technologies

MDPI

*Article*

# An Image Deraining Network Integrating Dual-Color Space and Frequency Domain Prior

**Luxia Yang [1,2], Yiying Hou [1,2] and Hongrui Zhang [1,2,*]**

[1] School of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, China
[2] Shanxi Provincial Key Laboratory of Intelligent Optimization Computing and Blockchain Technology, Jinzhong 030619, China
* Correspondence: zhanghongrui@tynu.edu.cn

**Abstract**

Image deraining is a crucial preprocessing task for enhancing the robustness of high-level vision systems under adverse weather conditions. However, most of the existing methods are limited to a single RGB color space, and it is difficult to effectively separate high-frequency rain streaks from low-frequency backgrounds, resulting in color distortion and detail loss in the restored image. Therefore, a rain removal network that combines dual-color space and frequency domain priors is proposed. Specifically, the devised network employs a dual-branch Transformer architecture to extract color and structural features from the RGB and YCbCr color spaces, respectively. Meanwhile, a Hybrid Attention Feedforward Block (HAFB) is constructed. HAFB achieves feature enhancement and regional focus through a progressive perception selection mechanism and a multi-scale feature extraction architecture, thereby effectively separating rain streaks from the background. Furthermore, a Wavelet-Gated Cross-Attention module is designed, including a Wavelet-Enhanced Attention Block (WEAB) and a Dual Cross-Attention module (DCA). This design enhances the complementary fusion of structural information and color features through frequency-domain guidance and bidirectional semantic interaction. Finally, experimental results on multiple datasets (i.e., Rain100L, Rain100H, Rain800, Rain12, and SPA-Data) demonstrate that the proposed method outperforms other approaches.

**Keywords:** image deraining; dual-color space; frequency domain prior; transformer; YCbCr

## 1. Introduction

Image deraining is a fundamental low-level vision task in computer vision [1], which aims to recover clear background content from rain-degraded images. This technology is crucial for improving the robustness and reliability of high-level vision systems, such as autonomous driving [2–5], video surveillance [6–8], and object detection [9–12], in adverse weather conditions.

Early deraining approaches mainly rely on traditional image processing techniques and manually designed priors to achieve rain streaks separation, using filtering or physical models such as sparse coding [13] and Gaussian mixture models [14]. However, these methods impose strong assumptions on the shape and distribution of rain streaks, which limits their adaptability in complex real-world scenarios. With the rise of deep learning, CNN-based methods have achieved notable progress by introducing strategies like residual learning [15], joint detection and removal [16], and multi-stage progressive refinement [17], substantially improving deraining performance. Although some progress has been made,



Academic Editor: Sikha Bagui

Received: 28 December 2025
Revised: 27 January 2026
Accepted: 31 January 2026
Published: 4 February 2026

these models are still limited by the local receptive field of CNNs, which makes it difficult to capture long-term dependence of irregular and non-local rain streaks in nature.

More recently, Transformers have demonstrated considerable potential in image deraining due to their powerful global modeling capability. Methods such as Restormer [18], IDT [19], and DRSformer [20] effectively model global contextual information. At the same time, it has been shown that the selection of color space significantly influences deraining performance. Different color spaces provide distinct advantages in terms of luminance-chrominance separation and detail preservation, offering further room for improvement. Nevertheless, most existing methods still perform feature extraction and fusion in a single color space, failing to fully exploit the complementary characteristics of multiple color representations. In particular, establishing an effective deep feature interaction mechanism across color spaces remains an open challenge.

To address these limitations, we propose an image deraining network that integrates dual-color space and frequency domain prior. The architecture of this network is dual-branch Transformer, processing image information in both RGB and YCbCr color spaces. This allows the advantages of the color richness of RGB and the structural sensitivity of YCbCr to be fully utilized. Meanwhile, a Hybrid Attention Feedforward Block is devised to enhance the capability of the model to perceive and suppress multi-scale rain streaks. Furthermore, we design a Wavelet Gated Cross-Attention module to enable refined feature interaction across color spaces in the frequency domain. The main contributions of this paper are summarized as follows:

(1) We propose a dual-branch Transformer that processes RGB and YCbCr information simultaneously. By extracting color-rich features from the RGB branch and structure-sensitive cues from the YCbCr branch, the model's ability to discriminate between rain streaks and background content is enhanced.

(2) We construct a Hybrid Attention Feedforward Block (HAFB) by cascading a Progressive Perception Selection Block (PPSB) with a Multi-scale Adaptive Fusion Feedforward Network (MAFFN). This design enables adaptive suppression of multi-scale rain streaks while preserving image details and structures.

(3) We design a Wavelet Gated Cross-Attention mechanism that achieves frequency domain feature decomposition and enhancement via wavelet transform. It also establishes cross-color-space feature alignment and semantic guidance through bidirectional cross-attention, thereby improving the fusion of color and structural information.

## 2. Related Work

### *2.1. Image Deraining*

Image deraining has undergone a series of evolutions, progressing from early model-based approaches to advanced data-driven techniques. Based on convolutional neural network methods, such as JORDER [16], the rain removal problem is constructed as two collaborative learning subtasks: rain streak detection and rain streak removal. Furthermore, its sequential strategy enhances the model's ability to handle complex rainfall patterns. Following this, RESCAN [21] employed a context dilated network and an attention mechanism to remove rain streaks in a localized manner, while PreNet [17] enhanced model interpretability by incorporating prior knowledge about the repeatability of rain streak distributions into network design. Subsequently, research expanded into multi-scale representation learning. For example, MSPFN [22], employed a Gaussian pyramid framework, progressively refining results through coarse-to-fine and fine-to-fine modules, significantly enhancing its ability to represent rain streaks of varying sizes and shapes. Similarly, MPRNet [23] utilized a multi-stage architecture to progressively reconstruct images. It enhanced detail restoration capabilities through feature fusion among stages, thereby achieving

clearer details. In addition, CCN (Complementary Cascaded Network) [24] unified the processing of rain streaks and raindrops within a single framework for the first time. It employed a complementary dual-branch architecture optimized via neural architecture search and introduced the real-world rain image dataset RainDS. This work marked a significant breakthrough in convolutional neural network-based methods for complex real-world deraining. feature fusion among stages, thereby achieving clearer details.

The principal limitation of convolutional neural network is the inherent nature of their local receptive fields, which restricts the model's ability to capture long-range dependencies. To address this, researchers introduced the Transformer architecture, whose self-attention mechanism enables global context modeling. In this direction, IPT [25] pioneered the application of large-scale pre-trained Transformers to image processing and demonstrated their effectiveness in tasks such as rain removal. Its multi-task adaptive architecture and pre-training paradigm established a new research direction in this field that integrates global modeling with pre-training. Subsequently, a series of Transformer-based deraining methods emerged. For example, Restormer [18] achieved high-quality deraining on high-resolution images by leveraging a multi-headed transposed attention mechanism and a gated feed-forward network. Similarly, Uformer [26] adopted a hierarchical design similar to U-Net and integrated a window-based local attention mechanism, which achieved an effective balance between global context integration and computational feasibility. Further extending this line, TransWeather [27] proposed a single encoder–decoder Transformer network that achieved joint perception and recovery of different degradation types by introducing learnable weather type queries and an internal block attention mechanism. In addition, APANet [28] employed External Attention Moudle (EAM) to model inter-sample correlations. It also introduced an Asymmetric Parallax Attention Module (APAM), which utilized stereoscopic geometric constraints to filter out invalid cross-view interactions, thereby improving deraining performance. In summary, the success achieved by Transformer-based models demonstrates the critical significance of global contextual information for complex rain removal tasks, providing strong evidence for selecting Transformers as the backbone network.

### *2.2. Color Spaces in Image Restoration*

In the field of image processing, the selection of color space is a critical factor influencing enhancement outcomes, as different color spaces represent luminance, color, and contrast information in distinct ways, thereby directly shaping the design and performance of subsequent algorithms. For low-light image enhancement, Yan et al. [29] introduced the HVI color space, which builds upon the HSV space by incorporating a polarized Hue-Saturation mapping and a learnable intensity compression function. This approach effectively alleviates red discontinuity and black-plane noise while achieving efficient decoupling between luminance and color. In underwater image enhancement, Zhang et al. [30] adopted the HSV color space, applying adaptive equalization separately to the saturation and value channels, and further combined it with multi-scale fusion techniques to significantly improve image contrast and detail visibility while preserving color naturalness. The aforementioned studies demonstrated that different color spaces could enhance the quality of image restoration. However, despite the significant progress achieved in their respective tasks, their core focus remained on optimizing a single color space. In contrast, our method leveraged the RGB space to preserve color naturalness and structural coherence, while the YCbCr space, through its linear and orthogonal separation, decomposed the image into luminance and chrominance components, thereby enabling independent and targeted enhancement. By effectively integrating features from both color

spaces, our approach achieved more balanced color fidelity and superior detail restoration in complex degraded scenarios.

### *2.3. Wavelet Transform in Vision Tasks*

Frequency-domain methods have become a valuable tool for enhancing the performance of deep neural networks in image restoration. A key advantage lies in their ability to decompose an image into distinct frequency sub-bands, each encapsulating different types of information. Wavelet-SRNet [31] first shifted the restoration task into the wavelet domain, directly predicting both high- and low-frequency coefficients to jointly reconstruct the image. MWCNN [32] further advanced this approach by developing the discrete wavelet transform into a general network component, replacing pooling operations to achieve lossless multi-scale feature extraction. Subsequent frequency-aware decomposition and enhancement methods [33], though not explicitly wavelet-based, designed a two-stage network grounded in the observation that noise exhibits distinct behaviors across frequency bands. This design follows a sequential strategy of first denoising and enhancing the low-frequency components, and then restoring the high-frequency details. Most recently, DiffLL [34] innovatively integrated wavelet transform with diffusion models, performing the computationally intensive generation process only on highly compressed low-frequency sub-bands, while recovering high-frequency details in parallel via a lightweight module, significantly improving efficiency. In summary, the aforementioned works fundamentally rely on the multi-resolution analysis capability of wavelet transform to achieve targeted decoupling and reconstruction of image content. This has continuously advanced the development of image restoration techniques.

These methods are grounded in a common physical prior. Specifically, in the frequency-domain representation, low-frequency components convey the main structure and background information of the image, while high-frequency components concentrate on edges, details, and noise. This property aligns closely with the intrinsic requirements of tasks such as rain removal, thereby offering clear guidance for network design. Building on this consensus, we further introduce an approach that integrates cross-color-space attention. While preserving the advantage of frequency-domain decoupling, this method enhances the distinction between rain streaks and background through cross-channel feature interaction.

## 3. Method

In this section, we first describe the overall design of the proposed deraining network. Then, the details of each designed module are presented.

### *3.1. Overview*

As shown in Figure 1, our network mainly contains three key designs: Dual-color Space Branches (DSB), Hybrid Attention Feedforward Block (HAFB), and Wavelet-Gated Cross-Attention (WGCA) module. Specifically, the DSB employs a siamese branch Transformer architecture to establish parallel processing streams for the RGB and YCbCr color spaces, thereby making full use of the complementary characteristics of different color spaces. HAFB enhances the ability of the model to perceive and suppress complex rain streaks through the Progressive Perception Selection Block (PPSB) and the Multi-scale Adaptive Fusion Feedforward Network (MAFFN), while simultaneously promoting the restoration of background details. WGCA is composed of a Wavelet-Enhanced Attention Block (WEAB) and a Dual Cross-Attention Module (DCA), which aims to establish a bidirectional semantic guidance mechanism.

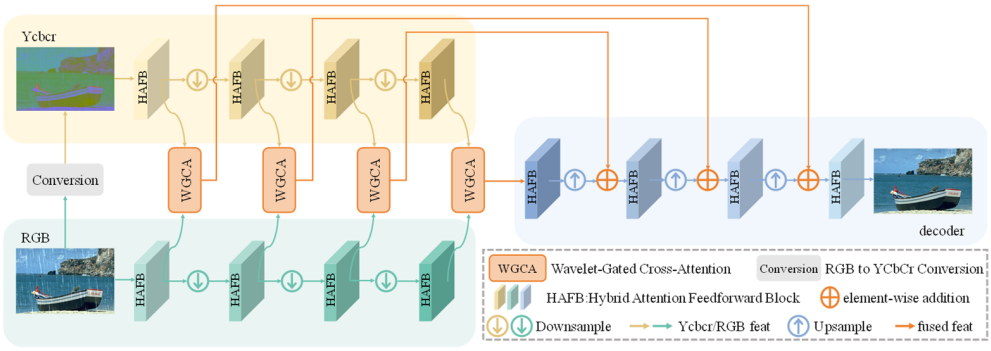


**Figure 1.** Overall Framework of the Proposed Model.

### *3.2. Dual-Color Space Branches*

The intrinsic coupling of luminance and chrominance in the RGB color space constitutes a fundamental limitation in deep learning-based image deraining, leading to persistent color distortion and loss of detail. Although the HSV color space is intuitive for tasks such as video processing, it is suboptimal for single-image deraining because its nonlinear transformation lacks a physical basis for separating rain streaks. In contrast, the YCbCr color space linearly and orthogonally separates image data into a luminance component (Y) and chrominance components (Cb, Cr). This separation provides a representation with a clear physical interpretation and high computational efficiency. However, the process can still compromise color fidelity, particularly in highly saturated regions. Therefore, a deraining network capable of leveraging the complementary advantages of both RGB and YCbCr spaces is desired.

Accordingly, we design a novel deraining network with a parallel dual-branch architecture to process information from both the RGB and YCbCr color spaces. This design fully leverages the strengths of the RGB color space for color reproduction and the YCbCr color space for structural perception. By doing so, it alleviates the feature representation constraints inherent in using a single color space, thereby improving color consistency and preserving fine details in the output. Specifically, the network employs a dual-branch encoder for separately extracting features from the two color spaces. Each branch is built upon Hybrid Attention Feedforward Block (HAFB) to enhance multi-scale rain streak perception and suppression. Additionally, a Wavelet-Gated Cross-Attention (WGCA) module is designed to achieve deep feature interaction and semantic guidance across color spaces in frequency domain, thereby improving the deraining performance and detail reconstruction quality.

### *3.3. Hybrid Attention Feedforward Block*

In image deraining tasks, rain streaks typically exhibit multi-scale characteristics and an uneven spatial distribution. The feedforward network in conventional Transformer architectures tends to have a homogeneous structure, which is often insufficient for modeling such complex rain patterns. Furthermore, existing attention mechanisms exhibit specific limitations. Channel attention mechanisms primarily capture interdependencies among channels, yet they frequently neglect the spatial distribution of rain streaks. Conversely, spatial attention mechanisms focus on local or global contextual information but they may not adequately differentiate the importance of features across channels. Although standard self-attention can model long-range dependencies, they impose higher computational costs and lack coordination between channel and spatial dimensions, which can lead to the loss of structural details during rain suppression.

To solve this issue, we design a Hybrid Attention Feedforward Block (HAFB), as illustrated in Figure 2. The proposed module features a progressive channel-spatial attention

mechanism that first utilizes half of the input channels to generate a data-dependent mask, thereby suppressing redundant features while enhancing informative representations. Subsequently, it performs joint channel and spatial recalibration to emphasize multi-scale features intrinsically fused with scene content. When integrated with a multi-scale feedforward network, the HAFB adaptively mitigates multi-scale rain streak interference and facilitates background detail recovery. As a result, the overall design contributes to notable improved image reconstruction quality under complex rainy conditions.

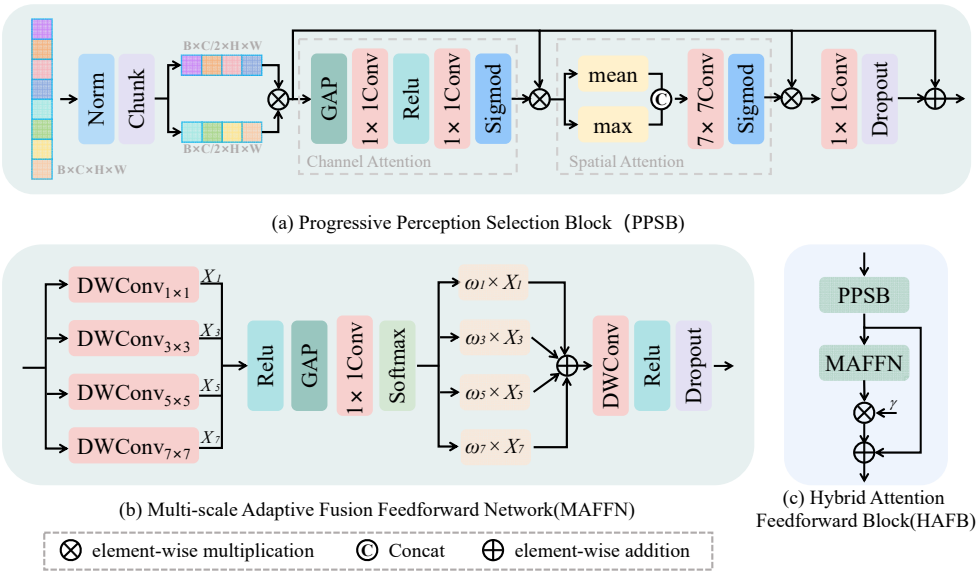


**Figure 2.** Structure of Hybrid Attention Feedforward Block (**a**) Progressive Perception Selection Block (PPSB); (**b**) Multi-scale Adaptive Fusion Feedforward Network (MAFFN); (**c**) Hybrid Attention Feedforward Block (HAFB).

The HAFB module integrates two components: the Progressive Perception Selection Block (PPSB) and the Multi-scale Adaptive Fusion Feedforward Network (MAFFN).

As shown in Figure 2a, the PPSB incorporates a serial dual-stage attention mechanism to enhance the discriminative feature response and focus on critical structural areas from the channel dimension and the spatial dimension. PPSB begins with the implementation of feature compression and nonlinear enhancement via channel splitting and interaction. In this way, it promotes the separation of rain pattern features from the background content. Subsequently, the channel weight is generated using global average pooling and a multi-layer perceptron to effectively suppress the channel response dominated by rain streaks. Then, the statistical information of channel weighted features in spatial dimension is fused and the weight of spatial structure is extracted by convolution operation, so as to accurately identify the area of rain streak aggregation and implement spatial suppression. Finally, the residual connection preserve the integrity of background structures, which enhances the feature expression ability and reduces the rain streak interference.

The MAFFN, depicted in Figure 2b, systematically processes multi-scale rain features using a parallel architecture of depthwise separable convolutions. This module employs four convolutional kernels of different sizes to extract features at various levels, ranging from local details to global contextual information. These features undergo dynamic weighting and fusion through adaptive weighting, which facilitates the collaborative suppression and removal of multi-scale rain noise. Through the aforementioned design, the model autonomously adjust its strategy for utilizing multi-scale information based on the features of the input, effectively eliminating rain interference while preserving essential image details.

### *3.4. Wavelet-Gated Cross-Attention Module*

In image deraining, the core challenge lies in effectively separating rain streaks from the image background. Rain streaks typically appear as localized high-frequency textures in the spatial domain, whereas the primary structure and smooth background of an image mainly reside in low-frequency components. Therefore, feature decoupling in the frequency domain offers a distinct advantage. With its multiresolution decomposition capability, the Discrete Wavelet Transform (DWT) naturally separates features into different subbands: the low-frequency subband (LL) conveys global contours and background content, while the high-frequency subbands (LH, HL, HH) capture various high-frequency details such as edges, fine textures, and noise. This inherent frequency separation allows the model to guide low-frequency components toward background reconstruction while directing high-frequency components to suppress noise such as rain streaks and simultaneously preserve or enhance meaningful image details. However, real-world scene structures exhibit consistent features across different color spaces, whereas rain streaks as additive noise typically lack such consistency. Relying solely on frequency-domain processing is insufficient to fully extract complementary information between RGB and YCbCr color spaces. To address this limitation, we propose the Wavelet-Gated Cross-Attention (WGCA) module. By jointly modeling frequency-domain decomposition and cross-space consistency, WGCA aligns and fuses features from complementary color spaces, thereby achieving enhanced integration of cross-color-space representations. The WGCA architecture is shown in Figure 3.

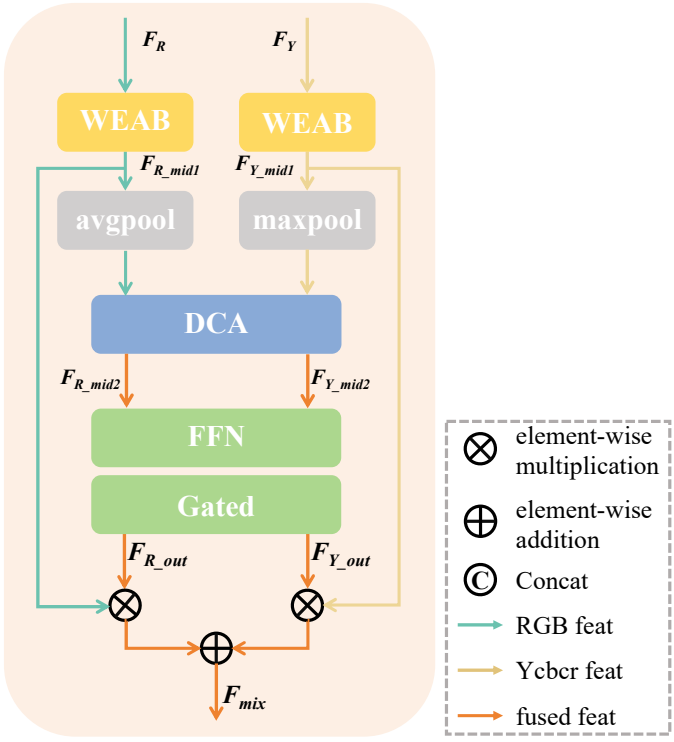


**Figure 3.** Architecture of Wavelet-Gated Cross-Attention Module.

The WGCA comprises two core sub-modules: (1) the Wavelet-Enhanced Attention Block (WEAB), which focuses on low-frequency background and high-frequency texture details by performing frequency-domain decoupling and enhancement of input features. (2) the Dual Cross-Attention Module (DCA) establishes bidirectional semantic dependencies and guides the interaction and reconstruction of features between the two color spaces.

3.4.1. Wavelet-Enhanced Attention Block

In image deraining tasks, rain streaks exhibit complex distribution characteristics in both spatial and frequency domains. In the proposed dual-branch architecture, the feature attributes of the RGB and YCbCr color spaces demonstrate a complementary relationship. Therefore, to achieve effective deep fusion, we design a preprocessing mechanism called the Wavelet-Enhanced Attention Block (WEAB) that enhances the semantic consistency and expression capacity of features in each branch. Using this mechanism, the input features are decoupled and enhanced in the frequency domain to provide high-quality input for subsequent fusion. Figure 4 depicts the workflow of the designed WEAB.

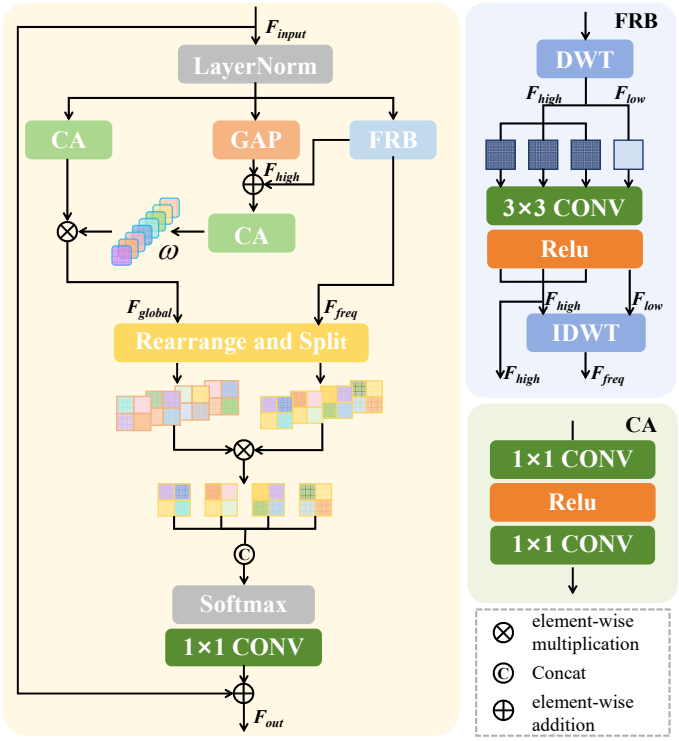


**Figure 4.** Detailed architecture of Wavelet-Enhanced Attention Block.

First, the input features $F_{input}$ undergoes wavelet decomposition to yield the low-frequency component $F_{low}$ that characterizes background structure and the high-frequency component $F_{high}$ carrying detailed rain streak information. Each component then undergoes separate convolutional processing. For the low-frequency component, convolution enhances structural information to maintain global image coherence. For high-frequency components, convolution is used for edge enhancement and noise suppression, highlighting the detailed features of rain streaks. Subsequently, the frequency domain enhancement feature $F_{freq}$ is obtained by inverse wavelet transform reconstruction, which corresponds to the frequency domain enhancement branch output in the module. This process can be expressed by the following equations.

$$F_{low}, F_{high} = \mathrm{DWT}(F_{input}) \tag{1}$$

$$F_{freq} = \mathrm{IDWT}(\mathrm{CONV}(F_{low}, F_{high})) \tag{2}$$

where DWT denotes the Discrete Wavelet Transform and IDWT refers to the Inverse Discrete Wavelet Transform.

Meanwhile, global average pooling is applied to the input features $F_{input}$ to extract global contextual information, which is then fused with the energy statistics feature $E_{high}$

of the high-frequency components. Next, the channel attention module generates weight coefficient $\omega$ with frequency domain perception capabilities. The weight coefficient further modulate the channel features processed by the channel attention module, yielding the channel-modulated features $F_{mod}$. By doing this, adaptive enhancement is accomplished for rain-streak-sensitive channels, which can be expressed as follows:

$$E_{high} = \left\| \|F_{high}\|_{3,4} \right\|_2 \tag{3}$$

$$\omega = \mathrm{CA}(\mathrm{GAP}(F_{input}) + E_{high}) \tag{4}$$

$$F_{mod} = CA(F_{input}) \otimes \omega \tag{5}$$

where $(\|\cdot\|_{3,4})$ denotes the L2-norm computed along the spatial dimensions of the high-frequency components $F_{high}$, $(\|\cdot\|_2)$ denotes the L2-norm computed along the directional dimensions, and CA denotes the Channel Attention module.

After that, the frequency-enhanced feature $F_{freq}$ and channel-modulated feature $F_{mod}$ are recombined and spatially partitioned into a multi-head representation divided into multiple sub-regions. In each sub-region, feature interaction is performed through element-wise multiplication to enhance local rain streak detail capture. The outputs from all sub-regions are then concatenated and spatially normalized via a Softmax function to establish cross-region global dependencies, ensuring spatial consistency in rain streak removal.

Finally, the fused features $F_{fused}$ are mapped back to the original channel dimension via a convolutional layer and connected to the module input $F_{input}$ through a residual connection. As a result, the original background structure is retained while removing the rain streaks, which not only avoids the loss of details caused by excessive smoothing, but also ensures the stability of network training and the quality of restored images. The above process can be represented by the following equations.

$$\left\{F_{freq}^1, \ldots, F_{freq}^K\right\} = \mathrm{Split}(\mathrm{Rearrange}(F_{freq})) \tag{6}$$

$$\left\{F_{mod}^1, \ldots, F_{mod}^K\right\} = \mathrm{Split}(\mathrm{Rearrange}(F_{mod})) \tag{7}$$

$$F_{fused} = \mathrm{Softmax}(\mathrm{Conv}(\{F_{freq}^i \otimes F_{mod}^i \,|\, i = 1, \ldots, k\})) \tag{8}$$

$$F_{out} = \mathrm{Conv}(F_{fused}) + F_{input} \tag{9}$$

#### 3.4.2. Dual Cross-Attention Module

The WEAB module improves the quality of feature representation within each branch by using frequency-domain enhancement. However, there are significant differences in feature distribution between the RGB and YCbCr color spaces. Here, the RGB branch tends to lose high-frequency edges and texture details under rain degradation. In contrast, the YCbCr branch is more effective at analyzing and preserving structural information, but suffers from insufficient color information and weaker hue reproduction capabilities. This asymmetry in characteristics poses significant challenges for cross-space information fusion and limits further improvements in rain removal performance.

To address these limitations, we propose a Dual Cross Attention (DCA) module to promote deep interaction across color spaces via a bidirectional semantic guidance mechanism. As shown in Figure 5, the DCA fully utilizes the enhanced features from WEAB to establish bidirectional attention flows of RGB→YCbCr and YCbCr→RGB. This design improves the quality of feature fusion by strengthening structural and color information in a complementary manner.

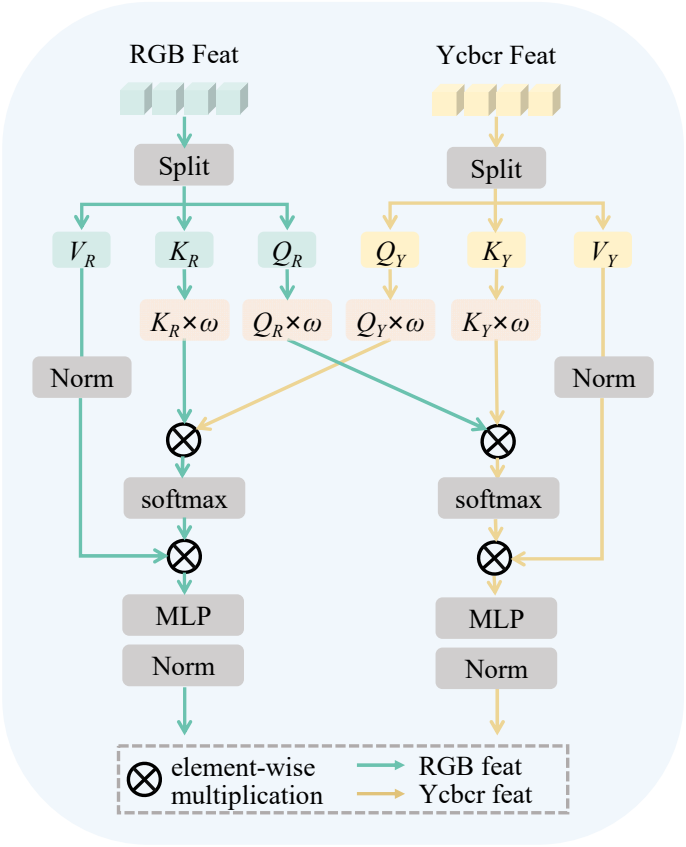


**Figure 5.** Architecture of Dual Cross Attention Module.

First, the enhanced features from both color spaces are transformed into a unified semantic embedding space via linear projection, where corresponding query vectors, key vectors, and value vectors are generated. The process is represented as follows:

$$Q_R, K_R, V_R = \text{Split}(\text{Linear}(F_{RGB})) \tag{10}$$

$$Q_Y, K_Y, V_Y = \text{Split}(\text{Linear}(F_{YCbCr})) \tag{11}$$

Subsequently, the keys and query vectors undergo normalization processing, and learnable frequency-domain attention weights $\omega_f$ are introduced to enhance the model's ability to perceive high-frequency rain-induced texture structures. This process can be expressed by the following equation.

$$\hat{Q}_R = Q_R \otimes \omega_f, \hat{K}_R = K_R \otimes \omega_f \tag{12}$$

$$\hat{Q}_Y = Q_Y \otimes \omega_f, \hat{K}_Y = K_Y \otimes \omega_f \tag{13}$$

Then, the structure guidance and color compensation across color spaces are achieved through bidirectional cross-attention. Specifically, RGB→YCbCr aligns color information with YCbCr's structural priors to enhance edge details, while YCbCr→RGB utilizes structural information to guide color reconstruction in the RGB space. Color consistency is improved through cross-guidance between features across both color spaces. The above process can be represented as follows:

$$A_{R \to Y} = \text{Softmax}\left(\frac{\hat{Q}_R \hat{K}_Y^T}{\sqrt{d_k}}\right) V_Y \tag{14}$$

$$A_{Y\to R} = \text{Softmax}\left(\frac{\hat{Q}_Y \hat{K}_R^T}{\sqrt{d_k}}\right) V_R \tag{15}$$

Finally, the attention outputs from both directions are fused through a multi-layer perceptron and concatenated. Subsequently, a linear projection yields the fused result. This process can be expressed by the following equation.

$$O = \text{Linear}([\text{MLP}(A_{R\to Y}), \text{MLP}(A_{Y\to R})]) \tag{16}$$

In summary, the DCA module establishes a bidirectional feature interaction mechanism between the RGB and YCbCr color spaces. Using this design, color and structural information can be effectively combined to generate semantically rich fusion features for subsequent image reconstruction. Consequently, the model's restoration capability in complex rainy scenarios is improved.

## 4. Experiments

### *4.1. Experimental Setup*

Datasets. To validate our approach, we trained and evaluated models on widely used datasets Rain100L [16], Rain100H [16], Rain800 [35], and Rain12 [14]. Rain100L [16] contains images with sparse rain streaks, consisting of 200 training and 100 test images. Rain100H [16] features dense rain streaks, with 1800 training and 100 test images. Rain800 [35] is a dataset of moderate complexity, comprising 700 training and 100 test images. Rain12 [14] is a small-scale test set containing 12 test images.

Evaluation Metrics. The Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index (SSIM) are adopted as the primary evaluation metrics to quantify model performance in terms of signal recovery accuracy and structural consistency, respectively. To mitigate boundary effects, a 4-pixel-wide border is cropped from each image before metric calculation. Additionally, all metrics were computed based on the luminance channel (Y channel) of the YCbCr color space to better assess the recovery of structural details.

Implementation Details. Our model was trained using the Adam optimizer with an initial learning rate of $4 \times 10^{-4}$, with the learning rate dynamically adjusted through cosine annealing warm restart scheduling. The training process consisted of a 15-epoch linear warm-up phase, during which the learning rate increased linearly from $1 \times 10^{-6}$ to $4 \times 10^{-3}$, followed by the cosine annealing phase. This phase commenced with an initial cycle length $T_0 = 50$, a cycle multiplication factor T_mult = 2, and a minimum learning rate of $1 \times 10^{-6}$. Furthermore, we trained using a batch size of 6 and ran the model for 300 epochs. The loss function was defined as 1-SSIM, and all experiments were conducted on an NVIDIA RTX 4090 GPU.

### *4.2. Ablation Study*

To systematically validate the effectiveness of the proposed components, we conducted a series of ablation experiments. Specifically, the study evaluated the component designs of the HAFB and WGCA modules, the configuration of the overall framework including color space selection and the integration of key modules, as well as the impact of the loss function and learning rate scheduling strategy. All models were trained and tested under identical experimental settings to ensure a fair comparison.

#### 4.2.1. Ablation Study of the HAFB Module

To systematically evaluate the effectiveness of the HAFB module and its components, we conducted an ablation study, as summarized in Table 1. First, to validate the design of the attention mechanism within HAFB, we compared the performance of different attention

schemes. A baseline model utilizing only RGB input served as the reference, achieving a PSNR of 37.12 dB and an SSIM of 0.969. The results showed that employing channel attention yielded a PSNR of 37.19 dB and an SSIM of 0.941. Using spatial attention achieved a PSNR of 37.21 dB and an SSIM of 0.947, while the standard self-attention mechanism attained a PSNR of 37.25 dB and an SSIM of 0.949. In contrast, the proposed PPSB achieved the best performance with a PSNR of 37.30 dB and an SSIM of 0.958, outperforming all other attention mechanisms in both metrics. This demonstrated that PPSB can more effectively model and separate rain streaks from background information for the deraining task.

**Table 1.** Quantitative Results of Ablation Study on HAFB Module.

| Module Composition | PSNR | SSIM |
|---|---|---|
| Baseline (Only RGB) | 37.12 | 0.969 |
| Channel Attention | 37.19 | 0.941 |
| Spatial Attention | 37.21 | 0.947 |
| Self Attention | 37.25 | 0.949 |
| PPSB | 37.30 | 0.958 |
| MAFFN | 37.28 | 0.953 |
| HAFB (PPSB + MAFFN) | 37.36 | 0.970 |

Second, to further analyze the synergistic interaction between the components within HAFB, we performed module ablation studies. Using only the PPSB module yielded a PSNR of 37.30 dB and an SSIM of 0.958, while employing only the MAFFN achieved a PSNR of 37.28 dB and an SSIM of 0.953. In contrast, the complete HAFB module, which integrates both PPSB and MAFFN, attained the optimal performance with a PSNR of 37.36 dB and an SSIM of 0.970. This demonstrated that the multi-scale feed-forward network can effectively capture rain streaks of various sizes and complements the attention mechanism. The result confirmed that the integrated design of PPSB and MAFFN was essential to HAFB's performance.

#### 4.2.2. Ablation Study of the WGCA Module

To systematically evaluate the effectiveness of the proposed WGCA module and its components, we conducted an ablation study, with the results summarized in Table 2. The model using both RGB and YCbCr color spaces served as the baseline, achieving a PSNR of 37.58 dB and an SSIM of 0.973. Building on this baseline, WEAB alone increased the PSNR to 37.75 dB. In contrast, a conventional cross-attention mechanism yielded a PSNR of 37.68 dB, while the proposed DCA achieved a higher performance with a PSNR of 37.84 dB and an SSIM of 0.972. This result verified that DCA could more effectively model cross-color-space consistency, thereby better distinguishing the structured content of real scenes from the irregular patterns of rain streaks.

**Table 2.** Quantitative Results of Ablation Study on WGCA Module.

| Module Composition | PSNR | SSIM |
|---|---|---|
| RGB + YCbCr | 37.58 | 0.973 |
| RGB + YCbCr + WEAB | 37.75 | 0.965 |
| RGB + YCbCr + Cross Attention | 37.68 | 0.961 |
| RGB + YCbCr + DCA | 37.84 | 0.972 |
| RGB + YCbCr + WGCA (WEAB + DCA) | 37.96 | 0.978 |

Finally, integrating WEAB with DCA to form the complete WGCA module attained the optimal performance, with a PSNR of 37.96 dB and an SSIM of 0.978. This demonstrates that combining wavelet-domain frequency guidance with cross-color-space cooperative

attention most effectively accomplishes background reconstruction and the suppression of rain streaks.

#### 4.2.3. Ablation Study of the Proposed Framework

To systematically evaluate the effectiveness of the proposed framework, an ablation study was conducted, with the results presented in Table 3. In terms of component effectiveness, the experiment started with a baseline model using only RGB input, which achieved a PSNR of 37.12 dB and an SSIM of 0.969. The introduction of a dual-branch architecture combining RGB and YCbCr led to improved performance, reaching a PSNR of 37.58 dB and an SSIM of 0.973, indicating that integrating complementary information from different color spaces is beneficial. Further incorporation of the WGCA module (RGB + YCbCr + WGCA) increased the PSNR to 37.96 dB and the SSIM to 0.978, as it leverages wavelet-based enhancement and attention-based fusion of dual-color space features. Finally, the complete model, which integrates the HAFB module, achieved the best performance with a PSNR of 38.64 dB and an SSIM of 0.989, confirming the synergistic effect between HAFB and WGCA and leading to further improvements in rain streak modeling and background reconstruction accuracy. The gradual performance gains systematically validate the effectiveness of the dual-branch structure and the collaborative relationships among the modules.

**Table 3.** Quantitative Results of Ablation Study on Proposed Framework.

| Module Composition | PSNR | SSIM |
|---|---|---|
| Baseline (Only RGB) | 37.12 | 0.969 |
| RGB + HSV | 37.33 | 0.962 |
| RGB + YCbCr | 37.58 | 0.973 |
| RGB + HSV + WGCA | 37.65 | 0.976 |
| RGB + YCbCr + WGCA | 37.96 | 0.978 |
| RGB + HSV + WGCA + HAFB | 38.17 | 0.980 |
| RGB + YCbCr + WGCA + HAFB | 38.64 | 0.989 |

Regarding color space comparison, a systematic evaluation was performed between HSV and YCbCr as the second branch. Under the basic dual-branch configuration, RGB + YCbCr (PSNR 37.58 dB, SSIM 0.973) outperformed RGB + HSV (PSNR 37.33 dB, SSIM 0.962). This advantage was maintained after adding the WGCA module: RGB + YCbCr + WGCA achieved a PSNR of 37.96 dB, higher than the 37.65 dB of RGB + HSV + WGCA. In the complete framework, the RGB + YCbCr combination also yielded better results, with a PSNR of 38.64 dB. These findings indicate that the YCbCr space offers clearer advantages in separating luminance and chrominance information, a property which leads to features with stronger complementarity to RGB features. Consequently, YCbCr integrates more effectively with modules such as WGCA and HAFB, contributing to enhanced overall deraining performance.

#### 4.2.4. Ablation Study of the Training Settings

To systematically evaluate the impact of different training configurations, we conducted an ablation study on the loss function and the learning rate schedule. The results are summarized in Table 4. When using the Cosine baseline scheduler, the 1-SSIM loss achieved a PSNR of 38.41 dB and an SSIM of 0.986, surpassing both the L1 loss (37.62 dB, 0.968) and the L1 + SSIM combination (37.91 dB, 0.985). This indicates that the 1-SSIM loss offers a clearer and more effective optimization target by prioritizing structural fidelity alone.

**Table 4.** Quantitative Results of Ablation Study on Training Settings.

| Loss Function | Learning Rate Schedule | PSNR | SSIM |
|---|---|---|---|
| L1 | Cosine (Baseline) | 37.62 | 0.968 |
| L1 + SSIM | Cosine (Baseline) | 37.91 | 0.985 |
| 1-SSIM (Ours) | Fixed LR | 37.85 | 0.979 |
| 1-SSIM (Ours) | Step Decay | 38.12 | 0.982 |
| 1-SSIM (Ours) | Cosine (Baseline) | 38.41 | 0.986 |
| 1-SSIM (Ours) | Warm-up + Restart (Ours) | 38.64 | 0.989 |

Furthermore, with the loss fixed to 1-SSIM, we compared learning rate strategies. The fixed learning rate (37.85 dB, 0.979) and step decay (38.12 dB, 0.982) delivered suboptimal results. In contrast, our proposed Warm-up + Restart scheduler elevated the performance to 38.64 dB and 0.989 SSIM, achieving the highest scores across all configurations. These findings demonstrate that conventional schedulers are limited in maintaining stable convergence. Conversely, our Warm-up + Restart mechanism effectively complements the 1-SSIM loss by stabilizing the early training phase and periodically escaping local optima. The consistent gains across all metrics confirm the effectiveness of this combined strategy.

### *4.3. Comparative Experiments*

#### 4.3.1. Experiments on Synthetic Datasets

To comprehensively evaluate the performance of the proposed method, we conducted comparative experiments on four benchmark datasets: Rain100L [16] Rain100H [16], Rain800 [35], and Rain12 [14]. The methods performed for comparison included JORDER [16], PReNet [17], MPRNet [23], MFDNet [36], DRAN [37], M2PN [38], CLIP-RPN [39], GKC-Net [40], and MDRDANet [41]. These competing methods encompassed both traditional CNN-based models and recently proposed Transformer architectures, providing a broad and representative comparison. All experiments adhered to a unified evaluation protocol. The restoration performance was quantitatively compared using both PSNR and SSIM, with the results summarized in Table 5.

**Table 5.** Quantitative Comparison on Synthetic Datasets.

| Method | Rain100L [16] | | Rain100H [16] | | Rain800 [35] | | Rain12 [14] | | Params (M) | FLOPs (G) | FPS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | | | |
| JORDER [16] | 36.11 | 0.970 | 26.54 | 0.835 | 26.73 | 0.868 | 36.02 | 0.960 | 4.169 | 27.2 | 0.68 |
| PReNet [17] | 36.99 | 0.977 | 29.46 | 0.899 | 27.17 | 0.874 | 36.16 | 0.961 | 0.168 | 66.44 | 1.81 |
| MPRNet [23] | 36.40 | 0.965 | 30.41 | 0.890 | 27.43 | 0.882 | 36.63 | 0.968 | 3.637 | 141.5 | 5.56 |
| MFDNet [36] | 37.61 | 0.973 | 30.48 | 0.898 | 28.78 | 0.914 | 36.89 | 0.947 | 4.74 | 156.26 | 8.09 |
| DRAN [37] | 37.83 | 0.981 | 30.65 | 0.901 | 27.31 | 0.856 | 36.02 | 0.958 | — | — | 0.28 |
| M2PN [38] | 38.36 | 0.985 | 29.36 | 0.899 | — | — | — | — | 0.168 | — | 9.09 |
| CLIP-RPN [39] | 36.21 | 0.921 | 29.66 | 0.842 | 28.40 | 0.812 | — | — | 32.72 | — | 3.09 |
| GKC-Net [40] | 37.97 | 0.980 | 29.10 | 0.881 | — | — | — | — | 109.91 | 70.84 | — |
| MDRDANet [41] | 38.01 | 0.986 | 29.90 | 0.904 | 27.48 | 0.899 | 37.01 | 0.975 | — | — | 12.50 |
| Ours | 38.64 ± 0.34 | 0.989 ± 0.002 | 30.51 ± 0.23 | 0.907 ± 0.001 | 28.56 ± 0.27 | 0.911 ± 0.002 | 37.28 ± 0.41 | 0.978 ± 0.003 | 20.40 | 56.49 | 10.20 |

As shown in Table 5, the proposed method achieved competitive performance across multiple datasets. Specifically, on the Rain100L dataset, it attained a PSNR of 38.64 dB, surpassing the second-best method by 0.63 dB, with a corresponding improvement in SSIM of 0.003. On the Rain12 dataset, our method achieved a PSNR of 37.28 dB, which was 0.27 dB higher than the suboptimal method, and obtained a SSIM gain of 0.003. Moreover, the proposed method also demonstrated robust generalization in more challenging dense-rain

scenarios, as evidenced by its stable results on the Rain100H and Rain800 datasets. Overall, the proposed method effectively enhance the model's performance in detail recovery and color fidelity under complex rainy conditions.

In terms of computational efficiency, the proposed method contained 20.40 M parameters and required 56.49 GFLOPs of computational complexity, with an inference speed of 10.20 FPS. Although its parameter count and computational cost were not the lowest among compared methods, it achieved optimal or leading performance across multiple image quality metrics. Notably, the proposed method achieves high PSNR and SSIM scores while keeping model complexity and computational overhead within reasonable bounds. Furthermore, it preserves high processing efficiency. Therefore, the proposed method effectively enhanced detail recovery and color fidelity under complex rainy conditions, while maintaining favorable practicality and efficiency for real-world deployment.

Visual comparison results of different methods on synthesized rainy images are shown in Figure 6. As shown in the figure, our method effectively removed rain streaks across various scenes while preserving background details and edge structures. Regions marked with red and blue boxes highlight that comparative methods exhibited residual rain streaks, along with blurred texture details and insufficient color restoration. In comparison, our method demonstrated relative advantages in color naturalness, texture preservation, and structural integrity, leading to more balanced overall performance in rain removal. Specifically, both JORDER and PReNet exhibited discernible rain streak residuals in their restored images, while our method achieved cleaner results with fewer residual streaks. Regarding detail reconstruction, results from MPRNet appeared excessively smoothed and lost original texture information. This was particularly evident in the second grassland image, where our method recovered finer grass textures more distinctly, highlighting its advantage in preserving and reconstructing image texture details.

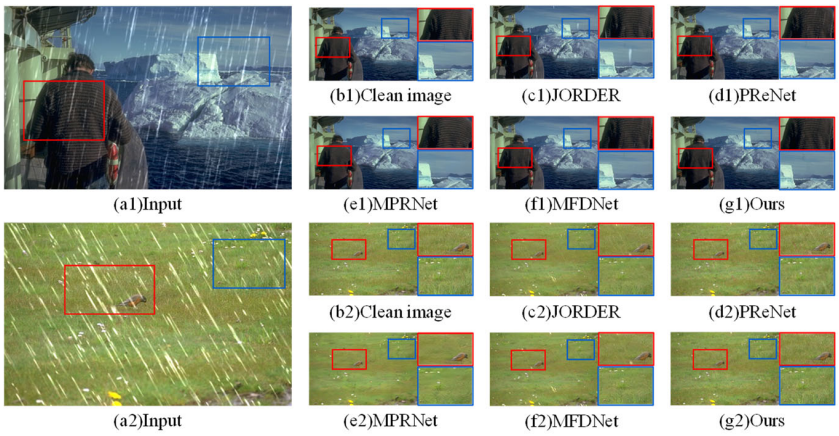


**Figure 6.** Visualization Results of Different Algorithms on Synthetic Datasets.

#### 4.3.2. Experiments on Real-World Datasets

To verify the generalization capability and robustness of the proposed method in real-world scenarios, comparative experiments were conducted on the real rainy image dataset SPA-Data [42]. This dataset contained a large number of real rainy images captured in complex natural scenes. Specifically, the rain streaks exhibited diverse patterns and uneven distributions, often accompanied by compound degradations such as haze and motion blur, which posed greater challenges to a model's adaptability and restoration capability. Four representative deraining methods were selected for comparison, including JORDER [16], PReNet [17], MPRNet [23], and MFDNet [36]. All models were tested using their officially released pre-trained weights to ensure a fair comparison. The quantitative results are presented in Table 6.

**Table 6.** Quantitative Comparison on Real-World Datasets.

| Metric | JORDER [16] | PReNet [17] | MPRNet [23] | MFDNet [36] | Ours |
|---|---|---|---|---|---|
| PSNR | 38.16 | 39.21 | 40.36 | 40.64 | $41.18 \pm 0.35$ |
| SSIM | 0.962 | 0.974 | 0.977 | 0.981 | $0.986 \pm 0.004$ |

As shown in Table 6, the proposed method achieved favorable results on both PSNR and SSIM metrics in real-world rainy scenarios. It attained a PSNR of 41.18 dB, which was 0.54 dB higher than the second-best method, and an SSIM of 0.986, representing an improvement of 0.005. In particular, compared to models trained and tested on synthetic datasets, rain streaks in real scenes are often more intricately blended with image content and are significantly influenced by environmental factors such as illumination and depth of field. These results thus demonstrate the reliability of the proposed method in practical settings.

The visual comparison results of different methods on real rain images are shown in Figure 7. Our method demonstrated stable robustness under rain streaks of varying densities and directions. Regions marked with red and blue boxes indicate that comparative methods still exhibited residual rain streaks and insufficient restoration of texture details in corresponding scenes. In comparison, our approach achieved better visual consistency and content fidelity than most existing methods. This performance can be attributed to the collaborative dual-color-space design, which helped maintain overall image harmony and local realism even in scenes with complex rain streaks and cluttered backgrounds. As shown in the first image, our method achieved more complete rain streak removal compared to the visible residuals left by JORDER, PReNet, and MPRNet. Moreover, although MFDNet also effectively reduced rain streaks, our approach better restored the original texture within the region marked by the blue bounding box. These observations collectively demonstrated the overall superiority of our method in terms of both rain streak suppression and detail recovery.

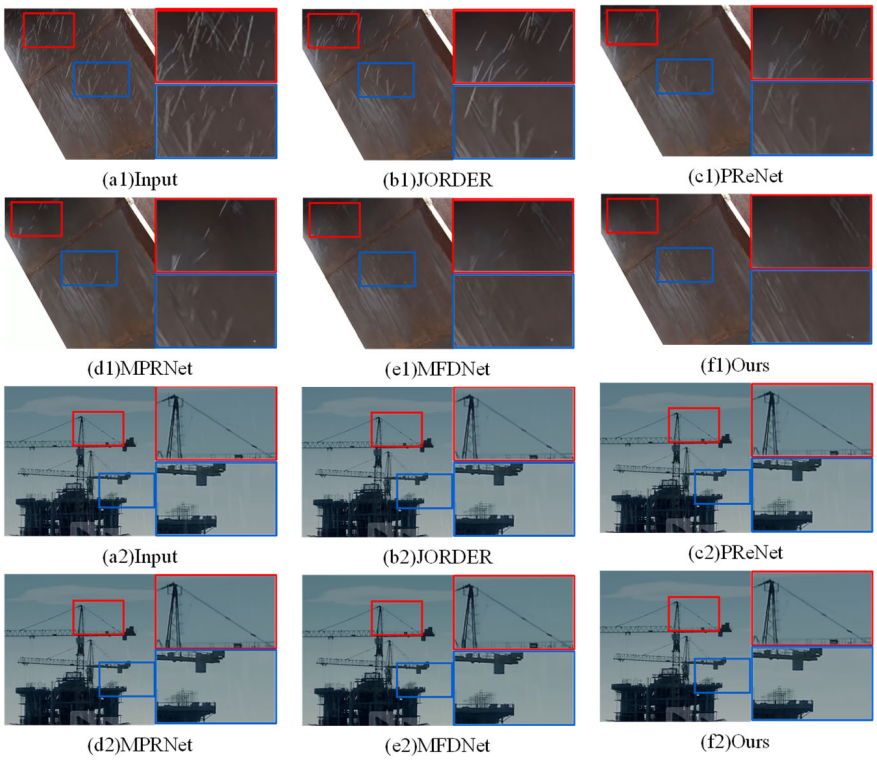


**Figure 7.** Visualization Results of Different Algorithms on Real-World Datasets.

### *4.4. Downstream Task Application*

To evaluate the practical value of the proposed algorithm, we introduced the YOLOv11 object detection algorithm into an autonomous driving scenario for validation. First, the YOLOv11n model was trained using the UA-DETRAC dataset [43] to obtain its optimal weights. Subsequently, rain streaks were synthetically added to the test images of this dataset to simulate real-world rainy conditions. Following this, vehicle detection was performed on both rainy and derained images using the trained model, with the results shown in Figure 8.

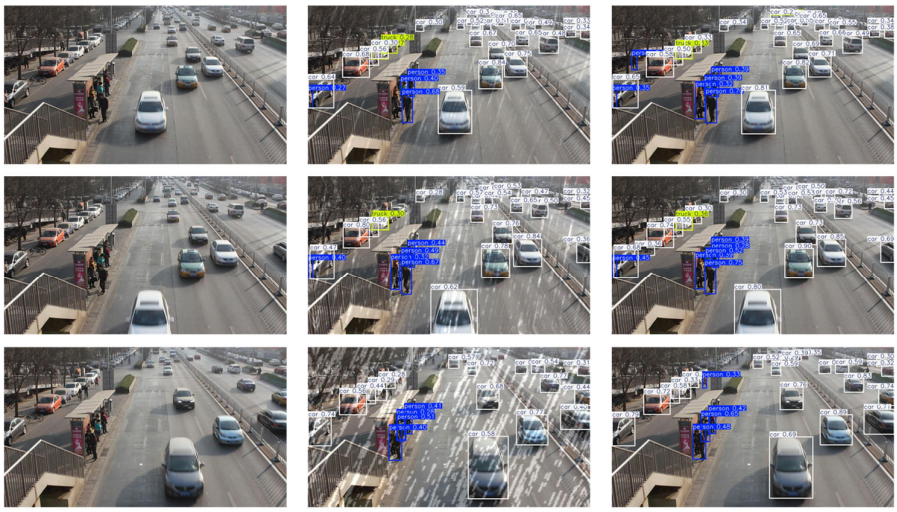


**Figure 8.** Object detection results in rainy weather.

It can be observed from Figure 8 that the rain removal process substantially improved object detection performance. As shown in the first row of Figure 8, the number of detected pedestrians at the bus stop increased from three in the rainy image to four in the derained image. The results presented in the second and third rows further demonstrated a significant enhancement in vehicle detection accuracy after rain removal. These improvements confirmed that the proposed method effectively removed rain streaks while preserving image details, thereby increasing the detection confidence for both nearby and distant targets, including pedestrians and vehicles. In conclusion, the proposed method demonstrates promising potential for practical applications such as autonomous driving and intelligent transportation systems.

## 5. Conclusions

We propose a novel image deraining network that integrates dual-color space processing with frequency-domain prior information to address the image deraining problem. By constructing a dual-branch Transformer architecture for RGB and YCbCr, we take full advantage of the complementary strengths of both color spaces in terms of color richness and structural sensitivity. Based on this approach, a Hybrid Attention Feedforward Block (HAFB) is designed. By integrating channel and spatial attention mechanisms with multi-scale convolutional feature extraction, it effectively enhances the perception and suppression capabilities for multi-scale rain streaks. We further design the Wavelet-Gated Cross-Attention (WGCA) module, which employs wavelet transforms to achieve frequency-domain feature decomposition and enhancement. Combined with bidirectional cross-attention, it establishes cross-color-space feature alignment and semantic guidance. Quantitative and qualitative evaluations across multiple benchmark datasets demonstrate that the proposed method achieves improved performance in both PSNR and SSIM metrics,

while visually exhibiting enhanced detail preservation and color consistency. In future work, we plan to integrate large-scale model technology to further enhance the structural detail fidelity and color authenticity of restored images.

**Author Contributions:** Conceptualization, L.Y. and Y.H.; methodology, H.Z. and L.Y.; software, H.Z.; validation, L.Y. and Y.H.; formal analysis, H.Z. and Y.H.; writing—original draft preparation H.Z. and Y.H.; funding acquisition, H.Z. and L.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Key Research and Development Program of Shanxi Province (No. 202102010101008), the Higher Education Institution Science and Technology Innovation Project of Shanxi Province (No. 2024L295), the Key Project of Science and Technology Strategy Research Special Project of Shanxi Province (No. 202304031401011), the Basic Research Program Project of Shanxi Province (Free Exploration Category) (No. 202403021222276, No. 202503021212249) and Shanxi Province Postgraduate Education Reform Research Project (No. 2023JG163).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Please contact the corresponding author if needed.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Khan, S.; Naseer, M.; Hayat, M.; Zamir, S.W.; Khan, F.S.; Shah, M. Transformers in vision: A survey. *ACM Comput. Surv. (CSUR)* **2022**, *54*, 1–41. [CrossRef]
2. Brophy, T.; Mullins, D.; Cormican, R.; Ward, E.; Glavin, M.; Jones, E.; Deegan, B. The impact of rain on image quality from sensors on connected and autonomous vehicles. *IEEE Open J. Veh. Technol.* **2025**, *6*, 632–646. [CrossRef]
3. Samani, M.A.; Farrokhi, M. Adverse to Normal Image Reconstruction Using Inverse of StarGAN for Autonomous Vehicle Control. *IEEE Access* **2025**, *13*, 77305–77316. [CrossRef]
4. Zhu, B.; Huang, Y.; Zhao, J.; Zhang, P.; Han, J.; Song, D. Synthetic Image Generation Model for Intelligent Vehicle Camera Function Testing in Rain and Fog. *IEEE Trans. Intell. Transp. Syst.* **2025**, *26*, 5332–5347. [CrossRef]
5. Shi, C.; Fang, L.; Wu, H.; Xian, X.; Shi, Y.; Lin, L. Nitedr: Nighttime image de-raining with cross-view sensor cooperative learning for dynamic driving scenes. *IEEE Trans. Multimed.* **2024**, *26*, 9203–9215. [CrossRef]
6. Xue, X.; He, J.; Ma, L.; Meng, X.; Li, W.; Liu, R. ASF-Net: Robust video deraining via temporal alignment and online adaptive learning. *Pattern Recognit.* **2025**, *158*, 110973. [CrossRef]
7. Sun, S.; Ren, W.; Zhou, J.; Wang, S.; Gan, J.; Cao, X. Semi-Supervised State-Space Model with Dynamic Stacking Filter for Real-World Video Deraining. In Proceedings of the Computer Vision and Pattern Recognition Conference, Nashville, TN, USA, 11–15 June 2025; pp. 26114–26124.
8. Zhang, Y.; Wang, J.; Weng, W.; Sun, X.; Xiong, Z. Egvd: Event-guided video deraining. *IEEE Trans. Neural Netw. Learn. Syst.* **2025**, *36*, 14092–14106. [CrossRef]
9. Li, X.; Fan, W.; Shen, Y.; Wang, C.; Wang, W.; Yang, X.; Zhang, Q.; Zhou, D. Iterative Optimal Attention and Local Model for Single Image Rain Streak Removal. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 5024515. [CrossRef]
10. Zhang, Y.; Xuan, S.; Li, Z. Robust object detection in adverse weather with feature decorrelation via independence learning. *Pattern Recognit.* **2026**, *169*, 111790. [CrossRef]
11. Le, T.-H.; Huang, S.-C.; Hoang, Q.-V.; Lokaj, Z.; Lu, Z. Amalgamating Knowledge for Object Detection in Rainy Weather Conditions. *ACM Trans. Intell. Syst. Technol.* **2025**, *16*, 1–20. [CrossRef]
12. Li, P.; Ni, J.; Tao, D. End-to-End Cascaded Image Restoration and Object Detection for Rain and Fog Conditions. *IET Comput. Vis.* **2025**, *19*, e70021. [CrossRef]
13. 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.
14. 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.
15. Fu, X.; Huang, J.; Ding, X.; Liao, Y.; Paisley, J. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Trans. Image Process.* **2017**, *26*, 2944–2956. [CrossRef]

16. 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. 1685–1694.
17. 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. 3932–3941.
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/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5718–5729.
19. Xiao, J.; Fu, X.; Liu, A.; Wu, F.; Zha, Z.-J. Image de-raining transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 12978–12995. [CrossRef]
20. 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.
21. 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. 262–277.
22. 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. 8343–8352.
23. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.-H.; Shao, L. Multi-stage progressive image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 14816–14826.
24. Quan, R.; Yu, X.; Liang, Y.; Yang, Y. Removing raindrops and rain streaks in one go. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 9143–9152.
25. Chen, H.; Wang, Y.; Guo, T.; Xu, C.; Deng, Y.; Liu, Z.; Ma, S.; Xu, C.; Xu, C.; Gao, W. Pre-trained image processing transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 12299–12310.
26. 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/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 17662–17672.
27. Valanarasu, J.M.J.; Yasarla, R.; Patel, V.M. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 2343–2353.
28. Wang, C.; Yan, T.; Huang, W.; Chen, X.; Xu, K.; Chang, X. APANet: Asymmetrical Parallax Attention Network for Efficient Stereo Image Deraining. *IEEE Trans. Comput. Imaging* **2025**, *11*, 101–115. [CrossRef]
29. Yan, Q.; Feng, Y.; Zhang, C.; Pang, G.; Shi, K.; Wu, P.; Dong, W.; Sun, J.; Zhang, Y. Hvi: A new color space for low-light image enhancement. In Proceedings of the Computer Vision and Pattern Recognition Conference, Nashville, TN, USA, 11–15 June 2025; pp. 5678–5687.
30. Zhang, J.; Su, H.; Zhang, T.; Tian, H.; Fan, B. Multi-Scale Fusion Underwater Image Enhancement Based on HSV Color Space Equalization. *Sensors* **2025**, *25*, 2850. [CrossRef]
31. Huang, H.; He, R.; Sun, Z.; Tan, T. Wavelet-srnet: A wavelet-based cnn for multi-scale face super resolution. In Proceedings of the IEEE International Conference on Computer Vision, Honolulu, HI, USA, 21–26 July 2017; pp. 1698–1706.
32. Liu, P.; Zhang, H.; Zhang, K.; Lin, L.; Zuo, W. Multi-level wavelet-CNN for image restoration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 886–895.
33. Xu, K.; Yang, X.; Yin, B.; Lau, R.W. Learning to restore low-light images via decomposition-and-enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 2278–2287.
34. Jiang, H.; Luo, A.; Fan, H.; Han, S.; Liu, S. Low-light image enhancement with wavelet-based diffusion models. *ACM Trans. Graph. (TOG)* **2023**, *42*, 1–14. [CrossRef]
35. Zhang, H.; Sindagi, V.; Patel, V.M. Image de-raining using a conditional generative adversarial network. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 3943–3956. [CrossRef]
36. 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.* **2024**, *33*, 191–204. [CrossRef]
37. Chang, Y.; Chen, M.; Yu, C.; Li, Y.; Chen, L.; Yan, L. Direction and residual awareness curriculum learning network for rain streaks removal. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *35*, 8414–8428. [CrossRef]
38. Zhong, J.; Li, Z.; Xiang, D.; Han, M.; Li, C.; Gan, Y. A Lightweight Multi-domain Multi-attention Progressive Network for Single Image Deraining. In Proceedings of the 32nd ACM International Conference on Multimedia, Melbourne, VIC, Australia, 28 October–1 November 2024; pp. 2001–2010.

39. Guan, C.; Yoshie, O. CLIP-driven rain perception: Adaptive deraining with pattern-aware network routing and mask-guided cross-attention. *Pattern Recognit.* **2026**, *173*, 112886. [CrossRef]
40. Gong, M.; Yu, J.; Wang, Z.; Chi, S. GKC-Net: Gated KAN with Channel-Position attention mechanism for image deraining. *Pattern Recognit.* **2026**, *170*, 112080. [CrossRef]
41. Li, Y.; Kong, W. Single-image deraining method based on multi-stage dense residuals network. *Comput. Eng. Des.* **2025**, *46*, 2134–2141.
42. Wang, T.; Yang, X.; Xu, K.; Chen, S.; Zhang, Q.; Lau, R.W.H. 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 (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 12262–12271.
43. Wen, L.; Du, D.; Cai, Z.; Lei, Z.; Chang, M.-C.; Qi, H.; Lim, J.; Yang, M.-H.; Lyu, S. UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking. *Comput. Vis. Image Underst.* **2020**, *193*, 102907. [CrossRef]