From Coarse to Fine: A Stage-Wise Deraining Net

CONG WANG1, MAN ZHANG1, ZHIXUN SU1,2, (Member, IEEE), GUANGLE YAO3, YAN WANG1, XIYAN SUN2, AND XIAONAN LUO4

1School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China
2Guangxi Experiment Center of Information Science, Guilin University of Electronic Technology, Guilin 541004, China
3School of Cyber Security, Chengdu University of Technology, Chengdu 610059, China
4School of Computer and Information Security, Guilin University of Electronic Technology, Guilin 541004, China

Corresponding authors: Zhixun Su (zxsu@dlut.edu.cn) and Xiyan Sun (sxy@guet.edu.cn)

The work of Z. Su was supported in part by the Natural Science Foundation of China under Grant 61572099, and in part by the Major National Science and Technology Project of China under Grant 2018ZX04011001-007. The work of X. Sun was supported in part by the Natural Science Foundation of China under Grant 61561016 and Grant 11603041, and in part by the Department of Science and Technology of Guangxi Zhuang Autonomous Region under Grant AC16380014, Grant AA17202048, and Grant AA17202033.

ABSTRACT In this paper, we propose a novel deep learning the based deraining method. The proposed method is motivated by the idea that an effective deraining algorithm should have the ability to remove various remaining rain streaks, which have been processed by the deraining method, in a repeated way. So, we design the deraining network in a coarse-to-fine manner that is multi-stage processing procedure and the parameters are shared in each stage. As the spatial contextual information is important for single image deraining, a densely connected dilation convolution block is proposed to deal with rain streaks with different sizes. Moreover, outer dense connections are used to guide the subsequent deraining procedures by fusing all the previous estimated rain-free images. The quantitative and qualitative experimental results demonstrate the superiority of the proposed method compared with recent state-of-the-art deraining methods on Rain100H, Rain1200, and Rain1400 datasets, while the number of parameters of our proposed method is greatly reduced due to the shared parameters strategy.

INDEX TERMS Deraining, deep learning, stage-wise, dense connections.

I. INTRODUCTION

Rain degrades visibility significantly and causes many computer vision systems to likely fail, so restoring rainy images is important for many computer vision applications. Deraining can be divided into two categories: video-based methods and single image-based methods. As video-based methods [1]–[4] can leverage temporal information by analyzing the difference between adjacent frames, they are easier than single image-based methods. In this paper, we explore the difficult one: single image deraining.

In the past decades, most of deraining methods are based on the physical model of rainy images:

\[ O = B + R, \]  

where \( O, B \) and \( R \) denote rainy images, background images (also called rain-free images) and rain streaks, respectively. Eq. 1 is a challenging ill-posed problem, which has a number of solutions of \( B, R \) for a given \( O \), theoretically. To solve this problem, some priors about rain streaks or clean images are proposed to constrain the solution space. Among them, image decomposition [5], sparse code [6], low-rank [7] and Gaussian mixture model [8] are widely employed. However, as these prior based methods usually assume that rain steaks should be sparse and have similar characters in falling directions and shapes, they only work in specific cases.

Recently, with convolutional neural networks (CNNs) having achieved great success in many computer vision tasks, e.g., object detection [9], object tracking [10], semantic segmentation [11], super resolution [12], [13], style transfer [14], [15], the CNNs have been developed to solve image deraining [16]–[21]. These deraining methods generally model the problem as a pixel-wise image regression process which directly learns to map an input rainy image to its clean one or a negative residual map in an end-to-end trainable CNNs. Although they have been demonstrated the effectiveness for deraining, some problems are still existing.

Firstly, most real-world restored images by deep learning based methods usually remain rain streaks, which makes these algorithms fail to estimate rain-free images.
Some methods improve the capability of rain streaks representation at the cost of increasing the number of network parameters, but we do not believe that this is a right direction for scientific development to get well-pleasing performance by blinding to add the model size. A light-weight model, which also can process various rain streaks, is needed. Secondly, spatial contextual information is important for single image deraining [22] that rain streaks with different sizes need different spatial contextual information to process. So a more adaptive deraining approach, which contains kinds of spatial contextual information, should be introduced.

To solve the above problems, we propose a coarse-to-fine deraining network that is multi-stage processing procedure. Our method is inspired by a natural idea that a good deraining method should have the ability to process various rain streaks in a repeated manner until the rain streaks are removed out cleanly. This also implies that estimated deraining image, in which there are remaining rain streaks, can be further processed by the deraining method. So our proposed deraining approach is in a coarse-to-fine manner that is multiple stages. What’s more, the model size keeps in a small weight due to that the parameters are shared in each stage. To deal with kinds of rain streaks, we propose a densely connected dilation convolution block. The dilation convolution can enlarge the receptive field, which can acquire more spatial contextual information, to handle with rain streaks with different sizes. In the block, a densely connected style is adopted, which can maintain important features from different levels [23]. Moreover, we find that the fusion among all the deraining results in previous each stage can promote the subsequent deraining performance. We think that the deraining results in previous stages can guide the subsequent deraining procedures. In the end, We find that there are existing some artifacts in the final estimated images, as shown in Fig. 1 (g). To solve this problem, we use a simper convolution network to refine the final results.

The contributions of this paper are summarized as follows:

- We propose a coarse-to-fine deraining method, which refines the deraining performance step by step with the parameters almost unchanged.
- Densely connected dilation convolution block is proposed, which can acquire more spatial contextual information, to handle with rain streaks with different sizes.
- We find that the fusion among different deraining results in previous each stage can promote the subsequent deraining performance by dense connections.
- Quantitative and qualitative experimental evaluations on both synthetic datasets and real-world datasets demonstrate that the proposed algorithm outperforms the state-of-the-art methods using the smallest model size.

II. RELATED WORK

Existing deraining methods can be divided into two categories: video-based methods and single image-based methods. As video-based methods [1]–[4] can leverage temporal information by analyzing the difference between adjacent frames, they are easier than single image-based methods. In this paper, we explore a more difficult task: single image deraining.

Single image based methods can also be split into two categories: prior based and deep-learning based methods. Next, we review the two categories briefly.

A. PRIOR BASED METHODS

Before deep-learning based methods, many priors about rainy images were proposed. Kang et al. [5] separated the rain streaks from high-frequency layer by sparse coding from HOG features. Luo et al. [6] proposed a discriminative sparse coding framework, which was based on image patches and separated rain streaks from background images. Chen and Hsu [7] proposed a generalized low-rank model, where the rain streaks layer was assumed to be low-rank. Kim et al. [24] used the non-local mean filter to detect rain streaks and remove them. Li et al. [8] decomposed a rainy image into background and rain streaks layer using Gaussian mixture models.

B. DEEP-LEARNING BASED METHODS

In recent years, deep-learning methods have achieved great success in many low-level vision tasks, e.g., dehazing [25]–[27], super-resolution [12], [13], denoising [28], style transfer [14], [15] and also deraining. Fu et al. [16], [17] firstly proposed deep-learning methods for single image deraining, where they firstly decomposed rainy images into low- and high-frequency parts then mapped high-frequency parts to rain streaks and lastly obtained clean images via Eq. 1. Yang et al. [18] designed a recurrent contextual network to jointly detect and remove rain streaks. Zhang et al. [29] proposed a conditional generative adversarial network for single image deraining and utilized perceptual loss to refine the last results. Li et al. [20] came

III. PROPOSED METHOD

In this section, we introduce the proposed method in details, including the overall network framework, densely connected dilation convolution block, refinement stage and loss function.

A. OVERALL NETWORK FRAMEWORK

The overall network framework shown in Fig. 2 is in a coarse-to-fine manner that is a multi-stage deraining procedure. Each stage contains three parts, including feature space, densely connected dilation convolution block and rain space. The feature space is to convert image space to feature space and the rain space is to map rain feature space to rain space. After the multi-stage process, the refinement stage is utilized to refine the multi-stage results. The refinement stage will be introduced in Sec. III-C. The densely connected dilation convolution block, called DCDCB, is to process rain feature space, which will be introduced in Sec. III-B. Moreover, the outer dense connections are used, which have the guiding function for the subsequent deraining procedures by fusing previous deraining results.

We describe the overall network mathematically as follows:

\[
X_t = f_{st}(\text{Conv}_{1 \times 1}(\text{Cat}[O_{t-1}, O_{t-2}, \cdots, O_0])),
\]

\[
Y_t = \text{DCDCB}(X_t),
\]

\[
R_t = r_{st}(Y_t),
\]

where \( t = 1, 2, \cdots, T \). \( T \) denotes the number of stages and \( O_0 \) denotes the original input rainy image. \( \text{Cat} \) and \( \text{Conv}_{1 \times 1} \) denote the concatenation at the dimension of channel and \( 1 \times 1 \) convolution operation, respectively. \( f_{st} \) and \( r_{st} \) denote the conversion operator from image space to feature space and from rain feature space to rain space at the t-th stage, respectively. \( \text{DCDCB}_t \) is our proposed densely connected dilation convolution block, which is to process rain streaks with different sizes. Please note that the parameters in \( f_{st} \), \( \text{DCDCB}_t \) and \( r_{st} \) are shared in each stage. Here, we can obtain the estimated rain-free image via Eq. 1: \( O_t = O_0 - R_t \), where we regard the estimated rain-free images as a fusion part of input rainy image in the next stage.

B. DENSELY CONNECTED DILATION CONVOLUTION BLOCK

As spatial contextual information is important for single image deraining [22] that rain streaks with different sizes need different spatial contextual information to process. To deal with the problem, we propose a densely connected dilation convolution block, called DCDCB, which consists of several convolution layers with exponential dilation factors.

The used dilation convolution can enlarge the receptive field, which can acquire more spatial contextual information and keep the number of parameters unchanged. Moreover, inner dense connections are utilized that can maintain and fuse important features from different levels [23].

This block can be expressed as:

\[
z_l = \text{Conv}_1(\text{Conv}_{1 \times 1}(\text{Cat}[z_{l-1}, z_{l-2}, \cdots, z_0])),
\]

\[
z_{\text{fusion}} = \text{Conv}_1(\text{Cat}[z_l, z_{l-1}, \cdots, z_0]),
\]

where \( l = 1, 2, \cdots, L \). \( L \) denotes the number of dilation convolution layers and \( \text{Conv}_1 \) denotes \( 3 \times 3 \) convolution with dilation factor \( 2^{l-1} \). \( z_{\text{fusion}} \) is the final output of the block. Here, we set \( L = 4 \), because we find that \( L = 4 \) is enough to obtain the satisfactory results.

C. REFINEMENT STAGE

We find that there are existing some artifacts in the final estimated images, see Fig. 1 and Fig. 9. To solve this problem, we use a simper convolution network to refine the final results. It is defined as:

\[
\hat{B} = \text{Conv}_{1 \times 1}(\sigma(\text{Conv}_{3 \times 3}(\sigma(\text{Conv}_{3 \times 3}(\hat{B})))),
\]

where \( \hat{B} \) and \( \hat{B} \) denote the estimated rain-free images by the multi-stage and the refinement results, respectively. Here we select LeakyReLU with \( \alpha = 0.2 \) as \( \sigma \).

D. LOSS FUNCTION

We use \( \text{MSE} \) as error measure for rain streaks in each deraining stage:

\[
\mathcal{L}_{\text{rain}} = \sum_{t=1}^{T} \| \hat{R}_t - R_t \|_2^2,
\]

where \( \hat{R}_t \) and \( R_t \) denote the estimated rain streaks and the ground-truth of rain streaks at the t-th stage, respectively.

For final refinement stage, we use the \( L_1 \) and \( \text{VGG} \) function as error measure:

\[
\mathcal{L}_{\text{refine}} = \| \hat{B} - B \|_1 + \| \text{VGG}(\hat{B}) - \text{VGG}(B) \|_1,
\]

where \( \hat{B} \) and \( B \) denote the estimated rain-free image and the ground-truth of rain-free images, respectively.

The final loss function is defined as:

\[
\mathcal{L} = \mathcal{L}_{\text{rain}} + \mathcal{L}_{\text{refine}}.
\]

IV. EXPERIMENTAL RESULTS

In this section, we demonstrate the effectiveness of the proposed method by conducting a mass of experiments on three synthetic datasets and a real-world dataset. All the results are compared with six state-of-the-art methods: DSC [6] (ICCV15), LP [8] (CVPR16), DDN [17] (CVPR17), JORDER [18] (CVPR17), RESCAN [20] (ECCV18), DID [21] (CVPR18). Moreover, more analysis about the proposed method will be discussed to further verify the superior of our designed network.
C. Wang et al.: From Coarse to Fine: Stage-Wise Deraining Net

FIGURE 2. The proposed network framework, which is multi-stage and the parameters are shared in each stage. For each stage, it contains three parts: Feature Space, Densely Connected Dilation Convolution Block (DCDCB) and Rain Space. In the end, we refine the multi-stage result by a refinement procedure.

TABLE 1. Quantitative experiments evaluated on three synthetic datasets. The best and the second best results are boldfaced and underlined, respectively. Number in parentheses indicates the parameter reduction compared our method.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rain100H</td>
<td>15.66</td>
<td>0.42</td>
<td>14.26</td>
<td>0.54</td>
<td>22.26</td>
<td>0.69</td>
<td>23.45</td>
<td>0.74</td>
<td>25.92</td>
<td>0.84</td>
<td>26.12</td>
<td>0.83</td>
<td>26.24</td>
<td>0.85</td>
</tr>
<tr>
<td>Rain1200</td>
<td>21.44</td>
<td>0.79</td>
<td>22.46</td>
<td>0.80</td>
<td>30.95</td>
<td>0.86</td>
<td>29.75</td>
<td>0.87</td>
<td>32.35</td>
<td>0.89</td>
<td>29.65</td>
<td>0.90</td>
<td>31.55</td>
<td>0.90</td>
</tr>
<tr>
<td>Rain1400</td>
<td>22.03</td>
<td>0.80</td>
<td>26.53</td>
<td>0.83</td>
<td>29.99</td>
<td>0.89</td>
<td>28.90</td>
<td>0.90</td>
<td>29.84</td>
<td>0.90</td>
<td>31.18</td>
<td>0.91</td>
<td>30.18</td>
<td>0.91</td>
</tr>
<tr>
<td>Parameters</td>
<td>-</td>
<td>-</td>
<td>58.175 (.70%)</td>
<td>369.792 (.95%)</td>
<td>54.735 (.68%)</td>
<td>372.839 (.95%)</td>
<td>17.305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A. IMPLEMENT DETAILS
We select $T = 12$ and the reason will be illustrated in Sec. IV-F. The number of channels is set to 16. We use the LeakyReLU with $\alpha = 0.2$ as the non-linear activation function. We randomly crop image patches with size of $100 \times 100$ pixels from the training image datasets as the inputs and set the mini-batch size to be 10 to train the network. The ADAM [30] optimizer is used. The learning rate is initialized to be 0.001, and it will be divided by 10 at 240K and 320K iterations. We train the network using 400K iterations on a PC with an NVIDIA GTX 1080Ti GPU.

B. EVALUATION CRITERIA
We use the peak signal to noise ratio (PSNR) [31] and structure similarity index (SSIM) [32] to evaluate the quality of the restored images on synthetic datasets. As there are no ground truth images for real-world images, we only show visual comparisons on the real-world datasets.

C. RESULTS ON SYNTHETIC DATASETS
We conduct deraining experiments on three widely used synthetic datasets: Rain100H [18], Rain1200 [21] and Rain1400 [17]. These three datasets include various rain streaks with different sizes, shapes and directions. Rain100H has 1800 images for training and 200 images for testing. Rain1200 has 12000 images for training and 1200 images for testing. Rain1400 has 13600 images for training and 1400 images for testing. It is ensured that all the testing datasets have different background images with training datasets. We select Rain100H as our analysis dataset.

FIGURE 3. An example in synthetic datasets compared with prior based deraining methods. The proposed method generates a much clear image.
FIGURE 4. Several examples in synthetic datasets compared with deep-learning based deraining methods. It is obvious that our results shown in (f) are the best than others.

FIGURE 5. One example in synthetic datasets compared with deep-learning based de-raining methods under the condition of the almost equivalent number of parameters.

on Rain100H. For the Rain1200 and Rain1400 datasets, our algorithm gains the highest value of SSIMs. And the PSNRs are also comparable with the other state-of-the-art methods even surpass most of them. The most notably key point is that our method only has 17,305 parameters, which are a large margin reduction compared with other deep-learning based deraining algorithms. Moreover, we give some visual examples as comparison. Fig. 3 shows the results compared prior based methods, including DSC [6] and LP [8]. It is obvious that our method shown in Fig. 3 (d) obtains the best result and the others gain the unacceptable results. We also compare the visual examples with deep-learning based methods: JORDER [18], DDN [17], RESCAN [20] and DID [21]. From Fig. 4, we can see that our method restores cleaner and clearer results and always gain the highest values of PSNRs and SSIMs, while the others’ results either have more artifacts or remain some rain streaks. Specifically, the JORDER [18] always obtain darker results.

To further verify the effectiveness of our proposed method, we add the number of parameters by enlarging the number of the channels and the dilation convolution layers in DCDCB to meet the almost equal number of parameters.
D. RESULTS ON REAL-WORLD DATASETS
We also provide some real-world challenging examples to further demonstrate the robustness of our method. Fig. 6 illustrates the results compared with prior based methods, including, DSC [6] and LP [8]. It is obvious that the prior based methods generate unacceptable results. This also demonstrates that the prior based methods do not work in some conditions. Our method shown in Fig. 6 (d) obtains the best performance and almost removes all the rain streaks. We also illustrate some examples compared with deep-learning based methods in Fig. 7. For the first example, our method almost removes all the rain streaks, while the others remain some rain streaks. For the other examples, our method produces

with RESCAN [20] and DDN [17]. The results are shown in Tab. 2 and we can see that we greatly improved the deraining performance in term of the PSNRs and SSIMs compared with RESCAN [20] and DDN [17]. This illustrates that our method is indeed effective and the deraining results will be better with the model size increasing. Further, we also provide a visual example in Fig. 5. It can be observed that our result is far better than other results. The others either have remaining rain streaks or some artifacts.

TABLE 2. The results under almost equal number of parameters.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>25.92</td>
<td>27.55</td>
<td></td>
</tr>
<tr>
<td>SSIM</td>
<td>0.84</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>Parameters</td>
<td>54,735</td>
<td>54,705</td>
<td></td>
</tr>
</tbody>
</table>
clearer results than others. Especially, in the third example, our algorithm removes all rain streaks and the others are worse. This also demonstrates that our light-weight model is also effective and surpasses other state-of-the-art methods.

### E. ABLATION STUDY

It is meaningful to discuss the effectiveness of each our proposed component, including dilation convolution that can obtain more spatial contextual information to deal with rain streaks with different sizes, inner dense connections that can maintain and fuse important features from different levels, outer dense connections that can guide subsequent deraining procedures by fusing all the previous estimated rain-free images and the refinement stage that can further refine the estimated deraining results.

- **No-dilation**: our final network without dilation convolution.
- **No-inner-dense**: the DCDCB without dense connections.
- **No-outer-dense**: our proposed network without outer dense connections, i.e., the input of each stage is the deraining result of previous stage.
- **No-all-dense**: our proposed network without inner dense connections and outer dense connections.
- **No-refinement**: our proposed network without refinement stage.
- **Ours**: our proposed method, which is multi-stage with refinement.

The results are shown in Tab. 3 and we can observe that every component is useful for the final results. Specially, the dense connections, including inner connections and outer connections, play very important roles, which greatly improve the final results in term of PSNRs and SSIMs. And the refinement stage improves the PSNR and SSIM by 0.40 dB and 2%, respectively, which also illustrates the effect of our added refinement stage.

### F. ANALYSIS ON THE NUMBER OF THE STAGE

A natural question is why we select $T = 12$ as our default setting of the stages. For this purpose, we show the results along with the number of the stages increasing in Fig. 10. We can see that the results are the best when $T = 12$ and the performance get worse along with $T$ increasing. So we set $T = 12$ as our network setting.

### G. ANALYSIS ON THE COARSE-TO-FINE MANNER

Due to our method is in a coarse-to-fine manner that is based on multi-stage, it is necessary to explore the effectiveness of this way. For this purpose, we carry out the experiments of single stage, multi-stage without refinement and multi-stage with refinement, respectively. The results are illustrated in Tab. 4. We can see that the result of the single stage is unsatisfactory, which obtains far worse performance than other state-of-the-art methods by combining with Tab. 1. Furthermore, our multi-stage manner is effective that PSNR and SSIM are increased by 4.84 dB and 22%, respectively, while the number of parameters is only increased by 6.8%. So the multi-stage manner is worthy of promotion. what's more, our refinement also plays a very import role to the results that the PSNR and SSIM are increased by 0.4 dB and 2%, respectively.
We also provide a visual example to compare the coarse-to-fine manner, shown in Fig. 9. It is observed that our result, i.e., refinement, obtains the best performance, while the others remain some rain streaks or artifacts.

V. CONCLUSION
In this paper, we propose a de-raining method in a coarse-to-fine manner, which is performed via multi-stage processing procedures. The proposed method is motivated by the idea that a good de-raining method, which should have the ability to remove all various rain streaks, can process rain streaks repeatedly until the rain streaks are removed out cleanly. So, a multi-stage deraining method is presented by us and the effectiveness has been demonstrated in this paper. Quantitative and qualitative experimental results demonstrate the superiority of the proposed method compared with several state-of-the-art deraining methods on Rain100H, Rain1200 and Rain1400 datasets, while the number of the parameters of our proposed method is greatly reduced.

ACKNOWLEDGMENT
(Cong Wang and Man Zhang contributed equally to this work.)

REFERENCES

CONG WANG received the B.S. degree from Inner Mongolia University, China, in 2017. He is currently pursuing the master’s degree in computational mathematics with the School of Mathematical Sciences, Dalian University of Technology, China, under the supervision of Prof. Z. Su. His current research interests include deep learning, computer vision, and image processing.

MAN ZHANG received the B.S. degree from the Dalian University of Technology, in 2018. She is currently pursuing the master’s degree in computational mathematics with the School of Mathematical Sciences, Dalian University of Technology, China, under the supervision of Prof. Z. Su. Her current research interests include deep learning, computer vision, and image processing.
ZHIXUN SU received the Ph.D. degree in computational mathematics from the Dalian University of Technology, China, in 1993, where he is currently a Professor. He is also a Visiting Professor with the Guilin University of Electronic Technology. His current research interests include computer graphics, computer vision, computational geometry, and machine learning. He is currently a member of the Standing Committee of China Society of Computational Mathematics.

GUANGLE YAO received the Ph.D. degree from the University of Electronic Science and Technology of China. During his Ph.D. study, he was with the Institute of Optics and Electronics, Chinese Academy of Sciences. He is currently with the Chengdu University of Technology. His research interests include action recognition, object detection, visual tracking, and deep learning.

YAN WANG is currently pursuing the Ph.D. degree in computational mathematics with the Dalian University of Technology. His research interests include computer vision and machine learning.

XIYAN SUN received the Ph.D. degree from the Guilin University of Electronic Technology, Guilin, China. She was with the National and Local Joint Engineering Research Center of Satellite Navigation and Location Service, Guilin University of Electronic Technology, where she is currently a Professor. She is also an Academic Leader of information and communication, enjoys the special government allowances of the State Council of the People’s Republic of China, and a Guangxi Specially Invite Professor. Her research interests include satellite navigation, high-precision positioning, real-time kinematic positioning, and navigation applications.

XIAONAN LUO was the Director of the National Engineering Research Center of Digital Life, Sun Yat-sen University, Guangzhou, China. He is currently a Professor with the School of Computer and Information Security, Guilin University of Electronic Technology, Guilin, China. He received the National Science Fund for Distinguished Young Scholars granted by the National Natural Science Foundation of China. His current research interests include computer graphics, machine learning, and pattern recognition.

* * *