Semi-supervised Single Image Deraining with Discrete Wavelet Transform

Xin Cui[1], Wei Shang[2], Dongwei Ren[2(✉)], Pengfei Zhu[1], and Yankun Gao[3]

[1] Tianjin University, Tianjin 300350, China
2019216101@tju.edu.cn
[2] Harbin Institute of Technology, Harbin 150001, China
[3] Beijing Institute of Computer Technology and Application, Beijing, China

Abstract. In recent years, single image deraining has received considerable research interests. Supervised learning is widely adopted for training dedicated deraining networks to achieve promising results on synthetic datasets, while limiting in handling real-world rainy images. Unsupervised and semi-supervised learning-based deranining methods have been studied to improve the performance on real cases, but their quantitative results are still inferior. In this paper, we propose to address this crucial issue for image deraining in terms of backbone architecture and the strategy of semi-supervised learning. First, in terms of network architecture, we propose an attentive image deraining network (AIDNet), where residual attention block is proposed to exploit the beneficial deep feature from the rain streak layer to background image layer. Then, different from the traditional semi-supervised method by enforcing the consistency of rain pattern distribution between real rainy images and synthetic rainy images, we explore the correlation between the real clean images and the predicted background image by imposing adversarial losses in wavelet space I_{HH}, I_{HL}, and I_{LH}, resulting in the final AID-DWT model. Extensive experiments on both synthetic and real-world rainy images have validated that our AID-DWT can achieve better deraining results than not only existing semi-supervised deraining methods qualitatively but also outperform state-of-the-art supervised deraining methods quantitatively. All the source code and pre-trained models are available at https://github.com/cuiyixin555/DeRain-DWT.

Keywords: Single image deraining · Semi-supervised learning · Attention · Discrete wavelet transform

1 Introduction

Single image deraining is a challenging task, and has a board application prospect in object detection, outdoor recognition and automatic driving [4,11] when facing bad weather condition. Image deraining can be regarded as an image decomposition problem that rainy image can be separated into rain pattern space $\mathcal{R}$ and clean background image space $\mathcal{X}$. Previously, traditional optimization algorithms, *e.g.* low-rank model, sparse code model, and Gaussian mixture model

D. N. Pham et al. (Eds.): PRICAI 2021, LNAI 13033, pp. 265–278, 2021.
https://doi.org/10.1007/978-3-030-89370-5_20

[1,15,18,21,25], etc., are adopted as the priors of rain streak layer and background image layer. However, these handcrafted designed priors are limited in modeling the complicated composition pattern of real-world rainy images, and also they are very time-consuming. With the rapid development of deep learning in recent years, learning-based deraining methods have achieved great progress. Supervised learning is introduced to address image deraining problem, and many Convolutional Neural Networks (CNNs)-based methods for single image deraining have been proposed [6,8,24,26,28–30,33,41]. These methods employ deep networks to automatically extract features of layers, enabling them to model more complex mappings from rainy images to clean images. Albeit great quantitative results on synthetic datasets, they cannot well deal with real-world rainy images. Then, unsupervised learning and semi-supervised learning are suggested to exploit real-world rainy images, leading to better generalization in practical applications. But unsupervised deraining method is quantitatively inferior to supervised deraining methods. In [31,35], transfer learning is introduced to transfer deraining model trained on synthetic images to real rainy images. As for heavy rainy image, these semi-supervised deraining methods can not process it, and there is leeway to improve deraining visual quality.

In this paper, we adopt semi-supervised strategy that we design a residual attention image deraining network and introduce real clean images to make our network learn the similarity of image texture in discrete wavelet space. In particular, we design the main network into two parts, where one is used to extract rainy streak layer, and the other is used to recover clean background image. As training iterations increasing, each convolution attention block is used as a coefficient unit for rainy pattern feature aggregation in image space. As for semi-supervised stage, to make the predicted clean background more natural, we design a discriminator which contains multiple convolutional layers to calculate adversarial losses between real-world clean images and predicted clean images in three subband f_{HH}, f_{HL} and f_{LH}. As for residual attention modules, we adopt 4 kinds of attention modules to explore its performance advantages in our deraining backbone.

Extensive experiments have been conducted on both synthetic and real-world rainy benchmark datasets. Our model quantitatively outperforms not only semi-supervised deraining method qualitatively but also state-of-the-art supervised deraining methods quantitatively.

Our contributions can be summarized from three aspects:

- We propose a simple yet effective semi-supervised deraining approach by Discrete Wavelet Transform (DWT), via which real-world clean images can be easily used to benefit the generalization ability of trained deraining model.
- We design a residual attention-based image deraining model to enhance the separation of the rain streak layer and the corresponding background layer; We use a ordinary convolution attention block for rain streak feature extraction, and compare with the other three attention modules, such as self-calibration block [17], attention feature fusion block [36] and self-attention block [39].

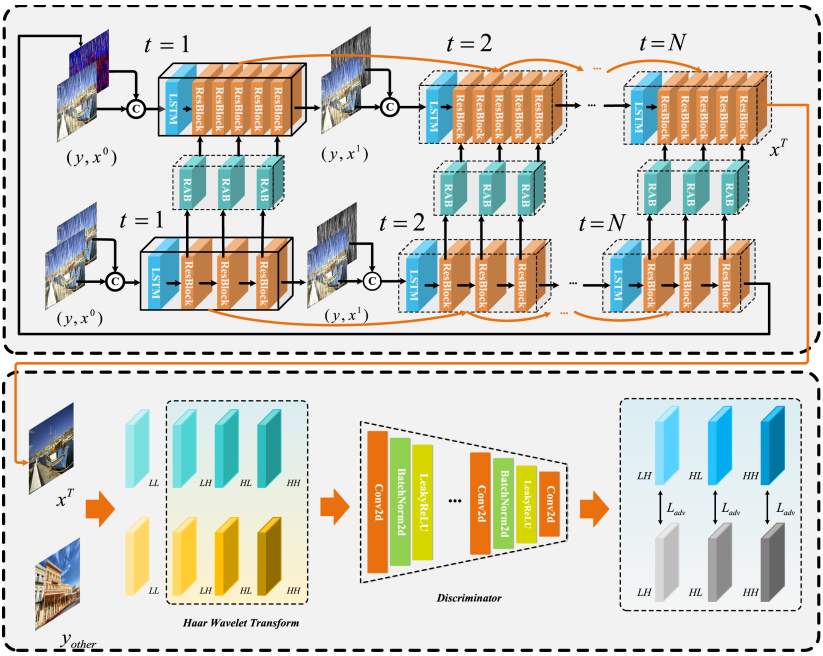

Fig. 1. The architecture of Semi-supervised deraining with Haar Wavelet Transform.

– Extensive experiments on synthetic and real-world rainy images have been conducted to validate that our model is superior to both supervised and semi-supervised deraining methods.

2 Related Works

Deep learning-based image deraining has been widely studied with the supervised learning manner, where various network architectures are designed to learn the mapping from rainy image to clean background image. In pioneer works [5,6], CNN and ResNet are first adopted to predict clean background image, outperforming conventional deraining methods. Subsequently, more complicated network architectures are proposed to better extract deep features from rainy images. In [28,30], multi-scale strategy can help model learn image features under different scales and enhance its robustness; Especially the application of dilation convolution is proposed, which is benefited in detecting and removing rain streaks simultaneously. In [14,26], recurrent networks are proposed to handle heavy rain streak accumulation. In [29,37], densely connected CNN is adopted for jointly estimating rain density and removing rain streaks. Besides, there are several works to incorporate lightweight networks in a cascaded scheme [3] or in a Laplacian pyramid framework [7].

Moreover, in [38], the authors propose to take advantage of adversarial learning to enhance the texture details in derained images. Most recently, pre-trained transformer [8] is introduced to significantly improve the quantitative metrics for image deraining. To sum up, supervised learning-based deraining methods have achieved excellent performance on paired synthetic datasets, but the trained deraining model are likely to poorly generalize to real-world rainy images. Then, unsupervised learning and semi-supervised learning are suggested to exploit real-world rainy images, leading to better generalization in practical applications. In [42], Zhu et al. proposed to adopt CycleGAN [43] to exploit unpaired real rainy images, which can improve generalization ability to real rainy images. In [31], SIRR is proposed to transfer deraining model trained on synthetic images to real rainy images. In [35], Syn2Real is proposed by adopting Gaussian processes to exploit both synthetic and real rainy images. However, these semi-supervised and unsupervised deraining methods may also be inferior to supervised methods in terms of quantitative metrics, and there is leeway to improve deraining visual quality.

3 Semi-supervised Image Deraining by DWT

In this section, we first present the proposed semi-supervised deraining framework by discrete wavelet transform in Sect. 3.1, and then give the details of residual attention framework in Sect. 3.2, finally the realization of our semi-supervised training method on the discrete wavelet transform is explained in Sect. 3.3.

3.1 Methodology Overview

As shown in Fig. 1, we propose to exploit real-world rainy images without paired ground-truth when training deraining networks, which is a semi-supervised approach. Different from [31,35], we propose a simple yet effective discriminative learning strategy by DWT to enforce the feature consistency of clean background from synthetic and real-world clean images on three subband f_{HH}, f_{HL} and f_{LH}. As shown in Fig. 1, the entire network structure is divided into two parts; One is LSTM followed by 5 resblocks for background prediction, and the other is LSTM followed by 3 resblocks to rain streak extracted. Between the upper and lower parts, the residual attentive blocks is applied to converge rain pattern feature removal in image space. Formally, the procedure is described as

$$\begin{aligned} Res_{r_i}^{\ t} &= F_{res_i}({h_r}^{t-1}), i = 1, 2, 3 \\ Res_{x_j}^{\ t} &= F_{res_j}({h_x}^{t-1}), j = 1, 2, 3, 4, 5 \\ Rab_{x_k}^{\ t} &= F_{rab_k}(res_{r_i}^{\ t}) \times res_{x_j}^{\ t} + res_{x_j}^{\ t}, k = 1, 2, 3 \end{aligned} \tag{1}$$

where hidden state h_r is from LSTM in space $\mathcal{R}$, hidden state h_x is from LSTM in space $\mathcal{X}$, $res_{r_i}^{\ t}$ indicates the residual map of i-th ResBlock in space $\mathcal{R}$, $res_{x_j}^{\ t}$ denotes the residual map of j-th ResBlock in space $\mathcal{X}$. And thus there are three

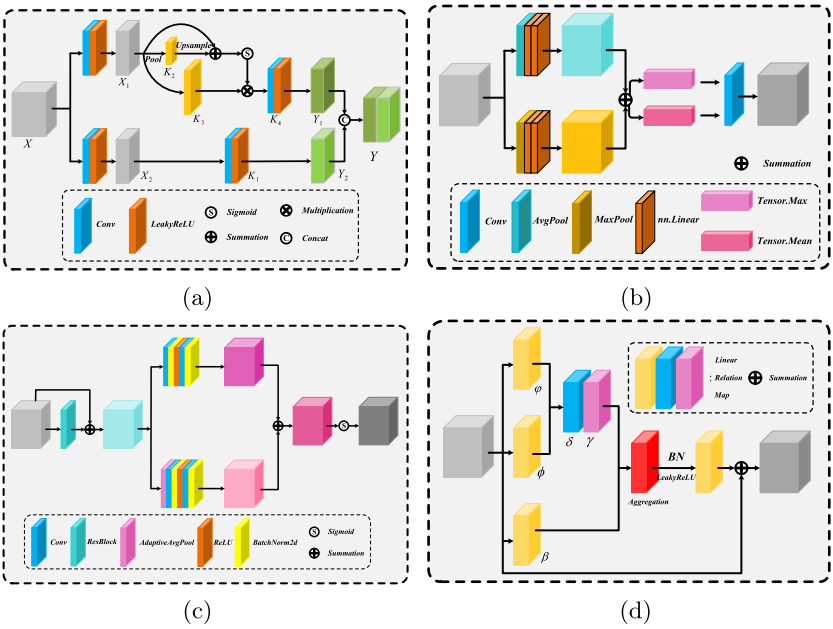

Fig. 2. *Top row*: (a) self-calibrated attention, (b) convolutional attention. *Bottom row*: (c) feature fusion attention, (d) self-attention.

RAB modules, where $rab_{x_k}^{t}$ connects the i-th ResBlock in $\mathcal{R}$ and j-th ResBlock in $\mathcal{X}$, exploiting the beneficial deep features between $\mathcal{R}$ and $\mathcal{X}$. The last crucial issue is how to determine the connections of i and j. To answer this question, we conducted experiments on Rain200H datasets [28] to validate the effectiveness of different connections. We will explain it specificly.

Finally, the predicted background X^g is used to initialize discrete wavelet transform discriminator network. When training deraining model shown on bottom row by only using real-world clean images, adversarial losses calculated in predicted background X^g and real clean image X_c in f_{HH}, f_{HL} and f_{LH} that is adopted to enforce its the consistency of feature distribution. Our model AID-DWT can achieve better results than existing semi-supervised and supervised deraining methods.

Overall, AID-DWT model consists of three parts: (i) Recurrent training for rain streak extracted and clean background image predicted with ResLSTM framework; (ii) Applying attention block for extracting rainy pattern feature strongly in image space; (iii) Calculating three subband adversarial losses on f_{HH} and f_{HL} and f_{LH}.

3.2 Residual Attentive Network Architecture

The good design of attention block can describe the rain pattern feature to the maximum extent. With the structure of deep CNNs becoming more and more

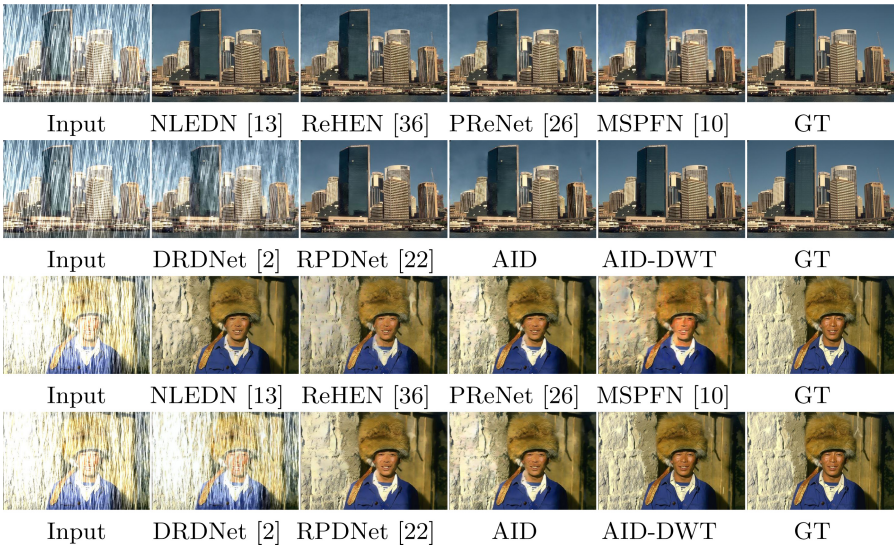

Fig. 3. Examples about the comparison of our method with other methods on Rain200H dataset.

complicated, extracting mean feature and max feature by average pool and max pool is not enough to achieve satisfactory results. Thanks to variety of attention module [17,32,36,39], as shown in Fig. 2, we will explain them respectively.

Self-Calibrated Block (SC): Different from the traditional attention mechanism, its usually performs operations in the dimension of the feature to obtain the average feature and the maximum feature. SC block has four parts of filters, i.e., $[K_1, K_2, K_3, K_4]$. Through splitting filters, the input X with channel C is split into X_1 and X_2 through 1×1 convolution with the channel C/2. Reviewing the entire self-calibrated convolution, it enables each spatial position to adaptively encode the context, which make difference between it and traditional attention block (CA).

Self-Attention Block (SAN): In [39], Zhao *et al.* firstly introduce feature aggregation into pairwise self-attention block. The whole procedure is describe as

$$y_i = \sum_{j \in R(i)} \alpha(x_i, x_j) \odot \beta(x_j), \tag{2}$$

where x_i and x_j are feature maps with indexes i and j, $\odot$ is the Hadamard product called aggregated with R(i). In order to utilize more surrounding pixels, the size of footprint is set 5×5.

Attention Feature Fusion Block (AFF): Similar to traditional convolution attention block (CA), AFF block [36] extracts the key pixels, as well as processes the residual information further. The whole process can be described as

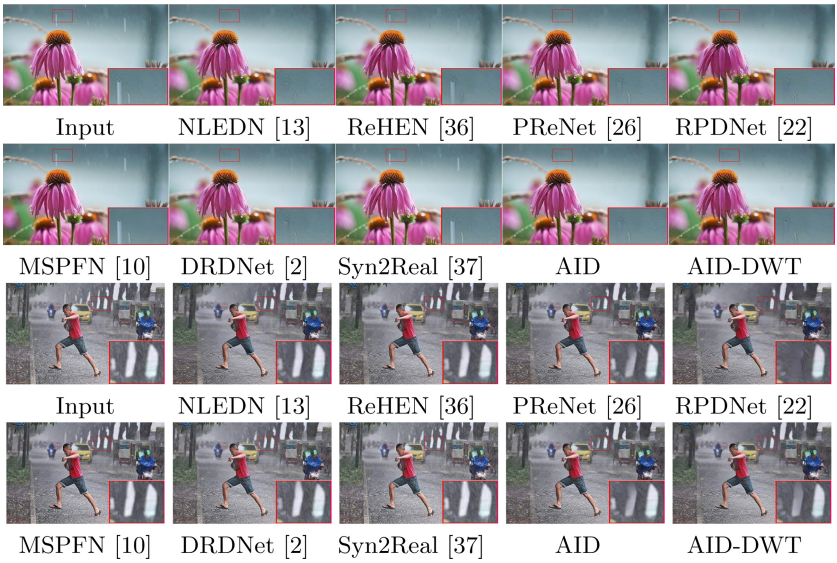

Fig. 4. Examples about the comparison of our method with other methods on real-world datasets.

$$\alpha_i = Local(x_i) + Global(x_i), \tag{3}$$

$$y_i = x_i \times \alpha + Res(x_i) \times (1 - \alpha), \tag{4}$$

where α_i is a attentional factor that is realized through a local attention block and a global attention block; A local attention block include multi-layer convolution that help model to learn feature based on dimension, while a global attention contains global average pool to extracted feature on space. Finally, the original feature x_i and the residual feature $Res(x_i)$ are proportionally distributed through attentional factor, which effectively solves the problem of information decreased as the number of convolution layers increases.

Convolution Attention Block (CA): As a traditional attention block, CA is widely used in feature extraction. As for input feature map F (H×W×C), it will be dealt with global max pooling and global average pooling, respectively to obtain two $1 \times 1 \times C$ features. And then, its will be sent to a two-layer neural network (MLP) that the number of neurons in the first layer is C/r (r is the reduction rate), the activation function is Relu, and the number of neurons in the second layer is C; The two-layer neural network is shared. After that, the MLP output features are subjected to an element-wise addition operation, as well as the sigmoid activation operation is performed to generate the final channel attention feature, namely M_c. Finally, the M_c and the input feature map F are subjected to an element-wise multiplication operation to generate the input features required by the Spatial attention module.

3.3 Discriminator by DWT for Semi-supervised Method

In 2D Discrete wavelet transform (DWT), four filters, i.e. f_{LL}, f_{LH}, f_{HL}, and f_{HH}, are used to convolve with an image x [19]. To illustrate the three subband filter, we first give the definition of f_{LH}, f_{HL}, and f_{HH},

$$\mathbf{f}_{LH} = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}, \mathbf{f}_{HL} = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}, \mathbf{f}_{HH} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}. \tag{5}$$

Given an image $\mathbf{x}$ with size of $m \times n$, the (i, j)-th value of $\mathbf{x}_1$ after 2D discrete transform can be written as $\mathbf{x}_1(i, j) = \mathbf{x}(2i-1, 2j-1) + \mathbf{x}(2i-1, 2j) + \mathbf{x}(2i, 2j-1) + \mathbf{x}(2i, 2j)$. Even though the downsampling operation is deployed, due to the biorthogonal property of DWT, the original image $\mathbf{x}$ can be accurately reconstructed by the inverse wavelet transform (IWT), i.e., $\mathbf{x} = IWT(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)$.

In order to make our predicted clean background image have similar texture distribution with real-world clean image, we introduce adversarial loss in the three subband image f_{HH}, f_{HL}, f_{LH}. First, we achieve X_{hh}, X_{hl} and X_{lh} by Hadamard product with above three subband. Our goal is to train the discriminator with the above subband images as

$$L_{hh} = Adv(X^i_{hh}, X^c_{hh}), \tag{6}$$

$$L_{hl} = Adv(X^i_{hl}, X^c_{hl}), \tag{7}$$

$$L_{lh} = Adv(X^i_{lh}, X^c_{lh}), \tag{8}$$

where Adv is WGAN-GP Loss, as well as X^i and X^c are denoted as a predicted background and a real-world clean image, respectively. So, we can treat X^c_{hh}, X^c_{hl} and X^c_{lh} as pseudo label for corresponding predicted output.

4 Experimental Results

In this section, we conduct extensive experiments to demonstrate the effectiveness of the proposed method on widely used four synthetic datasets and two real-world datasets. Eight state-of-the-art baseline are compared in this paper. Next, we will introduce the datasets and measurements in details as in Sect. 4.1, implementation details in Sect. 4.2, results on synthetic datasets and real-world datasets in Sect. 4.3 and ablation study in Sect. 4.4, respectively.

4.1 Datasets and Measurements

Our experiment is verified on four synthetic datasets and two real-world datasets, such as Rain200H [28], Rain1200 [37], Rain1400 [6] and Rain12 [16] for synthetic, as well as SPADatasets [27] and $Real_{275}$ for real-world datasets. Rain200H has heavy rain with different shapes, directions and sizes, which is the most challenging dataset including 1800 images for training and 200 images for testing. Rain200L contains the same number of pictures, which has light rain and easy to

trained. Rain1200 has three different level of rain images, including heavy rain, medium rain and light rain, which contains 12000 training images and 1200 testing images. Rain1400 has medium level rain images, which includes 12600 images for training and 1400 images for testing. Rain12 has 12 images for testing. SPA Datasets include 1000 testing images with labels. In addition, we has achieve 275 real rainy images from Internet. We has trained our proposed model on different datasets for verifying its robustness.

Table 1. The values of PSNR, SSIM and NIQE on two real datasets. **Red**, blue and cyan colors are used to indicate top **1st**, 2nd and 3rd rank, respectively.

	NLEDN [13]		ReHEN [34]		PReNet [26]		RPDNet [22]		AID		AID-HWT	
Dataset	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SPA	30.596	0.9363	32.652	0.9297	32.720	0.9317	32.803	0.9337	31.721	0.9359	**33.263**	**0.9375**
Dataset	Derain	GT	Derain	GT	Derain	GT	Derain	GT	Derain	GT	Derain	GT
$Real_{275}$	3.5554	–	3.7355	–	3.7745	–	3.8957	–	3.6013	–	**3.5519**	–
	MSPFN [10]		DRDNet [2]		SIRR [31]		Syn2Real [35]		AID		AID-HWT	
Dataset	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SPA	29.538	0.9193	28.083	0.9126	22.666	0.7474	31.824	0.9307	31.721	0.9359	**33.263**	**0.9375**
Dataset	Derain	GT	Derain	GT	Derain	GT	Derain	GT	Derain	GT	Derain	GT
$Real_{275}$	3.8616	–	3.6634	–	3.5592	–	4.0372	–	3.6013	–	**3.5519**	–

4.2 Implementation Details

Our AID-DWT networks are implemented using Pytorch [23] framework, adopt ADAM [12] algorithm for optimization, and are trained on PC equipped with two NVIDIA GTX 2080Ti GPUs. In our experiments, all the network shared the same training setting. We trained the network for 100 epochs. Each pair of training of samples will be randomly cropped 100×100 pixels. Adam optimizer is used with a learning rate of 0.001 which is divided by 5 after the 30th epochs, 50th epochs and 80th epochs.

4.3 Results and Analysis

Quantitative Comparsion. We compare our proposed model with NLEDN [13], ReHEN [34], PReNet [26], RPDNet [22], MSPFN [10] and DRDNet [2], which baseline models adopted supervised pattern, and SIRR [31], Syn2Real [35] with semi-supervised methods under the three metrics of PSNR [9], SSIM [40] and NIQE [20]. We trained our models on the synthetic datasets Rain200H, Rain1200 and Rain1400, and compare the quantitative results obtained with the training methods under the corresponding dataset, respectively. The metric results are presented on Table 2 and Table 1. Our proposed method has achieved the highest PSNR, SSIM and NIQE in all datasets.

Qualitative Comparsion. Figure 3 exhibits some synthetic examples on Rain200H dataset. We can see that we proposed model can achieve the best

Table 2. The values of PSNR and SSIM on four synthetic datasets. **Red**, blue and cyan colors are used to indicate top **1st**, 2nd and 3rd rank, respectively.

	NLEDN [13]		ReHEN [34]		PReNet [26]		AID-HWT	
Dataset	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Rain200H	27.315	0.8904	27.525	0.8663	27.883	0.8908	**28.903**	**0.9074**
Rain1200	30.799	0.9127	30.456	0.8702	27.307	0.8712	**31.960**	**0.9136**
Rain1400	30.808	0.9181	30.984	0.9156	30.609	0.9181	**31.001**	**0.9246**
Rain12	33.028	0.9615	35.095	0.9400	34.7912	0.9644	**35.587**	**0.9679**
	RPDNet [22]		MSPFN [10]		DRDNet [2]		AID-HWT	
Dataset	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Rain200H	27.909	0.8923	25.554	0.8039	22.825	0.7114	**28.903**	**0.9074**
Rain1200	26.486	0.8401	30.390	0.8862	28.386	0.8275	**31.960**	**0.9136**
Rain1400	30.772	0.9178	30.016	0.9164	28.360	0.8574	**31.001**	**0.9246**
Rain12	35.055	0.9657	34.253	0.9469	25.199	0.8497	**35.587**	**0.9679**

results, while other baseline models also are left some artifacts or remaining rain streaks. Especially, DRDNet [2] fails to work on Rain200H datasets. In addition, we also provide some examples shown in Fig. 4 of real-world datasets to prove the superiority of the proposed algorithm comparing with others. Expecially the rainy scene in forth column, our method can well recognize the cropped area that presents interspace between a trunk and the other, while the RPDNet [22] model removed it and regard it as a rain streak; In addition, our model can remove most of the rain streaks in the background while MSPFN [10] and DRDNet [2] even leaves behind traces of rain streaks. To sum up, our proposed model can adapt various rainy condition and restore image details and texture information better.

4.4 Ablation Study

In this section, we analyse the proposed model by conducting various experiments on Rain200H [28] datasets.

We analyze the network designment that consist of different attention modules, different location for residual attention connection, different recurrent stage numbers, different unsupervised losses. The experiment results are illustrated in Table 3, Table 4 and Table 5;

We adopt four attention block in ablation experiments on condition of the residual attention location fixed on $j = 1, 3, 5$, such as SAN block [39], AFF block [36], SC block [17] and CA block [32]; Further more, as for different combinations of location for residual attentive connection (RC), we set four experiments on different location, such as $j = 1, 2, 3$, $j = 2, 3, 4$, $j = 3, 4, 5$ and $j = 1, 3, 5$; To explore whether unsupervised losses with HWT can play a key role in the clean background predication, we set four experiments with L_{HH}, L_{HL}, L_{LH} and without HWT operation.

Table 3. The results of different modules on Rain200H. The best results are highlighted in boldface.

Experiments	E1	E2	E3	E4	E5	E6	E7	E8
Single SAN Block	✓							
Single AFF Block		✓						
Single SC Block			✓					
Single CA Block				✓	✓	✓	✓	✓
RA Connection on Location 1, 2, 3					✓			
RA Connection on Location 2, 3, 4						✓		
RA Connection on Location 3, 4, 5							✓	
RA Connection on Location 1, 3, 5	✓	✓	✓	✓				✓
PSNR	28.871	27.423	28.581	28.903	28.646	28.587	28.398	**28.903**
SSIM	0.9066	0.8882	0.9027	0.9074	0.9054	0.9046	0.9042	**0.9074**

Table 4. The results of different stage numbers on Rain200H. The best results are highlighted in boldface.

Experiments	E9	E10	E11	E12	E13	E14	E15	E16	E17
1 Recurrent Stage	✓								
2 Recurrent Stage		✓							
3 Recurrent Stage			✓						
4 Recurrent Stage				✓					
5 Recurrent Stage					✓				
6 Recurrent Stage						✓			
7 Recurrent Stage							✓		
8 Recurrent Stage								✓	
9 Recurrent Stage									✓
PSNR	25.753	26.392	27.501	27.761	28.073	28.277	28.346	**28.903**	28.760
SSIM	0.8732	0.8817	0.8914	0.8949	0.8986	0.8991	0.9006	**0.9074**	0.9062

Analysis on Single Attentive Block. We adopted four different blocks, such as self-attention block [39] (SAN), attention fusion feature block [36] (AFF), self-calibrated block [17] (SC) and convolution attention block [32] (CA); When the residual attention position is fixed on $j = 1, 3, 5$, the experiment results proves that CA block can achieve better performance.

Analysis on Different Residual Attentive Connection. In order to explore the connection of residual attention output from rain space to image space, we conducted experiments on Rain200H to validate the effectiveness of different connections. We set four experiments on different connection location, namely $j = 1, 2, 3$, $j = 2, 3, 4$, $j = 3, 4, 5$ and $j = 1, 3, 5$. The experimental results show that the location $j = 1, 3, 5$ can achieve best results on Table 3.

Analysis on the Number of Recurrent Stage. As the number of recurrent stage increasing, the separation of a background layer and its rain streak layer

Fig. 5. Examples about the comparison of different unsupervised losses on $Real_{275}$.

tends to be obvious. In order to discover the optimal value of the recurrent stage number, we set the stage $T = 1, 2, 3, 4, 5, 6, 7, 8, 9$; The experiments results verified that the stage $T = 8$ achieve the best performance in terms of PSNR and SSIM, whose specific results are shown on Table 4.

Analysis on Unsupervised Losses. At the stage of semi-supervised training, in order to show the effectiveness of DWT discriminative loss, we conduct experiments with $Loss_{HH}$, $Loss_{HL}$, $Loss_{LH}$ and No-DWT that is our AID model. As shown in Table 5, we set four different experiments to verify the unsupervised losses effectiveness on $Real_{275}$ datasets. Our final results also confirm that calculating the adversial losses of f_{HH}, f_{HL} and f_{LH} between real-world clean image and the generated clean background are more beneficial to image restoration.

Table 5. The analysis on unsupervised losses.

Experiments	L_{HH}	L_{HL}	L_{LH}	AID	AID-DWT
PSNR	28.557	28.574	28.491	28.548	**28.903**
SSIM	0.9040	0.9030	0.9028	0.9042	**0.9074**

5 Conclusion

In this work, we proposed a semi-supervised approach with residual attention based on Haar wavelet transform to tackle image deraining, i.e., AID-DWT. We design two sets of multi-layer residual block combined with the LSTM network to divide the rainy image into streak layer space and image layer space, and connect the two spaces through the residual attention block to accelerate the convergence and removal of rain features in image layer space. Moreover, we simultaneously calculate the adversarial losses on f_{HH}, f_{HL} and f_{LH} between real-world clean image and restorated background image to better predict clean background image. Extensive experiments on synthetic and real-world benchmark datasets have validated the effectiveness of our AID-DWT, which quantitatively and qualitatively outperforms existing semi-supervised deraining methods

and state-of-the-art supervised deraining methods. In future work, the proposed semi-supervised framework has the potential to be extended to other relevant low-level vision tasks, e.g., blind image denoising.

References

1. Chen, Y., Hsu, C.: A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In: IEEE ICCV (2013)
2. Deng, S., et al.: Detail-recovery image deraining via context aggregation networks. In: IEEE CVPR, pp. 14548–14557 (2020)
3. Fan, Z., Wu, H., Fu, X., Hunag, Y., Ding, X.: Residual-guide feature fusion network for single image deraining. In: ACM MM, pp. 1751–1759 (2018)
4. Fu, X., Liang, B., Huang, Y., Ding, X., Paisley, J.: Lightweight pyramid networks for image deraining. In: IEEE TNNLS (2020)
5. Fu, X., Huang, J., Ding, X., Liao, Y., Paisley, J.: Clearing the skies: a deep network architecture for single-image rain removal. IEEE TIP **26**(6), 2944–2956 (2017)
6. Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., Paisley, J.: Removing rain from single images via a deep detail network. In: IEEE CVPR, pp. 1715–1723 (2017)
7. Fu, X., Liang, B., Huang, Y., Ding, X., Paisley, J.: Lightweight pyramid networks for image deraining. IEEE TNNLS **31**(6), 1–14 (2019)
8. Chen, H., et al.: Pre-trained image processing transformer. arXiv:2012.00364 (2021)
9. Huynh-Thu, Q., Ghanbari, M.: Scope of validity of PSNR in image/video quality assessment. Electron. Lett. **44**(13), 800–801 (2008)
10. Jiang, K., et al.: Multi-scale progressive fusion network for single image deraining. In: IEEE CVPR, pp. 8343–8352 (2020)
11. Kang, L., Lin, C., Fu, Y.: Automatic single-image-based rain streaks removal via image decomposition. In: IEEE TIP (2012)
12. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: ICLR (2015)
13. Li, G., He, X., Zhang, W., Chang, H., Dong, L., Lin, L.: Non-locally enhanced encoder-decoder network for single image de-raining. In: ACM MM, pp. 1056–1064 (2018)
14. 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.) ECCV 2018. LNCS, vol. 11211, pp. 262–277. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_16
15. Li, Y., Tan, R.T., Guo, X., Lu, J., Brown, M.S.: Rain streak removal using layer priors. In: IEEE CVPR, pp. 2736–2744 (2016)
16. Li, Y., Tan, R.T., Guo, X., Lu, J., Brown, M.S.: Rain streak removal using layer priors (2016)
17. Liu, J.J., Hou, Q., Cheng, M.M., Wang, C., Feng, J.: Improving convolutional networks with self-calibrated convolutions. In: IEEE CVPR, pp. 10093–10102 (2020)
18. Luo, Y., Xu, Y., Ji, H.: Removing rain from a single image via discriminative sparse coding. In: IEEE ICCV (2015)
19. Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE TPAMI **11**(7), 674–693 (1989)
20. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" Image quality analyzer. IEEE Sig. Process. Lett. **20**, 209–212 (2013)
21. Pan, J., Hu, Z., Su, Z., Yang, M.: l_0-regularized intensity and gradient prior for deblurring text images and beyond. In: IEEE TPAMI (2017)

22. Pang, B., Zhai, D., Jiang, J., Liu, X.: Single image deraining via scale-space invariant attention neural network. In: ACM MM, pp. 375–383 (2020)
23. Paszke, A., et al.: Automatic differentiation in pytorch. In: NIPS Autodiff Workshop The Future of Gradient-based Machine Learning Software and Techniques
24. Ren, D., Shang, W., Zhu, P., Hu, Q., Meng, D., Zuo, W.: Single image deraining using bilateral recurrent network. IEEE TIP **29**, 6852–6863 (2020)
25. Ren, D., Zuo, W., Zhang, D., Zhang, L., Yang, M.: Simultaneous fidelity and regularization learning for image restoration. IEEE TPAMI **43**(1), 284–299 (2019)
26. Ren, D., Zuo, W., Hu, Q., Zhu, P., Meng, D.: Progressive image deraining networks: a better and simpler baseline. In: IEEE CVPR (2019)
27. Wang, T., et al.: Spatial attentive single-image deraining with a high quality real rain dataset. In: IEEE CVPR (2019)
28. Yang, W., et al.: Deep joint rain detection and removal from a single image. In: IEEE CVPR, pp. 1357–1366 (2017)
29. Wang, C., Wu, Y., Su, Z., Chen, J.: Joint self-attention and scale-aggregation for self-calibrated deraining network. In: ACM MM, pp. 2517–2525 (2019)
30. Wang, C., Xing, X., Wu, Y., Su, Z., Chen, J.: DCSFN: deep cross-scale fusion network for single image rain removal. In: ACM MM, pp. 1643–1651 (2019)
31. Wei, W., Meng, D., Zhao, Q., Xu, Z., Wu, Y.: Semi-supervised transfer learning for image rain removal. In: IEEE CVPR, pp. 3872–3881 (2019)
32. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: CBAM: convolutional block attention module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 3–19. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_1
33. Li, X., et al.: Recurrent squeeze-and-excitation context aggregation net for single image deraining. In: ECCV, pp. 2736–2744 (2016)
34. Yang, Y., Lu, H.: Single image deraining via recurrent hierarchy enhancement network. In: ACM MM, pp. 1814–1822 (2019)
35. Yasarla, R., Sindagi, V.A., Patel, V.M.: Syn2real transfer learning for image deraining using gaussian processes. In: IEEE CVPR, pp. 2723–2733 (2020)
36. Dai, Y., Gieseke, F., Oehmcke, S., Wu, Y., Barnard, K.: Attentional feature fusion. In: IEEE WACV (2021)
37. Zhang, H., Patel, V.M.: Density-aware single image de-raining using a multi-stream dense network. In: IEEE CVPR, pp. 695–704 (2018)
38. Zhang, H., Sindagi, V., Patel, V.M.: Image de-raining using a conditional generative adversarial network. IEEE TCSVT **30**(11), 3943–3956 (2019)
39. Zhao, H., Jia, J., Koltun, V.: Exploring self-attention for image recognition. In: IEEE CVPR (2020)
40. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE TIP **13**(4), 600–612 (2004)
41. Zhu, H., Wang, C., Zhang, Y., Su, Z., Zhao, G.: Physical model guided deep image deraining. In: IEEE ICME, pp. 1–6 (2020)
42. Zhu, H., et al.: Singe image rain removal with unpaired information: a differentiable programming perspective. In: AAAI, pp. 9332–9339 (2019)
43. Zhu, J., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: IEEE ICCV, pp. 2242–2251 (2017)