Computers & Graphics (2023)

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Image Deraining Based on Dual-Channel Component Decomposition

Xiao Lin[a,b,c], Duojiu Xu[a,b], Peiwen Tan[a,b], Lizhuang Ma[d], Zhi-Jie Wang[e]

[a]*Shanghai Engineering Research Center of Intelligent Education and Bigdata, Shanghai Normal University, Shanghai 200234,China*
[b]*The College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 200234, China*
[c]*The Research Base of Online Education for Shanghai Middle and Primary Schools, Shanghai 200234, China*
[d]*School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*
[e]*The College of Computer Science and the Ministry of Education Key Laboratory of Dependable Service Computing in Cyber Physical Society, Chongqing University, Chongqing 400044, China.*

ARTICLE INFO

Article history:
Received xxx

Keywords: Deraining, Transformer, Component decomposition, Background detail recovery

ABSTRACT

Image deraining aims to remove rain streaks from images and reduce information loss in outdoor images caused by rain. As a fundamental task in image processing, image deraining not only enhances the visibility of images but also provides necessary image restoration for advanced vision tasks. Existing image deraining models mostly train end-to-end models by minimizing the similarity between the output image of the model and the rain-free ground truth. Although these methods have achieved significant results, they often perform poorly in the face of dense and changing rain streak scenes. In this paper, we propose a novel method, called **D**ual-Channel **C**omponent **D**ecomposition **Net**work (DCD-Net). The basic idea of DCD-Net is to leverage the separability prior of rainy images, treats the rain-free background layer and the rain streak mask layer as two parallel component extraction tasks. To this end, it builds a dual-branch parallel networks that extract the rain-free background image and decouple the reconstruction information of the rain streak mask, respectively. It finally applies a composite multi-level contrastive supervision to the output of the above dual-branch parallel network, thereby achieving rain streak removal. Extensive experiments on various datasets demonstrate that the proposed model outperforms existing methods in deraining dense rain streak images.

1. Introduction

Rainy weather photography inevitably leads to visual quality degradation, image distortion and other issues [1, 2, 3]. This degradation has a negative impact on advanced tasks, including outdoor monitoring, autonomous driving, object detection, and so on [4, 5, 6, 7, 8]. Therefore, it is important to remove rain streaks from rainy images without losing the original image information. Usually this task is called *image deraining*. Many existing methods (e.g., model-driven and/or based on the statistical characteristics of rain streaks/ background scenes) can handle light rain well. In the case of heavy rain, they, however, often blur the background scene. Consider the example shown in Fig. 1, dense rain streaks interfere with the model's judgment of the background details of the rainy image, while the model lacks the acquisition of rain streak information; thereby rain streaks and image details are usually deleted at the same time, obtaining poor deraining result, see e.g., Figs. 1(b~d).

In the past few decades, many researchers have been dedicated to solving the rain image restoration problem (i.e., image deraining). The proposed methods can be roughly divided into two categories. The One is based on image priors, treating the rain removal problem as a decomposition problem between the rain layer and the background layer. This line of works include morphological component analysis [9], non-local mean filtering [10], sparse coding [11, 12], and sparsity and low

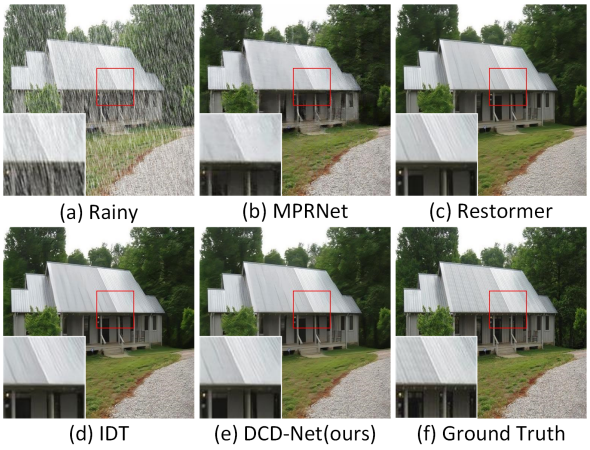

Fig. 1. The results of image deraining between our method and other state-of-the-art deraining methods. The derained images obtained by our method have higher preservation of details and textures.

rank methods [13, 14, 15]. The other involves the construction of end-to-end rain removal networks by integrating deep neural networks. This line of methods extract background details through stronger feature analysis. In recent years, many network architectures with excellent rain removal performance have been proposed [16, 17, 18, 19]. However, the above-mentioned methods face the following challenges: 1) Model constructions based on image priors often rely on well-defined mathematical models, which limits their generalization ability. 2) Models constructed based on deep learning gradually attempt to achieve rain removal advantages through more complex end-to-end models or training on larger rain image datasets. Consequently, they require significant computational power and data requirements. Moreover, due to the ill-posed nature of end-to-end model training, image details can still be lost. Therefore, it is of great significance to study how to effectively extract background detail features from rainy images and make full use of image priors to reduce detail loss.

To this end, we propose a component decomposition-based image deraining method, called the Dual-Channel Component Decomposition Network (DCD-Net). The proposed network adopts a multi-task modeling approach. It builds a dual-branch parallel network, and combines Swin-Transformer [1] and UNet [3] to construct a background component decomposition network (BCDN), in order to extract the rain-free background image. Meanwhile, it combines an attention module and UNet to construct a mask component decomposition network (MCDN), in order to decouple the reconstruction information of the rain streak mask (which are from the multi-scale feature maps). Furthermore, to enable the model to fully understand the semantic information of the rain mask (while learning the mapping relationship from rainy images to rain-free images), we employ a composite multi-level contrastive supervision on the model output. Briefly speaking, we use ground truth rain-free images to constrain the model's extraction of background images, and use real rain mask to constrain the model's extraction of rain mask. These two parallel branches of the network share kernels to enable information interaction, in which the proposed method considers the commonality between different components contained in rainy images as well as the characteristics of each component.

In summary, the main contributions of this work are as follows:

- We propose a new method called Dual-Channel Component Decomposition Network (DCD-Net) for image deraining.
- We conduct extensive experiments based on both real and synthetic datasets to demonstrate the competitiveness of our proposed model.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 presents the prelimianry. Section 4 covers our proposed method in details. Section 5 presents experimental settings and discusses experimental results. Finally, Section 6 concludes this paper.

2. Related Work

Existing image deraining methods can be generally categorized into two types: traditional methods and deep learning-based methods. In what follows, we first review traditional methods (Section 2.1), and then review deep learning-based methods (Section 2.2).

2.1. Tranditional Methods

Most traditional methods decompose the rain removal task into the separation of rain layer and background layer. For example, Kang et al. [9] utilized bilateral filtering to decompose the input image into low-frequency and high-frequency components and then extracted the rain-free component from the high-frequency component. Luo et al. [12] proposed sparse coding, which could better separate rain streaks and background layers from rainy images. Our approach also considers the rain removal task as a separation problem between the rain layer and the background layer. However, we enhance the extraction of detailed features for the background image by combining Swin-Transformer and CNN. This integration provides a more robust capability to capture fine-grained details in the background layer.

Additionally, researchers in this field developed rain removal methods based on prior knowledge of rain and background. For example, Li et al. [20] proposed a Gaussian Mixture Model (GMM) as a prior method to decompose the input image into rain streaks and background layers. Zhu et al. [21] first detected image regions dominated by rainwater and then used the detected regions as a guided image. Zhang et al. [15] utilized the low-order characteristics of rain streaks to separate rain streaks and background layers. Our approach also inherits this merit, i.e., utilizing the separability prior of rainy images. Besides, our approach incorporates deep neural networks to enhance the decomposition capability.

2.2. Deep Learning-based Methods

Fu et al. [22] pioneered the development of a deep convolutional neural network for image deraining. Gradually, deep learning-based approaches have dominated the research in the field. Our method basically belongs to this branch.

In recent years, researchers have proposed many advanced networks. For instance, Li et al. [23] introduced a multi-stage network based on a recurrent neural network architecture to capture rain streak informatio. Zhang et al. [24] proposed a density-aware multi-flow fusion network for rain removal. Our method also employs deep neural networks, but it is different from theirs. In brief, our method integrates the rain streak extraction task into the rain removal task, employing parallel deep neural networks to handle each task.

Additionally, practical priors have been incorporated into deep learning-based methods. For example, Yang et al. [25] used a deep recurrent network to decompose the rain layer into different layers corresponding to different types of rain streaks for effective removal. Liu et al. [26] introduced a dual-residual connection network that leverages the advantages of paired operations to remove rain streaks. Our method combines both the separability prior of rainy images and deep neural networks to enhance the rain removal capability of the model. By leveraging the complementary advantages of these two components, our approach improves the effectiveness of rain removal.

Some researchers also considered the integration of adversarial learning into rain removal tasks. For example, Zhang et al. [27] proposed a conditional generative adversarial network (GAN) for rain streak removal and combined it with perceptual loss for supervision optimization. Wang et al. [28] introduced an encoder-decoder architecture with a conditional generator to enhance rain removal performance. Our method also employs the encoder-decoder architecture, but it builds a dual-branch parallel network, which is not covered in this line of works.

To handle real-world rainy images, Wang [29] constructed a real rain image dataset and designed SPANet using spatial attention mechanisms. Wei et al. [30] utilized semi-supervised training by incorporating real rainy images to represent the residual between input rainy images and their expected rain-free results as Gaussian mixtures. Although our method does not currently incorporate semi-supervised training with real images, it effectively learns the characteristics of rain streaks by utilizing a synthetic image dataset. As a result, our approach demonstrates robustness in rain removal, even when applied to real rainy images, as shown later experiemnts.

Leveraging the significant success of Transformers [31] in natural language processing (NLP) and advanced visual tasks, more researchers have attempted to integrate Transformers into the field of image rain removal. By leveraging the advantages of Transformers in capturing global dependencies, superior performance has been achieved, compared to previous CNN-based methods. Yi et al. [32] proposed a rain removal network based on a Wavelet-based Multi-Level Module, incorporating a residual channel prior (RCP) guidance mechanism to preserve more background details. Xiao et al. [33] proposed an image rain removal network with a dual-transformer architecture based on window-based and spatial-based Transformers, achieving excellent rain removal results. Our approach combines Swin-Transformer and UNet to enhance the model's advantages in extracting both long-term and short-term dependencies as well as fine-grained details. By integrating these two architectures, we aim to reduce the loss of details, thereby preserving important information in the output.

3. Preliminary

To better understand the proposed method, we first introduce some preliminaries related to Transformer [34] and its variant called Swin-Transformer [1].

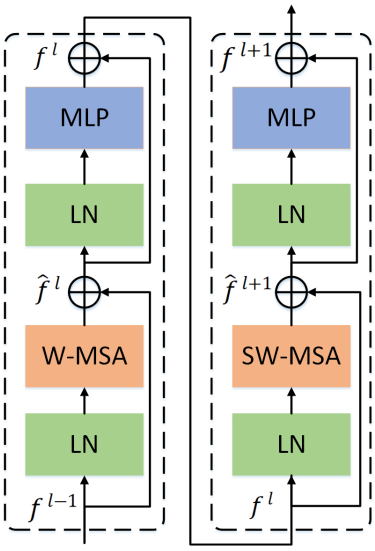

Fig. 2. Swin-Transformer Block (×2) [1].

In brief, Transformer is a deep learning model that uses self-attention mechanism to process sequential data, such as natural language text or time series data. It is notable for its ability contextual relationships and requiring less training time compared to older models like LSTM (Long Short Term Memory). Transformers are widely used in various applications, particularly in natural language processing, computer vision, and so on.

As mentioned earlier, Swin-Transformer [1] is a variant of Transormer. Generally speaking, it utilizes a sliding window mechanism to facilitate the learning of cross-window information. It also introduces a downsampling mechanism, allowing the model to be trained on high-resolution images, while significantly reducing computational costs. Unlike the multi-head self-attention (MSA) module used in Vision Transformer [31], the Swin-Transformer block is built based on a shifted window. The Swin-Transformer Block (×2) is illustrated in Fig. 2 and consists of LayerNorm (LN) layer, multi-head self-attention module, residual connections, and a 2-layer MLP with GELU non-linearity. The window-based multi-head self-attention (W-MSA) module and the shifted window-based multi-head self-attention (SW-MSA) module are applied to the first and second Swin-Transformer blocks, respectively. Based on this dual-layer framework, the Swin-Transformer Block (×2) can be represented mathematically as follows:

$$\hat{f}^{l} = \mathrm{W-MSA}\left(\mathrm{LN}\left(f^{l-1}\right)\right) + f^{l-1} \tag{1}$$

$$f^l = \mathrm{MLP}\left(\mathrm{LN}\left(\hat{f}^l\right)\right) + \hat{f}^l \tag{2}$$

$$\hat{f}^{l+1} = \mathrm{SW-MSA}\left(\mathrm{LN}\left(f^l\right)\right) + f^l \tag{3}$$

$$f^{l+1} = \mathrm{MLP}\left(\mathrm{LN}\left(\hat{f}^{l+1}\right)\right) + \hat{f}^{l+1} \tag{4}$$

where $\hat{f}^l$ and f^l represent the outputs of the (S)W-MSA module and MLP module of the l-th block, respectively. The calculation of self-attention is as follows:

$$Attention(Q, K, V) = SoftMax\left(\frac{QK^T}{\sqrt{d}} + B\right)V \tag{5}$$

where $Q, K, V \in \mathbb{R}^{M\times d}$ represents the query, key, and value vector matrices; M and d represent the number of patches in the window and the dimension of the (query/key) matrix, respectively. Since the relative positions along each axis range from $[-M+1, M-1]$, one can set a bias matrix $\hat{B} \in \mathbb{R}^{(2M-1)\times(2M-1)}$ and extract the value B from $\hat{B}$.

4. Proposed Method

In this section, we first present an overview of our model (Section 4.1). Then, we cover the details of our model (Sections 4.2~ 4.4).

4.1. Architecture Overview

The dense rain streaks are a significant factor causing blurriness in the details of rainy images. However, individual rain streaks are small and have a minor impact on the light, which is reflected from objects to the camera. Therefore, when removing rain streaks, the rain-line map I can be represented as a linear combination of a clean background component B and a sparse rain-streak layer R. Thus, the image degradation caused by rain streaks can be modeled as follows [35]:

$$I = B + R \tag{6}$$

where I is the rainy image, B is the background image, R is the rain streak mask, which describes the distribution and motion of rain streaks. Based on the separability prior of rainy images, we constructed a dual-channel component decomposition network (DCD-Net), whose overall structure is shown in Fig. 3.

DCD-Net consists of two branches. On one hand, the background component decomposition network (BCDN) constructs an encoder-decoder structure with Swin-Transformer as a unit to extract multi-scale feature maps from rainy images and to reconstruct rain-free background images. BCDN servers as the backbone of our model. On the other hand, the mask component decomposition network (MCDN) constructs an encoder-decoder structure with attention modules as a unit to extract the reconstruction information of rain streaks mask from multi-scale feature maps. The significance of MCDN lies in its ability to enhance the model's recognition capability of rain-streak regions. The interaction between these two branches is achieved through shared kernels, which are used to provide feedback to the BCDN, improving the model's attention to rain-streak regions and minimizing the loss of background details.

4.2. Background Component Decomposition Network

The BCDN is an encoder-decoder network that is built with Swin-Transformer as the node. The role of the BCDN is to extract multi-scale feature maps from the input rainy images, and to preserve the reconstruction information of the rain-free background image as possible as it can. By combining transformer and Unet, the BCDN can fully extract long-range dependencies and detail features of the input degraded image. The skip connection is added to fuse the multi-scale features of the encoder stage with the upsampled features of the decoder stage, and thus it reduces the information loss of the feature maps, during the downsampling process.

Each block in the encoder part of the BCDN consists of a *Swin-Transformer block* and a *patch fusion block*. The Swin-Transformer block is responsible for information extraction at the feature level, where the resolution and feature dimension remain unchanged. The patch fusion block is responsible for downsampling, i.e. reducing height and width of the feature mapping, and increasing the number of channels (2× downsampling).

In addition, following the modeling idea of the transfer UNet, we also design a decoder that is symmetrical to the encoder. Each block of the decoder consists of a *Swin-Transformer block* and a *patch extension block*. The patch extension block is responsible for upsampling and reducing the number of channels (2× upsampling).

The decoder receives the multi-scale features passed through the skip connection from the encoder. Then, it fuses the multi-scale features with its own extracted features, and passes them to the next layer. This way, it thus reduces the spatial information loss caused by downsampling during the encoding process. Finally, the decoder output feature maps are restored to a clean background image, through an image reconstruction kernel.

4.3. Mask Component Decomposition Network

We introduce a parallel branch network for reconstructing rain streak masks, and build an encoder-decoder network with attention modules as nodes (Attention Model Encoder/Decoder), in order to fully understand the semantic information of rain streak masks, and to avoid confusion between rain streaks and background.

Using feature decoupling methods, heterogeneous feature maps of rain streak masks are extracted from the multi-scale feature maps of rainy images. The attention module is employed to extract detailed information from these heterogeneous feature maps, aiming to restore the rain streak mask.

The encoder part of the MCDN decouples and extracts heterogeneous feature maps from the multi-scale feature maps $FM_i(1 \le i \le 10)$ of the BCDN. The input of each unit is composed of the following two parts: 1) heterogeneous feature maps $MFM_i(1 \le i \le 5)$, which is decoupled from the multi-scale maps of the same layer unit of the BCDN; and 2) the output feature $EF_j(1 \le j \le 5)$ of the previous unit. The attention module extracts the features required for component reconstruction from the combined feature matrix of the input, using a self-attention kernel. And it down-samples the extracted features using maximum pooling.

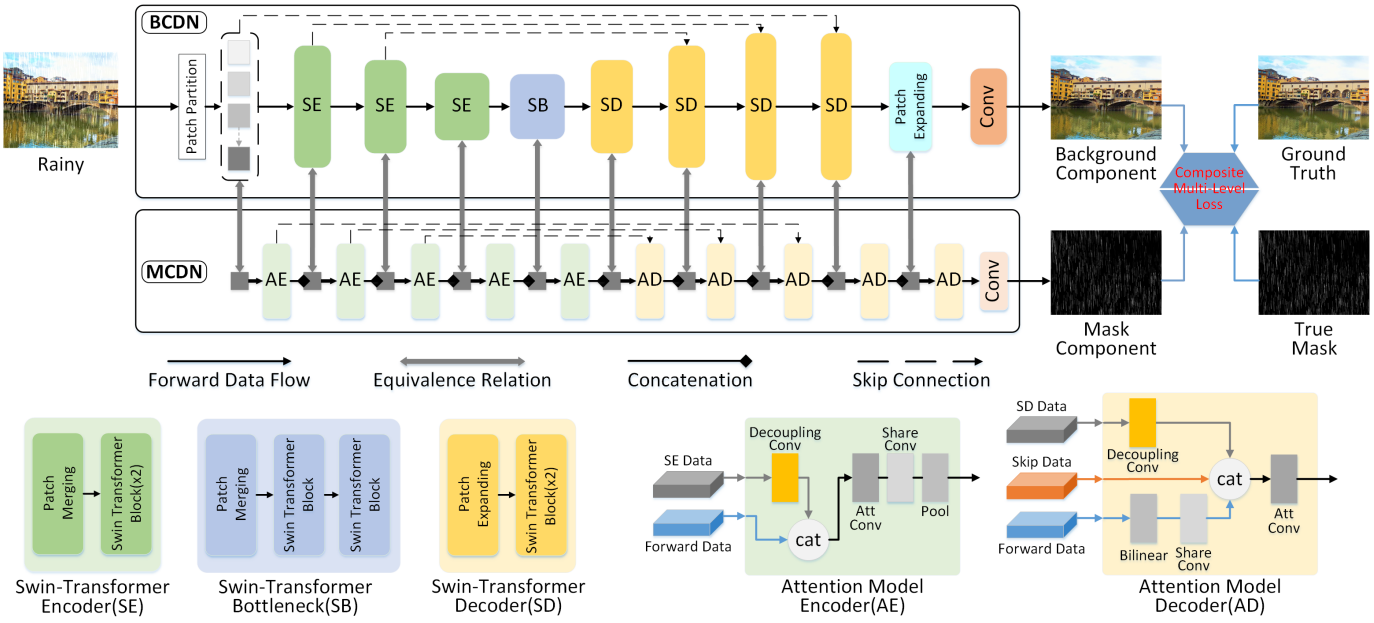

Fig. 3. Overall architecture of proposed DCD-Net. DCD-Net consists of the **B**ackground **C**omponent **D**ecomposition **N**etwork (BCDN) and the **M**ask **C**omponent **D**ecomposition **N**etwork (MCDN). The BCDN is an encoder (SE) decoder (SD) network built with Swin-Transformer as the node, which extracts multi-scale feature maps from the input rainy image and retain the reconstruction information of the rain-free background image. The final output background component is the target derained image. The MCDN is an encoder (AE) decoder (AD) network built with attention module as a node, which extracts unique representations of rain mask components from multi-scale feature maps in a decoupled manner, and ultimately reconstructs the rain streak mask. In order to enhance the model's attention to the rain pattern area and improve the model's ability to recognize rain streaks, we respectively impose full-supervision constraint on the output (Background Component & Mask Component) of the parallel dual-branch network. We design a composite multi-level loss function to calculate the loss values of the two supervision tasks. And then we add the two values with weights to train the network.

In the decoder part, each unit's input is composed of the following three elements: 1) the heterogeneous feature map $MFM_i(6 \le i \le 10)$, which is decoupled from the multi-scale mapping of the same layer unit in the BCDN; 2) The encoder output features that are passed through the shortcut connection at the same scale (skip connection); 3) the output feature $DF_j(1 \le j \le 5)$ of the previous unit. The attention module parses the features of the components from the combined feature matrix, using a self-attention kernel. And it performs bilinear interpolation upsampling on the parsed features. The output $EF_{\mathcal{E}}$ of the encoder in the network can be obtained using the following formula:

$$MFM_i = decoupling\left(FM_i\right), (1 \le i \le 10) \tag{7}$$

$$EF_{\mathcal{E}} = Enc\left(EF_{\mathcal{E}-1}, MFM_{\mathcal{E}}\right), (1 \le \mathcal{E} \le 5) \tag{8}$$

$$Enc(\cdot) = down(att(conv(cat_N(\cdot)))) \tag{9}$$

where $EF_{\mathcal{E}}$ represents the output of the $\mathcal{E}$-th encoder in the MCDN. $EF_0 = Null$ represents the input of the first encoder has no forward data, $att(\cdot)$ represents the self-attention kernel, $down(\cdot)$ represents the maximum pooling downsampling, and $decoupling(\cdot)$ represents the decoupling kernel.

The output $DF_{\mathcal{D}}$ of the decoder in the network can be obtained based on the following formulas:

$$DF_{\mathcal{D}} = Dec\left(\Phi\left(DF_{\mathcal{D}-1}\right), EF_{4-\mathcal{D}}, MFM_{\mathcal{D}+5}\right) \tag{10}$$

$$\Phi(\cdot) = up(conv(\cdot)) \tag{11}$$

$$Dec(\cdot) = att(cat_N(\cdot)) \tag{12}$$

where $DF_{\mathcal{D}}(1 \le \mathcal{D} \le 5)$ represents the output of the $\mathcal{D}$-th decoder of the MCDN, and $DF_0 = EF_5$ denotes that the forward data of the first decoder is the output of the last encoder. $EF_{4-\mathcal{D}}$ represents the output of the encoder passed through the shortcut. When $\mathcal{D} \ge 4$, $EF_{4-\mathcal{D}} = Null$, which means that only the first three decoders have skip connection. $\Phi(\cdot)$ represents dimension adjustment operation, and $up(\cdot)$ represents bilinear interpolation upsampling.

4.4. Loss Function

During network training, we set two supervision tasks and calculate the loss values for each supervision task via a composite multi-level loss function. The two supervision tasks are: 1) supervising the predicted background image (B_{pre}) with the ground truth image (B_{true}); 2) supervising the predicted rain streak mask (R_{pre}) with the true rain streak mask ($R_{true} = I - B_{true}$).

The supervision task for rain streak masks will assist the model in better learning the semantic information of rain streaks during training, and thereby it can improve the model's ability to recognize rain streaks.

The composite multi-level loss function includes pixel-level smooth L1 loss and image-level structural similarity loss. The pixel-level smooth L1 loss provides fine-grained supervision for image restoration from a micro-feature perspective. The image-level structural similarity loss controls the model from a macro-image perspective to avoid mistakenly removing the

background area of the original image. The definition of pixel-level smooth L1 loss is as follows:

$$\mathcal{L}_{\alpha\_SmoothL1} = \begin{cases} 0.5{E_\alpha}^2 & if\ |E_\alpha| < 1 \\ |E_\alpha| - 0.5 & otherwise, \end{cases} \tag{13}$$

where $E_\alpha = \alpha_{true} - \alpha_{pre}$, $\alpha = B$ or R. The loss of structural similarity is obtained based on the SSIM values of each group of compared images. The goal is to improve the structural similarity between predicted and real images. It can be defined as follows:

$$\mathcal{L}_{\alpha\_SSIM} = 1 - SSIM\left(\alpha_{true}, \alpha_{pre}\right), (\alpha = B, R) \tag{14}$$

where $SSIM(\cdot)$ is the function used to calculate structural similarity. The total loss of each supervision task can be defined as follows:

$$\mathcal{L}_{\alpha\_total} = \mathcal{L}_{\alpha\_SmoothL1} + \mathcal{L}_{\alpha\_SSIM}, (\alpha = B, R) \tag{15}$$

The total loss of the model can be summarized as follows:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{B\_total} + \lambda_2 \mathcal{L}_{R\_total} \tag{16}$$

where λ_1 and λ_2 are hyperparameters. They can control the contribution of the loss value (of each supervision task) to the total loss. λ_1 represents the relative importance for the supervised task of predicting the background image, while λ_2 represents the relative importance for the supervised task of predicting the rain streak mask.

In our model, the overall loss of the model considers the extraction of background components as the main objective, and the restoration of rain streak components as the secondary objective. In later experiments, we set λ_1 and λ_2 to 1 and 0.2, respectively.

5. Experiments

In this section, we first introduce experimental settings such as the datasets, training details, and evaluation metrics (Sections 5.1~5.2). Then, we cover the experimental results (Section 5.3). Finally, we conduct ablation study to verify the effectiveness of various parts of the proposed method (Section 5.4).

5.1. Datasets and Evaluation Metrics

Datasets: We evaluate the performance of our method on both real and synthetic datasets. Three widely used synthetic datasets are employed: Rain200L [39], Rain200H [39], and DID-Data [24]. Among them, Rain200L is a small synthetic rain dataset, Rain200H is a large synthetic rain dataset, both of which contain 1800 training image pairs and 200 testing image pairs. DID-Data consists of synthetic rain images with different rain directions and density levels, including 12000 training image pairs and 1200 testing image pairs. We conduct ablation experiments on the Rain200L dataset to verify the effectiveness of the DCD-Net structure. In addition, the large-scale real-world dataset SPA-Data [29] is employed, in order to further evaluate the robustness of DCD-Net. The rainy images in SPA-Data are extracted from real rainy videos, consisting of 638,492 image pairs for training and 1,000 image pairs for testing.

Evaluation Metrics: We use two widely used evaluation metrics, Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM), for image deraining algorithms. PSNR is based on the pixel error between two images and is inversely proportional to the error. PSNR has a range of $(0, +\infty)$. A higher PSNR value indicates a higher overall similarity between the two images, meaning that the derained result image is closer to the ground truth rain-free image. On the other hand, SSIM measures the similarity of content and textures between two images, and the SSIM value is ranging from 0 to 1. A higher SSIM value also indicates a better deraining effect. By using these two metrics, we can quantitatively compare the deraining performance of the algorithms.

5.2. Training Details

We used the Pytorch framework to build the network and conducted experiments on the NVIDIA RTX 8000 GPU. We used the Adam optimizer [40] with a learning rate initialized to 0.0001, which was automatically adjusted to 0.2 times the original value every 40 epochs. The batch size was set to 8, and the number of training epochs was determined based on the dataset size: 300 epochs were trained on both the Rain200L and Rain200H datasets, while 100 epochs were trained on the DID-Date dataset. During training, we randomly cropped the input image into patches of size 256×256 for rain removal. In each epoch, we extracted 20% of the training data as the validation set.

5.3. Comparisons with the State-of-the-Arts

We compared our method with several state-of-the-art deraining methods on the test dataset. The comparison was divided into numerical evaluation and visual evaluation. Details are presented as follows.

5.3.1. Synthetic Datasets

We compared DCD-Net with six advanced methods, PReNet [19], MPRNet [36], DualGCN [37], SPDNet [32], Restormer [38], and IDT [33], on three synthetic datasets.

Table 1 reports the quantitative evaluation results of each method on different datasets, with the best performance shown in bold and the second best performance shown underlined. From Table 1 , it can be observed that our method outperforms other rain removal methods and exhibits the best rain removal performance. For example, on the Rain200L/Rain200H datasets, DCD-Net improves by 0.28dB/1.03dB in PSNR compared to the second-best method. On the DID-Data dataset, DCD-Net improves by 3.07dB and 0.0058 in the PSNR and SSIM metrics, respectively, compared to the second-best method.

Fig. 4 presents visual comparisons between our method and these advanced methods. From this figure, it can be observed that models purely based on CNN, such as PReNet and MPRNet, tend to confuse rain streaks with the background in heavy rain scenes, resulting in the inability to restore clear images. On the other hand, rain removal methods based on Transformers are limited by their ability to extract detailed features, which leads

Table 1. Quantitative results on synthetic datasets. **Bold** and underline represent the best and second-best results.

Methods	Publication	Rain200L		Rain200H		DID-Data		Overall	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
PReNet [19]	CVPR-19	37.71	0.9798	29.02	0.8985	33.17	0.9480	33.30	0.9421
MPRNet [36]	CVPR-21	39.37	0.9810	30.65	0.9107	33.99	0.9591	34.67	0.9503
DualGCN [37]	AAAI-21	40.63	0.9875	31.15	0.9125	34.38	0.9621	35.39	0.9540
SPDNet [32]	ICCV-21	40.35	0.9866	31.28	0.9189	34.57	0.9558	35.40	0.9538
Restormer [38]	CVPR-22	40.87	**0.9880**	32.02	0.9215	35.29	0.9641	36.06	0.9579
IDT [33]	TPAMI-22	40.65	0.9873	32.11	**0.9226**	34.85	0.9622	35.87	0.9574
DCD-Net(ours)	—	**41.19**	0.9875	**33.14**	0.9221	**38.36**	**0.9699**	**37.56**	**0.9598**

Table 2. Quantitative results in terms of parameter sizes and time expense.

Methods	PReNet [19]	MPRNet [36]	DualGCN [37]	SPDNet [32]	Restormer [38]	IDT [33]	DCD-Net(ours)
Params	0.17M	20.10M	2.73M	3.04M	25.31M	18.30	11.51M
Time	1.8ms	217.3ms	29.5ms	32.9ms	273.6ms	197.8ms	124.4ms

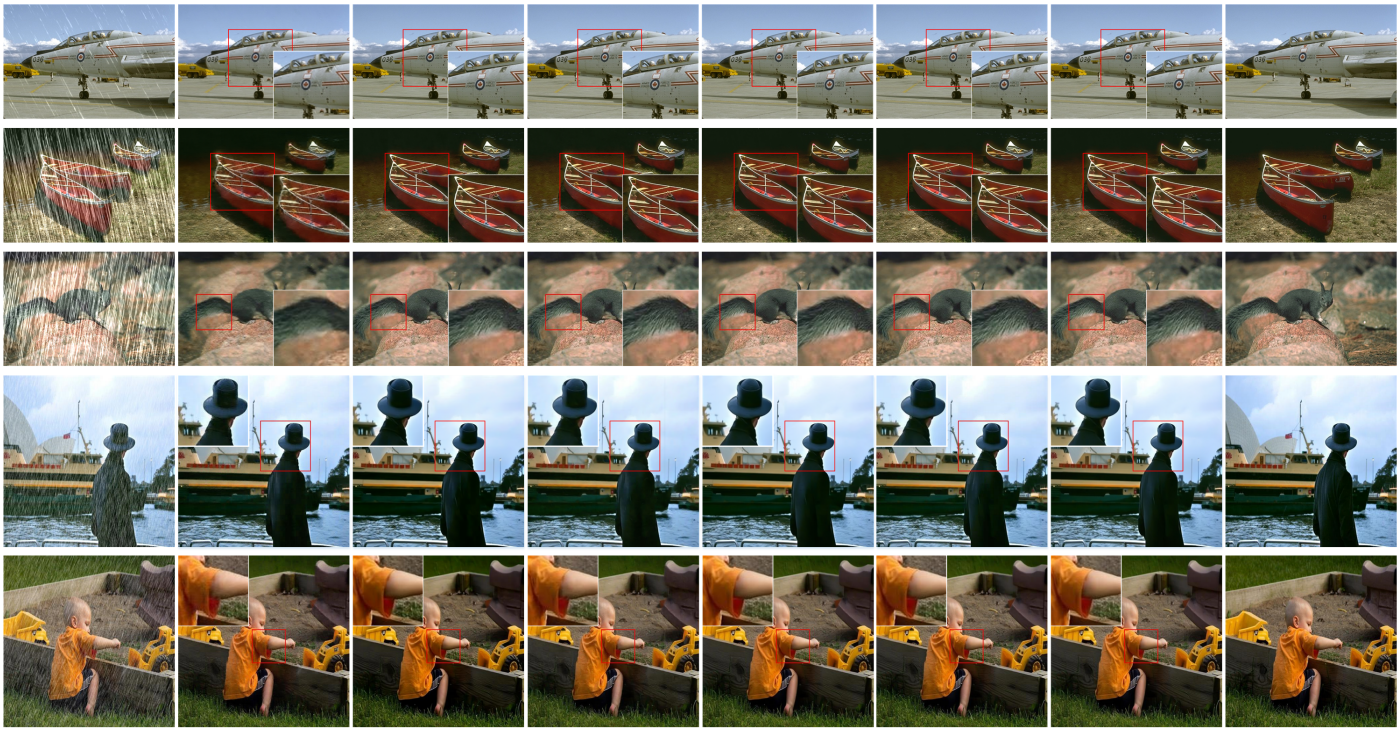

Fig. 4. Visual quality comparison with other state-of-the-art methods on the Rain200L/H and DID-Date datasets. Zooming in the figures offers a better view at the deraining performance.

to drawbacks in image details and texture restoration. DCD-Net, which combines the advantages of Swin-Transformer and UNet in extracting long-range dependencies and detailed features, can generate high-quality rain-free background images that are closer to ground truth. These results further demonstrate the effectivess of our proposed method.

Besides, Table 2 displays the parameter sizes of each model as well as the average time expenditure for processing a single image in the Rain200L dataset. From this table, we can see that the time cost of our method is moderate, compared with all these competitors. In addition, comparing with large-scale deep learning models such as MPRNet, Restormer, and IDT, it is evident that our approach has advantages in terms of parameter size and time cost. Although it may not have an advantage over

Fig. 5. Visual quality comparison with other state-of-the-art methods on the SPA-Data datasets. Zooming in the figures offers a better view at the deraining performance.

models like PReNet, DualGCN, and SPDNet in these aspects, our method exhibits better rain removal performance and can achieve clearer rain-free images with tolerable resource consumption (recall Fig. 4).

5.3.2. Real-world Datasets

In this subsection, we compare our method against state-of-the-art methods based on real datasets called SPA-Data [29]. The rainy images in SPA-Data are extracted from real rainy videos. The visual quality comparisons are shown in Fig. 5. From this figure, we can see that our method not only removes the majority of rain streaks but also minimizes the residual rain streak contours. Also, it reduces false judgments and removes artifacts caused by rain streaks (which cover the background details). All in one, this set of results demonstrate that our method can also obtain competitive results in real rain images.

5.4. Ablation Studies

To verify the effectiveness of each component in the DCD-Net network structure, we conducted ablation studies based on the Rain200L dataset.

5.4.1. Effectiveness of Parallel Dual-branch Structure

As mentioned earlier, the parallel dual-branch structure is used to enhance the model's ability in distinguishing rain streak masks and background textures. To verify the effectiveness of the mask component decomposition network (MCDN) in the parallel dual-branch structure, we compared its rain removal performance with the model under the single-branch structure. Here the former refers to our proposed model DCD-Net, while the later refers to the model that uses the single branch structure, denoted as DCD-Net*, for ease of presentation.

The quantitative comparison results on the Rain200L dataset are shown in Table 3. From the comparison results, we can see that, although the model with the dual-branch structure has more parameters than the model with the single-branch structure, it has better rain removal performance and stronger robustness when facing different types of rainy images.

Table 3. Comparative experimental results of branch structures.

Structures	Params	PSNR	SSIM
DCD-Net*	5.06M	40.25	0.9857
DCD-Net	11.51M	**41.19**	**0.9875**

5.4.2. Effectiveness of Loss Function

To investigate the impact of the proposed composite multi-level loss function on the model's rain removal performance, we conducted experiments on two separate supervised tasks and the composite loss function. The same parallel dual-branch network structure was used for all experiments, and model performance was evaluated based on two metrics, PSNR and SSIM. Table 4 shows the performance obtained by training the network with different loss functions, which were set using a controlled variable method. The experimental abbreviations are as follows:

- **DBS_single-Loss**: Supervised constraint only on the background image extracted by the dual-branch network.
- **DBS_multi-Loss**: The dual-supervision task set in this paper refers to the full-supervision constraint applied separately to the background image and rain mask components extracted by the dual-branch network.

By analyzing the results, it can be concluded that the dual-supervision task proposed in this paper helps improve the model's rain removal performance. The full supervision constraint on the rain mask can propagate semantic information about the rain to the model. This improves its ability to recognize rain, and helps the background separation network to focus on rain areas for removal, thereby reducing misremoval of background texture. In terms of performance metrics, the dual-supervision task improved the PSNR and SSIM values by 0.73dB and 0.0026, respectively, compared to the single-supervision task. The composite multi-level loss function also greatly improves the model's rain removal performance, since it achieves the best SSIM value when this loss function is used.

Furthermore, when using $\mathcal{L}_{SmoothL1}$ and $\mathcal{L}_{SSIM}$ as the loss functions for the dual-supervision task, the model not only achieves a better SSIM value but also increases the PSNR by 0.0036dB. Since $\mathcal{L}_{SSIM}$ calculates the loss value of the content structure from the image level, while the smooth L1 loss calculates the loss value of texture details from the pixel level, they form a complementary joint supervision effect, providing more targeted supervision for model training.

Table 4. Comparative experimental results of loss functions.

loss-functions	$\mathcal{L}_{SmoothL1}$	$\mathcal{L}_{SSIM}$	PSNR	SSIM
DBS_single-Loss	✓	—	40.40	0.9841
	—	✓	40.25	0.9841
	✓	✓	<u>40.46</u>	<u>0.9849</u>
DBS_multi-Loss	✓	—	32.21	0.9021
	—	✓	40.22	0.9839
	✓	✓	**41.19**	**0.9875**

5.4.3. Effectiveness of Attention Module

In order to evaluate the impact of the attention module used in MCDN, we compared the rain removal performance of the model with and without the attention modules. For ease of discussion, we use DCD-Net_Na to denote that the model constructed using Non-attention modules as building blocks. Table 5 presents the quantitative comparison results on the Rain200L dataset. From the table, it can be observed that the attention module plays a crucial role in enhancing the rain removal performance.

Table 5. Experimental results of with/without attention modules.

Models	PSNR	SSIM
DCD-Net_Na	37.41	0.9675
DCD-Net	**41.19**	**0.9875**

6. Concluding Remarks

In this paper, we proposed an effective single-image rain removal method, based on component decomposition. A dual-channel decomposition network (DCD-Net) is built according to the separability prior of rainy images, which extracts the background image and rain mask from the rainy image through a parallel dual-branch network architecture. To fully extract the underlying background image from the rainy image, a U-shaped background component decomposition network (BCDN) is constructed, using the Swin-Transformer as the basic unit. The rainy image is projected into a multi-scale feature space to reconstruct a clean background image, and then a U-shaped mask component decomposition network (MCDN), using attention modules as the basic unit, is built to decouple the reconstruction features of the rain mask from the multi-scale feature maps. The information interaction between the two branches is achieved through shared kernels. In addition, to improve the model's recognition ability for rain mask, we introduced dual-contrast supervision to the output of the dual-branch network, and added a supervision task specifically for rain mask to assist the model in better recognizing and removing rain. Qualitative and quantitative experiments demonstrate that the proposed method achieves good results on benchmark datasets and real images, compared to state-of-the-art methods.

Limitations. Our approach aims to minimize the loss of details in the image deraining process and thereby to improve the performance of image deraining. The inclusion of the mask component decomposition network increases the number of model parameters. Specifically, our model has about 11.51 million parameters, which inevitably requires high computational power when deployed in real applications. Additionally, the model is trained on a synthetic rainy weather dataset, and its effectiveness in handling more *complex* real rainy scenarios is uncertain. In the future, we would like to optimize our model and explore the feasibility of incorporating real rainy images for semi-supervised training.

Acknowledgments

We would like to thank all the anonymous reviewers for their helpful comments. This work is supported by the National Natural Science Foundation of China (61775139, 61972425, 62072126, 61772164, 61872242, 61972157), Shanghai Municipal Commission of Economy and Information (XX-RGZN-01-19-6348) and Opening Topic of Key Laboratory of Embedded Systems and Service Computing of Ministry of Education (ES-SCKF 2019-03).

References

[1] Liu, Z, Lin, Y, Cao, Y, Hu, H, Wei, Y, Zhang, Z, et al. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of ICCV. 2021, p. 10012–10022.

[2] Lin, X, Ma, L, Sheng, B, Wang, Z, Chen, W. Utilizing two-phase processing with FBLS for single image deraining. IEEE Trans Multim 2021;23:664–676.

[3] Ronneberger, O, Fischer, P, Brox, T. U-net: Convolutional networks for biomedical image segmentation. In: Proceedings of MICCAI. 2015, p. 234–241.

[4] Wang, Z, Ma, L, Lin, X, Wu, X. MSGC: A new bottom-up model for salient object detection. In: Proceedings of ICME. 2018, p. 1–6.

[5] Tan, X, Xu, K, Cao, Y, Zhang, Y, Ma, L, Lau, RWH. Night-time scene parsing with a large real dataset. IEEE Transactions on Image Processing 2021;30:9085–9098.

[6] Lin, X, Wang, Z, Ma, L, Li, R, Fang, M. Salient object detection based on multiscale segmentation and fuzzy broad learning. The Computer Journal 2022;65(4):1006–1019.

[7] Wang, Z, Ma, L, Lin, X, Zhong, H. Saliency detection via multi-center convex hull prior. In: Processings of ICASSP. 2018, p. 1867–1871.

[8] Tan, X, Lin, J, Xu, K, Pan, C, Ma, L, Lau, RWH. Mirror detection with the visual chirality cue. IEEE Transactions on Pattern Analysis and Machine Intelligence 2022;.

[9] Kang, LW, Lin, CW, Fu, YH. Automatic single-image-based rain streaks removal via image decomposition. IEEE Transactions on Image Processing 2012;21(4):1742–1755.

[10] Kim, JH, Lee, C, Sim, JY, Kim, CS. Single-image deraining using an adaptive nonlocal means filter. In: Proceedings of ICIP. 2013, p. 914–917.

[11] Chen, DY, Chen, CC, Kang, LW. Visual depth guided color image rain streaks removal using sparse coding. IEEE transactions on circuits and systems for video technology 2014;24(8):1430–1455.

[12] Luo, Y, Xu, Y, Ji, H. Removing rain from a single image via discriminative sparse coding. In: Proceedings of ICCV. 2015, p. 3397–3405.
[13] Chang, Y, Yan, L, Zhong, S. Transformed low-rank model for line pattern noise removal. In: Proceedings of ICCV. 2017, p. 1726–1734.
[14] Gu, S, Meng, D, Zuo, W, Zhang, L. Joint convolutional analysis and synthesis sparse representation for single image layer separation. In: Proceedings of ICCV. 2017, p. 1708–1716.
[15] Zhang, H, Patel, VM. Convolutional sparse and low-rank coding-based rain streak removal. In: Proceedings of WACV. 2017, p. 1259–1267.
[16] Fu, X, Liang, B, Huang, Y, Ding, X, Paisley, J. Lightweight pyramid networks for image deraining. IEEE Transactions on Neural Networks and Learning Systems 2019;31(6):1794–1807.
[17] Hu, X, Fu, CW, Zhu, L, Heng, PA. Depth-attentional features for single-image rain removal. In: Proceedings of CVPR. 2019, p. 8022–8031.
[18] Li, G, He, X, Zhang, W, Chang, H, Dong, L, Lin, L. Non-locally enhanced encoder-decoder network for single image de-raining. In: Proceedings of ACM Multimedia. 2018, p. 1056–1064.
[19] Ren, D, Zuo, W, Hu, Q, Zhu, P, Meng, D. Progressive image deraining networks: A better and simpler baseline. In: Proceedings of CVPR. 2019, p. 3937–3946.
[20] Li, Y, Tan, RT, Guo, X, Lu, J, Brown, MS. Rain streak removal using layer priors. In: Proceedings of CVPR. 2016, p. 2736–2744.
[21] Zhu, L, Fu, CW, Lischinski, D, Heng, PA. Joint bi-layer optimization for single-image rain streak removal. In: Proceedings of ICCV. 2017, p. 2526–2534.
[22] Fu, X, Huang, J, Zeng, D, Huang, Y, Ding, X, Paisley, J. Removing rain from single images via a deep detail network. In: Proceedings of CVPR. 2017, p. 3855–3863.
[23] Li, X, Wu, J, Lin, Z, Liu, H, Zha, H. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In: Proceedings of ECCV. 2018, p. 254–269.
[24] Zhang, H, Patel, VM. Density-aware single image de-raining using a multi-stream dense network. In: Proceedings of CVPR. 2018, p. 695–704.
[25] Yang, W, Tan, RT, Feng, J, Liu, J, Guo, Z, Yan, S. Joint rain detection and removal via iterative region dependent multi-task learning. CoRR 2016;abs/1609.07769. URL: `http://arxiv.org/abs/1609.07769`. `arXiv:1609.07769`.
[26] Liu, X, Suganuma, M, Sun, Z, Okatani, T. Dual residual networks leveraging the potential of paired operations for image restoration. In: Proceedings of CVPR. 2019, p. 7007–7016.
[27] Zhang, H, Sindagi, V, Patel, VM. Image de-raining using a conditional generative adversarial network. IEEE transactions on circuits and systems for video technology 2019;30(11):3943–3956.
[28] Wang, G, Sun, C, Sowmya, A. Erl-net: Entangled representation learning for single image de-raining. In: Proceedings of ICCV. 2019, p. 5644–5652.
[29] Wang, T, Yang, X, Xu, K, Chen, S, Zhang, Q, Lau, RW. Spatial attentive single-image deraining with a high quality real rain dataset. In: Proceedings of CVPR. 2019, p. 12270–12279.
[30] Wei, W, Meng, D, Zhao, Q, Xu, Z, Wu, Y. Semi-supervised transfer learning for image rain removal. In: Proceedings of CVPR. 2019, p. 3877–3886.
[31] Dosovitskiy, A, Beyer, L, Kolesnikov, A, Weissenborn, D, Zhai, X, Unterthiner, T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In: Proceedings of ICLR. 2021,.
[32] Yi, Q, Li, J, Dai, Q, Fang, F, Zhang, G, Zeng, T. Structure-preserving deraining with residue channel prior guidance. In: Proceedings of ICCV. 2021, p. 4238–4247.
[33] Xiao, J, Fu, X, Liu, A, Wu, F, Zha, ZJ. Image de-raining transformer. IEEE Transactions on Pattern Analysis and Machine Intelligence 2022;.
[34] Vaswani, A, Shazeer, N, Parmar, N, Uszkoreit, J, Jones, L, Gomez, AN, et al. Attention is all you need. In: Proceedings of NIPS. 2017, p. 5998–6008.
[35] Deng, S, Wei, M, Wang, J, Feng, Y, Liang, L, Xie, H, et al. Detail-recovery image deraining via context aggregation networks. In: Proceedings of CVPR. 2020, p. 14560–14569.
[36] Zamir, SW, Arora, A, Khan, S, Hayat, M, Khan, FS, Yang, MH, et al. Multi-stage progressive image restoration. In: Proceedings of CVPR. 2021, p. 14821–14831.
[37] Fu, X, Qi, Q, Zha, ZJ, Zhu, Y, Ding, X. Rain streak removal via dual graph convolutional network. In: Proceedings of AAAI; vol. 35. 2021, p. 1352–1360.
[38] Zamir, SW, Arora, A, Khan, S, Hayat, M, Khan, FS, Yang, MH. Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of CVPR. 2022, p. 5728–5739.
[39] Yang, W, Tan, RT, Feng, J, Liu, J, Guo, Z, Yan, S. Deep joint rain detection and removal from a single image. In: Proceedings of CVPR. 2017, p. 1357–1366.
[40] Loshchilov, I, Hutter, F. Decoupled weight decay regularization. In: Proceedings of ICLR. 2019,.