

Article

MPINet: Multi-Stage Progressive Illumination-Aware Network for Image Deraining

Zhengwen Qian, Xiaoxiong Dong *, Mudong Li and Xuewen Miao

National Key Laboratory of Complex Aviation System Simulation, Beijing 100076, China; afeuzqw@163.com (Z.Q.); afeuzqaz@163.com (M.L.); afeudh@163.com (X.M.)

* Correspondence: afeudxx@126.com

Abstract

Image deraining remains a critical challenge in computer vision, as rain streaks significantly degrade visual quality by introducing complex patterns that overlay scene content, impeding applications from autonomous driving to surveillance systems. Existing methods often struggle to balance global scene understanding with precise removal of rain streaks, resulting in either residual rain artifacts or over-smoothed textures. To address this, we propose MPINet, a Multi-Stage Progressive Illumination-Aware Network that integrates illumination awareness and global-context modeling, which is specifically designed for rain-removal tasks. Our architecture features a novel illumination-aware module that generates illumination maps to enhance robustness in the varying lighting conditions commonly encountered during rainy scenarios. The UniMetaFormer core adaptively incorporates global semantic information through dynamic transformations and attention mechanisms, effectively distinguishing between rain streaks and underlying image content. Inspired by MPRNet's multi-stage restoration framework, our network employs a hierarchical approach with progressive patch-based processing and deep supervision across three stages, enabling efficient integration of our illumination-aware modules while maintaining reasonable model complexity. Experimental results verify the effectiveness of MPINet for rain-removal tasks, demonstrating superior capability in removing rain streaks of varying densities while preserving original image textures. On average, MPINet outperforms MPRNet by approximately 6.5% in terms of PSNR and 1.3% in terms of SSIM across all datasets.

Keywords: global-context modeling; illumination awareness; image deraining

MSC: 94A08; 68T07



Academic Editors: Jianping Gou,
Weihua Ou and Lan Du

Received: 11 April 2026

Revised: 15 May 2026

Accepted: 9 June 2026

Published: 1 July 2026

Copyright: © 2026 by the authors.
Licensee MDPI, Basel, Switzerland.
This article is an open access article
distributed under the terms and
conditions of the [Creative Commons
Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Image deraining remains a central challenge in computer vision. Its aim is to recover clear and detail-rich images from rain-degraded inputs, and it has critical applications spanning autonomous driving, surveillance systems, and outdoor photography. Although deep learning methods, such as U-Net [1] and transformer architectures [2], have made significant strides, existing techniques still face limitations when dealing with complex, non-uniform rain streaks that vary in density, direction, and appearance under different illumination conditions [3]. These limitations primarily stem from three aspects: first, traditional feature-extraction methods struggle to distinguish rain streaks from texture details, making it difficult to effectively balance the removal of rain and the preservation of image

content [4]; second, most networks lack explicit perception and adaptation mechanisms for scene illumination conditions, resulting in poor robustness when processing rainy images captured in low light or extreme contrast scenarios [5]; finally, simple fusion of multi-scale features often fails to leverage the intrinsic properties of rain-streak patterns at different scales, thereby limiting the final deraining accuracy [6].

Recent efforts have sought to mitigate these issues through multi-stage architectures and advanced feature-fusion strategies. Image deraining differs from many other restoration tasks, such as denoising and deblurring, because the visual appearance of rain streaks is strongly influenced by scene illumination. Under daytime conditions, rain streaks may appear as thin translucent structures, whereas under nighttime, backlit, or high-contrast conditions, they often exhibit enhanced brightness, specular reflections, and locally non-uniform visibility. As a result, the same rain pattern can interact with the background in markedly different ways across lighting environments. This property makes illumination awareness particularly important for deraining in real-world scenarios. For instance, MPRNet [7] introduces a progressive multi-stage design that decomposes the restoration process into manageable steps, augmented by perpixel adaptive attention for feature reweighting. This approach effectively balances spatial fidelity and semantic understanding in tasks like deraining. However, MPRNet's insufficient modeling of rain streaks under varying illumination conditions leads to suboptimal performance in complex rainy scenes. Similarly, Restormer [3] leverages transformer-based extraction to enhance global context yet struggles with efficient texture recovery due to self-attention complexity and insufficient handling of illumination variations. Its computational overhead fails to incorporate illumination-aware adaptations for non-uniform lighting. The NightRain [8] achieves effective nighttime video deraining through its two-stage approach combining adaptive rain removal and illumination correction. NiteDR [9] achieves nighttime image deraining through cross-view sensor cooperative learning for dynamic driving scenes. Existing deraining methods still face three major limitations. First, many models do not sufficiently distinguish rain streaks from fine image textures, which often leads to either residual rain artifacts or the loss of structural details. Second, most existing architectures lack an explicit mechanism to perceive and adapt to scene illumination, making them less robust in low-light, high-contrast, or non-uniformly illuminated rainy scenes. Third, conventional multi-scale feature-fusion strategies often fail to fully exploit the scale-dependent characteristics of rain streaks, thereby limiting restoration accuracy. To address these issues, our method introduces illumination-aware guidance for lighting-adaptive restoration, a global-context modeling mechanism for more reliable rain-content discrimination, and a multi-stage progressive architecture for hierarchical refinement across scales. These shortcomings highlight the need for a robust framework that preserves high-frequency details via wavelet transforms, dynamically modulates features with global and illumination-aware mechanisms, and achieves seamless multi-scale integration.

To address these challenges, we propose MPINet, a novel multi-stage progressive illumination-aware network specifically designed for image deraining. The choice of this architecture is driven by a core principle: effective deraining demands the coordinated processing of both global structures and local details while accounting for illumination variations. We, therefore, adopt a coarse-to-fine, multi-stage strategy inspired by MPRNet's framework, designed to first establish a global contextual model in the early stages, which then guides the fine-grained rain removal in subsequent stages. MPINet extends MPRNet's capabilities by seamlessly integrating illumination awareness and global-context modeling, mechanisms absent or underdeveloped in prior designs that often struggle with rain streaks under varying lighting conditions. The core contribution of MPINet lies in the deep integration of two synergistic mechanisms to boost deraining quality: Global-Context

Modeling and Illumination-Aware Regulation. In the central block, we design the Unified Meta-Transformer (UniMetaFormer) block, which incorporates a Global-Context Extractor and dynamic transformations to efficiently inject the semantic information of the entire image into the local-feature-processing stream, achieving effective discrimination between rain streaks and underlying image content. Concurrently, a dedicated illumination-aware module estimates the scene illumination map and utilizes the ICAB (Illumination Channel Attention Block) to embed illumination priors into the multi-stage feature-refinement process, significantly enhancing the model's resilience to diverse lighting environments in rainy scenes—addressing the poor adaptability of prior networks to non-uniform illumination conditions during rain. MPINet adopts a progressive three-stage encoder–decoder architecture based on MPRNet's proven structure, complemented by deep supervision to prevent gradient vanishing and promote stable convergence. Experimental results demonstrate that our method achieves superior performance on image-deraining benchmarks, with an average improvement of 6.5% of PSNR compared to MPRNet across datasets, and particularly excels at removing rain streaks of varying densities while preserving original image textures. The main contributions of this work are as follows:

- (1) **Illumination-Aware Rain Removal:** We introduce illumination awareness to generate illumination maps that guide ICABs for targeted adaptation to local lighting variations in rainy scenes, significantly enhancing performance under diverse illumination conditions.
- (2) **Adaptive Global-Context Modeling:** We propose the UniMetaFormer block with dynamic transformations and attention mechanisms to effectively distinguish between rain streaks and underlying image content, enabling precise rain removal without over-smoothing textures.
- (3) **Enhanced Progressive Architecture:** We improve upon MPRNet's multi-stage framework by integrating our illumination-aware modules while maintaining reasonable model complexity, demonstrating significant performance gains in deraining benchmarks.

The remainder of this paper is organized as follows. Section 2 reviews related work on image deraining and illumination-aware restoration. Section 3 presents the proposed MPINet framework and its key components. Section 4 describes the experimental settings and reports the quantitative results, qualitative comparisons, parameter analysis, and ablation studies. Section 5 concludes the paper and discusses the limitations and future research directions.

2. Related Work

2.1. Image Deraining

Image deraining aims to remove rain streaks from images while preserving original content details. Rain streaks manifest as high-frequency components with varying densities, shapes, and directions, often coupled with illumination changes that complicate the restoration process [10]. Early deraining methods employed sparse coding [11] and Gaussian mixture models [12] to separate rain layers from background content. These approaches relied on handcrafted priors about rain-streak properties but struggled with complex real-world rain patterns and computational efficiency.

Deep learning has revolutionized deraining with end-to-end trainable networks that learn to distinguish rain streaks from image content. Initial CNN-based methods [10] utilized residual architectures to extract rain streaks directly. Subsequent approaches incorporated recurrent frameworks [13], attention mechanisms [14], and multi-scale feature extraction [15] to handle varying rain densities and preserve structural details. Recent advances include GAN-based frameworks [16], graph neural networks for spatial rea-

soning [17], and specialized networks for nighttime scenarios where rain streaks interact with limited visibility and uneven illumination [8,9]. Despite these advances, deraining remains challenging due to the diversity of rain patterns across different scenes, the difficulty in modeling rain accumulation effects, and the complex interaction between rain and scene illumination.

Multi-stage networks decompose the restoration task into a coarse-to-fine progression, a structure highly effective for establishing a global context before refining local details [18–23]. This design allows initial stages to focus on broader contextual cues, which then guide subsequent stages of detailed reconstruction. However, simply using identical subnetworks across stages can yield suboptimal results [7]. For instance, MPRNet [7] introduces a multi-stage progressive architecture where encoder–decoder branches are crucial for building contextual information at a global scale, while a high-resolution branch preserves spatial details. Its supervised attention modules (SAM) and cross-stage feature fusion (CSFF) ensure that this global context effectively informs local-feature refinement across stages, achieving strong performance on deraining, deblurring, and denoising.

Recent advances have further emphasized the use of multi-stage designs for integrating powerful global modeling mechanisms. DEMNet [24] leverages this structure to generalize across degradations [24], while MSTNet [25] explicitly integrates local–global transformers in its multi-stage pipeline to enhance long-range dependencies for deraining [25]. These methods build on the multi-stage paradigm not just for gradual refinement, but as a strategic framework to inject global awareness into the restoration process. Meanwhile, recent deraining research in 2024–2025 has increasingly explored architectures beyond conventional CNN-only designs. Representative transformer-based methods include SFformer [26], which introduces adaptive sparse attention and frequency-guided feed-forward design for single-image deraining, and DBSTNet [27], which combines a sparse-transformer branch with a U-shaped local recovery branch. In parallel, state-space-model-based deraining methods such as DFSSM [27] have further expanded the design space by improving long-range dependency modeling and multi-scale representation under complex rain degradations. These developments reflect a broader trend toward integrating global modeling, frequency-aware processing, and efficiency-oriented design in modern deraining networks.

2.2. Illumination-Aware Restoration

Illumination variations present significant challenges for image restoration tasks, particularly in deraining scenarios where low-light conditions and uneven lighting interact with rain degradations. Early approaches to handling illumination issues employed histogram equalization [28] and Retinex theory [29] to decompose images into illumination and reflectance components. With deep learning, specialized architectures have emerged for joint illumination enhancement and restoration. Enlighten-GAN [30] introduced an unsupervised approach using attention-guided enhancement, while KinD [31] decomposed the restoration process into illumination-adjustment and detail-refinement stages. Recent methods like IDT [32] leverages illumination-guided dual transformers for low-light deraining. Despite these advances, many deraining methods still struggle with complex illumination patterns, as they often prioritize rain removal without adequately modeling the interaction between lighting conditions and rain appearances. This limitation highlights the need for illumination-aware architectures that can explicitly reason about lighting variations while performing restoration tasks.

3. Framework

Our proposed MPINet is a multi-stage progressive network designed for robust image deraining with illumination awareness, building upon the MPRNet framework [7]. The architecture decomposes the rain-removal process into sequential refinement stages, enabling hierarchical feature learning from coarse contextual recovery to fine-grained detail enhancement. At its core, MPINet integrates illumination adjustment modules that normalize lighting variations before feature extraction, illumination channel attention blocks (ICABs) that adaptively process features based on brightness characteristics, and UniMetaFormer blocks with dynamic activation functions (DyT) for improved rain pattern modeling. These components are strategically integrated to enhance performance in varying illumination conditions. The overall architecture of MPINet is illustrated in Figure 1. The details of these illumination-aware enhancements are presented in Sections 3.1–3.3.

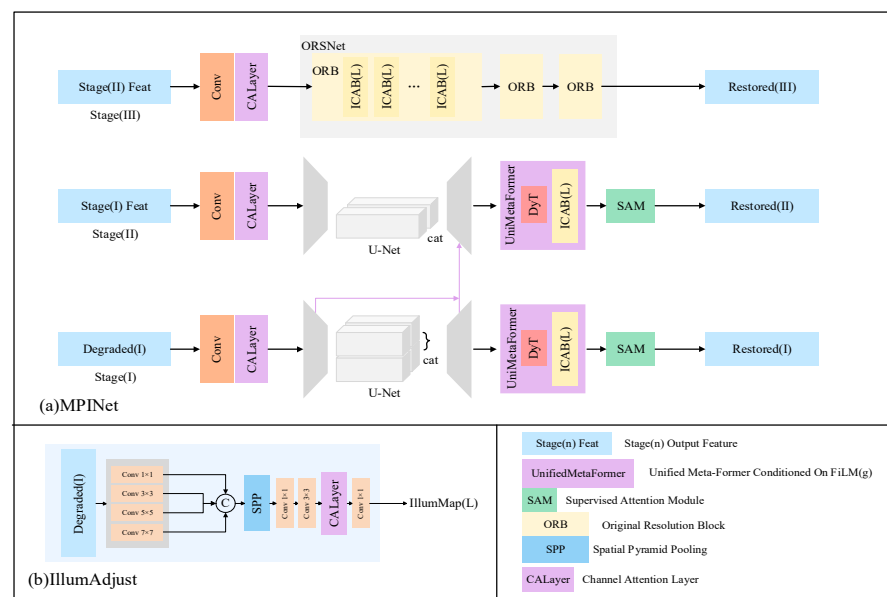


Figure 1. Overall architecture of the proposed MPINet and its illumination adjustment module. (a) MPINet adopts a three-stage progressive illumination-aware restoration framework. Each stage starts with a Conv and CALayer for shallow feature extraction. The first two stages use U-Net-style encoder-decoder networks enhanced with UniMetaFormer blocks (DyT + ICAB) and Supervised Attention Modules (SAM). The third stage employs an Original Resolution Subnetwork (ORSNet) composed of ICAB-based ORBs for global detail refinement. (b) Illumination channel attention block produces the illumination map L through multi-scale convolutions and a CALayer to provide illumination guidance across all stages. In the diagram, different colors denote different components: light blue boxes are input/output feature maps (e.g., Degraded(I), Stage(n) features, and restored outputs); orange denotes convolutional (Conv) layers; purple denotes the Channel Attention Layer (CALayer); magenta denotes the UniMetaFormer module (Unified Meta-Former conditioned on FiLM(g)), within which the red block is DyT; yellow denotes the ICAB/Original Resolution Block (ORB); green denotes the Supervised Attention Module (SAM); blue denotes Spatial Pyramid Pooling (SPP); and gray denotes the U-Net encoder-decoder structure.

3.1. Illumination-Aware Module

The illumination-aware module is used to generate spatially variant illumination guidance maps to inform subsequent processing stages about lighting conditions across the image plane. Given an input image $I \in \mathbb{R}^{3 \times H \times W}$, the module computes a single-channel illumination map $L \in \mathbb{R}^{1 \times H \times W}$ through a hierarchical feature-extraction process.

The illumination estimation is formally defined as

$$L = \sigma(W_1 \circ A \circ \varphi \circ W_5(I)) \quad (1)$$

where W_5 represents a convolutional operation with 5×5 kernel size for capturing broad illumination patterns, φ denotes the LeakyReLU [33] activation function with negative slope parameter 0.2, A implements a channel-attention mechanism with reduction ratio 5 for emphasizing illumination-critical regions, W_1 indicates a point-wise convolutional layer that produces the single-channel output, and σ signifies the sigmoid activation function that normalizes illumination values to the bounded interval $[0, 1]$. The architectural implementation follows a sequential processing pathway where the input RGB image undergoes progressive transformation through convolutional filtering with extended receptive fields for spatial-context capture, non-linear activation, channel-wise attention weighting for selective feature enhancement, and dimensional reduction, culminating in sigmoidal normalization that yields probability-like values suitable for multiplicative feature modulation in subsequent network layers.

The resultant illumination map L encodes the spatial intensity distribution of the input image, serving as an explicit prior that characterizes local lighting conditions. This illumination representation provides essential guidance for adaptive processing throughout the network hierarchy, enabling context-aware restoration that respects the physical constraints of scene lighting. The module establishes a foundation for illumination-informed computation that enhances network sensitivity to lighting variations and promotes visually coherent results across diverse illumination scenarios.

3.2. Illumination-Conditioned Attention Block

The Illumination-Aware Channel Attention Block (ICAB) implements a dual-attention mechanism that synergistically integrates channel-wise feature recalibration with illumination-aware spatial modulation. Given an input feature map $X \in \mathbb{R}^{C \times H \times W}$ and an illumination guidance map $L \in \mathbb{R}^{1 \times H \times W}$, the block performs sequential processing through a residual pathway enhanced by complementary attention mechanisms. The fundamental operation begins with a dual-convolutional body that transforms the input features as follows: $F = W_2(\phi(W_1(X)))$, where W_1 and W_2 represent convolutional layers with kernel size k , and ϕ denotes the activation function. This transformed feature map subsequently undergoes channel-wise recalibration through the channel attention layer: $F_{CA} = CA(F)$, which computes adaptive scaling factors based on global channel statistics to emphasize semantically meaningful feature dimensions.

The core innovation of ICAB manifests in the illumination-aware spatial attention mechanism, formulated as $S = \sigma(W_5([L \parallel X]))$, where $\parallel$ indicates channel-wise concatenation and W_5 denotes a convolutional operation with 5×5 kernel size. This spatial attention map $S \in \mathbb{R}^{1 \times H \times W}$ encodes illumination-dependent weighting coefficients that modulate the original input features. The complete forward propagation is, thus, expressed as $Y = F_{CA} + X + X \odot S$, where $\odot$ represents element-wise multiplication. This formulation establishes a comprehensive feature enhancement framework that simultaneously addresses channel-wise feature importance and spatially variant illumination conditions, enabling contextually adaptive representation learning for illumination-robust image restoration.

The motivation for illumination-conditioned processing is grounded in the photometric properties of rain streaks. Rain streaks are relatively consistent bright elongated patterns whose photometric signature does not fundamentally change with scene illumination. In contrast, scene content—particularly textures and edges—exhibits significant appearance variation under different lighting conditions. In low-light or high-contrast environments,

CNN features tend to conflate rain-induced intensity variations with illumination-induced ones, leading to either residual rain artifacts or over-smoothing of scene content. By conditioning feature processing on the estimated illumination map L , ICAB applies differentiated spatial weighting across bright and dark regions of the image, where the photometric relationship between rain streaks and background content differs. This illumination-conditioned modulation provides the network with an explicit inductive bias to maintain stable rain-content discrimination across varying lighting conditions, offering a principled architectural motivation beyond purely empirical observation.

3.3. Unified Meta-Transformer Block

The Unified Meta-Transformer (UniMetaFormer) block employs a residual architecture that integrates dynamic feature conditioning with illumination-aware processing. The core innovation lies in the Dynamic Tanh Transformation module, which applies a parameterized non-linear projection defined by $D(\mathbf{X}) = \tanh(\alpha \cdot \mathbf{X}) \odot \mathbf{w} + \mathbf{b}$. This transformation enables adaptive feature conditioning through learnable scaling and channel-wise affine parameters.

The processed features then undergo illumination-aware attention through the ICAB module, which incorporates spatial illumination guidance to enable context-adaptive feature enhancement. Following the attention mechanism, a compact MLP branch performs feature refinement through channel expansion and contraction operations. The complete transformation maintains residual connections throughout, ensuring stable gradient propagation while effectively integrating illumination-aware processing into the feature representation. This design achieves robust performance across varying lighting conditions through a synergistic combination of dynamic conditioning and spatial adaptation.

3.4. Overview of the MPINet

The proposed MPINet architecture builds upon multi-stage progressive restoration while integrating comprehensive illumination-awareness. The framework processes images through three hierarchical stages with explicit illumination guidance.

The illumination modeling begins with the IllumAdjust module, which generates a global illumination map L through a convolutional encoder with channel attention. This illumination prior undergoes hierarchical downsampling to produce multi-scale representations $\{L^{(1)}, L^{(2)}, \text{ and } L^{(3)}\}$ that guide feature processing at the corresponding network levels. The illumination modeling begins with the IllumAdjust module, which generates a global illumination map L through a convolutional encoder with channel attention. This illumination prior undergoes hierarchical downsampling to produce multi-scale representations $\{L^{(1)}, L^{(2)}, \text{ and } L^{(3)}\}$ that guide feature processing at the corresponding network levels.

Stage 1 employs a quad-patch processing strategy, dividing the input image into four non-overlapping subregions. Each patch undergoes independent processing through dedicated encoder–decoder pathways composed of Unified MetaFormer blocks. These blocks integrate Dynamic Tanh Transformation for feature conditioning and Illumination Channel Attention Blocks (ICAB) for illumination-aware feature enhancement.

Stage 2 processes two larger patches through enhanced encoder–decoder networks that incorporate Cross-Stage Feature Fusion (CSFF). This stage receives concatenated features from Stage 1 through Supervised Attention Modules (SAM), which generate intermediate predictions while preserving spatial correspondence. This is essentially similar to the feature-information-transfer method in the intermediate stages of MPINet, but the difference is that multi-scale illumination information is integrated into the transferred features.

Stage 3 processes the full-resolution image through an ORSNet module composed of cascaded ORB blocks. Each ORB incorporates multiple ICAB units with residual connections, enabling iterative refinement under global illumination guidance. The final output is generated through a tail convolution that aggregates multi-scale features while maintaining a residual connection to the original input.

The complete architecture demonstrates a hierarchical progression from local patch processing to global refinement, with illumination-awareness consistently integrated through spatially adaptive attention mechanisms and multi-scale guidance maps.

4. Experiments and Analysis

To evaluate the effectiveness of our proposed MPINet, we conducted comprehensive experiments on several benchmark datasets for image deraining tasks. We compared MPINet against famous methods and performed ablation studies to validate the contributions of key components. All experiments are implemented in PyTorch 2.1.0 and conducted on an NVIDIA RTX 4090 GPU (NVIDIA Corporation, Santa Clara, CA, USA) with 24 GB memory.

4.1. Datasets

We evaluate MPINet through three widely used deraining-capability-assessment datasets, specifically:

- (1) Rain100L/H [34]: A set of synthetic datasets containing light rain (Rain100L) and heavy rain (Rain100H), both including 100 image pairs (a rainy image and its corresponding clean ground truth). Rain100L features low-density rain streaks for evaluating models under light rain conditions. Rain100H presents a more challenging benchmark with higher-density and more complex patterns (e.g., multi-directional) streaks, suitable for evaluating performance under severe rain conditions.
- (2) RealRain1kL/H [35]: A real-world rain dataset comprising light rain (RealRain1kL) and heavy rain (RealRain1kH), each containing 1000 rainy-clean image pairs. These images are collected from real scenes and combined with synthetic enhancement. RealRain1kL provides realistic visual characteristics for light rain, while RealRain1kH contains dense, large rain streaks and severe visibility degradation, designed to evaluate model robustness in heavy-rain scenarios.
- (3) SPA-Data [35]: A real-world dataset with 95,000 rainy-clean image pairs collected under various weather conditions, including light rain and night scenes. We use 90,000 pairs for training and 5000 for testing.

These datasets allow us to assess MPINet's generalization across synthetic/real, day-time/nighttime, and single/multi-degradation settings.

4.2. Implementation Details

MPINet is trained end-to-end using the AdamW optimizer with an initial learning rate of 1×10^{-4} , $\beta_1 = 0.9$, and $\beta_2 = 0.999$. We employ a cosine annealing scheduler with a 3-epoch warmup, reducing the learning rate to 1×10^{-6} over 1000 epochs. Mixed-precision training (AMP) is used with a gradient scaler for efficiency. The batch size is 16, and patches of size 128×128 are randomly cropped with horizontal flips for augmentation. The loss function is a multi-term combination: Charbonnier loss (L_{char}) for pixel-wise reconstruction, edge-aware loss (L_{edge}) for structural preservation, and LPIPS (L_{LPIPS}) for perceptual quality

$$L = L_{\text{char}} + 0.05L_{\text{edge}} + 0.1L_{\text{LPIPS}} \quad (2)$$

Multi-stage supervision is applied to intermediate outputs. Models are trained for 1000 epochs, with checkpoints saved every 25 epochs and the best model selected based on validation PSNR. Pre-trained weights from ImageNet are not used, as MPINet learns priors directly from degradation data. For evaluation, we report Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) on the Y-channel for color images. While LPIPS is incorporated as a perceptual-loss term during training to improve visual quality, it is not reported as a standalone benchmark metric in Table 1 as most prior methods do not provide corresponding LPIPS values for direct comparison. Inference is conducted at full resolution without augmentation.

Table 1. Quantitative comparison with other image deraining methods using the Rain100L, Rain100H, RealRain1kL, and RealRain1kH benchmarks.

Datasets	Rain100L		Rain100H		RealRain1kL		RealRain1kH	
Metrics	PSNR (dB) ↑	SSIM ↑	PSNR (dB) ↑	SSIM ↑	PSNR (dB) ↑	SSIM ↑	PSNR (dB) ↑	SSIM ↑
Input	29.6	0.838	13.56	0.371	22.27	0.7657	25.93	0.8651
CNN-based Methods								
DerainNet	27.03	0.8841	14.92	0.5923	27.09	0.925	22.88	0.8886
DDN	32.38	0.9259	24.64	0.849	31.18	0.9172	29.17	0.8783
PreNet	37.48	0.979	29.46	0.899	-	-	-	-
SPANet	35.33	0.969	25.11	0.827	30.43	0.947	25.76	0.9095
MPRNet	36.4	0.965	30.41	0.89	36.29	0.972	34.74	0.964
NAFNet	37	0.978	29.66	0.9	38.8	0.986	36.11	0.976
MPINet (Ours)	39.93	0.9755	31.22	0.912	40.11	0.9800	36.11	0.9710
Transformer-based Methods								
Restormer	38.99	0.978	31.46	0.904	40.9	0.9849	39.57	0.9812
PromptIR	38.34	0.983	28.69	0.877	36.99	0.973	33.61	0.953
DRSFormer	41.32	0.9887	32.07	0.9316	-	-	-	-
NeRD-Rain-S	42.00	0.9900	32.86	0.9320	38.64	0.9790	36.69	0.9700

Note: We report the average PSNR (dB) and SSIM values. The results of us are marked in bold.

4.3. Main Results

To evaluate the effectiveness of our proposed MPINet model for image deraining, we conducted extensive experiments on four widely used benchmark datasets: Rain100L [34], Rain100H [34], RealRain1kL [35], and RealRain1kH [35]. These datasets encompass a variety of rain-streak patterns, ranging from synthetic light rain (Rain100L) to heavy and realistic rain scenarios (RealRain1kH). We compare our method against state-of-the-art CNN-based approaches, including DerainNet [36], DDN [37], PreNet [19], SPANet [28], MPRNet [7], and NAFNet [38], as well as transformer-based methods such as Restormer [3], PromptIR [39], DRS-Former [40], and NeRD-Rain-S [41].

The quantitative results are presented in Table 1. As shown, our proposed MPINet demonstrates competitive performance among CNN-based methods across the benchmark datasets. Specifically, MPINet attains the highest PSNR among CNN-based approaches on Rain100L (39.93 dB), Rain100H (31.22 dB), and RealRain1kL (40.11 dB), and achieves comparable PSNR to NAFNet on RealRain1kH (36.11 dB). Notably, MPINet achieves this with only 10.40 M parameters compared to NAFNet's 17.11 M—a 39% reduction in model size—demonstrating a more favorable parameter–performance trade-off overall.

Furthermore, when compared with transformer-based methods, MPINet remains highly competitive. Although recent transformer architectures like NeRDRain-S and DRS-Former achieve higher scores on some benchmarks, our CNN-based approach offers advantages in terms of computational efficiency while maintaining comparable performance.

Notably, on RealRain1kL, MPINet's PSNR of 40.11 dB is competitive with Restormer's 40.9 dB. Our model maintains a favorable balance between fidelity and structural consistency, as reflected by its consistently high SSIM scores across all datasets. The visualization effects are shown in Figure 2, which demonstrates the different rain-removal results in both daytime and nighttime conditions.

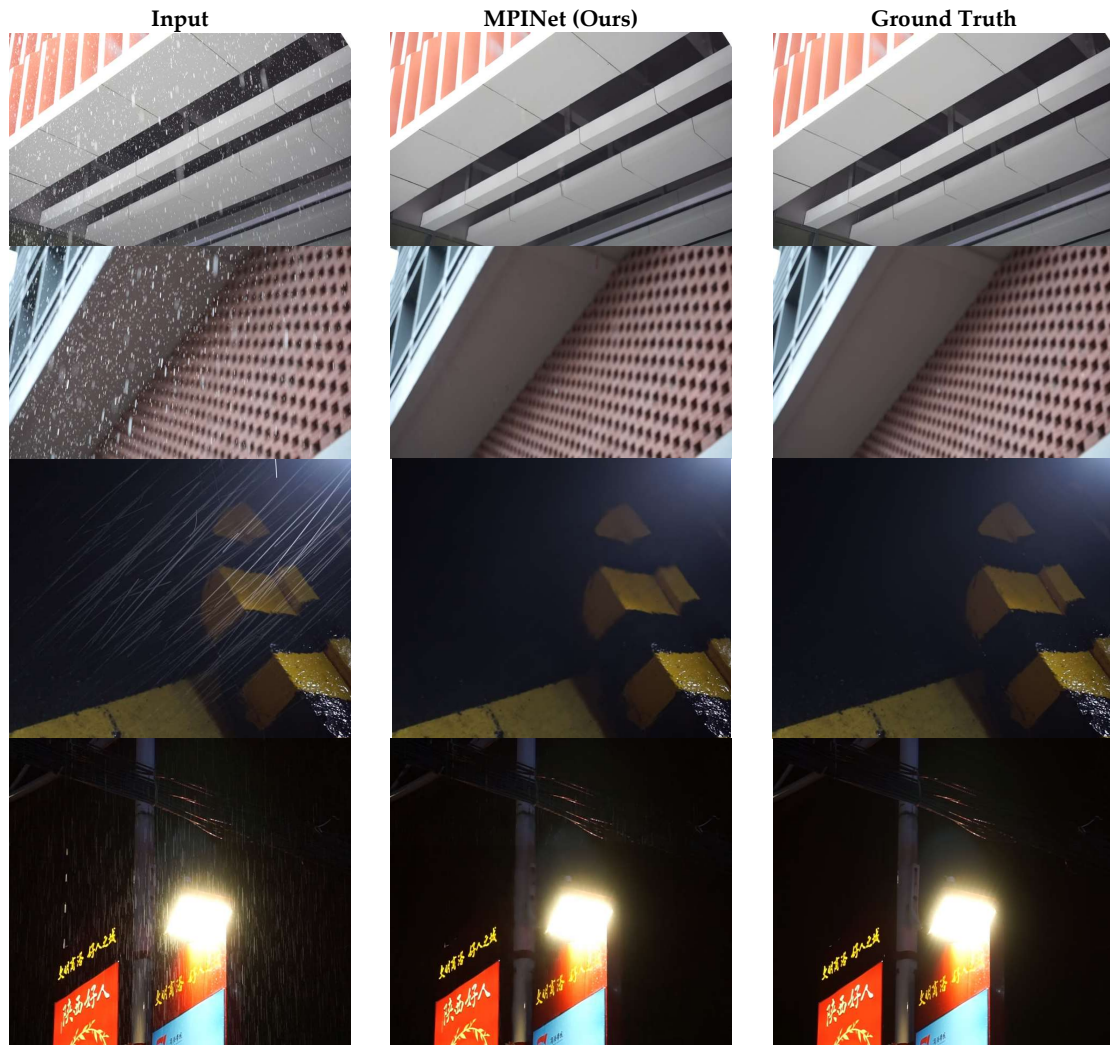


Figure 2. Qualitative results of MPINet on the RealRain1kL dataset under different illumination conditions. The first two rows correspond to daytime scenes, while the bottom two rows correspond to nighttime scenes. MPINet effectively removes rain streaks while preserving structural details and maintaining visual consistency across both well-lit and low-light environments.

It is worth noting that the advantage of illumination-aware modeling becomes less pronounced on RealRain1kH. A possible reason is that this dataset mainly represents heavy-rain scenarios, where severe rain occlusion and dense streak accumulation become the dominant degradation factors. Under such conditions, the marginal benefit of illumination guidance may be reduced because the restoration difficulty is increasingly governed by large-scale rain interference rather than illumination variation alone. Nevertheless, MPINet still maintains competitive performance, indicating that the proposed design remains effective even when the contribution of illumination awareness is partially saturated in heavy-rain scenes.

4.4. Parameter Analysis

As evidenced in Table 2, the evolution of deraining approaches demonstrates a clear trade-off between model complexity and representational capacity. Early CNN methods like DDN (0.06 M) and PreNet (0.28 M) pursued extreme parameter efficiency but lacked sufficient modeling capacity for complex rain patterns, particularly in real-world scenarios with varying illumination conditions.

Table 2. Comparison of model parameters for different image deraining networks.

	Model	Params(M)	FLOPs (G) @ 256×256
CNN	DerainNet	0.75	51.48
	DDN	0.06	3.79
	PreNet	0.28	59.37
	SPANet	0.28	-
	MPRNet	3.64	141.28
	NAFNet	17.11	-
	MPINet (Ours)	10.40	158.4
Transformer	Restormer	26.10	174.7
	PromptIR	32.11	158.1
	DRSFormer	33.66	242.9
	NeRD-Rain-S	10.53	79.2

The parameter distribution in our MPINet (10.40 M) reflects a deliberate architectural strategy that allocates computational resources more effectively than recent transformer-based models. While transformer architectures like Restormer (26.10 M) and DRSFormer (33.66 M) invest heavily in self-attention mechanisms to capture global dependencies, our analysis suggests that much of this capacity becomes redundant for rain-removal tasks. Instead, MPINet's multi-perspective integration approach achieves comparable or superior performance with approximately one-third of the parameters by focusing computational resources on modeling critical illumination–rain interactions. The strategic deployment of illumination-aware modules in MPINet demonstrates that architectural innovations targeting specific physical characteristics of the degradation process can yield greater efficiency gains than simply scaling model size. In terms of computational cost, MPINet requires 158.4G FLOPs at an input size of 256×256 , which remains lower than Restormer and DRSFormer while maintaining a moderate parameter scale. This finding challenges the prevailing trend toward ever-larger models being used in image restoration tasks and suggests that domain-specific insights can significantly improve the parameter–performance curve for specialized applications like deraining.

4.5. Ablation Studies

We conduct comprehensive ablation studies to evaluate the contribution of each proposed component in our MPINet architecture. As shown in Table 3, we progressively integrate our illumination-aware enhancements to the baseline MPRNet and evaluate performance on the challenging RealRain1kL and RealRain1kH datasets. The addition of the Illumination-Aware Module (IAM) provides substantial improvements of +1.12 dB and +0.44 dB PSNR on the respective datasets, demonstrating the importance of explicit illumination modeling.

Further incorporating the Illumination-Conditioned Attention Block (ICAB) yields additional performance gains, with PSNR improvements reaching +2.24 dB and +0.73 dB over the baseline. The integration of UniMetaFormer with Dynamic Tanh (DyT) activation further enhances feature-representation capabilities, contributing to consistent performance

improvements. The full MPINet architecture achieves significant gains of +3.82 dB and +1.37 dB PSNR on RealRain1kL and RealRain1kH, respectively, compared to MPRNet, validating the effectiveness of our illumination-aware design for robust rain removal under various lighting conditions.

Table 3. Ablation study on the components of MPINet compared to the baseline MPRNet.

Model Configuration	RealRain1kL		RealRain1kH	
	PSNR	SSIM	PSNR	SSIM
MPRNet (Baseline)	36.29	0.972	34.74	0.964
+Illumination-Aware Module	37.41	0.974	35.18	0.967
+Illumination-Conditioned Attention (ICAB)	38.53	0.976	35.47	0.969
+UniMetaFormer	39.42	0.978	35.83	0.970
MPINet	40.11	0.980	36.11	0.971

Note: Experiments conducted on RealRain1kL and RealRain1kH datasets.

4.6. Generalization Studies

To investigate the cross-dataset generalization capabilities of our proposed MPINet, we conducted experiments by training the model on different rain datasets and evaluating its performance on the unseen SPA-Data benchmark. As shown in Table 4, the model trained on RealRain1kL demonstrates the strongest generalization ability, achieving 33.38 dB PSNR and 0.9501 SSIM on SPA-Data. This significantly outperforms models trained on synthetic datasets Rain100L and Rain100H, which achieve lower PSNR values of 31.57 dB and 30.33 dB, respectively.

Table 4. Generalization capability of MPINet when trained on different datasets and evaluated on SPA-Data.

Training Dataset	SPA-Data	
	PSNR	SSIM
RealRain1kL	33.38	0.9501
RealRain1kH	32.32	0.9455
Rain100L	31.57	0.9107
Rain100H	30.33	0.8989

The superior generalization performance of models trained on real-world rain datasets (RealRain1kL and RealRain1kH) highlights the importance of using realistic training data for practical deraining applications. Notably, the illumination-aware design of MPINet enables it to better handle the diverse lighting conditions present in real-world scenarios, facilitating more effective knowledge transfer across different datasets. These results demonstrate that our model not only performs well on its training distribution but also generalizes effectively to unseen rain patterns and scenes.

5. Conclusions

We present MPINet, a novel illumination-aware enhancement of the MPRNet framework that significantly improves deraining performance through the integration of multi-perspective feature extraction, illumination-guided modulation, and efficient context modeling. By explicitly addressing the critical relationship between illumination conditions and rain-streak appearance, our proposed architecture advances beyond MPRNet's multi-stage progressive design while strategically balancing model capacity and computational efficiency.

Extensive experiments validate the effectiveness of our improvements. Compared to the base MPRNet architecture, our illumination-aware modifications achieve consistently better performance across multiple deraining benchmarks. MPINet maintains strong quantitative metrics on standard datasets while significantly outperforming previous methods on real-world rainy images with challenging illumination variations. The performance gains are particularly evident in complex scenes where rain characteristics vary with lighting conditions, confirming that our illumination-aware adaptations successfully address limitations in existing progressive-restoration frameworks.

Limitations remain. Despite our improvements, MPINet still has several failure cases. In extremely low-light scenes accompanied by motion blur, the illumination estimation may become less reliable, which can weaken the subsequent feature-modulation process. Similarly, in cases with strong specular reflections or highly irregular rain patterns, the network may confuse bright rain streaks with fine background structures, leading to either incomplete rain removal or slight detail smoothing. These failure cases suggest that more robust illumination estimation and degradation disentanglement remain important directions for future work.

Overall, MPINet demonstrates how targeted, illumination-aware modifications to progressive restoration architectures can yield significant performance improvements for image deraining. By optimizing MPRNet's foundation with our proposed enhancements, we are advancing the field toward more robust restoration techniques that are applicable in real-world scenarios with varied lighting conditions.

Author Contributions: Conceptualization, Z.Q.; methodology, X.D.; validation, M.L.; writing—original draft, X.M.; and writing—review and editing, X.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the 18th International Conference Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241. [\[CrossRef\]](#)
2. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *arXiv* **2017**, arXiv:1706.03762. [\[CrossRef\]](#)
3. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H. Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5728–5739. [\[CrossRef\]](#)
4. Chen, H.; Wang, Y.; Guo, T.; Xu, C.; Deng, Y.; Liu, Z.; Ma, S.; Xu, C.; Xu, C.; Gao, W. Pre-trained image processing transformer. *arXiv* **2020**, arXiv:2012.00364. [\[CrossRef\]](#)
5. Ismail Fawaz, H.; Forestier, G.; Weber, J.; Idoumghar, L.; Muller, P.A. Deep learning for time series classification: A review. *Data Min. Knowl. Discov.* **2019**, *33*, 917–963. [\[CrossRef\]](#)
6. Li, J.; Fang, F.; Mei, K.; Zhang, G. Multi-scale residual network for image super-resolution. In Proceedings of the 15th European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 517–532. [\[CrossRef\]](#)
7. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H.; Shao, L. Multi-stage progressive image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 14821–14831. [\[CrossRef\]](#)
8. Lin, B.; Jin, Y.; Yan, W.; Ye, W.; Yuan, Y.; Zhang, S.; Tan, R.T. Nightrain: Nighttime video deraining via adaptive-rain-removal and adaptive-correction. In Proceedings of the 38th Annual AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 26–27 February 2024; pp. 3378–3385. [\[CrossRef\]](#)

9. Shi, C.; Fang, L.; Wu, H.; Xian, X.; Shi, Y.; Lin, L. NiteDR: Nighttime image deraining with cross-view sensor cooperative learning for dynamic driving scenes. *IEEE Trans. Multi-Media* **2024**, *26*, 9203–9215. [\[CrossRef\]](#)
10. Wang, H.; Xie, Q.; Wu, Y.; Zhao, Q.; Meng, D. Single image rain streaks removal: A review and an exploration. *Int. J. Mach. Learn. Cybern.* **2020**, *11*, 853–872. [\[CrossRef\]](#)
11. Kang, L.W.; Lin, C.W.; Fu, Y.H. Automatic single-image-based rain streaks removal via image decomposition. *IEEE Trans. Image Process.* **2012**, *21*, 1742–1755. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Li, Y.; Tan, R.T.; Guo, X.; Lu, J.; Brown, M.S. Rain streak removal using layer priors. In Proceedings of the IEEE Conference Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2736–2744. [\[CrossRef\]](#)
13. Cao, X.; Hao, S.; Xu, L. Single image deraining by fully exploiting contextual information. *Neural Process. Lett.* **2022**, *54*, 853–870. [\[CrossRef\]](#)
14. Zhang, H.; Xie, Q.; Lu, B.; Gai, S. Dual attention residual group networks for single image deraining. *Digit. Signal Process.* **2021**, *116*, 103106. [\[CrossRef\]](#)
15. Jiang, K.; Wang, Z.; Yi, P.; Chen, C.; Huang, B.; Luo, Y.; Ma, J.; Jiang, J. Multi-scale progressive fusion network for single image deraining. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 8346–8355. [\[CrossRef\]](#)
16. Yadav, S.; Mehra, A.; Rohmetra, H.; Ratnakumar, R.; Narang, P. DerainGAN: Single image deraining using wasserstein GAN. *Multimed. Tools Appl.* **2021**, *80*, 36491–36507. [\[CrossRef\]](#)
17. Fu, X.; Qi, Q.; Zha, Z.J.; Ding, X.; Wu, F.; Paisley, J. Successive graph convolutional network for image deraining. *Int. J. Comput. Vis.* **2021**, *129*, 1691–1711. [\[CrossRef\]](#)
18. Nah, S.; Hyun Kim, T.; Mu Lee, K. Deep multi-scale convolutional neural network for dynamic scene deblurring. In Proceedings of the IEEE Conference Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 257–265. [\[CrossRef\]](#)
19. Ren, D.; Zuo, W.; Hu, Q.; Zhu, P.; Meng, D. Progressive image deraining networks: A better and simpler baseline. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 3937–3946. [\[CrossRef\]](#)
20. Suin, M.; Purohit, K.; Rajagopalan, A. Spatially-attentive patch-hierarchical network for adaptive motion deblurring. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 3606–3615. [\[CrossRef\]](#)
21. Zhang, H.; Dai, Y.; Li, H.; Koniusz, P. Deep stacked hierarchical multi-patch network for image deblurring. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 5978–5986. [\[CrossRef\]](#)
22. Zheng, Y.; Yu, X.; Liu, M.; Zhang, S. Residual multiscale based single image deraining. In Proceedings of the 30th British Machine Vision Conference, Cardiff, UK, 9–12 September 2019; pp. 1–12.
23. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Vedaldi, A. Gather-excite: Exploiting feature context in convolutional neural networks. *arXiv* **2018**, arXiv:1810.12348. [\[CrossRef\]](#)
24. Yang, Y.; Wang, X.; Lin, X.; Chen, H. Demnet: A degradation difference enabled multi-stage network for image restoration. *Knowl.-Based Syst.* **2025**. preprint. [\[CrossRef\]](#)
25. Liu, R.; Wang, L.; He, J.; Wang, J.; Zhang, J.; Liu, X.; Wang, C.; Zhang, H.; Dai, S. Mstnet: A multi-stage progressive network with local-global transformer for image restoration. *Appl. Intell.* **2025**. preprint. [\[CrossRef\]](#)
26. Wang, X.; Zhang, H.; Cai, K.; Miao, D.; Zhang, Q.; Li, M. SFformer: Adaptive Sparse and Frequency-Guided Transformer Network for Single Image Derain. In *Pattern Recognition and Computer Vision—PRCV 2024*; Lin, Z., Zha, H., Cheng, M.-M., He, R., Liu, C.-L., Ubul, K., Silamu, W., Zhou, J., Eds.; Lecture Notes in Computer Science; Springer: Singapore, 2025; Volume 15038, pp. 482–496. [\[CrossRef\]](#)
27. Yamashita, S.; Tsubota, K.; Ikeda, H.; Matsuo, H. Image deraining with frequency-enhanced state space model. *arXiv* **2024**, arXiv:2405.16470.
28. Pizer, S.M.; Amburn, E.P.; Austin, J.D.; Cromartie, R.; Geselowitz, A.; Greer, T.; ter Haar Romeny, B.; Zimmerman, J.B.; Zuiderveld, K. Adaptive histogram equalization and its variations. *Comput. Vis. Graph. Image Process.* **1987**, *39*, 355–368. [\[CrossRef\]](#)
29. Land, E.H.; McCann, J.J. Lightness and retinex theory. *J. Opt. Soc. Am.* **1971**, *61*, 1–11. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Gong, Y.; Liao, P.; Zhang, X.; Zhang, L.; Chen, G.; Zhu, K.; Tan, X.; Lv, Z. Enlighten-gan for super resolution reconstruction in mid-resolution remote sensing images. *Remote Sens.* **2021**, *13*, 1104. [\[CrossRef\]](#)
31. Zhang, Y.; Zhang, J.; Guo, X. Kindling the dark-ness: A practical low-light image enhancer. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 1632–1640. [\[CrossRef\]](#)
32. Xiao, J.; Fu, X.; Liu, A.; Wu, F.; Zha, Z.J. Image deraining transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 12978–12995. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Xu, J.; Li, Z.; Du, B.; Zhang, M.; Liu, J. ReLUpex made more practical: Leaky ReLU. In Proceedings of the 2020 IEEE Symposium on Computers and Communications, Rennes, France, 7–10 July 2020; pp. 1–7. [\[CrossRef\]](#)

34. Yang, W.; Tan, R.T.; Feng, J.; Liu, J.; Guo, Z.; Yan, S. Deep joint rain detection and removal from a single image. In Proceedings of the IEEE Conference Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1357–1366. [\[CrossRef\]](#)
35. Wang, T.; Yang, X.; Xu, K.; Chen, S.; Zhang, Q.; Lau, R.W. Spatial attentive single-image de-raining with a high quality real rain dataset. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 12270–12279. [\[CrossRef\]](#)
36. Fu, X.; Huang, J.; Ding, X.; Liao, Y.; Paisley, J. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Trans. Image Process.* **2017**, *26*, 2944–2956. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Fu, X.; Huang, J.; Ding, X.; Liao, Y.; Paisley, J. Removing rain from single images via a deep detail network. In Proceedings of the IEEE Conference Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1715–1723. [\[CrossRef\]](#)
38. Chen, L.; Chu, X.; Zhang, X.; Sun, J. Simple baselines for image restoration. *arXiv* **2022**, arXiv:2204.04676. [\[CrossRef\]](#)
39. Zhang, K.; Liang, J.; Van Gool, L.; Timofte, R. PromptIR: Prompting for image restoration. *arXiv* **2023**, arXiv:2301.02127. [\[CrossRef\]](#)
40. Chen, X.; Li, M.; Ren, W.; Zhang, J.; Zhang, L.; Tao, D. Learning a sparse transformer network for effective image deraining. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 1190–1200. [\[CrossRef\]](#)
41. Chen, X.; Pan, J.; Dong, J. Bidirectional Multi-Scale Implicit Neural Representations for Image Deraining. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 25627–25636. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.