Jointly Learning of Heavy Rain Removal and Super-Resolution in Single Images

Dac Tung Vu, 김문철
한국과학기술원 전기 및 전자 공학과
tungvu@kaist.ac.kr, mkim@ee.kaist.ac.kr

Jointly Learning of Heavy Rain Removal and Super-Resolution in Single Images

Dac Tung Vu, Munchurl Kim
Korea Advanced Institute of Science and Technology Dep. Of Electronic Engineering

요 약

Images were taken under various weather such as rain, haze, snow often show low visibility, which can dramatically decrease accuracy of some tasks in computer vision: object detection, segmentation. Besides, previous work to enhance image usually downsample the image to receive consistency features but have not yet good upsample algorithm to recover original size. So, in this research, we jointly implement removal streak in heavy rain image and super resolution using a deep network. We put forth a 2-stage network: a multi-model network followed by a refinement network. The first stage using rain formula in the single image and two operation layers (addition, multiplication) removes rain streak and noise to get clean image in low resolution. The second stage uses refinement network to recover damaged background information as well as upsample, and receive high resolution image. Our method improves visual quality image, gains accuracy in human action recognition task in datasets. Extensive experiments show that our network outperforms the state of the art (SoTA) methods.

1. Introduction

Nowadays, rain removal in the single image becomes an important task to preprocess images. To overcome this problem, many methods [1,3] have been proposed to recover rain-free image from rain image. However, these methods only achieve good results in light rain images, in the heavy rain most approaches fail to recover the background scene. Because when dense veiling effect appears, rain and fog can become entanglement with each other and difficult to separate using existing methods. This is leading to a combine deraining and dehazing method but cannot address the problem properly. Besides, the information in the background scene can be seriously broken and some of them in dense rain are not represented by the model. In other words, the model cannot describe all happen in the real scene cause failure to solve problems. Moreover, to enhance image quality most works usually downsample the image to receive consistency features but have not yet good upsample algorithm to recover original size. So, in this research, we jointly implement removal streak in heavy rain image and super resolution using deep network.

To resolve these existing problems in dense rain scene in the single image, we propose a network with the following contributions. Firstly, we introduce a network including two stages to simultaneously address heavy rain removal and super resolution. The first stage using operator architecture to reconstruct background scenes and remove noise components in low resolution. The second stage uses base

on channel attention network to upsample image and reconstruct visual damaged information. Secondly, we propose a new approach extract the MSER features to improve the learning ability. Finally, our experimental results outperform the SoTA methods qualitatively and quantitatively.

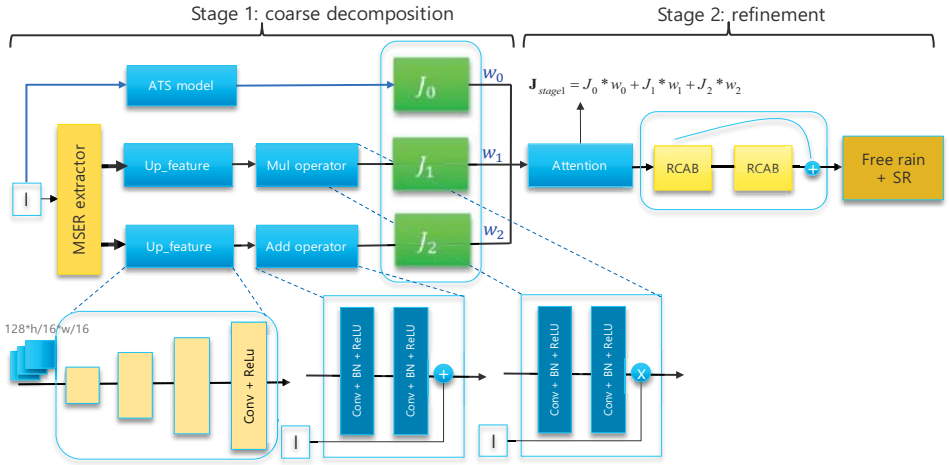

Figure 1. The architecture of the proposed network. The detail of ATS model is depicted in Fig. 2

2. Method

This section shows 2-stages network. Firstly, we present overall of architecture, as well as the input and output of each stage. The first stage is coarse decomposition, this stage decomposes rainy input image and result output is rain-free image at low resolution. Coarse decomposition stage applies MSER feature extractor, up feature module to combine features and then predicts output image from model and operation layers. However, as present in introduction part, some background information in result from first stage are dramatically damaged and cannot represented by model, so we then propose second stage, refinement stage. The result image from stage 1 firstly are fed into attention mechanism. After that we implement a RCAB architecture base on channel attention to refinement image, generate high frequency information.

2.1 Stage 1: Coarse decomposition

- MSER extractor

MSER[2] (maximum stable extremal regions) is a method for blob detection in images. The MSER algorithm extracts from an image a number of co-variant regions that are stable across a range of thresholds in algorithm. In this work, we extract patch with size H/16×W/16 (H, W: height, width of rain image) corresponding with coordinates of MSER features. So, these regions almost have the same characteristic, related to easy to remove rain. Besides, dividing image into small parts (where rain have same density and direction) and then combine them also increase capability residual in rain image. After that, MSER features is taken as input of two operation layers, see Fig.1 for more detail.

- Prediction from operation layer

Rain removal task in single image is pretty complicated and it is difficult to removal rain from single layer. In this regard, we combine deraining result from separate layers to improve performance because these layers can learn complementary information from rain image. Hence, we implement two specific layers (addition, multiplication) to decompose rain from input image, then use attention mechanism in stage 2 to combine them and choose better information.

Firstly, we consider to multiplication layer for rain decomposition model:

$$J_1(p) = I(p) * R_1(p) \quad (1)$$

where I is the rainy input; J_1 and R_1 denote the two layers, which are learned to decompose input I using the Eq. (1), p is the pixel location. In particular, we use two convolution layers with 3x3 kernel and then calculate the result J_1 followed Eq. (1) with input image I and layer R_1.

Secondly, addition mechanism is explored to separated layer:

$$J_2(p) = I(p) + R_2(p) \quad (2)$$

Similarly, J_2 and R_2 denote the two layer decompose input using Eq. (2), see figure 1 for more detail. We apply two 3x3 convolution and then addition R_2 and I in layer-by-layer manner.

- Why only two layers

The main purpose of our layer separation models is removal rain from the input rainy image to recover background scene. In other word, the model separate image combined by rain streaks, atmospheric light, transmission map and background scene. According to heavy rain model in [5], the heavy rain model consists of two mathematic

operations, addition and multiplication. Hence, we propose the two separator models (see Figure 1) containing two operations for two-layer combinations, and they are multiplication "∗"in J_1, addition "+" in J_2. Moreover, for better approximating in the presence of rain, we use the channel mechanism to create weighting maps for linearly combining all these two models in the final rain-free image, and then these weights are optimized when train in the loss function of network. Our superior performance shows that the effectiveness of our two-layer separator models for image deraining.

- Prediction from ATS model

To predict free-rain image from rain model [5], we together predict rain streaks, atmospheric light and transmission map. We estimated rain streak using two dense blocks and three 3x3 convolution layers. Transmission map is estimated by global average pooling, fully connected layer. And atmospheric light is predicted from likely U-net architecture. Then we combine three components using formula:

$$J(p) = \frac{I(p) - (1 - T(p)) \odot A(p)}{T(p)} - \sum_i^n S_i(p)$$

Where I is the observed rainy image. J is the background scene to be recover. p is pixel location. S_i is the rain layer, with n as the total number of rain streak layers. T is the transmission map, which represents the distance-dependent factor affecting the fraction of light that reaches the camera sensor. A is the global atmospheric light, indicating the ambient light intensity.

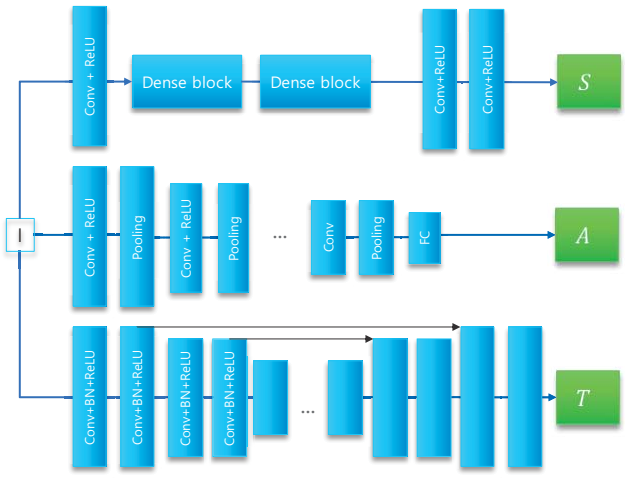

Figure 2. The ATS model

2.2 Stage 2: Refinement module

- Attention mechanism

After receive result from two operation layers, we concatenate 2 layers and result prediction from ATS rain model. We apply attention mechanism to three output of stage 1. Specifically, two 1x1 convolutions, two 3x3 convolution layers and softmax function is implemented. Obviously, the information in background scene are automatically highlighted, which are better recovered rain-free image. Furthermore, since there is complementary information in the results of the three models (two operator layers and prediction from ATS rain model), the attention mechanism can select the best one among all the three results. Hence, our method combines these three models by using these learned attention maps resulted in better performance in image deraining.

As mentioned in introduction part, the existing methods have not yet upsample algorithm to recover original size of original image, consequently the restoration image do not achieve good result. Hence, we apply RCAB[4] (Residual channel attention block) base on channel attention. These block allow low frequency information to be bypass through skip connection and focus on learning high frequency background, which information is damaged from stage 1 – coarse decomposition. Which architecture jointly refinement damaged information and upsample resolution of output result in stage 1. See figure 1 to illustrate design of stage 2 – refinement.

2.3 Loss function and training detail

- Loss functions

The total loss in this work consists of 2 losses is in stage 1 and in stage 2. The loss in stage 1 is MSE loss between ground truth image low resolution and output of ATS model and two operation layers as follow:

$$L_{stage1} = \lambda_0 L_{ATS} + \lambda_1 L_{Mul} + \lambda_2 L_{add}$$
$$L_{ATS} = L2(J_{GT-LR}, J_0)$$
$$L_{Mul} = L2(J_{GT-LR}, J_1)$$
$$L_{Add} = L2(J_{GT-LR}, J_2)$$

The loss in stage 2 is MSE loss, perceptual loss between output super resolution and ground truth high resolution.

$$L_{stage2} = L_{SR} + \lambda_3 L_{VGG}$$
$$L_{SR} = L2(J_{GT-HR}, J_{SR})$$
$$L_{VGG} = L2(VGG(J_{GT-HR}), VGG(J_{SR}))$$

The whole loss function can be expressed as:

$$L_{total} = L_{stage1} + L_{stage2}$$

Where J_{GT-LR}, J_{GT-HR} is ground truth image low resolution, high resolution respectively. J_0, J_1, J_2 is output low resolution result of ATS model, multiplication, addition layer in stage 1, J_{SR} is output super resolution result of proposed network. The perceptual loss base on VGG16 is pretrained on ImageNet dataset and $\lambda_{0,1,2,3}$ is hyper parameter.

- Training dataset

There are several large-scale synthetic datasets available for training deraining networks; however almost dataset do not contain rain accumulation effects. We choose Outdoor-Rain dataset [5] which is on a set of outdoor clean images. This dataset renders proper rain streaks and rain accumulation effects and estimate the depth of the scene using the method in single image depth estimation. This dataset contains 9000 training samples and 1,500 test samples.

- Training strategy

We initialize the parameters of the network by Xavier initial. We randomly cropped image with patch size 200x300 on all the training images and apply the Adam optimizer for training. The learning rate is adjusted by the poly policy with the initial learning rate of 5e-5 and divided by 2 after 100 epochs in total 500 epochs. We use a mini-batch size of 4 and train our network using two NVIDIA TITAN X GPU based on the PyTorch library.

Figure 3. The comparison results in Outdoor-Rain dataset.

3. Experiments

3.1 Ablation study

To study the role of the components in the coarse decomposition module, we compare four different modules: (a) there is no module prediction rain from model (only addition layer and multiplication layer), denoted as "Add+Mul". (b) there is no addition layer from two operation layers (only prediction from rain model and multiplication layer), denoted as "ATS+Mul". Similarly, (c) denoted "ATS+Add", there is only addition layer and prediction rain from model. Finally, (d) is fully coarse module proposed in this work.

Table 1. The PSNR and SSIM results for ablation study

Method	Add+Mul (a)	ATS+Mul (b)	ATS+add (c)	Proposed (d)
PSNR	22.37	22.25	22.24	22.67
SSIM	0.748	0.748	0.748	0.755

We training these four methods on testing dataset and evaluated quantitative results in table 1. The results of model (a-c) are approximate according to PSNR/SSIM values. It is shows that the role of layer prediction from model and two operation layer are similar. However, the proposed model (d) combined detailed information in other modules by attention mechanism, leading to significant improvement in the prediction.

3.2 Quantitative and qualitative results

Table 2. The comparison results with SoTAs

Method	PSNR	SSIM
Heavy_rain + RCAN	21.01	0.707
SPANet + RCAN	20.46	0.736
PreNet + RCAN	16.26	0.619
RESCAN + RCAN	21.72	0.674
DID-MDN + RCAN	19.18	0.646
Our result	22.67	0.755

The quantitative performance of our algorithm shown in table 2 outperforms the baseline methods in PSNR and SSIM metrics. Our network achieves 22.67 dB in PSNR and 0.755 in SSIM compare with 21.01dB and 0.707 in SoTA. Moreover, figure 3 demonstrates the qualitative results produced by our algorithm are comparable to other baseline methods. One can see, these methods still haze and some of them get black regions in results. Meanwhile, clean images in our result is more realism.

4. Conclusions

In this paper, we have proposed a 2-stage method that is able to jointly remove rain and super resolution by considering operation layer attention and refinement network. Firstly, in the coarse decomposition stage, we implement prediction from rain model and two operations (addition, multiplication) to remove heavy rain and noise to generate a clean image in low resolution. In the refinement stage, we proposed a network base on channel attention to upscale, reconstructed scenes from previous stage simultaneously and produce the final clean super resolution image. Comprehensive experiments demonstrate our method outperforms the state-of-the-art methods on datasets.

Acknowledgement

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No. 2017-0-00419, Intelligent High Realistic Visual Processing for Smart Broadcasting Media).

References

[1] X. Li, et al, “Recurrent Squeeze-and-Excitation Context Aggregation Net for Single Image Deraining,” Proceedings of the European Conference on Computer Vision (ECCV), 2018.

[2] P. Forssén, et al, “Maximally stable colour regions for recognition and matching.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007.

[3] T. Wang, et al, “Spatial Attentive Single-Image Deraining with a High Quality Real Rain Dataset,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[4] Yulun Zhang, et al, "Image Super-Resolution Using Very Deep Residual Channel Attention Networks", Proceedings of the European Conference on Computer Vision (ECCV) 2018.

[5] R. Li, et al, "Heavy Rain Image Restoration: Integrating Physics Model and Conditional Adversarial Learning," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.