



ARTICLE

Efficient Image Deraining through a Stage-Wise Dual-Residual Network with Cross-Dimensional Spatial Attention

Tiantian Wang^{1,2} and Zhihua Hu^{3,*}

¹School of Computer and Software Engineering, SIAS University, Zhengzhou, 451150, China

²Henan Province Engineering Research Center for Intelligent Manufacturing and Digital Twin, Zhengzhou, 451150, China

³School of Computer Science and Artificial Intelligence, Huanggang Normal University, Huanggang, 438000, China

*Corresponding Author: Zhihua Hu. Email: huzhihua@hgnu.edu.cn

Received: 22 September 2025; Accepted: 29 October 2025; Published: 26 November 2025

ABSTRACT: Rain streaks introduced by atmospheric precipitation significantly degrade image quality and impair the reliability of high-level vision tasks. We present a novel image deraining framework built on a three-stage dual-residual architecture that progressively restores rain-degraded content while preserving fine structural details. Each stage begins with a multi-scale feature extractor and a channel attention module that adaptively emphasizes informative representations for rain removal. The core restoration is achieved via enhanced dual-residual blocks, which stabilize training and mitigate feature degradation across layers. To further refine representations, we integrate cross-dimensional spatial attention supervised by ground-truth guidance, ensuring that only high-quality features propagate to subsequent stages. Inter-stage feature fusion modules are employed to aggregate complementary information, reinforcing reconstruction continuity and consistency. Extensive experiments on five benchmark datasets (Rain100H, Rain100L, RainKITTI2012, RainKITTI2015, and JRSRD) demonstrate that our method establishes new state-of-the-art results in both fidelity and perceptual quality, effectively removing rain streaks while preserving natural textures and structural integrity.

KEYWORDS: Image deraining; stage-wise network; deep learning; feature aggregation; image processing

1 Introduction

Images captured on rainy days often contain a large number of rain streaks, which obscure important visual details and degrade overall image quality. This degradation not only affects human visual perception but also disrupts the performance of high-level computer vision tasks such as image classification [1], object detection [2], and image segmentation [3]. To address these issues, image deraining has attracted significant research interest in recent years [4–6]. The goal of image deraining is to recover a clean background from rain-affected images, an inherently challenging inverse problem due to the diverse patterns, intensities, and distributions of rain streaks. To address this inverse problem, which can be broadly categorized into two main types: video-based deraining algorithms [7] and single-image deraining algorithms [4,8]. Video-based methods can use information from adjacent frames for fusion, greatly reducing the difficulty of removing rain. In contrast, single-image deraining algorithms cannot leverage such temporal information, and the random distribution of rain streaks makes it difficult to determine their exact positions. Therefore, deraining from a single image has long been a particularly challenging task. Since a rainy video is essentially composed of multiple rainy frames, single-image deraining methods can be applied to each frame of a video to achieve



video deraining. As a result, single-image deraining has not only theoretical importance but also broad practical applications. Initial research efforts predominantly employed image prior-based approaches for single-image deraining. Kang et al. [9] utilized dictionary learning and sparse coding techniques to separate rain streaks from non-rain components within high-frequency image regions. Kim et al. [10] developed adaptive non-local mean filtering methods to reconstruct rain-free images. Li et al. [11] employed Gaussian mixture models to construct prior information, allowing separation of background and rain streak layers through predefined constraints. However, these traditional approaches suffer from fundamental limitations due to the inherent complexity of rain patterns and the diversity of background content, resulting in incomplete rain removal and suboptimal image quality.

The advancement of deep learning in computer vision has catalyzed significant progress in single-image deraining algorithms based on convolutional neural networks (CNNs). Ahn et al. [12] pioneered the application of the learning-based model to single-image deraining. Fu et al. [13] introduced a deep detail network that employs guided filtering to separate high-frequency details, subsequently utilizing deep residual networks to map these details to the layers of rain streaks for removal. Zhang and Patel [14] developed a rain density information acquisition network, incorporating rain density labels as guidance for more effective removal of rain streaks. Ren et al. [15] designed a progressive recurrent network based on residual architecture, composed of simplified network structures. Qian et al. [16] implemented adversarial generative networks with integrated attention mechanisms in the generator, combining attention maps with input images to produce rain-free output. Wei et al. [17] adapted CycleGAN's cycle consistency to image deraining, enhancing model consistency and stability through complementary transformation processes while achieving the first unsupervised deraining implementation. Huang et al. [18] proposed an effective semi-supervised deraining method that incorporates supervised and unsupervised data through memory storage mechanisms, and achieved superior deraining performance in transfer learning scenarios.

The aforementioned methods implement deraining tasks within single-scale network architectures. However, the complexity of rain patterns and background diversity in rainy images renders single-scale networks inadequate for achieving both thorough rain removal and high image quality simultaneously. Consequently, researchers have developed multi-scale network architectures for image deraining. Yang et al. [19] employed dilated convolutions to obtain receptive fields at different scales, constructing a multi-channel single-image deraining network that removes rain streaks through closed-loop feedback mechanisms across multiple stages. Nevertheless, this approach utilizes fixed weights and simple fusion strategies without considering the varying importance and contribution of different scale features across diverse scenarios, affecting the balance and stability of deraining performance. Fu et al. [20] utilized lightweight Laplacian pyramid networks to extract feature information at different scales, constructing recursive residual networks for rapid rain streak removal. While this model demonstrates advantages in parameter count and computational complexity, its deraining performance remains limited, failing to completely eliminate rain streaks from images. Jiang et al. [21] proposed a multi-scale progressive fusion network that constructs Gaussian pyramid structures through Gaussian kernel downsampling, followed by feature extraction and fusion for deraining tasks. However, this method inadequately considers the information correlations between different scales, resulting in feature redundancy and diminished deraining effectiveness. Wang et al. [22] replaced convolutional layers in U-Net with transformer modules, proposing the Uformer structure that enhances the model's ability to capture global dependencies, introducing transformers to image deraining for the first time and demonstrating their potential in this domain. However, Transformers typically rely on query-key mechanisms to obtain token information for attention-based feature aggregation, and token mismatches can interfere with subsequent image restoration clarity [23]. Due to these limitations in attention-based architectures, while most deraining methods focus on monocular inputs, recent work by Zhang et al. [24]

explored alternative approaches through stereo image deraining by leveraging semantic priors through a parallel stereo deraining network, demonstrating that joint semantic and multi-view fusion can significantly improve restoration quality.

Despite the excellent deraining performance achieved by most multi-scale algorithms, several critical issues persist: (1) Direct cross-scale information fusion without considering inter-scale relationships leads to background detail loss; (2) Low-scale information during upsampling processes readily causes aliasing effects [25], resulting in image distortion; (3) The varying importance of features at different scales requires careful fusion strategies to ensure information utilization while preventing redundancy, presenting an ongoing challenge for researchers.

Contributions:

- We propose a novel multi-stage dual-residual deraining network that progressively removes rain artifacts by decomposing the complex restoration task into tractable sub-problems across three sequential stages.
- A key innovation lies in the inter-stage feature fusion strategy, which integrates features across processing phases to preserve critical image details and enhance information continuity throughout the network.
- The core of our framework is a dual-residual reconstruction module with enhanced identity mappings, which improves feature propagation and mitigates degradation during deep learning.
- To effectively capture multi-scale rain streak features while mitigating the influence of extraneous information, we propose a cross-dimensional spatial attention module. This module augments the network's capacity to prioritize rain-affected regions, thereby enhancing reconstruction accuracy.
- Extensive experiments on multiple synthetic and real-world datasets demonstrate that our method achieves state-of-the-art performance, offering superior rain streak removal, better detail preservation, and improved outcomes in downstream vision tasks.

2 Related Work

Over the past decade, significant progress has been made in single-image deraining, and existing methods can be broadly categorized into two types: traditional algorithms and deep learning-based models.

2.1 Classic Deraining Algorithms Driven by Image Priors

Early image deraining methods based on hand-crafted priors typically relied on sparse coding [26], Gaussian models [27], and low-rank representations [28]. These approaches aimed to separate rain streaks from the background by exploiting statistical and structural assumptions about their appearance. Sun et al. [29] introduced a wavelet-based sparse optimization framework that modeled rain streaks using a directional multiscale transform. Their method incorporated convex optimization with dual sparsity regularizers: one capturing the directional sparsity of rain streaks, and another preserving background structure. Building on this, Deng et al. [30] proposed a global sparse model that included multiple sparsity constraints to account for both the directional nature of rain and background details. Luo et al. [31] addressed the limitations of synthetic rainy datasets by proposing a nonlinear screen blend model to simulate more realistic rain effects. They developed a dictionary-guided sparse coding scheme that emphasized mutual exclusivity between rain and clean image components, enabling more accurate separation. Similarly, Zhu et al. [32] observed that rain streaks typically exhibit consistent orientation. By analyzing local image gradients, they estimated dominant rain directions and proposed a sparsity-based model that alternates between removing rain and restoring background texture.

Other methods approached deraining from a statistical modeling perspective. Li et al. [11] utilized a patch-wise Gaussian Mixture Model (GMM) to capture the diverse orientations and scales of rain streaks

while preserving texture details in the background. Chen and Hsu [33] proposed a tensor-based low-rank representation that exploits the repetitive nature of rain streaks in space and time. Unlike earlier approaches, this method avoids pixel-wise operations and dictionary learning, offering improved efficiency for both image and video deraining. Chang et al. [34] further advanced low-rank modeling by analyzing rain patterns in the transform domain and employing directional total variation to distinguish rain from background layers.

While these traditional approaches have achieved promising results, they remain limited in several respects. Many methods struggle to fully remove all rain streaks, particularly when rain resembles background textures. In addition, the variability and ill-posed nature of real-world rain patterns challenge the generalizability of models based on Gaussian assumptions or dictionary learning. These methods also tend to be computationally intensive and require manual tuning of parameters, which reduces their scalability and practical applicability.

2.2 CNN-Based Image Deraining Methods

In recent years, CNNs have gained considerable attention for their adaptive learning capabilities, which allow them to automatically extract relevant features from input data, much like the human brain. These networks have achieved remarkable success in various computer vision tasks. As CNNs rapidly evolved, their strengths became increasingly evident compared to traditional methods, especially in addressing the limitations of conventional algorithms in image deraining. This has motivated researchers to explore CNN-based approaches for more effective rain removal. Fu et al. [35] are among the first to integrate CNNs with traditional techniques for image deraining. They decomposed rainy images into a low-frequency base layer and a high-frequency detail layer using a low-pass filter. A CNN is then used to extract background features from the detail layer, which is combined with an enhanced base layer to reconstruct the final derained image. Building on this idea, Deng et al. [36] later proposed an end-to-end deep residual network designed to learn the negative residual mapping between rainy and clean images. By focusing solely on rain streak features and excluding background content, their model simplified the learning process and reduced training complexity. Yang et al. [37] introduced a multi-task deep learning framework targeting challenging scenarios involving heavy rain and dense streaks. Their approach employed binary masks to localize rain streaks in both the rain layer and the background. A context aggregation network is used to expand the receptive field and capture broader contextual information. A recurrent rain detection and removal module then progressively identified and eliminated rain streaks. While effective in preserving background details, this method often led to distortions in color and contrast. To address uneven rain density, Liu et al. [38] developed a multi-stream densely connected network capable of both estimating rain density and removing streaks. The model utilized a residual-aware classifier to assign density labels to rainy images, which guided the network in removing corresponding streaks more accurately. Jiang et al. [39] are the first to consider the visual impact of rain on scene depth. They proposed a rain imaging model that incorporates both rain streaks and haze, and introduced a depth-guided attention mechanism to enhance the extraction of depth-related features. These features are then used to regress a residual map, enabling the removal of both streaks and haze. Zhang et al. [40] proposed DerainMamba, a locally enhanced visual state space model for image deraining, which integrates a global-aware state space mechanism, a direction-aware symmetrical scanning module, and a local-aware mixture of experts to effectively improve deraining performance while maintaining model efficiency. Recognizing the scarcity of real-world paired rainy and clean images, Wang et al. [41] proposed a semi-automatic method that combines temporal priors with human supervision. From sequential frames in real rainy videos, they generated clean images and trained a spatial attention network to identify and remove rain streaks in both local and global contexts. Gao et al. [42] introduced a model that addresses single-image

deraining by disentangling rain streaks through a dual-branch framework: a self-supervised macroscopic branch models rain statistics across low-frequency regions, while a supervised microscopic branch captures pixel-level rain characteristics, combined with an adversarial restoration module to preserve sharp edges.

Beyond the dominant trends of CNNs and Transformers, recent research has explored novel and highly efficient architectural components to achieve a more favorable trade-off between performance and computational cost. For instance, Cui et al. [43] introduced a universal image restoration network that leverages the inherent properties of standard pooling operations. By treating average pooling and max pooling as implicit extractors of low-frequency and high-frequency components, respectively, their method facilitates effective dual-domain (frequency and spatial) representation learning with minimal parameters, demonstrating strong performance across a wide range of degradation tasks. In a similar vein, the pursuit of efficiency is evident in domain-specific works such as EENet [44] for image dehazing. The model explicitly designs separate modules for frequency and spatial processing, followed by a dedicated dual-domain interaction module to enhance information exchange. This structured approach to spatial-spectral learning allows EENet to achieve state-of-the-art results in dehazing while maintaining strong generalization to other tasks like desnowing and low-light enhancement.

Although CNN-based deraining methodologies have demonstrated substantial performance improvements over conventional prior-based techniques, fundamental challenges persist. CNN architectures exhibit limited generalization capabilities when confronted with the heterogeneous and intricate distribution patterns characteristic of rain streak phenomena, frequently yielding residual visual artifacts in reconstructed imagery. Contemporary approaches commonly employ progressively deeper network topologies to enhance restoration quality; however, architectural complexity escalation can prove detrimental to overall system performance. Such depth augmentation introduces computational pathologies including gradient vanishing and explosion phenomena, elevated computational overhead, and compromised training convergence stability. Furthermore, prevailing methodologies rely predominantly on synthetic training corpora with constrained resolution specifications, thereby limiting their adaptability to diverse real-world imaging scenarios characterized by variable spatial resolutions and environmental conditions.

3 Proposed Methodology

The network consists of three sequential processing stages, each designed to progressively remove rain artifacts while preserving underlying image content. Each stage begins with a feature extraction module that identifies rain streak patterns and captures features at multiple scales from the input imagery. Following feature extraction, a Channel Attention (CA) module evaluates the importance of different feature channels, selectively emphasizing those that contribute most effectively to rain removal while suppressing less relevant information. The core restoration process is performed by a dual residual image reconstruction network, which employs enhanced residual connections to facilitate stable training and prevent information degradation. To ensure quality control during the restoration process, the first two stages incorporate a cross-dimensional spatial attention that uses ground-truth images to guide feature selection, allowing only features that contribute to improved restoration quality to advance to subsequent processing stages. A critical aspect of the proposed architecture is its information integration strategy across the three processing stages. Rather than employing simple sequential processing, the network incorporates feature fusion modules positioned between stages to combine information from different processing phases. This design ensures that important image details extracted in earlier stages are preserved and utilized throughout the entire restoration pipeline, preventing the gradual loss of fine-grained information that commonly occurs in deep network architectures. To improve the efficiency of the network, we have designed lightweight processing blocks that significantly reduce computational overhead without compromising performance. The network

utilizes lightweight feature extraction modules and cross-dimensional attention mechanisms with low computational cost, ensuring minimal resource consumption during both training and inference stages. Additionally, the stage-wise processing architecture avoids redundant computations, as each stage focuses on progressively refining the rain removal process without reprocessing already extracted features. This results in a more efficient network architecture capable of operating on devices with constrained resources, while maintaining high-quality deraining performance. The overall architecture of the proposed method is presented in Fig. 1.

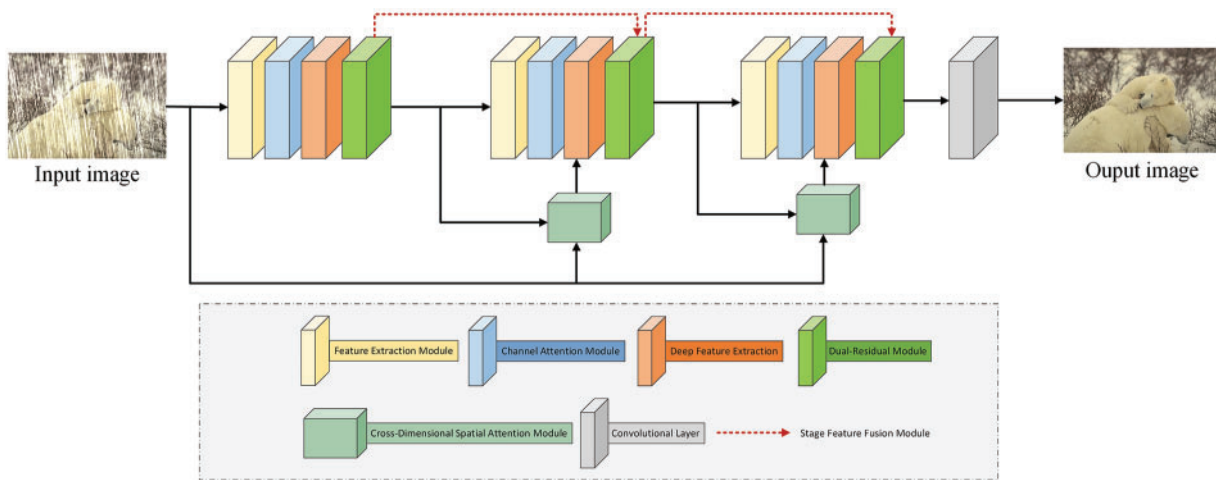


Figure 1: Graphical representation of stage-wise dual-residual architecture

3.1 Dual-Residual Module

Deep neural networks face fundamental challenges when layer depth increases excessively, leading to training difficulties and performance degradation rather than improved learning capacity [45]. To address these limitations while maintaining the benefits of deep architectures, we propose a dual residual module that enhances information flow through the network. As illustrated in Fig. 2, our architecture extends traditional residual connections by incorporating additional pathways between consecutive processing blocks, creating multiple routes for information propagation and reducing the risk of feature loss during training. The reconstruction network integrates conventional and dilated convolutions to achieve both local detail preservation and expanded spatial context awareness, following approaches established by Liu et al. [46]. The network incorporates SE-ResNet modules [47] to capture global feature relationships across the entire image. The complete architecture consists of six processing blocks organized into two distinct groups: three Dual-Residual Module (DRM) followed by three DRM with SE-ResNet (DRMS), with residual connections maintained from input to output across the entire network. The DRM implement a progressive refinement approach to rain streak detection and removal. Each DRM combines two standard convolutions with two dilated convolutions, where all standard convolutions employ 3×3 kernels for consistent local feature extraction. The dilated convolutions utilize decreasing dilation rates of 12, 8, and 6 across the three DRMs, respectively, enabling the network to progressively focus from broad spatial patterns to fine-grained rain streak details. This coarse-to-fine processing strategy allows for comprehensive rain pattern recognition across multiple spatial scales. The DRMS modules extend the basic dual residual design by incorporating SE-ResNet components at their final layers, enabling the network to integrate global spatial relationships with the locally processed features. These modules maintain the standard 3×3 convolution structure while adding global context awareness essential for distinguishing between rain artifacts and legitimate image textures. All

convolutions throughout the reconstruction network utilize ReLU activation functions to maintain training stability and computational efficiency. The complete module processes feature maps from the preceding extraction stage and forwards refined representations to subsequent network components.

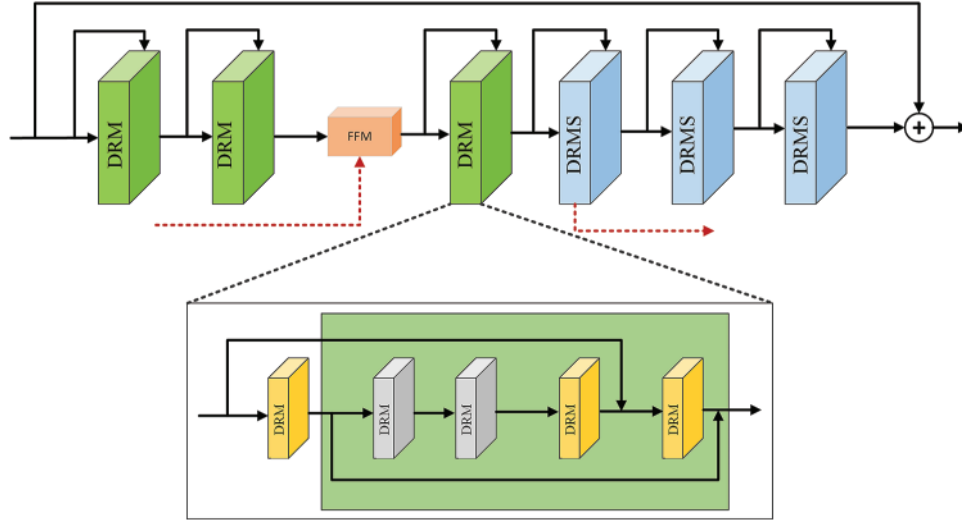


Figure 2: Visual representation of dual-residual module

3.2 Feature Extraction Module

Shallow network architectures, characterized by limited layer depth, exhibit constrained receptive fields that yield feature representations enriched with localized spatial information. This architectural property facilitates superior extraction of low-level visual characteristics, including textural patterns and fine-grained detail features. Conversely, deep network configurations possess expanded receptive fields that enable the capture of global contextual information, thereby extracting high-level semantic features such as structural contours and geometric shape representations. The proposed methodology incorporates a shallow feature extraction module at the initialization of each processing stage. The extracted shallow features are subsequently processed through a CA module before propagation to the deep feature extraction component for high-level feature computation. The resultant deep feature representations are then forwarded to the dual residual network architecture for further processing.

To extract shallow features, we apply a 3×3 convolutional layer at the input stage. This is followed by a CA, which enhances useful shallow features. For deeper feature extraction, we employ a residual network composed of 8 residual blocks. Each block consists of two convolutional layers and a ReLU activation function. The overall structure of the feature extraction module (Fig. 3) can be expressed as follows:

$$RB(x) = C(\delta(C(x))) \quad (1)$$

$$FEM(x) = RB_t(RB_{t-1}, RB_{t-2}, \dots, RB_0(x)) \quad (2)$$

Here, C denotes convolution, δ is the ReLU activation function, RB represents a residual block, and FEM denotes the feature extraction module. To effectively integrate features from the previous stage of the dual residual network, we introduce a Feature Fusion Module (FFM), illustrated in Fig. 4. This module receives feature maps from the second DRB and the first DRBs of the prior stage. Within the FFM, these two feature maps are first processed by individual CA modules, which assign weights to different channels to highlight important features. The outputs are then concatenated along the channel dimension and fused using a 1×1 convolution. ReLU activation is applied after each convolution operation in the FFM.

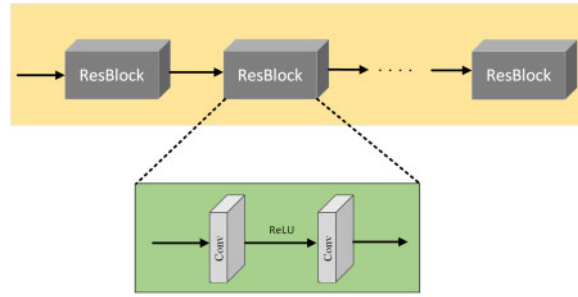


Figure 3: Overall structure of feature extraction module

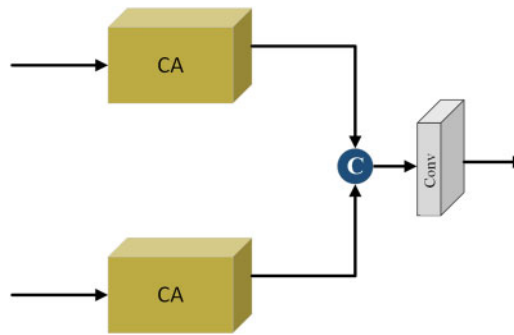


Figure 4: Overall structure of feature fusion module

Additionally, skip connections within the residual blocks are employed to pass information directly to deeper layers. This strategy mitigates the problem of vanishing gradients in deep networks and enhances the overall learning capacity of the model.

3.3 Cross-Dimensional Spatial Attention Module

To effectively extract rain streak features across multiple scales while suppressing irrelevant information, we propose a cross-dimensional spatial attention module (CDSAM), as illustrated in Fig. 5. This module comprises three parallel branches, each equipped with a convolutional layer followed by a Sigmoid activation function.

Following shallow feature extraction, the resulting feature map is processed through three parallel convolutional branches with kernel sizes of 1×1 , 3×3 , and 5×5 , respectively. Each branch is designed to capture spatial features at a different scale. The 1×1 convolution captures fine-grained details, the 5×5 convolution enlarges the receptive field to extract broader contextual information, and the 3×3 convolution, serving as the main branch, captures comprehensive spatial features.

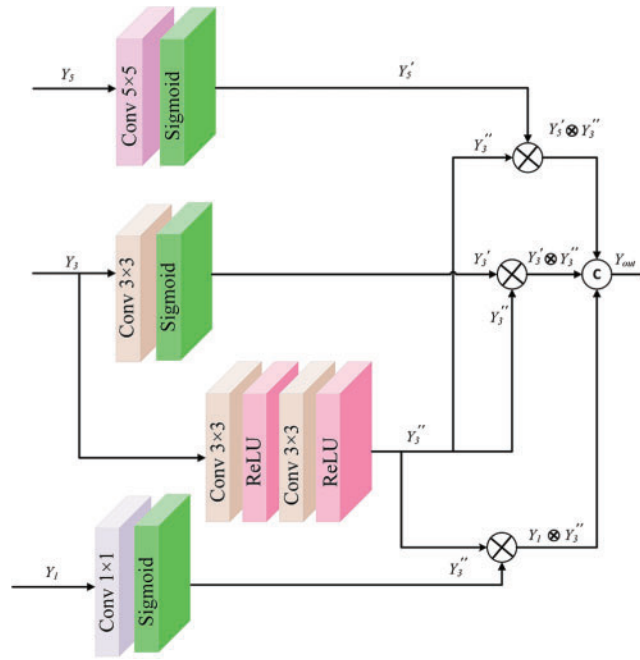


Figure 5: Architecture of cross-dimensional spatial attention module

Each branch output is activated using the Sigmoid function, defined as:

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}}. \quad (3)$$

To promote inter-scale correlation and reduce feature loss, we integrate larger-scale features into adjacent smaller-scale ones using element-wise multiplication. This operation expands the receptive field for lower-scale features while enhancing feature continuity across scales.

While convolutional operations are effective for extracting spatial features, relying solely on a single activation function may hinder convergence and introduce vanishing gradient issues. To address this, we introduce a ReLU activation function after the 3×3 convolution in the main branch. ReLU preserves positive activations and suppresses negative ones, facilitating feature sparsity and improving network generalization. The use of ReLU is governed by:

$$\text{ReLU}(x) = \max(0, x). \quad (4)$$

The main branch combines convolution and ReLU (Conv & ReLU) to reduce feature dimensionality, ease optimization, and enhance feature expressiveness by employing convolutional operations identical in size to those in the main branch. This increases feature dimensionality consistently across the three parallel paths.

To further refine feature modulation, we combine ReLU and Sigmoid activations. This hybrid strategy suppresses non-informative features and enhances weight modulation. The joint activation function is defined as:

$$\text{ReLU} \odot \text{Sigmoid} = \begin{cases} 0, & x < 0, \\ \frac{1}{1 + \exp(-x)}, & x \geq 0, \end{cases} \quad (5)$$

where \odot denotes the composite application of both activation functions.

The final output of the module is computed as:

$$Y_{\text{out}} = \text{Cat}(Y'_5 \otimes Y''_3, Y'_3 \otimes Y''_3, Y'_1 \otimes Y''_3), \quad (6)$$

where \otimes denotes element-wise multiplication, and $\text{Cat}(\cdot)$ represents channel-wise concatenation. Here, Y'_1 , Y'_3 , and Y'_5 are the outputs from the 1×1 , 3×3 , and 5×5 branches after Conv & ReLU, respectively, and Y''_3 is the output of the 3×3 branch after Conv & ReLU.

3.4 Loss Function

A commonly used loss function in image restoration tasks is the Mean Squared Error (MSE). However, due to its quadratic penalization of errors, MSE often leads to overly smoothed outputs, which suppress fine image details—an undesirable effect for image deraining, where preserving structural clarity is critical.

To overcome this limitation and enhance perceptual quality, we adopt the Structural Similarity Index (SSIM) as our loss function. SSIM aligns better with human visual perception by comparing luminance, contrast, and structural information between the predicted derained image and the corresponding ground truth.

The SSIM between a predicted image X and the ground truth image Y is defined as:

$$\text{SSIM}(X, Y) = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)}, \quad (7)$$

where μ_X and μ_Y are the mean intensities of X and Y , σ_X^2 and σ_Y^2 denote the variances, and σ_{XY} is the covariance between X and Y . The constants $C_1 = 0.01$ and $C_2 = 0.03$ are introduced to avoid division by zero.

To guide training, we define the overall loss function as the negative sum of SSIM values computed at each stage of the multi-stage deraining network. The total loss is given by:

$$L = - \sum_{s=1}^3 \text{SSIM}(X_s, Y), \quad (8)$$

where s indicates the stage index, X_s denotes the derained image output from the s -th stage, and Y is the ground truth clean image. This formulation encourages the network to produce structurally consistent and perceptually sharper outputs across all stages.

4 Experimental Results

4.1 Network Configuration

The proposed network architecture is implemented using the PyTorch deep learning framework. All experiments are conducted on a computational platform equipped with an Intel Core i9-10700K processor operating at 5.1 GHz, 32 GB of RAM, and a single NVIDIA GeForce RTX 3090 GPU running on Windows 11 operating system. For data preprocessing, training images underwent random cropping augmentation, resulting in input tensors of dimensions 128×128 pixels. The model optimization is performed using the ADAM optimizer with momentum coefficients $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate is initialized at 2×10^{-4} and decreased to 1×10^{-6} following a cosine annealing schedule over 1000 training iterations. A mini-batch size of 4 is employed throughout the training process.

4.2 Datasets

Training effective image deraining models requires paired datasets containing corresponding rainy and clean images. However, acquiring sufficient real-world paired data for training purposes presents significant practical challenges. To address this limitation, researchers commonly employ synthetically generated image pairs as training data. Our experimental evaluation utilizes five established synthetic datasets: Rain100L [19], Rain100H [19], RainKITTI2012 [48], RainKITTI2015 [48], and JRSRD [49]. These datasets provide complementary characteristics for comprehensive model training and assessment. Rain100L contains images with single-direction rain streaks, while Rain100H incorporates rain patterns oriented in five distinct directions. The RainKITTI datasets simulate complex precipitation scenarios including raindrop interference on camera lenses, motion trajectories of falling rain, and varying raindrop densities. JRSRD extends beyond traditional rain streak simulation by incorporating both streak patterns and raindrop effects that occur simultaneously in natural rainfall conditions. While real precipitation rarely exhibits the multi-directional streak patterns and diverse raindrop configurations present in these synthetic datasets, training on such varied artificial conditions enhances network robustness and generalization capabilities. This approach enables models to handle a broader spectrum of precipitation scenarios than would be possible with naturally occurring data alone.

To evaluate the generalization performance of our trained network on authentic scenarios, we employ two real-world datasets: Raindrop [16] and a newly constructed real-world rainy images dataset. The Raindrop dataset provides genuine images containing natural raindrop artifacts, serving as a valuable resource for assessing raindrop detection and removal capabilities. Our real-world rainy images dataset encompasses diverse authentic precipitation scenarios captured under varying environmental conditions, offering comprehensive coverage of naturally occurring rainy situations. These real-world datasets present the inherent complexity and variability found in authentic precipitation scenarios, providing a more realistic and challenging evaluation environment compared to synthetic alternatives. This diversity enables thorough assessment of algorithm adaptability across different practical scenarios, ultimately demonstrating improved performance and broader applicability in real-world deraining applications.

4.3 Quantitative Metrics

Quality assessment constitutes a fundamental aspect of image processing methodologies, offering quantitative measures to evaluate algorithmic performance in maintaining or enhancing image fidelity and perceptual quality. Hence, the performance evaluation of deraining methods employs three established quantitative metrics: PSNR [50], SSIM [51], and FSIM [52]. All quantitative assessments are computed using the luminance channel to ensure consistent evaluation across different methods. These metrics provide complementary perspectives on reconstruction quality. PSNR measures pixel-level accuracy by quantifying the difference between predicted and ground truth images, while SSIM evaluates perceptual similarity by considering luminance, contrast, and structural information. FSIM extends this assessment by incorporating human visual system characteristics through feature-based similarity computation. Typically, elevated values of SSIM, FSIM, and PSNR correspond to enhanced visual fidelity and superior reconstruction quality in the processed outputs.

4.4 Qualitative Assessment

4.4.1 Visual Assessment of Synthetic Rain Effects

In this section, we present qualitative comparisons between proposed and state-of-the-art deraining methods using sample results from the Rain100L and Rain100H datasets as shown in Fig. 6. The evaluation

reveals three key findings. Firstly, compared to the conventional DSC [31], our model demonstrates superior rain streak removal while maintaining better image fidelity. Secondly, when evaluated against recent methods, our approach shows significant advantages in preserving background details without introducing unnatural artifacts commonly observed in competing methods. Most notably, in comparison with LSN [53], proposed model exhibits enhanced performance in texture preservation. As shown in the third test case, our method better maintains rock textures and vegetation details. These visual comparisons collectively demonstrate that our method achieved two critical improvements: (1) more complete rain streak removal, and (2) superior preservation of fine image details, resulting in cleaner outputs with more natural appearance.

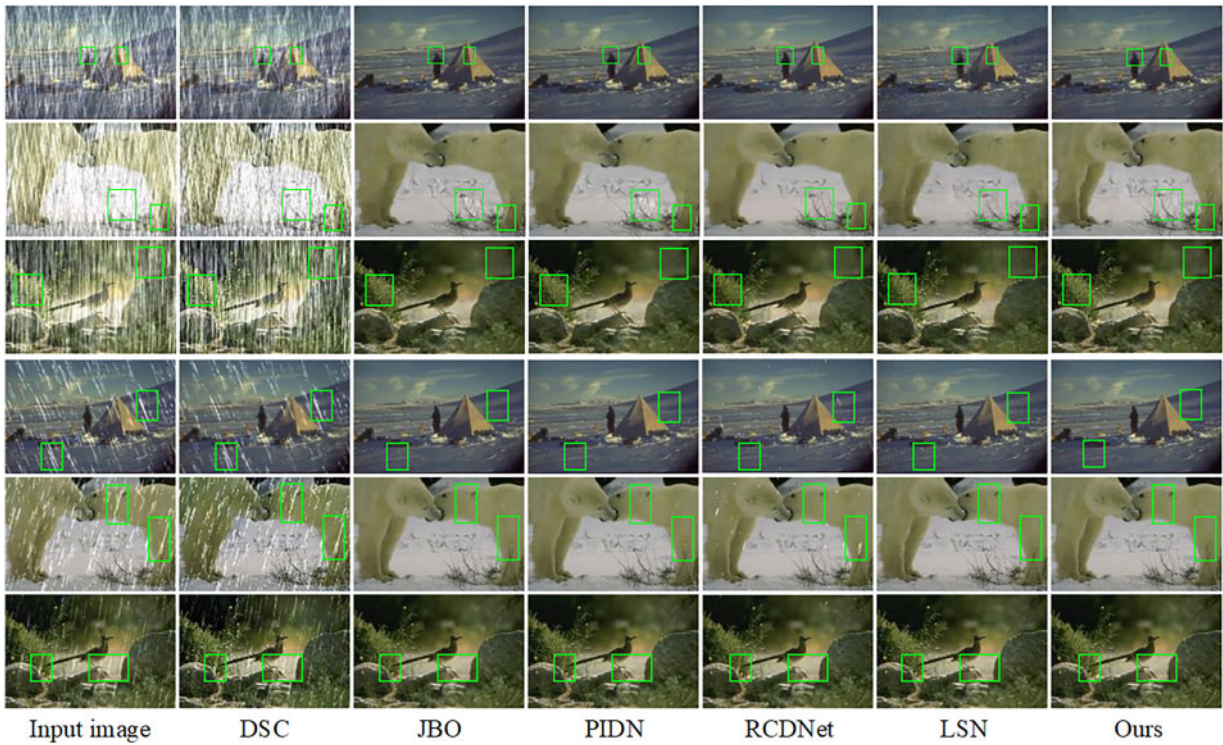


Figure 6: Comparative visual results of the proposed model and prior methods on Rain100L and Rain100H datasets

Furthermore, Fig. 7 presents a comparative analysis of synthetic rain image deraining performance between the proposed algorithm and established methods including RESCAN [47], JBO [32], PIDN [15], and RCDNet [54]. The proposed method demonstrates superior adaptability across varying rain streak distributions, effectively processing both densely distributed rain patterns in the second test image and sparsely distributed streaks in the third image. In contrast, existing algorithms such as JBO [32] exhibit limited effectiveness when confronted with diverse rain streak distributions. Additional analysis of the seventh test image reveals that the proposed algorithm achieved superior discrimination between rain artifacts and background content, resulting in more thorough rain streak removal compared to competing methods. Notably, while alternative algorithms introduce varying degrees of blur artifacts in structural elements such as horse legs in the second test image, the proposed method preserves fine-scale image details effectively, producing deraining results that more closely approximate the ground truth images. These experimental observations demonstrate that the proposed algorithm achieves comprehensive rain streak removal across diverse precipitation patterns while maintaining enhanced detail preservation, resulting in superior visual quality compared to existing methodologies.



Figure 7: Comparative visual results of the proposed model and prior methods on datasets [14,35,55]

4.4.2 Visual Assessment of Real Rain Images

To verify the deraining performance of proposed model on real rainy images, comparison of recent prior algorithms are tested on real-world datasets. Several representative comparison results are selected from the test data for analysis. Fig. 8 shows the comparison results in light rain scenarios. Among the various algorithms, only proposed and SIDBRN [56] completely removed the rain streaks on the car body surface in the locally enlarged area. However, like TRNR [57] RWSRN [58], and HCN [59] also lost the white detail of the bicycle's front wheel hub on the left side of the image. Only DRR-Net is able to more thoroughly remove the rain streaks on the car body while effectively preserving the details of the bicycle wheel hub. In the second test image, the texture of the wooden board and the rain streaks have a similar appearance. Although SIDBRN [56] and VSSM [40] are able to retain some of the texture details of the board, a considerable amount of rain streaks remained in the output image. TRNR [57] and PHMN [60] severely damaged the board's texture structure. RWSRN [58] removed some rain streaks but also lost much of the wooden board's texture details. Our model outperformed the comparison algorithms in deraining performance and preserved most of the texture details.



Figure 8: Visual results analysis by proposed and prior deraining models using real-world rainy dataset

To further evaluate the deraining capability of the proposed model on real-world rainy images, comparative results from a real-world rainy dataset are presented in Fig. 9. As shown in the figure, SIDBRN [56] removed only a limited portion of the rain streaks, while HCN [59] and RWSRN [58] significantly degraded the color fidelity of the output image compared to the original. Additionally, TRNR [57], SIDBRN [56], and VSSM [56] are unable to completely eliminate the rain streaks within the magnified region. In contrast, our approach exhibits superior performance, effectively removing the rain streaks while preserving image quality in both test images in the given figure.

4.5 Quantitative Analysis

This section presents a comprehensive experimental framework designed to provide objective performance assessment of the proposed model through systematic comparison with established methodologies. The evaluation employs SSIM, PSNR, and FSIM as quantitative metrics, selected for their widespread adoption and demonstrated effectiveness in measuring image processing quality. Table 1 presents a quantitative comparison of various state-of-the-art deraining methods on the Rain100H and Rain100L datasets using PSNR, SSIM, and FSIM. Classical models such as DSC [31] and JBO [32] demonstrate limited capability in handling complex rain patterns, achieving low performance with PSNR values of 15.25 and 14.50, and SSIM scores of 0.388 and 0.396 on Rain100H, respectively. These models also underperform in FSIM, highlighting their inability to preserve structural information. With the advent of deep learning, significant improvements are observed. For example, Fu et al. [35] improved performance to 22.91 PSNR and 0.729

SSIM on Rain100H. RESCAN [47] and PIDN [15] further enhance results, achieving 28.93 and 29.52 PSNR and SSIM values of 0.868 and 0.902, respectively. On the simpler Rain100L dataset, these methods show strong results, with RESCAN and PIDN achieving 36.95 and 37.05 PSNR, and SSIM values of 0.976 and 0.980. Further advancements are seen in models such as RCDNet [54], EAGNet [12], and LSN [53]. Notably, EAGNet [12] delivers 32.81 PSNR and 0.928 SSIM on Rain100H, and reaches 41.86 PSNR and 0.990 SSIM. While newer models like RRSI [61] exhibit some variability, it still performs competitively, attaining 28.71 PSNR and 0.890 SSIM, and 36.89 PSNR and 0.980 SSIM. Recently introduced methods such as CCAL [62], MFFDNet [63], NSB [64], and UC-Former [65] show further improvements, with CCAL achieving 34.60 PSNR and 0.950 SSIM, and 41.80 PSNR and 0.992 SSIM. MFFDNet reports 35.10 PSNR and 0.953 SSIM, reaching 42.20 PSNR and 0.993 SSIM. NSB performs with 34.80 PSNR and 0.952 SSIM, and 42.10 PSNR and 0.993 SSIM. UC-Former shows 34.90 PSNR and 0.954 SSIM, and 42.30 PSNR and 0.994 SSIM. The proposed model surpasses all previous approaches across all metrics. It achieved a PSNR of 32.12, SSIM of 0.952, and FSIM of 0.947 on Rain100H, while delivering 41.95 PSNR, 0.995 SSIM, and 0.994 FSIM on Rain100L. These results clearly demonstrate the proposed model's superior capability in reconstructing high-quality images while preserving structural and perceptual fidelity under varying rain conditions.

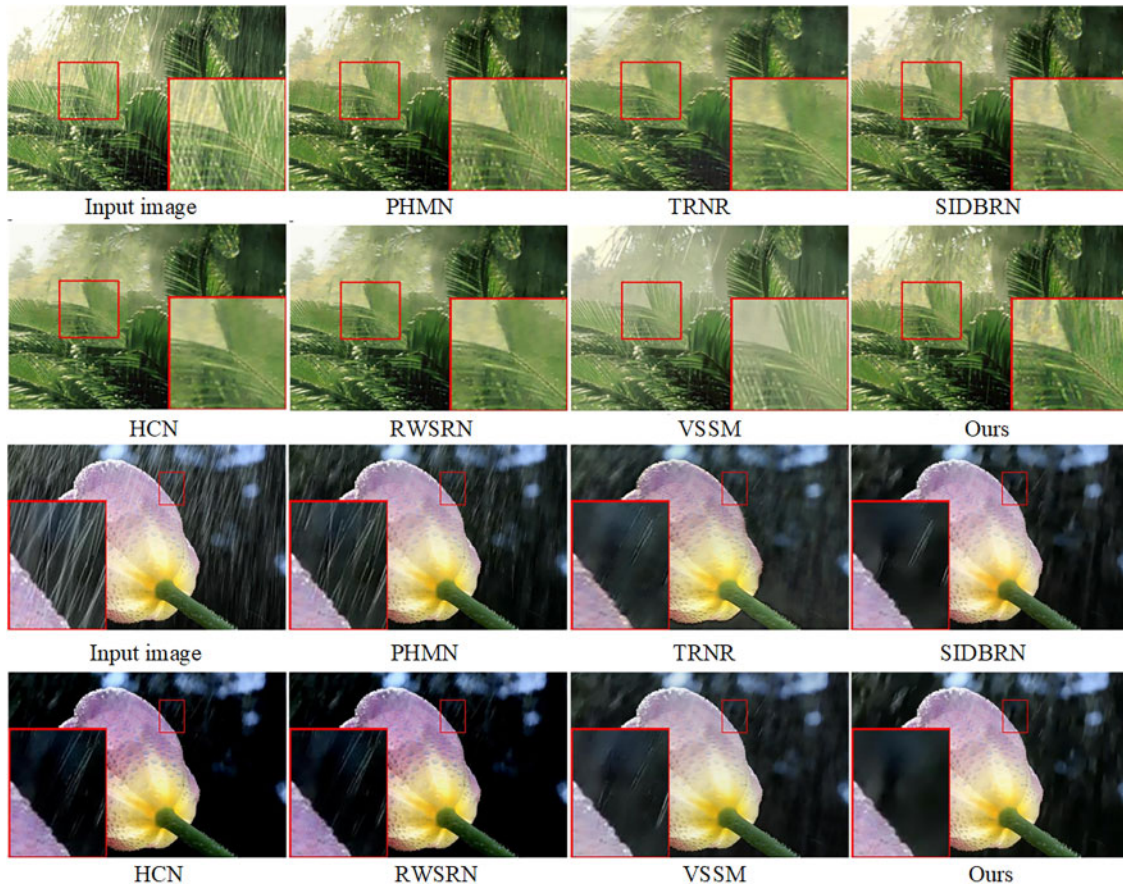


Figure 9: Visual assessment of proposed and several prior models using real rainy images

Table 1: Quantitative comparison of deraining methods on the Rain100H and Rain100L datasets

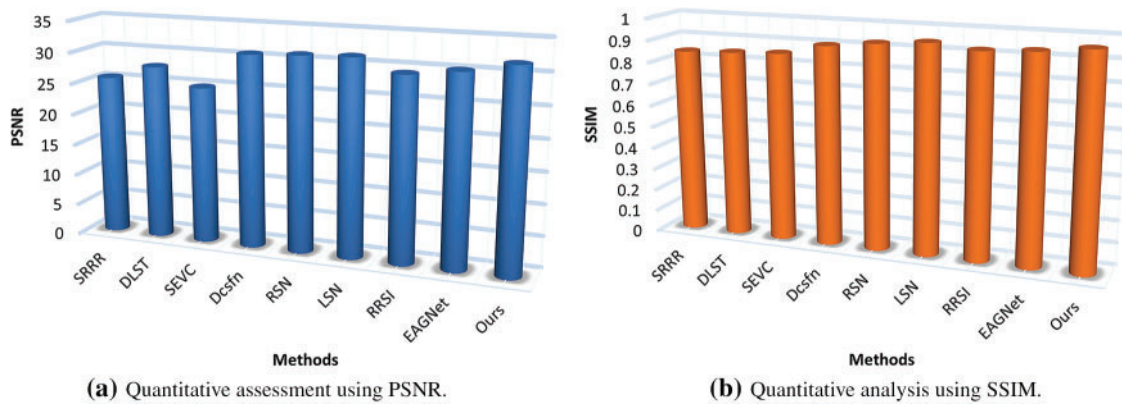
Methods	Year	Rain100H			Rain100L		
		PSNR	SSIM	FSIM	PSNR	SSIM	FSIM
DSC [31]	2015	15.25	0.388	0.621	28.73	0.869	0.891
JBO [32]	2017	14.50	0.396	0.624	27.82	0.871	0.874
Fu et al. [35]	2017	22.91	0.729	0.822	32.45	0.929	0.941
RESCAN [47]	2018	28.93	0.868	0.886	36.95	0.976	0.977
PIDN [15]	2019	29.52	0.902	0.914	37.05	0.980	0.982
SPANet [41]	2019	25.18	0.837	0.860	35.40	0.971	0.956
RCDNet [54]	2020	31.34	0.911	0.928	38.91	0.986	0.988
EAGNet [12]	2021	32.81	0.928	0.941	41.86	0.990	0.991
RRSI [61]	2023	28.71	0.890	0.916	36.89	0.980	0.981
RWSRN [58]	2023	22.47	0.709	0.809	33.50	0.940	0.947
LSN [53]	2023	30.48	0.893	0.892	36.48	0.967	0.924
CCAL [62]	2024	34.60	0.950	0.940	41.80	0.992	0.993
MFFDNet [63]	2024	35.10	0.953	0.944	42.20	0.993	0.994
NSB [64]	2024	34.80	0.952	0.945	42.10	0.993	0.994
UC-Former [65]	2024	34.90	0.954	0.946	42.30	0.994	0.995
Ours	–	36.00	0.960	0.950	43.00	0.995	0.996

Furthermore, to evaluate the effectiveness of the proposed method in enhancing the visibility of rainy images, we conducted performance assessments on two benchmark datasets: RainKITT2012 and RainKITT2015. The quantitative results, presented in Table 2, demonstrate that our approach significantly improves image visibility under rainy conditions. Traditional methods such as DSC [31] and JBO [32] exhibit limited performance, with PSNR values below 20 and relatively low SSIM scores on both datasets that indicate their difficulty in preserving image structure under challenging weather conditions. With the emergence of deep learning, approaches such as Fu et al. [35], RESCAN [47], and PIDN [15] show substantial improvements. For instance, Fu et al. achieved over 31 PSNR and SSIM scores above 0.96, while RESCAN reaches 34.52 PSNR and 0.978 SSIM on RainKITT2012, and 35.74 PSNR and 0.983 SSIM on RainKITT2015. Similarly, PIDN attains 34.48 and 35.69 PSNR with SSIM values nearing 0.99. Recent advancements continue this trend. Models like Dcsfn [66], EAGNet [12], and RRSI [61] demonstrate robust deraining capabilities. Notably, RRSI achieves 35.49 PSNR and 0.986 SSIM on RainKITT2012, and 36.91 PSNR on RainKITT2015. RWSRN [58] also performs strongly, delivering 38.33 PSNR and 0.990 SSIM on RainKITT2015. Our proposed method surpasses all existing techniques across all three metrics on both datasets. Specifically, it achieves the highest PSNR of 35.97 and SSIM of 0.988 on RainKITT2012, and 38.63 PSNR and 0.995 SSIM on RainKITT2015. Furthermore, it records the best FSIM values of 0.991 and 0.996, confirming its superior capability in restoring both perceptual and structural fidelity. These results underscore the model's effectiveness in enhancing visual quality under adverse weather conditions, an essential feature for downstream tasks in autonomous driving systems.

Table 2: Quantitative evaluation of deraining performance on RainKITTI2012 and RainKITTI2015 datasets

Methods	Year	RainKITTI2012			RainKITTI2015		
		PSNR	SSIM	FSIM	PSNR	SSIM	FSIM
DSC [31]	2015	19.24	0.683	0.793	19.81	0.693	0.811
JBO [32]	2017	18.91	0.679	0.793	20.85	0.699	0.823
Fu et al. [35]	2017	31.82	0.968	0.973	31.56	0.965	0.976
RESCAN [47]	2018	34.52	0.978	0.984	35.74	0.983	0.987
PIDN [15]	2019	34.48	0.978	0.980	35.69	0.987	0.990
SPANet [41]	2019	33.16	0.975	0.983	34.47	0.974	0.980
Dcsfn [66]	2020	33.44	0.978	0.987	34.77	0.979	0.988
EAGNet [12]	2021	32.15	0.968	0.975	33.51	0.978	0.985
RWSRN [58]	2023	35.81	0.985	0.989	38.33	0.990	0.994
RRSI [61]	2023	35.49	0.986	0.989	36.91	0.985	0.989
LSN [53]	2023	32.72	0.976	0.979	34.79	0.977	0.984
VSSM [40]	2024	35.42	0.970	0.986	35.27	0.978	0.988
Ours	–	35.97	0.988	0.991	38.63	0.995	0.996

Moreover, rain streaks and raindrops are distinct natural phenomena that degrade image visibility in different ways. Most existing deep learning-based deraining methods treat these two problems separately, failing to address their combined effects effectively. To evaluate the performance of our proposed in enhancing visibility under raindrop-distorted conditions, we conducted experiments on the Raindrop dataset. As demonstrated in Fig. 10, the quantitative results confirm that our method significantly outperforms existing approaches in raindrop removal while maintaining superior image quality.

**Figure 10:** Comparative evaluation on the Raindrop dataset

In real-world rainy conditions, rain streaks and raindrops typically coexist, presenting a complex challenge for image restoration. To evaluate our method's effectiveness in handling this combined degradation, we assessed our model's performance on the JRSRD dataset. The quantitative results, presented in Fig. 11, demonstrate our approach's capabilities in simultaneous rain streak and raindrop removal. Based on the graph data, a clear progression in performance across deraining methods is observed when evaluated using PSNR and SSIM metrics. Traditional or earlier methods such as SRRR [67], DLST [3], and SEVC [68] show

comparatively lower PSNR scores and SSIM values, indicating limited restoration quality and structural fidelity. Mid-tier methods like Dcsfn [66] and RSN [69] offer moderate improvements, while more recent approaches such as LSN [53], RRSI [61], and EAGNet [12] demonstrate significant gains in both metrics, with PSNR values nearing 28 dB and SSIM exceeding 0.92. The proposed method outperforms all baselines, achieving the highest PSNR of 28.55 and SSIM of 0.942, indicating superior detail preservation and structural accuracy in the derained outputs. These results validate the effectiveness and robustness of the proposed approach in enhancing image quality under challenging rain conditions.

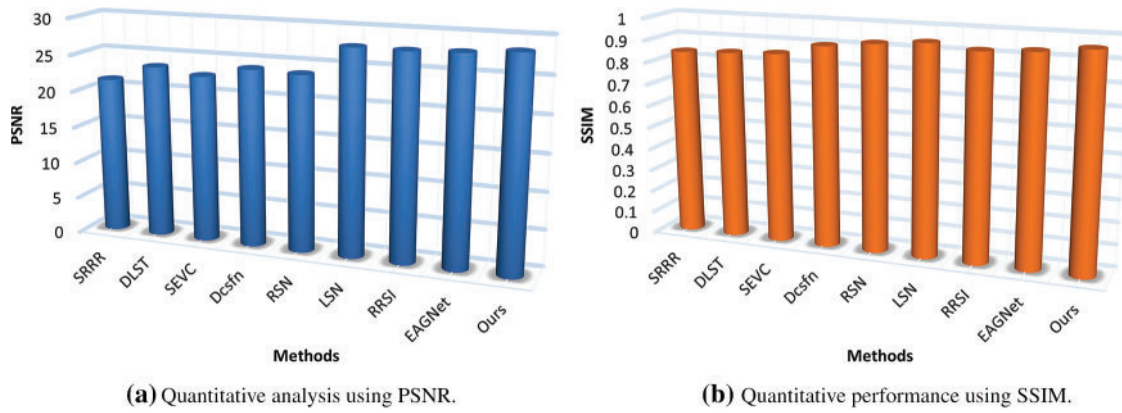


Figure 11: Comparative evaluation on the JRSRD dataset

4.6 Ablation Study

To evaluate the contribution and effectiveness of each component in the proposed algorithm, comprehensive ablation studies are conducted on the Rain100H dataset. These experiments systematically assess the influence of key design choices, including the number of deraining stages, various combinations of network modules, and the choice of loss functions. The details are provided below.

4.6.1 Number of Deraining Stages

In this section, we investigate the influence of the number of deraining stages on the performance of the proposed. The evaluation is carried out on the Rain100H dataset using two quantitative metrics: PSNR and SSIM. As shown in Fig. 12, increasing the number of stages from one to three results in a consistent and substantial enhancement in both metrics. Specifically, PSNR improves from 31.40 to 33.45, while SSIM increases from 0.904 to 0.939, indicating improved image fidelity and structural detail preservation. The most significant performance gain occurs when progressing from a single-stage to a two-stage architecture, highlighting the effectiveness of multi-stage feature refinement. In contrast, extending the network to four stages yields only marginal improvements. Hence, PSNR increases slightly to 33.50 and SSIM to 0.940 that demonstrating diminishing returns. Given the increased computational burden and model complexity associated with additional stages, the four-stage configuration provides limited practical advantage. Therefore, a three-stage structure is adopted as the optimal balance between deraining performance and computational efficiency.

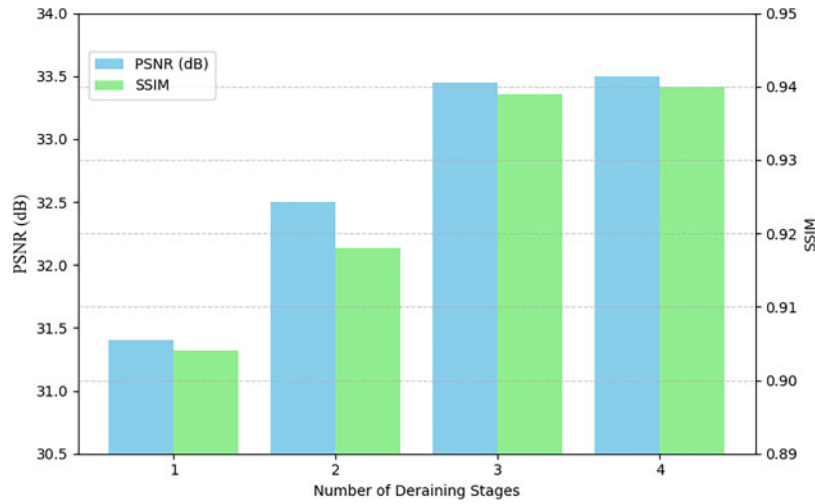


Figure 12: Performance analysis of the proposed model with respect to the number of deraining stages

4.6.2 Analysis of Loss Functions

The choice of loss function significantly impacts the performance of image deraining models. To validate our multi-stage SSIM loss, we conduct an ablation study comparing it with other common losses, measuring both reconstruction accuracy (PSNR) and perceptual quality (SSIM). As shown in Table 3, our method achieves the best results by enforcing structural consistency across all network stages. While MSE leads to oversmoothing (32.94 PSNR) and single-stage SSIM lacks hierarchical refinement (0.941 SSIM), our multi-stage SSIM loss outperforms hybrid alternatives (e.g., L1+SSIM) with a PSNR of 34.12 and SSIM of 0.958. This demonstrates that progressive SSIM optimization is critical for preserving fine details in derained images.

Table 3: Ablation study on loss functions for rain removal with computational overhead analysis

Loss function	PSNR (dB) ↑	SSIM ↑	Ave. inf. time (s) ↓
L2 (MSE)	32.94	0.924	0.182
Single-stage SSIM	33.43	0.941	0.195
Stage-wise SSIM (ours)	34.12	0.958	0.218
L1 + SSIM	33.65	0.943	0.197
Charbonnier + SSIM	33.70	0.945	0.201

Note: We obtain the average inference time of deraining on images with size of 512×384 . All models share the same architecture; only loss functions differ during training.

4.6.3 Module-Wise Performance Evaluation

An ablation study is conducted to systematically assess the contribution of each component within the proposed network architecture. Configuration 1, which includes only the basic DRM, establishes the baseline performance. Subsequent configurations progressively incorporate critical modules such as

the CA, CDSAM, deeper feature extraction layers, skip connections, and cross-stage fusion. The results demonstrate a consistent improvement in both PSNR and SSIM with the integration of these components. Notably, Configuration 4, which includes all components except cross-stage fusion, achieves a substantial performance boost, underscoring the importance of attention-based feature enhancement. The full model (Configuration 8), which combines all components, yields the highest performance with a PSNR of 33.58 and an SSIM of 0.944, validating the synergistic effect of the complete architecture. These findings highlight the critical role of each module and justify their inclusion in the final network design.

A detailed analysis of Table 4 provides valuable insights into the interactions between the network components. For example, Configuration 5, which lacks deep layers, exhibits a significant performance decline (PSNR: 30.08), highlighting that shallow networks are insufficient for effectively removing complex rain patterns. Similarly, Configuration 6, which uses only 16 deep layers, underperforms compared to Configurations 4 and 8, suggesting that 32 layers offer an optimal balance between model capacity and overfitting. Configuration 7, which substitutes the proposed DRM with RDB, yields lower performance (PSNR: 30.48), indicating that the DRM design is more suitable for deraining tasks. Importantly, Configuration 4 demonstrates strong performance without cross-stage fusion (CSF), underscoring the effectiveness of CA and CDSAM in refining features within individual stages. However, the full model (Configuration 8) outperforms Configuration 4 by +0.13 PSNR, confirming that CSF facilitates the exchange of complementary information across stages, thereby enhancing coarse-to-fine prediction refinement. This synergistic interaction, where attention mechanisms improve local feature representation, deep layers capture hierarchical context, and CSF propagates multi-scale cues, collectively drives the superior performance of the final architecture.

Table 4: Ablation study on component combinations

Config.	CA	CDSAM	Deep layers	Skip conn.	Rest. module	CSF	PSNR (dB)	SSIM
1	–	–	–	–	DRM	–	28.03	0.847
2	–	✓	32	✓	DRM	✓	32.18	0.911
3	✓	–	32	✓	DRM	✓	32.46	0.925
4	✓	✓	32	✓	DRM	–	33.55	0.939
5	✓	✓	–	✓	DRM	✓	30.08	0.886
6	✓	✓	16	✓	DRM	✓	31.74	0.905
7	✓	✓	32	✓	RDB	✓	30.48	0.894
8 (Proposed)	✓	✓	32	✓	DRM	✓	33.68	0.944

5 Limitations

Despite the strong performance demonstrated by our proposed deraining framework, certain limitations remain. The current model is primarily trained and evaluated on datasets that predominantly feature linear and sparse rain streaks. As a result, while the network excels at removing rain streaks, its effectiveness diminishes when applied to more complex rainy conditions, such as dense rain mist or occluding raindrops on the camera lens. These scenarios introduce additional degradation patterns, such as scattering and blur that are not adequately represented in existing training datasets. Consequently, the model struggles to generalize in such cases.

This limitation highlights the need for more diverse and realistic data, including high-quality real-world images or synthetic datasets that better simulate the full spectrum of rain types. Addressing this gap will be an important direction for future research to enhance the model's robustness in challenging weather conditions.

6 Conclusion

In this work, we proposed a novel multi-stage dual-residual image deraining framework that effectively decomposes the complex task of rain streak removal into progressive sub-stages. By incorporating a dual-residual reconstruction design and an inter-stage feature fusion strategy, the network ensures stable feature propagation and minimizes the loss of critical image details during training. To further enhance rain streak localization and suppression, we introduced a cross-dimensional spatial attention module, enabling the model to focus on rain-affected regions across multiple scales. Extensive experiments on both synthetic and real-world datasets validate the effectiveness of our method, which consistently outperforms existing approaches in terms of both visual quality and quantitative metrics. Moreover, our results confirm that the proposed deraining model can serve as an effective preprocessing step for downstream tasks in computer vision, significantly enhancing detection accuracy.

While the proposed framework shows promising results, there is still room for further improvement. In future work, we aim to explore lightweight architectures for real-time deployment on resource-constrained devices, such as mobile phones or embedded systems. Additionally, integrating unsupervised or self-supervised learning strategies could help reduce dependence on paired training data, making the model more adaptable to diverse and unstructured environments.

Acknowledgement: None.

Funding Statement: This project is supported by Key Scientific and Technological Research Program of Henan Province (Grant No. 252102211111).

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Tiantian Wang, Zhihua Hu; data collection and analysis: Tiantian Wang; interpretation of results: Zhihua Hu; draft manuscript preparation: Tiantian Wang, Zhihua Hu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The authors will make the data available upon request to the corresponding author.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Obeso AM, Benois-Pineau J, Vázquez MSG, Acosta AA.R. Visual vs internal attention mechanisms in deep neural networks for image classification and object detection. *Pattern Recognit.* 2022;123:108411. doi:10.1016/j.patcog.2021.108411.
2. Sharma T, Debaque B, Duclos N, Chehri A, Kinder B, Fortier P. Deep learning-based object detection and scene perception under bad weather conditions. *Electronics.* 2022;11(4):563. doi:10.3390/electronics11040563.
3. Yang Q, Yu M, Xu Y, Cen S. Single image rain removal based on deep learning and symmetry transform. *Symmetry.* 2020;12(2):224. doi:10.3390/sym12020224.
4. Li S, Ren W, Wang F, Araujo IB, Tokuda EK, Junior RH, et al. A comprehensive benchmark analysis of single image deraining: current challenges and future perspectives. *Int J Comput Vis.* 2021;129(4):1301–22. doi:10.1007/s11263-020-01416-w.

5. Du S, Liu Y, Zhao M, Shi Z, You Z, Li J. A comprehensive survey: image deraining and stereo-matching task-driven performance analysis. *IET Image Process.* 2022;16(1):11–28. doi:10.1049/ipr2.12347.
6. Su Z, Zhang Y, Shi J, Zhang XP. A survey of single image rain removal based on deep learning. *ACM Comput Surv.* 2023;56(4):1–35.
7. Yue Z, Xie J, Zhao Q, Meng D. Semi-supervised video deraining with dynamical rain generator. In: *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2021 Jun 20–25; Nashville, TN, USA. p. 642–52.
8. Li S, Araujo IB, Ren W, Wang Z, Tokuda EK, Junior RH, et al. Single image deraining: a comprehensive benchmark analysis. In: *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2019 Jun 15–20; Long Beach, CA, USA. p. 3838–47.
9. Kang LW, Lin CW, Fu YH. Automatic single-image-based rain streaks removal via image decomposition. *IEEE Transact Image Process.* 2011;21(4):1742–55. doi:10.1109/tip.2011.2179057.
10. Kim HG, Seo SJ, Song BC. Multi-frame de-raining algorithm using a motion-compensated non-local mean filter for rainy video sequences. *Visual Communicat Image Represent.* 2015;26:317–28. doi:10.1016/j.jvcir.2014.10.006.
11. Li Y, Tan RT, Guo X, Lu J, Brown MS. Rain streak removal using layer priors. In: *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*; 2016 Jun 27–30; Las Vegas, NV, USA. p. 2736–44.
12. Ahn N, Jo SY, Kang SJ. EAGNet: elementwise attentive gating network-based single image de-raining with rain simplification. *IEEE Transact Circ Syst Video Technol.* 2021;32(2):608–20. doi:10.1109/tcsvt.2021.3068985.
13. Fu X, Huang J, Zeng D, Huang Y, Ding X, Paisley J. Removing rain from single images via a deep detail network. In: *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*; 2017 Jul 21–26; Honolulu, HI, USA. p. 3855–63.
14. Zhang H, Patel VM. Density-aware single image de-raining using a multi-stream dense network. In: *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition*; 2018 Jun 18–23; Salt Lake City, UT, USA. p. 695–704.
15. Ren D, Zuo W, Hu Q, Zhu P, Meng D. Progressive image deraining networks: a better and simpler baseline. In: *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2019 Jun 15–20; Long Beach, CA, USA. p. 3937–46.
16. Qian R, Tan RT, Yang W, Su J, Liu J. Attentive generative adversarial network for raindrop removal from a single image. In: *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition*; 2018 Jun 18–23; Salt Lake City, UT, USA. p. 2482–91.
17. Wei Y, Zhang Z, Wang Y, Xu M, Yang Y, Yan S, et al. Deraincyclegan: rain attentive cyclegan for single image deraining and rainmaking. *IEEE Transact Image Process.* 2021;30:4788–801. doi:10.1109/tip.2021.3074804.
18. Huang H, Yu A, He R. Memory oriented transfer learning for semi-supervised image deraining. In: *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2021 Jun 20–25; Nashville, TN, USA. p. 7732–41.
19. Yang W, Tan RT, Feng J, Liu J, Guo Z, Yan S. Deep joint rain detection and removal from a single image. In: *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*; 2017 Jul 21–26; Honolulu, HI, USA. p. 1357–66.
20. Fu X, Liang B, Huang Y, Ding X, Paisley J. Lightweight pyramid networks for image deraining. *IEEE Transact Neural Netw Learn Syst.* 2019;31(6):1794–807. doi:10.1109/tnnls.2019.2926481.
21. Jiang K, Wang Z, Yi P, Chen C, Huang B, Luo Y, et al. Multi-scale progressive fusion network for single image deraining. In: *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2020 Jun 13–19; Seattle, WA, USA. p. 8346–55.
22. Wang Z, Cun X, Bao J, Zhou W, Liu J, Li H. Uformer: a general u-shaped transformer for image restoration. In: *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2022 Jun 18–24; New Orleans, LA, USA. p. 17683–93.
23. Chen X, Li H, Li M, Pan J. Learning a sparse transformer network for effective image deraining. In: *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2023 Jun 17–24; Vancouver, BC, Canada. p. 5896–905.

24. Zhang K, Luo W, Yu Y, Ren W, Zhao F, Li C, et al. Beyond monocular deraining: parallel stereo deraining network via semantic prior. *Int J Comput Vis.* 2022;130(7):1754–69. doi:10.1007/s11263-022-01620-w.
25. Liu JJ, Hou Q, Cheng MM, Feng J, Jiang J. A simple pooling-based design for real-time salient object detection. In: *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2019 Jun 15–20; Long Beach, CA, USA. p. 3917–26.
26. Yu L, Wang B, He J, Xia GS, Yang W. Single image deraining with continuous rain density estimation. *IEEE Transact Multim.* 2021;25(1):443–56. doi:10.1109/tmm.2021.3127360.
27. Li Y, Tan RT, Guo X, Lu J, Brown MS. Single image rain streak decomposition using layer priors. *IEEE Transact Image Process.* 2017;26(8):3874–85. doi:10.1109/tip.2017.2708841.
28. Wang Y, Huang TZ. A tensor-based low-rank model for single-image rain streaks removal. *IEEE Access.* 2019;7:83437–48. doi:10.1109/access.2019.2924447.
29. Sun G, Leng J, Cattani C. A particular directional multilevel transform based method for single-image rain removal. *Knowl Based Syst.* 2020;200:106000. doi:10.1016/j.knsys.2020.106000.
30. Deng LJ, Huang TZ, Zhao XL, Jiang TX. A directional global sparse model for single image rain removal. *Appl Mathema Model.* 2018;59(1):662–79. doi:10.1016/j.apm.2018.03.001.
31. Luo Y, Xu Y, Ji H. Removing rain from a single image via discriminative sparse coding. In: *Proceedings of the 2015 IEEE International Conference on Computer Vision*; 2015 Dec 7–13; Santiago, Chile. p. 3397–405.
32. Zhu L, Fu CW, Lischinski D, Heng PA. Joint bi-layer optimization for single-image rain streak removal. In: *Proceedings of the 2017 IEEE International Conference on Computer Vision*; 2017 Oct 22–29; Venice, Italy. p. 2526–34.
33. Chen YL, Hsu CT. A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In: *Proceedings of the 2013 IEEE International Conference on Computer Vision*; 2013 Dec 1–8; Sydney, NSW, Australia. p. 1968–75.
34. Chang Y, Yan L, Zhong S. Transformed low-rank model for line pattern noise removal. In: *Proceedings of the 2017 IEEE International Conference on Computer Vision*; 2017 Oct 22–29; Venice, Italy. p. 1726–34.
35. Fu X, Huang J, Ding X, Liao Y, Paisley J. Clearing the skies: a deep network architecture for single-image rain removal. *IEEE Transact Image Process.* 2017;26(6):2944–56. doi:10.1109/tip.2017.2691802.
36. Deng S, Wei M, Wang J, Feng Y, Liang L, Xie H, et al. Detail-recovery image deraining via context aggregation networks. In: *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2020 Jun 13–19; Seattle, WA, USA. p. 14560–9.
37. Yang W, Tan RT, Feng J, Guo Z, Yan S, Liu J. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE Transact Pattern Anal Mach Intell.* 2019;42(6):1377–93. doi:10.1109/tpami.2019.2895793.
38. Liu D, Huang Y, Fang Y, Zuo Y, An P. Multi-stream dense view reconstruction network for light field image compression. *IEEE Transact Multim.* 2022;25:4400–14. doi:10.1109/tmm.2022.3175023.
39. Jiang Y, Zhu B, Zhao X, Deng W. Pixel-wise content attention learning for single-image deraining of autonomous vehicles. *Expert Syst Appl.* 2023;224:119990. doi:10.1016/j.eswa.2023.119990.
40. Zhang Y, He X, Zhan C, Li J. Visual state space model for image deraining with symmetrical scanning. *Symmetry.* 2024;16(7):871. doi:10.3390/sym16070871.
41. Wang T, Yang X, Xu K, Chen S, Zhang Q, Lau RW. Spatial attentive single-image deraining with a high quality real rain dataset. In: *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2019 Jun 15–20; Long Beach, CA, USA. p. 12270–9.
42. Gao X, Wang Y, Wang M. Macroscopic-and-microscopic rain streaks disentanglement network for single-image deraining. *IEEE Transact Image Process.* 2023;32:2663–77. doi:10.1109/tip.2023.3272173.
43. Cui Y, Ren W, Knoll A. Exploring the potential of pooling techniques for universal image restoration. *IEEE Transact Image Process.* 2025;34:3403–16. doi:10.1109/tip.2025.3572788.
44. Cui Y, Wang Q, Li C, Ren W, Knoll A. EENet: an effective and efficient network for single image dehazing. *Pattern Recognit.* 2025;158:111074. doi:10.1016/j.patcog.2024.111074.

45. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition; 2016 Jun 27–30; Las Vegas, NV, USA. p. 770–8.
46. Liu X, Suganuma M, Sun Z, Okatani T. Dual residual networks leveraging the potential of paired operations for image restoration. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019 Jun 15–20; Long Beach, CA, USA. p. 7007–16.
47. Li X, Wu J, Lin Z, Liu H, Zha H. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In: Proceedings of the European Conference on Computer Vision (ECCV); Cham, Switzerland: Springer; 2018. p. 254–69.
48. Geiger A, Lenz P, Stiller C, Urtasun R. Vision meets robotics: the kitti dataset. *Int J Robot Res.* 2013;32(11):1231–7. doi:10.1177/0278364913491297.
49. Zhang K, Li D, Luo W, Ren W. Dual attention-in-attention model for joint rain streak and raindrop removal. *IEEE Transact Image Process.* 2021;30:7608–19. doi:10.1109/tip.2021.3108019.
50. Huynh-Thu Q, Ghanbari M. Scope of validity of PSNR in image/video quality assessment. *Elect Lett.* 2008;44(13):800–1. doi:10.1049/el:20080522.
51. Bakurov I, Buzzelli M, Schettini R, Castelli M, Vanneschi L. Structural similarity index (SSIM) revisited: a data-driven approach. *Expert Syst Appl.* 2022;189:116087. doi:10.1016/j.eswa.2021.116087.
52. Zhang L, Zhang L, Mou X, Zhang D. FSIM: a feature similarity index for image quality assessment. *IEEE Transact Image Process.* 2011;20(8):2378–86. doi:10.1109/tip.2011.2109730.
53. Jiang N, Luo J, Lin J, Chen W, Zhao T. Lightweight semi-supervised network for single image rain removal. *Pattern Recognit.* 2023;137:109277. doi:10.1016/j.patcog.2022.109277.
54. Wang H, Xie Q, Zhao Q, Meng D. A model-driven deep neural network for single image rain removal. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2020 Jun 13–19; Seattle, WA, USA. p. 3103–12.
55. Zhang H, Sindagi V, Patel VM. Image de-raining using a conditional generative adversarial network. *IEEE Transact Circ Syst Video Technol.* 2019;30(11):3943–56. doi:10.1109/tcsvt.2019.2920407.
56. Tejaswini M, Sumanth TH, Naik KJ. Single image deraining using modified bilateral recurrent network (modified_BRN). *Multim Tools Appl.* 2023;11:1–24. doi:10.1007/s11042-023-15276-2.
57. Ran W, Yang B, Ma P, Lu H. TRNR: task-driven image rain and noise removal with a few images based on patch analysis. *IEEE Transact Image Process.* 2023;32:721–36. doi:10.1109/tip.2022.3232943.
58. Hsu WY, Chang WC. Recurrent wavelet structure-preserving residual network for single image deraining. *Pattern Recognit.* 2023;137:109294. doi:10.1016/j.patcog.2022.109294.
59. Nanba Y, Miyata H, Han XH. Dual heterogeneous complementary networks for single image deraining. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2020 Jun 13–19; New Orleans, LA, USA. p. 568–77.
60. Yu X, Zhang G, Tan F, Li F, Xie W. Progressive hybrid-modulated network for single image deraining. *Mathematics.* 2023;11(3):691. doi:10.3390/math11030691.
61. Das B, Saha A, Mukhopadhyay S. Rain removal from a single image using refined inception ResNet v2. *Circ Syst Signal Process.* 2023;42(6):3485–508. doi:10.1007/s00034-022-02279-x.
62. Zhao C, Cai W, Hu C, Yuan Z. Cycle contrastive adversarial learning with structural consistency for unsupervised high-quality image deraining transformer. *Neural Netw.* 2024;178:106428. doi:10.1016/j.neunet.2024.106428.
63. Tao W, Yan X, Wang Y, Wei M. MFFDNet: single image deraining via dual-channel mixed feature fusion. *IEEE Transact Instrument Measur.* 2024;73:1–13. doi:10.1109/tim.2023.3346498.
64. Wen Y, Gao T, Chen T. Neural Schrödinger bridge for unpaired real-world image deraining. *Inf Sci.* 2024;682:121199. doi:10.1016/j.ins.2024.121199.
65. Zhou W, Ye L. UC-former: a multi-scale image deraining network using enhanced transformer. *Comput Vis Image Underst.* 2024;248:104097. doi:10.1016/j.cviu.2024.104097.
66. Wang C, Xing X, Wu Y, Su Z, Chen J. Dcsfn: deep cross-scale fusion network for single image rain removal. In: Proceedings of the 28th ACM International Conference on Multimedia; 2020 Oct 12–16; Seattle, WA, USA. p. 1643–51.

67. Hu X, Zhu L, Wang T, Fu CW, Heng PA. Single-image real-time rain removal based on depth-guided non-local features. *IEEE Transact Image Process.* 2021;30:1759–70. doi:10.1109/tip.2020.3048625.
68. Bhutto JA, Zhang R, Rahman Z. Symmetric enhancement of visual clarity through a multi-scale dilated residual recurrent network approach for image deraining. *Symmetry.* 2023;15(8):1571. doi:10.3390/sym15081571.
69. Wang C, Zhu H, Fan W, Wu XM, Chen J. Single image rain removal using recurrent scale-guide networks. *Neurocomputing.* 2022;467:242–55. doi:10.1016/j.neucom.2021.10.029.