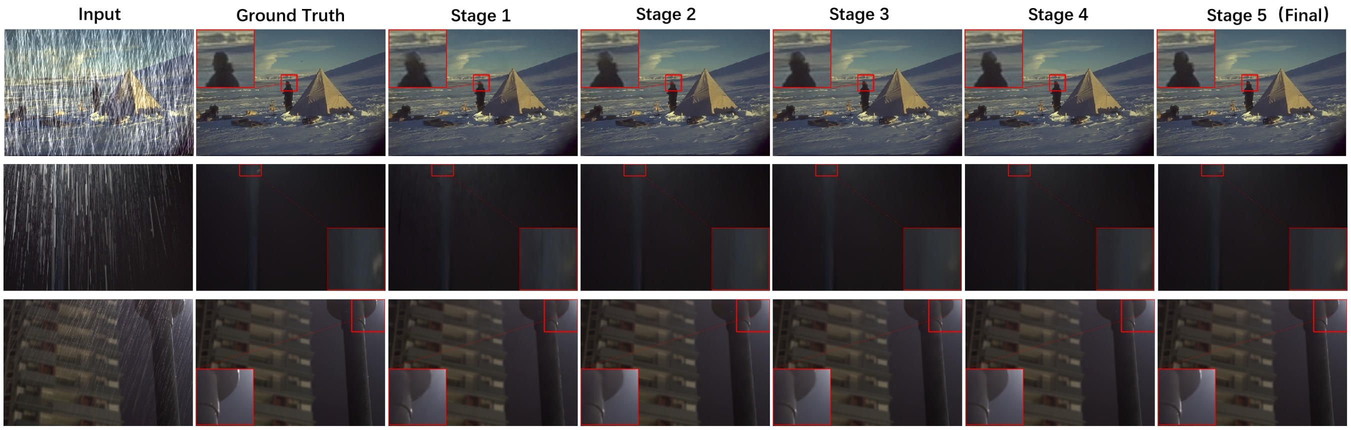

Figure 3: Comparison of restoration results across different stages on three datasets (from top to bottom): Rain100H (Yang et al. 2019) and two RealRain-1k-H (Li et al. 2022) samples. Stage 1 represents Stage In, Stages 2-4 correspond to Stage Mid, and Stage 5 shows the final restoration from ORStage.

Method	Metrics	Training Dataset
	PSNR	
MPRNet	25.55	Rain13K
Restormer	26.59	Rain13K
NeRD-Rain-S	26.67	Rain200L
SD-PSFNet	**26.98**	**Rain100H**

Table 3: Generalization comparison from synthetic to real datasets on the RealRain-1k-L dataset.

Model Configuration	RealRain-1k-L	
	PSNR	SSIM
MPRNet (Baseline)	37.24dB	0.9754
- Patch Split	37.20dB	0.9755
+ Gate	38.65dB	0.9773
+ H Inter-stage Updates	40.41dB	0.9792
+ Enhanced CSFF	41.41dB	0.9850
+ 1-channel PSF	41.63dB	0.9855
+ 40-channel PSF (Ours)	42.28dB	0.9872

Table 4: Ablation study results on RealRain-1k-L dataset. (components added incrementally)

the size of the model in different numbers of stages.

Incremental analysis on the RealRain-1k-L dataset shows substantial performance gains with each module in Table 4. The baseline, MPRNet with patch-splitting, confirms its advantage over the variant without it. Ablation results indicate that removing patch-splitting leads to decreased performance. The analysis shows significant gains with each module, with patch-splitting removal decreasing performance. The gate mechanism improves performance by 1.45dB, while hierarchical inter-stage updates add 1.76dB, highlighting the value of progressive feature refinement. The enhanced CSFF module contributes an additional 1.0dB, and the 1-channel PSF offers a modest 0.22dB gain. The final architecture achieves optimal performance of 42.28dB PSNR and 0.9872 SSIM, reflecting a substantial 5.04dB improvement over the baseline. These results highlight the meaningful contributions of each proposed component.

Parameter	PSNR	SSIM	Params (M)	MACs (G)
$\tau = 0$	40.81dB	0.9832	3.64	135.92
$\tau = 1$	40.86dB	0.9835	5.64	170.67
$\tau = 2$	40.97dB	0.9843	7.63	206.97
$\tau = 3$	41.54dB	0.9862	9.63	244.83

Table 5: Ablation study on different values of parameter τ.

Parameter τ examination over 1000 epochs reveals consistent performance improvements as τ increases from 0 to 3 in Tabel 5. Without Stage Mid (τ=0), the model achieves 40.81dB PSNR and 0.9832 SSIM with 3.64M parameters and 135.92G MACs. Increasing to τ=3 delivers optimal performance (41.54dB PSNR, 0.9862 SSIM), demonstrating a significant 0.57dB improvement over the baseline configuration, though at the cost of increased model size (9.63M parameters) and computational complexity (244.83G MACs). Our CNN-based architecture efficiently manages these parameter increases while outperforming similar-sized alternatives. This analysis confirms the importance of multiple Stage Mid phases for enhancing restoration capability while illustrating the associated computational trade-offs.

Conclusion

This paper proposes a sequential model incorporating dynamic physics-aware Point Spread Function (PSF) mechanisms. The approach integrates innovative multi-scale PSF modules into a multi-stage deraining architecture with multi-level gated fusion mechanisms, enhancing degradation feature modeling while suppressing detail loss through cross-stage feature memory transfer. Experimental results show that SD-PSFNet achieves a better balance between performance and computational efficiency compared to existing deraining methods, providing new insights for physics-aware image restoration.

Acknowledgments

This work was supported by the Zhejiang Provincial Natural Science Foundation of China (Grant NO. LY23F020025), the Science and Technology Program of Huzhou (Grant NO. 2024GZ09), the Zhejiang Province Leading Geese Plan (Grant NO. 2025C02025), and the National Natural Science-Foundation of China (Grant NO. 62576292).

References

Chen, C.; and Li, H. 2021. Robust representation learning with feedback for single image deraining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7742–7751.

Chen, L.; Chu, X.; Zhang, X.; and Sun, J. 2022a. Simple baselines for image restoration. In *European conference on computer vision*, 17–33. Springer.

Chen, L.; Deng, H.; Liu, G.; Law, R.; Li, D.; Wu, E. Q.; and Zhu, L. 2025a. Retinex-guided illumination recovery and progressive feature adaptation for real-world nighttime UAV-based vehicle detection. *Expert Systems with Applications*, 129476.

Chen, L.; Lu, X.; Zhang, J.; Chu, X.; and Chen, C. 2021. Hinet: Half instance normalization network for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 182–192.

Chen, X.; Li, H.; Li, M.; and Pan, J. 2023. Learning a sparse transformer network for effective image deraining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5896–5905.

Chen, X.; Pan, J.; and Dong, J. 2024. Bidirectional multi-scale implicit neural representations for image deraining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 25627–25636.

Chen, X.; Pan, J.; Dong, J.; and Tang, J. 2025b. Towards unified deep image deraining: A survey and a new benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Chen, X.; Pan, J.; Jiang, K.; Li, Y.; Huang, Y.; Kong, C.; Dai, L.; and Fan, Z. 2022b. Unpaired deep image deraining using dual contrastive learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2017–2026.

Dey, R.; and Bhattacharjee, D. 2020. Single image deraining using GAN for accurate video surveillance. In *Intelligence Enabled Research: DoSIER 2019*, 7–11. Springer.

Ding, Y.; Li, M.; Yan, T.; Zhang, F.; Liu, Y.; and Lau, R. W. 2021. Rain streak removal from light field images. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(2): 467–482.

Fu, X.; Huang, J.; Ding, X.; Liao, Y.; and Paisley, J. 2017a. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Transactions on Image Processing*, 26(6): 2944–2956.

Fu, X.; Huang, J.; Zeng, D.; Huang, Y.; Ding, X.; and Paisley, J. 2017b. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3855–3863.

Gao, H.; Yang, J.; Zhang, Y.; Wang, N.; Yang, J.; and Dang, D. 2023. A mountain-shaped single-stage network for accurate image restoration. *arXiv preprint arXiv:2305.05146*.

Hochreiter, S.; and Schmidhuber, J. 1997. Long short-term memory. *Neural computation*, 9(8): 1735–1780.

Jiang, K.; Wang, Z.; Wang, Z.; Yi, P.; Jiang, J.; Xiao, J.; and Lin, C.-W. 2022. Danet: Image deraining via dynamic association learning. In *IJCAI*, 980–986.

Jin, X.; Chen, Z.; Lin, J.; Chen, Z.; and Zhou, W. 2019. Unsupervised single image deraining with self-supervised constraints. In *2019 IEEE International Conference on Image Processing (ICIP)*, 2761–2765. IEEE.

Kang, L.-W.; Lin, C.-W.; and Fu, Y.-H. 2011. Automatic single-image-based rain streaks removal via image decomposition. *IEEE transactions on image processing*, 21(4): 1742–1755.

Li, W.; Zhang, Q.; Zhang, J.; Huang, Z.; Tian, X.; and Tao, D. 2022. Toward real-world single image deraining: A new benchmark and beyond. *arXiv preprint arXiv:2206.05514*.

Li, X.; Wu, J.; Lin, Z.; Liu, H.; and Zha, H. 2018. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *Proceedings of the European conference on computer vision (ECCV)*, 254–269.

Potlapalli, V.; Zamir, S. W.; Khan, S. H.; and Shahbaz Khan, F. 2023. Promptir: Prompting for all-in-one image restoration. *Advances in Neural Information Processing Systems*, 36: 71275–71293.

Qian, R.; Tan, R. T.; Yang, W.; Su, J.; and Liu, J. 2018. Attentive generative adversarial network for raindrop removal from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2482–2491.

Ren, D.; Zuo, W.; Hu, Q.; Zhu, P.; and Meng, D. 2019. Progressive image deraining networks: A better and simpler baseline. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3937–3946.

Rossmann, K. 1969. Point spread-function, line spread-function, and modulation transfer function: tools for the study of imaging systems. *Radiology*, 93(2): 257–272.

Shao, M.; Li, L.; Wang, H.; and Meng, D. 2021. Selective generative adversarial network for raindrop removal from a single image. *Neurocomputing*, 426: 265–273.

Su, Z.; Zhang, Y.; Shi, J.; and Zhang, X.-P. 2023. A survey of single image rain removal based on deep learning. *ACM Computing Surveys*, 56(4): 1–35.

Sun, H.; Ang, M. H.; and Rus, D. 2019. A convolutional network for joint deraining and dehazing from a single image for autonomous driving in rain. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 962–969. IEEE.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Wang, H.; Xie, Q.; Wu, Y.; Zhao, Q.; and Meng, D. 2020a. Single image rain streaks removal: a review and an exploration. *International Journal of Machine Learning and Cybernetics*, 11(4): 853–872.

Wang, K.; Wang, T.; Qu, J.; Jiang, H.; Li, Q.; and Chang, L. 2022a. An end-to-end cascaded image deraining and object detection neural network. *IEEE Robotics and Automation Letters*, 7(4): 9541–9548.

Wang, T.; Yang, X.; Xu, K.; Chen, S.; Zhang, Q.; and Lau, R. W. 2019. Spatial attentive single-image deraining with a high quality real rain dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 12270–12279.

Wang, Y.-T.; Zhao, X.-L.; Jiang, T.-X.; Deng, L.-J.; Chang, Y.; and Huang, T.-Z. 2020b. Rain streaks removal for single image via kernel-guided convolutional neural network. *IEEE Transactions on Neural Networks and Learning Systems*, 32(8): 3664–3676.

Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; and Li, H. 2022b. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 17683–17693.

Wei, Y.; Zhang, Z.; Wang, Y.; Xu, M.; Yang, Y.; Yan, S.; and Wang, M. 2021. Deraincyclegan: Rain attentive cyclegan for single image deraining and rainmaking. *IEEE Transactions on Image Processing*, 30: 4788–4801.

Wu, Z.; Zhang, J.; Fu, Y.; Zhang, Y.; and Yuan, X. 2025. Restoration of Images Taken Through a Dirty Window Using Optics-guided Transformer. *IEEE Transactions on Image Processing*.

Xu, Q.; Shi, Y.; Bamber, J.; Tuo, Y.; Ludwig, R.; and Zhu, X. X. 2023. Physics-aware machine learning revolutionizes scientific paradigm for machine learning and process-based hydrology. *arXiv preprint arXiv:2310.05227*.

Yang, W.; Tan, R. T.; Feng, J.; Guo, Z.; Yan, S.; and Liu, J. 2019. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE transactions on pattern analysis and machine intelligence*, 42(6): 1377–1393.

Yang, W.; Tan, R. T.; Feng, J.; Liu, J.; Guo, Z.; and Yan, S. 2017. Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1357–1366.

Yang, W.; Tan, R. T.; Wang, S.; Fang, Y.; and Liu, J. 2020. Single image deraining: From model-based to data-driven and beyond. *IEEE Transactions on pattern analysis and machine intelligence*, 43(11): 4059–4077.

Yuan, W.; Meng, C.; and Bai, X. 2025. Image restoration driven by dual-scale prior. *Neural Networks*, 108138.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M.-H. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5728–5739.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; Yang, M.-H.; and Shao, L. 2021. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 14821–14831.

Zhang, H.; and Patel, V. M. 2018. Density-aware single image de-raining using a multi-stream dense network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 695–704.

Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; and Fu, Y. 2018. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, 286–301.

Zheng, X.; Liao, Y.; Guo, W.; Fu, X.; and Ding, X. 2013. Single-image-based rain and snow removal using multi-guided filter. In *International conference on neural information processing*, 258–265. Springer.

Zhu, H.; Wang, C.; Zhang, Y.; Su, Z.; and Zhao, G. 2020. Physical model guided deep image deraining. *arXiv preprint arXiv:2003.13242*.