Deep joint rain and haze removal
from a single image

Liang Shen*, Zihan Yue*, Quan Chen, Fan Feng and Jie Ma

Institute of Image Recognition and Artificial Intelligence

School of Automation, Huazhong University of Science and Technology, China

{shenliang,yuezihan,chenquan,fengfan,majie}@hust.edu.cn

Abstract—Rain removal from a single image is a challenge which has been studied for a long time. In this paper, a novel convolutional neural network based on wavelet and dark channel is proposed. On one hand, we think that rain streaks correspond to high frequency component of the image. Therefore, haar wavelet transform is a good choice to separate the rain streaks and background to some extent. More specifically, the LL subband of a rain image is more inclined to express the background information, while HL, LH subband tend to represent the rain streaks and the edges respectively. On the other hand, the accumulation of rain streaks from long distance makes the rain image look like haze veil. We extract dark channel of rain image as a feature map in network. By increasing this mapping between the dark channel of input and output images, we achieve haze removal in an indirect way. All of the parameters are optimized by back-propagation. Experiments on both synthetic and real-world datasets reveal that our method outperforms other state-of-the-art methods from a qualitative and quantitative perspective.

I. INTRODUCTION

Most computer vision tasks assume the sufficient high-quality of images. However, various degradations often occur in realistic scenes. For example, rainy weather becomes an inevitable situation when these tasks are applied to outdoor scenes. The rain in image can be roughly divided into two cases. Rain streaks near to the camera lens can be considered as noise in the image, whereas rain from long distance looks like fog.

In this paper, a novel convolutional neural network based on wavelet [1] and dark channel [2] has been proposed to remove different types of rain. On one hand, considering that rain streaks mainly correspond to the high frequency components in the image, we think that the wavelet-based approach is probably a good choice. Firstly, the rain image and the ground truth are transformed into four sub-images (low-low, low-high, high-low, high-high frequency) by using Haar wavelet [1] respectively. Then we try to train an end-to-end mapping between these different sub-images in wavelet domain to remove the light rain. On the other hand, the accumulation of rain streaks from long distance makes the image overcast as if covered by haze. In this condition, the dark channel prior proposed by He *et al.*[2] can still be considered as a good approach to remove the veil from an image. However, in this model we regard dark channel as a

feature map in convolutional neural network. By combining above two different methods in a consistent framework, the final model is considered as a multi-task optimization problem and all parameters are optimized by back-propagation.

II. RELATED WORKS

Similar to many image restoration problems such as image denoising [3], image deblurring [4], super resolution [5] and image dehazing [6], image deraining is also a noticeable field. The common point of these methods is to solve the inverse problem by using the degraded images. As we all know, inverse problems are the pathological problems with infinite solutions. To overcome this difficulty, many prior-based and regularization methods have been proposed. In particular, rain removal can be divided into two groups: Video based methods and single image based methods.

A. Video based methods

Removing rain from video has been widely explored. K-shitiz *et al.*[7] analysed the visual effects of rain on an imaging system. They developed a physics-based blur model that explained the photometry of rain. Barnum *et al.*[8] studied the phenomenon of rain in frequency domain. They revealed that dynamic weathers such as rain and snow have a significant effect in frequency space. Bossu *et al.*[9] separated the foreground from background in image sequences by using a classical Gaussian mixture model. The histogram of orientation of rain streaks maked it possible to detect the pixel of rain in the foreground image. Chen *et al.*[10] proposed a novel low-rank model from matrix to tensor structure to capture the correlated rain streaks. Recently, Jiang *et al.*[11] proposed a novel tensor-based approach by considering the inherent property of rain streaks and cleared the videos. All of these methods make full use of temporal in adjacent frames to figure out rain streaks in video.

B. Single image methods

Compared with multi-frame rain removal, single image deraining is more difficult due to the lack of temporal information. Traditional methods are usually based on image decomposition, sparse coding or dictionary learning. For example, Fu *et al.*[12] treated the rain removal as an image decomposition problem by using morphological component analysis. Li *et al.*[13] tried to use simple patch-based priors for

---

* indicates equal contribution by authors.

(a) Ground truth

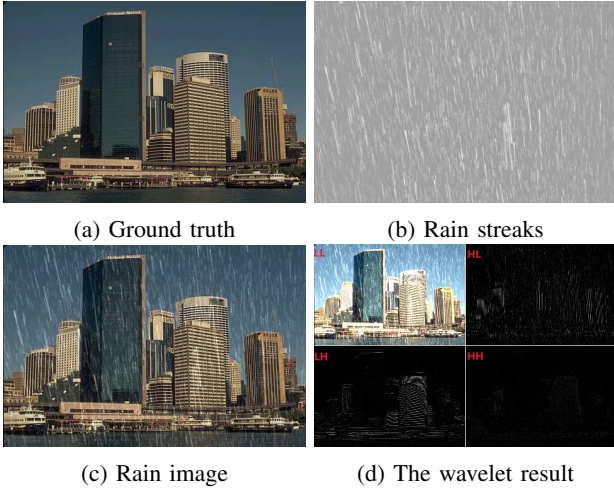(b) Rain streaks

(c) Rain image

(d) The wavelet result

Fig. 1: (a)ground truth,(b)image with rain streaks,(c) rain image formed by (a) and (b).(d) are the four sub-images of haar wavelet transform result. Obviously, HL subband contains more raindrops due to the fact that rain is falling down from the top, by the contrast, LH subband contains more the ground truth edge.(Best zoom in the LH and HL subband on screen)

both foreground and background. Xu *et al.*[14] used filtering method to remove rain by guiding image such as guider filter proposed by He *et al.*[2]. In [15], Luo proposed a dictionary learning method for single image deraining. Besides these, deep learning makes a great achievement in many low-level vision tasks. Dong *et al.*[5] attempted to use convolutional neural network in image super-resolution for the first time and achieved remarkable improvement. After this, a large number of similar methods spewed out. For instance, Chakrabarti and Ayan [16] proposed a novel neural approach for blind motion deblurring which uses the trained network to compute sharp image patches. Cai *et al.*[6] proposed an end-to-end system named dehazeNet which is used to estimate the medium transmission. Fu *et al.*[17] directly learned the mapping between rain image and high-frequency detail image by using the residual structure proposed in [18]. Due to the fact that rain streaks removal in an image is almost an identity mapping, residual structure will make the learning process easier. Yang *et al.*[19] constructed a multi-task that solved the inverse problem through an end-to-end learning. They also proposed a novel network for extracting the rain discriminative feature to leverage more content.

III. PROPOSED METHOD

In this section, we elaborate that rain in image can be roughly divided into two situations. Rain streaks near to the lens look like noise in an image, whereas rain from distance looks like haze veil. Our model takes the above two aspects into account. At last, by combining above two separated structures to one network, the final model can be considered as an end-to-end framework for rain and haze removal.
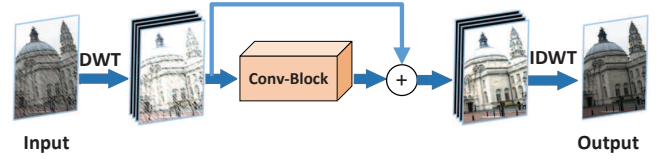

Fig. 2: The simple rain removal network architecture(SRR-net) for removing the rain image without the haze veil

A. Rain model in an image

Traditional rain model is composed of two components: rain streaks and background. Mathematically, it can be expressed as:

$$O = B + R \tag{1}$$

Where O represents the observed degraded image, B is the background scene and R are the rain streaks. However, in many cases where the accumulation of rain streaks from long distance makes the image overcast as if covered by haze veil, which causes the model perform not well. Based on this foundation, many modified models have been proposed. For example, considering the dense rain and fog phenomenon, Yang *et al.*[19] extended Equation 1 to create a new model to accommodate them:

$$O = \alpha(B + \sum_{t=1}^{s} R_t) + (1 - \alpha)A \tag{2}$$

As we can see, the first item of above equation consists of several different layers of rain streaks. A and α in the second item are global atmospheric light and scene transmission respectively, which have been described in haze removal papers such as [2].

B. Wavelet for rain streaks removal

Fourier transform is a nice tool for analysing images in the frequency domain, however, the spectrum of an image loses a lot of great properties such as local receptive field, which makes it difficult to use convolutional neural network. Fortunately, the other frequently-used method called wavelet transform is now making it easier to analyse the images. Different from Fourier transform based on sinusoids, wavelet transform is based on small waves, which is more convenient to train. In this paper, we attempt to use one of the most commonly used wavelet: Haar wavelet.

Figure 1 gives an example of discrete wavelet transform using Haar basis function. As we can see, the four sub-images correspond to the approximation subband LL, horizontal detail LH, vertical detail HL and diagonal detail HH, respectively. LL subband represents the main content of an image, whereas LH, HL and HH represent the detail information of an image. More specifically, the LL subband of a rain image is more inclined to express the background information, and HL subband contains more raindrops due to the fact that the rain is falling down from the top, while the LH subband includes the more edge information of the background. Therefore, this decomposition is not only helpful to get rid of the rain noise but also to protect the edge details.
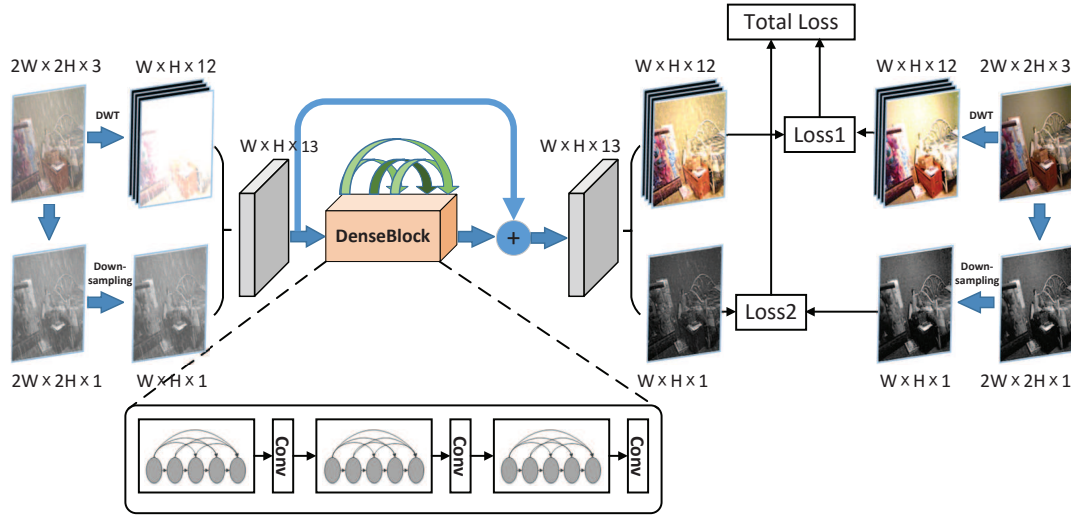
Fig. 3: Deep joint rain and haze removal network(DJRHR-net), which is training for removing the rain and the haze veil.

The goal of rain removal is to recover low-quality(LQ) images to high-quality(HQ) images. Considering that wavelet subbands can better represent the shape of rain streaks and edges, our network tries to fit these subbands. Firstly, we convert original image pairs (LQ, HQ) to wavelet domain, which are considered as input image and train label respectively:

$$X : \{LQ_{LL}, LQ_{LH}, LQ_{HL}, LQ_{HH}\} = DWT(LQ)$$
$$Y : \{HQ_{LL}, HQ_{LH}, HQ_{HL}, HQ_{HH}\} = DWT(HQ)$$
$$(3)$$

Inspired by the residual structure [18], Fu *et al.*[17] proposed an end-to-end network between rain image and high-frequency detail image. Compared with their method which directly learns the mapping between original image pairs, our model attempts to fit the mapping between these sub-images generated by wavelet as shown in Figure 2. Mathematically, the loss function of SRR-net can be defined as:

$$L = \frac{1}{N} \sum_{i=1}^{N} \|Y_i - X_i - f(X_i)\|_F^2 + \lambda \|W\|_F^2 \qquad (4)$$

Where X_i and Y_i represent the tensors that concatenate four wavelet subbands of LQ and HQ images respectively, N is equal to the number of the images on training dataset. $f(\cdot)$ represents the nonlinear mapping of the neural network, and in this paper we use ordinary convolution and ReLU layers. W corresponds to the parameters of the whole model which are optimized by back-propagation, and $\|\cdot\|_F$ represents the Frobenius norm. More experimental results and training parameter setting will be elaborated in section IV.

To summmarize, after the whole network(SRR-net) is trained, the whole rain removal process is as follows:

Firstly, we convert the rain image to wavelet domain by Haar wavelet and concatenate these four subbands to a tensor X with 12 channels.

$$X : \{LQ_{LL}, LQ_{LH}, LQ_{HL}, LQ_{HH}\} = DWT(LQ) \quad (5)$$

Next, we put the wavelet subbands X to the trained residual network.

$$Y = f(X) + X \qquad (6)$$

At last, the inverse wavelet transform is used to generate final high-quality result.

$$HQ = IDWT(Y) \qquad (7)$$

C. Dark channels for rain accumulation

In the previous subsection, we have explained the role of wavelet transform in deraining. However, the simple rain removal network can't handle the situation very well, where rain streaks are dense and these make the image have the haze veil. Haze removal is a traditional research direction which has been studied for a long time. One of the most classical methods is based on dark channel prior [2], which is a statistic of the haze-free images. By using this extra strong prior, the thickness of haze can be estimated and the high-quality image can be recovered directly.

In order to integrate the de-hazing into deep learning framework, we extract dark channel of an image as a feature map in convolutional neural network to contribute to the removal of this noise. It is more effective to add the artificial feature directly than the features learned by the deep network. So we increase a mapping between the dark channel of input and output images, which helps achieve haze removal through indirect means. Figure 3 shows our final deep joint rain and haze removal network(DJRHR-net) which is designed as a multi-task architecture.

As we can see, the original LQ(rain image) and HQ(ground truth) should be converted into wavelet subbands and dark channel firstly.

$$X : \{DWT(LQ), dark\_channel(LQ)\}$$
$$Y : \{DWT(HQ), dark\_channel(HQ)\}$$
$$(8)$$

(a) Ground truth   (b) Synthetic data   (c) SIRR-net[20]   (d) Detail-net[17]   (e) JORDER[19]   (f) Our SRR-net

Fig. 4: Results using different methods on synthesized test images

TABLE I: Quantitative measurement results using PSNR/SSIM/NIQE on synthesized test images

Method	1st row			2nd row			3rd row		
Metric	PSNR	SSIM	NIQE	PSNR	SSIM	NIQE	PSNR	SSIM	NIQE
Synthetic image	26.69	0.85	4.10	29.73	0.86	4.00	21.22	0.72	3.33
SIRR-net[20]	26.86	0.89	3.51	29.22	0.93	2.99	20.99	0.74	3.04
Detail-net[17]	28.61	0.91	4.02	30.20	0.95	2.95	21.88	0.76	2.93
JORDER[19]	21.90	0.86	3.34	22.03	0.88	2.73	21.03	0.74	3.23
Ours	**32.44**	**0.95**	**2.32**	**34.38**	**0.96**	**2.48**	**22.29**	**0.77**	**2.66**

Where

$$dark\_channel(R,G,B) = \min\{R,G,B\} \qquad (9)$$

For simplicity, we define

$$
\begin{aligned}
X_1 &= DWT(LQ), X_2 = dark\_channel(LQ) \\
Y_1 &= DWT(HQ), Y_2 = dark\_channel(HQ) \\
X &= \{X_1, X_2\}, Y = \{Y_1, Y_2\}
\end{aligned}
\qquad (10)
$$

In consideration of the two above aspects, we propose a novel training method. On one hand, four wavelet subbands and dark channel of low-quality image pass through the same convolutional layers:

$$\hat{Y} = f(\{X_1, X_2\}) + \{X_1, X_2\} \qquad (11)$$

Where $\{\cdot, \cdot\}$ means concatenating the two tensors to a tensor and $f(\cdot)$ represents the convolutional architecture. In this paper we use dense-net proposed in [21], which is better to represent the features of the image and obtains the faster convergence.

On the other hand, the loss function of these two results are evaluated separately. As for the first task, we should make the wavelet transform of rain image Y_1 restore to this of ground truth $\hat{Y}_1$ under the criteria of Frobenius norm. We set the loss L_1 to:

$$L_1 = \frac{1}{N} \sum_{i=1}^{N} \left\| Y_{1i} - \hat{Y}_{1i} \right\|_F^2 \qquad (12)$$

Meanwhile, we make the dark channel feature of the rain image Y_2 and this feature of ground truth $\hat{Y}_2$ as close as possible, which is useful to detect the area of haze veil and remove it. The loss L_2 is set:

$$L_2 = \frac{1}{N} \sum_{i=1}^{N} \left\| Y_{2i} - \hat{Y}_{2i} \right\|_F^2 \qquad (13)$$

We combine above two loss functions, L_1 and L_2, and obtain the final goal of optimization L_{total} which is:

$$L_{total} = L_1 + \alpha L_2 \qquad (14)$$

Where α plays the balance role of L_1 and L_2, we empirically set α to 1 in this paper because we find that the result of our method is insensitive to the different value of α in a large range. The parameters in this model are optimized by back-propagation. After DJRHR-net has been trained, the whole rain removal process is similar to last section:

Firstly, we generate the wavelet subbands and dark channel of a rain image.

$$X_1 = DWT(LQ), X_2 = dark\_channel(LQ) \qquad (15)$$

Then, we concatenate the X_1 and X_2 to a tensor with 13 channels and pass it through the trained residual network(DJRHR-net).

$$\{\hat{Y}_1, \hat{Y}_2\} = f(\{X_1, X_2\}) + \{X_1, X_2\} \qquad (16)$$

At last, the inverse wavelet transform is used to generate final high-quality result.

$$HQ = IDWT(\hat{Y}_1) \qquad (17)$$

We have to emphasize that the dark channel feature in this model is just used for removing the haze in an indirect way. In training process, the dark channel error is used for updating the weights of convolutional network. But in test process,

| (a) Real-world image | (b) SIRR-net[20] | (c) Detail-net[17] | (d) JORDER[19] | (e) Our DJRHR-net |

Fig. 5: Results using different methods on real-world images

considering that the train result of dark channel feature is different from the dark channel of wavelet subbands, we preserve the wavelet subbands and discard dark channel, as Equation 17 shows.

IV. EXPERIMENT

In last section, we propose two networks, SRR-net and DJRHR-net, which process the sparse and dense rain respectively. To evaluate the performance of our method, we use both the synthetic test data and the real-world images to compare our approch with two recent state-of-the-art deraining methods based on network, which contains Detail-net [17], JORDER [19] and SIRR-net[20], which removes the final enhancement for fair.

A. Dataset generation

For learning the parameters of the SRR-net and DJRHR-net, we construct two datasets to deal with different situations. As for the rain images without the haze veil, we simply make use of the dataset from [17] as ground truth and add 12 types of rain streaks [13] to obtain **TrainSet A**.

Furthermore, in order to train the parameters of our DJRHR-net, we create a new dataset as **TrainSet B**, which contains a number of low-quality(LQ) and high-quality(HQ) image pairs with rain and haze veil noises. In view of the fact that the formation of haze is based on the depth information of the images, we firstly select 1449 RGBD images from the NYU Depth Dataset V2 [22] and generate the haze according to the atmospheric scattering model. Next, we also increase the 12 types of rain streaks [13] to these foggy images. Figure 6 shows the part of the TrainSet B.

Fig. 6: Some images in TrainSet B. The top row is the ground truth in NYU dataset, the middle row is the image with haze veil and the bottom row is the final image with rain and haze.

Besides the experiment on our own synthetic dataset, Train-Set A and TrainSet B, we also choose real-world rain data [13] to evaluate our method.

B. Training setup

For the SRR-net, we simply set the depth of the network to 20. We spend about 8 hours on training the SRR-net by using the Caffe [23] and use Adam with weight decay of 10^{-6} and mini-batch size of 64. For DJRHR-net, we remove the batch normalization and pooling layer to get better regression effect. Besides, we set the growth rate K to 12, the number of the denseblocks L is 3. We use the pytorch to construct the network and use Adam with weight decay of 10^{-4} and mini-batch size of 10. We start with a learning rate of 10^{-3} and the learning rate decay of 0.95.

C. Experiment on synthetic rain data

Figure 4 shows the visual comparison for several methods on synthesized rain images. As we can see, the results of SIRR-net [20], Detail-net [17] and JORDER [19] look unnatural and remove the rain streaks not well, while our method achieves better performance.

Considering that the ground truth is known for the synthetic test data, we use PSNR, SSIM [24] and NIQE [25] for a quantitative evaluation. A higher PSNR or SSIM indicates that the image is closer to the ground truth, but the lower NIQE means a higher image quality. All the best results are boldfaced. As shown in Table I, our SRR-net obtains higher PSNR/SSIM and lower NIQE average than other methods for 200 test images.

D. Experiment on real-world rain data

Figure 5 also shows the results of several state-of-the-art methods on the real-word images. As shown in each row, our method DJRHR-net always achieves better performance than others in the aspect of rain and haze veil removal. As for the heavy rain, DJRHR-net is valid to remove these noises.

E. Study of SRR-net and DJRHR-net parameters

The number of the denseblocks L and the growth rate K are the main hyper-parameters in our DJRHR-net. As shown in Table II, we know that the deeper structure can improve the learning ability. For the better performance, we set the $K = 12$ and $L = 3$ for our experiments above.

TABLE II: Average PSNR for different network parameters

	$K = 8$	$K = 10$	$K = 12$
$L = 1$	14.85	14.90	15.40
$L = 2$	15.10	15.20	15.69
$L = 3$	15.50	15.55	15.75

V. CONCLUSION

In this paper, we propose a novel convolutional neural network based on wavelet and dark channel. Considering that rain streaks correspond to high frequency component of the image, we attempt to use wavelet transform to separate the rain streaks and background. More specifically, HL, LH of the rain image are more inclined to represent the raindrops and the edges of the ground truth respectively. However, the dense rain makes the image look like haze veil. So we extract dark channel as a feature map in network, which plays an important role in removing the haze veil. Finally, we design two architectures, SRR-net and DJRHR-net to process the sparse and dense rain streaks respectively and test our model on both synthetic and real-world images, all of which obtain very impressive performance.

REFERENCES

[1] B. Vidakovic and P. Mller, *An Introduction to Wavelets*. Springer New York, 1999.

[2] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2011.

[3] D. Liu, B. Wen, X. Liu, and T. S. Huang, "When image denoising meets high-level vision tasks: A deep learning approach," *arXiv preprint arXiv:1706.04284*, 2017.

[4] I. Vasiljevic, A. Chakrabarti, and G. Shakhnarovich, "Examining the impact of blur on recognition by convolutional networks," *arXiv preprint arXiv:1611.05760*, 2016.

[5] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2016.

[6] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.

[7] K. Garg and S. K. Nayar, "When does a camera see rain?" in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 2. IEEE, 2005, pp. 1067–1074.

[8] P. C. Barnum, S. Narasimhan, and T. Kanade, "Analysis of rain and snow in frequency space," *International journal of computer vision*, vol. 86, no. 2, pp. 256–274, 2010.

[9] J. Bossu, N. Hautière, and J.-P. Tarel, "Rain or snow detection in image sequences through use of a histogram of orientation of streaks," *International journal of computer vision*, vol. 93, no. 3, pp. 348–367, 2011.

[10] Y.-L. Chen and C.-T. Hsu, "A generalized low-rank appearance model for spatio-temporally correlated rain streaks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1968–1975.

[11] T.-X. Jiang, T.-Z. Huang, X.-L. Zhao, L.-J. Deng, and Y. Wang, "A novel tensor-based video rain streaks removal approach via utilizing discriminatively intrinsic priors," in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR17)*, 2017.

[12] Y.-H. Fu, L.-W. Kang, C.-W. Lin, and C.-T. Hsu, "Single-frame-based rain removal via image decomposition," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1453–1456.

[13] Y. Li, R. T. Tan, X. Guo, J. Lu, and M. S. Brown, "Rain streak removal using layer priors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2736–2744.

[14] J. Xu, W. Zhao, P. Liu, and X. Tang, "An improved guidance image based method to remove rain and snow in a single image," *Computer and Information Science*, vol. 5, no. 3, p. 49, 2012.

[15] Y. Luo, Y. Xu, and H. Ji, "Removing rain from a single image via discriminative sparse coding," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3397–3405.

[16] A. Chakrabarti, "A neural approach to blind motion deblurring," in *European Conference on Computer Vision*. Springer, 2016, pp. 221–235.

[17] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1715–1723.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[19] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan, "Deep joint rain detection and removal from a single image," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1685–1694.

[20] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley, "Clearing the skies: A deep network architecture for single-image rain streaks removal," *IEEE Transactions on Image Processing*, vol. PP, no. 99, pp. 1–1, 2016.

[21] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, "Densely connected convolutional networks," *arXiv preprint arXiv:1608.06993*, 2016.

[22] P. K. Nathan Silberman, Derek Hoiem and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *ECCV*, 2012.

[23] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.

[24] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[25] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a completely blind image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.