

# DCSFN: Deep Cross-scale Fusion Network for Single Image Rain Removal

Cong Wang\*

Dalian University of Technology  
supercong94@gmail.com

Xiaoying Xing\*

Tsinghua University  
xing-xy17@mails.tsinghua.edu.cn

Yutong Wu

Dalian University of Technology  
ytongwu@mail.dlut.edu.cn

Zhixun Su

Dalian University of Technology  
zxsu@dlut.edu.cn

Junyang Chen†

University of Macau  
yb77403@umac.mo

## ABSTRACT

Rain removal is an important but challenging computer vision task as rain streaks can severely degrade the visibility of images that may make other visions or multimedia tasks fail to work. Previous works mainly focused on feature extraction and processing or neural network structure, while the current rain removal methods can already achieve remarkable results, training based on single network structure without considering the cross-scale relationship may cause information drop-out. In this paper, we explore the cross-scale manner between networks and inner-scale fusion operation to solve the image rain removal task. Specifically, to learn features with different scales, we propose a multi-sub-networks structure, where these sub-networks are fused via a cross-scale manner by Gate Recurrent Unit to inner-learn and make full use of information at different scales in these sub-networks. Further, we design an inner-scale connection block to utilize the multi-scale information and features fusion way between different scales to improve rain representation ability and we introduce the dense block with skip connection to inner-connect these blocks. Experimental results on both synthetic and real-world datasets have demonstrated the superiority of our proposed method, which outperforms over the state-of-the-art methods. The source code will be available at <https://supercong94.wixsite.com/supercong94>.

## CCS CONCEPTS

- Computing methodologies → Reconstruction.

## KEYWORDS

Image Rain Removal, Cross-scale Fusion, Inner-scale Connection, Gate Recurrent Unit

\*Both authors contributed equally to this research.

†The corresponding author is Junyang Chen.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

MM '20, October 12–16, 2020, Seattle, WA, USA.

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-7988-5/20/10...\$15.00

<https://doi.org/10.1145/3394171.3413820>

## ACM Reference Format:

Cong Wang, Xiaoying Xing, Yutong Wu, Zhixun Su, and Junyang Chen. 2020. DCSFN: Deep Cross-scale Fusion Network for Single Image Rain Removal. In *28th ACM International Conference on Multimedia (MM '20), October 12–16, 2020, Seattle, WA, USA..* ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3394171.3413820>

## 1 INTRODUCTION

Images with rain captured by outdoor surveillance equipment significantly degrade the performance of some existing computer vision systems and also result in poor visual experience for some multimedia applications. Hence, removing rain from a single image is important for these multimedia applications. In this paper, we aim to solve the image deraining problem.

A rainy image  $\mathbf{O}$  can be expressed as the linear sum between a background image  $\mathbf{B}$  and a rain layer  $\mathbf{R}$ :

$$\mathbf{O} = \mathbf{B} + \mathbf{R}, \quad (1)$$

Deraining is a highly ill-posed problem because there are numerous  $\mathbf{B}$  and  $\mathbf{R}$  pairs for a given rainy image. To restrain the solution space, many priors about the background image or rain layer are proposed [3, 16, 26, 31]. Although these methods solve the deraining task to some extent, they always fail to work when the priors are invalid due to the simpler assumption on shapes, directions, and sizes of rain. Hence, a more effective deraining approach, which can deal with various rain in most conditions, is needed.

Benefiting from stronger feature representation ability, Convolution Neural Networks (CNNs, also called Deep Learning) has achieved great success in many computer vision problems, e.g., object detection [41], object tracking [55], pose estimation [4], optical flow [36], semantic segmentation [30], deblurring [33], dehazing [1, 20, 38, 52], super-resolution [6, 7, 49] and also deraining [8–12, 21, 23, 25, 34, 35, 37, 39, 46, 47, 53, 54]. These deraining algorithms consider different manners to remove rain. Researchers design various deep network structures to solve the image deraining problem. To improve the computational efficiency and decrease model sizes, light-weight networks [8, 11] came out in turn. By learning rain mask or estimating rain density to guide the deraining process, rain detection and removal [46] and multi-stream network [53] are proposed and present a better deraining performance. And other various network modules are developed by designing kinds of mechanisms, e.g., channel attention by squeeze-and-excitation [13], pixel-wise attention using nonlocal mean network [40] and adversarial learning [23, 35, 54]. Although these deraining methods have achieved better performance, they still

neglect some important details. Previous works mainly focused on feature extraction and processing or neural network structure, while the current rain removal methods can already achieve remarkable results, training based on single network structure without considering the cross-scale relationship may cause information drop-out.

To address this problem, we propose a deep network with cross-scale learning architecture to remove rain from a single image. Firstly, we design an inner-scale connection block to fuse features with different scales by building the correlation between each scales so that better learn the rain features. Secondly, to maximize information flow and enable the computation of long-range spatial dependencies as well as efficient usage of the feature activation of proceeding layers, we introduce encoder and decoder with dense connection structure and skip connection to inner-connect these blocks. Lastly, we propose a cross-scale fusion network to learn features at different scales, where the proposed cross-scale manner connects features at different scales by Gate Recurrent Unit (GRU) which makes full use of information at different scales.

In summary, this paper has the following contributions:

- We propose an inner-scale connection block by building the correlation between each scale so that it better learns the rain features.
- We introduce encoder and decoder with dense connection structure and skip connection to inner-connect these blocks to improve the deraining performance.
- We propose a cross-scale fusion network to learn features at different scales, where the proposed cross-scale manner connects features at different scales by Gate Recurrent Unit that makes full use of information at different scales.
- Experimental results on both synthetic and real-world datasets have demonstrated the superiority of our proposed method, which outperforms over the state-of-the-art methods.

## 2 RELATED WORK

In this section, we review some image deraining methods and the multi-scale learning.

### 2.1 Single Image Deraining

Single image rain removal can be more difficult compared to video-based methods [2, 22, 28, 45]. Video-based deraining can be promoted by extracting the features from serial frames that contain correlated information, while single image deraining is appropriate for an independent single input so that it is more difficult to restore the background image without frame information. In this section, we only review the image deraining problem.

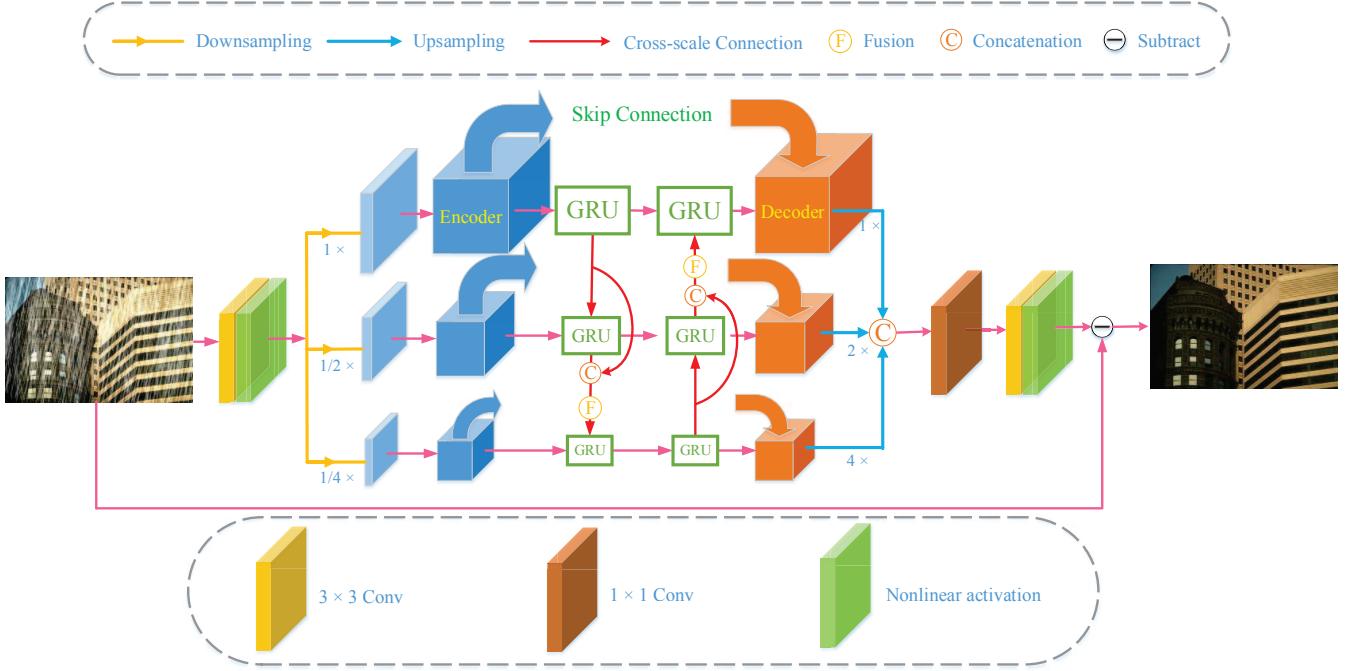
**Prior-Based Methods:** Many early works attempt to solve the problem with image priors [3, 16, 17, 26, 31, 51]. Considering that rain streaks dominant in high-frequency structure, Kang *et al.* [16] decompose the input image into high frequency and low-frequency layers, removing the high-frequency rain streaks by dictionary learning. Based on the observation that rain is sparse, Luo *et al.* [31] propose a discriminative sparse coding structure to separate the rain layer from a clean image. Moreover, since rain in a local patch is low-rank, some methods restrain the solution space from this point. Chen *et al.* [3] propose a low-rank representation-based method

which promotes the deraining performance by taking advantage of the low-rank model. In [17], they apply kernel regression to the deraining framework by using a non-local mean filter. Wang *et al.* [42] propose a hierarchical structure for single image deraining and desnowing. Zhu *et al.* [56] ingeniously consider rain direction and propose a joint optimization process.

**Deep learning-Based Methods:** Deep learning-based methods achieve considerably better performance than prior-based methods, which verifies the efficiency of deep learning on computer vision tasks. Fu *et al.* [9, 10] first apply deep learning to single image deraining. They manage to extract high-frequency structures via a guide filter and use a residual network to obtain the rain layer. The learned rain layer is used to obtain a clean image via Eq.1. Yang *et al.* [46] consider the hazy condition in some rainy situations and apply a dehazing-deraining-dehazing algorithm. They propose joint rain detection and removal method to deal with some complex situations. Li *et al.* [21] design a multi-scale non-local enhanced encoder-decoder network, which uses the pixel-wise attention mechanism to learn the residual rain. Considering the guide role, Fan *et al.* [8] design a light-weight residual-guide network for deraining in a stage-wise manner. Li *et al.* [25] believe spatial contextual information is important for single image deraining that they propose dilated convolution to capture more contextual information and remodel the rainy model by utilizing the squeeze-and-excitation operation. Zhang *et al.* [53] realize that some approaches have the performance of excessive deraining or residual rain. Based on this, they propose a multi-stream densely connected convolutional neural network and estimate rain density to guide the deraining procedure. Ren *et al.* [37] present a better baseline model by investigating the input, output, and loss function focus on network architecture and achieve better results than previous works. Wang *et al.* [39] design a spatial attentive network to remove rain streaks in a local-to-global manner. Li *et al.* [24] study and evaluate existing single image deraining methods elaborately. They develop a new large-scale benchmark consisted of both synthetic and real-world rainy images by considering diverse possible situations.

### 2.2 Multi-scale Learning

The hierarchy architecture can generate features of different scales that contain abundant layer information. Moreover, exploiting the inner correlations of multi-scale features can lead to a deeper insight into image layout and boost the feature extraction performance to a great extent. A representative architecture on multi-scale learning is proposed in [27], which is a top-down convolutional network structure that is applied lateral connections for extracting multi-scale semantic features for the object detection task. Inspired by the considerable performance on feature map construction, multi-scale learning had been applied on face image processing [29], super-resolution [19], and deraining [11][48] as well. Yang *et al.* [48] propose a recurrent hierarchy enhancement network considering the correlations between neighboring stages, where features extracted from the previous stage are transmitted to the later stage as guidance. Although some works consider the correlations of different convolutional layers, the cross-scale information is still ignored and it should be explored.



**Figure 1:** The proposed network. The network consists of three sub-networks with different scales. At each scale, they have the same structure with densely connected encoder and decoder, shown in Fig. 2. At the end of each encoder, features at different scales are fused by GRU via a cross-scale manner to make full use of information at different scales. At last, all features at different scales are fused to generate estimated rain and we gain the final estimated rain-free images via Eq. 1.

### 3 PROPOSED METHOD

We detail the proposed network in Sec. 3.1 and introduce the encoder and decoder in Sec. 3.2. The proposed Inner-scale Connection Block and applied loss function are described in Sec. 3.3 and Sec. 3.4, respectively.

#### 3.1 Proposed Network (DCSFN)

Our proposed network is illustrated in Fig. 1. The network consists of three subnetworks to learn information with different scales and they are fused via a cross-scale manner by Gate Recurrent Unit (GRU) at the end of each encoder to make full use of information at different scales. We obtain the estimated rain by fusing features with different scales in these sub-networks and generate the final estimated rain-free image via Eq. 1.

#### 3.2 Encoder and Decoder with Dense Connection

Dense connection is proposed in [14], which has achieved great success in many computer vision problems. The dense connection structure alleviates the vanishing-gradient problem, strengthens feature propagation, encourages feature reuse. One big advantage of dense connection is that it maximizes information and gradients flow throughout the network that makes it easy to train. Moreover, as each layer has direct access to the gradients from the loss function and the original input signal, it leads to implicit deep supervision

for these layers. In this paper, we design densely connected encoder and decoder with skip connection as the style of our network.

Mathematically, the encoder can be expressed as:

$$\mathbf{F}_l^E = \mathcal{T}\left(\mathcal{F}(C[\mathbf{F}_0^E, \mathbf{F}_1^E, \dots, \mathbf{F}_{l-1}^E])\right), l = 1, 2, \dots, L, \quad (2)$$

where  $\mathbf{F}_0^E$  denotes the input features and  $\mathbf{F}_l^E$  refers to the output of  $l^{th}$  layer at the encoder stage.  $C[\cdot]$  and  $\mathcal{F}(\cdot)$  denotes the concatenation at channel dimension and fusion operation, respectively. We use  $1 \times 1$  convolution as the fusion operation.  $\mathcal{T}(\cdot)$  refers to the proposed inner-scale connection block detailed in Sec. 3.3.

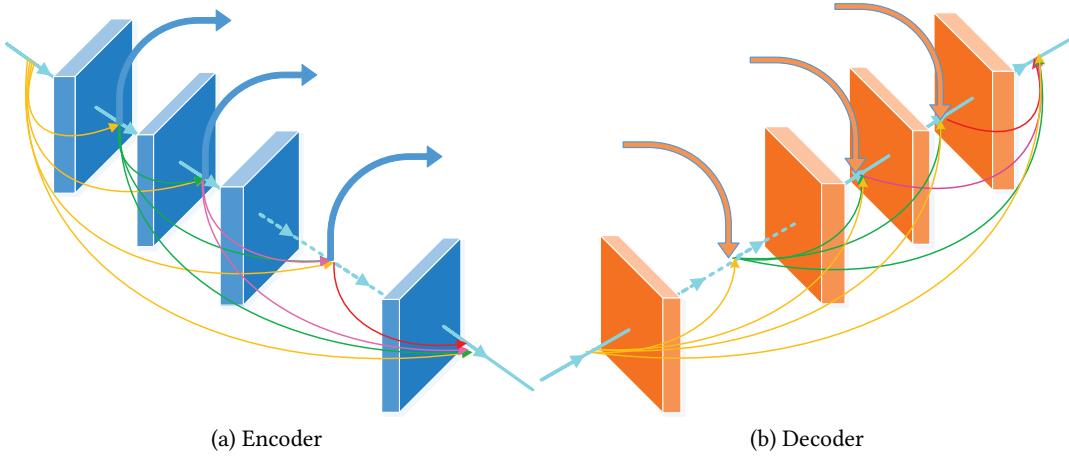
In the decoder stage, skip connection between Encoder and Decoder is utilized to enable the computation of long-range spatial dependencies as well as efficient usage of the feature activation of proceeding layers:

$$\mathbf{F}_l^D = \mathcal{T}\left(\mathcal{F}(C[\mathbf{F}_0^D, \mathbf{F}_1^D, \dots, \mathbf{F}_{l-1}^D])\right) + \mathbf{F}_l^E, l = 1, 2, \dots, L, \quad (3)$$

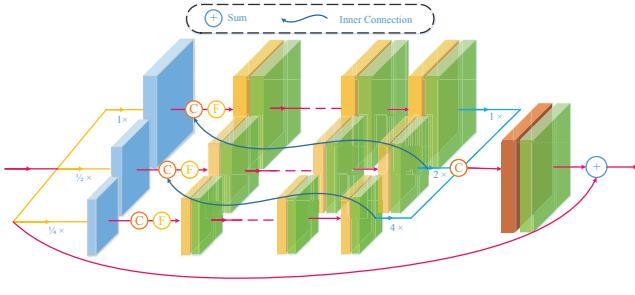
where  $\mathbf{F}_0^D$  denotes the GRU output of the current scale and  $\mathbf{F}_l^D$  refers to the output of  $l^{th}$  layer at the decoder stage.

#### 3.3 Inner-scale Connection Block

Multi-scale learning has been applied to a number of vision problems and obtains huge development. Although various multi-scale manners are proposed and designed, the inner correlation between



**Figure 2:** (a) Encoder. The encoder consists of a series of Inner-scale Connection Block (as illustrated in Fig. 3) with dense connection. (b) Decoder. The Decoder has the same structure with encoder and skip connection is utilized to enable the computation of long-range spatial dependencies as well as efficient usage of the feature activation of proceeding layers.



**Figure 3: Inner-scale Connection Block.** Firstly, *Global Max Pooling* is utilized to obtain features with different scales. Secondly, features after several convolutions are connected between different scales to boost the correlation information exploration. Lastly, all features at different scales are fused to learn main features.

different scales never is explored. In this paper, we explore the inner correlation between different scales by designing the proposed inner-scale connection block. The proposed block is shown in Fig. 3.

Firstly, we utilize *Global Max Pooling* operation to obtain multi-scale features:

$$\mathbf{P}_{\frac{1}{k}} = \mathcal{P}_k(x), k = 1, 2, \dots, 2^{K-1}, \quad (4)$$

where  $x$  is the input feature.  $K$  is the maximum scales we used.  $\mathcal{P}_k(\cdot)$  is *Global Max Pooling* operation with  $k \times k$  kernels and  $k \times k$  strides.  $\mathbf{P}_{\frac{1}{k}}$  is the output features at  $\frac{1}{k}$  scale.

Then, features at different scales are connected via inner way:

$$\tilde{\mathbf{P}}_{\frac{1}{k}} = \mathcal{F}\left(C[\tilde{\mathbf{P}}_{\frac{1}{2^k}}, \mathcal{U}_2(\mathbf{Y}_{\frac{1}{k}})]\right). \quad (5)$$

Here,  $\tilde{\mathbf{P}}_{\frac{1}{k}} = \mathbf{P}_{\frac{1}{k}}$  when  $k = 1$ .  $\mathbf{Y}_{\frac{1}{k}} = \mathcal{H}_n(\tilde{\mathbf{P}}_{\frac{1}{k}})$ .  $\mathcal{H}_n(\cdot)$  denotes  $3 \times 3$  convolution layer followed a nonlinear activation with  $n$  groups.  $\mathcal{U}_k$  denotes the Up-sampling operation with  $k \times$  factor.

At last, the features at different scales are fused and the residual learning is introduced:

$$\mathbf{Z} = \mathcal{F}\left(C[\mathbf{Y}_1, \dots, \mathcal{U}_{2^{K-1}}(\mathbf{Y}_{\frac{1}{2^{K-1}}})]\right) + x. \quad (6)$$

Here,  $\mathbf{Z}$  is the output of our proposed Inner-scale Connection Block.

### 3.4 Loss Function

SSIM loss has been proved to the best loss function in the deraining task [37]. In the network, we use it as our error loss:

$$\mathcal{L} = -\text{SSIM}(\hat{\mathbf{B}}, \mathbf{B}). \quad (7)$$

where  $\hat{\mathbf{B}}$  and  $\mathbf{B}$  are the de-rained image and the corresponding ground-truth background.

## 4 EXPERIMENTS RESULTS

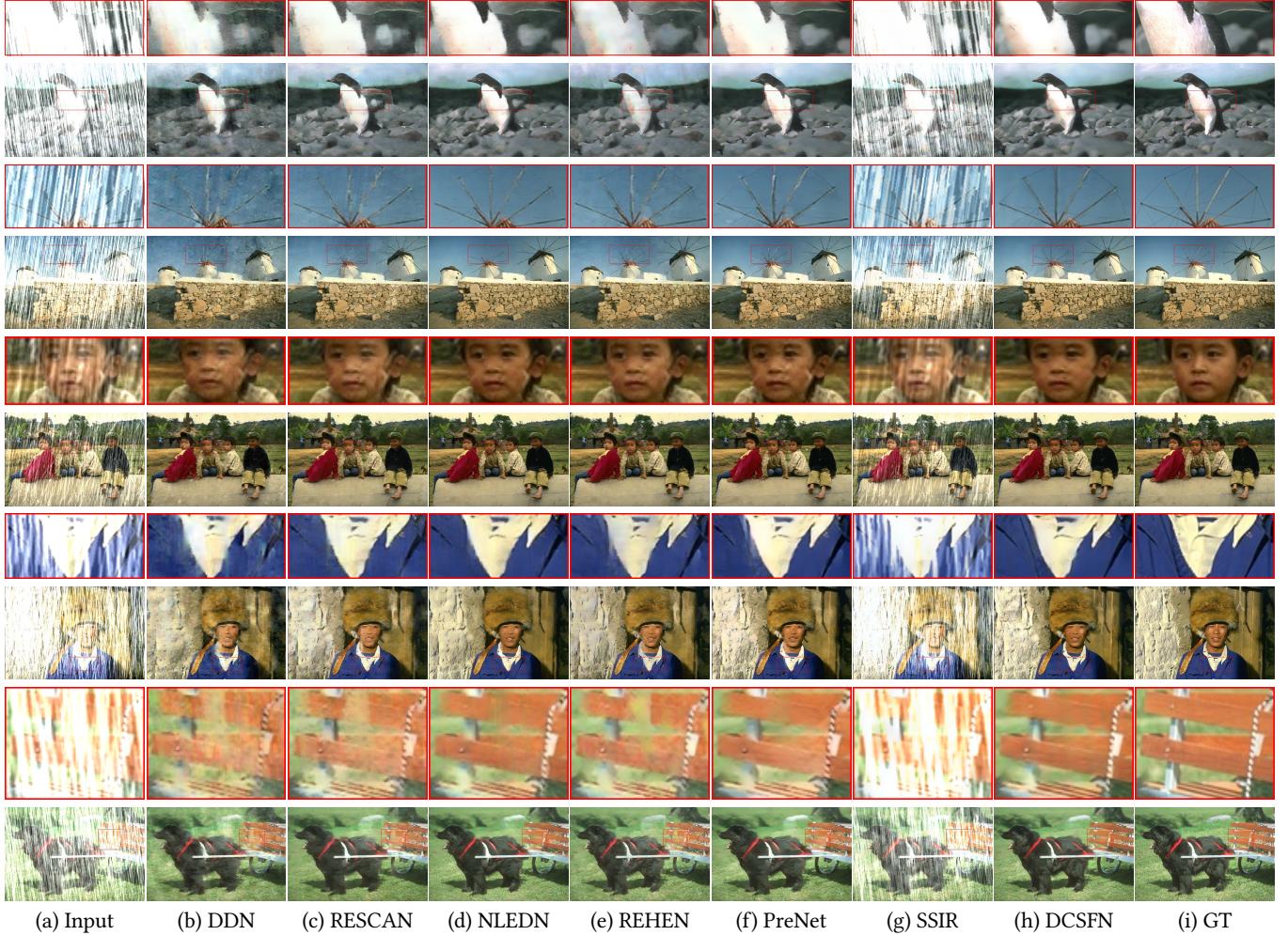
In this section, we conduct extensive experiments to demonstrate the effectiveness of the proposed method on four synthetic widely used datasets and a lot of real-world images. Six state-of-the-art methods are compared in this paper, DDN [10] (CVPR17), RESCAN [25] (ECCV18), NLEDN [21] (ACM MM18), REHEN [48] (ACM MM19), PreNet [37] (CVPR19), SSIR [44] (CVPR19). Next, we will introduce detailedly the datasets and measurements in Sec. 4.1, implementing details in Sec. 4.2, results on synthetic datasets in Sec. 4.3, results on real-world datasets in Sec. 4.4 and ablation study in Sec. 4.5.

### 4.1 Datasets and Measurements

**4.1.1 Synthetic Datasets.** We carry out experiments to verify the effectiveness of our network on synthetic datasets: Rain 100L [46], Rain 100H [46], Rain1200 [53]. Rain100L has light rain and therefore is relatively the easiest dataset. The dataset has 1800 image pairs for training and 200 image pairs for testing. Rain 100H has the same number of training and testing, while has heavy rain with different shapes, directions, and sizes and is also the most challenging dataset. Rain1200 dataset has 12000 images for training and 1200 images

**Table 1: Quantitative experiments evaluated on three synthetic datasets. The best results are highlighted in boldface.**

	DDN [10] CVPR'17		RESCAN [25] ECCV'18		NLEDN [21] ACM MM'18		REHEN [48] ACM MM'19		PreNet [37] CVPR'19		SSIR [44] CVPR'19		Ours	
Dataset	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Rain100H	22.26	0.69	25.92	0.84	28.42	0.88	27.52	0.86	27.89	0.89	22.47	0.71	<b>28.81</b>	<b>0.90</b>
Rain100L	34.85	0.95	36.12	0.97	38.84	0.98	37.91	0.98	36.69	0.98	32.37	0.92	<b>38.96</b>	<b>0.99</b>
Rain1200	30.95	0.86	32.35	0.89	32.98	0.92	32.51	0.91	32.38	0.92	29.32	0.89	<b>33.19</b>	<b>0.93</b>

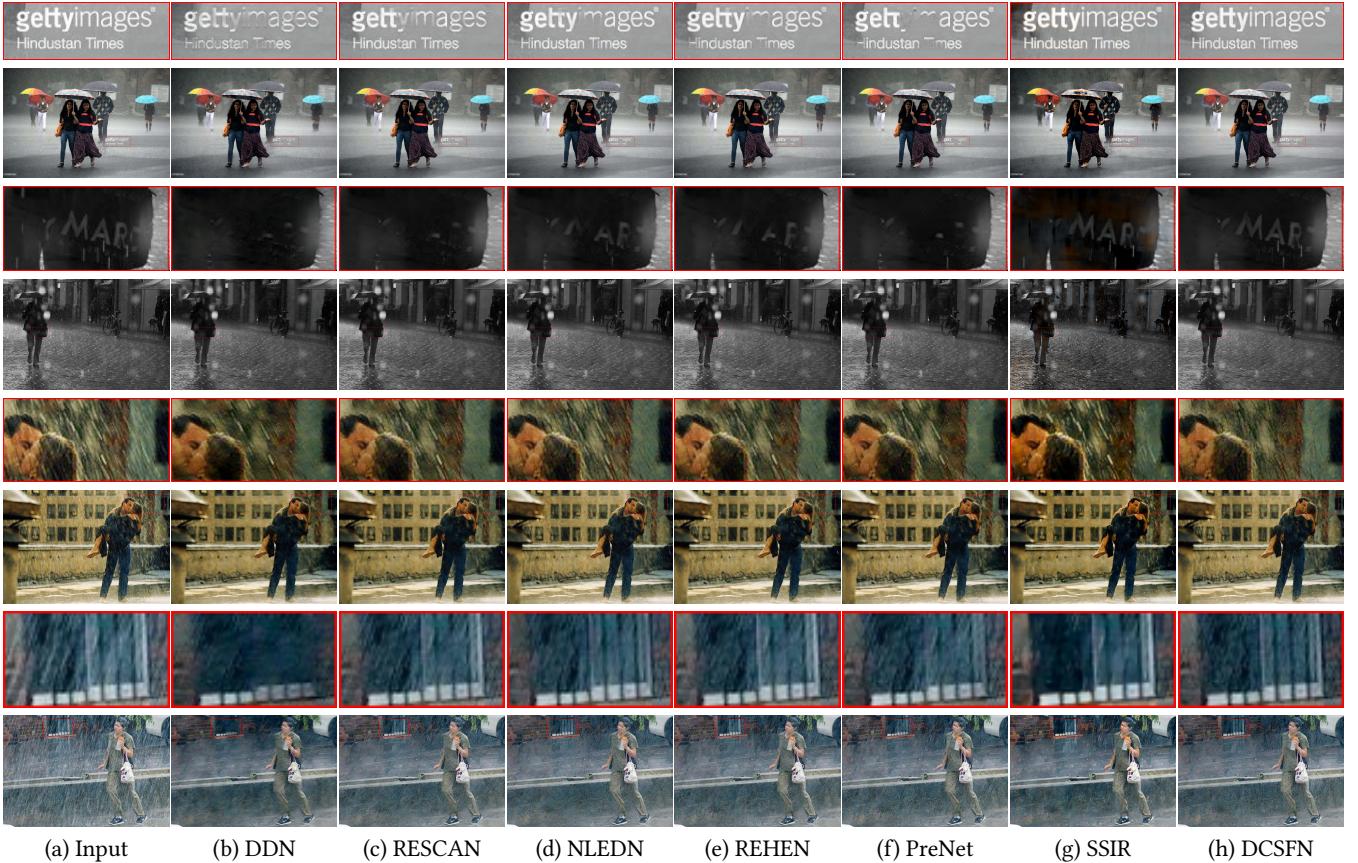


**Figure 4: The results from synthetic datasets. Our method restores better deraining results, while other method always hand down some artifacts or rain. Especially, SSIR [44] fails to restore rain-free images and its results are unacceptable.**

for testing. The testing set has different background images with training setting, assuring the experiment results are convincing.

**4.1.2 Real-world Datasets.** We collect some real-world rainy images from the Internet and select carefully some rainy images from [24]. We use them as our real-world dataset.

**4.1.3 Measurements.** To verify the de-raining performance we use peak signal to noise ratio (PSNR) [15] and structure similarity index (SSIM) [43]. The two measurements are used to compare the restored results with the corresponding ground-truth. We use the two measurements to evaluate the deraining quality on synthetic datasets. Since the performance on real-world images can



**Figure 5:** The results from real-world datasets. For the first two examples, our method can restore clearer texture information, especially in cropped boxes, while other methods blur the texture. For the last two examples, other methods hand down a number of rain, while our method is able to restore better rain-free images. Again, we note that the SSIR [44] is less effective to restore rain-free images.

not be evaluated in the same way due to without ground-truth of real-world rainy images, we evaluate the restored image visually. Moreover, we also use the Natural Image Quality Evaluator (NIQE) and visual comparisons to evaluate the performance of our method. The NIQE calculates the no ground-truth image quality score for the real-world hazy image. The lower the image quality is, the higher NIQE is.

## 4.2 Implementation Details

We set  $L = 16$ ,  $K = 3$  and  $n = 4$ . The number of channels is 32. We randomly select  $64 \times 64$  patch pairs from training image datasets as input. We use *LeakyReLU* with  $\alpha = 0.2$  as our non-linear activation function. The network is implemented using PyTorch and trained with Adam algorithm [18] using two NVIDIA 1080ti GPUs. We initialize the learning rate to 0.0005, and divide it by 10 at 480 epochs and 640 epochs, and stop the training at 800 epochs.

## 4.3 Results on Synthetic Datasets

We evaluate the proposed method on three widely used datasets: Rain100H, Rain100L, and Rain1200 in Tab. 1. Compared with other

**Table 2: The NIQE results. The best result is marked in bold.**

Methods	NIQE
DDN [10]	4.14
RESCAN [25]	3.98
NLEDN [21]	3.62
REHEN [48]	3.98
SSIR [44]	3.86
PreNet [37]	4.13
DCSFN (Ours)	<b>3.54</b>

state-of-the-art methods, we achieve the newest state-of-the-art results on these datasets. Especially, SSIR [44], a semi-supervised deraining network, is less effective on synthetic datasets.

We show some deraining examples on synthetic datasets in Fig. 4. We can observe that our method can restore better results from the global perspective and obtain clearer texture and cleaner background, especially in the cropped boxes. We note that the SSIR

**Table 3: Ablation study on different components. The ✓ symbol denotes that the corresponding component is adopted.**

Experiments	$M_1$	$M_2$	$M_3$	$M_4$
w/o Skip Connection	✓			
Skip Connection		✓	✓	✓
w/o Dense Connection		✓		
Dense Connection	✓		✓	✓
Single			✓	
Cross-scale	✓	✓		✓
PSNR	27.86	27.78	26.89	28.81
SSIM	0.8979	0.8898	0.8770	0.9040

fails to restore rain-free images, which illustrates their designed semi-supervised method cannot work on synthetic datasets.

#### 4.4 Results on Real-world Datasets

We verify the robustness on real-world datasets in Fig. 5. For the first two examples, our method can restore clearer texture information, especially in cropped boxes, while other methods blur the texture. For the last two examples, other methods hand down a number of rain, while our method is able to restore better rain-free images. Especially, we note that the SSIR [44] can not restore rain-free and always hand down lots of rain. The above examples illustrate our proposed method is a better deraining algorithm that can restore cleaner and clearer rain-free images and also can preserve better details and texture information.

We computer the NIQE values on our collected real-world datasets and the results are shown in Tab. 2. We observe that our method has lower NIQE value, which illustrates our method is able to generate more natural deraining results. This also demonstrates that the proposed method is a more robust derainer than others on real-world rainy conditions.

**Table 4: Discussion on Inner-scale Connection.**

Metric	w/o inner-scale Connection	w inner-scale Connection
PSNR	28.32	28.81
SSIM	0.9031	0.9040

#### 4.5 Ablation Study

We analyze the proposed model by conducting various experiments.

**4.5.1 Analysis on Network Designment.** We analyze the network designment that consists of skip connection, dense connection, single model, and cross-scale network. The results are illustrated in Tab. 3 and the abbreviation of these experiments are as follows:

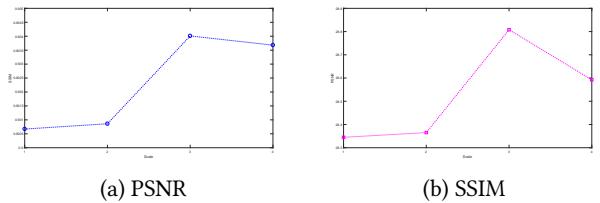
- $M_1$ : The proposed model without skip connection.
- $M_2$ : The proposed model without dense connection.
- $M_3$ : Single model that only consists of a single network.
- $M_4$ : Our proposed network with skip connection, dense connection, inner connection, and three sub-networks.

We can observe that both skip connection and dense connection have improvements for the deraining results. Compared to the single model that is a single deraining network with a single scale network, the proposed cross-scale manner has significant improvement for the deraining performance. This also shows our proposed cross-scale method is a better designment that a remarkable promotion to the deraining results.

**4.5.2 Discussion on Inner-scale Connection.** In this paper, we propose an inner-scale connection manner that improves the correlation between different scales. Whether the inner-connection is meaningful for the deraining result is worth discussing. We show the results in Tab. 4 and we can see that the designed inner-scale connection can improve the deraining performance. This demonstrates that our proposed inner-scale connection manner is meaningful.

**Table 5: The analysis of the number of scale (K).**

Metric	K = 1	K = 2	K = 3	K = 4
PSNR	28.34	28.35	28.81	28.59
SSIM	0.9006	0.9007	0.9040	0.9036



**Figure 6: Curve graph about the number of scale.**

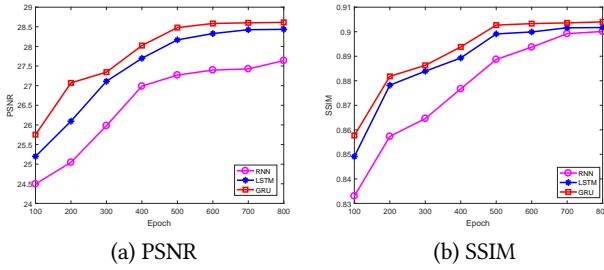
**4.5.3 The analysis of the number of scales.** We provide an analysis of the number of scales in the inner-scale connection block. The results are given in Tab. 5 Compare the single scale, the multi-scale manner improves deraining results and it reaches the best performance in terms of PSNR and SSIM when  $K = 3$ . So we select  $K = 3$  as the number of the scale. To better see the effect of the scale number on the results, we provide the curve graph about

the number of scales in Fig. 6. We can see that it has a significant improvement when  $K = 3$ .

**4.5.4 Analysis of recurrent unit.** There are three recurrent units can be selected: ConvRNN [32], ConvGRU [5] and ConvLSTM [50]. It needs to explore which the best recurrent unit is. We conduct three experiments on recurrent units, as shown in Tab. 6 and Fig. 7. We can see that the ConvGRU performs better than other recurrent units: ConvLSTM and ConvRNN. So, we select ConvGRU as our default recurrent unit.

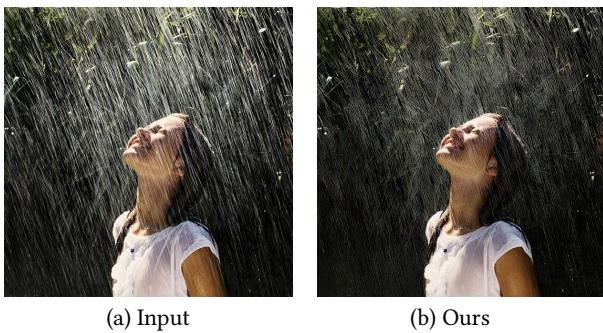
**Table 6: Analysis on recurrent unit.**

Metric	RNN	LSTM	GRU
PSNR	27.63	28.43	28.81
SSIM	0.9001	0.9017	0.9040



**Figure 7: Curve graph about on recurrent unit.**

**4.5.5 Failure Case.** We show a failure example in Fig. 8. We can see that our method can not handle the complex rainy image with heavy rain streaks. This may be that the samples of the synthetic dataset can not model the real-world image with other rainy condition so that the proposed model trained on synthetic datasets are not robust to real-world rainy images. This also stimulates us to propose a more robust derainer in the future. Maybe we can design a trainable model on real-world images to learn different rainy conditions.



**Figure 8: A failure example.**

## 5 CONCLUSION

In this paper, we propose a cross-scale fusion network (called DCSFN for short) to solve the image deraining problem. Different from other multi-scale methods, an inner-scale connection manner has been explored and verified to improve the deraining performance. Densely connected encoder and decoder with skip connections have been developed to maximize information flow and to enable the computation of long-range spatial dependencies as well as efficient usage of the feature activation of proceeding layers and boost the rain representation of the network. A cross-scale fusion manner between subnetworks with different scales is proposed to inner-connect and make full use of information at different scales. Quantitative and qualitative experimental results demonstrate the superiority of the proposed method compared with several state-of-the-art deraining methods on Rain100H, Rain100L, and Rain1200 datasets.

## REFERENCES

- [1] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. 2016. DehazeNet: An End-to-End System for Single Image Haze Removal. *IEEE TIP* 25, 11 (2016), 5187–5198.
- [2] Jie Chen, Cheen-Hau Tan, Junhui Hou, Lap-Pui Chau, and He Li. 2018. Robust Video Content Alignment and Compensation for Rain Removal in a CNN Framework. In *CVPR*. 6286–6295.
- [3] Jie-Li Chen and Chiou-Ting Hsu. 2013. A Generalized Low-Rank Appearance Model for Spatio-temporally Correlated Rain Streaks. In *ICCV*. 1968–1975.
- [4] Yilun Chen, Zhicheng Wang, Yuxiang Peng, Zhiqiang Zhang, Gang Yu, and Jian Sun. 2018. Cascaded Pyramid Network for Multi-Person Pose Estimation. In *CVPR*. 7103–7112.
- [5] Kyunghyun Cho, Bart van Merriënboer, Çağlar Gülcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In *EMNLP*. 1724–1734.
- [6] Zhen Cui, Hong Chang, Shiguang Shan, Bineng Zhong, and Xilin Chen. 2014. Deep Network Cascade for Image Super-resolution. In *ECCV*. 49–64.
- [7] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. 2016. Image Super-Resolution Using Deep Convolutional Networks. *IEEE TPAMI* 38, 2 (2016), 295–307.
- [8] Zhiwei Fan, Huafeng Wu, Xueyang Fu, Yue Huang, and Xinghao Ding. 2018. Residual-Guide Network for Single Image Deraining. In *ACM MM*. 1751–1759.
- [9] Xueyang Fu, Jiabin Huang, Xinghao Ding, Yinghao Liao, and John Paisley. 2017. Clearing the Skies: A Deep Network Architecture for Single-Image Rain Removal. *IEEE TIP* 26, 6 (2017), 2944–2956.
- [10] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. 2017. Removing Rain from Single Images via a Deep Detail Network. In *CVPR*. 1715–1723.
- [11] X. Fu, B. Liang, Y. Huang, X. Ding, and J. Paisley. 2019. Lightweight Pyramid Networks for Image Deraining. *IEEE TNNLS* (2019), 1–14.
- [12] Xueyang Fu, Qi Qi, Yue Huang, Xinghao Ding, Feng Wu, and John W. Paisley. 2018. A Deep Tree-Structured Fusion Model for Single Image Deraining. *CoRR* abs/1811.08632 (2018).
- [13] Jie Hu, Li Shen, and Gang Sun. 2018. Squeeze-and-Excitation Networks. In *CVPR*. 7132–7141.
- [14] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. 2017. Densely Connected Convolutional Networks. In *CVPR*. 2261–2269.
- [15] Q. Huynh-Thu and M. Ghanbari. 2008. Scope of validity of PSNR in image/video quality assessment. *Electronics Letters* 44, 13 (2008), 800–801.
- [16] Li-Wei Kang, Chia-Wen Lin, and Yu-Hsiang Fu. 2012. Automatic Single-Image-Based Rain Streaks Removal via Image Decomposition. *IEEE TIP* 21, 4 (2012), 1742–1755.
- [17] Jin-Hwan Kim, Chul Lee, Jae-Young Sim, and Chang-Su Kim. 2013. Single-image deraining using an adaptive nonlocal means filter. In *ICIP*. 914–917.
- [18] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *ICLR*.
- [19] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. 2019. Fast and Accurate Image Super-Resolution with Deep Laplacian Pyramid Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 41, 11 (2019), 2599–2613.
- [20] Boyi Li, Xiulan Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. 2017. AOD-Net: All-in-One Dehazing Network. In *ICCV*. 4780–4788.
- [21] Guanbin Li, Xiang He, Wei Zhang, Huiyou Chang, Le Dong, and Liang Lin. 2018. Non-locally Enhanced Encoder-Decoder Network for Single Image De-raining.

- In ACM MM. 1056–1064.
- [22] Minghan Li, Qi Xie, Qian Zhao, Wei Wei, Shuhang Gu, Jing Tao, and Deyu Meng. 2018. Video Rain Streak Removal by Multiscale Convolutional Sparse Coding. In *CVPR*. 6644–6653.
- [23] Ruoteng Li, Loong-Fah Cheong, and Robby T. Tan. 2019. Heavy Rain Image Restoration: Integrating Physics Model and Conditional Adversarial Learning. In *CVPR*. 1633–1642.
- [24] Siyuan Li, Iago Brejo Araujo, Wenqi Ren, Zhangyang Wang, Eric K. Tokuda, Roberto Hirata Junior, Roberto Cesar-Junior, Jiawan Zhang, Xiaoqie Guo, and Xiaochun Cao. 2019. Single Image Deraining: A Comprehensive Benchmark Analysis. In *CVPR*. 3838–3847.
- [25] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. 2018. Recurrent Squeeze-and-Excitation Context Aggregation Net for Single Image Deraining. In *ECCV*. 262–277.
- [26] Yu Li, Robby T. Tan, Xiaoqie Guo, Jiangbo Lu, and Michael S. Brown. 2016. Rain Streak Removal Using Layer Priors. In *CVPR*. 2736–2744.
- [27] Tsung-Yi Lin, Piotr Dollár, Ross B. Girshick, Kaiming He, Bharath Hariharan, and Serge J. Belongie. 2017. Feature Pyramid Networks for Object Detection. In *CVPR*. 936–944.
- [28] Jiaying Liu, Wenhua Yang, Shuai Yang, and Zongming Guo. 2018. Erase or Fill? Deep Joint Recurrent Rain Removal and Reconstruction in Videos. In *CVPR*. 3233–3242.
- [29] Zhilei Liu, Yunpeng Wu, Le Li, Cuicui Zhang, and Baoyuan Wu. 2020. Joint Face Completion and Super-resolution using Multi-scale Feature Relation Learning. *CoRR* abs/2003.00255 (2020). arXiv:2003.00255 <https://arxiv.org/abs/2003.00255>
- [30] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *CVPR*. 3431–3440.
- [31] Yu Luo, Yong Xu, and Hui Ji. 2015. Removing Rain from a Single Image via Discriminative Sparse Coding. In *ICCV*. 3397–3405.
- [32] Danilo P. Mandic and Jonathon Chambers. 2001. Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability. *Adaptive Learning Systems for Signal Processing Communications Control* (2001).
- [33] Janne Mustaniemi, Juho Kannala, Simo Särkkä, Jiri Matas, and Janne Heikkilä. 2018. Inertial-aided Motion Deblurring with Deep Networks. In *CoRR*, Vol. abs/1810.00986. arXiv:1810.00986
- [34] Jinshan Pan, Sifei Liu, Deqing Sun, Jiawei Zhang, Yang Liu, Jimmy S. J. Ren, Zechao Li, Jinhui Tang, Huchuan Lu, Yu-Wing Tai, and Ming-Hsuan Yang. 2018. Learning Dual Convolutional Neural Networks for Low-Level Vision. In *CVPR*. 3070–3079.
- [35] Jinshan Pan, Yang Liu, Jiangxin Dong, Jiawei Zhang, Jimmy S. J. Ren, Jinhui Tang, Yu-Wing Tai, and Ming-Hsuan Yang. 2018. Physics-Based Generative Adversarial Models for Image Restoration and Beyond. *CoRR* abs/1808.00605 (2018). arXiv:1808.00605
- [36] Anurag Ranjan and Michael J. Black. 2017. Optical Flow Estimation Using a Spatial Pyramid Network. In *CVPR*. 2720–2729.
- [37] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. 2019. Progressive Image Deraining Networks: A Better and Simpler Baseline. In *CVPR*. 3937–3946.
- [38] Wenqi Ren, Si Liu, Hua Zhang, Jin-shan Pan, Xiaochun Cao, and Ming-Hsuan Yang. 2016. Single Image Dehazing via Multi-scale Convolutional Neural Networks. In *ECCV*. 154–169.
- [39] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson W. H. Lau. 2019. Spatial Attentive Single-Image Deraining With a High Quality Real Rain Dataset. In *CVPR*. 12270–12279.
- [40] Xiaolong Wang, Ross B. Girshick, Abhinav Gupta, and Kaiming He. 2017. Non-local Neural Networks. In *CoRR*, Vol. abs/1711.07971. arXiv:1711.07971
- [41] Xiaolong Wang, Abhinav Shrivastava, and Abhinav Gupta. 2017. A-Fast-RCNN: Hard Positive Generation via Adversary for Object Detection. In *CVPR*. 3039–3048.
- [42] Yinglong Wang, Shuaicheng Liu, Chen Chen, and Bing Zeng. 2017. A Hierarchical Approach for Rain or Snow Removing in a Single Color Image. *IEEE TIP* 26, 8 (2017), 3936–3950.
- [43] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE TIP* 13, 4 (2004), 600–612.
- [44] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. 2019. Semi-Supervised Transfer Learning for Image Rain Removal. In *CVPR*. 3877–3886.
- [45] Wenhua Yang, Jiaying Liu, and Jiashi Feng. 2019. Frame-Consistent Recurrent Video Deraining With Dual-Level Flow. In *CVPR*.
- [46] Wenhua Yang, Robby T. Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. 2017. Deep Joint Rain Detection and Removal from a Single Image. In *CVPR*. 1685–1694.
- [47] W. Yang, R. T. Tan, J. Feng, J. Liu, S. Yan, and Z. Guo. 2019. Joint Rain Detection and Removal from a Single Image with Contextualized Deep Networks. *IEEE TPAMI* (2019).
- [48] Youzhao Yang and Hong Lu. 2019. Single Image Deraining via Recurrent Hierarchical Enhancement Network. In *ACM MM*. 1814–1822.
- [49] Jiahui Yu, Yuchen Fan, Jianchao Yang, Ning Xu, Zhaowen Wang, Xincho Wang, and Thomas Huang. 2018. Wide Activation for Efficient and Accurate Image Super-Resolution. *CoRR* abs/1808.08718. arXiv:1808.08718
- [50] Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals. 2014. Recurrent Neural Network Regularization. *CoRR* abs/1409.2329 (2014). arXiv:1409.2329
- [51] He Zhang and Vishal M. Patel. 2017. Convolutional Sparse and Low-Rank Coding-Based Rain Streak Removal. In *WACV*. 1259–1267.
- [52] He Zhang and Vishal M. Patel. 2018. Densely Connected Pyramid Dehazing Network. In *CVPR*. 3194–3203.
- [53] He Zhang and Vishal M. Patel. 2018. Density-Aware Single Image De-Raining Using a Multi-Stream Dense Network. In *CVPR*. 695–704.
- [54] H. Zhang, V. Sindagi, and V. M. Patel. 2019. Image De-raining Using a Conditional Generative Adversarial Network. *IEEE TCSVT* (2019).
- [55] Yunhua Zhang, Lijun Wang, Jinqing Qi, Dong Wang, Mengyang Feng, and Huchuan Lu. 2018. Structured Siamese Network for Real-Time Visual Tracking. In *ECCV*. 355–370.
- [56] Lei Zhu, Chi-Wing Fu, Dani Lischinski, and Pheng-Ann Heng. 2017. Joint Bi-layer Optimization for Single-Image Rain Streak Removal. In *ICCV*. 2545–2553.