SDNet: A Novel Separable Structure for Single Image Deraining

Xuheng Chen
Key Laboratory of Underwater Acoustic Communication and Marine Information Technology, School of Informatics, Xiamen University, 361005, Fujian, China
chenxuheng1995@stu.xmu.edu.cn

Abstract. Rain streaks degrade visual quality of images and severely affect the performance of computer vision system. Recently, deep convolutional neural networks (DCNNs) have shown the potential in single image deraining. However, their complex structures with large-scale of parameters lead to strong requirement of large storage, which still limits the applications of deraining systems. In this paper, we propose a separable deraining network (SDNet) which provides a more flexible and simply baseline of deraining network by using a cascaded framework to achieve automatic trade-off between model performance and computational resources. Besides, a novel module called group-atrous spatial pyramid (G-ASP) and a network separation strategy (NSS) are proposed to extract more discriminative multi-scale information with fewer parameters and adapts our model to different levels rain-streaks for reasonable allocation of computational resources automatically. The experimental results on public datasets demonstrate that our SDNet achieves promising performances with much fewer parameters than other methods.

1. Introduction
Rain streaks degrade the visual quality and then severely affect the performance of outdoor computer vision tasks such as object detection, segmentation, video surveillance [1]. Thus, effective methods for removing rain streaks are required for a wide range of practical applications, which aims at restoring clean background image from a rainy image.

Several methods based on traditional optimization separate rain streaks from object details by the use of low-level image features [2, 3]. However, it is difficult to remove the rain streaks and preserve structure information simultaneously, when the object and rain streak have similar structures and orientation. Fortunately, inspired by the unprecedented success of deep convolutional networks (DCNNs) in low level vision tasks [3, 4, 5]. Fu et al. [6, 7] utilized a relative shallow network to remove rain streaks which brought great performance promotion than traditional methods. Zhang et al. [8] proposed an algorithm based on GAN apply to the deraining and [9] used low-rank coding-based method to remove rain streak.

However, existing models for single image rain streaks removal tend to design a complex model to learn different modalities and intensities negative residual with tones of parameters which is wasteful to utilize a resource-hungry model to meet diverse rainy conditions. The practical applications require that models should follow: light-weight, adaptive ability and performance simultaneously.

To address above problems, we propose a separable deraining network (SDNet) with cascaded structure. The SDNet consists of cascaded several residual learning blocks and each block can be regarded as an independent rain steaks removal network. The visual results are shown in Fig.1.
We further propose a network separation strategy (NSS) to help SDNet to be able to meet diverse rainy conditions at test phase automatically. Investigations show that representing features at multiple scales is of great importance for numerous vision tasks [10]. Therefore, we propose a novel module for multi-scale features extraction called group-atrous spatial pyramid (G-ASP) which adopts group strategy and introduces atrous convolutions with different rates to gain stronger scale adaptability. Our contributions can be summarized as follows:

(1): We build a single image deraining network with both light-weight and promising performance, which strongly reduces the storage requirement in practical applications.

(2): A network separation strategy (NSS) is proposed at the testing procedure, which achieves adaptability across different rainy conditions.

(3): We propose a novel multi-scale features extraction module which is called G-ASP to enhance modeling ability to extract discriminative features and a parameter reuse strategy is adopted to G-ASP to reduce model complexity while maintain a decent performance.

2. Methods
In this part, we first describe the structure of G-ASP and its parameter reuse strategy. Then, we further propose NSS. Finally, the overall architecture of SDNet and loss functions are discussed.

2.1. Group-atrous Spatial Pyramid
In recent years, improvement of model performance brought by multi-scale information has attracted attention [10]. To obtain more abundant multi-scale features, we propose a novel atrous spatial pyramid (G-ASP) in Fig.2.

The original features are first expanded double size on channel by 1 × 1 convolutional kernel. Then, we evenly split new features into 4 groups, denoted as $x_i$, where $i \in \{1,2,3,4\}$. Each feature group $x_i$ has same spatial size but 1/4 number of channels compared the original features. Accordingly, each $x_i$ has $j$ atrous convolutions, denoted as $f_{i,j}(\cdot)$ except $x_1$, where $j \in \{1,2,3\}$. We present $y_i$ as the output of groups. Then, $y_i$ can be calculated as Eq.1.

$$ y_i = \begin{cases} x_i & i = 1, \\ f_{i,j}(x_i) & i = 2, j = 1, \\ f_{i,j}(f_{i,j}(x_i + y_{i-1}) + y_{i-1}) & i = 3, j \in \{1,2\}, \\ f_{i,j}(f_{i,j}(f_{i,j}(x_i + y_{i-1}) + y_{i-1}) + y_{i-1}) & i = 4, j \in \{1,2,3\}. \end{cases} $$ (1)
Where $f_i(x_i), i \in \{2,3,4\}$ represents the function mapping of atrous convolutions with rate of 1, 2, 3, respectively. The output of G-ASP can be described as $y_{\text{out}} = \text{concat}(y_1, y_2, y_3, y_4)$. The function of $\text{concat}(\cdot)$ denotes concatenation operation. We can find each group has different number of layers whereas the convolutional kernels share same size, by which we can maintain a steady increase of receptive field on same group while enlarging feature diversities from different groups to extract higher quality features with abundant multi-scale information. Meanwhile, the hierarchical information connections (HIC) among different groups (as shown by the red lines in Fig. 2) are also built. Thus, the multi-scale features from different groups can achieve hierarchical feature fusion and information guidance along these connections.

![Residual learning block](image)

**Fig. 2. Group-atrous spatial pyramid (G-ASP)**

In addition, we further propose a parameter reuse strategy, of which atrous convolutions in same group share parameters. Moreover, the recursive computation method [5] is also adopted. In this paper, we set each G-ASP as a recursive block, the inference of G-ASP at recursion $t$ can be formulated as $x_{t+1} = H(x_t)$, where $H(\cdot)$ denotes the function mapping of each recursive block, $t$ is recursion stage and we set recursion as 5 in each forward propagation. Network parameters are reused across different stages to reduce the model complexity significantly.

**Algorithm 1:** Our proposed algorithm for adaptive network separation, see section 3.2 for details

```
Require: $I$ : the input images
Require: $O$ : the output images
Require: $\text{SSIM} (I, O)$ : the function of SSIM
Require: $V$ : the threshold of activation value
Require: $f_t(I)$ : the function mapping of $\text{block}_t$
Require: $t$ : the index of block

$t \leftarrow t_{\text{init}}$
while $\text{SSIM}(f_{t+1}(I) - f_t(I)) < V$ do
    $t \leftarrow t + 1$
end while

return $f_{t+1}(I), t + 1$
```
2.2. Progressive Network Separation Strategy
In order to meet the requirements of practical application, we further propose a network separation strategy (NSS) which aims at adjusting the model structure to adaptive diverse rainy conditions in test phase according to a particular activation threshold. In image processing field, SSIM [11] is a widely used index to evaluate image quality which mainly considers the image similarity in brightness, contrast and structure. Thus, we hereby choose SSIM value to construct activation threshold for network separation in Algorithm 1. We consider that there existing an obvious structural difference between the rainy images and deraining images which is reflected on SSIM value. Thus, if the increasement of blocks do not brings expected SSIM value fluctuations, it means that current network has already reach” effective capacity”, which causes less pays off with the increasing blocks budget.

2.3. Overview Architecture
The overall architecture of SDNet is shown in Fig.3 and Fig.4. Our proposed SDNet is cascaded by several residual learning blocks with repetitive structure. Fig.5 indicates that SDNet will obtain a more eye-pleasing reconstruction and higher PSNR and SSIM as block goes deeper.

![Fig. 3. Network architecture of our proposed SDNet](image)

![Fig. 4. Block(n) of our proposed SDNet](image)

![Fig. 5. Comparisons under rainy conditions using our SDNet, (a) is clean images. (b) is the synthetic rainy images (c) - (g) are the results from block 1 to block 5 of SDNet under heavy rainy conditions, respectively.](image)
means rainy images and deraining images, respectively. We use only one convolutional layer to constitute initial feature extraction, as shown in the formula $F_{\text{init}} = \text{Conv}(I_R)$, where $F_{\text{init}}$ means initial features and $\text{Conv}(\cdot)$ denotes convolution operator. Then, $F_{\text{init}}$ is used to extract multi-scale features by G-ASP, as shown in the formula $F_{\text{multi-scale}} = H_{\text{multi-scale}}(F_{\text{init}})$, where $H_{\text{multi-scale}}(\cdot)$ represents the function mapping of G-ASP. Finally, the $I_D$ is obtain by reconstructional layer, as shown in the formula $I_D = H_{\text{rec}}(F_{\text{multi-scale}})$, where $H_{\text{rec}}$ denotes the reconstruction part which has only one convolutional layer. In addition, we use dense connections between shallow blocks and deep blocks to compensate the information loss brought by cascaded scheme. The features produced by G-ASP from shallow blocks are concatenated into the corresponding input of G-ASP in deep blocks. Therefore, the deeper blocks can merge features extracted by all previous blocks to predict finer estimation. Considering most methods optimized with Euclidean distance generate blurry predictions since the per-pixel losses do not close to perceptual difference between output and ground-truth as human visual perception [12]. In particular, it is hard to distinguish between the rain streaks and objects structure relying on $L_2$ loss function solely when the rain streaks are blend with object edges and background scene. We adopt the combination of $L_2$ loss and SSIM loss [11] to preserve global structure better as well as maintaining per-pixel similarity in each block. Meanwhile, global supervision is introduced in each block by minimize the loss collection of all blocks. The overall loss function for $\text{Block}_k$ is defined as follow:

$$L_{\text{Block}_k} = L_{\text{MSE}_k} + \lambda \cdot L_{\text{SSIM}_k}$$  \hspace{1cm} (2)

where $\lambda$ is a hyper-parameter to balance the MSE loss and SSIM loss, $\lambda$ is set as 1 via cross-validation experiments. We can hereby calculate global loss function as Eq. 3.

$$L_{\text{Sum}} = \sum_{k=1}^{M} \omega_k \cdot L_{\text{Block}_k}$$  \hspace{1cm} (3)

The global loss $L_{\text{Sum}}$ is a loss collection of all blocks, $M$ is the number of blocks, $\omega_k$ is hyper-parameter to balance every $L_{\text{Block}_k}$ and we set $\omega_k$ as 1.

3. Experiments

3.1. Implementation Details

All experiments are conducted by using TensorFlow for the python environment on a NVIDIA GeForce GTX 1080 with 8GB GPU memory, we use the Xavier method to initialize the network parameters and the RMSprop optimizer is chosen. The initial learning rate is set to 0.001, the batch size is set to 16, 50000 iterations of training are required for training process and we set all kernels size of $3 \times 3$ and each convolution layer has 16 kernels. During training stage, we randomly generate 0.8 million clean and rainy image pairs with size of $128 \times 128$. In this paper, SDNet with 5 blocks as it’s conventional form for experiments. We use the Rain100H and Rain100L for training and test respectively. Rain100H and Rain100L are two synthetic datasets with heavy and light rain streaks in [13] and each of them has 100 rainy images for test.

3.2. Evaluation on Synthetic Dataset

Effects of G-ASP: We assess the impact of G-ASP. A network which uses one convolutional layer to replace G-ASP (set as SDNet_O) is trained as the baseline. Meanwhile, we set experiments to quantitative analysis of hierarchical information connections (HIC). Table 1 shows the atrous pyramid and HIC are essential parts for G-ASP which brings distinct improvement on PSNR [14] and SSIM [11]. We arguably demonstrate that multi-scale features and feature fusion brought by G-ASP possess positive affects for deraining task.
Effects of NSS: Table 2 shows that under most light rainy conditions, a light-weight (3 blocks) SDNet is able to construct a good result and the further increase of blocks do not bring obvious performance improvement. But for heavy rainy conditions, require a more powerful network to get better reconstructions.

Quantitative results on synthetic dataset: We compare our models (5 blocks) with several state-of-the-art deep and non-deep algorithms on synthetic in Table 3. The negative sign in the last column of Table 3 indicates the number of parameters not mentioned in the relevant paper. Compared with most other methods, our SDNet achieve comparable PSNR and SSIM with JORDER while outperforming other methods, which is in consistent with the visual results. We can find the intermediate result from SDNet_3 also has decent results. Table 3 indicates that SDNet has higher SSIM value but fewer parameters as the block increases when compared with ResGuideNet. It means our SDNet is a more efficient structure for deraining task. The visual comparisons are shown in Fig.6, benefited from multi-scale features extraction and hierarchical fusion brought by G-ASP, our model has a satisfactory reconstruction with less blurring and abundant details.

3.3. Evaluation on real-world dataset
The ultimate goal of image processing algorithms is to serve the applications of practical scenes. Therefore, the performance in real-world dataset is a significant evaluation indicator for deraining algorithms. In this section, SDNet is trained on synthetic dataset and we also implement other methods according to their optimal setting. The visual results are shown in Fig.7. Since no ground truth exist, we only show their reconstruction results, we can find that SDNet still works well on real-world images, and the results produced by SDNet has joyful performance on multiple kinds of rain streaks with less blurring and more details.

Table 1. Investigation of G-ASP. Averaged PSNR/SSIM on synthesized images. Red indicates the best and blue indicates the second-best performance.

| Number of blocks | SDNet_O | SDNet_O+G-ASP,w/o HIC | SDNet_O+G-ASP |
|------------------|---------|------------------------|----------------|
| 1                | 21.78/0.752 | 22.24/0.745 | 23.10/0.774 |
| 2                | 23.22/0.789 | 23.98/0.815 | 25.09/0.815 |
| 3                | 24.00/0.803 | 24.90/0.824 | 25.62/0.831 |
| 4                | 24.31/0.815 | 25.63/0.831 | 25.86/0.844 |
| 5                | 24.48/0.819 | 25.89/0.838 | 26.13/0.850 |

Table 2. The average SSIM at different depths with fixed number of blocks in Rain100H and Rain100L.

| Number of blocks | 1     | 2     | 3     | 4     | 5     |
|------------------|-------|-------|-------|-------|-------|
| Rain100H         | 0.774 | 0.8146| 0.831 | 0.844 | 0.851 |
| Rain100L         | 0.945 | 0.959 | 0.964 | 0.968 | 0.971 |

Table 3. Averaged PSNR/SSIM on synthesized images with their parameters number.

| Item                | Rain100H | Rain100L | Rain12 | Parameters |
|---------------------|----------|----------|--------|------------|
| Rainy images        | 13.56/0.379 | 26.90/0.838 | 30.14/0.855 | -          |
| GMM [15]            | 15.05/0.425 | 28.66/0.865 | 32.02/0.910 | -          |
| DDN [16]            | 21.92/0.764 | 32.16/0.936 | 31.78/0.900 | 57,369     |
| JORDER [13]         | 26.54/0.835 | 36.63/0.974 | 33.92/0.953 | 369,792    |
| DID-MDN [17]        | 17.39/0.612 | 25.70/0.858 | 29.43/0.904 | 135,800    |
| ResGuideNet [18]    | 25.25/0.841 | 33.16/0.963 | 29.45/0.938 | 37,065     |
| SDNet_3             | 25.86/0.844 | 34.02/0.969 | 29.10/0.939 | 14,769     |
| SDNet_5             | 26.13/0.850 | 34.18/0.971 | 29.26/0.942 | 27,879     |
4. Conclusion
In this paper, we propose a light-weight and separable network for single image deraining called SDNet which is a cascaded structure. In particular, we propose a network separation strategy (NSS) to achieve proper computational resource allocation at test phase to meet real-world computer vision system under diverse rainy conditions. Meanwhile, a novel module called group-atrous spatial pyramid (G-ASP) is constructed for multi-scale feature extraction, G-ASP reuses parameters to reduce the model complexity significantly while maintaining a good deraining performance. Extensive experiments on synthetic datasets and real-world dataset show that our proposed SDNet obtains promising reconstruction results with no more than 28K parameters.

Acknowledgments
This work was supported in part by the National Natural Science Foundation of China under Grants 61571382, 81671766, 61571005, 81671674, 61671309 and U1605252, in part by the Fundamental Research Funds for the Central Universities under Grants 20720160075 and 20720180059, in part by the CCF-Tencent open fund, and the Natural Science Foundation of Fujian Province of China (No.2017J01126).

References
[1] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” in Advances in neural information processing systems, 2015, pp. 91–99.
[2] Li-Wei Kang, Chia-Wen Lin, and Yu-Hsiang Fu, “Automatic single-image-based rain streaks removal via image decomposition,” IEEE transactions on image processing, vol. 21, no. 4,
Yu Luo, Yong Xu, and Hui Ji, “Removing rain from a single image via discriminative sparse coding,” in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3397–3405.

Chao Dong, Chen Change Loy, Kaiming He, and Xiaou Tang, “Image super-resolution using deep convolutional networks,” IEEE transactions on pattern analysis and machine intelligence, vol. 38, no. 2, pp. 295–307, 2015.

Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, “Deeply-recursive convolutional network for image super-resolution,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1637–1645.

Xueyang Fu, Jiabin Huang, Xinghao Ding, Yinghao Liao, and John Paisley, “Clearing the skies: A deep network architecture for single-image rain removal,” IEEE Transactions on Image Processing, vol. 26, no. 6, pp. 2944–2956, 2017.

Xueyang Fu, Borong Liang, Yue Huang, Xinghao Ding, and John Paisley, “Lightweight pyramid networks for image deraining,” IEEE transactions on neural networks and learning systems, 2019.

He Zhang, Vishwanath Sindagi, and Vishal M Patel, “Image de-raining using a conditional generative adversarial network,” IEEE Transactions on Circuits and Systems for Video Technology, 2019.

He Zhang and Vishal M Patel, “Convolutional sparse and low-rank coding-based rain streak removal,” in 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2017, pp. 1259–1267.

Shang-Hua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip Torr, “Res2net: A new multi-scale backbone architecture,” arXiv preprint arXiv:1904.01169, 2019.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al., “Image quality assessment: from error visibility to structural similarity,” IEEE transactions on image processing, vol. 13, no. 4, pp. 600–612, 2004.

Justin Johnson, Alexandre Alahi, and Li Fei-Fei, “Perceptual losses for real-time style transfer and superresolution,” in European conference on computer vision. Springer, 2016, pp. 694–711.

Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan, “Deep joint rain detection and removal from a single image,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1357–1366.

Wang Yuanji, Li Jianhua, Lu Yi, Fu Yao, and Jiang Qinzhong, “Image quality evaluation based on image weighted separating block peak signal to noise ratio,” in International Conference on Neural Networks and Signal Processing, 2003. Proceedings of the 2003. IEEE, 2003, vol. 2, pp. 994–997.

Yu Li, Robby T Tan, Xiaojie Guo, Jiangbo Lu, and Michael S Brown, “Rain streak removal using layer priors,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2736–2744.

Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley, “Removing rain from single images via a deep detail network,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3855–3863.

He Zhang and Vishal M Patel, “Density-aware single image de-raining using a multi-stream dense network,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 695–704.

Zhiwen Fan, Huafeng Wu, Xueyang Fu, Yue Hunag, and Xinghao Ding, “Residual-guide feature fusion network for single image deraining,” in 2018 ACM Multimedia Conference on Multimedia Conference. ACM, 2018, pp. 1751–1759.