Two-Stage Fusion Model for Heavy Rain Removal on Single Image

Xiaomiao Pan¹, Yueting Yang², Chuansheng Yang¹³, Chao Wang¹³* and Anhui Tan¹³

¹School of Information Engineering, Zhejiang Ocean University, Zhoushan, Zhejiang, 316022, China
²School of Mathematics and Statistics, Beihua University, Jilin, Jilin, 132013, P.R. China
³Key Laboratory of Oceanographic Big Data Mining and Application of Zhejiang Province, Zhoushan, Zhejiang, 316022, China

*Corresponding author’s e-mail: wangchao@zjou.edu.cn

Abstract. Current single image derain methods cannot solve the heavy rain situation well. In this paper, based on the physical model of a rainy image, we build a two-stage network, TSF-Net, which combines model-driven and data-driven methods. The first stage gets the rain streaks, atmospheric light, and transmission map to obtain the coarse rain-free image by the physical model. The second stage is a fully convolutional neural network with the structure of U-Net. The proposed Multi-Scale Projection Fusion Block (MSPFB) module, which can perceive the spatial information across the scale of images, is employed to remove the residual rain and fuzzy parts in the first stage and obtain a refined clean image. Extensive experiments show that our TSF-Net achieves better accuracy and visual improvements against state-of-the-art methods.

1. Introduction

In the images and videos obtained under heavy rain, the important information we need has been covered or degraded to a certain extent, which makes downstream computer vision tasks, e.g., target recognition and tracking, difficult to deal with. Therefore, we should give priority to removing the heavy rain and restoring clean images. At present, though most single image derain methods have become increasingly mature, it is still difficult removing the heavy rain streaks.

Current methods of single image rain removal can be mainly divided into two categories, namely model-driven [1,2,3,4] and data-driven [5,6,7]. It is mostly model-driven in the early rain removal methods, people obtain the artificial priors needed for rain through experimental research and statistical analysis of a large amount of information related to the rain environment, such as decomposition [1], sparse code [2], low-rank representation [3], gaussian mixture model [4] and so on. In recent years, with the rise of deep convolutional neural networks in the image processing field, many data-driven rain removal methods have produced amazing results. In the paper of Li et al. [7], the following physical model formula of heavy rain is proposed:

\[ I = T \odot (J + S) + (1 - T) \odot A, \]  

where \( I, J, S, T, \) and \( A \) represent a heavy rain image, a rain-free image, rain streaks, a transmission map, and atmospheric light respectively. \( \odot \) indicates element multiplication. Based on the atmospheric
scattering model formula, this formula takes not only the rain streaks but also the influence of rain and fog on the image into account, which well meets our demand for heavy rain removal and inspires our work deeply.

To solve the above problems, we build a new CNN method with the main contributions as follows:

- We introduce an end-to-end network that simultaneously fuses physics-based and model-free information to effectively remove heavy rain from single images, called TSF-Net.
- We design an MSPFB module to improve the effect of network optimization.
- We compare the proposed model with state-of-the-art (SOTA) rain removal methods. Our method significantly outperforms the SOTA methods.

2. Proposed method

Here, we briefly introduce the proposed two-stage network. The first stage is designed based on the physical model [7]. In this stage, a heavy rain image is an input, and the high-pass image and low-pass image are obtained by using the guided filter. We put the high-pass image into the residual network for extracting rain streaks, the low-pass one into the residual network for extracting atmospheric light, and the image mixed with the low-pass and high-pass information into the U-Net for transmission map. We get the rough clean image according to the formula of rain-free image:

\[
J = \frac{I - T \cdot OA}{T} - \sum S_i.
\]

The second stage is the convolutional neural network of U-Net architecture. In this stage, we put the rough clean image obtained in the first stage, the input heavy rain image and the difference between them with concatenate operation into U-Net for preventing the loss of original background details and promoting the efficiency of network learning, and get our refined clean image.

Figure 1. The overall architecture of the proposed network. The image J is reconstructed according to Eq. (2), I is the input rain image and R is the input I minus J. RB is a residual block and FC is the full connection layer. © represents the concatenate operation.

2.1. Physics-based restoration

In this part, we explain the first stage of the network in detail. Rain-Subnet which is used for extraction of rain streaks, consists of shallow feature extraction block 1×1 Conv and twelve residual blocks. Atmo-Subnet is used for extraction of atmospheric light, and it consists of a shallow feature extraction block 1×1 Conv, five convolutional blocks with ReLU activation function, a full connection layer, and bilinear up-sampling. As for Trans-Subnet, which is the extraction of transmission maps, is composed of shallow feature extraction block 1×1 Conv and U-Net.

In the U-Net, we use four convolution layers with ReLU as the encoder part, and carry out feature extraction and sub-sampling at each layer to reshape the image size. The decoder also contains four layers and deconvolution layers with a kernel size of 2 are used for up-sampling the image. After concatenating operation, 3×3 Conv with ReLU and feature enhancement are carried out to double the
number of channels. According to the atmospheric scattering model, we use the results of the three subnetworks to get a rough clean image J.

2.2. Model-free refinement

The second stage network is a convolutional neural network of U-Net architecture named TSF-Net. We first perform three parallel convolutions of 3×3, 5×5, and 7×7 kernels to obtain information of different receptive fields. The large receptive field is beneficial to learn rain streaks of different directions and sizes. Concatenate operation is carried out on the multi-scale information, and a 1×1 Conv is used to obtain a three-channel feature map, which is put into U-Net.

Figure 2. The Multi-Scales Projection Fusion Block. 1, 1/2, and 1/4 represent the scale of the original input feature. Down is a max-pooling operation and up is a bilinear up-sample operation.

The encoders are divided into three layers, each layer consists of two MSPFB modules and a max-pooling. The purpose of MSPFB, whose structure is shown in figure 2, is to reduce the computation cost and information missing, and improve multi-scale representation fusion effect. Max-pooling is used to reduce the width and height of images, and then bilinear interpolation is used to upsample the images. We pass the result of projection fusion operation through a 1×1 Conv and two 3×3 Conv blocks to get the residual image. In the decoder part, the first layer conducts concatenate operation with the upsampled result features from the second layer of the encoder, and then uses 1×1 Conv to make the channel number consistent. The processing of the next two layers is similar, and the final clean image is reconstructed by a 1×1 Conv.

2.3. Loss function

In stage one, we use the ℓ1 loss for S, T, A images estimated from the three subnetworks and use color loss and perceptual loss for rough image J. The loss function of stage one is as follows:

$$\mathcal{L}_{one} = \mathcal{L}_1(S, S_{gt}) + \mathcal{L}_1(T, T_{gt}) + \mathcal{L}_1(A, A_{gt}) + \mathcal{L}_{per}(J, C_{gt}) + \mathcal{L}_{color}(J, C_{gt}),$$  \hspace{1cm} (3)

where $\mathcal{L}_1$, $\mathcal{L}_{color}$, $\mathcal{L}_{per}$ represent the loss function of ℓ1 loss, color loss, and perceptual loss respectively. Perceptual loss is based on VGG16 pretrained on the ImageNet dataset. And S, T, A, J respectively represent the rain streaks, atmospheric light, transmission maps, and rough clean images extracted from the network. $S_{gt}$, $A_{gt}$, $T_{gt}$, and $C_{gt}$ represent the corresponding ground truth in the dataset.

In stage two, we will use ℓ1 loss and perceptual loss for clean images reconstructed from U-Net. The loss function of stage two is as follows:

$$\mathcal{L}_{two} = \mathcal{L}_1(C, C_{gt}) + \mathcal{L}_{per}(C, C_{gt}).$$  \hspace{1cm} (4)
3. Experiments
We evaluate our approach on several datasets across image restoration tasks. We report the standard metrics in image restoration including PSNR and SSIM [9]. The datasets used for training are described next.

3.1. Datasets
We selected the Outdoor-Rain dataset [7] which is generated on a set of clean outdoor images and contains 9k training samples and 1.5k test samples. This dataset renders proper rain streaks and rain accumulation effects based on scene depths, which are estimated by a pre-trained monocular depth estimation [8].

3.2. Implementation details
Before the training, we perform data augmentation, and randomly crop the patch of $256 \times 256$ with random flipping of horizontal or vertical for each image. The network training is mainly divided into two stages. For the first stage, we only train the physical model, setting the network hyperparameter Batch Size = 1, the initial Learning Rate $= 10^{-4}$, and using the cosine annealing algorithm to iterate 60 epochs. For the second stage, we froze the weights of the physical model and iterate 100 epochs.

Table 1 shows quantitative results of our method and three state-of-the-art deraining methods on Outdoor-Rain [7]. And we also show several challenging synthetic examples for visual comparisons in figure 3. As we can see that our proposed method outperforms these state-of-the-art approaches on this dataset.

Table 1. The quantitative results of our proposed network and the state-of-the-art (SOTA) methods on the Outdoor-Rain dataset [7].

| Method          | PSNR  | SSIM  |
|-----------------|-------|-------|
| TSF-Net (ours)  | 29.64 | 0.904 |
| Heavy Rain [7]  | 20.68 | 0.783 |
| RESCAN [5]      | 21.77 | 0.709 |
| DID-MDN [6]     | 22.64 | 0.807 |

Figure 3. Visual and quantitative comparisons of three synthetic examples. Obviously, the proposed method performs better than the other three deep learning-based methods.
To evaluate the robustness of our method on other synthetic rain images Test1 [7], we also provide the visual results in figure 4. Our method generates the cleanest result, while the other methods remain some obvious rain streaks and fog.

Figure 4. A comparison of our algorithm performed on Test 1 dataset.

3.3. Ablation study

Here, we present ablation experiments to analyze the contribution of the projection block of our model. Evaluation is performed on the Outdoor-Rain dataset [7] with our deraining model trained on image patches of size $256 \times 256$ for $1.3 \times 10^6$ iterations, and the results are shown in table 2.

|          | PSNR  | SSIM  |
|----------|-------|-------|
| W Projection | 29.64 | 0.904 |
| W/O Projection | 27.49 | 0.894 |

4. Conclusion

In this paper, we propose a two-stage network that combines a physical model and a fully convolutional network. The network uses MSPFB to enhance multi-scale features and improve performance. The results of quantitative and qualitative experiments show that our network has excellent performance.

Acknowledgments

This work was supported by Science and Technology Development Program of Jilin Province (20190303132SF); The Key Project of Natural Science Foundation Joint Fund of Jilin Province (2020122367JC); National Natural Science Foundation of China (12001005).

References

[1] Wang, Y., Liu, S., Chen, C., Zeng, B. (2017) A hierarchical approach for rain or snow removing in a single color image. IEEE Transactions on Image Processing, 26(8): 3936-3950.
[2] Deng, L.J., Huang, T.Z., Zhao, X.L., Jiang, T.X. (2018) A directional global sparse model for single image rain removal. Applied Mathematical Modelling, 59: 662-679.
[3] Chen, Y.L., Hsu, C.T. (2013) A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In Proceedings of the IEEE International Conference on Computer Vision. pp. 1968–1975.
[4] Li, Y., Tan, R. T., Guo, X., Lu, J., Brown, M.S. (2016) Rain streak removal using layer priors. In Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2736-2744.
[5] Li, X., Wu, J.L., Lin, Z.C., Liu, H., Zha, H.B. (2018) Recurrent squeeze-and-excitation context aggregation net for single image deraining. In European Conference on Computer Vision (ECCV). pp. 262–277.
[6] Zhang, H., Patel, V.M. (2018) Density-aware single image de-raining using a multi-stream dense network. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 695–704.
[7] Li, R., Cheong, L.F., Tan, R.T. (2019) Heavy rain image restoration: Integrating physics model
and conditional adversarial learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1633-1642.

[8] Godard, C., Mac Aodha, O., Brostow, G.J. (2017) Unsupervised monocular depth estimation with left-right consistency. In Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 270-279.

[9] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P. (2004) Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4): 600-612.