Rethinking Image Deraining via Rain Streaks and Vapors

Yinglong Wang[1], Yibing Song[2⋆], Chao Ma[3], and Bing Zeng[1⋆]

[1] University of Electronic Science and Technology of China
[2] Tencent AI Lab
[3] MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University
ylwanguestc@gmail.com yibingsong.cv@gmail.com
chaoma@sjtu.edu.cn eezeng@uestc.edu.cn

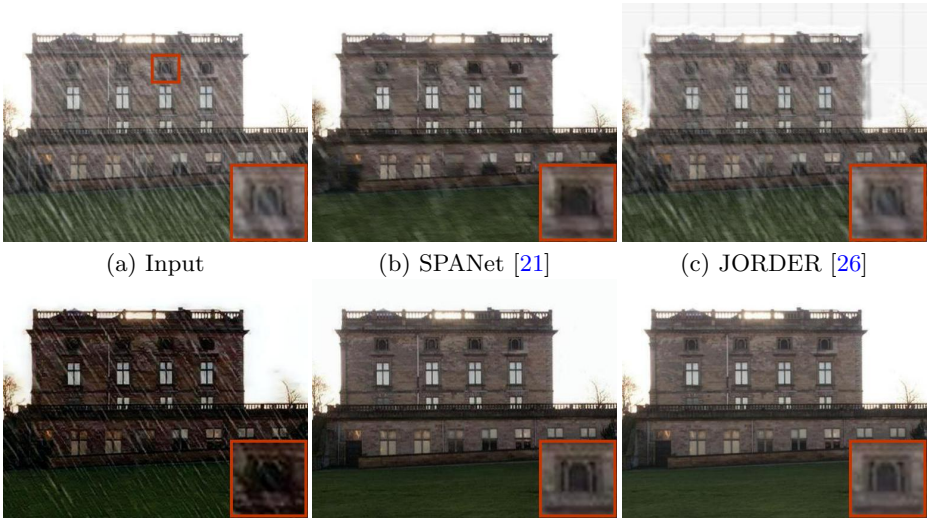
Abstract. Single image deraining regards an input image as a fusion of a background image, a transmission map, rain streaks, and atmosphere light. While advanced models are proposed for image restoration (i.e., background image generation), they regard rain streaks with the same properties as background rather than transmission medium. As vapors (i.e., rain streaks accumulation or fog-like rain) are conveyed in the transmission map to model the veiling effect, the fusion of rain streaks and vapors do not naturally reflect the rain image formation. In this work, we reformulate rain streaks as transmission medium together with vapors to model rain imaging. We propose an encoder-decoder CNN named as SNet to learn the transmission map of rain streaks. As rain streaks appear with various shapes and directions, we use ShuffleNet units within SNet to capture their anisotropic representations. As vapors are brought by rain streaks, we propose a VNet containing spatial pyramid pooling (SSP) to predict the transmission map of vapors in multi-scales based on that of rain streaks. Meanwhile, we use an encoder CNN named ANet to estimate atmosphere light. The SNet, VNet, and ANet are jointly trained to predict transmission maps and atmosphere light for rain image restoration. Extensive experiments on the benchmark datasets demonstrate the effectiveness of the proposed visual model to predict rain streaks and vapors. The proposed deraining method performs favorably against state-of-the-art deraining approaches.

Keywords: Deep Image Deraining

1 Introduction

Rain image restoration produces visually pleasing background (i.e.,scene content) and benefits recognition systems (e.g., autonomous driving). Attempts [10,4] of image deraining formulate rain image as the combination of rain streaks and background. These methods limit their restoration performance when the rain

⋆ Y. Song and B. Zeng are the corresponding authors. The results and code are available at https://github.com/yluestc/derain.



(a) Input (b) SPANet [21] (c) JORDER [26]

(d) PYM+GAN [13] (e) Ours (f) Ground Truth

Fig. 1. Rain image restoration results. Input image is shown in (a). Results of SPANet [21], JORDER [26], PYM+GAN [13] are shown in (b)-(d). Ground Truth is shown in (f). The proposed visual model is effective to formulate rain streaks and vapors, which brings high quality deraining result as shown in (e).

is heavy. The limitation occurs because heavy rain consisting of rain streaks and vapors causes severe visual degradation. When the rain streaks are clearly visible, a part of them accumulate to become vapors. The vapors produce the veiling effect which decreases image contrast and causes haze. Fig. 1 shows an example. Without considering vapors, existing deraining methods do not perform well to restore heavy rainy images as shown in Fig. 1(b) and Fig. 1(c).

Recent study [13] reformulates rain image generation via the following model:

$$\mathbf{I} = \mathbf{T} \odot (\mathbf{J} + \sum_{i=1}^{n} \mathbf{S}_i) + (\mathbb{1} - \mathbf{T}) \odot \mathbf{A} \tag{1}$$

where $\mathbf{I}$ is the rain image, $\mathbf{T}$ is the transmission map, $\mathbf{J}$ is the background to be recovered, $\mathbf{S}_i$ is the rain streak layer, and $\mathbf{A}$ is atmosphere light of the scene. Besides, $\mathbb{1}$ is a matrix whose pixel values are 1 and $\odot$ indicates element-wise multiplication. The transmission map $\mathbf{T}$ encodes influence from vapors to generate rain images. Based on this model, deraining methods propose various CNNs to predict $\mathbf{T}$, $\mathbf{A}$, and $\mathbf{S}$ to calculate background image $\mathbf{J}$.

Rain streaks and vapors are entangled with each other in practice. Removing them separately is not feasible [13]. Meanwhile, this entanglement makes Eq. (1) difficult to explicitly model both. The limitation raises that the transmission map and rain streaks are not estimated well. The incorrect estimation brings

unnatural illumination and color contrasts on the background. Although a generative adversarial network [13] is employed to refine background beyond model constraint, the illumination and color contrasts are not completely corrected as shown in Fig. 1 (d).

In this work, we rethink rain image formation by delving rainy model itself. We observe that in Eq. (1), both rain streaks and background are modeled to have the same properties. This is due to the meaning which two terms convey in Eq. (1). The first term $\mathbf{T} \odot (\mathbf{J} + \sum_{i=1}^{n} \mathbf{S}_i)$ indicates that both $\mathbf{S}_i$ and $\mathbf{J}$ are transmitted via $\mathbf{T}$. The rain streaks are regarded as part of the background to be transmitted. The second term $(\mathbb{1} - \mathbf{T}) \odot \mathbf{A}$ shows that rain streaks do not contribute to atmosphere light transmission because only vapors are considered in $\mathbf{T}$. As rain streaks and vapors are entangled with each other, the modeling of rain streaks as background is not accurate. Based on this observation, we propose a visual model which formulates rain streaks as transmission medium. The entanglement of rain streaks and vapors is modeled properly from the transmission medium perspective. We show the proposed model in the following:

$$\mathbf{I} = (\mathbf{T_s} + \mathbf{T_v}) \odot \mathbf{J} + [\mathbb{1} - (\mathbf{T_s} + \mathbf{T_v})] \odot \mathbf{A} \tag{2}$$

where $\mathbf{T_s}$ and $\mathbf{T_v}$ are the transmission map of rain streaks and vapors, respectively. In our model, all the variables are extended to the same size, so that we utilize element-wise multiplication to describe the relationship of variables.

Rain streaks appear in various shapes and directions. This phenomenon is more obvious in heavy rain. In order to effectively predict $\mathbf{T_s}$, we propose an encoder-decoder CNN with ShuffleNet units [28] named SNet. The group convolutions and channel shuffle improve network robustness upon diverse rain streaks. The learned multiple groups in ShuffleNet units are able to capture anisotropic appearances of rain streaks. Furthermore, we predict transmission map of vapors (i.e., $\mathbf{T_v}$) by using a VNet where there is a spatial pyramid pooling (SPP) structure. VNet takes the concatenation of $\mathbf{I}$ and $\mathbf{T_s}$ as input and use SPP to capture its global and local features in multi-scales for compact representation. On the other hand, we propose an encoder CNN named ANet to predict atmosphere light $\mathbf{A}$. ANet is pretrained by using training data in a simplified low transmission condition, under which we obtain estimated labels of $\mathbf{A}$ from rainy image $\mathbf{I}$. After pretraining ANet, we jointly train SNet, VNet and ANet by measuring the difference between the calculated background $\mathbf{J}$ and the ground truth background. The learned networks well predict $\mathbf{T_s}$, $\mathbf{T_v}$, and $\mathbf{A}$, which are further transformed to generate background images. We evaluate the proposed method on standard benchmark datasets. The proposed visual model is shown effective to model transmission maps of rain streaks and vapors, which are removed in the generated background images.

We summarize the contributions of this work as follows:

- We remodel the rain image formation by formulating rain streaks as transmission medium. The rain streaks and vapor contribute together to transmit both scene content and atmosphere light into input rain images.

- We propose SNet, VNet and ANet to learn rain streaks transmission map, vapor transmission map and atmosphere light. These three CNNs are jointly trained to facilitate the rain image restoration process.
- Experiments on the benchmark datasets show the proposed model is effective to predict rain streaks and vapors. The proposed deraining method performs favorably against state-of-the-art approaches.

2 Related Work

Single image deraining originate from dictionary learning [18] to solve the negative impact of various rain streaks on the background [23,22,11,1,29,9,8,2,17,16]. Recently, deep learning has obtained better deraining performances compared with the conventional methods. Prevalent deep learning based deraining methods can be categorized as direct mapping method, residual based method and scattering model based methods. Direct mapping based methods directly estimate rain-free background from the observed rainy images via novel CNN networks. It includes the work [21], in which a dataset is first built by incorporating temporal priors and human supervision. Then, a novel SPANet is proposed to solve the random distribution of rain streaks in a local-to-global manner.

A residual rain model is proposed in residual based methods to formulate a rainy image as a summation of the background layer and rain layers. It covers majority of existing deraining methods. For example, Fu et al. train their DerainNet in high-frequency domain instead of the image domain to extract image details to improve deraining visual quality [3]. In the meantime, inspired by the deep residual network (ResNet) [5], a deep detail network which is also trained in high-pass domain was proposed to reduce the mapping range from input to output, to make the learning process easier [4]. Yang et al. create a new model which introduces atmospheric light and transmission to model various rain streaks and veiling effect, but the rainy image is finally decomposed into a rain layer and background layer by their JORDER network. During the training, a binary map is learnt to locate rain streaks to guide the deraining network [25,26]. In [27], the density of rain streaks is classified into three classes and automatically estimated to guide the training of a multi-stream densely connected DID-MDN structure which can better characterize rain streaks with various shape and size. Li et al. regard the rain in rainy images as a summation of multiple rain streak layers, then use a recurrent neural network to remove rain streaks state-wisely [15]. Hu et al. study the visual effect of rain to scene depth, based on which fog is introduced to model the formation of rainy images and the depth feature is learned to guide their end-to-end network to obtain rain layer [7]. In [19], a better and simpler deraining baseline is proposed by considering the network structure, input and output of network, and the loss functions.

In the scattering model based methods, atmospheric light and transmission of vapor are rendered and learned to remove rain streaks as well as vapor effect, but rain streaks are treated the same as the background rather than the transmission medium [13]. Different from existing approaches, we reformulate rainy

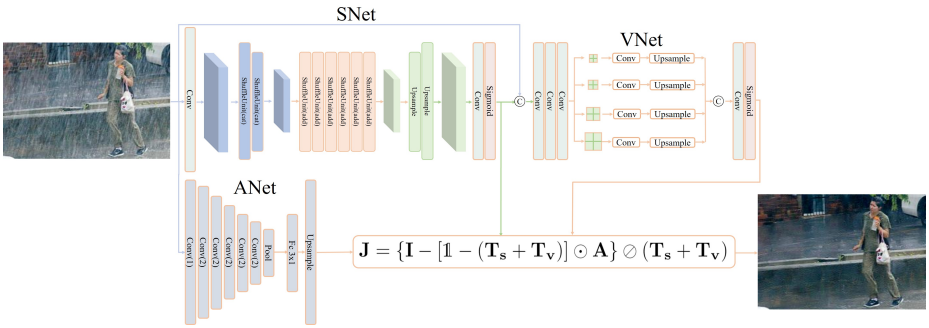


Fig. 2. This figure shows our network structure. The pool denotes adaptive average pooling operation. The Upsample operation after triangle-shaped network extends the atmospheric light $\mathbf{A}$ to the image size. The notations $\odot$ and $\oslash$ are the pixel-wise multiplication and division, respectively

image generation by modeling rain streaks as transmission medium instead of background content, and use two transmission maps to model the influence of rain streaks and vapor on the background. This formulation naturally models the entanglement of rain streaks and vapors and produce more robust results.

3 Proposed Algorithm

We show an overview of the pipeline in Fig. 2. It consists of SNet, VNet and ANet to estimate transmission maps and atmosphere light. The background image can then be computed as follows:

$$\mathbf{J} = \{\mathbf{I} - [\mathbb{1} - (\mathbf{T_s} + \mathbf{T_v})] \odot \mathbf{A}\} \oslash (\mathbf{T_s} + \mathbf{T_v}), \tag{3}$$

where $\oslash$ is the element-wise division operation. In the following, we first illustrate the network structure of SNet, VNet, and ANet, then we show how we train these three networks in practice and elucidate how these networks function in rain image restoration.

3.1 SNet

We propose a SNet that takes rain image as input and predicts rain streak transmission maps $\mathbf{T_s}$. SNet is an encoder-decoder CNN with ShuffleNet units that consist of group convolutions and shuffling operations. The input CNN features are partially captured by different groups and then shuffled to fuse together. We extend ShuffleNet unit to capture anisotropic representation of rain streaks as shown in Fig. 2. Our extension is shown in Fig. 3 where we increase the number of group convolutions and deep separable convolution. The features of different groups in single unit will be more discriminative by twice grouping

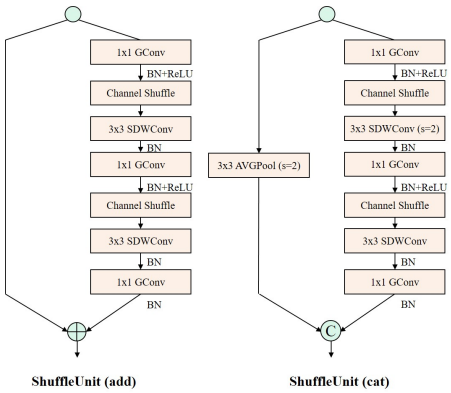


Fig. 3. This figure shows our revised ShuffleNet Units. ShuffleUnit(add) can keep image size unchanged, and ShuffleUnit(cat) downsamples image once. + and C mean addition and concatenation respectively.

to boost the global feature grouping. Moreover, the depthwise convolution is symmetrically padded (SDWConv) to decrease the influence of padded 0 on the image edges. Finally, we upsample the feature map to original size and convolution layers are followed to fuse multi-group features. The prediction of $\mathbf{T_s}$ on SNet can be written as:

$$\mathbf{T_s} = \mathcal{S}(\mathbf{I}) \tag{4}$$

where $\mathbf{I}$ is the rain image and $\mathcal{S}(\cdot)$ is the SNet inference.

3.2 VNet

We propose a VNet that captures multi-scale features to predict vapor transmission maps $\mathbf{T_v}$. VNet takes the concatenation of rain image $\mathbf{I}$ and $\mathbf{T_s}$ as input where $\mathbf{T_s}$ provides the global intensity information for $\mathbf{T_v}$ and $\mathbf{I}$ supplies the local background information, as different local areas have different vapor intensity. Compared with anisotropic rain streaks, vapor is locally homogeneous and the values of different areas have high correlation. VNet utilizes SPP structure to capture global and local features to provide compact feature representation for $\mathbf{T_v}$ as shown in Fig. 2. The prediction of $\mathbf{T_v}$ on VNet can be written as:

$$\mathbf{T_v} = \mathcal{V}(\text{cat}(\mathbf{I}, \mathbf{T_s})). \tag{5}$$

3.3 ANet

We propose an encoder network ANet to predict atmosphere light. Its structure is shown in Fig. 2. The network inference can be written as:

$$\mathbf{A} = \mathcal{A}(\mathbf{I}), \tag{6}$$

Algorithm 1 Pretraining ANet

Input: Rainy images $\{\mathbf{I}^{\{t\}}\}$.

for $i = 1$ to epoch **do**
  for $j = 1$ to batchnum **do**
    Locate rain pixels for $\mathbf{I}^{\{t\}}$ via [25].
    Based on Eq. (8), find highest rainy pixel as the ground truth atmospheric light $\mathbf{A}^{\{t\}}$.
    Calculate $\mathcal{A}(\mathbf{I}^{\{t\}})$ via Eq. (6).
    Updating $\mathcal{A}(\cdot)$ via loss (9).
  end for
end for

Output: Learned atmospheric light $\mathbf{A}$.

where $\mathcal{A}(\cdot)$ is the ANet inference. As atmosphere light is usually considered constant in the rain image, the output of the encoder is a 3×1 vector. We use an ideal form of our rain model Eq. (2) to create labels of atmospheric light from rain images to pretrain ANet. Then, we integrate ANet into the whole pipeline for joint training. The details are presented in the following.

3.4 Network Training

The pipeline of the whole network consists of a SNet, a VNet, and an ANet to predict $\mathbf{T_s}$, $\mathbf{T_v}$, and $\mathbf{A}$, respectively. Then, we will generate $\mathbf{J}$ according to Eq. (3). We first pretrain ANet using labels from a simplified condition and perform joint training of these three networks.

Pretraining ANet. Sample collection is crucial for pretraining ANet as the ground truth value of atmosphere light is difficult to obtain. Instead of empirically modeling $\mathbf{A}$ as a uniform distribution [13], we generate labels under a simplified condition based on our rain model Eq. (2), where the transmission maps of both rain streaks and vapors are 0. For one pixel x in rain image $\mathbf{I}$, $\mathbf{T_s}(x) + \mathbf{T_v}(\mathbf{x}) = 0$, our visual model of rain image formation can be written as:

$$\mathbf{I}(x) = \mathbf{A}(x) \tag{7}$$

where the pixel value of atmosphere light is equal to that of rain image. In practice, the values of transmission at rain pixel x with high intensity approach 0 (i.e., $\mathbf{T_s}(x) + \mathbf{T_v}(x) \approx 0$). Our model in Eq. 2 can be approximated by:

$$\mathbf{I}(x) = [1 - (\mathbf{T_s}(x) + \mathbf{T_v}(x))] * \mathbf{A}(x) \tag{8}$$

where the maximum value of $\mathbf{I}(x)$ at rain streak pixels is $\mathbf{A}(x)$. We use [25] to detect rainy pixels in $\mathbf{I}$ and identify the maximum intensity value as ground truth atmospheric light $\mathbf{A}$. In this simplified form, we obtain labels for ANet

Algorithm 2 Pretraining SNet

Input: Rainy images $\{\mathbf{I}^{\{t\}}\}$ and ground truth background $\{\mathbf{J}^{\{t\}}\}$.
Initialization: $\mathbf{T_v} = 0$ in Eq. (3), $\mathcal{A}(\cdot)$ is initialized with pretrained model in Alg. 1.
for $i = 1$ to epoch **do**
  for $j = 1$ to batchnum **do**
    Calculate $\mathcal{A}(\mathbf{I}^{\{t\}})$ for $\{\mathbf{I}^{\{t\}}\}$ via Eq. (6).
    Calculate $\mathcal{S}(\mathbf{I}^{\{t\}})$ via Eq. (4).
    Calculate $\mathcal{J}(\mathbf{I}^{\{t\}})$ via Eq. (3).
    Updating $\mathcal{S}(\cdot)$ and fine tuning $\mathcal{A}(\cdot)$ via loss (10).
  end for
end for
Output: Learned transmission map $\mathbf{T_s}$ of rain streaks.

and train it using the following form:

$$\mathcal{L}_{\mathbf{A}} = \frac{1}{N}\sum_{t=1}^{N} ||\mathcal{A}(\mathbf{I}_t) - \mathbf{A}_t||^2 \tag{9}$$

where N is the number of training samples. The algorithm is in Algorithm 1.

Pretraining SNet. We pretrain SNet by assuming an ideal case where vapors do not contribute to the transmission (i.e., $\mathbf{T_v} = 0$). We use the input rain image $\mathbf{I}$ and ground truth restoration image $\mathbf{J}$ to train SNet. The objective function can be written as follows:

$$\mathcal{L}_{\mathbf{S}} = \frac{1}{N}\sum_{t=1}^{N} ||\mathcal{J}(\mathbf{I}_t) - \mathbf{J}_t||_F^2 \tag{10}$$

where $\mathcal{J}(\mathbf{I}_t) = \{\mathbf{I}_t - [\mathbb{1} - \mathcal{S}(\mathbf{I}_t)] \odot \mathcal{A}(\mathbf{I}_t)\} \oslash \mathcal{S}(\mathbf{I}_t)$ derives from Eq. 3. More details are shown in Algorithm 2.

Joint Training After pretraining ANet and SNet, we perform joint training of the whole network. The overall objective function can be written as:

$$\mathcal{L}_{\text{total}} = \lambda_1 \cdot \frac{1}{N}\sum_{t=1}^{N} ||\nabla\mathcal{J}(\mathbf{I}_t) - \nabla\mathbf{J}_t||_F^2 + \lambda_2 \cdot \frac{1}{N}\sum_{t=1}^{N} ||\mathcal{J}(\mathbf{I}_t) - \mathbf{J}_t||_1$$

where ∇ is the gradient operator in both horizontal and vertical directions, λ_1 and λ_2 are constant weights. The value $\mathcal{J}(\mathbf{I}_t)$ is from Eq. (3) consisting of $\mathbf{T_s}$, $\mathbf{T_v}$, and $\mathbf{A}$. These variables are predicted from SNet, VNet, and ANet, respectively. We perform joint training to these networks. As VNet takes the concatenation of $\mathbf{I}$ and $\mathbf{T_s}$ as input, we back propagate the network gradient to SNet via VNet. The details of our joint training is shown in Algorithm 3.

Algorithm 3 Joint training

Input: Rainy images $\{\mathbf{I}^{\{t\}}\}$ and ground truth background $\{\mathbf{J}^{\{t\}}\}$.
Initialization: $\mathcal{A}(\cdot)$ is initialized with fine tuned model in Alg. 2, $\mathcal{S}(\cdot)$ is initialized with pretrained model in Alg. 2.

for $i = 1$ to epoch **do**
  for $j = 1$ to batchnum **do**
    Calculate $\mathcal{A}(\mathbf{I}^{\{t\}})$ for $\{\mathbf{I}^{\{t\}}\}$ via Eq. (6).
    Calculate $\mathcal{S}(\mathbf{I}^{\{t\}})$ via Eq. (4).
    Calculate $\mathcal{V}(\text{cat}(\mathbf{I}^{\{t\}}, \mathcal{S}(\mathbf{I}^{\{t\}})))$ via Eq. (5).
    Calculate $\mathcal{J}(\mathbf{I}^{\{t\}})$ via Eq. (3).
    Updating $\mathcal{V}(\cdot)$ and fine tuning $\mathcal{A}(\cdot)$ and $\mathcal{S}(\cdot)$ via loss (11).
  end for
end for

Output: Learned transmission map $\mathbf{T_s}$ of rain streaks.

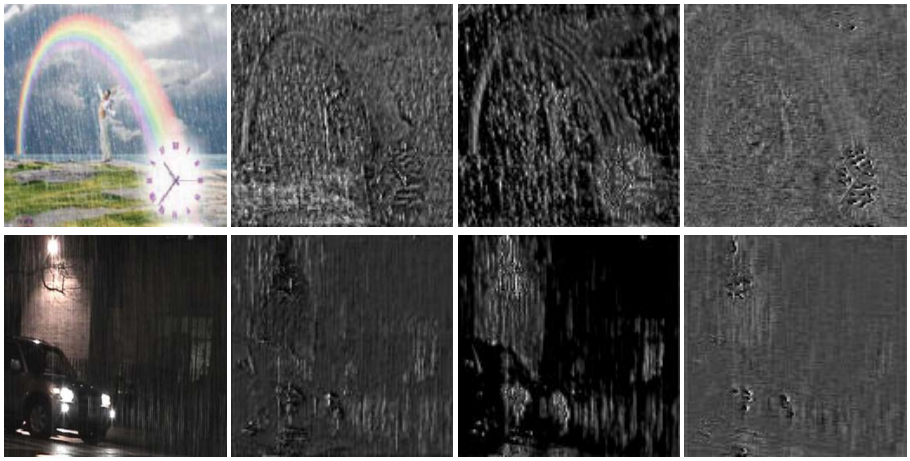
3.5 Visualizations

We visualize the intermediate results of our method to verify the effectiveness of our network. In Section 3.1, we extract the features of rainy image by 3 separate convolution groups. We show the learned feature maps of different convolution groups in Fig. 4. The (a)-(c) shows that different groups contain different features of rain streaks in various shapes and sizes. The first group extracts slim rain streaks and their shapes are similar, the second group contains wide rain features and the shapes are diversified. The third group captures homogeneous feature representations resembling vapors.

Our rain model allows for the anisotropic transmission map of rain streaks, the homogeneous transmission map of vapor and the atmospheric light of rainy scenes. In Fig. 5, we display the learned transmission map $\mathbf{T_s}$ of rain streaks, the transmission map $\mathbf{T_v}$ of vapor and the atmospheric light $\mathbf{A}$. We can see that $\mathbf{T_s}$ captures the various rain streak information and contains the anisotropy of rainy scenes. While $\mathbf{T_v}$ models the influence of vapor, it possesses the similar values in local areas and different areas are separated by object contours. $\mathbf{A}$ keeps relatively high values, which reflects the fact that atmospheric light possesses high illumination in rainy scenes.

4 Experiments

To assess the performance of our deraining method quantitatively, the commonly used PSNR and SSIM [24] are used as our metrics. In order to evaluate our deraining network more robustly, we measure the quality of deraining results by calculating their Frechet Inception Distance (FID) [6] to the ground truth background. FID is defined via the deep features extracted by Inception-V3 [20], smaller values of FID indicate more similar deraining results to the ground truth. For visual quality evaluation, we show some restored results of real-world and synthetic rainy images. Existing methods [15,26,13,21] are selected to make



(a) (b) (c) (d)

Fig. 4. Feature maps of different convolution groups. (a) Input rainy images. (b) Features in 1st group. (c) Features in 2nd group. (d) Features in 3rd group. The features of rain streaks in the first group is always slim, the second group extract rain streaks with relatively large size, and the third group contains features of homogeneous vapor.

Table 1. PSNR and SSIM of our ablation studies

Datasets	Rain-I		Rain-II	
Metric	PSNR	SSIM	PSNR	SSIM
C_1	27.15	0.772	25.48	0.793
C_2	27.49	0.806	28.57	0.844
C_3	31.30	0.897	33.86	0.930
Ours	**31.34**	**0.908**	**34.42**	**0.938**

complete comparisons in our paper. The comparisons with another two methods [27,4] are provided in the supplementary file. Except for [26,13,27] which need additional ground truth configuration, these methods are retrained on the same dataset for fair comparisons.

In the training process, we crop 256×256 patches from the training samples, and Adam [12] is used to optimize our network. The learning rate for pretraining ANet is 0.001. While learning $\mathbf{T_s}$, loss $\mathcal{L}_\mathbf{S}$ is to train SNet and fine tune ANet in a joint way, the learning rate for SNet is 0.001 and the learning rate for ANet is 10^{-6}. Similarly, in the stage of jointly learning $\mathbf{T_v}$, the learning rate for VNet is 0.001 and the learning rate for SNet and ANet is 10^{-6}. The hyper-parameters λ_1 and λ_2 in Eq. (11) are 0.01 and 1 respectively. Our network is trained on a PC with NVIDIA 1080Ti GPU based on PyTorch framework. The training is converged at the 20-*th* epoch. Our code will be released publicly.

Fig. 5. Transmission map of rain streaks and vapor. Input rainy images are shown in (a). Transmission maps $\mathbf{T_s}$ of rain streaks are in (b). Transmission maps $\mathbf{T_v}$ of vapor are in (c). Deraining results with only using $\mathbf{T_s}$ are in (d). Deraining results with $\mathbf{T_v}$ involved are in (e). The Removed rain streaks are shown in (f) and the removed vapors are shown in (g). $\mathbf{T_s}$ is shown to capture anisotropic rain streaks while $\mathbf{T_v}$ models homogeneous vapors.

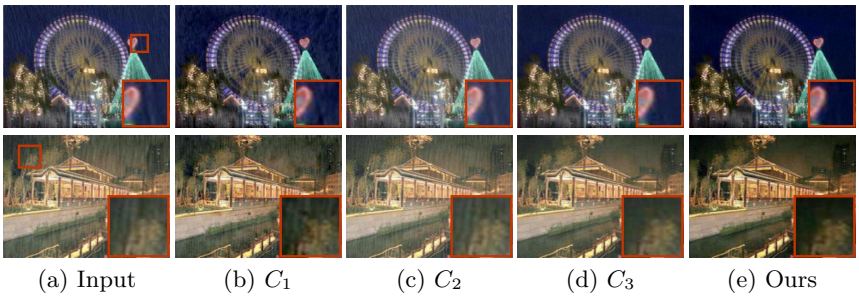


Fig. 6. Visual results of ablation studies. (a) Input rainy images. (b)-(e) Deraining results under C_1, C_2, C_3 and the whole pipeline, respectively.

4.1 Dataset Constructions

We follow [14] to prepare training dataset where there are 20800 training pairs. The rainy image in each pair is synthesized with ground truth and rendered rainy layer by using screen blend mode. Our evaluation datasets consists of three parts. First, we randomly select 100 images from each dataset in [27,15,4,25], which brings 400 images in total and named as Rain-I. Second, we synthesize 400 images[4] where the synthetic rainy images possess apparent vapor, which is named as Rain-II. Third, we follow the real-world dataset [21] and name it as Rain-III. The real-world rainy images are collected from either existing works or Internet data. The independence between our training and testing datasets ensures the generalization of proposed method.

[4] http://www.photoshopessentials.com/photo-effects/rain/

Table 2. PSNR/SSIM comparisons on our three datasets

Methods	[15]	[26]	[13]	[21]	Ours
Rain-I	27.51/0.897	27.69/0.898	17.96/0.675	28.43/0.848	**31.34/0.908**
Rain-II	26.68/0.830	29.97/0.893	17.99/0.605	30.53/0.905	**34.42/0.938**
Rain-III	34.78/0.943	28.39/0.902	18.48/0.747	35.10/0.948	**35.91/0.951**

Table 3. FID comparisons on our three datasets

Methods	[15]	[26]	[13]	[21]	Ours
Rain-I	62.71	101.74	104.08	81.54	**50.66**
Rain-II	97.30	134.54	118.10	88.15	**67.18**
Rain-III	81.42	89.63	134.34	80.68	**79.86**

4.2 Ablation Studies

Our network consists of SNet, ANet, and VNet. We show how these networks work together to gradually improve image restoration results. We first remove ANet and VNet. The atmosphere light is estimated via the simplified condition illustrated in Sec. 3.4 to train SNet. This configuration is denoted as C_1. On the other side, we incorporate a pretrained ANet and use its output for SNet training, which is denoted as C_2. Also, we perform joint training of ANet and SNet, which is denoted as C_3. Finally, we jointly train ANet, SNet, and VNet where ANet and SNet are initialized with pretrained models. This configuration is the whole pipeline of our method.

Fig. 6 and Table 1 show the qualitative and quantitative results. We observe that the results from C_2 are of higher quality than those from C_1. The higher quality indicates that estimating atmosphere light in ideal condition is not stable for effectively image restoration as shown in Fig. 6 (b). Compared to C_2, the results from C_3 are more effective to remove rain streaks, which indicates the importance of joint training. However, the vapors are not well removed in the results from C_3. In comparison, by adding VNet to model vapors, we observe haze is further reduced in Fig. 6 (e). The numerical evaluations in Table 1 also indicate the effectiveness of joint training and vapor modeling.

4.3 Evaluations with State-of-the-art

We compare our method with existing deraining methods on three rain datasets (i.e., Rain-I, Rain-II, and Rain-III). The comparisons are categorized as numerical and visual evaluations. The details are presented in the following:

Quantitative evaluation. Table 2 shows the comparison to existing deraining methods under PSRN and SSIM metrics. Overall, our method achieves favorable results. The PSNR of our method is about 4 dB higher than [21] on Rain-II dataset. In Table 3, we show the evaluations under the FID metric. This comparison shows that our method achieves lowest FID scores on all three datasets,

(a) Input (b) [15] (c) [26] (d) [13] (e) [21] (f) Ours (g) GT

Fig. 7. Qualitative comparisons of selected methods and our method on synthetic rainy images. (a) Input rainy images. (b)-(f) Deraining results of RESCAN [15], JORDER [26], PYM+GAN [13], SPANet [21] and our method. (g) Ground truth. These two samples are two failure cases of the state-of-the-art methods.

Table 4. Averaged time cost of comparison methods with a fixed image size of 512×512.

Methods	[15]	[26]	[13]	[21]	Ours
Time	0.47s	1.39s	0.45s	0.66s	0.03s

which indicates that our results resemble most to the ground truth images. The time cost of online inference of comparison methods is shown in Table 4. Our method is able to produce results efficiently.

Qualitative evaluation. We show visual comparison from aspects of synthetic data and real-world data. Fig. 7 shows two synthetic rain images where existing methods are able to restore effectively. In comparison, our method is effective to remove both rain streaks and vapors.

Besides synthetic evaluations, we show visual comparisons on real-world images in Fig. 8. When the rain streaks are heavy as shown on the first row of (a), existing methods do not remove these streaks completely. When the rain streaks are mild as shown on the fourth row, all the comparison methods are able to remove their appearance. When the streak edges are blur as shown on the second row, the results of RESCAN and ours are able to faithfully restore while PYM+GAN tends to change the whole color perceptions in Fig. 8(d). Meanwhile, there are artifacts and blocking effects appear on the third row of Fig. 8(d). The limitations also arise in JORDER and SPANET where details are missing shown in Fig. 8(c) and heavy rain streaks remain in Fig. 8(d). Compared to existing methods, our method is able to effectively model both rain streaks and vapors. By jointing training of three subnetworks, the parameters of our visual model is accurately predicted and produces visually pleasing results.

(a) Input (b) [15] (c) [26] (d) [13] (e) [21] (f) Ours

Fig. 8. Qualitative comparisons of comparison methods on real-world rainy images. Input images are shown in (a). Deraining results of RESCAN [15], JORDER [26], PYM+GAN [13], SPANet [21] and our method are shown from (b) to (f), respectively.

5 Concluding Remarks

"Rain is grace; rain is the sky condescending to the earth; without rain, there would be no life."

— John Updike

Rain nourishes daily life except visual recognition systems. Recent studies on rain image restoration propose models to calculate background images according to rain image formations. A limitation occurs that the appearance of rain consists of rain streaks and vapors, which are entangled with each other in the rain images. We rethink rain image formation by formulating both rain streaks and vapors as transmission medium. We propose two networks to learn transmission maps and atmosphere light that constitute rain image formation. These essential elements in the proposed model are effectively learned via joint network training. Experiments on the benchmark dataset indicate the proposed method performs favorably against state-of-the-art approaches.

Acknowledgement. This work was supported by NSFC (60906119) and Shanghai Pujiang Program.

References

1. Chen, D., Chen, C.c., Kang, L.: Visual depth guided color image rain streaks removal using sparse coding. IEEE Transactions on Circuits and Systems for Video Technology **24**(8), 1430–1455 (Aug 2014)
2. Chen, Y.L., Hsu, C.T.: A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In: IEEE International Conference on Computer Vision (ICCV 2013). pp. 1968–1975. IEEE, Sydney, Australia (Dec 2013)
3. Fu, X., Huang, J., Ding, X., Liao, Y., Paisley, J.: Clearing the skies: a deep network architecture for single-image rain removal. IEEE Transactions on Image Processing **26**(6), 2944–2956 (July 2017)
4. Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., Paisley, J.: Removing rain from single images via a deep detail network. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2017). pp. 1715–1723. IEEE, Honolulu, HI, USA (July 2017)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2017). IEEE (July 2015)
6. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: International Conference on Neural Information Processing Systems (NIPS 2017). pp. 6629–6640. Long Beach, USA (Dec 2017)
7. Hu, X., Fu, C.w., Zhu, L., Heng, P.a.: Depth-attentional features for single-image rain removal. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2019). IEEE, Long Beach CA, USA (Jun 2019)
8. Huang, D.A., Kang, L.W., Wang, Y.C.F., Lin, C.W.: Self-learning based image decomposition with applications to single image denoising. IEEE Transactions on Multimedia **16**(1), 83–93 (Jan 2014)
9. Huang, D.A., Kang, L., Yang, M.C., Lin, C.W., Wang, Y.C.F.: Context-aware single image rain removal. In: IEEE International Conference on Multimedia and Expo (ICME-2012). pp. 164–169. IEEE, Melbourne, Australia (July 2012)
10. Jiang, T.X., Huang, T.Z., Zhao, X.L., Deng, L.J., Wang, Y.: A novel tensor-based video rain streaks removal approach via utilizing discriminatively intrinsic priors. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2017). pp. 2818–2827. IEEE, Honolulu, HI, USA (July 2017)
11. Kang, L., Lin, C.w., Fu, Y.h.: Automatic single-image-based rain streaks removal via image decomposition. IEEE Transactions on Image Processing **21**(4), 1742–1755 (Apr 2012)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: the 3rd International Conference for Learning Representations(ICLR-2015). IEEE, San Diego (2015)
13. Li, R., Cheong, L.F., Tan, R.T.: Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2019). IEEE, Long Beach CA, USA (Jun 2019)
14. Li, S., Ren, W., Zhang, J., Yu, J., Guo, X.: Fast single image rain removal via a deep decomposition-composition network. Computer Vision and Image Understanding **186**, 48–57 (Apr 2018)
15. Li, X., Wu, J., Lin, Z., Liu, H., Zha, H.: Recurrent squeeze-and-excitation context aggregation net for single image deraining. In: European Conference on Computer Vision (ECCV-2018). IEEE, Munich, Germany (Sep 2018)

16. Li, Y., Tan, R.T., Guo, X., Lu, J., Brown, M.S.: Rain streak removal using layer priors. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016). pp. 2736–2744. Las Vegas, Nevada, USA (June 2016)
17. Luo, Y., Xu, Y., Ji, H.: Removing rain from a single image via discriminative sparse coding. In: IEEE International Conference on Computer Vision (ICCV 2015). pp. 3397–3405. Boston, MA, USA (Dec 2015)
18. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online learning for matrix factorization and sparse coding. Journal of Machine Learning Research pp. 19–60 (Mar 2010)
19. Ren, D., Zuo, W., Hu, Q., Zhu, P., Meng, D.: Progressive image deraining networks: a better and simpler baseline. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2019). IEEE, Long Beach CA, USA (Jun 2019)
20. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016). pp. 2818–2826. Las Vegas, NV, USA (Jun 2016)
21. Wang, T., Yang, X., Xu, K., Chen, S., Zhang, Q., Lau, R.W.: Spatial attentive single-image deraining with a high quality real rain dataset. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2019). IEEE, Long Beach CA, USA (Jun 2019)
22. Wang, Y., Chen, C., Zhu, S., Zeng, B.: A framework of single-image deraining method based on analysis of rain characteristics. In: IEEE International Conference on Image Processing (ICIP 2013). pp. 4087 – 4091. IEEE, Phoenix, USA (Sep 2016)
23. Wang, Y., Liu, S., Chen, C., Zeng, B.: A hierarchical approach for rain or snow removing in a single color image. IEEE Transactions on Image Processing **26**(8), 3936–3950 (August 2017)
24. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (April 2004)
25. Yang, W., Tan, R.T., Feng, J., Liu, J., Guo, Z., Yan, S.: Deep joint rain detection and removal from a single image. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2017). pp. 1685–1694. IEEE, Honolulu, HI, USA (July 2017)
26. Yang, W., Tan, R.T., Feng, J., Liu, J., Yan, S., Guo, Z.: Joint rain detection and removal from a single image with contextualized deep networks. IEEE Transactions on Pattern Analysis and Machine Intelligence (January 2019)
27. Zhang, H., Patel, V.M.: Density-aware single image de-raining using a multi-stream dense network. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2018). pp. 1685–1694. IEEE, Salt Lake City, UT (July 2018)
28. Zhang, X., Zhou, X., Lin, M., Sun, J.: Shufflenet: An extremely efficient convolutional neural network for mobile devices (2018)
29. Zhang, X., Li, H., Qi, Y., Leow, W.K., Ng, T.K.: Rain removal in video by combining temporal and chromatic properties. In: IEEE International Conference on Multimedia and Expo (ICME 2006). vol. 1, pp. 461–464. IEEE, Toronto, Ontario, Canada (July 2006)