De-raining Of Images Using Deconvolutional Neural Network

Rusafid Mirza Pottengad
Department of Computer Science
BITS-Pilani KK Birla Goa Campus
Zuarinagar,Goa
Email: f20150572@goa.bits-pilani.ac.in

***Abstract*—Severe weather conditions such as rain and snow adversely affect the visual quality of images captured under such conditions thus rendering them useless for further usage and sharing. In addition, such degraded images drastically affect performance of vision systems. Hence, it is important to solve the problem of single image de-raining/de-snowing. However, this is a difficult problem to solve due to its inherent ill-posed nature. Existing approaches attempt to introduce prior information to convert it into a well-posed problem. In this paper, we investigate a new point of view in addressing the single image de-raining problem**

***Keywords*—single image de-raining, de-snowing.**

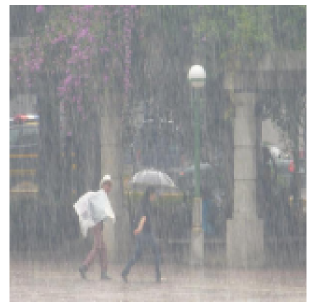

Fig. 1. Input rainy images.

I. INTRODUCTION

It has been widely acknowledged that unpredictable impairments such as illumination, noise and severe weather conditions (i.e. rain, snow and fog) adversely affect the performance of many computer vision algorithms such as detection, classification and tracking. This is primarily due to the fact that these algorithms are trained using images that are captured under well-controlled conditions. For instance, it can be observed from Figure 1, that the presence of heavy rain greatly impairs visual quality of the image, thus rendering face detection and verification algorithms ineffective for such degradations. A possible method to address this issue is to include images captured under unconstrained conditions in the training process of these algorithms. However, it may not be practical to collect such images for all classes in the training set, especially in a large scale setting. In addition, in this age of ubiquitous smartphone usage, images captured by smartphone cameras under difficult weather conditions undergo degradations that drastically affect the visual quality of images making the images useless for sharing and usage. In order to improve the overall quality of such degraded images for better visual appeal and to ensure enhanced performance of vision algorithms, it becomes essential to automatically remove undesirable artifacts arising due to difficult weather conditions discussed above. In this paper, we investigate deconvolution neural network to address this issue.

II. BACKGROUND

A. Single image de-raining

In this section, we briefly review the literature for existing single image de-raining methods, conditional GANs and perceptual loss A Single image de-raining is an extremely challenging task due to its ill-posed nature and unavailability of temporal information which could have been used as additional constraints. Hence, in order to generate optimal solutions to this problem, different kinds of prior information are enforced into the optimization function. Sparse coding-based clustering method [2] is among the first ones to tackle the single image de-raining problem where the authors proposed to solve it in the image decomposition framework. The network consists of two sub-networks: generator G and discriminator D. They first separated the input image into low frequency and high frequency images using a bilateral filter. The high frequency image is further decomposed into rain and non-rain components based on the assumption that learned dictionary atoms can sparsely represent clear background image and rain-streak image separately. An important assumption that is made in this approach is that rain streaks usually have similar edge orientations. This may result in the removal of non-rain component as rain. Also, the methods effectiveness is dependent on the performance of the bilateral filter and clustering of basis vectors for generating sparse representation. Similar to the above approach, Luo et al. in [3] propose a discriminative sparse coding based method that considers the mutual exclusive property into the optimization framework. Though the authors present significant improvements as compared to previous methods, their method is ineffective in removing large rain-streaks due to the assumption that rain streaks are high frequency components. In addition, due to the same assumption, their method generates artifacts around the rain-streak components in the resulting images. In another approach, Chen et al. proposed a low-rank representation-based method [5] that uses patch-rank as a prior to characterize unpredictable rain pattern. They use a low-rank model to capture correlated rain streaks. Observing that dictionary and low-rank based methods tend to leave too many rain pixels in the output image, Li et al. in [4] used the image decomposition framework to propose patch-based priors for background and rain image. These priors are based on GMMs which can accommodate multiple orientations and scales of rain streaks. These methods [4], [5] are based on the assumption that rain streaks have similar patterns and orientations. Due to this assumption, they tend to capture other global repetitive patterns such as brick and texture which results in removal of certain non-rain components from the background image. To address this issue, Zhang et al. recently proposed a convolutional coding-based method [11] that uses a set of learned convolutional low-rank filters to capture the rain pixels. Most recently, due to their immense success in learning non-linear functions, several CNN-based methods have also been proposed to directly learn an end-to-end mapping between input and its corresponding ground truth for de-raining [6], [7], [12].

B. General adversarial networks

Generative Adversarial Networks were proposed by Goodfellow et al. in [13] to synthesize realistic images by effectively learning the distribution of training images. The authors adopted a game theoretic min-max optimization framework to simultaneously train two models: a generative model G and a discriminative model D. The goal of GAN is to train G to produce samples from training distribution such that the synthesized samples are indistinguishable from actual distribution by the discriminator D. Unlike other generative models such as Generative Stochastic Networks [14], GANs do not require a Markov chain for sampling and can be trained using standard gradient descent methods [13]. Initially, the success of GANs was limited as they were known to be unstable to train, often resulting in artifacts in the synthesized images. Radford et al. in [8] proposed Deep Convolutional GANs (DCGANs) to address the issue of instability by including a set of constraints on their topology. Another limiting issue in GANs is that, there is no control on the modes of data being synthesized by the generator in case of these unconditioned generative models. Mirza et al. [15] incorporated additional conditional information in the model, which resulted in effective learning of the generator. The use of conditioning variables for augmenting side information not only increased the stability in learning but also improved the descriptive power of the generator G [16]. Recently, researchers have explored various aspects of GANs such as training improvements [17] and use of task specific cost function [18]. Also, an alternative viewpoint for the discriminator function is explored by Zhao et al. [19] where they deviate from the traditional probabilistic interpretation of the discriminator model. The success of GANs in synthesizing realistic images has led to researchers exploring the GAN framework for numerous applications such as style transfer [20], image inpainting [21], text to image translation [22], image to image translation [23], texture synthesis [24] and generating outdoor scenes from attributes [16]. Isola et al. proposed a general purpose solution for image-to-image translation using conditional adversarial networks. Apart from learning a mapping function, they argue that the network also learns a loss function, eliminating the need for specifying or hand designing a task specific loss function. Karacan et al. in [16] proposed a deep GAN conditioned on semantic layout and scene attributes to synthesize realistic outdoor scene images under different conditions. Recently, Jetchev et al. [24] proposed spatial GANs for texture synthesis. Deviating from traditional GANs, their input noise distribution constitutes a whole spatial tensor instead of a vector, thus enabling them to create architectures more suitable for texture synthesis.

C. Perceptual loss function

Loss functions form an important and integral part of learning process, especially in CNN-based reconstruction tasks. Several works [25][28] have explored different loss functions and their combinations for effective learning for tasks such as super-resolution, semantic segmentation, depth estimation, feature inversion and style transfer. Initial work on CNN based image translation or restoration optimized over pixelwise L2-norm (Euclidean loss) or L1-norm between the predicted and ground truth images [26], [27]. Since these losses operate at pixel level, their ability to capture high level perceptual/contextual details is limited and they tend to produce blurred results [10]. Hence, many authors argue and demonstrate through their results that it would be better to optimize a perceptual loss function where the aim is to minimize perceptual difference between reconstructed image and the ground truth image [29]. In a different approach, the conditional GAN framework can also be considered as an attempt to explore a structured loss function where, a generator network is trained to minimize the discriminators ability to correctly classify between the synthesized image and the corresponding ground truth image. Researchers have attempted to solve various reconstruction tasks such as image super-resolution and style transfer where conditional GAN framework augmented with perceptual and L2 loss function have been used to produce state-of-the-art results [9], [10].

III. Proposed method

In order to address the single image de-raining problem, we aim to directly learn a mapping from an input rainy image to a de-rained (background) image by constructing a CNN network with deconvolution and appropriate skip connections .The network directly learns an end-to-end mapping from input rainy image to its corresponding ground truth. The proposed network with a symmetric structure is shown in the top part of Figure 3. A set of convolutional layers (along with batch normalization and PReLU activation) are stacked in the front which act as a learned feature extractor or semantic attributes extractor. Then, three shrinking layers are stacked in the middle part serving for better computational efficiency. These three shrinking layers can be also regarded as performing linear combination within the learned features [30]. These are followed by a stack of deconvolutional 2 layers (along with batch normalization and ReLU activation function). Note that the deconvolutional layers are a mirrored version of the forward convolutional layers. For all layers, we use a stride of 1 and pad appropriate zeros to maintain the dimension of each feature map to be the same as that of the input. To make the network efficient in training and have better convergence performance, we involve symmetric skip connections into the proposed generator sub-network, similar to [1]. The generator network is as follows: CBP(K)-CBP(K)-CBP(K)-CBP(K)-CBP(K/2)-CBP(1)-DBR(K/2)-DBR(K)-DBR(K)-DBR(K)-DBR(K)-DBR(3)-Tanh where, CBP(K) is a set of K-channel convolutional layers followed by batch normalization and PReLU activation, DBR(K) is a set of K-channel deconvolutional layers followed by batch normalization and ReLU activation.

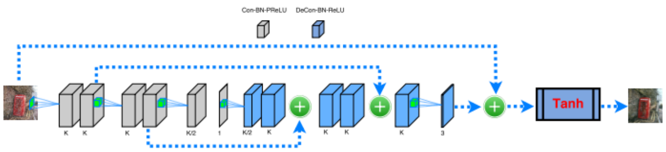

Fig. 2. Network Architecture

IV. Experiments and Results

A. Dataset, training and evaluation detail

The training set consists of a total of 700 images with each single image consisting of the noisy image and ground truth image. Due to lack of a test set, the training was done with taking 20 percent of the training set as validation set. The images were split into two images one with rain streaks and other without it. The images with rain streaks were fed into the network and after training for 30 epochs the accuracy with mean square error as the loss function came close to 90 percent with a validation set accuracy as 89 percent. All input images were resized to 128x128.

B. Model details and parameters

The entire network is trained on a Nvidia Ge-force 820M GPU using the keras framework [31]. We used a batch size of 1 and number of training iterations of 30. Rms algorithm [32] with a learning rate of 0.1 is used. During training, we set K = 64 for the proposed network. All the convolutional and deconvolutional layers in the network are composed of kernels of size 3x3 with a stride 1 and zero padding by 1.

V. Conclusion

In this paper, we proposed a deconvulation network for the removal of rain streaks form a single image. By designing a deep convolution neural network with deconvolution layers and appropriate symmetrically placed skip connection we have managed to achieve a training accuracy close 90 percent. We believe the performance could be further improved by tweaking the parameters of the algorithm or Improving the input image by preprocessing the image before feeding it into the network and then processing it again afterwards.

Acknowledgment

This work was supported by the instructor of the course BITS-F312 Neural Networks And Fuzzy Logic Mr.Tirthiraj Das.

References

[1] X.-J. Mao, C. Shen, and Y.-B. Yang, Image denoising using very deep fully convolutional encoder-decoder networks with symmetric skip connections, arXiv preprint arXiv:1603.09056, 2016.

[2] L.-W. Kang, C.-W. Lin, and Y.-H. Fu, Automatic single-image-based rain streaks removal via image decomposition, IEEE Transactions on Image Processing, vol. 21, no. 4, pp. 17421755, 2012.

[3] 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. 33973405.

[4] Y. Li, R. T. Tan, X. Guo, J. Lu, and M. S. Brown, Rain streak removal using layer priors, in IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 27362744.

[5] Y.-L. Chen and C.-T. Hsu, A generalized low-rank appearance model for spatio-temporally correlated rain streaks, in IEEE International Conference on Computer Vision, 2013, pp. 19681975.

[6] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley, Clearing the Skies: A deep network architecture for single-image rain removal, ArXiv eprints, Sep. 2016.

[7] D. Eigen, D. Krishnan, and R. Fergus, Restoring an image taken through a window covered with dirt or rain, in Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 633640.

[8] A. Radford, L. Metz, and S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, arXiv preprint arXiv:1511.06434, 2015.

[9] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, Image-to-image translation with conditional adversarial networks, arxiv, 2016.

[10] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang et al., Photo-realistic single image super-resolution using a generative adversarial network, arXiv preprint arXiv:1609.04802, 2016.

[11] H. Zhang and V. M. Patel, Convolutional sparse and low-rank codingbased rain streak removal, in 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2017, pp. 19.

[12] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan, Joint rain detection and removal via iterative region dependent multi-task learning, CoRR, vol. abs/1609.07769, 2016. [Online]. Available: http://arxiv.org/abs/1609.07769

[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative adversarial nets, in Advances in Neural Information Processing Systems, 2014, pp. 2672 2680.

[14] E. Thibodeau-Laufer, G. Alain, and J. Yosinski, Deep generative stochastic networks trainable by backprop, 2014.

[15] M. Mirza and S. Osindero, Conditional generative adversarial nets, arXiv preprint arXiv:1411.1784, 2014.

[16] L. Karacan, Z. Akata, A. Erdem, and E. Erdem, Learning to generate images of outdoor scenes from attributes and semantic layouts, arXiv preprint arXiv:1612.00215, 2016.

[17] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, Improved techniques for training gans, in Advances in Neural Information Processing Systems, 2016, pp. 22262234.

[18] A. Creswell and A. A. Bharath, Task specific adversarial cost function, arXiv preprint arXiv:1609.08661, 2016.

[19] J. Zhao, M. Mathieu, and Y. LeCun, Energy-based generative adversarial network, arXiv preprint arXiv:1609.03126, 2016.

[20] C. Li and M. Wand, Precomputed real-time texture synthesis with markovian generative adversarial networks, in European Conference on Computer Vision, 2016, pp. 702716.

[21] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, Context encoders: Feature learning by inpainting, in CVPR, 2016.

[22] H. Zhang, T. Xu, H. Li, S. Zhang, X. Huang, X. Wang, and D. Metaxas, Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks, arXiv preprint arXiv:1612.03242, 2016.

[23] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, Generative adversarial text to image synthesis, arXiv preprint arXiv:1605.05396, 2016.

[24] N. Jetchev, U. Bergmann, and R. Vollgraf, Texture synthesis with spatial generative adversarial networks, arXiv preprint arXiv:1611.08207, 2016.

[25] A. Dosovitskiy and T. Brox, Generating images with perceptual similarity metrics based on deep networks, arXiv preprint arXiv:1602.02644, 2016.

[26] J. Long, E. Shelhamer, and T. Darrell, Fully convolutional networks for semantic segmentation, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 34313440.

[27] D. Eigen and R. Fergus, Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture, in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 26502658.

[28] L. A. Gatys, A. S. Ecker, and M. Bethge, A neural algorithm of artistic style, arXiv preprint arXiv:1508.06576, 2015.

[29] J. Johnson, A. Alahi, and L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, arXiv preprint arXiv:1603.08155, 2016.

[30] C. Dong, C. C. Loy, and X. Tang, Accelerating the super-resolution convolutional neural network, in European Conference on Computer Vision. Springer, 2016, pp. 391407.

[31] P.W.D. Charles, Project Title, (2013), GitHub repository, https://github.com/charlespwd/project-title.

[32] https://www.reddit.com/r/MachineLearning/comments/2lmo0l