PRISM: PROGRESSIVE RAIN REMOVAL WITH INTEGRATED STATE-SPACE MODELING

Pengze Xue[1], *Shanwen Wang*[1], *Fei Zhou*[2], *Yan Cui*[3], *Xin Sun*[1*]

[1]Faculty of Data Science, City University of Macau, SAR Macao, China
[2]College of Oceanography and Space Informatics, China University of Petroleum (East China), China
[3]Zhuhai 4Dage Network Technology, China

ABSTRACT

Image deraining is an essential vision technique that removes rain streaks and water droplets, enhancing clarity for critical vision tasks like autonomous driving. However, current single-scale models struggle to the fine-grained recovery with global consistency. To address this challenge, we propose Progressive Rain removal with Integrated State-space Modeling (PRISM), a progressive three-stage framework: Coarse Extraction Network (CENet), Frequency Fusion Network (SFNet) and Refine Network (RNet). Specifically, CENet and SFNet utilize a novel Hybrid Attention UNet (HA-UNet) for multi-scale feature aggregation by combining channel attention with windowed spatial transformers. Moreover, we propose Hybrid Domain Mamba (HDMamba) for SFNet to jointly model spatial semantics and wavelet domain characteristics. Finally, RNet recovers the fine-grained structures via original-resolution subnetwork. Our model learns high-frequency rain characteristics while preserving structural details and maintaining global context, leading to improved image quality. Our method achieves competitive results on multiple datasets against the recent deraining methods.

Index Terms— Multi-Stage Deraining, Image Restoration, State-Space Models, Frequency-Domain Aware

1. INTRODUCTION

Image deraining removes or reduces rain streaks and water droplets from an image to improve the clarity and visibility [1, 2, 3]. Adverse weather often produces rain streaks and visibility loss that impair perception. Consequently, robust image deraining is crucial for autonomous driving and surveillance. This task is challenging due to the intricate diversity of rain patterns and difficulty on global consistency with fine-grained recovery. To address these challenges, recent methods mainly follow two directions, i.e., Convolutional Neural Network (CNN) and Transformer architectures. Representative CNN architectures like UNet employ symmetric encoder-decoder structures with skip connections, enabling multi-scale feature fusion and stable training. While CNN effectively leverages local convolutions for hierarchical feature representation, their limited receptive fields pose challenges for modeling long-range dependencies, particularly on complex rain distributions. This limitation motivates the adoption of global modeling mechanisms such as Transformers in the deraining framework. Transformer methods [4, 5, 6, 7] introduce self-attention to model the global context. Although transformers are good at low-frequency and long-range contexts, they are not sensitive to high-frequency rain cues. Beyond transformer-based global attention, state-space models (SSMs) are increasingly employed in the image deraining task [8, 9, 10, 11]. The structured state-space model [12] enables long-range dependency modeling with linear complexity. For instance, Mamba's selective scanning [13] builds correlations among image patches and propagates information across distant regions. However, Mamba models are more responsive to high-frequency information than transformers, they still lack a dedicated frequency-aware component [6, 14], as illustrated in Fig. 1. To effectively addressing this limitation, our HDMamba module introduces a semantic reordering branch (a) for strong global modeling and a dedicated wavelet-domain reordering branch (b) to provide frequency-aware.

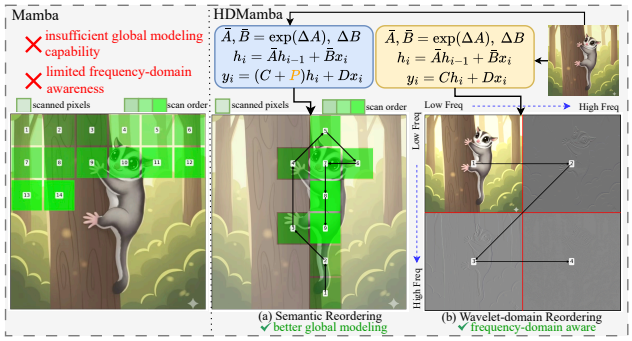

Fig. 1. HDMamba provides stronger global modeling and incorporates frequency-domain awareness.

In this work, we propose Progressive Rain removal with Integrated State-space Modeling (PRISM), which balances the global consistency with fine-grained recovery and frequency-domain fidelity. The PRISM mainly consists of three key stages, i.e., CENet, SFNet and RNet. Specifically, CENet performs preliminary rain removal with HA-UNet, which extracts shallow features and supplies coarse deraining information for the next stage. Then SFNet integrates HA-UNet with HDMamba to jointly model spatial semantics and wavelet domain characteristics. Finally, RNet recovers fine structures via original-resolution subnetwork (ORS). Overall, PRISM adopts a multi-stage architecture, which strengthens high-frequency reconstruction while maintaining global coherence. Our main contributions are summarized as follows:

• We introduce PRISM, a progressive multi-stage deraining framework that enables coarse-to-fine restoration. It not only effectively learns high-frequency rain cues, but also captures global context.
• We propose a HA-UNet for the CENet and SFNet stages, which

*This work is supported by the Science and Technology Development Fund, Macao SAR No.0006/2024/RIA1 and National Natural Science Foundation of China No.61971388.

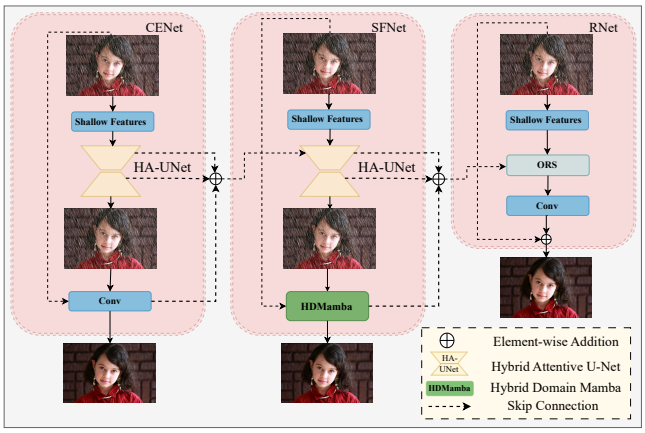

Fig. 2. The overall architecture of our proposed PRISM with CENet, SFNet, and RNet to achieve the rain-free image.

enhances multi-scale feature aggregation by combining Channel Attention (CA) with windowed spatial transformers.

• We propose HDMamba that couples state-space sequence modeling with wavelet-based decomposition, to strengthen frequency-aware modeling with efficient global context.

• Our experiments on widely used benchmarks demonstrate competitive results. At the same time, our method shows good deraining performance across various rain scenarios.

2. METHOD

As shown in Fig. 2, PRISM consists of three key stages: CENet, SFNet, and RNet. CENet and SFNet share a common design with same shallow feature extractor, HA-UNet backbone, and output head. The differences are that CENet uses several convolutional layers for coarse rain-streak removal and SFNet stacks HDMamba modules to fuse wavelet-domain features with global context. And RNet refines the reconstruction with ORS.

We invent CENet, which extracts shallow features and performs coarse-grained deraining, to model the local spatial context and short-range dependencies. SFNet, which is following and guided by CENet, further refines deraining. The shallow extractor and HA-UNet supply robust representations and multi-scale interactions. The proposed HDMamba achieves global semantic modeling and hybrid domain fusion. It combines semantic reordering for long-range dependencies with a wavelet branch for high-frequency enhancement. An adaptive gating module fuses the spatial and frequency cues, to improve detail recovery and suppress artifacts. RNet merges features and completes the fine-grained reconstruction with ORS. As discussed in prior work [15], the ORS is effective for fine-grained refinement at the original resolution.

2.1. Hybrid Attention UNet

The novel HA-UNet is the core module of our CENet and SFNet, as shown in Fig. 3. UNet has demonstrated strong performance on image deraining in many prior works [15, 16, 17, 18]. In this work, we enhance its expressive power by integrating a joint channel and spatial attention mechanism. In particular, the proposed spatial attention adopts a novel Swin-like windowed self-attention module to capture global spatial context [19]. Concretely, HA-UNet employs one basic unit composed of CA, alternating Window Multi-Head Self-Attention (W-MSA) and Shifted Multi-Head Self-Attention (SW-MSA), and one Convolutional Feed-Forward Network (ConvFFN). The specific calculation process of CA is as follows:

$$y = \mathrm{Conv}_{3\times3}\big(\phi(\mathrm{Conv}_{3\times3}(x_{\mathrm{in}}))\big), \quad x_{\mathrm{in}} \in \mathbb{R}^{B\times C\times H\times W}, \tag{1}$$

$$z = \mathrm{GAP}(y), \quad s = \sigma\big(\omega_2\, \delta_R(\omega_1 z)\big), \tag{2}$$

$$x_{\mathrm{ca}} = x_{\mathrm{in}} + y \odot s, \tag{3}$$

where, $x_{\mathrm{in}} \in \mathbb{R}^{B\times C\times H\times W}$ represents the input, B is the batch size, C is the number of channels, H is the image height, and W is the width. The operator $\mathrm{Conv}_{3\times3}$ uses padding of 1 to preserve spatial resolution and $\phi(\cdot)$ denotes the PReLU activation function. Global Average Pooling (GAP) averages over the $H\times W$ spatial dimensions to get $z \in \mathbb{R}^{B\times C}$. The CA signal $s \in \mathbb{R}^{B\times C}$ is broadcast spatially across H and W. The reduction ratio is r, with $\omega_1 \in \mathbb{R}^{\frac{C}{r}\times C}$ and $\omega_2 \in \mathbb{R}^{C\times\frac{C}{r}}$ implemented as 1×1 convolutions or linear layers. Here, σ denotes the Sigmoid activation function, and δ_R denotes the ReLU activation function. In addition, $\odot$ denotes element-wise multiplication.

$$x_{\mathrm{attn}} = \mathrm{WAttn}\big(\mathrm{LN}(\mathrm{CA}(\mathrm{LN}(x_{\mathrm{ca}})))\big), \tag{4}$$

$$x_{\mathrm{convfnn}} = \mathrm{ConvFNN}\big(\mathrm{LN}(x_{\mathrm{attn}})\big). \tag{5}$$

Here, WAttn is specifically implemented by alternately using W-MSA and SW-MSA, as shown in Fig. 3. LN denotes Layer Normalization. The calculation process of the final component ConvFNN in HA-UNet is as follows:

$$U = \delta_G\big(\mathrm{LN}(x_{\mathrm{convfnn}})\,\omega_1 + b_1\big), \tag{6}$$

$$U' = \mathrm{DWConv}(U), \tag{7}$$

$$x_{\mathrm{out}} = U'\,\omega_2 + b_2, \tag{8}$$

where the δ_G employs the GELU activation function. We use DWConv to denote a depthwise convolution to preserve spatial size and hidden channels. The ConvFNN projections use $\omega_1 \in \mathbb{R}^{C\times C_{\mathrm{ff}}}$ and $\omega_2 \in \mathbb{R}^{C_{\mathrm{ff}}\times C}$ with biases b_1, b_2, where C_{ff} denotes the FFN intermediate dimension and ff denotes the feed-forward hidden width. These parameters are not shared with the CA path. The final output of HA-UNet is denoted as y_{out}:

$$y_{\mathrm{out}} = x_{\mathrm{attn}} + x_{\mathrm{out}}. \tag{9}$$

2.2. Hybrid Domain Mamba

Rain streaks in the image mainly expresses as high-frequency bands. However, the traditional Mamba model cannot achieve frequency-aware modeling and global context aggregation. Specifically, the classic Mamba model achieves token interactions via a discretized state-space recurrence: SSMs map an input sequence x_i to outputs y_i through a latent state h_i updated by a stabilized, discretized linear recurrence. The computational process is detailed below:

$$\bar{A},\ \bar{B} = \exp(\Delta A),\ \Delta B, \tag{10}$$

$$h_i = \bar{A}\,h_{i-1} + \bar{B}\,x_i, \tag{11}$$

$$y_i = C\,h_i + D\,x_i, \tag{12}$$

where A and B are continuous parameters, while $\bar{A}$ and $\bar{B}$ are discrete parameters, and Δ is the sampling period, and exp denotes the matrix exponential operation. Specifically, $\bar{A}$ is the state matrix, which governs the system's intrinsic dynamics. Meanwhile, $\bar{B}$ is the input matrix, determining how the current input matrix influences and modifies the system state. In addition, h_i denotes the

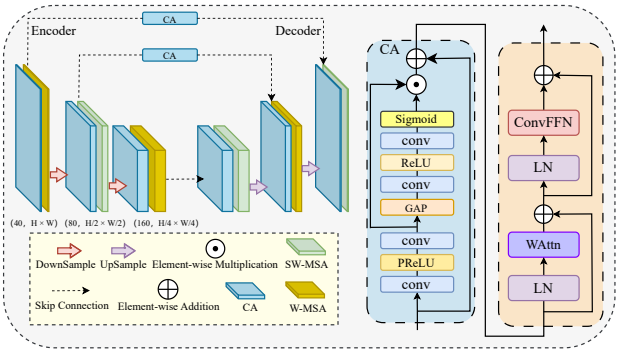

Fig. 3. HA-UNet architecture. A UNet encoder-decoder with hybrid attention and skip connections for feature fusion.

hidden state at the current time step i, and h_{i-1} represents the state from the previous time step. However, standard causal state-space models have limited global perception. Each token depends only on predecessors, therefore, long-range interactions decay even with multi-directional scans.

Accordingly, we propose HDMamba to enhance frequency-aware modeling and global context aggregation for image deraining, as illustrated in Fig.1. This framework employs a semantic reordering Mamba architecture that models spatial-domain semantic reordering. It simultaneously performs wavelet-domain frequency reordering, achieving coordinated cross-domain processing via an adaptive gating fusion mechanism. We follow prior non-causal designs and add a learnable prompt term P to the readout, allowing queries to unscanned positions to enable prompt guided single scan processing [8].

Specifically, in the spatial domain, HDMamba performs semantic reordering. A routing network predicts a semantic class for each spatial position and employs Gumbel-Softmax to realize discrete routing. These features are then rearranged based on semantic affinity. Finally, the prompt guided single scan captures the long-range dependencies within the reordered sequence. This semantic-driven reordering integrates semantically similar regions, improving sequence modeling efficiency and effectiveness.

Moreover, in the wavelet domain, we perform a standard Discrete Wavelet Transform (DWT) to achieve a two-level decomposition. We then model only the top-level subbands, where the four subbands are arranged as a 2×2 quad to form an $H\times W$ canvas and flattened into a one-dimensional sequence in a fixed order for a single scan. The input tensor $I \in \mathbb{R}^{H\times W\times C}$ is decomposed into: one low-frequency approximation map and three oriented high-frequency detail subbands. The detailed computation of DWT proceeds as follows:

$$I_{\mathrm{LL}},\, I_{\mathrm{LH}},\, I_{\mathrm{HL}},\, I_{\mathrm{HH}} = \mathrm{DWT}(I), \tag{13}$$

where each sub-band corresponds to different directional frequency details. The scanned coefficients are inverse-rearranged and mapped back to the spatial domain via Inverse Discrete Wavelet Transform (IDWT) to yield the wavelet-branch output X_w.

The two domains are processed in parallel. The spatial semantic reordering branch and the wavelet frequency reordering branch conduct single scans independently to avoid mutual interference. Finally, adaptive gating fuses the two domains, automatically adjusting their relative contributions according to content, so that their advantages can complement each other.

$$G = \sigma\big([X_s; X_w]\,\omega_g + b_g\big), \quad \omega_g \in \mathbb{R}^{2C\times C}, \;\; b_g \in \mathbb{R}^{C}, \tag{14}$$

$$X_{\text{out}} = G \odot X_s + (1 - G) \odot X_w, \tag{15}$$

where G is a per-channel gate vector of length C with entries in $(0, 1)$. X_s and X_w are the spatial-branch and wavelet-branch features, $[X_s; X_w]$ denotes channel-wise concatenation. ω_g and b_g are learnable parameters.

2.3. Loss Function

We adopt a multi-component loss that combines classical reconstruction terms with a wavelet domain regularization tailored to HDMamba. Following [16], the basic loss includes a Charbonnier loss [15] for robust image reconstruction and an edge loss to preserve sharp structures and fine details. For SFNet, we additionally employ a WaveletLoss to enforce accurate subband reconstruction. Given the prediction X_s at stage $s \in \{1, 2, 3\}$ and ground truth Y, then our global loss function is:

$$\mathcal{L} = \sum_{s=1}^{3} \Big[\mathcal{L}_{\text{char}}(X_s, Y) \;+\; \lambda\, \mathcal{L}_{\text{edge}}(X_s, Y) \;+\; \mu_s\, \mathcal{L}_{\text{wavelet}}(X_s, Y)\Big], \tag{16}$$

where, $\mathcal{L}_{char}$ is Charbonnier loss, $\mathcal{L}_{edge}$ is Edge loss, and $\mathcal{L}_{wavelet}$ is Wavelet-domain loss. In our implementation, the coefficient λ in Eq. 16 balances the Charbonnier and Edge losses and is set to 0.05 as suggested by [15]. For each stage, we set $\mu_s \in (0, 0.05, 0)$. The $\mathcal{L}_{\text{char}}$ is described as follows:

$$\mathcal{L}_{\text{char}}(X_s, Y) = \mathbb{E}\left[\sqrt{\|X_s - Y\|_2^2 + \varepsilon^2}\right], \tag{17}$$

where the constant ε is empirically set to 10^{-3} for all the experiments. The $\mathcal{L}_{edge}$ is described as follows:

$$\mathcal{L}_{\text{edge}}(X_s, Y) = \sqrt{\|\nabla^2(X_s) - \nabla^2(Y)\|_2^2 + \varepsilon^2}, \tag{18}$$

where ∇^2 denotes the Laplacian operator. The $\mathcal{L}_{wavelet}(X_s, Y)$ is described as follows:

$$\begin{aligned}\mathcal{L}_{\text{wavelet}}(X_s, Y) &= \mathcal{L}_1\big(LL(X_s),\, LL(Y)\big) \\ &\quad + \sum_{i=1}^{J} \sum_{b\in\{LH,HL,HH\}} \mathcal{L}_1\big(H_b^{(i)}(X_s),\, H_b^{(i)}(Y)\big),\end{aligned} \tag{19}$$

where, $\mathcal{L}_1$ is the mean of the absolute differences between predictions and targets. LL_x and LL_y denote the level-J low-pass approximation coefficients of the predicted and target images, respectively. $H_b^{(i)}$ denotes the b-th high-frequency subband at level i.

3. EXPERIMENTS

In this section, we train PRISM on mixed training deraining datasets and evaluate its performance against representative deraining baselines. The code is publicly available at `https://github.com/xuepengze/PRISM`.

3.1. Datasets

We follow a mixed training track that integrates classic synthetic deraining datasets. The datasets include Rain800 [26], Rain100L and Rain100H [25], Rain14000 (DDN-Data) [20], and Rain12 [27]. In total, this setup yields 13,712 paired rainy and clean images. Rain800 has 700 training images and 100 testing images with BSD and UCID backgrounds [26]. Rain100L and Rain100H provide

Table 1. Quantitative comparison for image deraining on five benchmark datasets.

Method	Test100		Rain100H		Rain100L		Test2800		Test1200		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DerainNet [20]	22.77	0.810	14.92	0.592	27.03	0.884	24.31	0.861	23.38	0.835	22.48	0.796
UMRL [21]	24.41	0.829	26.01	0.832	29.18	0.923	29.97	0.905	30.55	0.910	28.02	0.880
PReNet [22]	24.81	0.851	26.77	0.858	32.44	0.950	31.75	0.916	31.36	0.911	29.43	0.897
MPRNet [15]	30.27	0.897	30.41	0.890	36.40	0.965	33.47	0.938	32.91	0.916	32.69	0.921
Uformer [23]	29.17	0.880	30.95	0.884	36.34	0.966	33.34	0.936	31.98	0.909	32.36	0.915
Semi-SwinDerain [6]	28.54	0.893	28.79	0.861	34.71	0.957	32.68	0.932	30.96	0.909	31.14	0.910
DAWN [24]	29.86	0.902	29.89	0.889	35.97	0.963	-	-	32.76	0.919	32.12	0.918
PRISM (Ours)	30.29	0.900	30.06	0.889	36.88	0.966	33.73	0.939	32.56	0.913	32.70	0.921

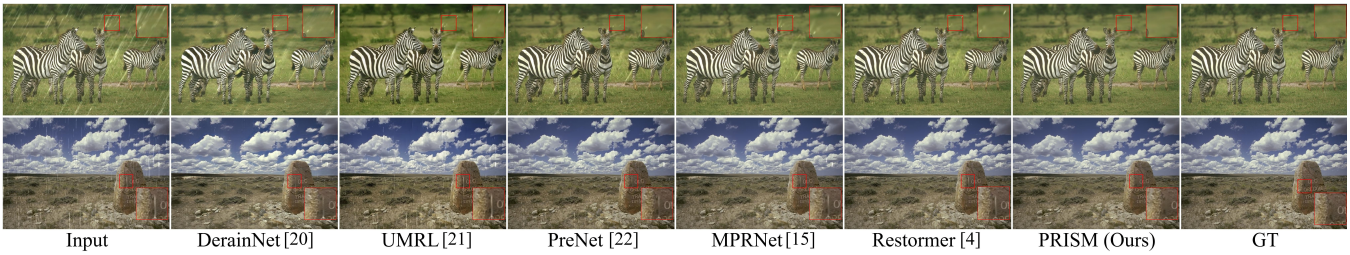

Fig. 4. Deraining results on the Rain100L dataset [25]. Close-up views highlight deraining details.

1,800 training pairs and 100 testing images [25]. Rain14000 offers 12,600 training images with BSD and UCID backgrounds [20]. Rain12 has 12 synthesized images, mainly for testing [27]. This mixed-track strategy improves generalization to varied rain patterns and intensities and better matches real-world deraining scenarios.

3.2. Evaluation Metrics

We adopt Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) as objective metrics. PSNR and SSIM reflect pixel-level fidelity and structural similarity, respectively, with higher values indicating better reconstruction and visual quality [28, 29].

Table 2. Ablation on the effectiveness of different components. Results are reported on the 100L dataset [25] using PSNR.

Stage Ablation				Module Ablation		
CENet	SFNet	RNet	PSNR	HDMamba	HA-UNet	PSNR
✓			34.49		✓	36.09
	✓		33.31			
		✓	33.33	✓		34.84
✓	✓	✓	**36.88**	✓	✓	**36.88**

3.3. Experimental Results

Following prior works [16, 8, 30], we calculate PSNR and SSIM on the Y channel in the YCbCr color space. Across five benchmarks in the mixed training track, PRISM achieves the highest overall average PSNR and SSIM, as shown in Tab. 1. It obtains better results than the recent Transformer-based Semi-SwinDerain [6] and wavelet-based DAWN [24], with pronounced gains on Rain100L and Test2800, and comparable performance on Test1200. Then, we further compare PRISM with widely used deraining baselines under the mixed-track setting and present visual results as shown in Fig. 4. In the first-row zebra example, our model removes slanted rain cues more thoroughly than others, especially in the top-right region, where recent methods leave small raindrop residues. In the second-row stone example, other models tend to misinterpret rain cues as letters on the stone, leading to an erroneous retention of these rain cues. In contrast, PRISM learns to recognize and remove thin and elongated rain streaks, while preserving the integrity of the actual stone letters. These observations indicate robust deraining quality across the rain patterns shown here.

3.4. Ablation Studies

Number of stages. We train CENet, SFNet, and RNet as independent single-stage models at the original resolution under the same settings to assess the benefit of a multi-stage design. As shown in Tab. 2, none of the single-stage variants matches the full three-stage PRISM. RNet further integrates and leverages the information from the preceding CENet and SFNet stages, leading to strong performance. Overall, these results demonstrate that the three-stage fused network is superior to any single-stage counterpart.

Module ablation. We replace HA-UNet with a standard convolutional UNet of the same depth and width. We further replace HDMamba with a stack of plain convolutional blocks , keeping other settings unchanged. As shown in Tab. 2, both substitutions degrade the performance, indicating that HA-UNet and HDMamba are effective and complementary.

4. CONCLUSION

We propose PRISM, an efficient three-stage progressive deraining network for fine-grained recovery with global consistency. The PRISM mainly consists of three key stages, i.e., CENet, SFNet and RNet. Specifically, CENet conducts initial rain removal using HA-UNet. Then SFNet integrates HA-UNet with HDMamba to extract shallow features and provide coarse deraining. Finally, RNet recovers fine structures via ORS. The integration of HA-UNet and HDMamba facilitates balanced preservation of local details and global consistency throughout the deraining process. Moreover, our

approach enhances restoration in local regions and captures critical high-frequency rain-streak clues to remove residual artifacts. PRISM achieves competitively quantitative results and stable visual quality on standard synthetic benchmarks, with restored images showing nearly no residual streaks.

5. REFERENCES

[1] Xiang Chen, Jinshan Pan, Jiangxin Dong, and Jinhui Tang, "Towards unified deep image deraining: A survey and a new benchmark," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 7, pp. 5414–5433, 2025.

[2] Ning Gao, Xingyu Jiang, Xiuhui Zhang, and Yue Deng, "Efficient frequency-domain image deraining with contrastive regularization," in *ECCV*. Springer, 2024, pp. 240–257.

[3] Wenyin Tao, Xuefeng Yan, Yongzhen Wang, and Mingqiang Wei, "Mffdnet: Single image deraining via dual-channel mixed feature fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–13, 2024.

[4] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *CVPR*, 2022, pp. 5728–5739.

[5] Yijin Liu, Guoqiang Xiao, Michael S Lew, and Song Wu, "Multi-scale fusion of gated neighborhood attention transformers for single image deraining," in *ICASSP*. IEEE, 2024, pp. 3130–3134.

[6] Chun Ren, Danfeng Yan, Yuanqiang Cai, and Yangchun Li, "Semi-swinderain: Semi-supervised image deraining network using swin transformer," in *ICASSP*. IEEE, 2023, pp. 1–5.

[7] Duolin Sun, Yimou Wang, Joey Zhaoyu Zuo, and Huan Zheng, "Haformer: Heterogeneous aggregation transformer for single image deraining," in *ICASSP*. IEEE, 2024, pp. 6050–6054.

[8] Hang Guo, Yong Guo, Yaohua Zha, Yulun Zhang, Wenbo Li, Tao Dai, Shu-Tao Xia, and Yawei Li, "Mambairv2: Attentive state space restoration," in *CVPR*, 2025, pp. 28124–28133.

[9] Zeyu Wang, Chen Li, Huiying Xu, Xinzhong Zhu, Xiao Huang, and Hongbo Li, "Restormamba: An enhanced synergistic state space model for image restoration," in *ICASSP*. IEEE, 2025, pp. 1–5.

[10] Haibo Li, Zhanshuo Liu, Tuo Zhao, Tingting Zhao, Yarui Chen, and Ning Xie, "Ms-rainmamba: Learning multi-scale state space models for single image deraining," in *ICASSP*. IEEE, 2025, pp. 1–5.

[11] Zhen Zou, Hu Yu, Jie Huang, and Feng Zhao, "Freqmamba: Viewing mamba from a frequency perspective for image deraining," in *ACM MM*, 2024, pp. 1905–1914.

[12] Albert Gu, Karan Goel, and Christopher Ré, "Efficiently modeling long sequences with structured state spaces," in *ICLR*, 2022.

[13] Albert Gu and Tri Dao, "Mamba: Linear-time sequence modeling with selective state spaces," in *COLM*, 2024.

[14] Zhaoyong Yan, Gang Li, Lingyu Si, and Hongwei Dong, "Fpgnet: Single image deraining with high-frequency channel and frequency domain prior guidance," in *ICASSP*. IEEE, 2024, pp. 4410–4414.

[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, pp. 14821–14831.

[16] Xiang Chen, Jinshan Pan, and Jiangxin Dong, "Bidirectional multi-scale implicit neural representations for image deraining," in *CVPR*, 2024, pp. 25627–25636.

[17] Xueyang Fu, Jie Xiao, Yurui Zhu, Aiping Liu, Feng Wu, and Zheng-Jun Zha, "Continual image deraining with hypergraph convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 8, pp. 9534–9551, 2023.

[18] Xiang Chen, Hao Li, Mingqiang Li, and Jinshan Pan, "Learning a sparse transformer network for effective image deraining," in *CVPR*, 2023, pp. 5896–5905.

[19] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *ICCV*, 2021, pp. 10012–10022.

[20] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley, "Removing rain from single images via a deep detail network," in *CVPR*, 2017, pp. 3855–3863.

[21] Rajeev Yasarla and Vishal M Patel, "Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining," in *CVPR*, 2019, pp. 8405–8414.

[22] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng, "Progressive image deraining networks: A better and simpler baseline," in *CVPR*, 2019, pp. 3937–3946.

[23] 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, pp. 17683–17693.

[24] Kui Jiang, Wenxuan Liu, Zheng Wang, Xian Zhong, Junjun Jiang, and Chia-Wen Lin, "Dawn: Direction-aware attention wavelet network for image deraining," in *ACM MM*, 2023, pp. 7065–7074.

[25] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan, "Deep joint rain detection and removal from a single image," in *CVPR*, 2017, pp. 1357–1366.

[26] Siyuan Li, Iago Breno Araujo, Wenqi Ren, Zhangyang Wang, Eric K Tokuda, Roberto Hirata Junior, Roberto Cesar-Junior, Jiawan Zhang, Xiaojie Guo, and Xiaochun Cao, "Single image deraining: A comprehensive benchmark analysis," in *CVPR*, 2019, pp. 3838–3847.

[27] Yu Li, Robby T Tan, Xiaojie Guo, Jiangbo Lu, and Michael S Brown, "Rain streak removal using layer priors," in *CVPR*, 2016, pp. 2736–2744.

[28] Shaodi You, Robby T Tan, Rei Kawakami, Yasuhiro Mukaigawa, and Katsushi Ikeuchi, "Adherent raindrop modeling, detection and removal in video," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 9, pp. 1721–1733, 2015.

[29] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[30] Guanglu Dong, Tianheng Zheng, Yuanzhouhan Cao, Linbo Qing, and Chao Ren, "Channel consistency prior and self-reconstruction strategy based unsupervised image deraining," in *CVPR*, 2025, pp. 7469–7479.