processes

MDPI

Article

Rain-Resilient Image Restoration for Reliable and Sustainable Visual Monitoring in Industrial Inspection and Quality Control

Miao Zhang [1], Shanqin Wang [1] and Shiqun Yin [2,*]

[1] School of Information Engineering, Chuzhou Polytechnic, Chuzhou 239000, China; zhangmiao@chzc.edu.cn (M.Z.); wangshanqin@chzc.edu.cn (S.W.)
[2] College of Computer and Information Science, Southwest University, Chongqing 400715, China
* Correspondence: qiongyin@swu.edu.cn

Abstract: Reliable visual monitoring is essential for industrial quality control systems under adverse weather conditions. Rain-induced degradation, such as occlusions, texture blurring, and depth distortions, can significantly hinder image clarity and compromise precision in surface defect detection. To address this, we propose a novel image deraining framework, the Degradation-Background Perception Network (DBPNet). DBPNet features a hierarchical encoder–decoder structure and incorporates two core modules: the Frequency Degradation Perception Module (FDPM) and the Depth Background Perception Module (DBPM). FDPM focuses on frequency decomposition to remove high-frequency rain streaks while retaining critical image features using cross-attention mechanisms. DBPM is proposed to integrate robust depth maps, which remain unaffected by rain degradation, as explicit constraints to guide the model in reconstructing clean scenes. Furthermore, we propose the Selective Focus Attention (SFA) module, which enhances interactions between frequency-domain features and background priors, ensuring accurate reconstruction and effective rain removal. Experimental results on five synthetic and real-world benchmark datasets demonstrate that our method outperforms state-of-the-art CNN and transformer-based approaches. This framework contributes to more robust visual input for process control, enabling better fault detection, predictive maintenance, and sustainable system operation.

Keywords: single image deraining; frequency domain; depth feature

Academic Editor: Iqbal M. Mujtaba

Received: 22 April 2025
Revised: 15 May 2025
Accepted: 19 May 2025
Published: 22 May 2025

Citation: Zhang, M.; Wang, S.; Yin, S. Rain-Resilient Image Restoration for Reliable and Sustainable Visual Monitoring in Industrial Inspection and Quality Control. *Processes* **2025**, *13*, 1628. https://doi.org/10.3390/pr13061628

1. Introduction

Image deraining is a critical research topic in the field of image restoration, aimed at recovering clean and rain-free images from degraded observations captured in rainy weather conditions. The interference caused by rain can significantly hinder the performance of various downstream tasks, such as object detection [1], autonomous driving systems, and video surveillance [2]. In safety-critical applications like self-driving cars, even minor rain-induced artifacts can distort traffic sign recognition or obstacle detection, potentially leading to catastrophic consequences. Rain-related degradation typically manifests in two primary forms: rain streaks, which are the elongated patterns caused by falling raindrops, and rain haze, resulting from the scattering effect of rain droplets on light, particularly in heavy rain or distant scenes. These degradations exhibit complex spatial-frequency characteristics: while rain streaks predominantly occupy high-frequency bands due to their sharp edges, rain haze introduces low-frequency global illumination shifts that obscure structural details.

The development of image deraining techniques has undergone significant evolution over the past decades. Early approaches primarily relied on traditional image processing

techniques, such as filtering [3], low-rank modeling [4], and sparse coding [5]. For instance, Gaussian mixture models [6] were employed to separate rain layers by leveraging chromatic consistency assumptions, while dictionary learning methods [7] attempted to model rain streaks as sparse outliers. These methods focused on utilizing handcrafted features to model rain streak patterns and their removal but were often limited in handling diverse rain scenarios and complex background textures [8]. Their reliance on simplistic priors frequently led to over-smoothing of textures or incomplete rain removal when confronted with overlapping rain layers or non-Gaussian noise distributions. With the advent of deep learning, the field witnessed a paradigm shift. Convolutional neural networks (CNNs) became the backbone of modern deraining frameworks, leveraging their powerful feature extraction capabilities to address the intricate spatial structures of rain streaks. Recent architectures like SwinDerain [9] have further incorporated transformer-based mechanisms to capture long-range dependencies critical for distinguishing rain patterns from similar-looking background edges. Subsequently, advanced architectures, such as attention mechanisms [10], generative adversarial networks (GANs) [11], and multi-scale learning models [12,13], further enhanced deraining performance by focusing on spatial and contextual dependencies. Notably, progressive learning frameworks have demonstrated success in handling heavy rain by iteratively removing rain streaks and haze through cascaded subnetworks. Despite remarkable progress, challenges remain in achieving robust performance across varying rain intensities, textures, and complex scenes. Current methods often struggle with three key issues: (1) simultaneous handling of spatially varying rain streaks and global haze effects, (2) preservation of high-frequency textures in occluded regions, and (3) generalization to real-world rain patterns that deviate from synthetic training data distributions.

Recent studies [8,14,15] have explored the use of frequency-domain approaches for image deraining. These methods leverage frequency transformations like Fourier and Wavelet transforms to isolate and process features from various frequency bands, addressing both global and local degradation patterns effectively. The frequency domain's inherent separation capability allows explicit modeling of rain components: high-frequency bands capture rain streak details while low-frequency components contain haze-induced illumination shifts. For example, Fu et al. [14] proposed a Differential Dependency Network (DDN) that decomposes a rainy image into low-frequency background and high-frequency details. The network focuses on high-frequency components, utilizing residual blocks for its architecture. By learning nonlinear mappings through a CNN, DDN effectively removes rain streaks from the high-frequency components, thereby predicting both the rain residuals and the clean image. In a similar vein, Zhang et al. [8] proposed the Density-aware Image De-raining method using a Multistream Dense Network (DID-MDN), which overcomes the limitations of previous methods that struggled with varying rain densities and sizes. DID-MDN incorporates a rain density estimation into the network, enabling it to remove rain streaks effectively across diverse conditions. FreqMamba [15] combines frequency-based techniques with state-space modeling, which employs a multi-branch Frequency-State Space Model (SSM) block that integrates Fourier spectrum modeling with local spatial refinements. However, existing frequency methods often treat high- and low-frequency components in isolation, neglecting their cross-band correlations that are crucial for reconstructing edge-continuous structures. Moreover, a key limitation of image deraining is the difficulty in separating rain from object edges and the background, as the rain often blends with these elements, making it challenging to directly learn deraining information in the image domain. This ambiguity is exacerbated in heavy rain scenarios where dense streaks form complex occlusions while haze reduces contrast, creating a coupled degradation that requires joint spatial-frequency reasoning.

To address the challenges posed by rain degradation, we draw inspiration from recent advances in multimodal learning and physics-aware vision systems. We propose the Degradation-Background Perception Network (DBPNet), a novel framework designed to effectively remove rain degradation by leveraging frequency-domain properties and depth background priors. Our method addresses both low-frequency structural degradations and high-frequency detail disruptions, guided by background prompts to generate high-quality, clean images. The proposed framework is centered around two key innovations. First, we introduce a Frequency Degradation Perception Module (FDPM), which explicitly decomposes image features into high-frequency and low-frequency components, enabling targeted extraction of degradation features. The FDPM employs a cross-attention mechanism to modulate and enhance interactions between frequency components, facilitating effective feature separation and refinement. Second, inspired by the robustness of Depth-Anything [16] in handling extreme cases, we integrate background priors extracted from Depth-Anything as guidance for background modeling, called the Depth Background Perception Module (DBPM). These priors enable our framework to better understand and reconstruct background details, even in challenging conditions. To further improve rain removal and image reconstruction, we propose a Selective Focus Attention (SFA) module, which enhances interactions between frequency-domain features and background priors. The SFA module selectively emphasizes key frequency components and aligns them with background features, ensuring more precise reconstructions and robust rain removal. Extensive experiments on synthetic and real-world rain degradation datasets demonstrate that DBPNet achieves state-of-the-art performance, especially excelling in real-world benchmarks. Our contributions are summarized as follows:

- We propose the Degradation-Background Perception Network (DBPNet), a comprehensive framework that addresses rain degradation by leveraging frequency-domain characteristics and depth-based background priors.
- The Frequency Degradation Perception Module (FDPM) is proposed to explicitly decompose image features into high- and low-frequency components, facilitating targeted degradation modeling and feature refinement through a cross-attention mechanism.
- We design the Depth Background Perception Module (DBPM), which integrates depth priors extracted from Depth-Anything to guide the reconstruction of background details, showcasing the robustness of depth-aware background modeling in adverse weather conditions.
- We propose the Selective Focus Attention (SFA) module, which aligns frequency-domain features with depth priors, selectively emphasizing key features to achieve more accurate rain removal and image reconstruction.

2. Related Work

2.1. Single Image Deraining

Single image deraining (SID) has become an important and challenging area within emerging vision applications. The task aims to recover clean background content from a rain-degraded image, which is crucial for downstream applications like autonomous driving, video surveillance, and industrial monitoring. Over the years, numerous efforts have been made to tackle this problem under both supervised and unsupervised settings.

With the advent of deep learning, convolutional neural networks (CNNs) became the mainstream solution for SID due to their strong capability in learning spatial features and non-linear mappings. Early works like DDN [14] designed residual architectures to directly predict rain residuals, but faced limitations in handling diverse rain densities or separating rain streaks from complex textures. Subsequent models adopted multi-stream, multi-scale, or recurrent architectures to improve performance. For instance, RESCAN and PReNet

introduced recurrent processing for progressive refinement, while MSPFN and MPRNet employed multi-scale modules to aggregate global and local information. Transformer-based methods such as Uformer [17], Restormer [18], and NeRD-Rain [19] have further advanced this field by capturing long-range dependencies and modeling global context. Despite these improvements, existing models still face challenges in maintaining high-frequency details, generalizing to real-world rain patterns, and separating rain streaks from visually similar structures. To address these limitations, recent research has explored more refined strategies that incorporate additional priors, cross-domain learning, or new optimization schemes. For example, Chen et al. [20] proposed an error-aware feedback mechanism to detect residual degradation and adaptively refine features. Their approach improves robustness in complex rain scenarios by explicitly modeling deraining errors. Furthermore, DCD-GAN [21] introduces dual contrastive learning in an unpaired setting, allowing the model to align content representations across different domains without requiring paired supervision. This strategy has shown promising results for real-world applications, where clean ground truth may be unavailable.

Recent advancements have further focused on improving derained image quality by integrating diverse strategies such as feature fusion, attention mechanisms, frequency enhancement, and patch-level recurrence modeling. These approaches aim to address the inherent challenges of distinguishing rain streaks from background textures and preserving fine details across complex scenes. For instance, MFFDNet [22] adopts a dual-channel mixed fusion scheme that captures both local textures and global semantics, enabling more comprehensive degradation modeling. Building on this idea, SEPC [23] introduces a synergistic ensemble framework that leverages both single-scale and multi-scale features, reinforced by contrastive learning to improve discrimination between rain and object boundaries. Similarly, DPAFNet [24] enhances feature representation through dual-path attention fusion, allowing more effective integration of multidimensional contextual information. Focusing on frequency-domain characteristics, Gabformer [25] incorporates Gabor filters into transformer blocks to preserve high-frequency textures critical for visual fidelity. In parallel, AFENet [26] adaptively adjusts frequency responses across scales to enhance structural consistency in restored images. FADformer [27] further advances frequency-aware modeling by combining spectral priors with transformer-based architectures, effectively bridging spatial and frequency representations. Beyond pixel-level enhancement, structural recurrence has also been explored. MSGNN [28] exploits graph neural networks to model both internal and external patch similarities, improving generalization across varied rain patterns. Furthermore, insights from real-world benchmarks are driving progress—Zhang et al. [29] summarized lessons from the GT-Rain Challenge, encouraging a shift toward realistic degradation scenarios and cross-domain robustness.

Despite these advances, most existing approaches still rely heavily on spatial or frequency cues alone, often overlooking the inherent coupling between rain artifacts and scene semantics. In contrast, our proposed method integrates both frequency decomposition and depth-aware priors to address these challenges jointly.

2.2. Frequency Domain in Image Restoration

According to the spectral convolution theorem, Fast Fourier Transform (FFT) serves as an effective tool for modeling global information [30]. In this context, high-frequency components capture image details and textures, while low-frequency components represent smooth and flat regions. This separation makes it convenient to handle different frequency sub-bands independently within the frequency domain. Leveraging these advantages, several deep learning frameworks have been proposed for image restoration in the spectral domain. For instance, Mao et al. [31] employed Fourier transforms to integrate both

high- and low-frequency residuals for motion deblurring. Guo et al. [32] introduced a window-based frequency channel attention mechanism based on FFT, which models global information while maintaining model consistency across training and inference stages. Li et al. [33] incorporated Fourier transforms into their model to enhance low-light images by separately processing amplitude and phase. Additionally, FFT has been utilized in designing loss functions aimed at preserving high-frequency details [13,34–36].

Moreover, wavelet transform has also been explored for image restoration tasks. Chen et al. [37] proposed a hierarchical desnowing network using dual-tree complex wavelet representation. Yang et al. [38] developed a wavelet-based U-Net to replace traditional up-sampling and down-sampling operations. Zou et al. [39] employed wavelet-transform-based modules to restore texture details. Yang et al. [40] designed a wavelet structure similarity loss function to enhance training.

3. Proposed Methodology

We propose the Degradation-Background Perception Network (DBPNet), designed to effectively extract structural information and texture details from degraded images and generate high-quality, clean images under the guidance of background prompts. DBPNet consists of two main components: the Frequency Degradation Perception Module (FDPM), which captures rich structural and detail features in the frequency domain, and the Depth Background Perception Module (DBPM), which leverages extracted background priors to guide background modeling. Finally, to enhance the interaction between frequency-domain features and background priors, we introduce the Selective Focus Attention (SFA) module, which emphasizes key frequency components and aligns them with background features.

As shown in Figure 1, the proposed DBPNet adopts a classic encoder–decoder architecture to effectively learn hierarchical representations. The entire structure consists of three scales of encoders and decoders, along with a latent layer. Both the encoders and decoders are implemented using the Swin Transformer block [41]. Given a rain-degraded image of size $X_d \in \mathcal{R}^{H\times W\times 3}$, where $H \times W$ and C represent spatial dimensions and channel count, respectively, shallow features of size $X_d \in \mathcal{R}^{H\times W\times C}$ are extracted using a 3×3 convolution layer. In the encoder stage, the proposed FDPM is used to consciously extract global structures and texture details in the frequency domain. In the latent layer, the SFA module integrates deep background information unaffected by degradation into the pipeline effectively. Finally, the decoder restores clean images across three different scales. To achieve upsampling, we utilize an inverse pixel rearrangement strategy, which redistributes channel-wise feature maps into higher-resolution spatial grids. This operation is followed by convolutional refinement to recover fine structural details.

3.1. Frequency Degradation Perception Module

The frequency spectrum of natural images typically follows a power-law distribution, where low-frequency components capture the overall structure and smooth regions of the image, while high-frequency components represent edges, textures, and fine details. However, rain droplets or streaks tend to couple with the natural frequency characteristics of the background, leading to the enhancement or attenuation of certain frequency components. To address this, we propose a Frequency Degradation Perception Module (FDPM) that decouples different frequency characteristics, enhances or aligns key frequency components, and effectively removes degradation effects while preserving essential image details.

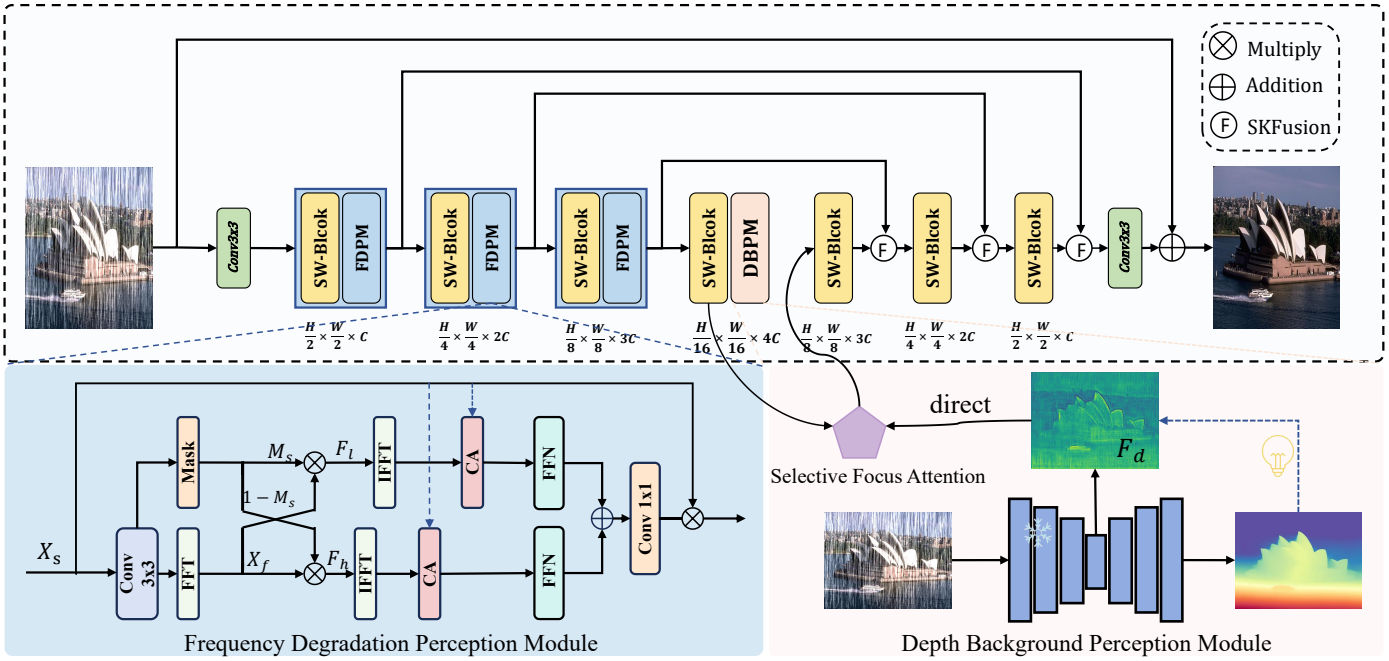

Figure 1. Overall architecture of the proposed Degradation-Background Perception Network (DBP-Net). It adopts a hierarchical encoder–decoder structure with three core components: the Frequency Degradation Perception Module (FDPM) for frequency-based rain removal, the Depth Background Perception Module (DBPM) for depth-guided background modeling, and the Selective Focus Attention (SFA) module for aligning frequency-domain features with depth priors. Swin Transformer blocks are used in the encoder and decoder paths to capture spatial and contextual dependencies.

As shown in Figure 1, for a given degraded feature $X_s \in \mathcal{R}^{H\times W\times C}$, we apply the Fast Fourier Transform (FFT) to convert the spatial representation into the frequency-domain representation $X_f \in \mathcal{R}^{H\times W\times C}$. The transformation can be expressed as

$$X_f(u,v,c) = \sum_{x=0}^{H-1}\sum_{y=0}^{W-1} X_s(x,y,c)\cdot e^{-j2\pi\left(\frac{ux}{H}+\frac{vy}{W}\right)}, \tag{1}$$

where (u,v) denotes the frequency-domain coordinates, c represents the channel index, and j is the imaginary unit.

To separate the high-frequency and low-frequency components of the data effectively, we design frequency masks based on the spectral properties of the input. The center of the frequency spectrum is identified by the coordinates c_h and c_w, where h and w represent the height and width of the input tensor. Using these central coordinates, we construct two masks: a low-frequency mask M_{low} and a high-frequency mask M_{high}. The size of the low-frequency region is determined by $low_\sigma = \frac{\min(c_h,c_w)}{m}$, and m is a scaling hyperparameter (e.g., $m = 4$). This definition ensures that the low-frequency mask encompasses approximately one-quarter of the smaller dimension of the spectrum, effectively isolating the dominant low-frequency components. Studies have shown that natural images follow a power-law spectral distribution, where the energy at frequency f satisfies $E(f) \propto 1/f^{2\alpha}$, with $\alpha \in [1,2]$. Under this distribution, the cumulative energy within a circular region of radius R (centered at the spectral origin) can be computed as

$$\text{Energy ratio} = \frac{\int_0^R f^{1-2\alpha}\,df}{\int_0^{f_{\max}} f^{1-2\alpha}\,df} = \left(\frac{R}{f_{\max}}\right)^{2-2\alpha} \tag{2}$$

For example, when $\alpha = 1.2$, which is typical for natural images, the setting $R = \frac{1}{4} f_{\max}$ yields Energy ratio $\approx \left(\frac{1}{4}\right)^{0.6} \approx 0.4$. This means the low-frequency mask retains approximately 40% of the total spectral energy, effectively capturing global structures and illumination patterns while discarding high-frequency rain artifacts.

The low-frequency mask M_{low} is set to 1 for all points within the specified low-frequency region around the spectral center and 0 elsewhere.

$$M_{\text{low}}(u', v') = \begin{cases} 1, & \text{if } \sqrt{(h')^2 + (w')} \leq low_\sigma, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

where $h' = u' - \text{center}_h$ and $w' = v' - \text{center}_w^2$. u' and v' represent the frequency-domain coordinates. The high-frequency mask M_{high} is then defined as the complement of the low-frequency mask $1 - M_{\text{low}}$. Then, we use these masks to extract low-/high-frequency components from the input tensor X_f:

$$F_l = X_f \otimes M_{\text{low}}, \quad F_h = X_f \otimes M_{\text{high}} \tag{4}$$

Finally, we apply the inverse Fourier transform to obtain the decoupled and adaptive frequency-domain features. Next, we use a cross-attention mechanism to highlight relevant features and suppress redundant ones from the extracted adaptive frequency-domain features. This process can be expressed as

$$Q_h = W_q^h X_s, \quad K_h = W_k^h F_h, \quad V_h = W_v^h F_h, \tag{5}$$

$$Q_l = W_q^l X_s, \quad K_l = W_k^l F_l, \quad V_l = W_v^l F_l, \tag{6}$$

$$\mathcal{CA}(Q_h, K_h, V_h) = \text{softmax}\left(\frac{Q_h K_h^T}{\sqrt{d_k}}\right) V_h, \tag{7}$$

$$\mathcal{CA}(Q_l, K_l, V_l) = \text{softmax}\left(\frac{Q_l K_l^T}{\sqrt{d_k}}\right) V_l, \tag{8}$$

where, $W_q^h, W_k^h, W_v^h, W_q^l, W_k^l, W_v^l \in \mathbb{R}^{C \times d}$ are learnable linear projection matrices that transform the input features into query, key, and value representations for the cross-attention operations in both high- and low-frequency branches. $\mathcal{CA}$ is the cross-attention operation. d_k is the dimension of the key vector, and the softmax operation ensures that the attention weights are normalized. Subsequently, we input the corresponding features into a feed-forward neural network to learn different inductive biases.

3.2. Depth Background Perception Module

In addition to the occlusion and blurring degradation caused by rain, adverse weather conditions also lead to distortions related to depth. These distortions create artifacts in areas with distant depth, further degrading the visual quality. Therefore, modeling scene content becomes another crucial factor for improving performance. Inspired by scene understanding [42,43], leveraging depth information has been shown to be an effective way to represent clean scenes. We observe that depth maps estimated by methods like Depth-Anything [16] remain largely unaffected by degradation, which partially demonstrates the robustness of intermediate hidden features in scene representation, as shown in Figure 2. Based on this insight, we propose a Depth Background Perception Module that applies an explicit constraint to the latent spatial features, enabling the model to focus more effectively on the background during image reconstruction.

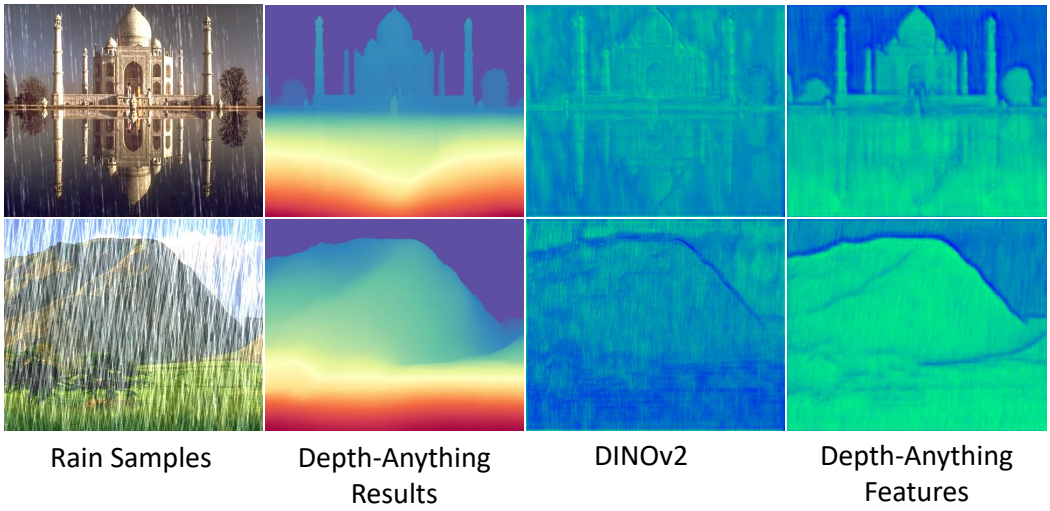

Figure 2. The motivation for using Depth-Anything [44] lies in its potential as a constraint condition. Unlike the traditional pre-trained network DINOv2 [45], Depth-Anything demonstrates superior resilience to degradation, with intermediate features exhibiting significantly enhanced robustness.

Building on this concept, the Depth Background Perception Module integrates depth cues into the reconstruction process by explicitly distinguishing background information from foreground elements. By leveraging the depth map's stable representation, the module enhances the model's ability to preserve the background's spatial coherence while mitigating the impact of degradation artifacts in distant regions. The depth information acts as a guide to refine the latent features, ensuring that background details are more accurately reconstructed, even in challenging weather conditions. Furthermore, the depth-guided constraint ensures that the reconstructed image retains natural depth relationships, resulting in a more realistic and artifact-free restoration.

Selective Focus Attention. To enhance the interaction between frequency-domain features and background priors, we introduce the Selective Focus Attention (SFA) module, which leverages an attention mechanism to integrate latent features from the Depth-Anything model, as shown in Figure 3.

Specifically, given the features F_s from the previous block and the latent features F_d from the Depth-Anything model, we first apply a linear transformation to F_d and extract the corresponding mean and maximum values to balance the feature distribution, allowing the integration of depth-based information, enabling the model to focus on relevant features while enhancing the background representation during the reconstruction process. We then multiply F_s by the transformed F_d, which can be expressed as

$$F_{di} = \text{Linear}(F_d) = W_d F_d + b_d \quad F_{si} = \mathcal{C}_3(F_s) \tag{9}$$

$$F_{di}^m = \text{mean}(F_{di}) + \max(F_{di}) \tag{10}$$

$$F_d^{\text{adjusted}} = F_{di}^m \cdot F_{si} \tag{11}$$

where Linear denotes a fully connected layer applied to the depth-guided features F_d. $W_d \in \mathbb{R}^{C\times C}$ and $b_d \in \mathbb{R}^C$ are learnable parameters. This projection aligns the depth features with the spatial feature space in the channel dimension. The operator $\mathcal{C}_3$ refers to a standard 3×3 convolutional layer applied to the spatial features F_s The mean and maximum values are used to normalize and balance the feature distribution.

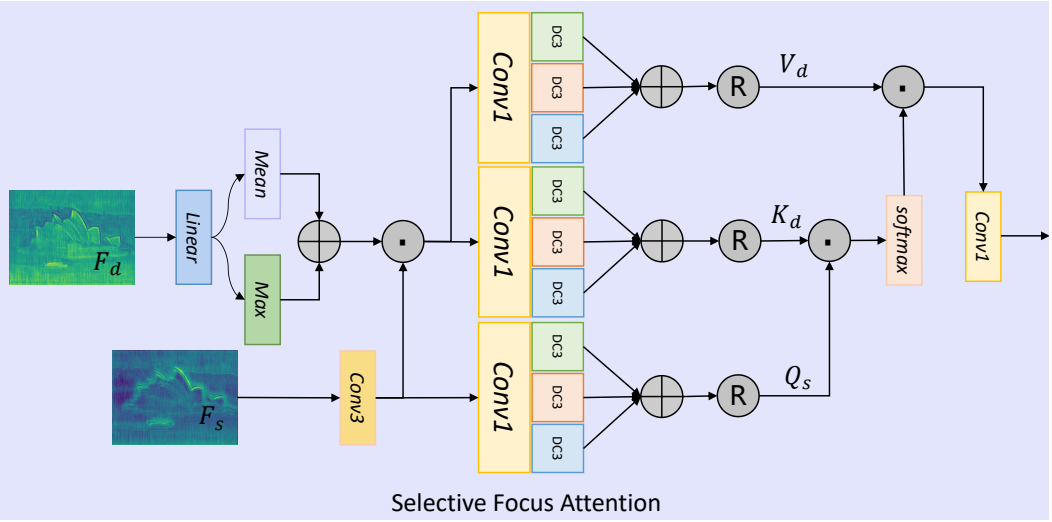

Figure 3. Pipeline of the proposed Selective Focus Attention (SFA) module. The SFA aligns frequency-domain features (F_s) with depth-guided priors (F_d) extracted from Depth-Anything. The depth features are enhanced via linear transformation, followed by mean-max fusion. A multi-scale cross-attention mechanism then integrates the refined depth and frequency information to guide more accurate background reconstruction.

Next, we define a cross-attention operation where the general features F_{si} serve as queries, while the keys and values are derived from the Depth-Anything features F_d^{adjusted}. Notably, the Selective Focus Attention (SFA) module incorporates multi-scale information for enhanced contextual understanding. To encode spatial positions, we use pointwise convolutions and apply depthwise separable convolutions with varying kernel sizes ($i = 1$, 3, and 5) to obtain the queries $Q_s^i \in \mathcal{R}^{C\times HW}$, keys $K_d^i \in \mathcal{R}^{HW\times C}$, and values $V_d^i \in \mathcal{R}^{C\times HW}$ needed for self-attention. This process can be expressed as

$$Q_s = \sum_{i=0} W_Q^i F_{si}, K_d = \sum_{i=0} W_K^i F_d^{\text{adjusted}}, V_d = \sum_{i=0} W_V^i F_d^{\text{adjusted}} \tag{12}$$

$$\mathcal{A}_s = \text{softmax}(Q_s \odot K_d) \odot V_d \tag{13}$$

where $\odot$ denotes the dot-product operation. By using this attention-based mechanism, the SFA module effectively leverages both the general features and the depth-guided features to improve image reconstruction, enhancing the model's ability to handle complex degradations and preserve important scene details. The multi-scale attention further contributes to capturing context from different spatial levels, ensuring that both fine-grained details and global background structures are appropriately restored.

3.3. Loss Function

Following the multi-resolution training strategy in [36], we design a hierarchical loss formulation that operates across three spatial scales to ensure comprehensive feature learning and structural preservation. The composite loss function integrates complementary constraints in both spatial and frequency domains, formulated as

$$\mathcal{L}_{\text{total}} = \sum_{s\in\{1,0.5,0.25\}} \left[\alpha\mathcal{L}_{\text{spectral}}^s + \beta\mathcal{L}_{\text{structural}}^s + \gamma\mathcal{L}_{\text{textural}}^s\right] \tag{14}$$

where $\alpha = 1.0$, $\beta = 0.5$, and $\gamma = 0.1$ denote adaptive weighting coefficients optimized through grid search on validation data. The multi-scale architecture processes images at full-resolution ($s = 1$), half-resolution ($s = 0.5$), and quarter-resolution ($s = 0.25$), enabling the network to learn both global contextual patterns and local texture details.

Spectral Consistency Loss ($\mathcal{L}_{\text{spectral}}$): The ℓ_1-norm-based component ensures pixel-level fidelity between derained results and ground truth:

$$\mathcal{L}^s_{\text{spectral}} = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{Y}^s_i - \hat{\mathbf{Y}}^s_i\|_1 \tag{15}$$

This serves as the foundation for color consistency preservation.

Structural Perception Loss ($\mathcal{L}_{\text{structural}}$): We adopt a multi-window SSIM formulation to enhance perceptual quality under misalignment scenarios:

$$\mathcal{L}^s_{\text{structural}} = 1 - \prod_{k\in\{3,5,7\}} \left[\text{SSIM}_k(\mathbf{Y}^s, \hat{\mathbf{Y}}^s)\right]^{w_k} \tag{16}$$

where w_k denotes window-size dependent weights. Here, SSIM denotes the Structural Similarity Index Measure, which evaluates the perceptual similarity between two image patches by jointly considering luminance, contrast, and structural components. It is defined as

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where μ_x, μ_y are the local means, σ_x^2, σ_y^2 the variances, and σ_{xy} the covariance between predicted and ground truth patches. C_1 and C_2 are small constants to avoid instability.

In our formulation, we adopt a multi-window SSIM loss, where SSIM is computed using sliding windows of sizes 3, 5, and 7, each weighted by window-dependent coefficients w_k. This multi-scale strategy enhances sensitivity to both fine textures and larger structural patterns.

Frequency Discriminative Loss ($\mathcal{L}_{\text{textural}}$): To address the spectral bias of conventional losses, we implement a phase-aware Fourier constraint:

$$\mathcal{L}^s_{\text{textural}} = \sum_{c\in\{\text{real,imag}\}} \|\mathcal{F}_c(\mathbf{Y}^s) - \mathcal{F}_c(\hat{\mathbf{Y}}^s)\|_2^2 + \lambda_{\text{HS}} \cdot \mathcal{L}_{\text{highpass}} \tag{17}$$

where $Y^s_i \in \mathbb{R}^{H_s\times W_s\times 3}$ are the ground truth clean image at spatial scale $s \in \{1, 0.5, 0.25\}$, i indexes the sample, and H_s, W_s denote the spatial resolution at scale s. $\hat{Y}^s_i \in \mathbb{R}^{H_s\times W_s\times 3}$ is the corresponding predicted image output by DBPNet at the same scale. $\mathcal{F}_c$ extracts real/imaginary components, and $\mathcal{L}_{\text{highpass}}$ penalizes deviations in high-frequency bands using Butterworth filtering. This dual-domain formulation bridges the gap between spatial reconstruction and spectral fidelity.

3.4. Implementation Details

The optimization process utilized the ADAM algorithm with cosine learning rate scheduling, initialized at 4×10^{-4} and progressively refined through 200 training epochs. Input images were processed as randomly cropped 256×256 patches with stochastic horizontal flipping for spatial augmentation. Implemented in PyTorch 2.50 with mixed-precision acceleration, the framework completed full training within 48 h on an NVIDIA RTX 3090 GPU, maintaining batch parallelism through gradient accumulation strategies for memory-efficient processing.

4. Experiments

4.1. Experiment Datasets

In this study, we utilized several benchmark datasets to evaluate the performance of our rain removal model across synthetic and real-world conditions, including Rain200L/H [46], DID-Data [8], DDN-Data [14], and SPA-Data [10]. The synthetic datasets, Rain200L [46] and Rain200H [46], consist of images with varied rain streak densities.

Rain200L includes relatively sparse rain streaks, while Rain200H contains dense rain patterns, providing challenges for different levels of rain removal. Each of these datasets comprises 1800 rainy images for training and 200 images for testing, allowing a robust assessment of model effectiveness under controlled conditions. Additionally, DID-Data [8] and DDN-Data [14] further contribute to synthetic evaluations by including 12,000 and 12,600 rainy images for training, respectively, with diverse rain orientations and densities. DID-Data provides 1200 test images, and DDN-Data includes 1400 test images, which capture a wide range of rain conditions to examine model generalizability.

To evaluate model performance in real-world scenarios, we also incorporated the SPA-Data [10] dataset, a large-scale, real-world rain dataset that includes 638,492 paired images for training and 1000 paired images for testing. SPA-Data's extensive collection of real rain conditions and diverse environmental settings presents a more practical evaluation, closely reflecting real-life application needs. Together, these datasets facilitate comprehensive testing across synthetic and authentic rain scenes, enabling a thorough evaluation of the proposed model's robustness and adaptability in diverse rain removal tasks.

4.2. Evaluation Metrics

To evaluate the effectiveness of our proposed model, we utilized two standard image quality metrics: Peak Signal-to-Noise Ratio (PSNR) [47] and Structural Similarity Index (SSIM) [48]. These metrics are widely used in image restoration tasks to assess both the overall fidelity and the structural integrity of restored images. Following previous deraining methods [49,50], we calculated PSNR and SSIM metrics in the Y channel of YCbCr space.

PSNR is a measure of the peak error between the ground truth and the restored image, providing an objective evaluation of image quality. It is calculated as the ratio of the maximum possible signal power to the power of the noise (error) introduced during the restoration process. On the other hand, SSIM is a perceptual metric that evaluates the similarity between two images based on luminance, contrast, and structure. Unlike PSNR, SSIM considers changes in structural information, which correlates more closely with human visual perception.

In addition to PSNR and SSIM, visual qualitative analysis was also performed to provide a more comprehensive evaluation of our model's performance.

4.3. Experiment Results

The proposed method was evaluated on five benchmark datasets in Table 1, each presenting unique rain removal challenges that assess the model's robustness across both synthetic and real-world rain conditions. We selected Prior-based methods [4,51], CNN-based methods [14,49,52–57], and transformer-based methods [17–19,58,59] for comparison.

On Rain200L, a synthetic dataset with sparse rain streaks, our method achieved a PSNR of 41.75 and SSIM of 0.9906, surpassing all competing methods. This result slightly outperformed NeRD-Rain, which achieved a PSNR of 41.71 and SSIM of 0.9903, demonstrating the model's effectiveness in maintaining image fidelity in light rain scenarios. Our approach benefits from the depth-guided background features, which provide a refined context to separate rain streaks from background structures, helping to preserve finer details in lightly degraded images. Figure 4 reports the visual comparison of our method with other advanced methods, which shows that our method can remove more rain degradation and shows advantages in restoring the local background.

Table 1. Comparison of quantitative results on synthetic and real datasets. **Bold** and underline indicate the best and second-best results.

Datasets		Rain200L		Rain200H		DID-Data		DDN-Data		SPA-Data	
Metrics		**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
Prior-based methods	DSC [51]	27.16	0.8663	14.73	0.3815	24.24	0.8279	27.31	0.8373	34.95	0.9416
	GMM [4]	28.66	0.8652	14.50	0.4164	25.81	0.8344	27.55	0.8479	34.30	0.9428
CNN-based methods	DDN [14]	34.68	0.9671	26.05	0.8056	30.97	0.9116	30.00	0.9041	36.16	0.9457
	RESCAN [52]	36.09	0.9697	26.75	0.8353	33.38	0.9417	31.94	0.9345	38.11	0.9707
	PReNet [53]	37.80	0.9814	29.04	0.8991	33.17	0.9481	32.60	0.9459	40.16	0.9816
	MSPFN [49]	38.58	0.9827	29.36	0.9034	33.72	0.9550	32.99	0.9333	43.43	0.9843
	RCDNet [54]	39.17	0.9885	30.24	0.9048	34.08	0.9532	33.04	0.9472	43.36	0.9831
	MPRNet [55]	39.47	0.9825	30.67	0.9110	33.99	0.9590	33.10	0.9347	43.64	0.9844
	DualGCN [56]	40.73	0.9886	31.15	0.9125	34.37	0.9620	33.01	0.9489	44.18	0.9902
	SPDNet [57]	40.50	0.9875	31.28	0.9207	34.57	0.9560	33.15	0.9457	43.20	0.9871
Transformer-based methods	Uformer [17]	40.20	0.9860	30.80	0.9105	35.02	0.9621	33.95	0.9545	46.13	0.9913
	Restormer [18]	40.99	0.9890	32.00	0.9329	35.29	0.9641	34.20	0.9571	47.98	0.9921
	IDT [58]	40.74	0.9884	32.10	0.9344	34.89	0.9623	33.84	0.9549	47.35	0.9930
	DRSformer [59]	41.23	0.9894	32.18	0.9330	35.38	0.9647	34.36	0.9590	48.53	0.9924
	NeRD-Rain-S [19]	41.30	0.9895	32.06	0.9315	35.36	0.9647	34.25	0.9578	48.90	0.9936
	NeRD-Rain [19]	41.71	0.9903	32.40	0.9373	35.53	0.9659	34.45	0.9596	49.58	**0.9940**
	Ours	**41.75**	**0.9906**	**32.42**	**0.9376**	**35.63**	**0.9703**	**34.49**	**0.9653**	**49.70**	0.9912

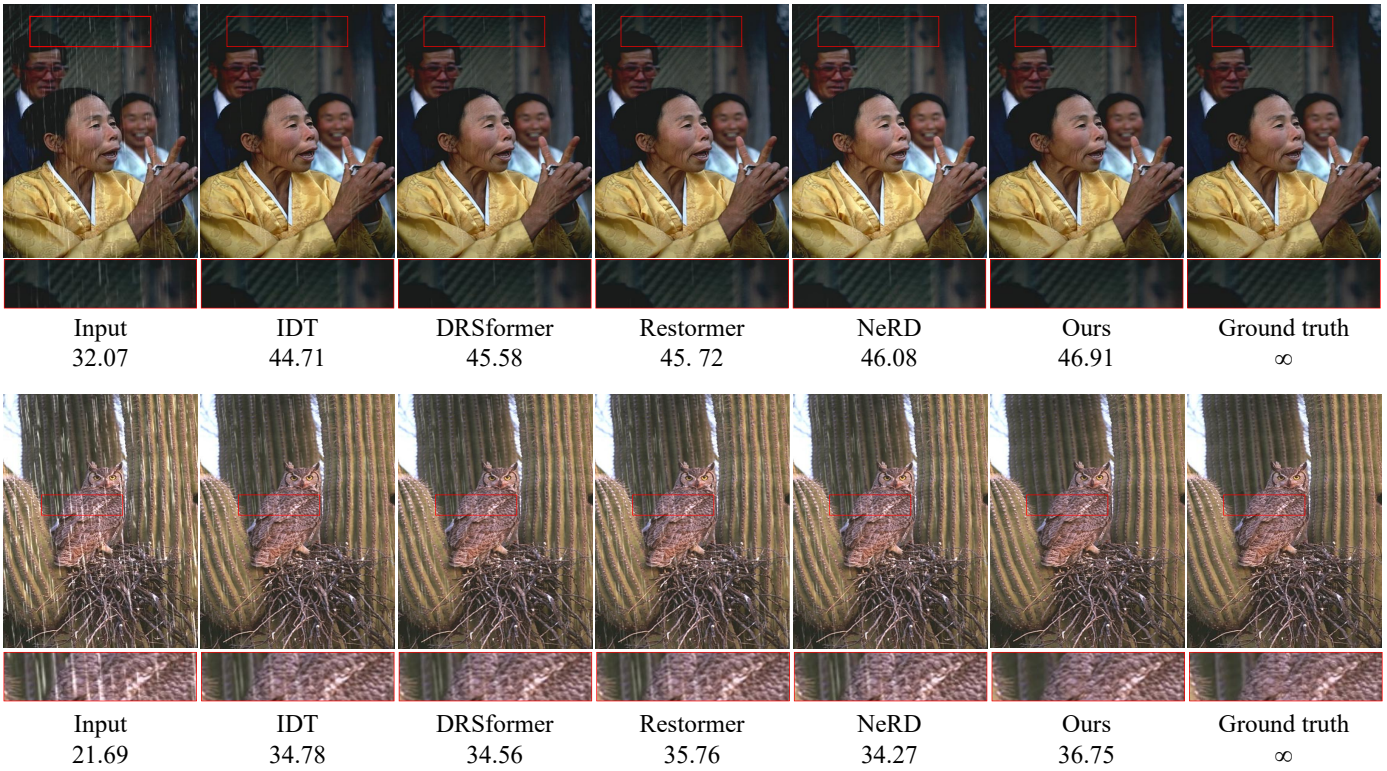

Figure 4. Visualization results on the Rain200L dataset. It is best to zoom in to see the difference.

For Rain200H, containing dense rain streaks and presenting a more challenging removal task, our approach achieved a PSNR of 32.42, with NeRD-Rain closely following at 32.40. This subtle improvement, along with our SSIM score of 0.9366, which nearly matched IDT's 0.9344, indicates the model's capacity to handle high-density rain by effectively capturing and preserving structural details. The integration of depth information into our model aids in accurately identifying background layers, even in heavily degraded

images, enabling improved separation of rain from crucial background textures. As shown in Figure 5, our results also show good restoration effects in distant local areas of the scene.

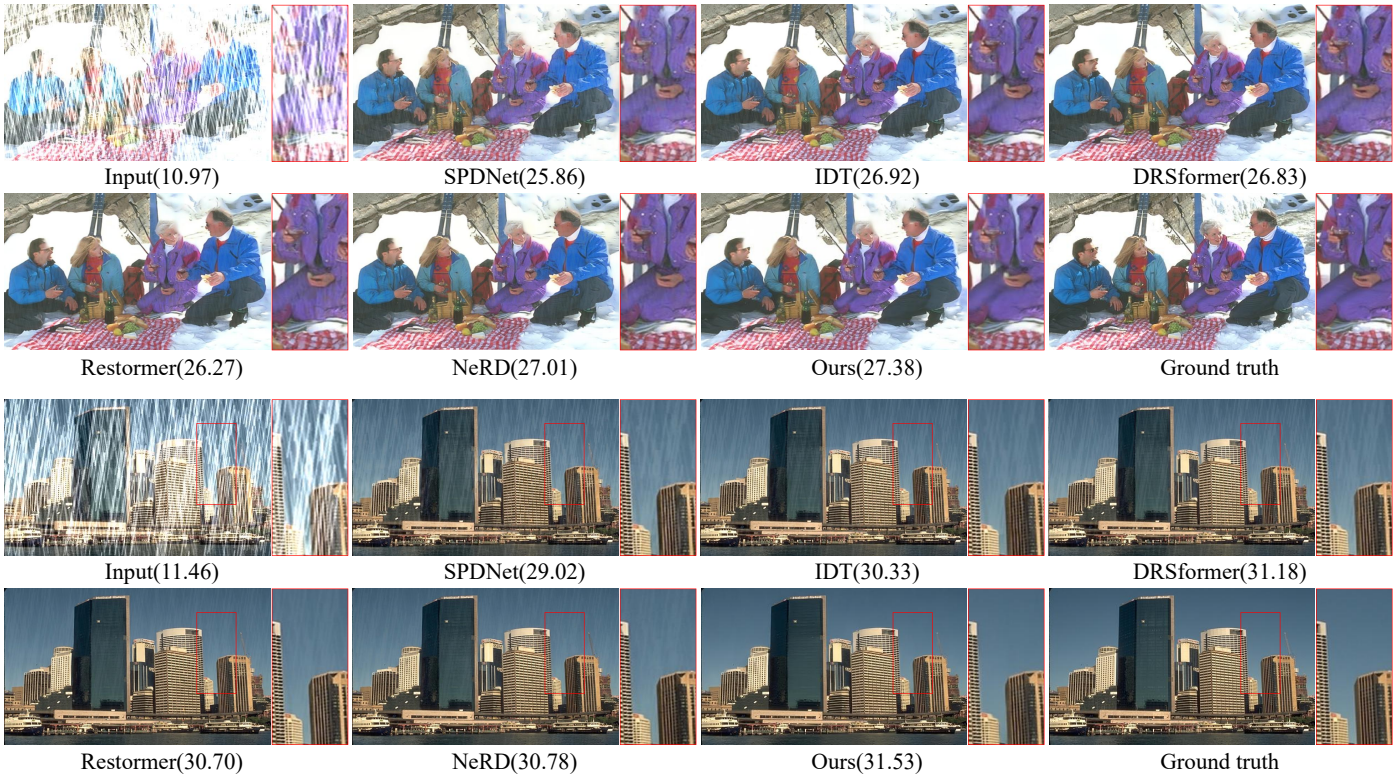

Figure 5. Visualization results on the Rain200H dataset. It is best to zoom in to see the difference. Our results are closer to ground truth and have clearer textures.

In the DID-Data dataset, which includes images with diverse rainfall directions and densities, our method achieved a PSNR of 35.63 and an SSIM of 0.9703, outperforming all other approaches. This PSNR represents a 0.3% increase over NeRD-Rain's 35.53, underscoring our model's robustness in processing complex rain patterns, where depth guidance enables enhanced consistency in background detail preservation. Similarly, on DDN-Data, another synthetic dataset with varying rain orientations and densities, our model achieved a PSNR of 34.49 and an SSIM of 0.9653, outperforming NeRD-Rain, which achieved 34.45 and 0.9596, respectively. Figures 6 and 7 showcase the visual results of our model's performance on the DID-Data and DDN-Data datasets, respectively. As seen in the figures, our method effectively restores images by removing rain degradation while preserving fine details, such as edges and textures, in both low- and high-frequency regions. Compared to NeRD-Rain, our model demonstrates better preservation of background structures and more accurate reconstruction of distant objects, especially in areas affected by varying rain directions and densities.

In the real-world SPA-Data dataset, our model attained a PSNR of 49.70, marking a 0.24% improvement over NeRD-Rain's 49.58, and an SSIM of 0.9912, closely following NeRD-Rain's 0.9940. SPA-Data's high diversity in real-world rain scenarios showcases our model's generalization capabilities. Figure 8 shows the visual comparison of various methods. It can be seen that the restored results of our method are closer to the actual ground truth and retain more details.

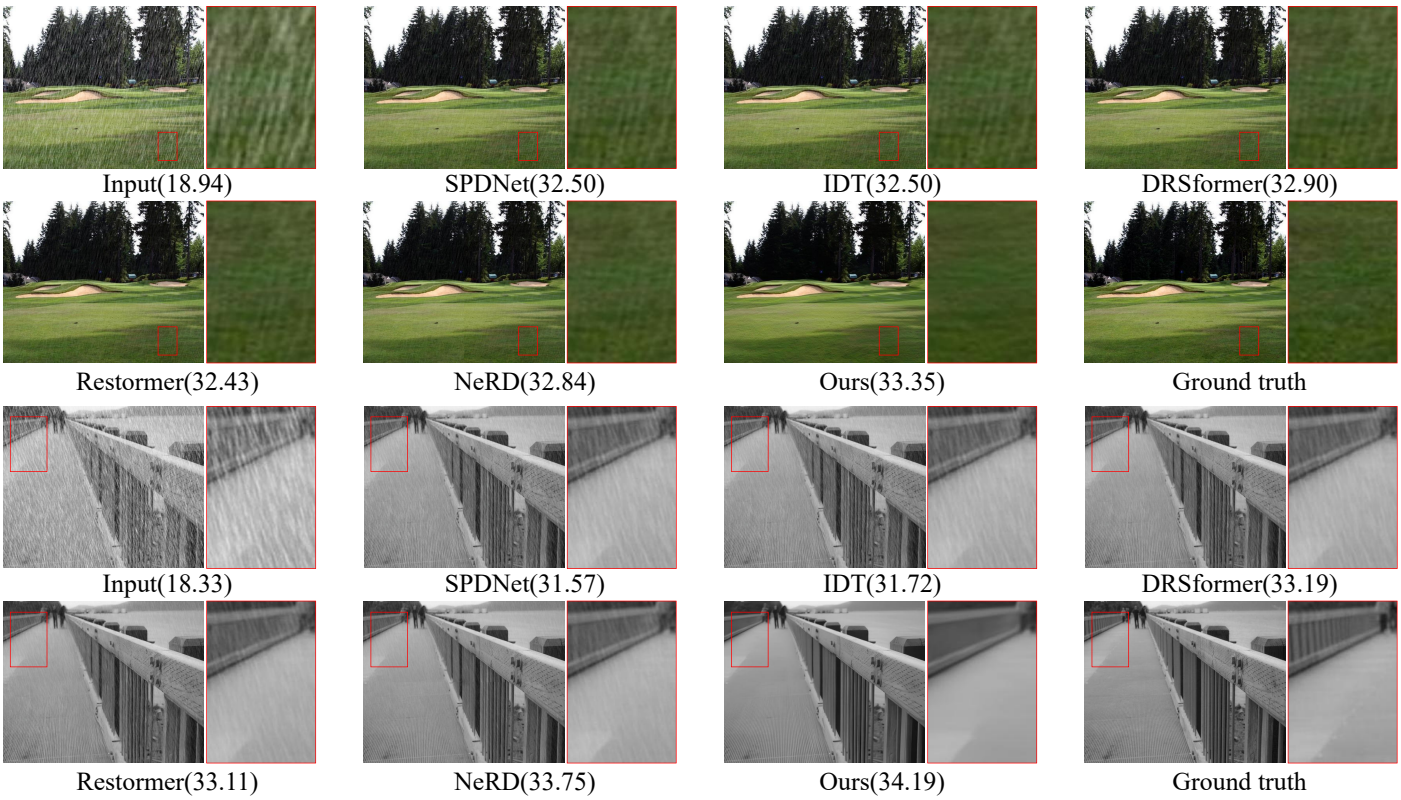

Figure 6. Visualization results on the DID-Data dataset. It is best to zoom in to see the difference.

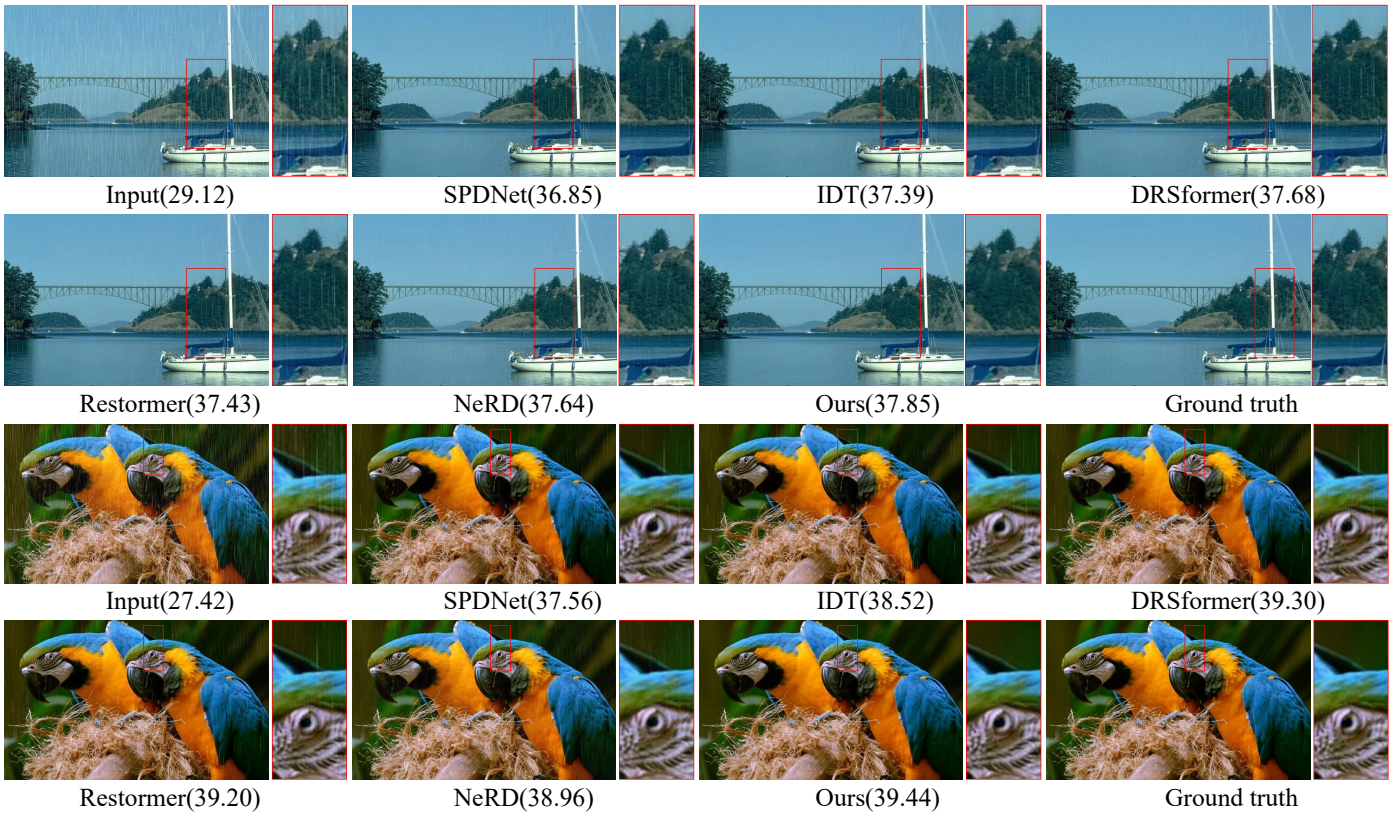

Figure 7. Visualization results on the DND-Data dataset. It is best to zoom in to see the difference.

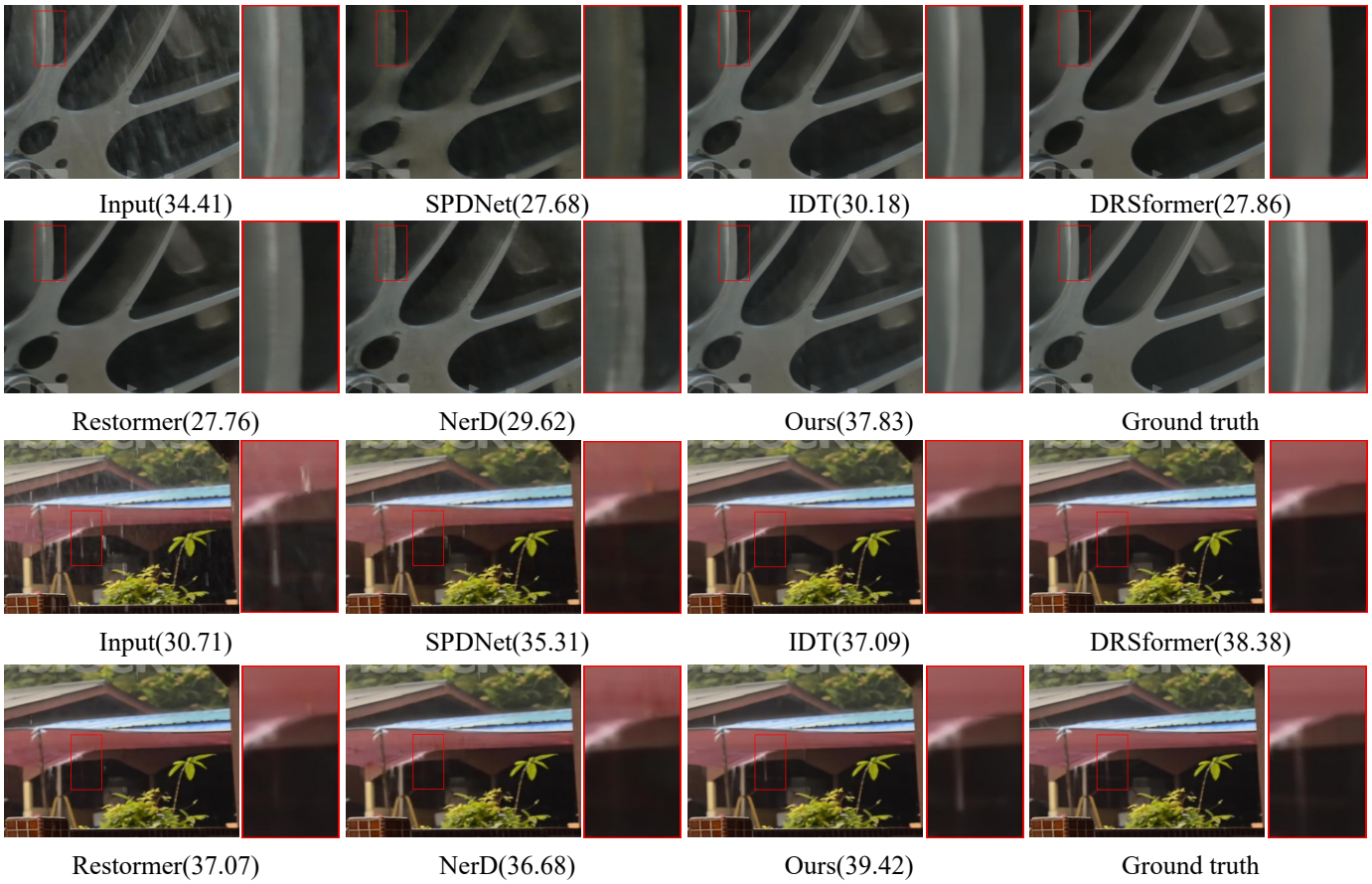

Figure 8. Visualization results on the SPA-Data dataset. It is best to zoom in to see the difference.

In summary, the consistent top performance across both synthetic and real datasets validates our model's advanced capabilities in rain removal. By leveraging depth-guided features to maintain background fidelity and frequency-domain analysis to identify and mitigate high-frequency degradations, our model achieves superior results, excelling in rain removal across diverse conditions and degradation complexities.

5. Ablation Study

5.1. The Effectiveness of Each Component

To validate the effectiveness of each proposed module in DBPNet, we conducted a comprehensive ablation study. The experiments are designed to progressively disable or replace key components and analyze their contributions. Specifically, we evaluated the following configurations:

1. Baseline (w/o FDPM and DBPM): A version of DBPNet that removes both the Frequency Degradation Perception Module (FDPM) and Depth Background Perception Module (DBPM), relying only on spatial features for rain removal.
2. Baseline + FDPM: Adds FDPM to the baseline to evaluate the impact of frequency-domain decomposition and cross-attention on degradation feature extraction.
3. Baseline + DBPM: Adds DBPM to the baseline to assess the role of depth-based background priors in guiding reconstruction.
4. Full DBPNet w/o SFA: Removes the Selective Focus Attention (SFA) module from the full DBPNet, to test the importance of selectively enhancing frequency-background interactions.
5. Full DBPNet: The complete model with all proposed modules enabled.

Discussion Impact of FDPM: Adding FDPM to the baseline significantly improved PSNR and SSIM on both datasets (+0.73 PSNR on DID-Data), highlighting the importance of explicitly decomposing frequency components for targeted degradation removal.

Impact of DBPM: Incorporating DBPM further boosted performance (+1.1 PSNR on DID-Data), demonstrating that depth-based priors enhance background consistency and reduce artifacts caused by depth-related distortions.

Role of SFA: The SFA module contributes the final refinement by aligning frequency-domain features with background priors. Removing SFA resulted in a notable drop in PSNR and SSIM, indicating its role in achieving robust and precise reconstructions.

Full Model: The complete DBPNet achieved state-of-the-art results on both datasets, showcasing the effectiveness of integrating all proposed components to handle complex rain patterns and preserve depth-consistent details.

This analysis confirms that each module in DBPNet is essential for optimal performance, with complementary contributions from FDPM, DBPM, and SFA (Table 2).

Table 2. Quantitative results of the ablation study on DID-Data and DDN-Data.

Datasets	**DID-Data**		**DDN-Data**	
Metrics	**PSNR**	**SSIM**	**PSNR**	**SSIM**
Baseline (w/o FDPM, DBPM)	34.12	0.9532	33.54	0.9487
Baseline + FDPM	34.85	0.9587	34.12	0.9556
Baseline + DBPM	35.22	0.9604	34.25	0.9563
Full DBPNet w/o SFA	35.48	0.9651	34.38	0.9594
Full DBPNet	**35.63**	**0.9703**	**34.49**	**0.9653**

5.2. Frequency Mask Design Analysis

We evaluated the impact of different frequency mask designs used in the Frequency Degradation Perception Module (FDPM) on the performance of our model. Specifically, we tested two frequency mask variants and compared their effectiveness in removing rain degradation on the Rain200L and Rain200H datasets. The following variants of the frequency mask were tested:

1. High-Frequency-Only Mask: A mask that only retains high-frequency components while discarding low-frequency components.
2. Low-Frequency-Only Mask: A mask that retains only low-frequency components, ignoring high-frequency details.

Table 3 shows the results of the decomposition experiments. The original model outperforms all variants on both the Rain200L and Rain200H datasets. It achieves the highest PSNR and SSIM values, demonstrating that the frequency separation of low- and high-frequency components is crucial for effective rain removal and image reconstruction. The High-Frequency-Only Mask shows a more noticeable decrease in both PSNR and SSIM. By neglecting the low-frequency components, it fails to retain important structural information and leads to poorer image quality, particularly for complex rain patterns where both low and high frequencies contribute to the overall image reconstruction. The Low-Frequency-Only Mask results in a moderate performance drop. While it helps retain structural integrity in the background, it loses finer details such as texture and edges, which are necessary for precise rain removal, particularly in high-frequency regions.

Table 3. Quantitative results of the frequency masks on Rain200L and Rain200H datasets.

Datasets	**Rain200L**		**Rain200H**	
Metrics	**PSNR**	**SSIM**	**PSNR**	**SSIM**
Full DBPNet (Original)	27.83	0.8967	29.52	0.9125
High-Frequency-Only Mask	26.85	0.8751	28.72	0.9009
Low-Frequency-Only Mask	27.05	0.8773	28.46	0.9044

5.3. Depth Representation Alternatives

We explored different alternatives for depth representation in our model. We examined different ways of incorporating a feature map. Specifically, we evaluated the following depth representation alternatives: 1. DINOV2: Utilizes feature maps derived from the DINOV2 model, which provides high-quality depth estimations in challenging conditions. 2. No Depth Information: A baseline where no depth information is used for reconstruction or background modeling. 3. Depth-Anything Depth: Incorporates depth information extracted from the Depth-Anything model.

Table 4 presents the quantitative results of the depth representation alternatives on the Rain200L and Rain200H datasets. The DBPNet model, which incorporates Depth-Anything Depth, achieves the best performance across both datasets. This confirms the effectiveness of integrating depth cues for improving rain removal and image reconstruction, helping the model better preserve essential details and separate background and foreground information. The No Depth Information model results in the lowest scores in both PSNR and SSIM. The model without depth fails to preserve critical details and improves significantly with depth cues. On both datasets, using DINOV2 Depth consistently improves performance compared to the No Depth Information baseline. Figure 9 shows the visual comparison results.

Table 4. Quantitative results of the depth on Rain200L and Rain200H datasets.

Datasets	**Rain200L**		**Rain200H**	
Metrics	**PSNR**	**SSIM**	**PSNR**	**SSIM**
DINOV2	27.60	0.8925	29.22	0.9095
No Depth Information	26.47	0.8654	28.36	0.8913
Depth-Anything	27.83	0.8967	29.52	0.9125

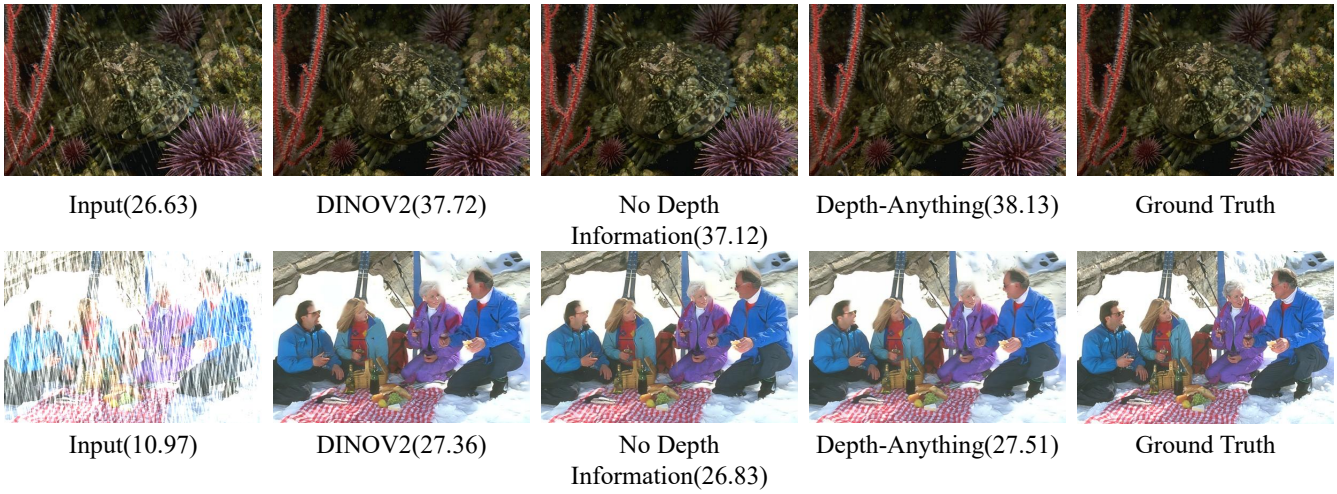

Figure 9. Visualization results of different feature-guided models.

6. Conclusions

In this paper, we proposed the Degradation-Background Perception Network (DBPNet), a novel framework designed to address the challenges of rain degradation in images. By combining frequency-domain analysis and depth-based background priors, DBPNet enhances the ability to effectively remove rain artifacts while preserving critical scene details. The framework includes two key components: the Frequency Degradation Perception Module (FDPM), which separates and refines high- and low-frequency components of the image, and the Depth Background Perception Module (DBPM), which leverages depth information to guide background reconstruction, thereby improving background detail retention. Additionally, we introduce the Selective Focus Attention (SFA) module, which strengthens the relationship between frequency-domain features and background priors, ensuring precise image reconstruction and efficient rain removal. Experimental results on the five real and synthetic rain datasets show that DBPNet outperforms existing methods, achieving superior PSNR and SSIM scores. Future directions could involve extending DBPNet to handle other image degradations and exploring the integration of additional scene understanding techniques to further enhance the model's generalization capabilities.

Author Contributions: Methodology, M.Z.; software, M.Z.; validation, S.W.; writing—original draft, M.Z.; writing—review and editing, S.W.; supervision, S.W. and S.Y. All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported by the Anhui Province Mid-Career and Young Teachers Training Initiative—Outstanding Young Teacher Cultivation Project (YQYB2023163), the Anhui Provincial Natural Science Research Project for Higher Education Institutions (2023AH053088,2024AH051439), and the Chuzhou Polytechnic Science and Technology Innovation Platform Project (YJP-2023-02).

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 conflicts of interest.

References

1. Fang, W.; Zhang, G.; Zheng, Y.; Chen, Y. Multi-Task Learning for UAV Aerial Object Detection in Foggy Weather Condition. *Remote Sens.* **2023**, *15*, 4617. [CrossRef]
2. Zhang, G.; Liu, T.; Fang, W.; Zheng, Y. Vision Transformer based Random Walk for Group Re-Identification. *arXiv* **2024**, arXiv:2410.05808.
3. Garg, K.; Nayar, S.K. Detection and removal of rain from videos. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004; CVPR 2004; IEEE: New York, NY, USA, 2004; Volume 1, p. I.
4. Li, Y.; Tan, R.T.; Guo, X.; Lu, J.; Brown, M.S. Rain streak removal using layer priors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2736–2744.
5. Ding, X.; Chen, L.; Zheng, X.; Huang, Y.; Zeng, D. Single image rain and snow removal via guided L0 smoothing filter. *Multimed. Tools Appl.* **2016**, *75*, 2697–2712. [CrossRef]
6. Reynolds, D.A. Gaussian mixture models. *Encycl. Biom.* **2009**, *741*, 3.
7. Tošić, I.; Frossard, P. Dictionary learning. *IEEE Signal Process. Mag.* **2011**, *28*, 27–38. [CrossRef]
8. Zhang, H.; Patel, V.M. Density-aware single image de-raining using a multi-stream dense network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 695–704.
9. Ren, C.; Yan, D.; Cai, Y.; Li, Y. Semi-swinderain: Semi-supervised image deraining network using swin transformer. In Proceedings of the ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; IEEE: New York, NY, USA, 2023; pp. 1–5.
10. Wang, T.; Yang, X.; Xu, K.; Chen, S.; Zhang, Q.; Lau, R.W. Spatial attentive single-image deraining with a high quality real rain dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 12270–12279.

11. Qian, R.; Tan, R.T.; Yang, W.; Su, J.; Liu, J. Attentive generative adversarial network for raindrop removal from a single image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2482–2491.
12. Wang, Q.; Jiang, K.; Wang, Z.; Ren, W.; Zhang, J.; Lin, C.W. Multi-scale fusion and decomposition network for single image deraining. *IEEE Trans. Image Process.* **2023**, *33*, 191–204. [CrossRef]
13. Zhang, G.; Fang, W.; Zheng, Y.; Wang, R. SDBAD-Net: A spatial dual-branch attention dehazing network based on meta-former paradigm. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *34*, 60–70. [CrossRef]
14. 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 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3855–3863.
15. Zou, Z.; Yu, H.; Huang, J.; Zhao, F. Freqmamba: Viewing mamba from a frequency perspective for image deraining. In Proceedings of the 32nd ACM International Conference on Multimedia, Melbourne, Australia, 28 October–1 November 2024; pp. 1905–1914.
16. Yang, L.; Kang, B.; Huang, Z.; Xu, X.; Feng, J.; Zhao, H. Depth anything: Unleashing the power of large-scale unlabeled data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 10371–10381.
17. 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 IEEE Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 17683–17693.
18. 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 Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5728–5739.
19. Chen, X.; Pan, J.; Dong, J. Bidirectional multi-scale implicit neural representations for image deraining. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 25627–25636.
20. Chen, C.; Li, H. Robust representation learning with feedback for single image deraining. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 7742–7751.
21. Chen, X.; Pan, J.; Jiang, K.; Li, Y.; Huang, Y.; Kong, C.; Dai, L.; Fan, Z. Unpaired Deep Image Deraining Using Dual Contrastive Learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 2017–2026.
22. Tao, W.; Yan, X.; Wang, Y.; Wei, M. MFFDNet: Single Image Deraining via Dual-Channel Mixed Feature Fusion. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 1–13. [CrossRef]
23. Jiang, Z.; Yang, S.; Liu, J.; Fan, X.; Liu, R. Multi-scale Synergism Ensemble Progressive and Contrastive Investigation for Image Restoration. *IEEE Trans. Instrum. Meas.* **2023**, *73*, 1–14.
24. Wei, B. DPAFNet: Dual Path Attention Fusion Network for Single Image Deraining. *arXiv* **2024**, arXiv:2401.08185.
25. He, S.; Lin, G. Gabor-guided transformer for single image deraining. *arXiv* **2024**, arXiv:2403.07380.
26. Yan, F.; He, Y.; Chen, K.; Cheng, E.; Ma, J. Adaptive Frequency Enhancement Network for Single Image Deraining. *arXiv* **2024**, arXiv:2407.14292.
27. Gao, N.; Jiang, X.; Zhang, X.; Deng, Y. Efficient Frequency-Domain Image Deraining with Contrastive Regularization. In Proceedings of the European Conference on Computer Vision, Milan, Italy, 29 September–4 October 2024; Springer: Cham, Switzerland, 2024; pp. 240–257.
28. Wang, C.; Wang, W.; Yu, C.; Mu, J. Explore Internal and External Similarity for Single Image Deraining with Graph Neural Networks. *arXiv* **2024**, arXiv:2406.00721.
29. Zhang, H.; Ba, Y.; Yang, E.; Upadhyay, R.; Wong, A.; Kadambi, A.; Guo, Y.; Xiao, X.; Wang, X.; Li, Y.; et al. GT-Rain Single Image Deraining Challenge Report. *arXiv* **2024**, arXiv:2403.12327.
30. Yu, H.; Huang, J.; Zhao, F.; Gu, J.; Loy, C.C.; Meng, D.; Li, C. Deep fourier up-sampling. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 22995–23008.
31. Mao, X.; Liu, Y.; Shen, W.; Li, Q.; Wang, Y. Deep residual fourier transformation for single image deblurring. *arXiv* **2021**, arXiv:2111.11745
32. Guo, S.; Yong, H.; Zhang, X.; Ma, J.; Zhang, L. Spatial-frequency attention for image denoising. *arXiv* **2023**, arXiv:2302.13598.
33. Li, C.; Guo, C.L.; Zhou, M.; Liang, Z.; Zhou, S.; Feng, R.; Loy, C.C. Embedding fourier for ultra-high-definition low-light image enhancement. *arXiv* **2023**, arXiv:2302.11831.
34. Cui, Y.; Tao, Y.; Ren, W.; Knoll, A. Dual-domain attention for image deblurring. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 479–487.
35. Cho, S.J.; Ji, S.W.; Hong, J.P.; Jung, S.W.; Ko, S.J. Rethinking coarse-to-fine approach in single image deblurring. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 4641–4650.

36. Fang, W.; Fan, J.; Zheng, Y.; Weng, J.; Tai, Y.; Li, J. Guided real image dehazing using ycbcr color space. *arXiv* **2024**, arXiv:2412.17496. [CrossRef]
37. Chen, W.T.; Fang, H.Y.; Hsieh, C.L.; Tsai, C.C.; Chen, I.; Ding, J.J.; Kuo, S.Y. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 4196–4205.
38. Yang, H.H.; Fu, Y. Wavelet u-net and the chromatic adaptation transform for single image dehazing. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, China, 22–25 September 2019; IEEE: New York, NY, USA, 2019; pp. 2736–2740.
39. Zou, W.; Jiang, M.; Zhang, Y.; Chen, L.; Lu, Z.; Wu, Y. Sdwnet: A straight dilated network with wavelet transformation for image deblurring. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1895–1904.
40. Yang, H.H.; Yang, C.H.H.; Tsai, Y.C.J. Y-net: Multi-scale feature aggregation network with wavelet structure similarity loss function for single image dehazing. In Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; IEEE: New York, NY, USA, 2020; pp. 2628–2632.
41. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
42. Jiang, H.; Larsson, G.; Shakhnarovich, M.M.G.; Learned-Miller, E. Self-supervised relative depth learning for urban scene understanding. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 19–35.
43. Chen, P.Y.; Liu, A.H.; Liu, Y.C.; Wang, Y.C.F. Towards scene understanding: Unsupervised monocular depth estimation with semantic-aware representation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2624–2632.
44. Hu, X.; Fu, C.W.; Zhu, L.; Heng, P.A. Depth-attentional features for single-image rain removal. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 8022–8031.
45. Oquab, M.; Darcet, T.; Moutakanni, T.; Vo, H.; Szafraniec, M.; Khalidov, V.; Fernandez, P.; Haziza, D.; Massa, F.; El-Nouby, A.; et al. Dinov2: Learning robust visual features without supervision. *arXiv* **2023**, arXiv:2304.07193.
46. 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 on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1357–1366.
47. Huynh-Thu, Q.; Ghanbari, M. Scope of validity of PSNR in image/video quality assessment. *Electron. Lett.* **2008**, *44*, 800–801. [CrossRef]
48. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
49. 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, 13–19 June 2020; pp. 8346–8355.
50. Fu, X.; Xiao, J.; Zhu, Y.; Liu, A.; Wu, F.; Zha, Z.J. Continual Image Deraining with Hypergraph Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 9534–9551. [CrossRef] [PubMed]
51. Luo, Y.; Xu, Y.; Ji, H. Removing rain from a single image via discriminative sparse coding. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 3397–3405.
52. 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), Munich, Germany, 8–14 September 2018; pp. 254–269.
53. 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, 15–20 June 2019; pp. 3937–3946.
54. Wang, H.; Xie, Q.; Zhao, Q.; Meng, D. A model-driven deep neural network for single image rain removal. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3103–3112.
55. 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 International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 14821–14831.
56. Fu, X.; Qi, Q.; Zha, Z.J.; Zhu, Y.; Ding, X. Rain streak removal via dual graph convolutional network. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; pp. 1352–1360.
57. Yi, Q.; Li, J.; Dai, Q.; Fang, F.; Zhang, G.; Zeng, T. Structure-Preserving Deraining with Residue Channel Prior Guidance. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 4238–4247.

58. Xiao, J.; Fu, X.; Liu, A.; Wu, F.; Zha, Z.J. Image De-raining Transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 12978–12995. [CrossRef] [PubMed]
59. Chen, X.; Li, H.; Li, M.; Pan, J. 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. 5896–5905.