Self-Learned Video Rain Streak Removal: When Cyclic Consistency Meets Temporal Correspondence (Supplementary Material)

Wenhan Yang[1], Robby T. Tan[2,4], Shiqi Wang[1], Jiaying Liu[3]
[1] City University of Hong Kong [2] National University of Singapore
[3] Peking University [4] Yale-NUS College

Abstract

This supplementary material presents the detailed configuration of the network architecture, shows more visual comparisons, and the visualization results of the immediate results. The compared methods include Uncertainty guided Multi-scale Residual Learning (UMRL) [10], Directional Global Sparse Model (UGSM) [2], Progressive Recurrent Network (PReNet) [8], Discriminatively Intrinsic Priors (DIP) [5], FastDeRain [4], Stochastic Encoding (SE) [9], Multi-Scale Convolutional Sparse Coding (MS-CSC) [6], Joint Recurrent Rain Removal and Reconstruction Network (J4RNet) [7], SuperPixel Alignment and Compensation CNN (SpacCNN) [1]. Video results are provided in the supplementary video.

1. Detailed Network Configuration

The specific network architecture is shown in Table 1.

2. Intermediate Results

2.1. Optical Flow

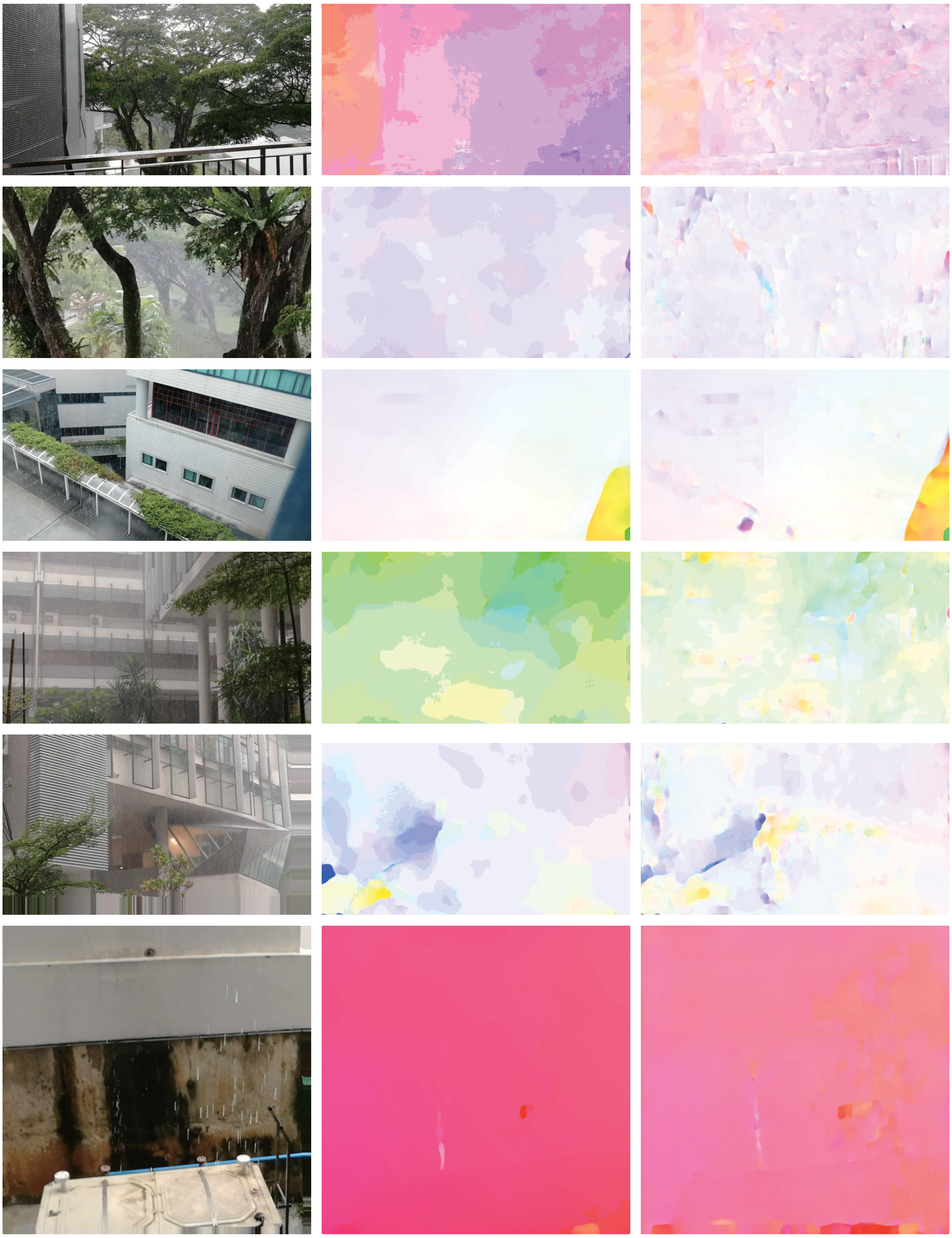
We first visualize the results of the pretrained optical flow extracted from FlowNet [3], and our finetuned optical flow in Fig. 1. It is observed that, compared to the results of FlowNet, our optical flow results tend to have moderate predictions (smaller flow values), more locally adaptive and consistent to the appearance of the video content. As demonstrated in Table 3 of our main submission, this locally adaptive optical flow estimation brings in large performance gains in PSNR and SSIM.

2.2. Non-Rain Masks

We also visualize the estimated non-rain masks of the adjacent and current rain frames. The non-rain masks of the adjacent rain frames M_i^{NA} and the current frame M_t^{NC} control which part information from the adjacent and current frames can be utilized. Therefore, it almost accurately detects the locations of the rain streaks and lowers their values to filter out their effects. Comparatively, the non-rain mask of the current rain frames M_t^{NC} focuses on denoting where the most reliable background regions are. Hence, the prediction is very conservative, namely predicting most regions as rain regions, to prevent from introducing the rain streaks from the current rain frame.

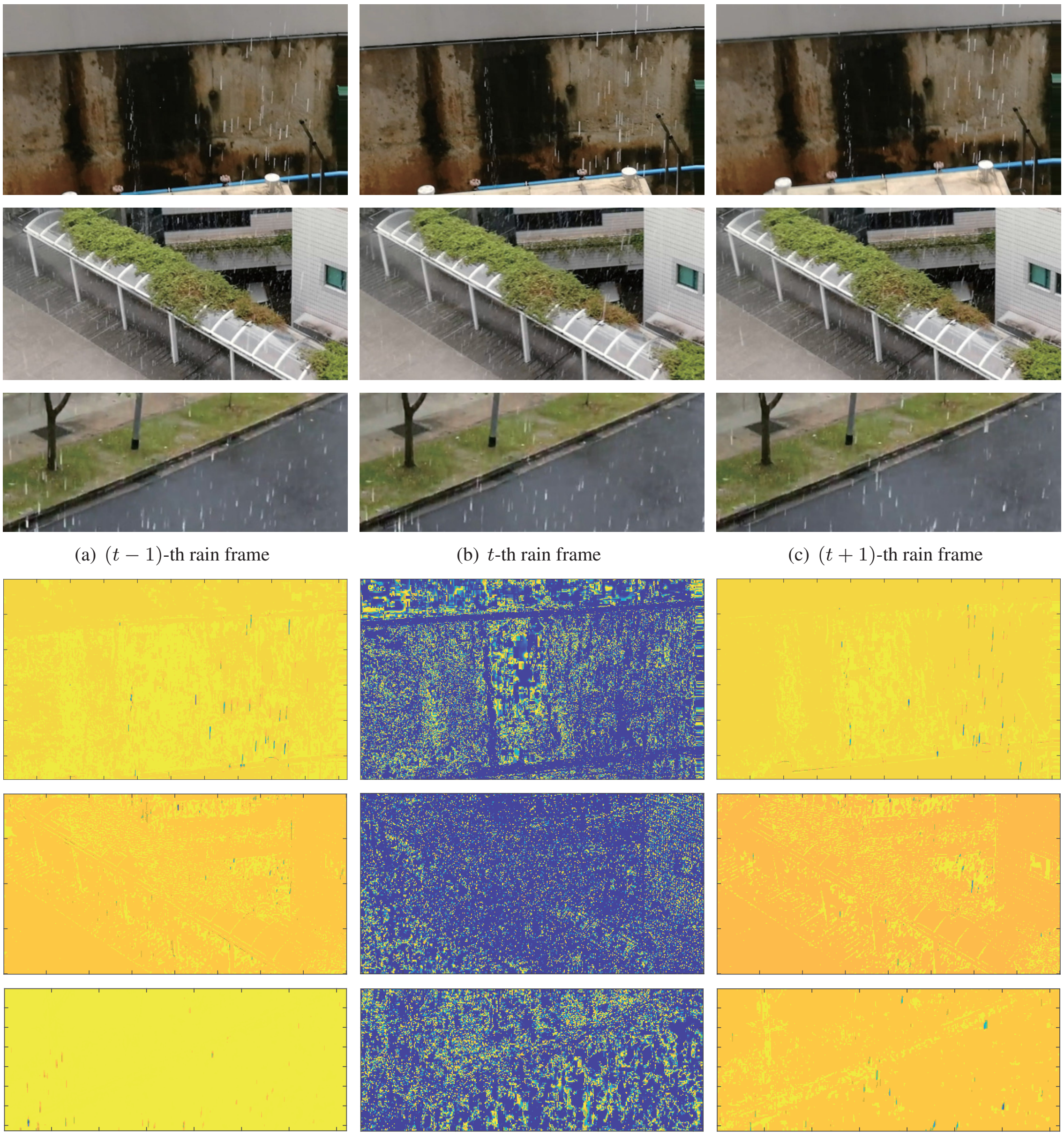
Table 1. Architecture of our self-learning deraining network. Ch denotes the output channel size of each module. The three dimensions of the kernel represent the height, width, and temporal dimensions, respectively.

Module	Layer and Output Name	Type	Kernel	Pad	Ch	Inputs
Flow Estimation	$\left\{C_{i\to t}^{I}\right\}_{i=t-s,t-s+1,\dots,t-1}$ $\left\{C_{i\to t}^{I}\right\}_{i=t+1,t+2,\dots,t+s}$	Flow Estimation Network	–	–	2	$\{I_i\}_{i=t-s,t-s+1,\dots,t+s}$
Warping	$\left\{\widetilde{I}_{i\to t}^{I}\right\}_{i=t-s,t-s+1,\dots,t-1}$ $\left\{\widetilde{I}_{i\to t}^{I}\right\}_{i=t+1,t+2,\dots,t+s}$	Warping Operation	–	–	3	$\{I_i\}_{i=t-s,t-s+1,\dots,t+s}$ $\left\{C_{i\to t}^{I}\right\}_{i=t-s,t-s+1,\dots,t-1}$ $\left\{C_{i\to t}^{I}\right\}_{i=t+1,t+2,\dots,t+s}$
PredNet	P_Conv1	3D Conv.	3×3×3	$[1,1,1]$	64	$\left\{\widetilde{I}_{i\to t}^{I}\right\}_{i=t-s,t-s+1,\dots,t-1}$ $\left\{\widetilde{I}_{i\to t}^{I}\right\}_{i=t+1,t+2,\dots,t+s}$
	P_ReLU1	ReLU	–	–	64	P_Conv1
	P_Conv2	3D Conv.	3×3×3	$[1,1,1]$	64	P_ReLU1
	P_ReLU2	ReLU	–	–	64	P_Conv2
	P_Conv3	3D Conv.	3×3×3	$[1,1,1]$	64	P_ReLU2
	P_ReLU3	ReLU	–	–	64	P_Conv3
	P_ADD3	ADD	–	–	64	P_ReLU3, P_ReLU1
	P_Conv4	3D Conv.	3×3×2	$[1,1,0]$	64	P_ADD3
	P_ReLU4	ReLU	–	–	64	P_Conv4
	P_Conv5	3D Conv.	3×3×3	$[1,1,1]$	64	P_ReLU4
	P_ReLU5	ReLU	–	–	64	P_Conv5
	P_Conv6	3D Conv.	3×3×3	$[1,1,1]$	64	P_ReLU5
	P_ReLU6	ReLU	–	–	64	P_Conv6
	P_ADD6	ADD	–	–	64	P_ReLU6, P_ReLU4
	...	...	...	...	...	...
	P_Conv19	3D Conv.	3×3×2	$[1,1,0]$	64	P_ADD18
	P_ReLU19	ReLU	–	–	64	P_Conv19
	P_Conv20	3D Conv.	3×3×3	$[1,1,1]$	64	P_ReLU19
	P_ReLU20	ReLU	–	–	64	P_Conv20
	P_Conv21	3D Conv.	3×3×3	$[1,1,1]$	64	P_ReLU20
	P_ReLU21	ReLU	–	–	64	P_Conv21
	P_ADD21	ADD	–	–	64	P_ReLU21, P_ReLU19
	$\hat{B}_t^1$	3D Conv.	3×3×3	$[1,1,1]$	3	P_ADD21
EHNet	E_Conv1	3D Conv.	3×3×3	$[1,1,1]$	64	I_t $\left\{\widetilde{I}_{i\to t}^{I}\right\}_{i=t-s,t-s+1,\dots,t-1}$ $\left\{\widetilde{I}_{i\to t}^{I}\right\}_{i=t+1,t+2,\dots,t+s}$
	E_ReLU1	ReLU	–	–	64	E_Conv1
	E_Conv2	3D Conv.	3×3×3	$[1,1,1]$	64	E_ReLU1
	E_ReLU2	ReLU	–	–	64	E_Conv2
	E_Conv3	3D Conv.	3×3×3	$[1,1,1]$	64	E_ReLU2
	E_ReLU3	ReLU	–	–	64	E_Conv3
	E_ADD3	ADD	–	–	64	E_ReLU3, E_ReLU1
	E_Conv4	3D Conv.	3×3×2	$[1,1,0]$	64	E_ADD3
	E_ReLU4	ReLU	–	–	64	E_Conv4
	E_Conv5	3D Conv.	3×3×3	$[1,1,1]$	64	E_ReLU4
	E_ReLU5	ReLU	–	–	64	E_Conv5
	E_Conv6	3D Conv.	3×3×3	$[1,1,1]$	64	E_ReLU5
	E_ReLU6	ReLU	–	–	64	E_Conv6
	E_ADD6	ADD	–	–	64	E_ReLU6, E_ReLU4
	...	...	...	...	...	...
	E_Conv19	3D Conv.	3×3×2	$[1,1,0]$	64	E_ADD18
	E_ReLU19	ReLU	–	–	64	E_Conv19
	E_Conv20	3D Conv.	3×3×3	$[1,1,1]$	64	E_ReLU19
	E_ReLU20	ReLU	–	–	64	E_Conv20
	E_Conv21	3D Conv.	3×3×3	$[1,1,1]$	64	E_ReLU20
	E_ReLU21	ReLU	–	–	64	E_Conv21
	E_ADD21	ADD	–	–	64	E_ReLU21, E_ReLU19
	$\Delta\hat{B}_t^1\left(\hat{B}_t^2=\Delta\hat{B}_t^1+\hat{B}_t^1\right)$	3D Conv.	3×3×3	$[1,1,1]$	3	E_ADD21



(a) Rain Frame (b) FlowNet [3] (c) Our Estimated Flow

Figure 1. The visualization results of FlowNet [3] and our finetuned optical flow results.



(a) $(t-1)$-th rain frame (b) t-th rain frame (c) $(t+1)$-th rain frame

(d) M_{t-1}^{NA} (e) M_{t}^{NC} (f) M_{t+1}^{NA}

Figure 2. The visualization results of the estimated non-rain masks. Yellow denotes the background regions and blue denotes the rain regions.

3. Visual Comparisons

We provide more visual comparisons in Figs. 3 to 7. It is demonstrated that, our results provide more effective results, with less remaining rain streaks, abundant details, and less blurring and artifacts. It is worth mentioning that, our method is self-learned and does not require any rain-streak-free ground truths.

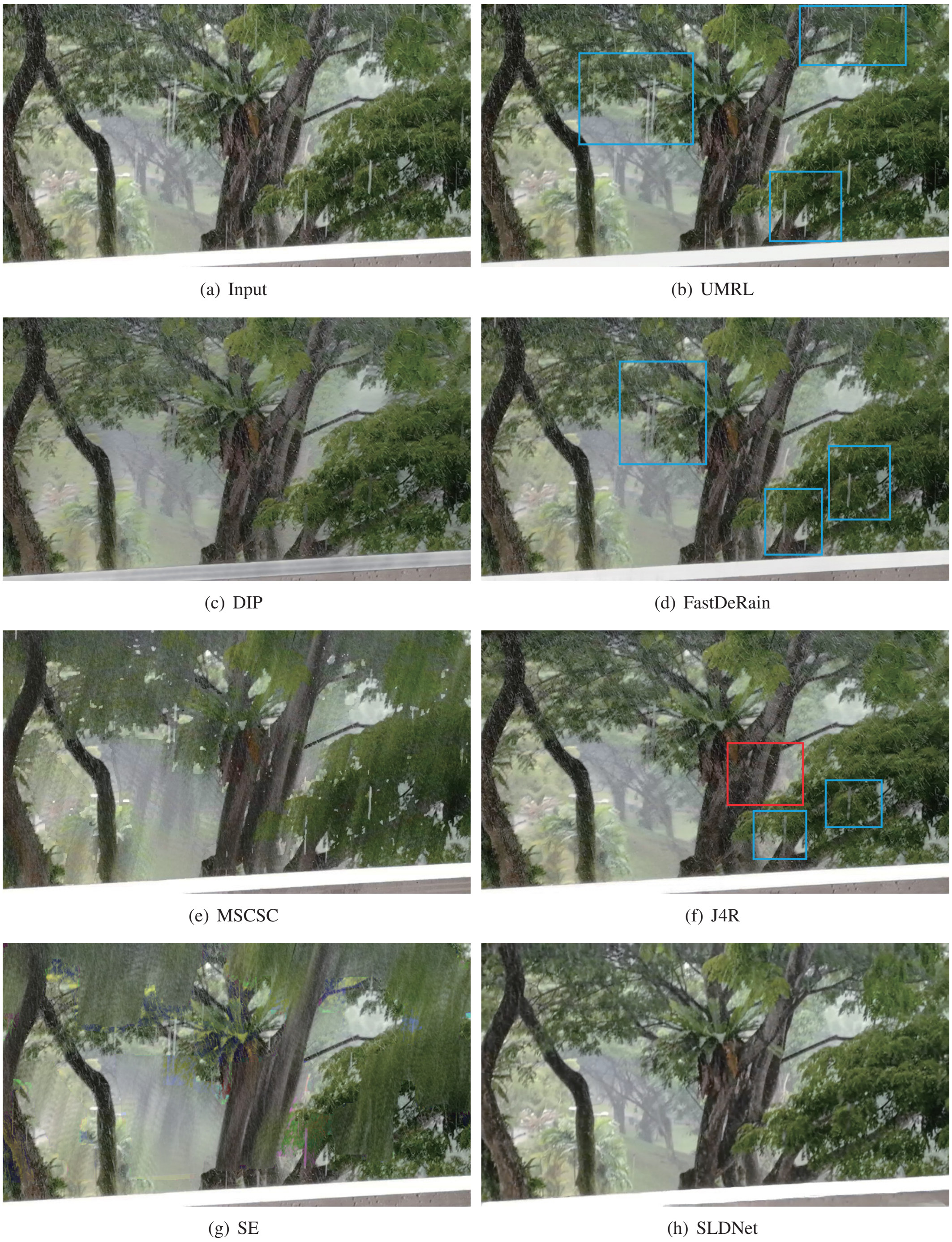
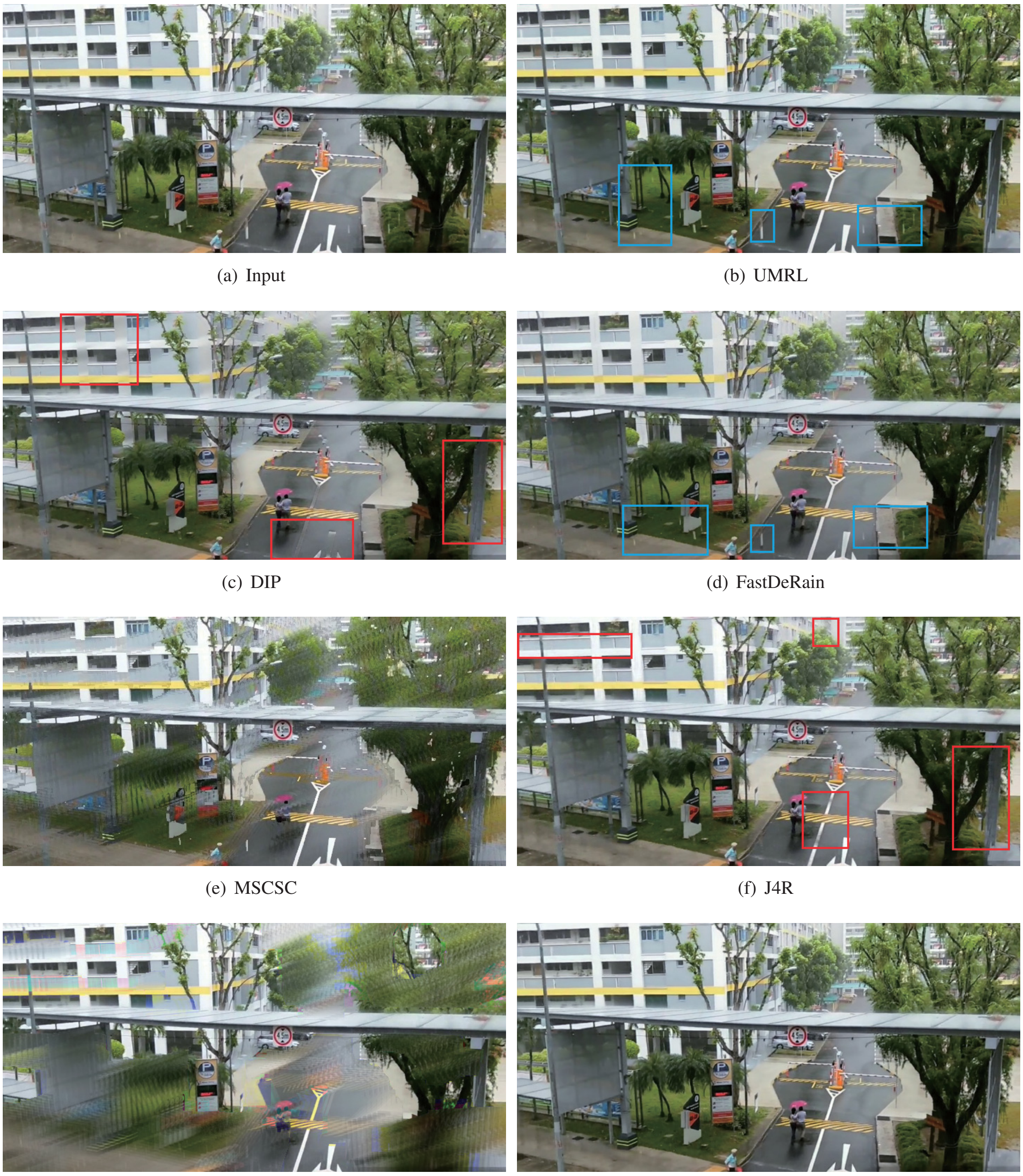
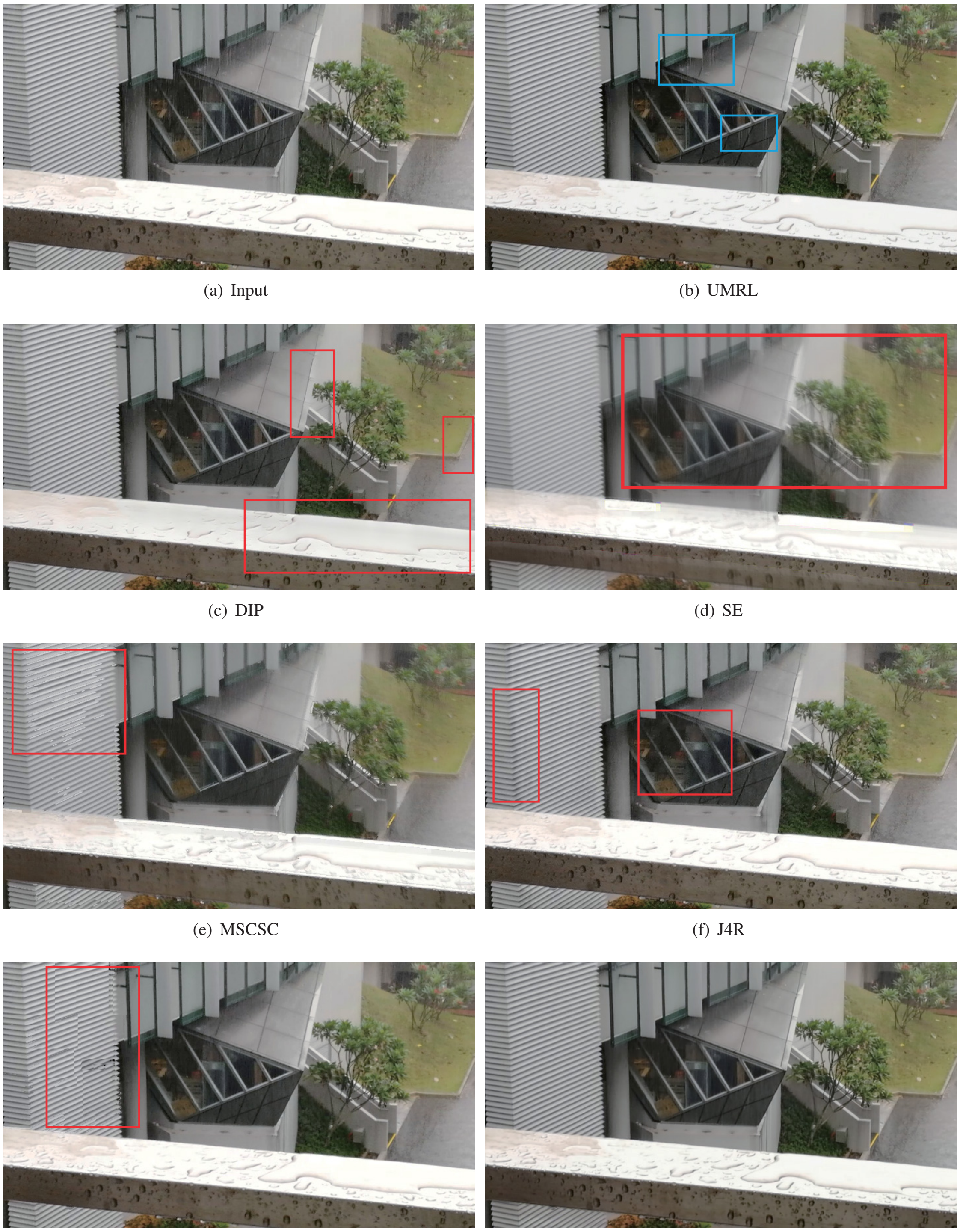


Figure 3. Visual comparison of different deraining methods on a real rain video sequence. The remaining rain streaks and artifacts are denoted with blue and red boxes, respectively. Note that, two white vertical lines in the center of the figure are parts of tree textures instead of rain streaks.



(a) Input (b) UMRL (c) DIP (d) FastDeRain (e) MSCSC (f) J4R (g) SE (h) SLDNet

Figure 4. Visual comparison of different deraining methods on a real rain video sequence. The remaining rain streaks and artifacts are denoted with blue and red boxes, respectively.



(a) Input (b) UMRL

(c) DIP (d) SE

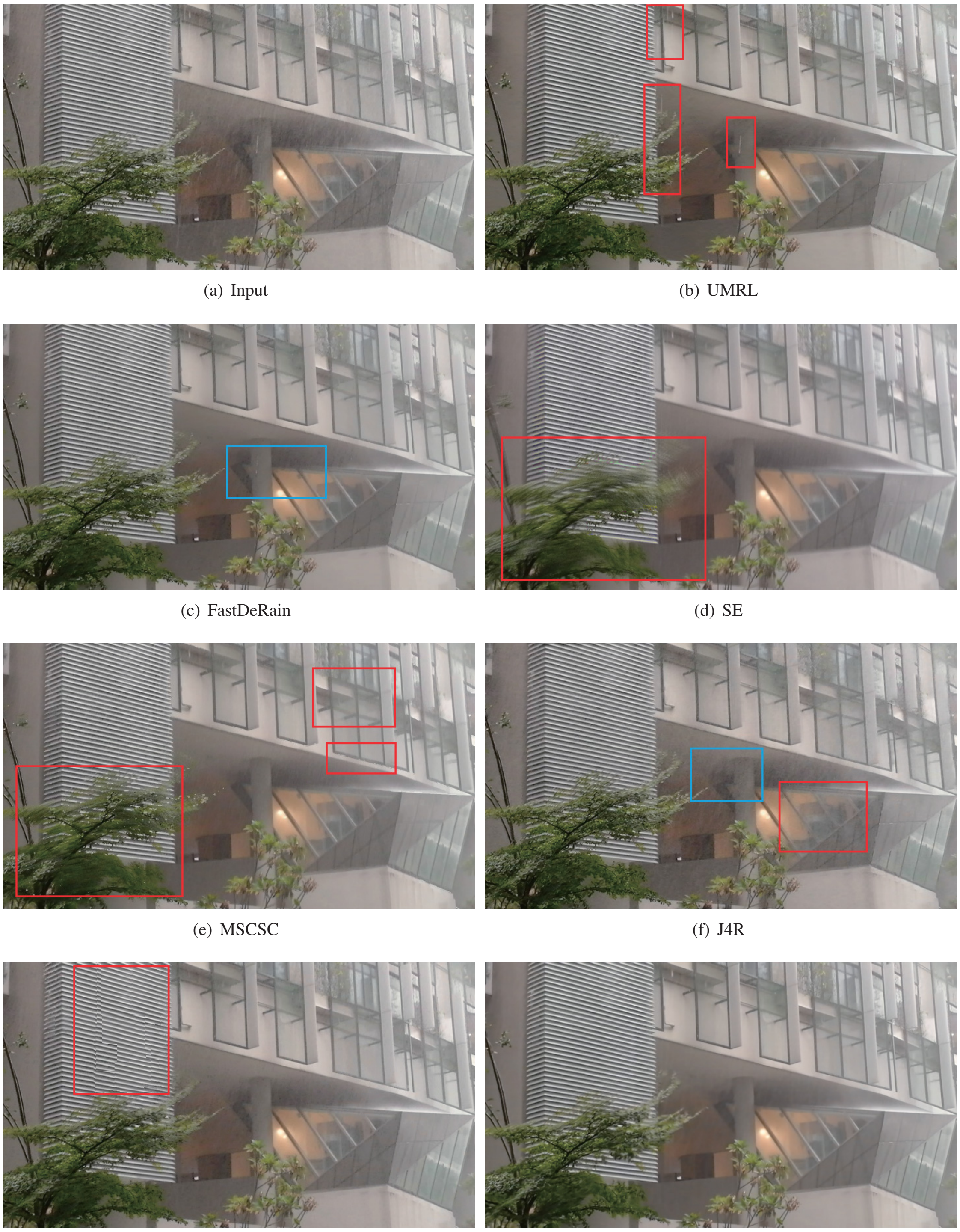
(e) MSCSC (f) J4R

(g) SpacCNN (h) SLDNet

Figure 5. Visual comparison of different deraining methods on a real rain video sequence. The remaining rain streaks and artifacts are denoted with blue and red boxes, respectively.

(a) Input (b) UMRL (c) FastDeRain (d) SE (e) MSCSC (f) J4R (g) SpacCNN (h) SLDNet

Figure 6. Visual comparison of different deraining methods on a real rain video sequence. The remaining rain streaks and artifacts are denoted with blue and red boxes, respectively.



(a) Input
(b) UMRL
(c) FastDeRain
(d) SE
(e) MSCSC
(f) J4R
(g) SpacCNN
(h) SLDNet

Figure 7. Visual comparison of different deraining methods on a real rain video sequence. The remaining rain streaks and artifacts are denoted with blue and red boxes, respectively.

References

[1] Jie Chen, Cheen-Hau Tan, Junhui Hou, Lap-Pui Chau, and He Li. Robust video content alignment and compensation for rain removal in a cnn framework. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, June 2018. 1

[2] Liang-Jian Deng, Ting-Zhu Huang, Xi-Le Zhao, and Tai-Xiang Jiang. A directional global sparse model for single image rain removal. *Applied Mathematical Modelling*, 59:662 – 679, 2018. 1

[3] Philipp Fischer, Alexey Dosovitskiy, Eddy Ilg, Philip Häusser, Caner Hazirbas, Vladimir Golkov, Patrick van der Smagt, Daniel Cremers, and Thomas Brox. Flownet: Learning optical flow with convolutional networks. *arXiv:1504.06852*. 1, 3

[4] T. Jiang, T. Huang, X. Zhao, L. Deng, and Y. Wang. Fastderain: A novel video rain streak removal method using directional gradient priors. *IEEE Trans. on Image Processing*, 28(4):2089–2102, April 2019. 1

[5] Tai-Xiang Jiang, Ting-Zhu Huang, Xi-Le Zhao, Liang-Jian Deng, and Yao Wang. A novel tensor-based video rain streaks removal approach via utilizing discriminatively intrinsic priors. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, July 2017. 1

[6] Minghan Li, Qi Xie, Qian Zhao, Wei Wei, Shuhang Gu, Jing Tao, and Deyu Meng. Video rain streak removal by multiscale convolutional sparse coding. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, June 2018. 1

[7] Jiaying Liu, Wenhan Yang, Shuai Yang, and Zongming Guo. Erase or fill? deep joint recurrent rain removal and reconstruction in videos. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, June 2018. 1

[8] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, June 2019. 1

[9] Wei Wei, Lixuan Yi, Qi Xie, Qian Zhao, Deyu Meng, and Zongben Xu. Should we encode rain streaks in video as deterministic or stochastic? In *Proc. IEEE Int'l Conf. Computer Vision*, Oct 2017. 1

[10] Rajeev Yasarla and Vishal M. Patel. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, June 2019. 1