Advances in Continuous and Discrete Models

RESEARCH Open Access

DWTN: deep wavelet transform network for lightweight single image deraining

Wenyin Tao[1], Xuefeng Yan[1,2*], Yongzhen Wang[3] and Mingqiang Wei[1]

*Correspondence: yxf@nuaa.edu.cn
[1]College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, Jiangsu, China
[2]Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, Jiangsu, China
Full list of author information is available at the end of the article

Abstract

On rainy days the uncertainty of the shape and distribution of rain streaks can cause the images captured by RGB image-based measurement tools to be blurred and distorted. Thanks to the wavelet transform ability to provide spatial and frequency domain information about an image and its multidirectional and multiscale nature, it is widely used in traditional image enhancement methods. In image deraining the distribution of rain streaks is not only related to spatial domain features but is also closely related to frequency domain spatial features. However, deep learning-based rain removal models mainly rely on the spatial features of the image, and RGB data can hardly distinguish rain marks from image details, which leads to the loss of crucial image information during rain removal. We have developed a lightweight single-image rain removal model called the deep wavelet transform network (DWTN) to address this limitation. This method separates image details from rain images and can more effectively remove rain marks. The proposed DWTN has three significant contributions. First, DWTN uses the feature components after the wavelet transform as the input to the model and assigns a separate frequency-aware enhancement block (FAEB) to each element. These blocks focus on specific frequency features that benefit the rain removal task. Second, we introduce a frequency feature fusion block (FFFB) that fuses different wavelet components to reduce noise and enhance the image background through a channel attention mechanism while attenuating rain streaks. Finally, we design a spatial feature enhancement block (SFEB), which uses a spatial attention mechanism to calibrate the spatial position of features to improve the rain removal performance. We evaluate the performance of DWTN using PSNR and SSIM on four synthetic datasets and NIQE and BRISQUE on two real datasets. The results of the evaluation of the six datasets and at least four performance metrics show that the proposed DWTN is superior to existing methods.

Keywords: DWTN; Image deraining; Wavelet transform; Signal processing

1 Introduction

The performance of high-level vision measurement task system is hightly dependent on the quality of input images in machine vision-based task system, such as autonomous driving [1] and object detection [2]. In rainy weather, rain streaks can cause severe degradation in the quality of images captured by vision measurement system [3]. Specifically, suppose the degradation model is $I_{clr} = I_r - R$, where I_{clr} represents a clean image in the image de-

raining problem, and I_r and R denote a rainy day image and streaks obtained through the network. According to that, deraining networks have become a popular tool for performing high-level vision tasks due to their ability to remove rain streaks from images [4]. Since the birth of JORDER [5] in 2017, human experts have conducated a huge number of experiments and consequently devised serval useful structures, such as SCAN [6] and PDR-Net [7]. However, designing an effective rain removal network has two main chanllenges. Firstly, the deranining task is the pretask of the high-level vision task, so the execution efficiency of the rain removal network affects the execution efficiency of the high-level vision task. Secondly, the output image quality of the rain removal network also directly affects the performance of the high-level vision task. The optimal network architecture and parameters depend on the specific rain streaks distribution, image content, and desired trade-off between removing rain streaks and preservation of image details [8].

Most existing single image rain removal methods work in the RGB domain. Jiang et al. [9] proposed a multiscale progressive fusion network for image deraining by excavating and exploiting the inherent correlations of rain streaks across different scales. However, the limited receptive field of the convolution and the characteristic of capturing the local features hinder the ability of the model to eliminate rain streaks. To alleviate such limitations, Xiao et al. [10] utilize a more general transformer to replace CNNs as the network backbone. Transformers can better model the nonlocal information for high-quality image reconstruction. RGB domain-based work has achieved significant success in single-image deraining tasks because of its complex and deeper network architecture. However, these methods have difficulty in separating the rain pattern from the real image content [11, 12].

Many traditional methods [13] have shown that reconstructing degraded images is more straightforward in the frequency domain. Wavelet-based approaches have been extensively studied in computer vision and have exhibited excellent performance in various tasks such as classification [14], image denoising [15], and image deraining [16]. Using wavelet transform in the image deraining task has four main advantages. First, wavelet transforms can decompose an image into components of different scales, and rain patterns also appear in images at various scales so that the model can analyze image features at different resolutions. This facilitates the removal of rain streaks of different sizes and shapes. Second, wavelet transforms can provide spatial and frequency information about an image, and rain streaks often cause local changes in space and frequency. Third, wavelet transforms can effectively separate noise (such as rain streaks) from the signal (image content) in an image. With appropriate thresholding, raindrop noise can be suppressed in the wavelet coefficient space while retaining the main details of the image. Fourth, wavelet transforms can provide multidirectional decompositions, making them very effective in detecting and processing raindrop marks with directionality. Liu et al. [17] constructed a U-Net structured image restoration network using multiscale features generated by the wavelet transform, an early study of using the wavelet transform for image restoration in deep learning. Hsu et al. [18] also referred to the U-Net structure to design the network and used a wavelet transform to divide the image into high- and low-frequency parts. Finally, the results are mixed and concatenated layer by layer to improve the image-deraining effect. In most wavelet transform-based image deraining studies, scholars have directly spliced image features of various frequencies after wavelet transformation and then used deep learning networks to recover images. Although these methods have achieved some

success in the rain removal task, they ignore the characteristics of the different frequency features and the correlation between them.

To resolve this problem, we propose a novel lightweight multistage image de-raining network using wavelet transform called **D**eep **W**avelet **T**ransform **N**etwork (DWTN). The input image features are converted into high- and low-frequency features using a wavelet transform, and the image is restored by removing the rain streaks from the structural information of the low-frequency features and the detailed information of the high-frequency features. After the wavelet transforms, the network module responsible for the low-frequency features learns and restores the structural and textural information of the image from the low-frequency features, whereas the network module responsible for the high-frequency features removes the rain pattern from the high-frequency features and preserves and enhances the edge detailed information of the image.

Unlike previous single-image deraining networks based on wavelet transforms, the DWTN proposed in this paper processes each feature component individually, allowing the subnetwork to focus on the feature components of one characteristic, resulting in the reconstruction of better image structures and the restoration of more detailed image textures. To improve the ability of the network to remove rain and preserve more detailed information, we have incorporated low-frequency feature information into the high-frequency network using the frequency feature fusion module (FFFB). This helps to enhance the high-frequency features. We have also introduced high-frequency feature information into the low-frequency network through FFFB to enhance the low-frequency features. By doing this, FFFB enables the network to effectively capture the relationship between high- and low-frequency features, resulting in more effective removal of rain streaks in the image. By adjusting the network parameters, DWTN can implement rain removal networks with different parameter sizes according to different working scenarios to meet the requirements of different working scenarios on model real time and performance. We show experimental results that demonstrate the effectiveness of DWTN, as shown in Fig. 1. The experimental results show that our method outperforms the state-of-the-art single-image rain deraining methods in standard benchmark tests. In the Rain200L, Rain200H, Rain800, and Rain1400 benchmark tests, our method achieves significantly

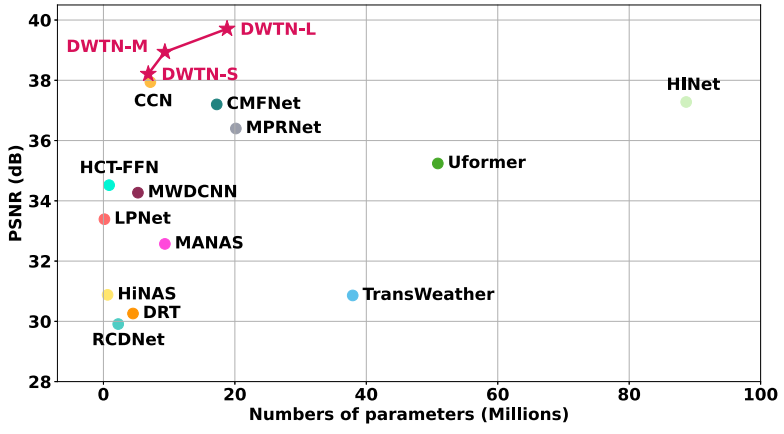


Figure 1 Single image deraining on the Rain200L dataset. Our method performs much better than other state-of-the-art algorithms

higher PSNR and SSIM scores, proving its effectiveness in removing rain streaks while preserving image structure and details. The advantages of our approach are mainly in the a priori knowledge of the wavelet transform, which allows the network to achieve stronger image deraining using fewer parameters, and the lightweight network enables higher real-time performance and allows subsequent high-level visual tasks to be performed more efficiently.

The contributions of this study are summarized as follows:

- We propose a novel method, Deep Wavelet Transform Network (DWTN). It implements a lightweight network and high-quality image deraining by introducing a prior knowledge of wavelets into the network. In addition, DWTN uses a Frequency-Aware Enhancement Block (FAEB) to extract feature information at different frequencies and then uses a Spatial Feature Enhancement Block (SFEB) to preserve more detailed information for image restoration.
- We have designed and implemented a frequency feature fusion mechanism (FFFB). By exploring the existence of latent relationships between high- and low-frequency features the network can remove streaks from images more efficiently.
- The experimental results demonstrate that the proposed DWTN delivers significant enhancements in performance on both synthetic and real datasets when compared to the most advanced methods currently available.

The subsequent sections of this work are structured as follows. Section 2 provides an overview of previous research, specifically focusing on deraining networks that utilize deep learning techniques and picture enhancement methods based on wavelet transforms. Section 3 presents the proposed DWTN network designed for single image deraining. Section 4 includes comparisons with state-of-the-art techniques, comments, and ablation studies. The report concludes in Sect. 5, summarizing findings and outlining potential areas for future research.

2 Related work

2.1 Deep learning-based deraining networks

A rain removal network is an end-to-end neural network where the input contains rain streaks, and the output is a clean image with the rain streaks removed. The rain removal network aims to remove rain streaks while preserving the structure and detailed information in the image [19, 20]. Convolutional neural networks (CNNs) are a widely used method trained on a dataset of paired rain images and clean images. It then uses a loss function to quantify the difference between the rain and clean images. Over the past few years, many methods have been invented to remove rain streaks from a single input image. These methods can be divided into two categories, CNN-based and transformer-based. Since deep residual networks [21] have performed well in complex visual tasks, Fu et al. [22] proposed a deep detail network (DDN), which uses a continuous network topology to improve the rain removal effect in the model. Unlike Fu et al., Ren et al. [23] proposed a simple and efficient progressive rain removal network that achieves image rain removal and reduces the number of model parameters by sharing parameters at multiple levels.

The main goal of low-level visual tasks is to improve the quality of input images for high-level visual tasks. Consequently, the effectiveness of complex visual tasks is impacted by the pace at which more straightforward visual tasks are performed. Fu et al. [24] introduced a lightweight deraining network called LPNet. This network utilizes the feature

pyramid and residual structures to extract multiscale features from a picture, resulting in image restoration. Xia et al. [6] introduced the RESCAN method to identify connections between different levels of picture characteristics to improve the performance of the progressive deraining network. The transformer model has the ability to capture correlations between features that are far apart and represent global features. This makes it highly capable of executing various complex tasks, including natural language processing [25] and computer vision [26].

Transformer-based network models such as IPT [27], Uformer [28], SwinIR [29], Restormer [30], and SDNet [31] have been used for low-level vision tasks. Chen et al. proposed the pretrained image processing transformer (IPT) model, which combines multiple head-tail structures and codec structures to solve low-level vision tasks. It has been trained and verified on many datasets, and the results show that IPT outperforms CNN-based models in multiple tasks such as image deraining, denoising, and quality enhancement, demonstrating the advantages of Transformer in model performance. Uformer utilizes a hierarchical codec framework similar to U-Net and uses nonoverlapping local windows to calculate self-attention while integrating local feature enhancement modules to improve the ability of Transformer models to process local information. In addition, some scholars have introduced neural architecture search (NAS) methods in the image restoration task [32, 33], which obtain deraining models through search networks with differentiable parameters. Zhang et al. [34] proposed a hierarchical neural architecture search model (HiNAS) for image denoising. This method uses a gradient-based network search algorithm and creates a hierarchical search space with an adaptive perception field to obtain a model with excellent performance. Quan et al. [35] proposed the CCN network, which is the first work to use NAS in the image deraining task. The network searches for a rain removal network and a raindrop removal network separately through NAS. It combines the two networks in a cross-way to achieve the effect of removing rain and raindrops simultaneously. However, because rain and image contents are mixed in the RGB domain, image details will inevitably be lost when rain is removed.

2.2 Wavelet transform-based image reconstruction

Over the past decades, wavelet-based methods have been explored in many computer vision tasks, including image classification [36], face aging [37], style transfer [38], etc. Among them, wavelet transforms are most widely used in low-level vision tasks, such as image superresolution [39], image denoising [15], and image deblocking [40]. Recently, some researchers have combined the wavelet transform with deep neural networks in image processing tasks [41, 42]. Liu et al. [43] proposed a Multilevel Wavelet CNN (MWCNN) network for image restoration. The network is a multilayered structure that applies wavelet transforms to different frequency subbands of the image. The authors designed a convolutional neural network that learns the coefficients of these wavelet transforms to improve the quality of image recovery. This combination of wavelet transforms and convolutional neural networks better preserves the structural information of the image while reducing the noise. Demirel et al. [44] performed image superresolution by interpolating the high-frequency subimage obtained from the discrete wavelet transform (DWT) with the original image. Several researchers [22, 45] have demonstrated that deep neural networks operating in the RGB domain are not successful in learning the transformation from images containing rain to images free of rain. Initial investigations into

rain removal typically employed decomposition [46] and filtering [22] techniques to isolate rain streaks from photos. Due to the more in-depth examination of rain removal jobs, the conventional methods are no longer appropriate. Scholars have successfully employed the wavelet transform to remove rain from images [47].

Unlike previous wavelet transform-based rain removal networks, the method proposed in this study utilizes the wavelet transform to process the image high-frequency details and low-frequency structures. It introduces novel frequency-aware enhancement block (FAEB), frequency feature fusion block (FFFB), and spatial feature enhancement block (SFEB). The streaks are removed while preserving the low-frequency information of the image. In the high-frequency part, detailed enhancement is achieved by dense residuals and low-frequency guidance, significantly improving model performance.

3 Method

In this study, our goal is to decompose the input image features into high- and low-frequency features by a wavelet transform, preserve the high-frequency details of the image, reconstruct the low-frequency structure of the image in the decomposed features, and finally achieve the effect of removing the rain streaks in the restored image. Section 3.1 introduces the overall structure of the Deep Wavelet Transform Network (DWTN). Then Sect. 3.2 provides the Frequency Awareness Enhancement Block (FAEB) implementation process. Then the design ideas and details of the frequency feature fusion block (FFFB) and spatial feature enhancement block (SFEB) are discussed in Sects. 3.3 and 3.4. Finally, the methods and loss functions for image reconstruction are introduced in Sect. 3.5.

This study proposes a deep wavelet transform network that is capable of removing rain streaks from an image while preserving image details and reconstructing the image structure, as shown in Fig. 2. The network consists of four parts: a shallow feature extraction network, frequency feature enhancement and cross-fusion network, spatial feature enhancement network, and image restoration network. To construct deep learning networks with various parameter scales, we combine these four parts into one stage, and by stacking multiple such stages the deep learning network can achieve various image rain removal performances.

3.1 Network architecture

In rain images, direct image recovery by deep learning network leads to loss of image detail information as the rain streaks overlap with the image background. In addition, different

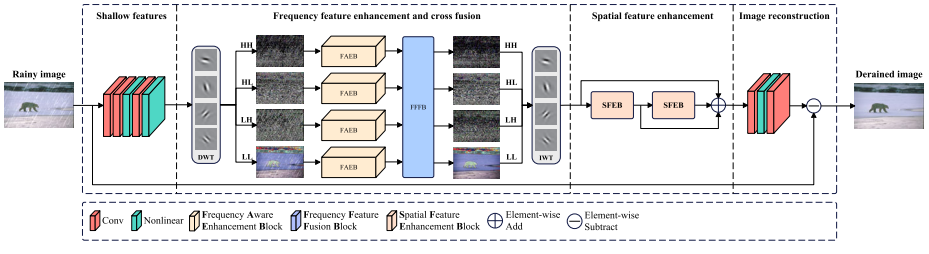


Figure 2 The DWTN architecture comprises four main components: the shallow feature extraction network, the frequency feature enhancement and cross-fusion network, the spatial feature enhancement network, and the image restoration network

rain streaks are also mixed with the texture details of the image, which makes it very difficult to remove the rain streaks directly from the image. To solve this problem, in this study the input image is decomposed into foreground rain streak images and background images by the 2D Discrete Wavelet Transform (DWT). The image features at different frequencies are extracted by an FAEB module. The FFFB module fuses the features at different frequencies. The image features without rain streaks are restored using 2D Inverse Wavelet Transform (IWT) and SFEB, and finally, a clean image without rain streaks is obtained using an image restoration network.

Specifically, given a degraded image $I_r \in \mathbb{R}^{3\times H\times W}$, DWTN first applies a 5×5 convolution layer to generate the shallow feature map $X_0 \in \mathbb{R}^{C\times H\times W}$, where C denotes the number of channels, and $H \times W$ represents spatial locations. Then the shallow features generate four components by 2D-DWT, which are low–low component (X_{LL}), low–high component (X_{LH}), high–low component (X_{HL}), and high–high component (X_{HH}), where X_{LL} can be regarded as the structural information of the image; X_{LH}, X_{HL}, and X_{HH} are the details and edges of the image. In this process the input feature map X_0 is first decomposed horizontally using a high-pass filter and a low-pass filter to obtain the low-frequency component $X_L \in \mathbb{R}^{C\times H\times \frac{W}{2}}$ and the high-frequency component $X_H \in \mathbb{R}^{C\times H\times \frac{W}{2}}$. Then the features $[X_L, X_H]$ obtained in the previous step are decomposed vertically using high- and low-pass filters to obtain the full components $X_{DWT} = [X_{LL}, X_{LH}, X_{HL}, X_{HH}] \in \mathbb{R}^{C\times \frac{H}{2}\times \frac{W}{2}}$ of the input feature map X_0. Since the decomposition process uses interval sampling, the output feature width and height are half of the input features, as shown in Fig. 3.

After the wavelet transform, the FAEB module extracts features from X_{DWT} separately, e.g., $X_{A_{LL}} = f_{FAEB_{LL}}(X_{LL})$. Each FAEB module consists of N residual dense blocks (RDBs) with activation function ReLU, and the size of output features is the same as that of input features.

It has been demonstrated that the feature components obtained after using wavelet transform are not independent but are latently correlated [18]. Therefore we use the FFFB module to fuse different feature components according to the rules to obtain new feature components. In this process, we first concatenate the three feature components of the input based on the channels, e.g., $X_{A_{cat}} = Concat\left(X_{A_a}, X_{A_b}, X_{A_c}\right) \in \mathbb{R}^{3\times C\times \frac{H}{2}\times \frac{W}{2}}$. Then $X_{A_{cat}}$

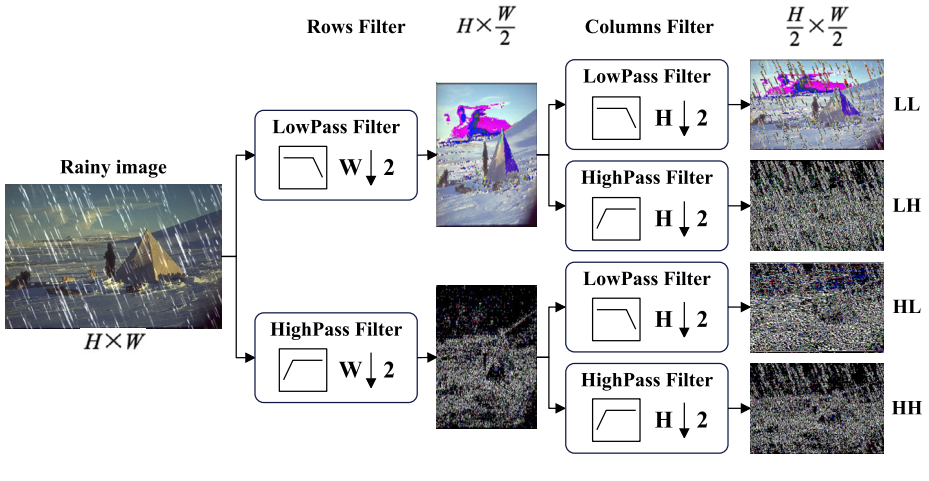


Figure 3 Procedure of 1-level 2D-DWT decomposition. ($W \downarrow 2$) indicates that the width is halved, and ($H \downarrow 2$) indicates that the height is halved

is passed into the FFFB module for feature fusion to obtain $X_{A_F} \in \mathbb{R}^{C\times\frac{H}{2}\times\frac{W}{2}}$. At this time, $X_{A_F} = \left[X_{A_{F\_LL}}, X_{A_{F\_LH}}, X_{A_{F\_HL}}, X_{A_{F\_HH}}\right]$ enhances the features that are advantageous for image restoration and weakens those that are not advantageous for image restoration. Finally, all the cross-fused feature components X_{A_F} are recovered to the complete image feature map $X_{IWT} \in \mathbb{R}^{C\times H\times W}$ using 2D-IWT.

After the inverse wavelet transform, the image feature X_{IWT} is passed into N SFEB modules to get $X_{SE_i} = \left[X_{SE_1}, X_{SE_2}, \ldots, X_{SE_N}\right] \in \mathbb{R}^{C\times H\times W}$. The SFEB module implements spatial enhancement of image features and fuses multiple enhanced image features through skip connections $X_E = Add\left(X_{IWT} + X_{SE_1} + X_{SE_2} + \cdots + X_{SE_N}\right)$.

Finally, the rain streak $R = Conv\left(X_{SE}\right) \in \mathbb{R}^{3\times H\times W}$ is obtained by a 5×5 convolution layer in the image reconstruction network, and the reconstructed image I_{clr} is obtained using the image degradation model $I_{clr} = I_r - R$.

3.2 Frequency awareness enhancement block

There are two main challenges to the single image rain removal task. Firstly, the rain streak in the image foreground is mixed with the image background, making it difficult to separate in the RGB domain. Therefore some scholars have proposed to convert the image data in the RGB domain into frequency domain data, which is beneficial for image denoising and reconstruction [48]. The reason is that high-frequency data contain rich information about image edges and details, while low-frequency data contain structural information about the image. Secondly, low-level visual tasks are usually used as pretasks for high-level visual tasks, so it is necessary to use hardware-friendly model design methods.

To solve the above two problems, we use 2D-DWT to decompose the rain image feature X_0 into multiple frequency components. We build a structurally consistent frequency-aware feature enhancement (FAEB) network for each feature component to predict the corresponding deraining feature $\left[X_{A_{LL}}, X_{A_{LH}}, X_{A_{HL}}, X_{A_{HH}}\right]$. The network mainly utilizes hardware-friendly residual dense block (RDB) structures to optimize the learning of complex features and backpropagation of gradients. RDB is a combination of residual network structure [21] and dense network structure [49]. In residual network structure, introducing a forward feedback connection between the input and output can effectively alleviate the problem of gradient disappearance caused by the increase of network depth, so that the deeper network can still maintain good performance and efficiency. Residual networks are suitable for image restoration because the similarity between low- and high-quality images is very high, and the residuals between them are very sparse, so the model can get high-quality images by learning only less information through residual networks. Each network layer in a dense network accepts the outputs of all previous layers as additional inputs, thus enabling feature taking, enhancing feature propagation, and reducing the number of parameters. Consequently, RDBs are extensively employed for image restoration and single-image superresolution (SISR) tasks [50].

The implementation of FAEB consists of three parts. The first is a set of RDB modules with ReLU, a combination that improves the performance of image feature extraction and prevents overfitting. The second is merging multiple RDB output features using a residual structure $X_{RDB_{merg}} \in \mathbb{R}^{C\times\frac{H}{2}\times\frac{W}{2}}$. Finally, 5×5 Conv+ReLU+Conv (CRC) is used to fuse the merged image features to output more effective image features, e.g., $X_{A_{LL}} \in \mathbb{R}^{C\times\frac{H}{2}\times\frac{W}{2}}$, as shown in Fig. 4. Suppose the input feature of FAEB is X_{LL}. The mathematical description

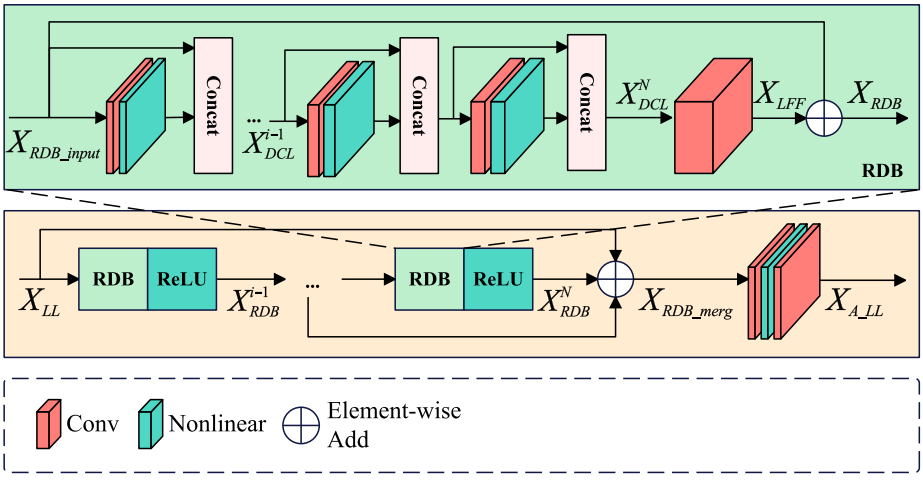


Figure 4 Architecture of frequency awareness enhancement block

of FAEB can be expressed as

$$\begin{aligned} &X_{RDB}^{i} = ReLU\left(f_{RDB_i}\left(X_{RDB}^{i-1}\right)\right) + X_{RDB}^{i-1}, 1 \le i \le N, \\ &X_{RDB_{merg}} = X_{LL} + \sum_{i=1}^{N} X_{RDB}^{i}, \\ &X_{A_{LL}} = Conv\left(ReLU\left(Conv\left(X_{RDB_{merg}}\right)\right)\right), \end{aligned} \tag{1}$$

where f_{RDB_i} denotes the ith RDB operation, X_{RDB}^{i-1} denotes the input of the RDB operation (if $i = 1$, then $X_{RDB}^{0} = X_{LL}$), $X_{RDB_{merg}}$ denotes the merging of the input of the FAEB and the output of the RDB module using the residual structure, and $X_{A_{LL}}$ denotes the output of the FAEB module.

The RDB implementation consists of a Dense Connected Layer (DCL), Local Feature Fusion (LFF), and global residual structure. Suppose there are N DCL modules in the RDB module, and the input feature of the RDB is $X_{RDB\_input} \in \mathbb{R}^{C\times H\times W}$. Through the concatenation operation in the DCL, the output of the Nth DCL module is $X_{DCL}^{N} \in \mathbb{R}^{N\times C\times H\times W}$. The LFF obtains feature $X_{LFF} \in \mathbb{R}^{C\times H\times W}$ of the same size as the input feature through the convolution of 1×1. This operation not only fuses high-dimensional features of the output of the last DCL but also reduces the number of channels of the features, which decreases the parameters of the model and provides correct input for the subsequent residual structure. The mathematical expression of RDB is described as follows:

$$\begin{aligned} &X_{DCL}^{i} = Concat\left(X_{DCL}^{i-1}, ReLU\left(Conv\left(X_{DCL}^{i-1}\right)\right)\right), \\ &X_{LFF} = Conv_{1\times 1}\left(X_{DCL}^{N}\right), \\ &X_{RDB} = X_{RDB\_input} + X_{LFF}, \end{aligned} \tag{2}$$

where X_{DCL}^{i} denotes the output of each DCL module (when $i = 1$, the input of the DCL module is the input of the RDB module), X_{LFF} denotes the output of the LFF after a 1×1 convolution operation, and X_{RDB} denotes the output of the RDB module obtained using the residual structure.

3.3 Frequency feature fusion block

2D-DWT divides the image features into four components: X_{LL}, X_{LH}, X_{HL}, and X_{HH}. Although X_{LL} and X_{HH} maintain the low- and high-frequency information of the image, respectively, X_{LH} and X_{HL} still keep a part of the high-frequency image detailed information and a part of the low-frequency image structure information. Therefore the fusion of feature components can strengthen the low-frequency information in X_{LL} and the high-frequency information in X_{HH}. Some scholars [48] have proposed that feeding image features directly into the CNN in image-denoising tasks is unreasonable. Because CNN is fair for each feature channel, the distribution of feature channels can be considered as the partition of image frequency, and the different channel weights respond to the other model choices for image frequency. Therefore, when performing feature component fusion, it is necessary to use the channel attention mechanism to assign different weights to different channels so that the model can select the parts that are advantageous to the image deraining task.

In this study, we propose a novel fusion mechanism for frequency feature components, as shown in Fig. 5. In Fig. 5, we tried three feature fusion methods: Methods A, B, and C. In Method A, in addition to considering that X_{LL}, X_{LH}, X_{HL}, and X_{HH} are latently related to

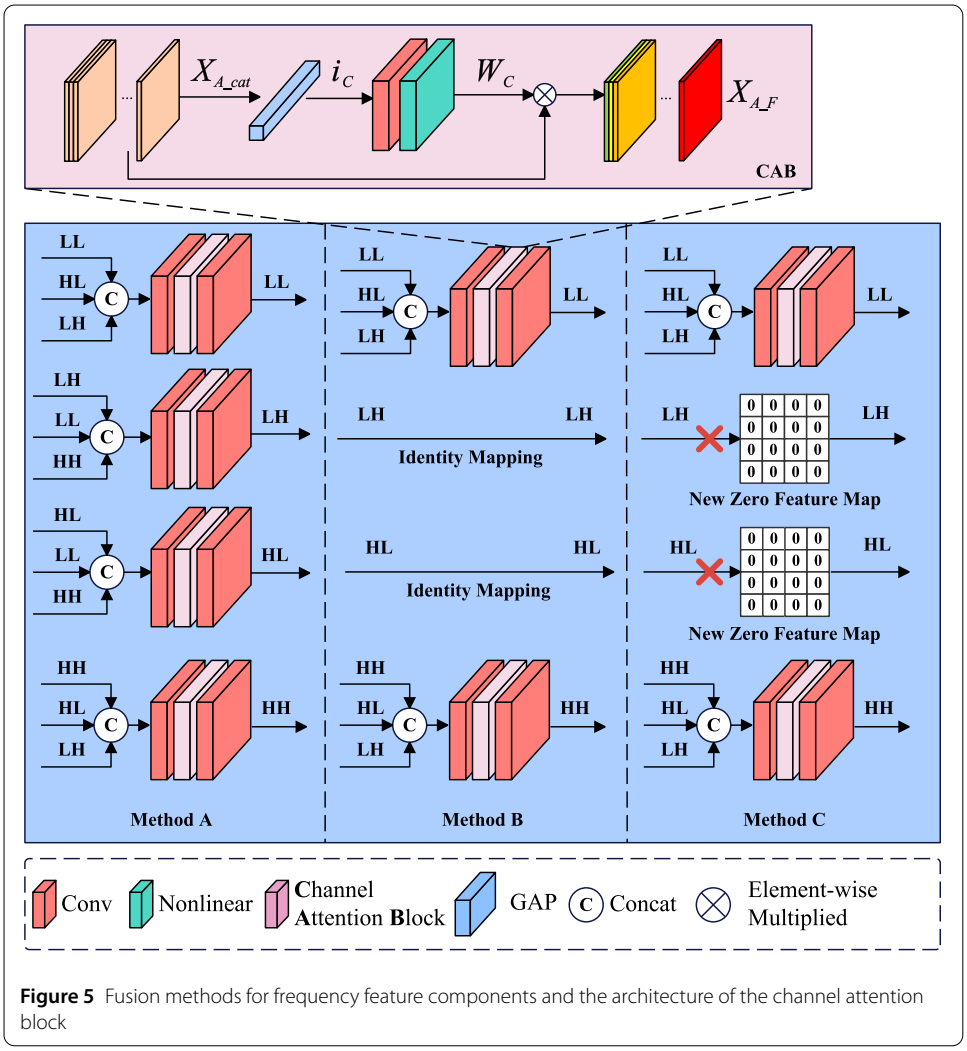


Figure 5 Fusion methods for frequency feature components and the architecture of the channel attention block

Table 1 The PSNR and SSIM results of various frequency feature fusion algorithms on the Rain200L dataset. The color red signifies the highest level of performance

Model	PSNR	SSIM
Model A	38.44	0.984
Model B	38.02	0.983
Model C	38.26	0.983

each other, the differences between the frequency feature components are also considered. For example, X_{LL} and X_{HH} contain completely different information, so fusing X_{LL} and X_{HH} would cause the generated frequency features to have incorrect information. Therefore, when fusing frequency feature components, we only fuse the relevant frequency feature components to enhance the information of the different frequency feature components. Instead of fusing the feature components LH and HL, the feature components LH and HL are obtained using the identity mapping in method B. In method C the feature components LH and HL are discarded and replaced with a feature of the same shape as LH and HL, which have zero values. In addition, based on the DWTN-S model, we evaluated the three fusion methods separately on the Rain200L dataset, as shown in Table 1. The results show that our proposed method A is more advantageous. The mathematical expression of FFFB is described as follows:

$$\begin{aligned} X_{A_{cat}} &= Concat\left(X_{A_a}, X_{A_b}, X_{A_c}\right), \\ X_{A_F} &= Conv_{1\times 1}\left(f_{CAB}\left(Conv\left(X_{A_{cat}}\right)\right)\right), \end{aligned} \tag{3}$$

where X_{A_a}, X_{A_b}, and X_{A_c} denote the frequency feature components according to Method A in Fig. 5, and these frequency feature components are concatenated according to the channel to obtain $X_{A_{cat}} \in \mathbb{R}^{3\times C\times \frac{H}{2}\times \frac{W}{2}}$. $Conv_{1\times 1}$ denotes the 1×1 convolution, f_{CAB} denotes the channel attention block, $X_{A_F} \in \mathbb{R}^{C\times \frac{H}{2}\times \frac{W}{2}}$, and $Conv$ denotes the fused frequency feature components.

Channel Attention Block (CAB) To assign different weights to the channels of the frequency feature components, we designed channel attention blocks based on ECA [51]. Suppose that the concatenated frequency feature components are given by $X_{A_{cat}} \in \mathbb{R}^{3\times C\times \frac{H}{2}\times \frac{W}{2}}$, where C represents the number of channels in the feature map, and H and W represent the height and width of the feature map, respectively. Firstly, the information for each channel $i_C \in \mathbb{R}^C$ is obtained by Global Average Pooling (GAP). The mathematical expression for GAP is as follows:

$$i_C = GAP\left(X_{A_{cat\_c}}^{H\times W}\right) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} X_{A_{cat\_c}}(i,j), \tag{4}$$

where $X_{A_{cat\_c}}(i,j)$ represents the feature value at position (i,j) of the feature map $X_{A_{cat}}$ in the C channel. GAP represents global average pooling.

Then we use the convolution operation $Conv$ to get the information of the remapped channels. Next, the activation function $Sigmoid$ is employed to activate or inhibit the different channels to obtain the final channel weights $W_c \in \mathbb{R}^{C\times 1\times 1}$. The mathematical ex-

pression is as follows:

$$W_c = Sigmoid\left(Conv\left(i_c\right)\right). \tag{5}$$

Finally, the channel weights W_c are elementwise multiplied with the input frequency feature components to obtain the feature frequency components $X_{A_F} \in \mathbb{R}^{C\times\frac{H}{2}\times\frac{W}{2}}$ with different channel weights. The mathematical expression is as follows:

$$X_{A_F} = W_c \otimes X_{A_{cat}}. \tag{6}$$

3.4 Spatial feature enhancement block

The goal of the single-image rain removal task is to recover a clear background from an image containing rain streaks; however, mixing rain streaks with the background makes recovering image details and textures difficult. In this study, we utilized the frequency feature fusion mechanism to enhance the feature channels advantageous for rain removal. However, detail and texture preservation in image restoration remain a challenge. Although frequency feature fusion can enhance the necessary features for rain removal, it does not offer enough resolution to effectively preserve image details and textures, mainly when rain streaks are similar or overlap with background details. This can result in blurring or losing details during rain removal.

Using the spatial attention mechanism enhances the model discernment of the significance of various locations within the image by dynamically assigning weights to the spatial attributes. More precisely, this feature allows the model to prioritize the crucial details and textures hidden by the rain streak while reducing the impact of background noise or irrelevant information that is less important for removing the rain. This is achieved by assigning greater importance to the areas of the image that are more relevant to rain removal [52]. This process enhances both the rain removal effect and the preservation of the original texture and details of the image to a certain degree, resulting in a more authentic rain removal effect, as depicted in Fig. 6.

We implement SFEB based on SGE Attention [53]. Specifically, given an input feature $X_{IWT} \in \mathbb{R}^{C\times H\times W}$, the RDB module is first used to extract the multilevel fusion feature $X_{IWT_R} \in \mathbb{R}^{C\times H\times W}$ of the input feature X_{IWT}. The latter is then fed into the spatial attention mechanism SGE to obtain the spatial feature-enhanced $X_{SE} \in \mathbb{R}^{3\times C\times H\times W}$. The mathemat-

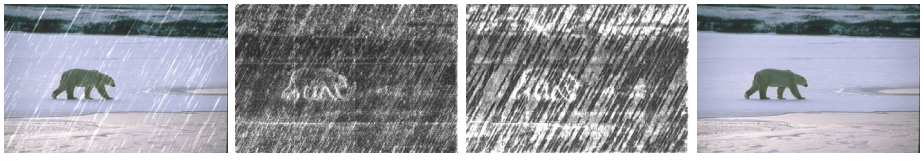


Figure 6 The spatial attention mechanism assigns distinct weights to each position of the input features. The recalibrated image features are obtained by multiplying the spatial features and weights of the input characteristics. As depicted in the diagram, the improved texture characteristics hold more significance compared to the spatial characteristics prior to augmentation

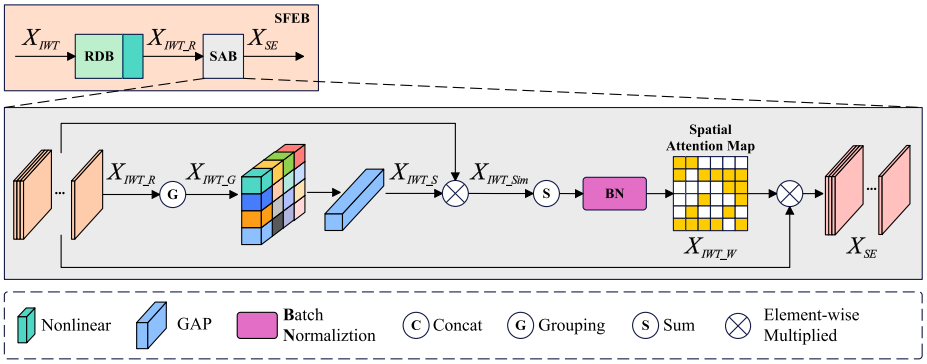


Figure 7 Architecture of the spatial feature enhancement block

ical expression is as follows:

$$\begin{aligned} X_{IWT_R} &= f_{RDB}\left(X_{IWT}\right), \\ X_{SE} &= f_{SGE}\left(X_{IWT_R}\right), \end{aligned} \tag{7}$$

where f_{RDB} denotes the RDB operation, and f_{SGE} denotes the spatial attention module SGE.

Spatial attention block As the rain streaks are mixed with the image background, it is possible for subfeatures that represent the details of the image texture to be distributed in each layer of the image features. Still, spatially, these features are affected by the rain streak noise, resulting in these features needing to be correctly localized and recognized. Unlike other spatial attention mechanisms [54] that use a global spatial model, SGE attention groups image features in channel dimensions. It computes attention weights for features within each group, which allows SGE to produce more detailed spatial weights, alleviating the problem of spatial mixing of rain patterns with the image background, as shown in Fig. 7.

Specifically, given an input feature $X_{IWT_R} \in \mathbb{R}^{C\times H\times W}$, X_{IWT_R} is first grouped on the channel to obtain $X_{IWT_G} \in \mathbb{R}^{G\times(C//G)\times H\times W}$, where G is the number of groups. Then GAP is used on X_{IWT_G} to get the semantic vector $X_{IWT_S} \in \mathbb{R}^{G\times 1\times H\times W}$ for the whole space. Next, the similarity $X_{IWT_{Sim}} \in \mathbb{R}^{G\times(C//G)\times H\times W}$ between X_{IWT_S} and X_{IWT_G} is computed using element multiplication, and $X_{IWT_{Sim}}$ is normalized to obtain the spatial attentional weight $X_{IWT_W} \in \mathbb{R}^{G\times 1\times H\times W}$. Finally, X_{IWT_W} is multiplied by the elements of X_{IWT_R} to get the weighted feature $X_{SE} \in \mathbb{R}^{C\times H\times W}$. The mathematical expression is as follows:

$$\begin{aligned} X_{IWT_G} &= f_{Grouping}\left(X_{IWT_R}\right), \\ X_{IWT_S} &= f_{GAP}\left(X_{IWT_G}\right), \\ X_{IWT_{Sim}} &= X_{IWT_S} \otimes X_{IWT_G}, \\ X_{IWT_W} &= BN\left(Reshape1\left(Sum\left(X_{IWT_{Sim}}\right)\right)\right), \\ X_{SE} &= X_{IWT_G} \otimes Sigmoid\left(Reshape2\left(X_{IWT_W}\right)\right), \end{aligned} \tag{8}$$

where $f_{Grouping}$ denotes the grouping operation, f_{GAP} represents the GAP operation, *Sum* denotes summing the input features by channel, *Reshape*1 indicates changing the input

feature with shape $(G \times 1 \times H \times W)$ to $(G \times H \times W)$, *Sigmoid* denotes the activation function, and *Reshape*2 denotes changing the input feature with shape $(G \times H \times W)$ to $(G \times 1 \times H \times W)$.

3.5 Image reconstruction network and loss function

Within the image reconstruction network, we employ a straightforward convolution operation to carry out the process of remapping image features and restoring the image. The mathematical expression is as stated:

$$R = Conv\left(BN\left(Conv\left(X_{SE}\right)\right)\right). \tag{9}$$

In single-image deraining tasks, L_1 and L_2 losses are usually used to optimize the network, but these loss functions are pixel-level losses. In this study, we want the restored image to be consistent with human evaluation, so we combine the peak signal-to-noise ratio (PSNR) loss with the structural similarity (SSIM) loss and add the edge loss [9]. The loss function we propose consists of the composite loss function $\mathcal{L}_{PS}$, which is based on the PSNR and SSIM loss, and the edge loss function $\mathcal{L}_{edge}$. The $\mathcal{L}_{PSNR}$ in the composite loss function is used to measure the difference between the I_{clr} of the clear image after rain removal and the I_{GT} of the real image. The higher the value, the closer the quality of the derained image is to that of the original image. $\mathcal{L}_{SSIM}$ measures the structural similarity between the derained and original images. The closer the value of SSIM to 1, the more similar the structure of the images. A tiny constant ϵ is used to avoid a zero denominator. The loss function that combines SSIM and PSNR can effectively balance structural similarity and overall image signal differences in the image deraining task. Using this combination, the model retains the visual effect of the derained image and ensures that the recovered image is as similar as possible in structure to the original image. The edge loss function $\mathcal{L}_{edge}$ measures the edge difference between the derained and original images. In the image deraining task, edge information is very important for visual effects, and maintaining edge details can prevent image blurring. Therefore an edge loss term is introduced so that the model can better preserve the edges and details of the image, thereby improving the visual image quality after rain removal. The final loss function $\mathcal{L}\left(I_{clr}, I_{GT}\right)$ combines the $\mathcal{L}_{PS}$ and $\mathcal{L}_{edge}$ loss functions, which can simultaneously optimize the global image structural similarity and local edge details so that the model can remove rain streaks while preserving the natural feel and details of the image. At the same time, this loss function also improves the robustness of the rain removal model so that it can better handle rain streaks of different types and intensities and improve the overall image quality. The mathematical expression is as follows:

$$\begin{aligned}
&\mathcal{L}_{PS}\left(I_{clr}, I_{GT}\right) = \frac{1 - \mathcal{L}_{SSIM}\left(I_{clr}, I_{GT}\right)}{\mathcal{L}_{PSNR}\left(I_{clr}, I_{GT}\right) + \epsilon}, \\
&\mathcal{L}_{edge}\left(I_{clr}, I_{GT}\right) = \sqrt{\left\|\Delta\left(I_{clr}\right) - \Delta\left(I_{GT}\right)\right\|^2 + \omega^2}, \\
&\mathcal{L}\left(I_{clr}, I_{GT}\right) = \mathcal{L}_{PS}\left(I_{clr}, I_{GT}\right) + \lambda \mathcal{L}_{edge}\left(I_{clr}, I_{GT}\right),
\end{aligned} \tag{10}$$

where $\mathcal{L}_{SSIM} \in [-1, 1]$ denotes the SSIM loss, and as the image restoration quality gets higher, $\mathcal{L}_{SSIM}$ gets closer to 1; $\mathcal{L}_{PSNR} \in [0, \infty)$ denotes the PSNR loss, and as the image

restoration quality gets higher, $\mathcal{L}_{PSNR}$ loss gets closer to ∞; Δ denotes the Laplace operation, ϵ is a constant that ensures that the denominator of $\mathcal{L}_{PS}$ does not have a zero, and ω is also a constant.

4 Experiments

This section provides a comprehensive evaluation of the performance of DWTN. First, we describe the data sets and performance metrics used in the comparative experiments. Next, we introduce the experimental setup, which is closely related to the experimental results. Then we quantitatively and qualitatively evaluate the performance of our proposed DWTN model using different data. At the same time, we use ablation experiments to verify the functions of various functional modules in the model. Finally, we also use DWTN as a pretask for the target detection task to show that DWTN can significantly improve the performance of advanced visual tasks.

4.1 Dataset and evaluation metrics

Synthetic datasets Yang et al. [5] collected two rain streak datasets, Rain200H and Rain200L, which are widely used to evaluate single-image rain removal tasks. Rain200H and Rain200L contain 1800 synthetic images for training and 200 pairs for testing, respectively. Rain200H mainly contains dense rain streaks, whereas the Rain200L dataset is relatively sparse compared to Rain200H. Yang et al. [55] collected a rain streak dataset called Rain800. The dataset consists of 800 pairs of rain images and corresponding clean images, of which 700 pairs of data are used for the training set, and the remaining 100 pairs are used for the test set. A larger rain streak dataset named Rain1400 was collected by Fu et al. [22] named Rain1400. Each clean image in this dataset corresponds to 14 rain images of different types. The training set contains 900 clean images, whereas the test set contains 100 clean images.

Real-world datasets To validate the effectiveness of the deraining model in the real world, Li et al. [56] and Wang et al. [57] collected real-world-based rain image datasets MPID and SPA, respectively. The MPID dataset contains 185 rain images, whereas the SPA dataset contains 146.

Evaluation metrics During experiments involving synthetic data, we evaluate the performance of the DWTN using two metrics: the peak signal-to-noise ratio (PSNR) [58] and the structural similarity index (SSIM) [59]. We evaluate the image luminance channel results, relying on previous research findings [60], as the human visual system is highly responsive to image brightness. Furthermore, because of the unavailability of ground truth for real-world rain images, we used the no-reference image quality measures NIQE [61] and BRISQUE [62] to evaluate the effectiveness of DWTN on real-world datasets.

A higher PSNR value indicates better image recovery from rainy images, as expressed mathematically in equation (11). The value of SSIM is in the range of $[-1, 1]$; when the value of SSIM is close to 1, the restored image is closer to the rain-free image, as expressed mathematically in equation (12).

$$PSNR\,(I, G) = 20 \times \log_{10}\left(\frac{MAX_I}{\sqrt{MSE\,(I, G)}}\right), \tag{11}$$

where I represents the derained image, G represents the ground truth (GT) image, $MSE(I, G)$ denotes the mean square error between the derained image and the GT image, and MAX_I denotes the maximum pixel value of the image.

$$SSIM(I, G) = \frac{(2\mu_I\mu_G + C_1)(2\sigma_I\sigma_G + C_2)}{(\mu_I^2 + \mu_G^2 + C_1)(\sigma_I^2\sigma_G^2 + C_2)}, \tag{12}$$

where I represents the image after rain removal, G represents the GT image, μ_I, μ_G, σ_I, and σ_G denote the mean and standard deviation of the input image, and C_1 and C_2 are constants used to prevent division by zero.

4.2 Implementation and training details

We implemented the proposed end-to-end single-image dewatering model DWTN using Pytorch 1.8 and did not use pretrained weights. We trained and inferred the model on a single NVIDIA GTX 3090. During training, we optimized the model using the Adam optimizer [63] with an initial learning rate of $1e^{-3}$ during training and a cosine annealing algorithm [64] to update the learning rate, thereby improving the training efficiency of the model. To adapt to different sizes of input images, we randomly cropped the images into 128×128 patches as the input to the model. In addition, we also use random flipping [65] of the images to increase the diversity of the input data.

By adjusting the number of modules included in the model, DWTN can derive three models with different parameter sizes: DWTN-S, DWTN-M, and DWTN-L. The difference between DWTN-S and DWTN-M is that the FAEB of the DWTN-S model contains two RDB+ReLU modules, whereas the FAEB of DWTN-M contains three RDB+ ReLU modules. DWTN-L consists of two DWTN-M models with the computation of the loss function added in the middle position of the network, which results in image deraining from coarse to fine images, as shown in Fig. 8.

4.3 Comparison with the state-of-the-arts

We compared DTWN-S, DTWN-M and DTWN-L with the state-of-the-art methods: HiNAS [34], HCT-FFN [66], RCDNet [67], DRT [68], MANAS [69], HINet [32], TransWeather [70], Uformer [28], and CMFNet [71], and the code for all the methods was obtained from open-source code provided in the paper. All source code was rerun on the

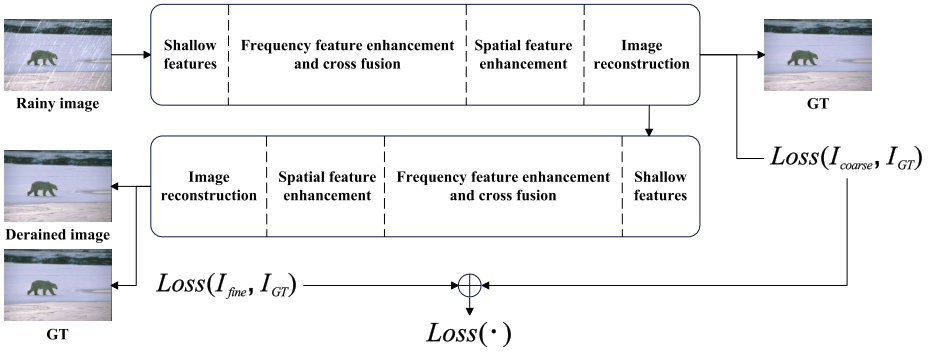


Figure 8 The architecture of DWTN-L. This network connects 2 DWTN-M in series and achieves better performance image deraining model by 2 times Loss computation

benchmark dataset. Table 2 shows the qualitative evaluation of different methods in terms of both PSNR and SSIM. The results demonstrate that our DTWN approach substantially benefits PSNR and SSIM measures. The Rain200L, Rain200H, and Rain800 datasets were used to evaluate the state-of-the-art performance. DTWN achieved the highest ranking in SSIM metrics and the second-highest PSNR metrics on the Rain1400 dataset. The quantitative analysis findings confirm the DTWN model efficiency in removing rain from a single image.

Furthermore, we conducted a comparison of two metrics, parameter size and frames per second (FPS), which have a direct correlation with the lightweight nature of the model. The results demonstrate that our proposed DTWN model exhibits superior performance regarding model inference speed compared to other models with similar parameter sizes. Additionally, DTWN-L outperforms the other comparison models in both inference speed and performance when the model parameters are smaller than those of the different models. Thus by amalgamating various assessment measures the overall efficacy of DWTN surpasses that of the thirteen aforementioned comparative models.

Results on synthetic datasets Fig. 9 shows the qualitative evaluation of thirteen rain removal methods on four samples from the Rain200L dataset. We can see in the figure that using HiNAS to remove the rain streaks leads to an overall distortion of the image colors, and the model has a weak ability to restore the image details, which affects the effectiveness of the image in removing the rain. HCT-FFN removes most of the streaks, but the image edge information is missing, so the recovered image is incomplete. Although RCDNet and CMFNet can remove some of the rain streaks, they produce artifacts when dealing with dense rain streaks, which affects the derain effect of the model. MANAS is a derain model designed based on NAS, which can remove most of the rain streaks but cannot remove the thicker or thinner rain streaks, which makes many rain streaks remain after deraining the image. HiNet and MPRNet have strong derain ability, but some image details are lost. TransWeather and Uformer are multitasking image restoration models, but the single-task performance differs from the performance of state-of-the-art models. CMFNet achieves a more realistic rain streak removal capability, but it cannot remove the thicker rain streaks, which is an unsatisfactory rain removal effect. Compared to these state-of-the-art (SOTA) methods, our proposed DWTN can remove rain streaks with different densities. In addition, the rain removal results of DWTN show more vivid colors and more detailed structural information, which are more advantageous than other rain removal methods.

The results demonstrate that our DWTN can generate rain-free images that are both more realistic and significantly sharper. The color red signifies the highest level of performance, whereas the color blue symbolizes the second position.

Results on real-world dataset The MPID and SPA datasets were evaluated using the no-reference image quality measures NIQE and BRISQUE, both quantitatively and qualitatively. Table 3 displays the evaluation metrics for multiple models on the MPID and SPA datasets. The deraining results of several approaches on three real rainy day samples are depicted in Fig. 10. Similarly to the results in Fig. 9, rain streaks remain in the images after deraining using CMFNet. DRT, HCT-FFN, and HiNAS failed to remove all the rain streaks, and the reconstructed images are missing a lot of details. In real environment deraining, MANAS cannot remove the rain streaks, and the structural information of the

Table 2 The PSNR and SSIM results of different rain removal methods on the four synthetic datasets. Red indicates the top performance, whereas blue represents the second place

Methods	Publication	FPS	Params (MB)	Rain200L		Rain200H		Rain800		Rain1400	
				PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
HiNAS	CVPR' 2020	26.60	0.63	30.88	0.935	26.31	0.869	23.04	0.781	27.07	0.857
HCT-FFN	AAAI' 2023	2.11	0.87	34.52	0.944	24.97	0.867	24.98	0.813	28.23	0.896
RCDNet	TNNLS' 2023	1.40	2.25	29.91	0.917	24.64	0.829	22.36	0.743	20.07	0.774
DWTN-S	-	131.34	5.53	38.29	0.984	29.53	0.871	25.74	0.815	27.56	0.898
DRT	CVPRW' 2022	0.60	4.50	30.26	0.879	24.54	0.740	21.66	0.658	26.44	0.778
MANAS	TCSVT' 2023	2.31	9.35	32.57	0.957	26.24	0.866	24.38	0.845	28.52	0.901
DWTN-M	-	102.47	12.76	38.86	0.984	31.25	0.921	28.32	0.893	29.77	0.911
HINet	CVPRW' 2021	7.12	88.67	37.28	0.970	30.65	0.892	28.01	0.870	33.77	0.939
MPRNet	CVPR' 2021	41.84	20.15	36.40	0.965	30.41	0.890	25.05	0.833	31.14	0.930
TransWeather	CVPR' 2022	10.31	37.93	30.86	0.939	23.35	0.782	23.93	0.790	30.19	0.901
Uformer	CVPR' 2022	78.31	50.88	35.24	0.933	23.93	0.813	23.42	0.781	29.55	0.916
CMFNet	ICIP' 2023	5.1	17.24	37.20	0.981	27.68	0.882	25.22	0.850	30.58	0.913
DWTN-L	-	90.12	26.71	39.70	0.987	31.27	0.917	28.77	0.903	34.04	0.944

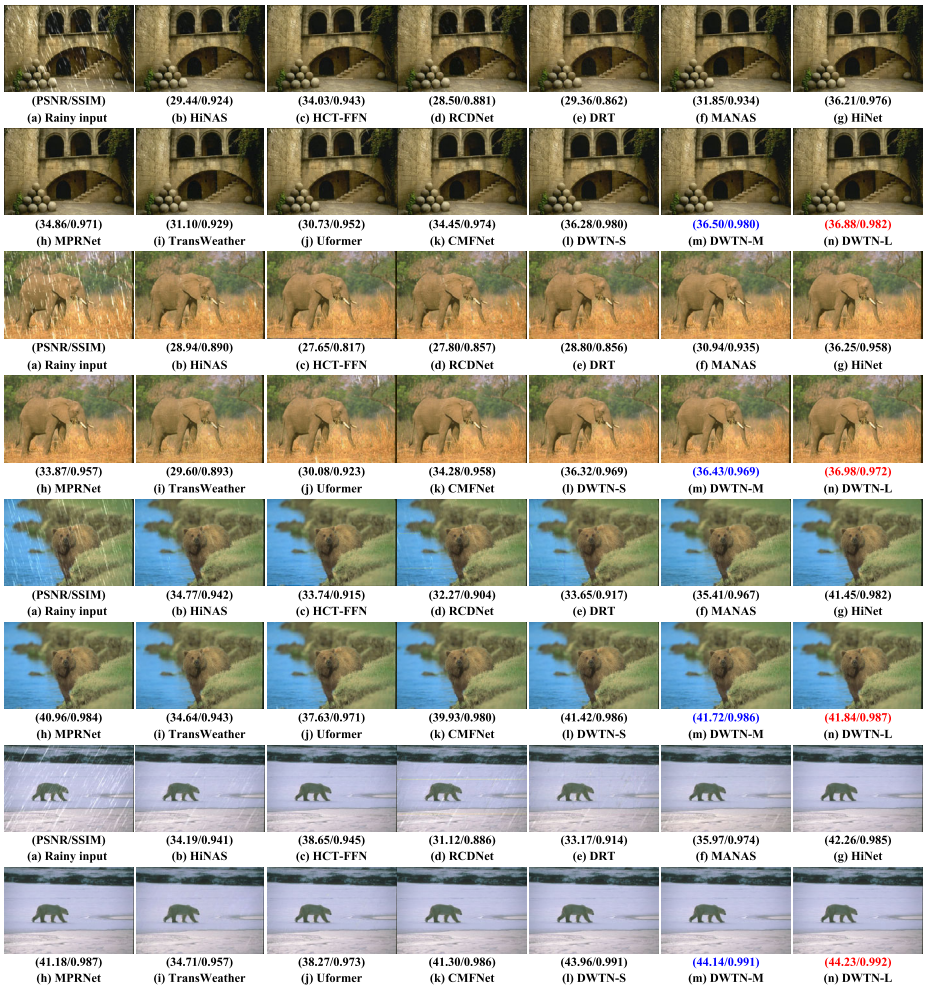


Figure 9 Qualitative comparisons of different rain removal models on the Rain200L dataset: (a) the input rain image, (b) HiNAS, (c) HCT-FFN, (d) RCDNet, (e) DRT, (f) MANAS, (g) HiNet, (h) MPRNet, (i) TransWeather, (j) Uformer, (k) CMFNet, (l) the DWTN-S, (m) the DWTN-M, and (j) DWTN-L are images after deraining. The results demonstrate that our DWTN can generate rain-free images that are both more realistic and significantly sharper. The color red signifies the highest level of performance, whereas the color blue symbolizes the second position

image is missing. RCDNet has made an error in processing the image texture. Compared with DWTN-S and DWTN-M, DWTN-L has the strongest deraining ability, which removes the rain streaks and preserves a lot of detailed information in the image reconstruction.

4.4 Ablation study

To confirm the efficiency of DWTN, we conducted ablation experiments on DWTN-S utilizing PSNR and SSIM metrics on the Rain200L dataset to assess the usefulness of various modules. The benchmark models were derived by excluding the FAEB, FFFB, and SFEB components. The training technique employed by the benchmark model is identical to that of the DWTN-S model and utilizes the loss function depicted in Eq. (10). The outcomes of the ablation trials are presented in Table 4 and Fig. 11.

Table 3 This study evaluates the NIQE and BRISQUE results of multiple rain removal techniques on two real datasets. The color red represents the highest level of performance, whereas the color blue signifies the second-highest level

	MPID		SPA	
Methods	NIQE	BRISQUE	NIQE	BRISQUE
HiNAS	3.381	27.572	4.647	26.620
HCT-FFN	3.860	91.208	4.801	59.057
RCDNet	3.300	25.379	4.693	24.753
DWTN-S	3.157	24.501	4.595	25.864
DRT	3.285	24.654	4.385	24.768
MANAS	3.111	24.803	4.394	26.128
DWTN-M	3.101	24.485	4.253	25.156
HINet	3.478	28.026	7.887	28.564
MPRNet	3.728	26.392	7.845	26.662
TransWeather	3.746	24.469	6.716	24.625
Uformer	3.447	27.121	7.443	28.121
CMFNet	3.197	25.054	4.586	25.940
DWTN-L	3.078	24.452	4.173	24.194

Figure 11 is drawn using the MulimgViewer tool [72] and illustrates the contribution of each component in the DWTN to the rain streak removal. The FAEB module mainly focuses on the rain streak feature extraction, so both the structural and detailed information of the recovered image is missing when the FAEB does not use the RDB, as shown in Fig. 11(c). The FFFB module mainly extracts different types of rain streaks through the reasonable feature fusion mechanism and the channel attention mechanism. Extracting different types of streaks makes the structural information of the recovered image better preserved, as shown in Fig. 11(d). The SFEB module, on the other hand, mainly uses the spatial attention mechanism to extract more texture information of the image, so that the detailed information of the recovered image can be better restored, as shown in Fig. 11(e). In summary, DWTN gives full play to the potential of each module, which ultimately has a significant advantage over the previous SOTA single-image derain method.

4.5 Application

To demonstrate that our DWTN can improve the performance of high-level vision applications, we first performed the deraining operation using the deraining model on the rain images in the RID dataset. Then we used YOLOv5 to perform object detection on 100 sets of images in the RID dataset, and the detection results are shown in Fig. 12. As we can see from the figure, the confidence and accuracy of object detection after deraining is significantly improved compared to that before the deraining operation.

5 Conclusions

This paper presents the development of a rain removal model named DWTN, designed explicitly for vision measuring systems using a single image approach. DWTN employs wavelet transforms to break down the image features into four distinct components with varying characteristics. Each feature component is then individually processed using FAEB while simultaneously extracting the beneficial features for removing rain using the dense residual module. Afterward, we combined the various feature components using the FFFB module based on the specified rules. We employed the channel attention method to enhance the image background characteristics and reduce the prominence of the foreground streaks. Subsequently, the feature components are obtained by applying the in-

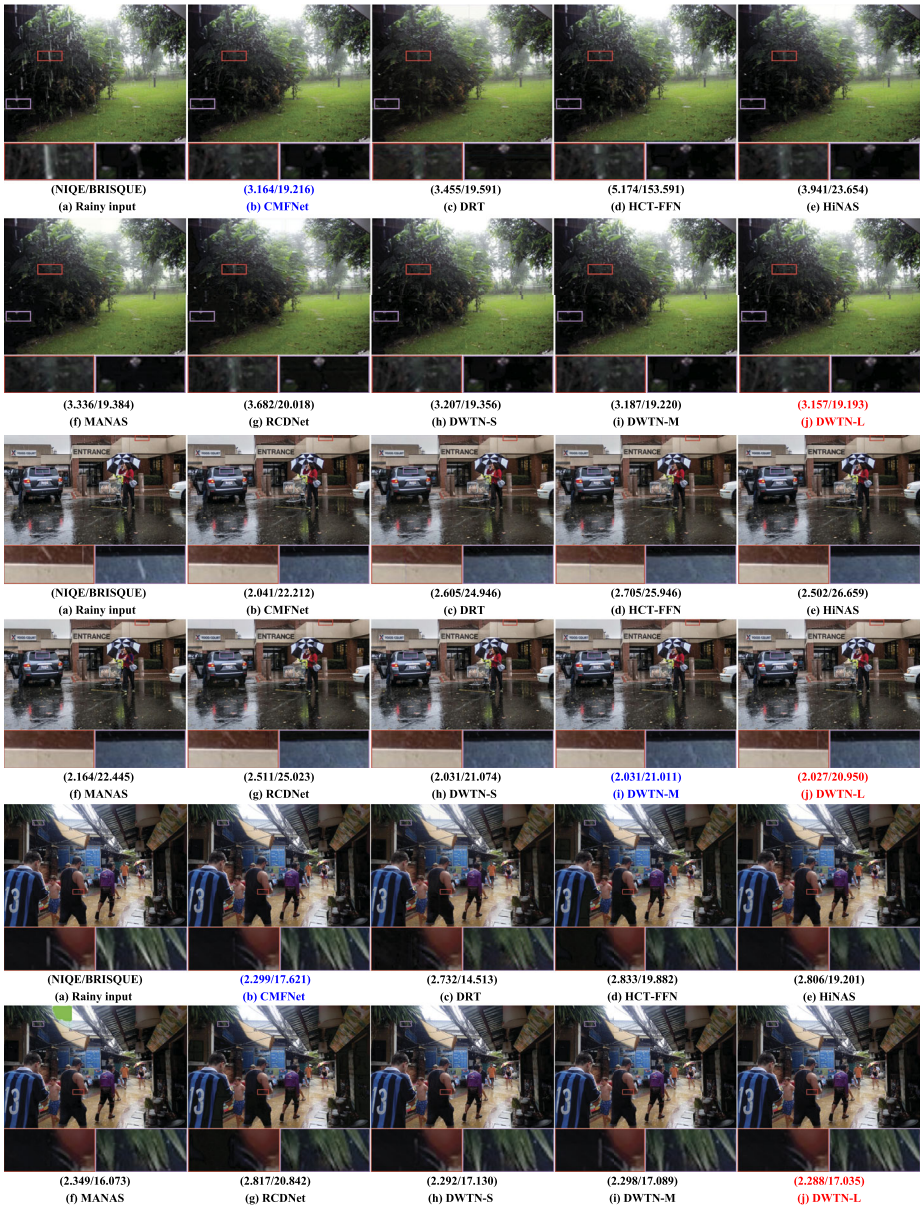


Figure 10 Performing qualitative comparisons on real-world images captured during rainy conditions. It has been noted that DWTN can provide visually appealing images with realistic colors and fewer artifacts caused by rain. The color red signifies the highest level of performance, whereas the color blue symbolizes the second position

Table 4 Ablation experiments with local search networks on the synthetic dataset Rain200L

Variants	FAEB	FFFB	SFEB	PSNR	SSIM
V_0	w/o	w/o	w/o	37.14	0.980
V_1	✓	w/o	w/o	38.04	0.983
V_2	✓	✓	w/o	38.28	0.984
V_3	✓	✓	✓	38.29	0.984

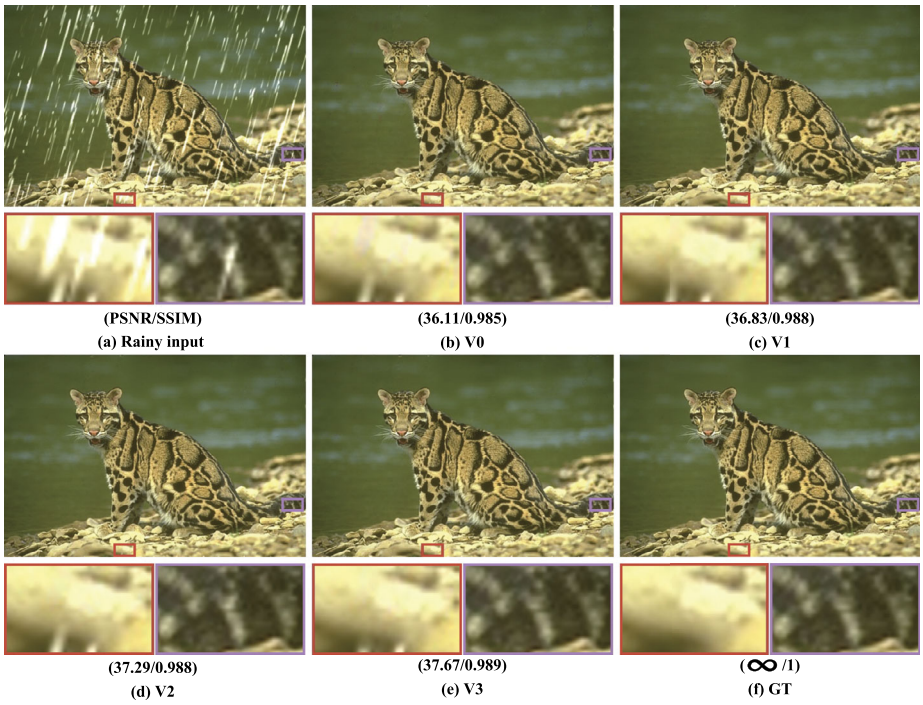
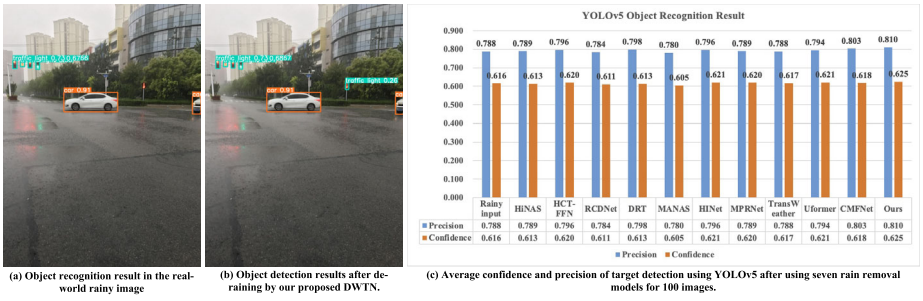


Figure 11 Results of the visual ablation experiments. (a) Rainy image, (b) result of the V_0 network, (c) result of the V_1 network, (d) result of the V_2 network, (e) result of the V_3 network, (f) clear image



	Rainy input	HiNAS	HCT-FFN	RCDNet	DRT	MANAS	HINet	MPRNet	TransWeather	Uformer	CMFNet	Ours
Precision	0.788	0.789	0.796	0.784	0.798	0.780	0.796	0.789	0.788	0.794	0.803	0.810
Confidence	0.616	0.613	0.620	0.611	0.613	0.605	0.621	0.620	0.617	0.621	0.618	0.625

Figure 12 Results from visual ablation experiments. The study investigates three primary components: (a) The result of object recognition in a real-world image captured during rainy weather; (b) the object detection results after applying our proposed DWTN to remove rain from the image; (c) the mean confidence and accuracy of target detection using YOLOv5 after employing seven different rain removal models on a dataset that includes 100 images

verse wavelet transform to the image, and the texture details of the image are enhanced using SFEB. A straightforward image reconstruction network is employed to restore the image affected by rain. This work involves a comprehensive experimental assessment of the performance of DWTN in comparison to thirteen advanced rain removal models. The evaluation is performed on four artificially generated datasets and two datasets derived from real-world scenarios. The testing results thoroughly showcase the cutting-edge capabilities of DWTN.

Acknowledgements
The authors thank the anonymous reviewers for their careful reading and valuable comments.

Author contributions
WT wrote the main manuscript text; WT and YW prepared the result of our experiments. All authors read and approved the final manuscript.

Funding
This work was supported by the Basic Research for National Defense under Grant No. JCKY2020605C003, the Innovation Platform Foundation of SIP under Grant Nos. YZCXPT2022107, YZCXPT2022106, and YZCXPT2023103, the Science Foundation of SISO under Grant No. KY-ZD202102, the Science Foundation of Jiangsu Provincial Department of Education under Grant No. 21KJB520037, Jiangsu Higher Education Teaching Reform Key Research Project Grant No. 2021JSJG194, and the Team Foundation of Jiangsu Provincial Department of Education under Grant No. 2022.13.

Data availability
The Rain200L and Rain200H dataset is available at https://github.com/hezhangsprinter/DID-MDN. The Rain800 dataset is available at https://github.com/hezhangsprinter/ID-CGAN. The Rain1400 dataset is available at https://xueyangfu.github.io/projects/cvpr2017.html. The SPA dataset is available at https://github.com/stevewongv/SPANet. The MPID dataset is available at https://github.com/panda-lab/Single-Image-Deraining.

Declarations

Ethics approval and consent to participate
The Rain200L, Rain200H, Rain800, Rain1400, MPID, and SPA dataset all are open-source datasets and are only used for noncommercial research purposes.

Competing interests
The authors declare no competing interests.

Author details
[1]College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, Jiangsu, China. [2]Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, Jiangsu, China. [3]College of Computer Science and Technology, Anhui University of Technology, Ma'anshan, 243032, Anhui, China.

Received: 18 August 2024 Accepted: 19 September 2024 Published online: 30 September 2024

References

1. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: transformers for image recognition at scale. In: 9th International Conference on Learning Representations (2021)
2. Wang, C., Bochkovskiy, A., Liao, H.M.: Yolov7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7464–7475 (2023)
3. Yang, W., Tan, R.T., Wang, S., Fang, Y., Liu, J.: Single image deraining: from model-based to data-driven and beyond. IEEE Trans. Pattern Anal. Mach. Intell. **43**(11), 4059–4077 (2021)
4. Wang, K., Wang, T., Qu, J., Jiang, H., Li, Q., Chang, L.: An end-to-end cascaded image deraining and object detection neural network. IEEE Robot. Autom. Lett. **7**(4), 9541–9548 (2022)
5. Yang, W., Tan, R.T., Feng, J., Liu, J., Guo, Z., Yan, S.: Deep joint rain detection and removal from a single image. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1685–1694. IEEE Comput. Soc., Honolulu (2017)
6. Li, X., Wu, J., Lin, Z., Liu, H., Zha, H.: Recurrent squeeze-and-excitation context aggregation net for single image deraining. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) Computer Vision - ECCV 2018 - 15th European Conference. Lecture Notes in Computer Science, vol. 11211, pp. 262–277. Springer, Munich (2018)
7. Zhao, J., Xie, J., Xiong, R., Ma, S., Huang, T., Gao, W.: Pyramid convolutional network for single image deraining. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 9–16. Computer Vision Foundation/IEEE, Long Beach (2019)
8. Wen, Y., Gao, T., Zhang, J., Zhang, K., Chen, T.: From heavy rain removal to detail restoration: a faster and better network. Pattern Recognit. **148**, 110205 (2024)
9. 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: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8343–8352. Computer Vision Foundation/IEEE, Seattle (2020)
10. Xiao, J., Fu, X., Liu, A., Wu, F., Zha, Z.: Image de-raining transformer. IEEE Trans. Pattern Anal. Mach. Intell. **45**(11), 12978–12995 (2023)
11. Wang, Y., Ma, C., Zeng, B.: Multi-decoding deraining network and quasi-sparsity based training. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 13375–13384. Computer Vision Foundation/IEEE, virtual (2021)
12. Tao, W., Yan, X., Wang, Y., Wei, M.: Mffdnet: single image deraining via dual-channel mixed feature fusion. IEEE Trans. Instrum. Meas. **73**, 1–13 (2024)
13. Portilla, J., Strela, V., Wainwright, M.J., Simoncelli, E.P.: Image denoising using scale mixtures of Gaussians in the wavelet domain. IEEE Trans. Image Process. **12**(11), 1338–1351 (2003)
14. Oyallon, E., Belilovsky, E., Zagoruyko, S.: Scaling the scattering transform: deep hybrid networks. In: IEEE International Conference on Computer Vision, pp. 5619–5628. IEEE Comput. Soc., Venice (2017)

15. Tian, C., Zheng, M., Zuo, W., Zhang, B., Zhang, Y., Zhang, D.: Multi-stage image denoising with the wavelet transform. Pattern Recognit. **134**, 109050 (2023)
16. Jiang, K., Liu, W., Wang, Z., Zhong, X., Jiang, J., Lin, C.: DAWN: direction-aware attention wavelet network for image deraining. In: Proceedings of the 31st ACM International Conference on Multimedia, pp. 7065–7074. ACM, Ottawa (2023)
17. Liu, P., Zhang, H., Lian, W., Zuo, W.: Multi-level wavelet convolutional neural networks. IEEE Access **7**, 74973–74985 (2019)
18. Hsu, W., Chang, W.: Wavelet approximation-aware residual network for single image deraining. IEEE Trans. Pattern Anal. Mach. Intell. **45**(12), 15979–15995 (2023)
19. Wang, Y., Yan, X., Wang, F.L., Xie, H., Yang, W., Wei, M., Qin, J.: UCL-Dehaze: Towards real-world image dehazing via unsupervised contrastive learning. CoRR (2022). arXiv:2205.01871
20. Shen, Y., Wei, M., Wang, Y., Fu, X., Qin, J.: Rethinking Real-world Image Deraining via an Unpaired Degradation-Conditioned Diffusion Model (2024)
21. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778. IEEE Comput. Soc., Las Vegas (2016)
22. Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., Paisley, J.W.: Removing rain from single images via a deep detail network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1715–1723. IEEE Comput. Soc., Honolulu (2017)
23. Ren, D., Zuo, W., Hu, Q., Zhu, P., Meng, D.: Progressive image deraining networks: a better and simpler baseline. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3937–3946. Computer Vision Foundation/IEEE, Long Beach (2019)
24. Fu, X., Liang, B., Huang, Y., Ding, X., Paisley, J.W.: Lightweight pyramid networks for image deraining. IEEE Trans. Neural Netw. Learn. Syst. **31**(6), 1794–1807 (2020)
25. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Burstein, J., Doran, C., Solorio, T. (eds.) Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 4171–4186. Assoc. Comput. Linguistics, Minneapolis (2019)
26. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: Guyon, I., Luxburg, U., Bengio, S., Wallach, H.M., Fergus, R., Vishwanathan, S.V.N., Garnett, R. (eds.) Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems, pp. 5998–6008 (2017)
27. 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: IEEE Conference on Computer Vision and Pattern Recognition, pp. 12299–12310. Computer Vision Foundation/IEEE, virtual (2021)
28. Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., Li, H.: Uformer: a general U-shaped transformer for image restoration. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 17662–17672. IEEE, New Orleans (2022)
29. Liang, J., Cao, J., Sun, G., Zhang, K., Gool, L.V., Timofte, R.: SwinIR: image restoration using swin transformer. In: IEEE/CVF International Conference on Computer Vision Workshops, pp. 1833–1844. IEEE, Montreal (2021)
30. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.: Restormer: efficient transformer for high-resolution image restoration. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5718–5729. IEEE, New Orleans (2022)
31. Tan, F., Kong, Y., Fan, Y., Liu, F., Zhou, D., Zhang, H., Chen, L., Gao, L., Qian, Y.: SDNet: mutil-branch for single image deraining using swin. CoRR (2021). arXiv:2105.15077
32. Chen, L., Lu, X., Zhang, J., Chu, X., Chen, C.: HINet: half instance normalization network for image restoration. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 182–192 (2021)
33. Lee, B., Ko, K., Hong, J., Ko, H.: Single cell training on architecture search for image denoising. CoRR (2022). arXiv:2212.06368
34. Zhang, H., Li, Y., Chen, H., Shen, C.: Memory-efficient hierarchical neural architecture search for image denoising. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3654–3663. Computer Vision Foundation/IEEE, Seattle (2020)
35. Quan, R., Yu, X., Liang, Y., Yang, Y.: Removing raindrops and rain streaks in one go. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, Virtual, June 19-25, 2021, pp. 9147–9156. Computer Vision Foundation/IEEE, virtual (2021)
36. Li, Q., Shen, L., Guo, S., Lai, Z.: Wavelet integrated CNNs for noise-robust image classification. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7243–7252. Computer Vision Foundation/IEEE, Seattle (2020)
37. Liu, Y., Li, Q., Sun, Z.: Attribute-aware face aging with wavelet-based generative adversarial networks. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 11877–11886. Computer Vision Foundation/IEEE, Long Beach (2019)
38. Ding, H., Fu, G., Yan, Q., Jiang, C., Cao, T., Li, W., Hu, S., Xiao, C.: Deep attentive style transfer for images with wavelet decomposition. Inf. Sci. **587**, 63–81 (2022)
39. Yu, Y., She, K., Liu, J., Cai, X., Shi, K., Kwon, O.: A super-resolution network for medical imaging via transformation analysis of wavelet multi-resolution. Neural Netw. **166**, 162–173 (2023)
40. Hsung, T., Lun, D.P., Siu, W.: A deblocking technique for block-transform compressed image using wavelet transform modulus maxima. IEEE Trans. Image Process. **7**(10), 1488–1496 (1998)
41. Cotter, F.: Uses of complex wavelets in deep convolutional neural networks. PhD thesis, University of Cambridge, UK (2019)
42. Huang, Y., Huang, J., Liu, J., Yan, M., Dong, Y., Lv, J., Chen, C., Chen, S.: Wavedm: wavelet-based diffusion models for image restoration. IEEE Trans. Multimed. **26**, 7058–7073 (2024)
43. Liu, P., Zhang, H., Zhang, K., Lin, L., Zuo, W.: Multi-level wavelet-CNN for image restoration. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 773–782. Computer Vision Foundation/IEEE Comput. Soc., Salt Lake City (2018)

44. Demirel, H., Anbarjafari, G.: IMAGE resolution enhancement by using discrete and stationary wavelet decomposition. IEEE Trans. Image Process. **20**(5), 1458–1460 (2011)
45. Fu, X., Huang, J., Ding, X., Liao, Y., Paisley, J.W.: Clearing the skies: a deep network architecture for single-image rain removal. IEEE Trans. Image Process. **26**(6), 2944–2956 (2017)
46. Kang, L., Lin, C., Fu, Y.: Automatic single-image-based rain streaks removal via image decomposition. IEEE Trans. Image Process. **21**(4), 1742–1755 (2012)
47. Huang, H., Yu, A., Chai, Z., He, R., Tan, T.: Selective wavelet attention learning for single image deraining. Int. J. Comput. Vis. **129**(4), 1282–1300 (2021)
48. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: Computer Vision - ECCV - 15th European Conference. Lecture Notes in Computer Science, vol. 11211, pp. 294–310 (2018)
49. Huang, G., Liu, Z., Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, July 21-26, pp. 2261–2269. IEEE Comput. Soc., Honolulu (2017)
50. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, pp. 2472–2481. Computer Vision Foundation/IEEE Comput. Soc., Salt Lake City (2018)
51. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: Eca-net: efficient channel attention for deep convolutional neural networks. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, pp. 11531–11539. Computer Vision Foundation/IEEE, Seattle (2020)
52. Wu, L., Wang, Y., Li, X., Gao, J.: Deep attention-based spatially recursive networks for fine-grained visual recognition. IEEE Trans. Cybern. **49**(5), 1791–1802 (2019)
53. Li, X., Hu, X., Yang, J.: Spatial group-wise enhance: Improving semantic feature learning in convolutional networks. CoRR (2019). arXiv:1905.09646
54. Woo, S., Park, J., Lee, J., Kweon, I.S.: CBAM: convolutional block attention module. In: Computer Vision - ECCV 2018, vol. 11211, pp. 3–19. Springer, Munich (2018)
55. Zhang, H., Sindagi, V., Patel, V.M.: Image de-raining using a conditional generative adversarial network. IEEE Trans. Circuits Syst. Video Technol. **30**(11), 3943–3956 (2020)
56. Li, S., Araujo, I.B., Ren, W., Wang, Z., Tokuda, E.K., Junior, R.H., Junior, R.M.C., Zhang, J., Guo, X., Cao, X.: Single image deraining: a comprehensive benchmark analysis. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3838–3847. Computer Vision Foundation/IEEE, Long Beach (2019)
57. 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: IEEE Conference on Computer Vision and Pattern Recognition, pp. 12270–12279. Computer Vision Foundation/IEEE, Long Beach (2019)
58. Huynh-Thu, Q., Ghanbari, M.: Scope of validity of PSNR in image/video quality assessment. Electron. Lett. **44**(13), 800–801 (2008)
59. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
60. Yang, W., Liu, J., Yang, S., Guo, Z.: Scale-free single image deraining via visibility-enhanced recurrent wavelet learning. IEEE Trans. Image Process. **28**(6), 2948–2961 (2019)
61. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Signal Process. Lett. **20**(3), 209–212 (2013)
62. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. **21**(12), 4695–4708 (2012)
63. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: Bengio, Y., LeCun, Y. (eds.) 3rd International Conference on Learning Representations, ICLR 2015 (2015)
64. Loshchilov, I., Hutter, F.: SGDR: stochastic gradient descent with warm restarts. In: 5th International Conference on Learning Representations, ICLR 2017. OpenReview.net, Toulon (2017)
65. Quan, Y., Deng, S., Chen, Y., Ji, H.: Deep learning for seeing through window with raindrops. In: 2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27–November 2, 2019, pp. 2463–2471. IEEE, Seoul (2019)
66. Chen, X., Pan, J., Lu, J., Fan, Z., Li, H.: Hybrid CNN-transformer feature fusion for single image deraining. In: Williams, B., Chen, Y., Neville, J. (eds.) Thirty-Seventh AAAI Conference on Artificial Intelligence, pp. 378–386. AAAI Press, Washington (2023)
67. Wang, H., Xie, Q., Zhao, Q., Li, Y., Liang, Y., Zheng, Y., Meng, D.: Rcdnet: an interpretable rain convolutional dictionary network for single image deraining. IEEE Trans. Neural Netw. Learn. Syst. (2023)
68. Liang, Y., Anwar, S., Liu, Y.: DRT: a lightweight single image deraining recursive transformer. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2022, June 19-20, pp. 588–597. IEEE, New Orleans (2022)
69. Cai, L., Fu, Y., Huo, W., Xiang, Y., Zhu, T., Zhang, Y., Zeng, H.: Multi-scale attentive image de-raining networks via neural architecture search. IEEE Trans. Circuits Syst. Video Technol. **33**(2), 618–633 (2022)
70. Valanarasu, J.M.J., Yasarla, R., Patel, V.M.: Transweather: transformer-based restoration of images degraded by adverse weather conditions. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2343–2353 (2022)
71. Fan, C., Liu, T., Liu, K.: Compound multi-branch feature fusion for image deraindrop. In: IEEE International Conference on Image Processing, ICIP 2023, October 8-11, pp. 3399–3403. IEEE, Kuala Lumpur (2023)
72. Liu, J.: MulimgViewer: A Multi-image Viewer for Image Comparison and Image Stitching. https://github.com/nachifur/MulimgViewer

Publisher's Note