Dual-stage Spatial-Frequency Domain Deraining Network based on Fast Fourier Transform

Yingzhi Wei

Abstract— **Recently, many deep learning-based deraining networks have been proposed, which directly extract rain streaks from the rainy images and subtract them to obtain deraining results. These methods often suffer from insufficient or excessive extraction of rain streaks, resulting in residual rain streaks or the loss of texture information in the deraining images. A Dual-stage Spatial-Frequency Domain Deraining Network Based on Fast Fourier Transform (FFT-DDN) is proposed, embedding the Fourier transform into the neural network. We divides the deraining task into two networks, namely the Image Deraining Network (IDN) and the Background Extraction Network (BEN). A Spatial-frequency domain Fourier Phase Enhancement Block (SFPEB) is designed as the fundamental block in both deraining networks, achieving parallel processing and fusion of the Fourier and spatial domains. Between the two networks, a Detail Attention Block (DAB) is designed to mine the intrinsic connection between background information and rain streak features, to restore richer texture information. Moreover, to fully utilize the complementary information between the spatial and frequency domains, a Feature Fusion Block (FFB) is designed to further enhance the overall performance of the network. Experimental results on synthetic and real datasets demonstrate that the proposed method achieves superior deraining effects both subjectively and objectively**

Index Terms— **Image deraining, Fourier transform, Attention mechanism.**

I. Introduction

Images captured during rainy conditions often suffer from significantly reduced visibility, posing substantial challenges to various outdoor visual tasks such as image segmentation[1], object detection[2], and video surveillance[3]. The objective of image deraining technology is to eliminate rain streaks from these images, thereby enhancing their quality and suitability as a preprocessing step for advanced computer vision applications. Despite its critical importance, the task of image deraining is an ill-posed problem that remains inadequately resolved, attracting considerable ongoing research attention. This persisting interest underscores the complexity and the essential nature of improving image clarity under adverse weather conditions, highlighting its pivotal role in the seamless execution of outdoor visual tasks[4].

To address these issues, we proposes a Dual-stage Spatial-Frequency Domain Deraining Network Based on Fast Fourier Transform (FFT-DDN). Unlike previous methods that only process features in the spatial domain, this network uses a multi-branch structure to decompose the image into the image

spatial domain and the image Fourier frequency domain, fully utilizing the frequency differences between the original and degraded images for deraining. Features are extracted by processing the phase spectrum in the Fourier domain because the phase spectrum represents the phase information of various frequency components in the image, which to some extent reflects the image's local structure, texture, edges, and other features, determining the image's relative position and shape in the frequency domain. And rain streaks are distributed throughout the image in different shapes and directions. Therefore, we propose the Spatial-Frequency Domain Fourier Phase Convolutional Block (SFPconv), which processes features in different domains through different branches. In the spatial branch, a Multi-Scale Convolutional Block (MSconv) is designed to extract spatial features of the image at different scales, and three learnable parameters are used to adjust the corresponding weights for specific scales. In the Fourier branch, features are extracted and integrated after Fourier decomposition and phase spectrum processing, and then, after inverse Fourier transform, they are input into the feature fusion attention module to dynamically select the information-rich frequency domain. Two Spatial-Frequency Domain Fourier Phase Convolution Blocks stacked with residual structures construct the Spatial-Frequency Domain Fourier Phase Enhancement Block (SFPEB) as the basic module of the network.

Furthermore, we explore the intrinsic correlation between the background layer and the rain streak layer, dividing the network into two parts: the Image Deraining Network (IDN) and the Background Extraction Network (BEN). Between the two networks, a Detail Information Attention Block (DAB) is proposed, which can generate degradation priors and generate degradation masks according to the rain streak distribution map predicted by the image deraining network, and then, through global correlation calculation, extract information-rich and complementary components from the rain streak image with degradation masks, thus mining the intrinsic connection between the background and rain streak layer to help achieve more accurate texture restoration. To further enhance deraining performance, we adopt a dense connected method to fully utilize shallow features from front to back to restore more background information. The main contributions of this method are summarized as follows:

1) A dual-stage image deraining architecture is proposed, including image deraining and background refinement stages. Guided by the coarsely separated rain streak layer

Manuscript received March 28, 2024

Yingzhi Wei, School of Computer Science and Technology, Tiangong University, Tianjin, China

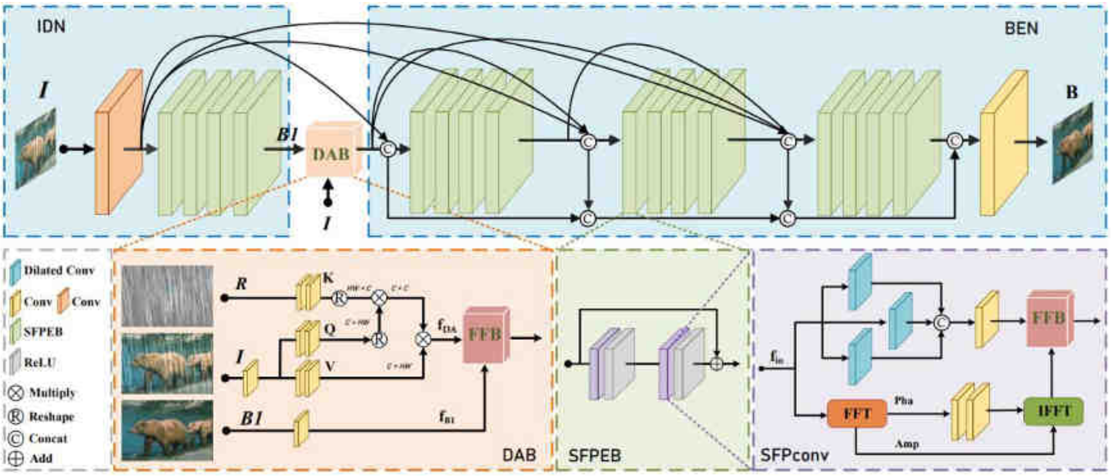

Figure 1: The architecture of the proposed FFT-DDN.

learned, this architecture can recover rain-free images with richer information.

2) A spatial-frequency domain deraining network is proposed, which automatically decomposes feature maps into the spatial and frequency domains through a multi-scale feature extractor and Fourier phase convolution, and dynamically selects the domain containing effective information for image deraining, preserving more texture information.
3) A detail information attention module is proposed to explore and uncover the potential correlation between the rain streak layer and the background layer, significantly reducing the learning burden and promoting the restoration of detail textures.
4) Extensive experiments prove that the proposed FFT-DDN achieves more advanced performance levels compared to some existing deraining methods on both synthetic and real datasets.

II. Related Work

A. *Optimization-based Deraining Methods*

In the realm of traditional deraining approaches, the process is often conceptualized as a signal decomposition task, transforming it into an optimization challenge with the goal of deriving rain-free images. Notably, Li et al.[5] introduced a methodology predicated on a Gaussian mixture model trained on image patches, aimed at enhancing the quality of the segregated images. Similarly, Luo et al.[6] advanced a method employing discriminative sparse coding to distinguish between rain layers and background layers within rainy images. Employing dictionary learning, Kang et al.[7] succeeded in obtaining rain-free images by extracting rain components from the images' high-frequency parts. Furthermore, Gu et al.[8] proposed a joint convolutional analysis and synthesis (JCAS) sparse representation framework, merging analysis sparse representation (ASR) and synthetic sparse representation (SSR), to dissect rainy images. These methods, grounded on the linear model, strive to efficaciously remove rain streaks. However, the actual degradation process experienced by images under rainfall conditions cannot be accurately described by a simple additive relationship between rain layers and background layers, thus limiting the effectiveness of these strategies in eradicating rain content and achieving clear, rain-free images.

B. *Deep Learning-Based Methods*

The emergence of deep learning technologies has given rise to a new generation of deraining algorithms, marking notable progress in the field. Li et al.[9] proposed a recurrent squeeze-and-excitation context aggregation network that leverages recurrent neural networks to remove rain streaks. Yang et al.[10] have developed a multi-task network designed to simultaneously detect and eliminate rain streaks. Through the deployment of prior knowledge, Fu et al.[11] have crafted a deep detail network capable of extracting more accurate rain streak information while minimizing background disturbances. A density-aware multi-stream CNN introduced by Zhang et al.[12] utilizes autonomous detection of rain density to efficiently remove rain streaks based on the detected rain density. Recognizing the importance of interaction between different processing stages, Ren et al.[13] constructed a simple yet effective progressive recurrent network for the phased removal of rain streaks. Further, Zamir et al.[14] have developed a multi-stage framework that progressively learns the mapping function from degraded inputs to reconstruct pristine rain-free images. Wang et al. [15]have innovated a rain convolutional dictionary network, employing the proximal gradient descent technique to formulate an optimization strategy for the model.

Despite the advancements these deep learning methods offer over optimization-based approaches, many still rely on the principle of linear image decomposition and the learning of residual mappings for the separation of rain streaks from the images. This approach can lead to the retention of residual rain streak information or its excessive removal, resulting in blurred edges and artifacts in the final deraining images.

III. Proposed Method

This chapter introduces a Dual-stage Spatial-Frequency Domain Deraining Network based on Fast Fourier Transform, dividing the network into two main parts: the Image Deraining

Network (IDN) and the Background Extraction Network (BEN), as illustrated in Figure 1.

Within the Image Deraining Network, we stack four Spatial-Frequency Domain Fourier Phase Enhancement Blocks (SFPEB) as a group, supplemented by a convolution block for extracting shallow features, constituting the core part of the network. The role of the Image Deraining Network is to produce a coarse deraining result, guiding the subsequent background refinement process. The Background Extraction Network utilizes three sets of Spatial-Frequency Domain Fourier Phase Enhancement Blocks for step-by-step background refinement, ultimately yielding a high-quality, rain-free background. Each Spatial-Frequency Domain Fourier Phase Enhancement Block comprises two Spatial-Frequency Domain Fourier Phase Convolution Blocks layered with a residual structure. Moreover, to achieve enhanced deraining effects, the network employs multi-level skip connections to integrate and utilize features from all levels optimally. The specific operational flow is as follows:

For a degraded image I of dimensions H × W × 3 , where 3 represents the channel number and H × W denotes the spatial coordinates, the degraded image serves as the input to the deraining network. It first passes through a 3 × 3 convolution layer to extract shallow features of the rainy image, obtaining a feature map of dimensions H × W × C. These features then go through a set of Spatial-Frequency Domain Fourier Phase Enhancement Blocks to generate a coarse derained result B1. Subsequently, the degraded image I minus the result B1 produces a rain streak map R, and then I, R, B1 serve as inputs to the Detail Attention Block, which establishes an intrinsic connection between rain streaks and the background. After passing through an image fusion module to enhance the features of the derained result B1, the enhanced features are fed into the Background Extraction Network. This network, through three sets of Spatial-Frequency Domain Fourier Phase Enhancement Blocks and a 3 × 3 convolution layer, generates the final restored image B. Throughout the deraining process, the output of each set of Spatial-Frequency Domain Fourier Phase Enhancement Blocks serves as the input for the next set, while the output features of each group are concatenated to produce a clearer background image. The entire network process can be described as follows:

$$B1 = IDN(I)\#(1)$$
$$R = I - B1\#(2)$$
$$f = DAB(I, R, B1)\#(3)$$
$$B = BEN(f)\#(4)$$

IDN represents the Image Deraining Network, BEN represents the Background Refinement Network, I denotes the input rainy image, B is the rain-free image restored by the Background Refinement Network, B1 is the coarse rain-free image recovered by the Image Deraining Network, and f represents the image features reweighted and distributed after processing by the Detail Attention Block.

A. *Spatial-Frequency Domain Fourier Phase Enhancement Block (SFPEB)*

The Spatial-Frequency Domain Fourier Phase Enhancement Block (SFPEB) comprises a spatial branch and a Fourier branch. After an image undergoes Fourier transform, it yields amplitude and phase spectra. The amplitude spectrum represents the brightness information of each pixel in the image, while the phase spectrum represents the spatial structure information of the image. To obtain derained images with more complete textures and richer details, this chapter processes the phase spectrum after Fourier transformation, making full use of the phase spectrum information of the features.

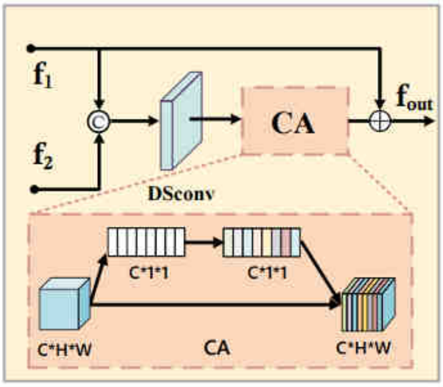

Figure 2: The architecture of FFB.

To integrate Fourier transform into the convolutional network, this chapter designs a Spatial-Frequency Domain Fourier Phase Enhancement Block. Since convolution operations are conducted in the spatial domain, directly operating on the amplitude or phase spectrum would alter the spatial structure of the image information. To prevent convolution operations from distorting the information characteristics of the amplitude and phase spectra, a multi-branch structure is introduced to refine the features after Fourier transformation. A Spatial-Frequency Domain Fourier Convolution Block is also designed to extract features from the phase spectrum, as shown in Figure 1.

In the spatial branch, this chapter uses three different scales of dilated convolution to extract spatial features, which can be represented as:

$$MSconv = Conv_{3x3}\left(Cat\begin{bmatrix} Conv_{3x3d1}(f_{in}), \\ Conv_{3x3d2}(f_{in}), \\ Conv_{3x3d3}(f_{in}) \end{bmatrix}\right)\#(5)$$

where $Cat(\cdot)$ denotes the channel-wise concatenation operation. $Conv_{kxdy}$ represents a dilated convolution with kernel size x and dilation rate y. In the Fourier branch, spatial features $f_{in} \in R^{C\times H\times W}$ first pass through Fast Fourier Transform to obtain the corresponding amplitude spectrum A and phase spectrum P, then the amplitude spectrum is input into two 1 × 1 convolutions to obtain the refined phase spectrum P, followed by calculating the corresponding features through Inverse Fast Fourier Transform. The Spatial-Frequency Domain Fourier Phase Convolution Block can be represented as:

$$A(f_{in}),P(f_{in}) = FFT(f_{in})\#(6)$$

$$SFPconv = FFB(IFFT\left(Conv_{1\times1}\left(Conv_{1\times1}(P(f_{in}))\right),P(f_{in})\right),MSconv(f_{in}))\#(7)$$

where A represents the amplitude spectrum, P represents the phase spectrum, $SFPconv$ represents the Spatial-Frequency Domain Fourier Phase Convolution Block, FFT represents Fast Fourier Transform, $IFFT$ represents Inverse Fast Fourier Transform, and FFB represents the Feature Fusion Module.

The Spatial-Frequency Domain Fourier Phase Enhancement Block connects two blocks in series with a residual connection, which can be represented as:

$$SFPEB = +ReLU\left(SFPconv\left(ReLU(SFPconv(f_{in}))\right)\right) \quad (8)$$

Table 1: Quantitative comparison on synthetic datasets (Red: rank 1st; Blue: rank 2nd)

Datasets	Rain100H		Rain200H		Rain200L		Rain800		Average	
Metrics	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
RESCAN	28.82	0.867	27.95	0.862	39.43	0.982	28.36	0.872	31.14	0.896
SIRR	22.03	0.714	22.17	0.726	32.21	0.931	22.73	0.762	24.79	0.783
PReNet	30.31	0.910	29.47	0.907	37.93	0.983	26.82	0.888	31.13	0.922
JORDER-E	30.22	0.898	29.23	0.894	39.13	0.985	27.92	0.883	31.63	0.915
RCDNet	31.26	0.912	30.18	0.909	39.49	0.986	28.66	0.893	32.40	0.925
BRN	31.32	0.924	30.27	0.919	38.86	0.985	28.31	0.896	32.19	0.931
MPRnet	31.71	0.917	30.62	0.914	40.68	0.988	28.78	0.887	32.95	0.927
ECNet	31.43	0.921	30.22	0.912	39.72	0.987	29.26	0.905	32.66	0.931
FFT-DDN	32.42	0.933	31.24	0.928	40.53	0.989	30.12	0.907	33.58	0.939

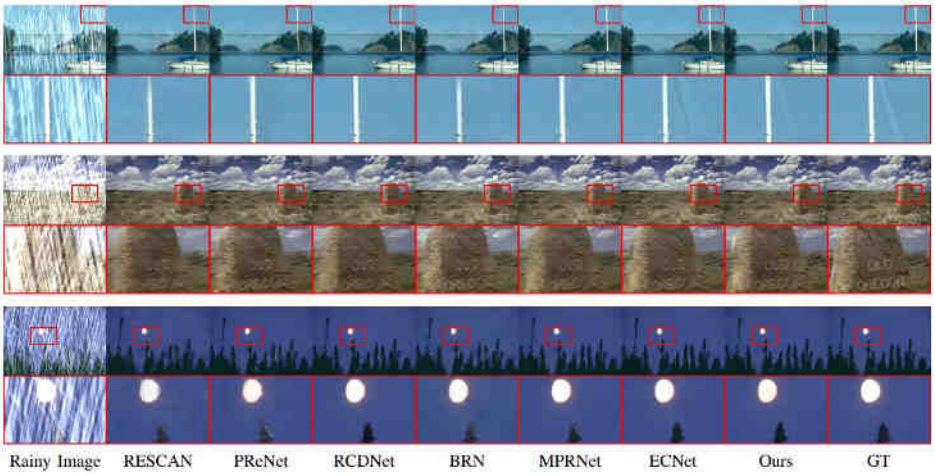

Figure 3: Visual comparisons on Rain100H dataset.

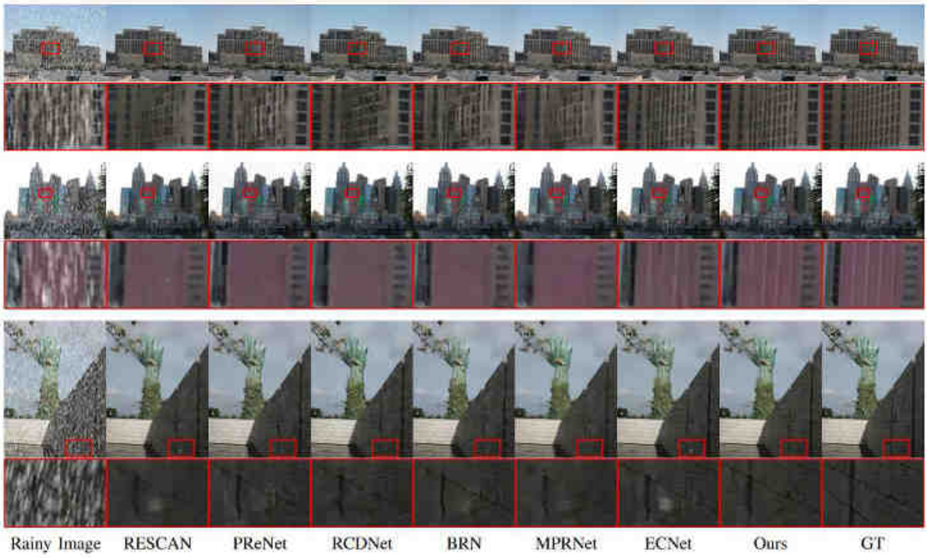

Figure 4: Visual comparisons on Rain800 dataset.

Table 2. Comparison of NIQE on the real dataset Internet-Data (red for optimal; Blue is suboptimal). ↓ Indicates that the lower the value, the better the rain removal effect

	RESCAN	PReNet	BRN	RCDNet	ECNet	MPRNet	Ours
NIQE↓	4.2963	4.1043	3.9758	3.8423	3.8592	3.9813	3.8252

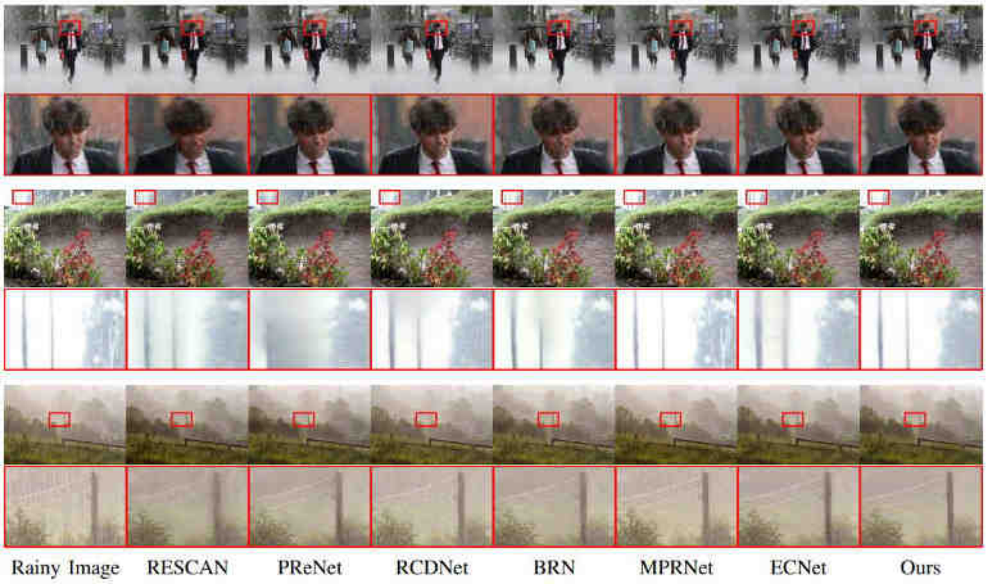

Figure 5: Visual comparisons on real-world dataset.

where f_{in} represents the input features, and $ReLU$ represents the Rectified Linear Unit activation function.

B. Detail Information Attention Block (DAB)

To better explore the complementary relationship between the rain streak and background layers, this chapter designs a Detail Information Attention Block. It generates degradation priors and performs attention operations on features based on the distribution of rain streaks, thus extracting complementary features from the background and rain streak layers to further optimize the texture details of the derained image. The specific structure is shown in Figure 1. Unlike standard self-attention modules that take the same feature as input, DAB takes the rain streak distribution map R, the coarse derained image B1, and the rainy image I as inputs, first learning their local context features, then projecting the features as Q, K, and V, respectively. Unlike the spatial attention that yields a feature map of size HW×HW, we reshape the projection maps of Q and K, generating a feature attention map of size C × C through the pointwise interaction between features, as shown in Figure 1. The attention map guides the network in extracting background detail and texture information from rainy image features. The process of the Detail Information Attention Block can be represented as:

$$f_{DA} = softmax\left(\left(F_K(R)\cdot F_Q(I)\right)\cdot F_V(I)\right)\#(9)$$

$$f_{out} = FFB(f_{DA}, f_{B1})\#(10)$$

where $F_K(R)\cdot F_Q(I)\cdot F_V(I)$ respectively denote the embedding equations used to generate the projection mappings. $(\cdot)$ denotes the dot product operation, and FFB represents the Feature Fusion Module, integrating the attention-focused features f_{DA} and background image features f_{B1} to achieve a background representation with richer details.

C. Feature Fusion Module (FFB)

In consideration of the redundancy and information differences between features across different domains, this chapter designs a Feature Fusion Block (FFB) to better integrate features from the Fourier and spatial domains. Specifically, the feature fusion module employs depth-wise separable convolutions and channel attention mechanisms to selectively aggregate features from different frequency domains across spatial and channel dimensions. This can be represented as:

$$f_{out} = CA(DSconv(Cat(f_1, f_2)) + f_1\#(11)$$

where $DSconv$ denotes depth-wise separable convolution, CA represents the channel attention module, and f_1, f_2 represent the two inputs to the feature fusion module.

Compared to simple skip connections or convolutional fusion methods, the feature fusion module designed in this chapter is more flexible and effective. The specific architecture is shown in Figure 2.

D. Loss Function

In the image deraining network, a coarse derained image B1 needs to be generated, whose texture structure should be consistent with that of the real rain-free image. The deraining result output by the Background Refinement Network needs to be consistent with the real rain-free image. Therefore, this chapter employs mean squared error loss and structural similarity loss to supervise the network, facilitating the training of the image network. The definitions of the two loss terms can be represented as:

$$L_{B1} = -\alpha_1 * SSIM(B1, B_{GT}) + ||B_{GT} - B1||_2 \#(12)$$

$$L_B = -\alpha_2 * SSIM(B.B_{GT}) + ||B_{GT} - B||_2 \#(13)$$

where L_{B1} and L_B represent the loss items for the image deraining network and the background refinement network, respectively, SSIM(·) denotes the operation for computing structural similarity, and $||\cdot||_2$ represents the L2 loss. α_1 and α_2 are two balancing parameters.

The dual-stage network is not trained separately for its two sub-networks; hence, we define a joint constraint loss to enhance the compatibility between the deraining model and background restoration. The total loss function can be represented as:

$$L = \lambda_1 * L_{B1} + \lambda_2 * L_B \#(14)$$

where λ_1 and λ_2 are the balancing parameters for the loss function.

IV. Experiments

A. *Dataset and Implementation Details*

This chapter evaluates the proposed method on four synthetic datasets: Rain100H[16], Rain200L[16], Rain200H[16], and Rain800[17]. Rain100H contains 1800 training images and 100 testing images. Rain200L/H each contain 1800 training images and 200 testing images. Rain800 contains 700 training images and 100 testing images. To test the method's performance in real scenarios, we utilized the Internet-Data[18] dataset, which includes 147 real-world rainy images.

Our network runs on a single NVIDIA 3090 GPU in a Pytorch environment. Network parameters are optimized using the Adam optimizer with settings: $\beta1 = 0.9$,$\beta2 = 0.999$. The learning rate is initially set to 1×10^{-3}, decaying to one-fifth of the current rate every 25 epochs. The model is trained on image blocks of size 96 × 96, with a batch size of 16, for a total of 100 epochs. The parameters λ1 and λ2 in equation (14) are set to 0.5 based on experience. α_1 and α_2 are both set to 10 to ensure the magnitude consistency of loss function terms. The performance of various comparison methods is quantitatively evaluated using PSNR and SSIM. Following the evaluation metrics used in previous deraining methods, PSNR[19] and SSIM[20]**.**metrics are calculated on the Y channel in YCbCr space.

B. *Comparative Experiment Results on Synthetic and Real Datasets*

To verify the effectiveness of the proposed FFT-DDN, we conducted quantitative comparisons with several state-of-the-art methods, including RESCAN[9], SIRR[18], PReNet[13], JORDER-E[10], RCDNet[15], BRN[21], ECNet[22], and MPRNet[20]. For these representative methods, where the authors provided pre-trained models, we directly used these models for testing. Otherwise, based on their provided codes, we retrained these models to ensure fairness in comparison. Table 1 lists the quantitative evaluation results for PSNR and SSIM. PSNR measures the content difference between rainy and rain-free images, while SSIM measures the structural similarity. Higher values of PSNR and SSIM indicate greater similarity. As can be seen from the table, our network outperforms other methods on average, demonstrating its strong adaptability across various rainy scenes.

To assess our method's applicability in real scenarios, we used the Rain100H dataset as the training set for our network and evaluated the performance of model on the Internet-Data real dataset. The real dataset contains real-world images in a rainfall environment, but lacks the corresponding rain-free images. In order to quantify the rain removal performance in the real world, this paper adopts the non-reference index NIQE [23]. The results are shown in Table 2. Our method obtains the best indicators, which indicates that the performance of the proposed method is superior to other comparison methods. Therefore, FFT-DDN also has good generalization for rain stripe removal in real environments.

Table 3 Results of ablation experiments on the Rain800 dataset

FFT-DDN	PSNR	SSIM
w/o DAB	29.32	0.903
w/o SFPEB	29.04	0.889
w/ Amp	29.83	0.907
w/ Amp and Pha	29.51	0.907
w/o FFB	30.06	0.904
Ours	**30.12**	**0.907**

Figure 3 and Figure 4 show the visual deraining results of various comparison methods on three images in the Rain100H and Rain800 datasets, respectively. As can be seen from the figure, compared with other methods, the results of the method in this paper have clearer edges and richer details. However, more texture information is lost in the results of other comparison methods, and obvious artifacts appear in the rain pattern area. Figure 5 shows the results of the visual comparison of rain removal on the real rain images in the Internet-Data dataset. It can be seen that other methods of removing rain can restore part of the details of the image, but there are still problems of image blur and residual rain streaks, while the rain removal results of this method can hardly see residual rain lines, and the reconstructed details are more realistic and the edges of objects are sharper, which indicates that the method proposed in this paper is significantly superior to other methods.

C. *Ablation Study*

In this section, we examine the impact of each module on the network and validate the effectiveness of the Fourier and spatial domain fusion mechanism and dense connections. All ablation experiments are conducted on the Rain800 dataset, using the quantitative metrics PSNR and SSIM for performance evaluation. The results are shown in Table 3 and Figure 6.

(1) Effectiveness of the Detail Attention Block (DAB)

The Detail Attention Block enhances the overall performance of the network by exploring the complementary relationship between the rain streak and background layers in images. In our experiments, we removed the DAB module and directly connected the IDN and BEN networks. The results, as shown in the 'w/o DAB' row of Table 3, indicate a decrease in PSNR by 0.8dB and SSIM by 0.004. This suggests that the DAB module indeed enhances the connection between the two models, improving the overall network performance.

(2) Effectiveness of the Spatial-Frequency Domain Fourier Phase Enhancement Block (SFPEB)

The SFPEB, being the core part of the network, effectively extracts and integrates features from the Fourier and spatial domains continuously. To validate the role of SFPEB, we designed three different experiments. First, SFPEB was replaced with traditional Resblocks to quantify its contribution, as shown in the 'w/o SFPEB' row. Then, we tested two different settings of SFPEB: (1) 'w/ Amp' – processing only the amplitude spectrum after Fourier decomposition, and (2) 'w/ Amp and Pha' – alternately processing the amplitude and phase spectra. In this setting, convolutional blocks in SFPEB alternately process only amplitude or phase spectrum information. According to Table 3, the current structure of solely processing the phase spectrum achieved the highest PSNR and SSIM values, proving its suitability for deraining tasks. The reduction in PSNR by 1.08dB after removing SFPEB highlights its effectiveness in mining features across different frequency domains.

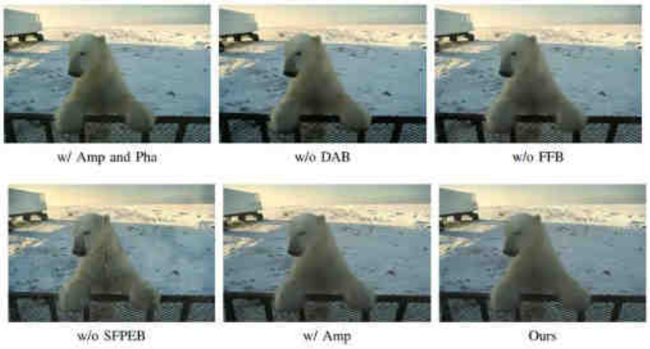

Figure 6 Comparison of deraining results of ablation experiment

Table.4 Quantitative comparison of parameters, FLOP and Runtime

Method	BRN	PReNet	RCDNet	MPRNet	ECNet	Ours
#Params(M)	0.375	0.169	2.958	3.637	14.998	1.475
FLOPs(G)	98.2	66.2	194.5	548.6	559.4	94.9
Runtime(s)	0.038	0.024	0.099	0.113	0.125	0.128

(3) Effectiveness of the Feature Fusion Module (FFB)

We designed experiments to verify the role of the FFB. By replacing FFB with concatenation operations and 1×1 convolutions, the 'w/o FFB' row in Table 3 shows the results of the network without FFB. The slight decrease in PSNR and more significant drop in SSIM indicate that FFB plays a major role in maintaining background structure, enhancing the structural similarity of images.

D. Model Complexity and Inference Time

To verify the efficiency of our method, we tested the FLOPs and inference time on an Nvidia 3090 GPU using 100 images of size 3 × 256 × 256. The inference time is the average testing time for these 100 images. Comparative results for FLOPs and running time are shown in Table 4. The results demonstrate that our proposed FFT-DDN can achieve efficient inference computation with less inference time.

E. Model Application

Eliminating the degrading effects of rain in rainy conditions while preserving reliable texture details is crucial for higher-level visual tasks, such as enhancing the accuracy of object recognition tasks. To further demonstrate the effectiveness of our proposed network, we used various deraining methods on the BDD350[24] dataset and performed object detection using YOLOv3[25]. The visual results are shown in Figure 7. Our method in the FFT-DDN deraining detection
results identified more vehicles and pedestrians compared to other methods. The rain-free images reconstructed by other methods contained more rain streaks, leading to lower accuracy in object recognition and instances of misidentification. Therefore, our proposed FFT-DDN is more beneficial for advanced computer vision tasks.

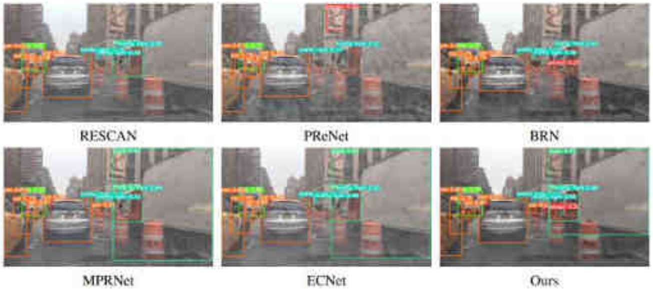

Figure 7 Deraining and object detection on BDD350 dataset.

V. CONCLUSION

To address the issue of poor deraining performance of current networks in real rainy environments, we propose a Dual-stage Spatial-Frequency Domain Deraining Network, utilizing a dual-stage architecture and Detail Attention Block (DAB) to achieve dual objectives of rain streak removal and background refinement. It enhances learning capabilities and explores the intrinsic connection between the image deraining and background refinement stages. By embedding Fourier transform into the neural network, we designed the Spatial-Frequency Domain Fourier Phase Enhancement Block (SFPEB), enabling the network to fuse features from multiple frequency domains for better generalization. Furthermore, we developed a Feature Fusion Module (FFB) to enhance feature fusion. Extensive experiments on synthetic and real datasets demonstrate that our method can restore rain-free images with richer content and details, outperforming some of the current advanced methods.

REFERENCES

[1] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask R-CNN. In *IEEE International Conference on Computer Vision (ICCV)*. Venice, Italy, pp. 2980-2988.

[2] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. 2017. Feature Pyramid Networks for Object Detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA, pp. 936-944.

[3] Khan Muhammad, Jamil Ahmad, Zhihan Lv, Paolo Bellavista, Po Yang, and Sung Wook Baik. 2019. Efficient Deep CNN-Based Fire Detection and Localization in Video Surveillance Applications. In *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 7, pp. 1419-1434.

[4] Yuntong Ye, Yi Chang, Hanyu Zhou, and Luxin Yan. 2021. Closing the Loop: Joint Rain Generation and Removal via Disentangled Image Translation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA, pp. 2053-2062.

[5] Yu Li, Robby T. Tan, Xiaojie Guo, Jiangbo Lu, and Michael S. Brown. 2016. Rain Streak Removal Using Layer Priors. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA, pp. 2736-2744.

[6] Yu Luo, Yong Xu, and Hui Ji. 2015. Removing Rain from a Single Image via Discriminative Sparse Coding. In *IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile, pp. 3397-3405.

[7] Li-Wei Kang, Chia-Wen Lin, and Yu-Hsiang Fu. 2012. Automatic Single-Image-Based Rain Streaks Removal via Image Decomposition. In *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1742-1755.

[8] Shuhang Gu, Deyu Meng, Wangmeng Zuo, and Lei Zhang. 2017. Joint Convolutional Analysis and Synthesis Sparse Representation for Single Image Layer Separation. In *IEEE International Conference on Computer Vision (ICCV)*. Venice, Italy, pp. 1717-1725.

[9] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. 2018. Recurrent Squeeze-and-Excitation Context Aggregation Net for Single Image Deraining. In *Computer Vision – ECCV 2018: 15th European Conference*. Munich, Germany, September 8–14, 2018, Proceedings, Part VII. Springer-Verlag, Berlin, Heidelberg, 262–277.

[10] Wenhan Yang, Robby T.Tan, Jiashi Feng, Zongming Guo, Shuicheng Yan, and Jiaying Liu. 2020. Joint Rain Detection and Removal from a Single Image with Contextualized Deep Networks. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 6, pp. 1377-1393.

[11] Xueyang Fu, Jiebin Huang, Delu Zeng, Yue Huang, Xianhao Ding, and John Paisley. 2017. Removing Rain from Single Images via a Deep Detail Network. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA, pp. 1715-1723.

[12] He Zhang and Vishal M Patel. 2018. Density-Aware Single Image De-raining Using a Multi-stream Dense Network. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City, UT, USA, pp. 695-704.

[13] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. 2019. Progressive Image Deraining Networks: A Better and Simpler Baseline. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA, pp. 3932-3941.

[14] Syed Waqas Zamir *et al*. 2021. Multi-Stage Progressive Image Restoration. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA, pp. 14816-14826.

[15] Hong Wang, Qi Xie, Qing Zhao, and Deyu Meng. 2020. A Model-Driven Deep Neural Network for Single Image Rain Removal. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA, pp. 3100-3109.

[16] Wenhan Yang, Robby T. Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. 2017. Deep Joint Rain Detection and Removal from a Single Image. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA, pp. 1685-1694.

[17] He Zhang, Vishwanath Sindagi, and Vishal M. Patel. 2020. Image De-Raining Using a Conditional Generative Adversarial Network. In *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 11, pp. 3943-3956.

[18] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. 2019. Semi-Supervised Transfer Learning for Image Rain Removal. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA, pp. 3872-3881.

[19] Quan Huynh-Thu and Mohammed Ghanbari. 2008. Scope of Validity of PSNR in Image/Video Quality Assessment. In Electronics Letters, vol. 44, pp. 800-801.

[20] Zhou Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” IEEE Trans. Image Process., vol. 13, no. 4, pp. 600-612, Apr. 2004.

[21] Zamir S. W., Arora A, Khan S, et al. Multi-Stage Progressive Image Restoration[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 2021: 14816-14826.

[22] Dongwei Ren, Wei Shang, Pengfei Zhu, Qinghua Hu, Deyu Meng, and Wangmeng Zuo. 2020. Single Image Deraining Using Bilateral Recurrent Network. In *IEEE Transactions on Image Processing*, vol. 29, pp. 6852-6863.

[23] Yizhou Li, Yusuke Monno, and Masatoshi Okutomi. 2022. Single Image Deraining Network with Rain Embedding Consistency and Layered LSTM. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Waikoloa, HI, USA, pp. 3957-3966.

[24] Anish. Mittal, Rajiv Soundararajan, and Alan C Bovik. 2013. Making a “Completely Blind” Image Quality Analyzer. In *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209-212.

[25] Fisher Yu *et al*. 2020. BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, pp. 2633-2642.

[26] Joseph Redmon and Ali Farhadi. 2018. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.