A Spatiotemporal Pre-processing Network for Activity Recognition under Rain

Minah Lee
minah.lee@gatech.edu
Burhan Mudassar
burhan.mudassar@gatech.edu
Taesik Na
taesik.na@gatech.edu
Saibal Mukhopadhyay
saibal.mukhopadhyay@ece.gatech.edu

Georgia Institute of Technology
Atlanta, GA 30332, USA

Abstract

This paper presents a deep neural network (DNN) based fully spatiotemporal rain removal network, MoPE-*Spatiotemporal*, to enhance accuracy of activity recognition in rainy videos. The proposed network utilizes spatiotemporal information of an image sequence to detect rain streaks and recover the non-rainy image. We also present rain alert network that detects the rain fall and informs the reduction of recognition confidence under rain. Experimental results show that heavy rain can highly degrade activity recognition accuracy. MoPE-*Spatiotemporal* removes heavy rain better than state-of-the-art methods, and significantly improves (0.15) activity recognition accuracy in rainy videos with minimal impact on recognition accuracy in clean videos.

1 Introduction

Activity recognition, which localizes and classifies activity on a video, is an important task in many applications including autonomous vehicle, surveillance and sports analysis. Many of these applications involve outdoor activities; where, adversarial weather condition such as rain, snow, or fog can significantly degrade activity detection accuracy. Since Garg and Nayar [6] presented rain streak analysis and model based on photometric properties, significant progress have been made in removing rain from video. Several prior works have developed physical model of rain streaks [6, 7, 8, 33], while learning based algorithms to remove rain have also been proposed [4, 15, 19, 31]. The prior efforts had mostly seek to recover rain-removed sequences from rainy videos. Hence, developed algorithms have been evaluated based on spatial similarity between clean and derained images using peak signal to noise ratio or structural similarity criteria. However, as we will show later, high spatial similarity to clean images does not guarantee high performance on activity recognition. Therefore, rain removal network for activity recognition remains an important problem.

In this paper, we study the effect of rain on deep learning based activity recognition, and propose a new rain removal network, hereafter referred to as, MoPE-*Spatiotemporal*,

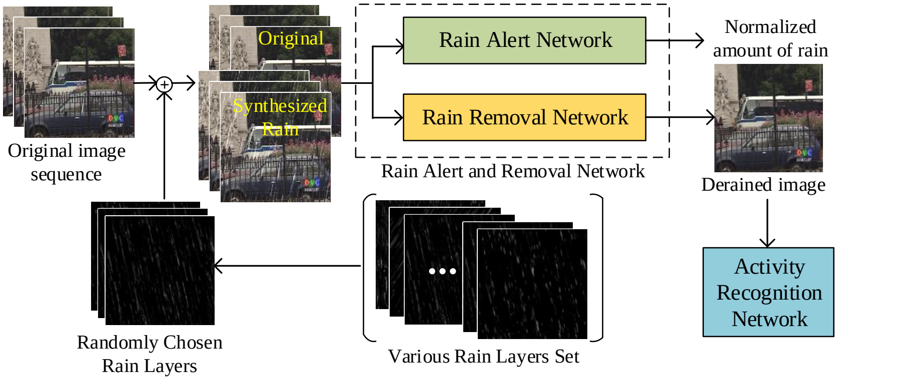


Figure 1: Framework of rain alert and removal network for activity recognition under rain.

for activity recognition under rain (Figure 1). The proposed network acts as a pre-processor and uses spatiotemporal information for detection and removal of rain streaks, and improve activity recognition accuracy. The paper makes following key contributions:

- To the best of our knowledge, this is the first work to enhance performance of activity recognition under rain.
- We propose a DNN based rain removal network that uses spatiotemporal features for both detection and reconstruction of rain images through end-to-end training.
- We propose a rain alert network that detects presence of rain and estimates (normalized) strength of rain to predict confidence reduction on activity recognition.

The MoPE-*Spatiotemporal* is trained with JHMDB dataset [12] and rain layers from *RainSynLight25* and *RainSynComplex25*, which are synthesized with rain streaks models [8, 32]. The end-to-end (MoPE-*Spatiotemporal* + activity recognition) network is evaluated on synthesized video from JHMDB dataset and *RainSynLight25*. The experimental results demonstrate that, the MoPE-*Spatiotemporal* improves activity recognition accuracy by 0.16 in rainy scenes with minimal (0.005) degradation in accuracy for clear scenes.

2 Related Work

Garg and Nayar [6] presented comprehensive analysis of the visual effects of rain and developed models that capture the dynamics and photometry of rain. They also proposed hardware-based scheme to remove rain streaks, such as exposure time or depth of field control [7]. Zhang *et al.* [33] presented rain removal method using k-means clustering based on chromatic and temporal properties of rain streaks. Barnum *et al.* [2] developed a model of rain streak shape and combined with statistical characteristics of rain in frequency domain. Chen *et al.* [5] proposed a low-rank model from matrix to tensor structure to capture the spatio-temporally correlated rain streaks. Jiang *et al.* [13] presented a tensor based approach by considering the overall directional tendency of rain streaks. Li *et al.* [17] presented multiscale convolutional sparse coding based on intrinsic characteristics of rain streaks, which multiscaled rain streaks sparsely scattered in repetitive local patterns. There have also been several work based on learning based approach. Chen *et al.* [4] developed deraining algorithm based on motion segmentation of dynamic scene using Gaussian mixture model. Kim

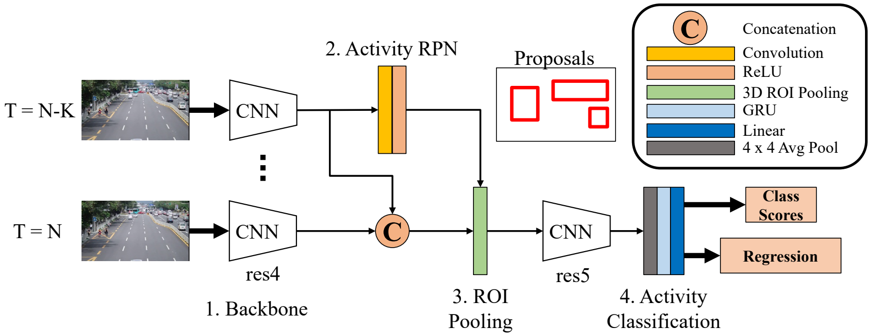


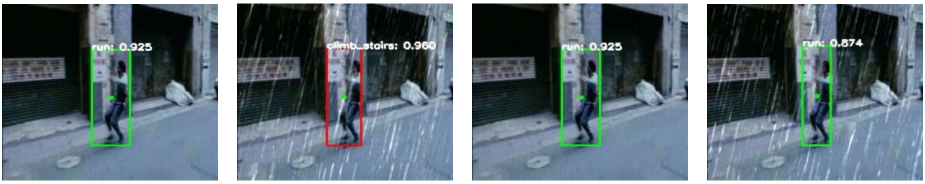
Figure 2: Faster R-CNN based tubelet activity recognition network [14].

et al. [15] proposed to generate rain map from temporal correlation and low rank matrix, and use support vector machine (SVM) to refine the rain map. Wei *et al.* [31] assumed rain streaks has patch-based mixture of Gaussian distribution and it showed good performance on various rain types. Most recently, Liu *et al.* [19] first presented a DNN based rain removal network, which uses spatiotemporal features for restoration of rain image sequence, and further extended [20] to deal with dynamically detected video contexts. In addition, many studies have been done to remove rain on a single image. Zhu *et al.* [34] suggested a joint bi-layer optimization method, and Luo *et al.* [22] utilized a wavelet tight frame and shape prior for fast rain removal. Shen *et al.* [25] presented a novel convolutional neural network based on wavelet and dark channel, and Chen *et al.* [3] proposed an end-to-end gated context aggregation network with dilated convolution. However, all of these studies aim to transform rain image sequence to non rain sequence, so they are evaluated by the spatial comparison between clean images and processed images. However, as we show later, higher spatial similarity does not guarantee better performance of activity recognition under rain.

3 Effect of Rain on Activity Recognition

The deep learning based activity recognition is primarily based on constructing class-specific activity tubes from 2D detection network such as faster R-CNN [16, 24]. Use of two separate convolutional networks for RGB and optical flow images and fusing them to improve recognition accuracy has also been proposed [27, 28]. Recently, integration of temporal dimension to spatial dimension using 3D convolutional networks have been demonstrated [29, 30]. However, all of these work consider only clean videos without considering adversarial weather conditions.

We first study the effect of rain on the accuracy of the baseline activity recognition network. We choose a recent tubelet activity recognition network and use its Faster R-CNN variant for activity recognition [14] as shown in Figure 2. At any time instant K set of images are passed, in a sliding window manner, through a convolutional backbone to generate K feature maps. The feature map corresponding to the last frame in the stack is used to generate activity proposals using an activity RPN. Proposals from the activity RPN are then used to pool features from the feature stack. The pooled features are then transformed using a gated recurrent unit (GRU). The output of the GRU is used to classify the action and perform regression on the proposal bounding box. We use ResNet-101 as the convolutional



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

Figure 3: Sample results of activity recognition on a video. (a), (c) are clean videos and (b), (d) are rain synthesized video from *RainSynLight25*. (a), (b) shows the activity recognition results without rain removal network and (c), (d) shows the result with MoPE-*Spatiotemporal*. Green rectangles indicate the recognized activity is correct and red rectangle indicates it is wrong.

(a) (b) (c)

Figure 4: Sample results of activity recognition under various rain without rain removal network. Green rectangle indicates the recognized activity is correct and red rectangles indicate it is wrong.

feature extractor. Features from res4 are used for generating activity proposals and generating pooled features. Features after ROI pooling are then passed through res5 as per standard practice for Faster RCNN. In our experiments, K is set to 3.

The activity recognition network is pre-trained on JHMDB dataset and is evaluated on synthesized rainy video from JHMDB dataset and *RainSynLight25*. Rain streaks on image sequence change the image content and affect activity recognition result as shown in Figure 3 (a), (b). Different rain configurations such as density, length, angle of rain streaks change the activity recognition results and heavy rain shows large impact (Figure 4). Our study shows that rain streaks removal network is necessary for activity recognition, especially under heavy rain.

4 Rain Alert and Removal Network

As shown in Figure 1, proposed network consists of *Rain Alert Network* and *Rain Removal Network*. The rain removal network detects the rain streaks and estimates the rain-removed image using spatiotemporal information. It outputs a derained image and feed to activity recognition network. The rain alert network estimates strength (normalized) of rain to inform the rain situation and notify that the confidence level of activity recognition is reduced due to rain. One of the challenges for using DNN to process rainy videos is that the spatiotemporal configuration of rain can be highly diverse, and it is impossible to train the network with infinite kinds of rainy videos. To tackle this problem, we generate a rain layers set which includes hundreds of spatially varying rain layers. For each batch, rain layers are randomly selected and added to original image sequence to emulate temporal variations of rain. Thus, the network can be trained with highly spatiotemporally various rain.

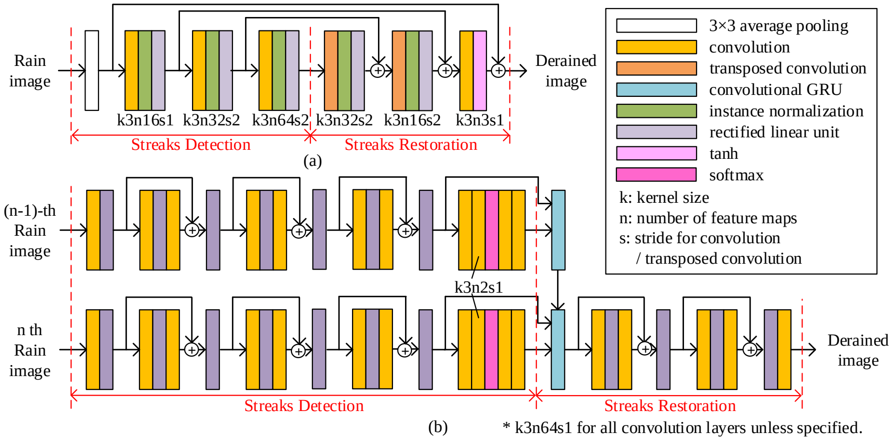


Figure 5: (a) Structure of MoPE-*Spatial* and (b) J4R-Net [19].

Figure 6: Sample results of MoPE-*Spatial* on rain removal and object detection. (a) is rain image and (b) is rain removal result of (a). We use Single Shot Detector [21] as object detection network with mobilenet v1 [10] as backbone feature extractor in Tensorflow API [11]. (c) is object detection result on clean image without rain removal and (d) is the result on rain image with MoPE-*Spatial*.

4.1 Spatial Pre-processing Based Rain Removal

We first aim to reduce the effect of rain by treating rain streaks as a spatial noise. Na *et al*. has proposed mixture of preprocessor experts (MoPE) to remove ideal Gaussian noise on images [23]. We adopt their MoPE structure of [23] and train it with MS-COCO dataset [18] and our rain layers as shown in Figure 5(a) (MoPE-*Spatial*). The MoPE-*Spatial* shows good performance on removing rain (Figure 6(b)) and thus good performance on object detection under rain (Figure 6(d)). However, MoPE-*Spatial* fails on activity recognition under rain as shown in Figure 7. It can detect human under rain, but recognizes as wrong activity. This is because MoPE-*Spatial* only use spatial features for rain removal. It brings limitation on both 1) detection and 2) restoration of rain streaks. 1) MoPE-*Spatial* cannot detect some rain streaks, especially under heavy rain where streaks are in high density and often overlapped to each others, thus result different spatial feature. Also it leads to distortion on non rain image, so requires additional gating network [23]. 2) MoPE-*Spatial* cannot recover rain image to non rain image, because rain drops have size and occlude some area of image. With spatial information only, it cannot see the information behind rain drops. On the other hand, J4R-Net [19], recent DNN based rain removal network, uses spatiotemporal information for restoration of occluded region by using recurrent unit (Figure 5(b)). However, J4R-Net still uses only spatial features for rain streaks detection. Therefore, fully spatiotemporal rain removal network is necessary for activity recognition purpose.

Figure 7: Sample results of MoPE-*Spatial* on activity recognition. (a) is the activity recognition result on clean video and (b) is the result on rain video with MoPE-*Spatial*. Green rectangle indicates the recognized activity is correct and red rectangle indicates it is wrong.

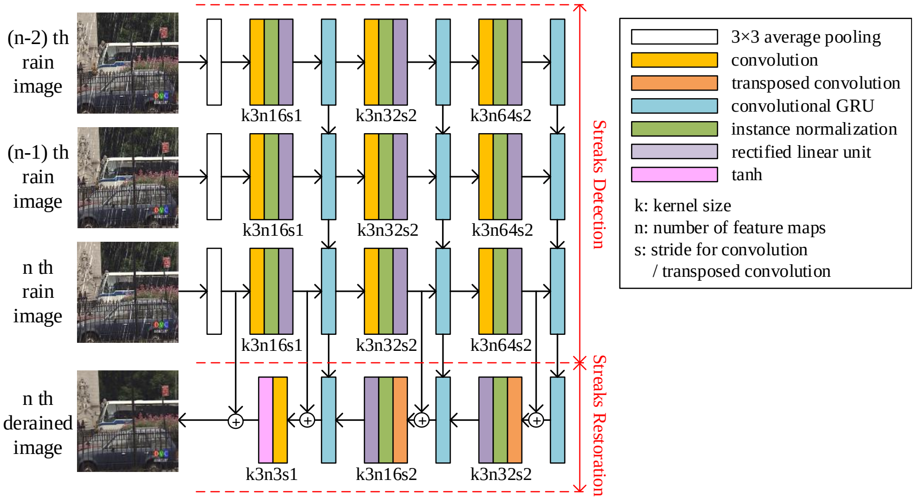


Figure 8: Rain removal network (MoPE-*Spatiotemporal*) structure.

4.2 Proposed Spatiotemporal Pre-processing Based Rain Removal

Figure 8 shows our rain removal network which exploits spatiotemporal information to detect and recover occluded regions (MoPE-*Spatiotemporal*). We use encoding-forecasting structure in [26] with 3 layers, 3 observations and 1 step predictions to generate one derained image from three rain images. Rain streaks are detected during encoding process, and restored during decoding. Reversing the order of encoding network during decoding helps to preserve both global and local representations. Convolution GRUs [1] are used to exploit spatiotemporal information over multiple images. 3×3 convolution or transposed convolution layers with stride 2 and rectified linear unit (ReLU), are used as downsampling and upsampling layers between ConvGRUs to capture spatial representation. Skip connections between convolution layers help preserve pixel wise information of original image. Compared to MoPE-*Spatial* and J4R-Net, our rain removal network uses spatiotemporal information for both detection and restoration through end-to-end training.

As our rain removal network uses spatiotemporal information for rain streaks detection, it does not require gating network to prevent distortion on non-rain images as in [23]. Instead, we adopt the gating network as rain alert network (Figure 9) to quantify the amount of rain and notify the confidence drop during activity recognition. We use 31×31 receptive field which is enough to detect rain streaks.

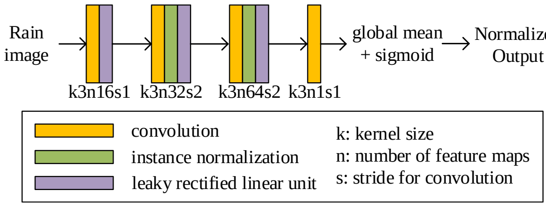


Figure 9: Rain alert network structure.

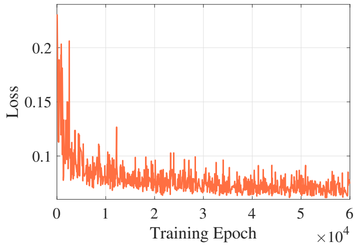


Figure 10: Training loss behavior of MoPE-*Spatiotemporal*.

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

Figure 11: Rain alert network outputs the normalized amount of rain. Yellow numbers at right bottom shows the output of rain removal network for each images.

4.3 Loss Function and Training

For the rain synthesis function $F : X \to Y$, derain function $G : Y \to X$ and its discriminator D, we use pixel difference loss (L1 loss), four level multi scale structural similarity (MS-SSIM) loss, and adversarial loss [9] to train rain removal network. We define the total loss function as follows:

$$\mathcal{L}_{\text{all}}(G,D|F) = \alpha\mathcal{L}_{\text{GAN}}(G,D|F) + \beta\mathcal{L}_{\text{pixel}}(G|F) + \gamma\mathcal{L}_{\text{structural}}(G|F), \tag{1}$$

$$\mathcal{L}_{\text{GAN}}(G,D|F) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{y\sim p_{\text{data}}(y)}[\log(1-D(G(y)))], \tag{2}$$

$$\mathcal{L}_{\text{pixel}}(G|F) = ||G(F(x)) - x||_2^2, \tag{3}$$

$$\mathcal{L}_{\text{structural}}(G|F) = 1 - \mathit{MS\text{-}SSIM}(G(F(x)), x). \tag{4}$$

α, β, and γ are set to be 1, 0.3, and 0.7 respectively. Figure 10 shows that the training loss converges after 30k epoch.

We consider the rain alert as a classification problem and we adopt softmax rain alert network. For the rain alert function H, the original image images x and the rainy images $F(x)$, the loss function is defined as follows:

$$\mathcal{L}_{\text{Rain alert}}(H) = -\log(H(x)) - \log(1 - H(F(x))) \tag{5}$$

where the output of H is the result from the sigmoid function.

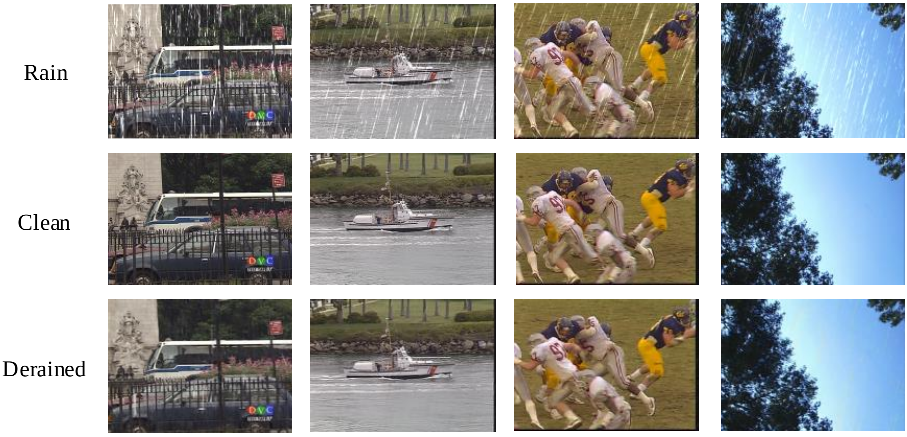


Figure 12: Sample results of rain removal on example of *RainSynLight25*, *RainSynComplex25*. Video is available in supplementary material.

Table 1: PSNR comparison on *RainSynLight25*, *RainSynComplex25*.

	JORDER [32]		J4R-Net [19]		MoPE-*Spatial*		This Work	
Dataset	Light	Complex	Light	Complex	Light	Complex	Light	Complex
PSNR (dB)	30.37	20.20	32.96	27.03	32.29	29.87	32.83	30.20

5 Experimental Results

We demonstrate experimental results of the proposed spatiotemporal pre-processing for rain removal and activity recognition in rainy scenes. MoPE-*Spatiotemporal* and rain alert network are trained with JHMDB dataset [12] and rain layers from *RainSynLight25* and *RainSynComplex25*, which are synthesized with rain streaks models [8, 32]. Rain removal is evaluated on synthesized rain images from *RainSynLight25*, *RainSynComplex25* to compare the results with previous work. Activity recognition network is trained with JHMDB dataset and evaluated on synthesized video from JHMDB dataset and *RainSynLight25*.

5.1 Rain Alert and Removal on Video

Rain alert network discriminates rain on an image and outputs the normalized amount of rain in range of [0, 1], where, '0' means heavy rain and '1' means no rain. Figure 11 shows the output of rain detection network on various strength of rain.

Figure 12 shows sample results of MoPE-*Spatiotemporal* on rain removal. It is compared with state-of-the-art DNN based rain removal methods, JORDER [32], J4R-Net [19], and MoPE-*Spatial* using peak signal-to-noise ratio (PSNR) as comparison criteria (Table 1). JORDER, MoPE-*Spatial* are single frame deraining methods and J4R-Net, MoPE-*Spatio temporal* are video deraining methods. On *RainSynLight25*, MoPE-*Spatiotemporal* and J4R-Net show comparable results which are better than MoPE-*Spatial* and JORDER. This is because both MoPE-*Spatiotemporal* and J4R-Net use temporal feature to restore rain streaks. Also, as most of the sequences in *RainSynLight25* have light rain, temporal feature is not critical on rain detection, which results proposed network and J4R-Net comparable. On the

Figure 13: Sample results of activity recognition on JHMDB synthesized with *RainSynLight25*, *RainSynComplex25* rain layers. Green rectangle indicates the recognized activity is correct and red rectangle indicates it is wrong. Video is available in supplementary material.

Table 2: Activity recognition comparison on clean and rain videos. Activity recognition network is trained only with clean videos.

mAP	Baseline	MoPE-*Spatial* w/o Gating Net.	MoPE-*Spatial* w/ Gating Net.	This Work
Clean	0.598	0.575	0.598	0.593
Rain	0.420	0.540	0.538	0.579

other hand, MoPE-*Spatiotemporal* shows the highest PSNR on *RainSynComplex25*. This is because some of the sequences in *RainSynComplex25* have heavy rain, so rain streaks cannot be detected spatially, but temporally. Only MoPE-*Spatiotemporal* can temporally detect rain streaks, thus shows the highest PSNR.

5.2 Activity Recognition under Rain

Figure 13 shows sample results of activity recognition. Table 2 shows the accuracy of activity recognition on clean and synthesized rainy videos with and without the rain removal network. Activity recognition without the rain removal network shows high recognition accuracy on clean videos, but the accuracy drops more than 0.15 under rain. MoPE-*Spatial* without gating network improves accuracy by 0.12 under rain, but the accuracy on clean video drops due to distortion on clean video. Gating network on MoPE-*Spatial* recovers the accuracy on clean video. On the other hand, the proposed network spatiotemporally detects rain streaks, so it maintains the accuracy on clean videos without gating network. Also, MoPE-*Spatiotemporal* results in less than 0.015 accuracy drop under rain, which is 0.04 higher than using MoPE-*Spatial*. Figure 3 shows sample results of activity recognition. Figure 3(a), (b) do not include rain removal network, so *run* activity is correctly recognized when there is no rain, but recognized as *climb stairs* with high confidence under rain. Figure 3(c), (d) include proposed network and the activity *run* is recognized regardless of rain. Figure 14 shows sample results of rain removal and activity recognition on a real world

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

Figure 14: Sample results on a real world rainy video from Youtube website. (a) is original practical rainy image and (b) is derained image using MoPE-*Spatiotemporal*. (c), (d) show activity recognition results without and with MoPE-*Spatiotemporal* respectively. Green rectangle indicates the recognized activity is correct and red rectangle indicates it is wrong. Video is available in supplementary material.

rainy video from Youtube website[1]. Practical rain streaks are also well removed by MoPE-*Spatiotemporal* (Figure 14(b)). Without proposed network, the activities are recognized as *jump* activities under rain (Figure 14(c)), and our network helps to recognize *run* activities well under rain (Figure 14(d)).

Moreover, Tables 1 and 2 show that MoPE-*Spatial* and MoPE-*Spatiotemporal* show similar PSNR on rain removal, but activity recognition accuracy of MoPE-*Spatiotemporal* under rain is better (0.04) than MoPE-*Spatial*. Hence, we observe that the spatial similarity between clean and derained images does not assure activity recognition performance under rain. To understand whether a rain removal network improves activity recognition accuracy, one needs to evaluate performance of the end-to-end network.

6 Conclusions

In this paper, we propose DNN based fully spatiotemporal rain removal network for activity recognition under rain. The proposed rain removal network, MoPE-*Spatiotemporal*, detects rain streaks and recovers the non-rainy image using spatiotemporal features. A rain alert network is integrated within MoPE-*Spatiotemporal* to estimate strength of the rain (heavy to light) to predict the degradation of confidence during activity recognition. Experimental results show better rain removal performance with MoPE-*Spatiotemporal* compared to state-of-the-art methods and improved activity recognition accuracy under rain. It implies that spatial similarity between clean and derained images does not guarantee activity recognition performance under rain and this work is the first work to improve performance of activity recognition under rain, to the best of our knowledge. For future work, we would like to train/evaluate MoPE-*Spatiotemporal* and activity recognition network with rainy videos in end-to-end manner, which may provide enhanced activity recognition under rain. Moreover, we would like to develop light-weighted MoPE-*Spatiotemporal* for real-time problem.

Acknowledgements The research reported here was supported in part by the Defense Advanced Research Projects Agency (DARPA) under contract number HR0011-17-2-0045. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA.

[1]https://www.youtube.com/

References

[1] N. Ballas, L. Yao, C. Pal, and A. Courville. Delving Deeper into Convolutional Networks for Learning Video Representations. In *International Conference on Learning Representations (ICLR)*, 2016.

[2] P. C. Barnum, S. Narasimhan, and T. Kanade. Analysis of Rain and Snow in Frequency Space. *International Journal of Computer Vision (IJCV)*, 2009.

[3] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and G. Hua. Gated Context Aggregation Network for Image Dehazing and Deraining. *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2019.

[4] J. Chen and L. Chau. A Rain Pixel Recovery Algorithm for Videos With Highly Dynamic Scenes. *IEEE Transactions on Image Processing*, 2014.

[5] Y. Chen and C. Hsu. A Generalized Low-Rank Appearance Model for Spatio-temporally Correlated Rain Streaks. In *IEEE International Conference on Computer Vision (ICCV)*, 2013.

[6] K. Garg and S. K. Nayar. Detection and removal of rain from videos. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004.

[7] K. Garg and S. K. Nayar. When does a camera see rain? In *IEEE International Conference on Computer Vision (ICCV)*, 2005.

[8] K. Garg and S. K. Nayar. Photorealistic Rendering of Rain Streaks. In *ACM SIGGRAPH*, 2006.

[9] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Nets. In *Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS)*, 2014.

[10] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. In *Computing Research Repository (CoRR)*, 2017.

[11] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. Murphy. Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[12] H. Jhuang, J. Gall, S. Zuffi, C. Schmid, and M. J. Black. Towards understanding action recognition. In *IEEE International Conference on Computer Vision (ICCV)*, 2013.

[13] T. Jiang, T. Huang, X. Zhao, L. Deng, and Y. Wang. 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.

[14] V. Kalogeiton, P. Weinzaepfel, V. Ferrari, and C. Schmid. Action tubelet detector for spatio-temporal action localization. *IEEE International Conference on Computer Vision (ICCV)*, 2017.

[15] J. Kim, J. Sim, and C. Kim. Video Deraining and Desnowing Using Temporal Correlation and Low-Rank Matrix Completion. *IEEE Transactions on Image Processing*, 2015.

[16] T. Lan, Y. Wang, and G. Mori. Discriminative figure-centric models for joint action localization and recognition. In *International Conference on Computer Vision*, 2011.

[17] M. Li, Q. Xie, Q. Zhao, W. Wei, S. Gu, J. Tao, and D. Meng. Video Rain Streak Removal by Multiscale Convolutional Sparse Coding. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.

[18] T. Lin, M. Maire, S. J. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft COCO: Common Objects in Context. In *European Conference on Computer Vision (ECCV)*, 2014.

[19] J. Liu, W. Yang, S. Yang, and Z. Guo. Erase or Fill? Deep Joint Recurrent Rain Removal and Reconstruction in Videos. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[20] J. Liu, W. Yang, S. Yang, and Z. Guo. D3R-Net: Dynamic Routing Residue Recurrent Network for Video Rain Removal. *IEEE Transactions on Image Processing*, 2019.

[21] W. Liu, D. Anguelov, D. Erhan, C. Szededy, S. Reed, C. Fu, and A. C. Berg. SSD: Single Shot MultiBox Detector. In *European Conference on Computer Vision (ECCV)*, 2016.

[22] Y. Luo, J. Zhu, J. Ling, and E. Wu. Fast Removal of Rain Streaks From a Single Image via a Shape Prior. *IEEE Access*, 2018.

[23] T. Na, M. Lee, B. Mudassar, P. Saha, J. Ko, and S. Mukhopadhyay. Mixture of Preprocessing Experts Model for Noise Robust Deep Learning on Resource Constrained Platforms. In *2019 International Joint Conference on Neural Networks (IJCNN)*, 2019.

[24] S. Saha, G. Singh, M. Sapienza, P.H. S. Torr, and F. Cuzzolin. Deep Learning for Detecting Multiple Space-Time Action Tubes in Videos. In *British Machine Vision Conference (BMVC)*, 2016.

[25] L. Shen, Z. Yue, Q. Chen, F. Feng, and J. Ma. Deep joint rain and haze removal from a single image. *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018.

[26] X. Shi, Z. Gao, L. Lausen, H. Wang, D. Yeung, W. Wong, and W. Woo. Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model. In *Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS)*, 2017.

[27] K. Simonyan and A. Zisserman. Two-Stream Convolutional Networks for Action Recognition in Videos. In *Advances in Neural Information Processing Systems (NIPS)*, 2014.

[28] G. Singh, S. Saha, M. Sapienza, P. Torr, and F. Cuzzolin. Online Real-Time Multiple Spatiotemporal Action Localisation and Prediction. In *IEEE International Conference on Computer Vision (ICCV)*, 2017.

[29] L. Sun, K. Jia, D. Yeung, and B. E. Shi. Human Action Recognition Using Factorized Spatio-Temporal Convolutional Networks. In *IEEE International Conference on Computer Vision (ICCV)*, 2015.

[30] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Learning Spatiotemporal Features with 3D Convolutional Networks. In *IEEE International Conference on Computer Vision (ICCV)*, 2015.

[31] W. Wei, L. Yi, Q. Xie, Q. Zhao, D. Meng, and Z. Xu. Should We Encode Rain Streaks in Video as Deterministic or Stochastic? In *IEEE International Conference on Computer Vision (ICCV)*, 2017.

[32] 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 (CVPR)*, 2017.

[33] X. Zhang, H. Li, Y. Qi, W. K. Leow, and T. K. Ng. Rain Removal in Video by Combining Temporal and Chromatic Properties. In *IEEE International Conference on Multimedia and Expo*, 2006.

[34] L. Zhu, C. Fu, D. Lischinski, and P. Heng. Joint Bi-layer Optimization for Single-Image Rain Streak Removal. In *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017.