Learning to Remove Rain in Video with Self-Supervision (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 estimated rain region masks and extracted features by our SLDNet+. The compared methods include Uncertainty guided Multi-scale Residual Learning (UMRL) [9], Directional Global Sparse Model (UGSM) [2], Progressive Recurrent Network (PReNet) [7], Discriminatively Intrinsic Priors (DIP) [4], FastDeRain [3], Stochastic Encoding (SE) [8], Multi-Scale Convolutional Sparse Coding (MS-CSC) [5], Joint Recurrent Rain Removal and Reconstruction Network (J4RNet) [6], SuperPixel Alignment and Compensation CNN (SpacCNN) [1]. Video results are provided in the supplementary video.

Contents

1. Detailed Network Configuration

The specific network architecture is shown in Table 1.

Table 1. The architecture of our self-learning deraining network (SLDNet+). 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
FlowNet	$\{C_{i\to t}\}$ $\{C_{t\to i}\}$	Flow Estimation Network	–	–	2	$\{I_i\}$
Warp	$\left\{\widetilde{I}_{i\to t}\right\}$	Warping Module	–	–	3	$\{I_i\}$ $\{C_{i\to t}\}$
RMNet	Conv1	3D Conv.	3×3×3	$[1,1,1]$	64	$\left\{\widetilde{I}_{i\to t}\right\}$
	ReLU1	ReLU	–	–	64	Conv1
	Conv2	3D Conv.	3×3×3	$[1,1,1]$	64	ReLU1
	ReLU2	ReLU	–	–	64	Conv2
	Conv3	3D Conv.	3×3×3	$[1,1,1]$	64	ReLU2
	ReLU3	ReLU	–	–	64	Conv3
	ADD3	ADD	–	–	64	ReLU3, ReLU1
	Conv4	3D Conv.	3×3×2	$[1,1,0]$	64	ADD3
	ReLU4	ReLU	–	–	64	Conv4
	Conv5	3D Conv.	3×3×3	$[1,1,1]$	64	ReLU4
	ReLU5	ReLU	–	–	64	Conv5
	Conv6	3D Conv.	3×3×3	$[1,1,1]$	64	ReLU5
	ReLU6	ReLU	–	–	64	Conv6
	ADD6	ADD	–	–	64	ReLU6, ReLU4
	...	...	...	...	...	...
	Conv19	3D Conv.	3×3×2	$[1,1,0]$	64	ADD18
	ReLU19	ReLU	–	–	64	Conv19
	Conv20	3D Conv.	3×3×3	$[1,1,1]$	64	ReLU19
	ReLU20	ReLU	–	–	64	Conv20
	Conv21	3D Conv.	3×3×3	$[1,1,1]$	64	ReLU20
	ReLU21	ReLU	–	–	64	Conv21
	ADD21	ADD	–	–	64	ReLU21, ReLU19
	$\hat{B}_t$	3D Conv.	3×3×3	$[1,1,1]$	3	ADD21
Warp	$\left\{\widetilde{B}_{t\to i}\right\}$	Warping Module	–	–	3	$\hat{B}_t$, $\{C_{i\to t}\}$
Rain Region Estimation	$\{M^{\text{NR}}_{t\to i}\}$, $\{M^{\text{NR}}_{i\to t}\}$, M^{NR}_t	Eq. (14), (15), and (17)	–	–	3	$\left\{\widetilde{B}_{t\to i}\right\}$, $\{I_i\}$, $\left\{\widetilde{I}_{i\to t}\right\}$, I_t, $\hat{B}_t$
Loss Function	$\mathcal{L}_{\text{Flow}}$	Eq. (10)	–	–	1	$\{M^{\text{NR}}_{i\to t}\}$, $\left\{\widetilde{I}_{i\to t}\right\}$, I_t
	$\mathcal{L}_{\text{Fid-Con}}$	Eq. (12)	–	–	1	$\{M^{\text{NR}}_{t\to i}\}$, $\left\{\widetilde{B}_{t\to i}\right\}$, $\{I_i\}$
	$\mathcal{L}_{\text{Fid-Back}}$	Eq. (16)	–	–	1	M^{NR}_t, $\hat{B}_t$, I_t

2. Visual Results of Rain Region Estimation

We also compare the visual results of rain region estimation on the *b1* sequence in *NTURain* dataset in Figs. 1-3. It is observed that, the rain/non-rain regions predicted by our method successfully locate the rain streak regions and play a role to weaken the effects of these regions.

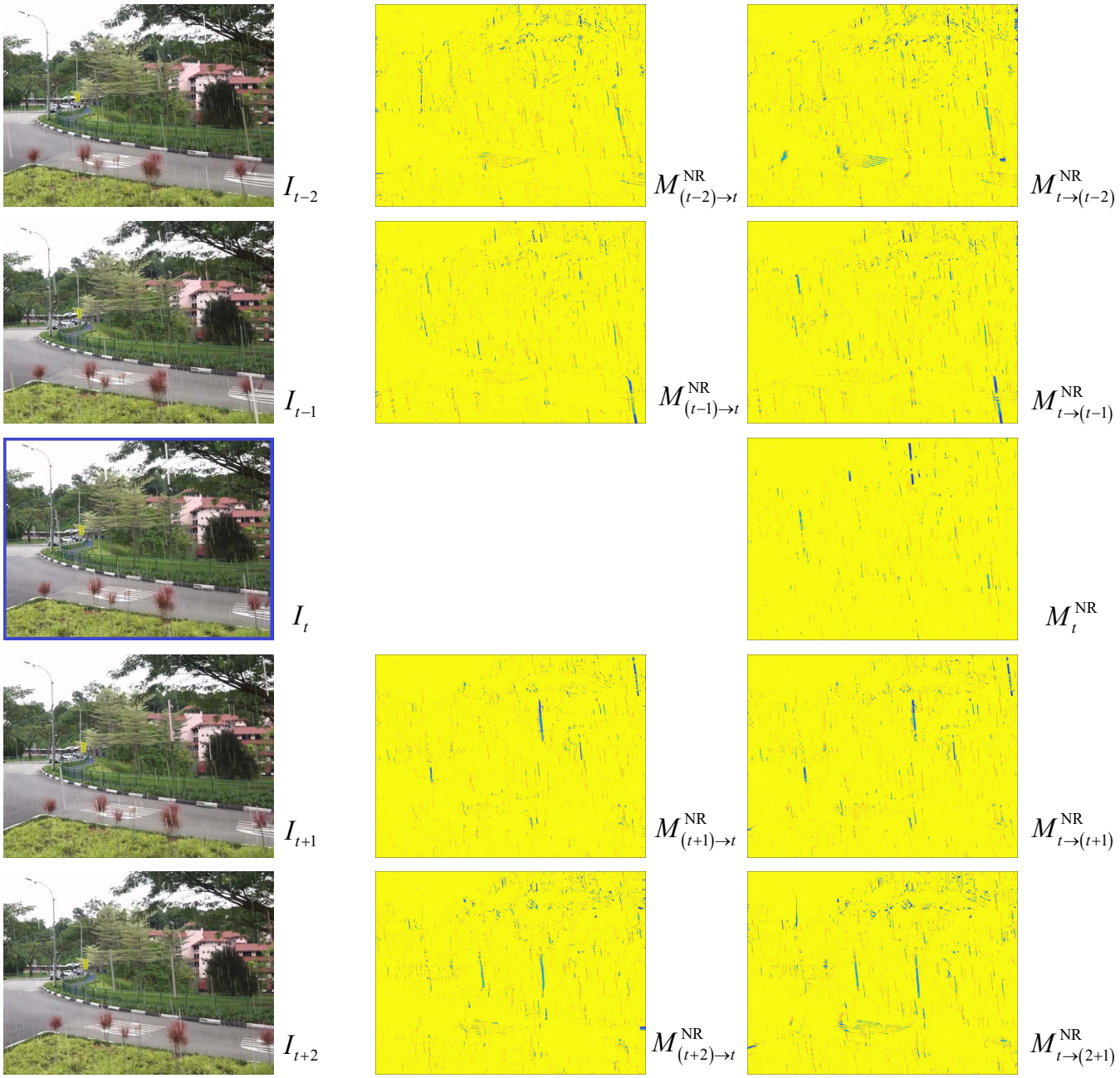


Figure 1. Visual results of rain region estimation on *b1* sequence in *NTURain* dataset. Top panel: rain input frame. Middle panel: rain masks used to train the optical flow network in Eq. (10). Bottom panel: rain masks used to train our deraining network in Eq. (12) and (16). Yellow color denotes the pixel value is close to 1 while blue color denotes the pixel value is close to 0.

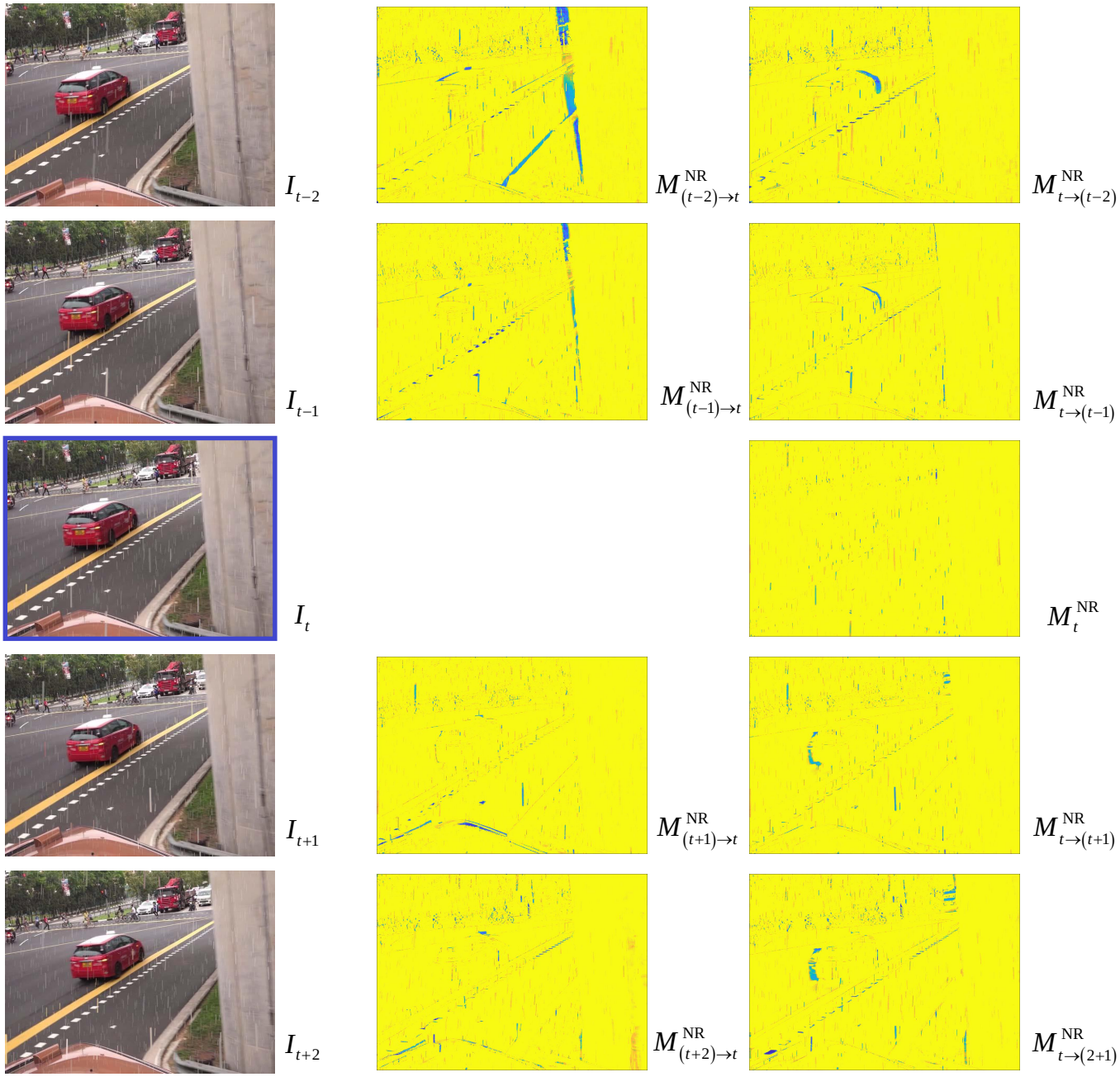


Figure 2. Visual results of rain region estimation on *b2* sequence in *NTURain* dataset. Top panel: rain input frame. Middle panel: rain masks used to train the optical flow network in Eq. (10). Bottom panel: rain masks used to train our deraining network in Eq. (12) and (16). Yellow color denotes the pixel value is close to 1 while blue color denotes the pixel value is close to 0.

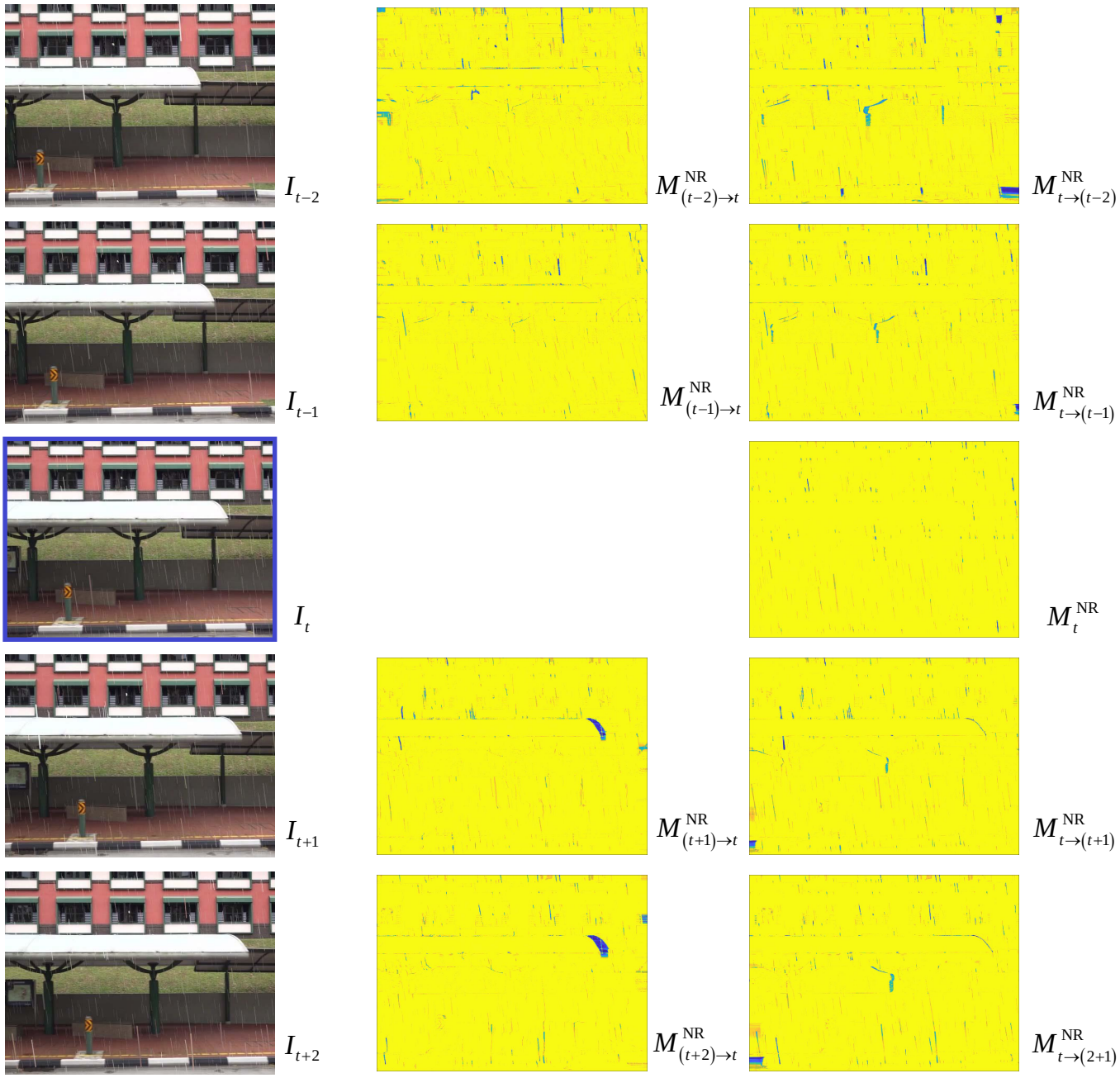


Figure 3. Visual results of rain region estimation on *b3* sequence in *NTURain* dataset. Top panel: rain input frame. Middle panel: rain masks used to train the optical flow network in Eq. (10). Bottom panel: rain masks used to train our deraining network in Eq. (12) and (16). Yellow color denotes the pixel value is close to 1 while blue color denotes the pixel value is close to 0.

3. Visualization of Extracted Features

We also visualize the extracted features from our SLDNet+ and a fully supervised network. The fully supervised model owns the same structure as SLDNet+. Differently, it is trained on the training set of NTURain. The visualized features are generated based on *b2* and *b3* sequences from the NTURain dataset. The feature maps are ranked by their variations and the feature maps with larger variations rank first. From Figs. 4-11, we can observe that, SLDNet+ extracts more sparse but diverse features compared with the fully supervised method. Among the features extracted by the fully supervised method,

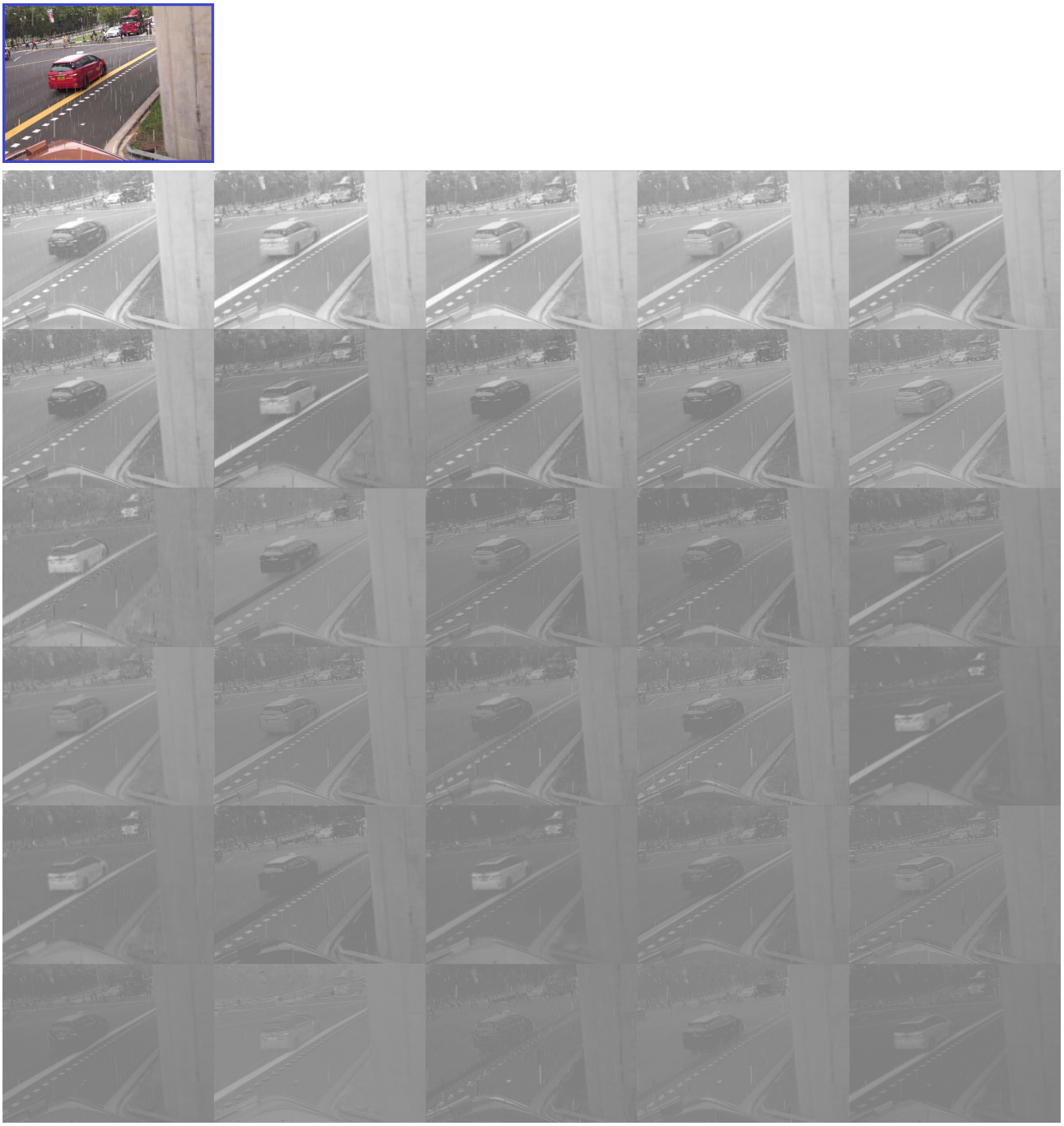


Figure 4. Visualization of the 1*st*-32*th* feature maps in the first layer by our SLDNet+ on *b2* sequence in NTURain dataset. The feature maps are ranked by their variations.

the top ones look similar and contain much background information. Comparatively, the extracted features by SLDNet+ capture look discriminative.

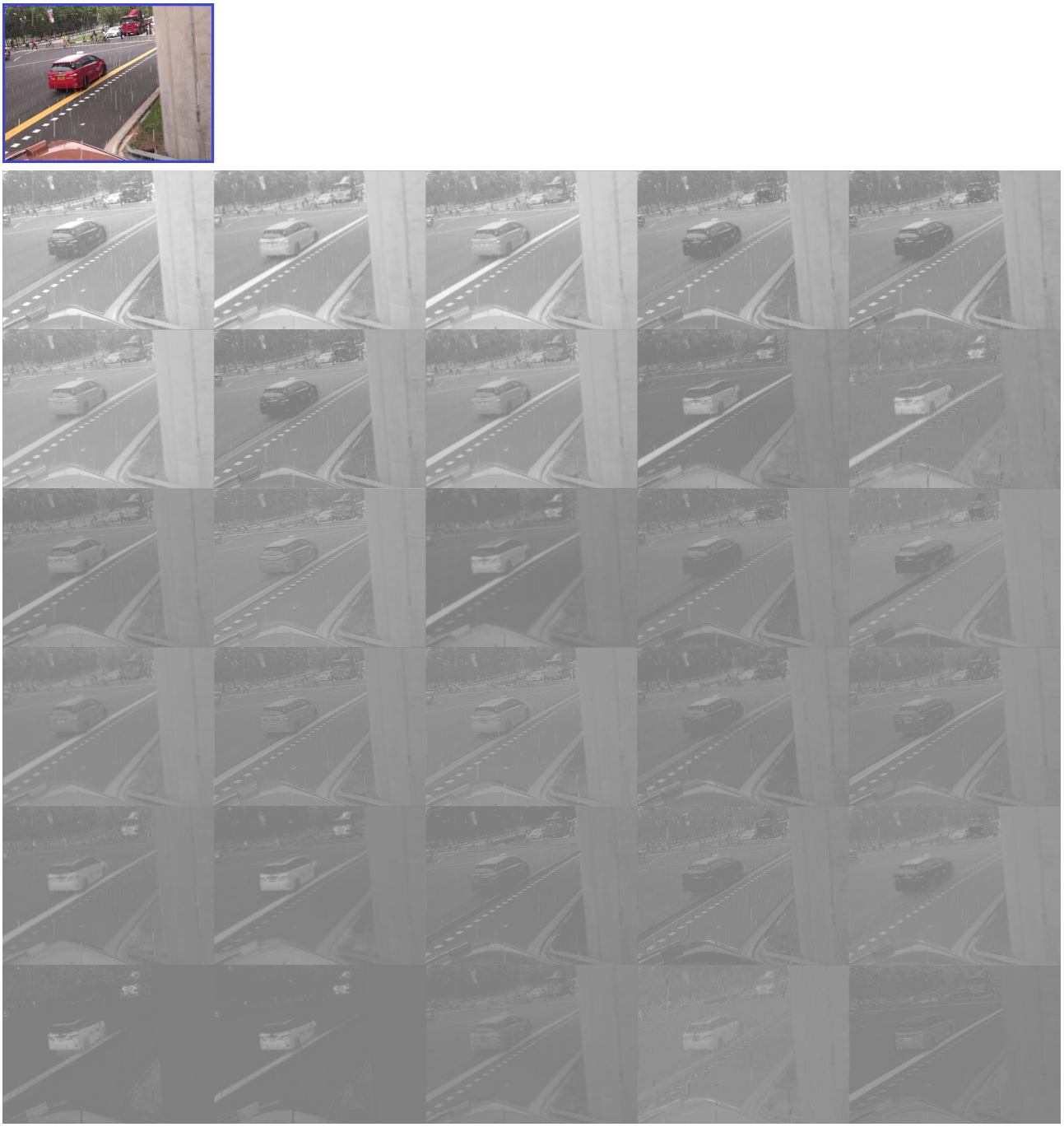


Figure 5. Visualization of the 1*st*-32*th* feature maps in the first layer by a fully-supervised network on *b2* sequence in NTURain dataset. The feature maps are ranked by their variations.

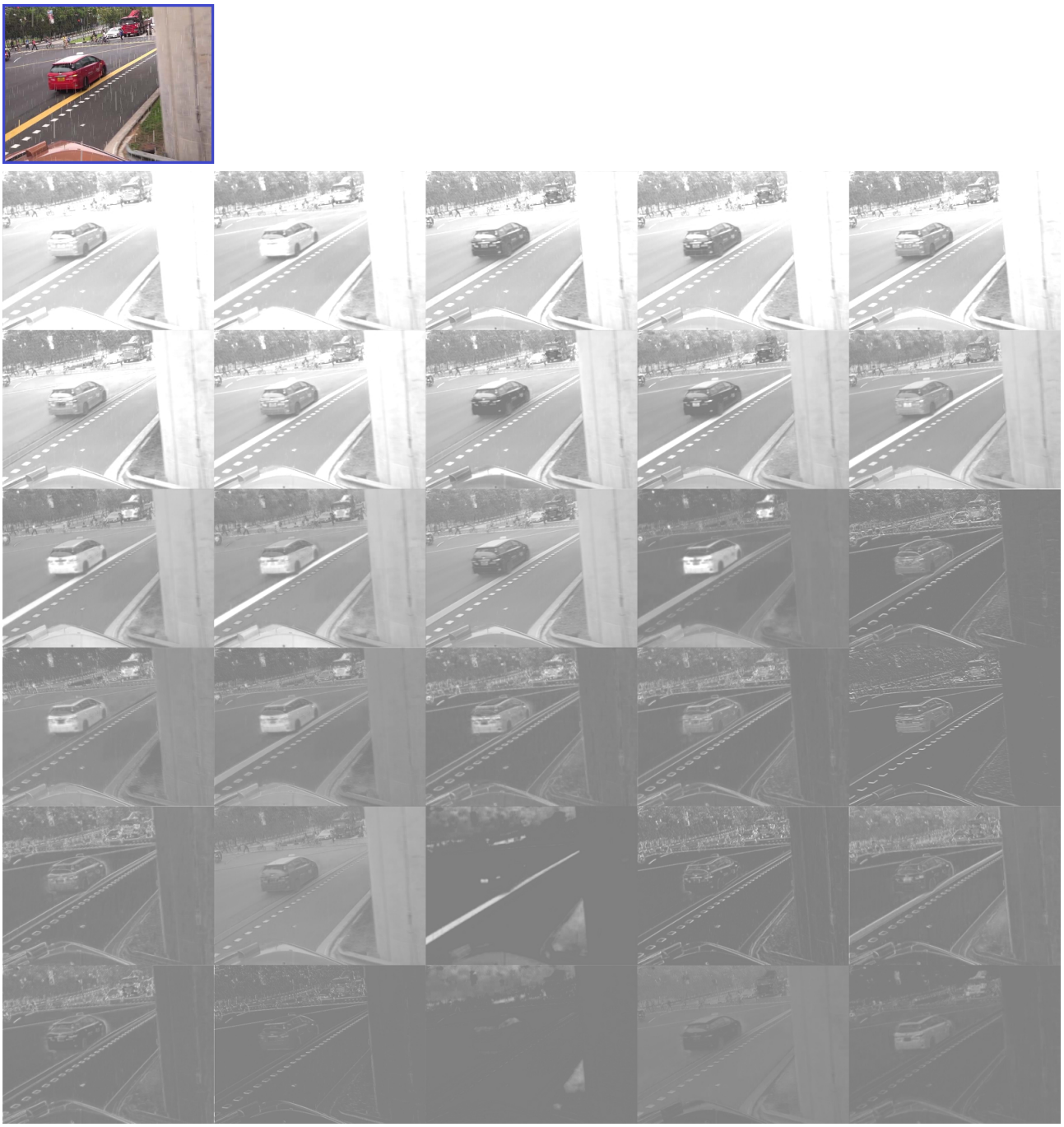


Figure 6. Visualization of the 1*st*-32*th* feature maps in the last layer by our SLDNet+ on *b2* sequence in NTURain dataset. The feature maps are ranked by their variations.

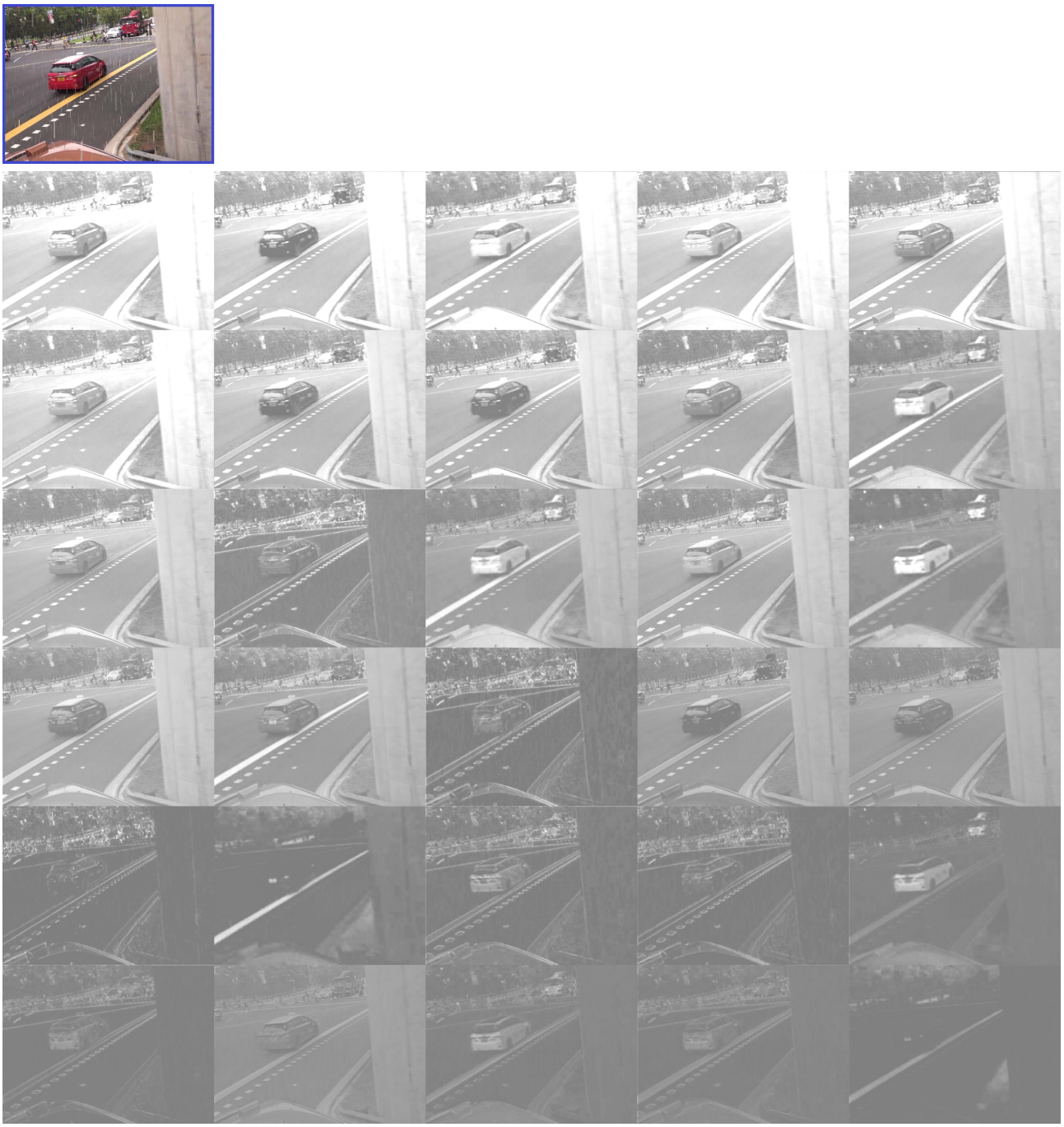


Figure 7. Visualization of the 1*st*-32*th* feature maps in the last layer by a fully-supervised network on *b2* sequence in NTURain dataset. The feature maps are ranked by their variations.

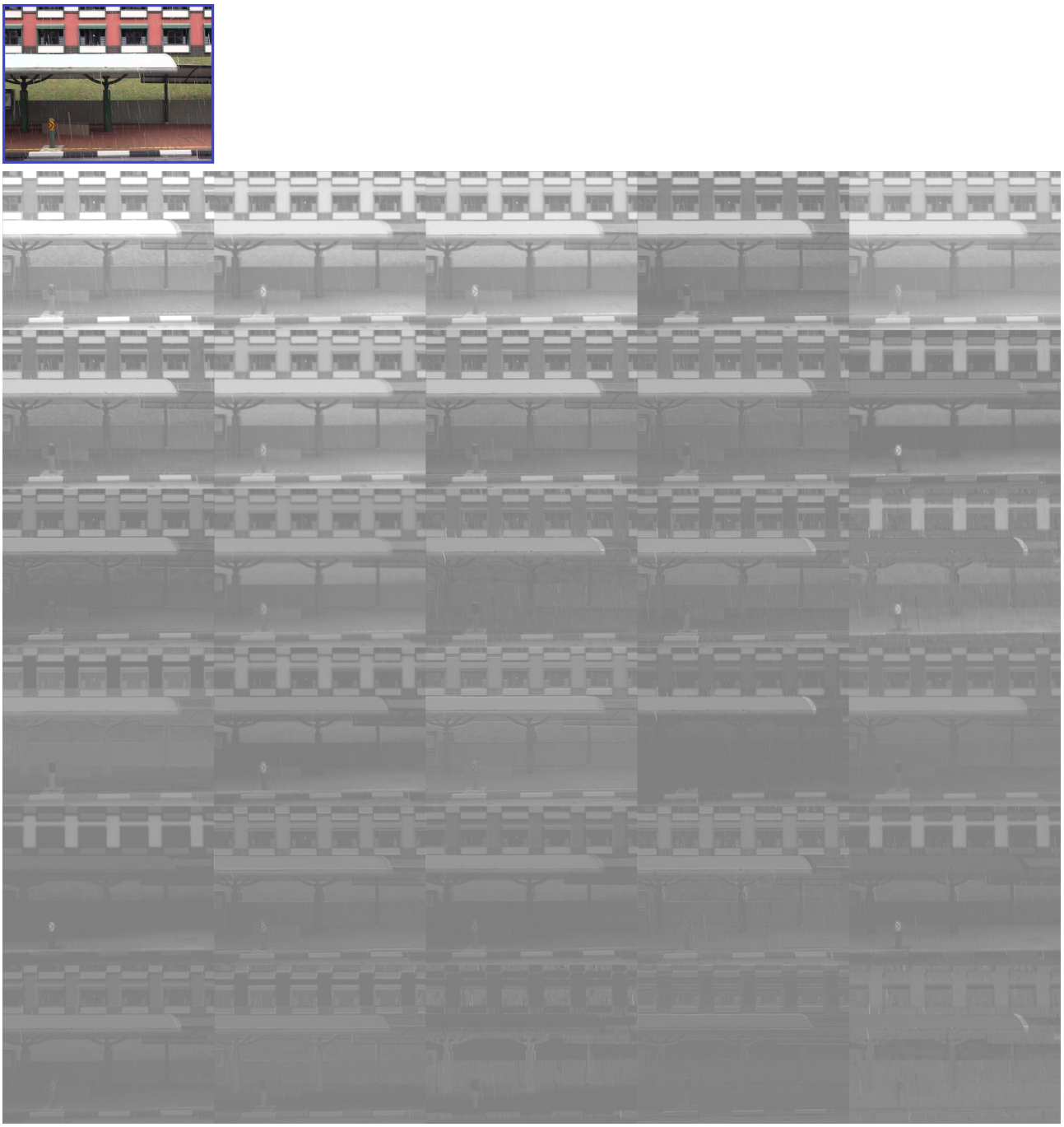


Figure 8. Visualization of the 1*st*-32*th* feature maps in the first layer by our SLDNet+ on *b2* sequence in NTURain dataset. The feature maps are ranked by their variations.

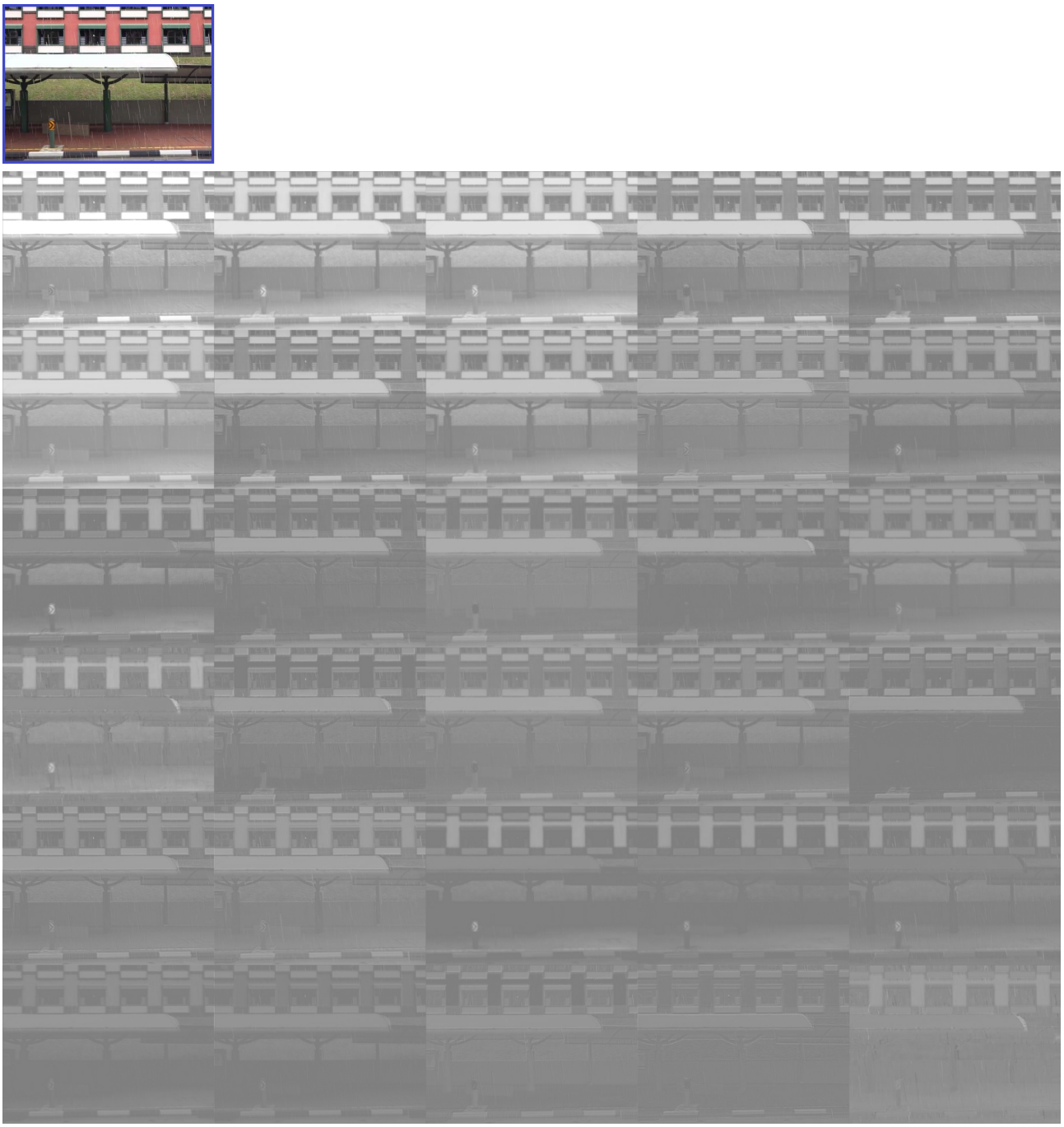


Figure 9. Visualization of the 1*st*-32*th* feature maps in the first layer by a fully-supervised network on *b2* sequence in NTURain dataset. The feature maps are ranked by their variations.

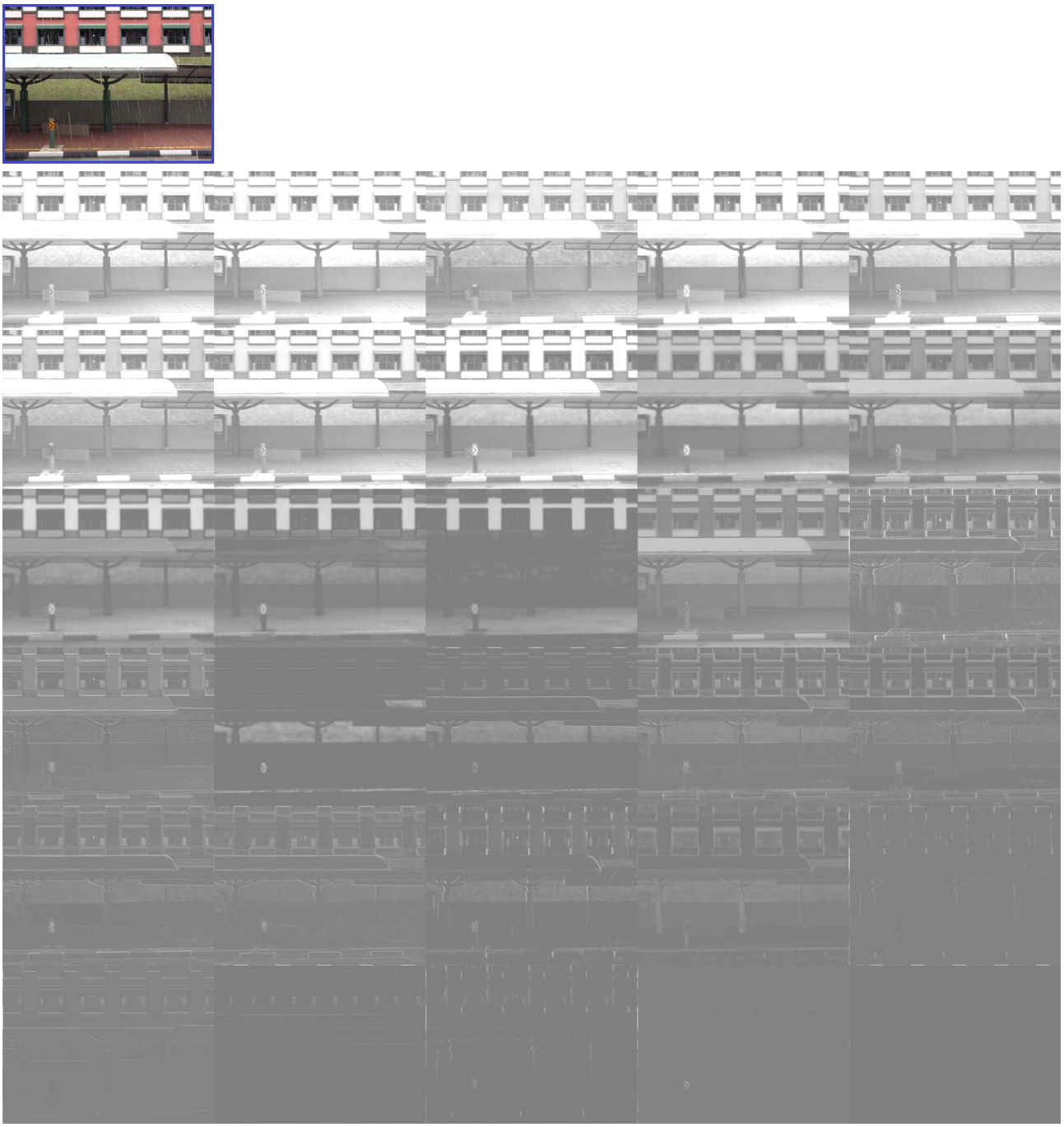


Figure 10. Visualization of the 1*st*-32*th* feature maps in the last layer by our SLDNet+ on *b2* sequence in NTURain dataset. The feature maps are ranked by their variations.

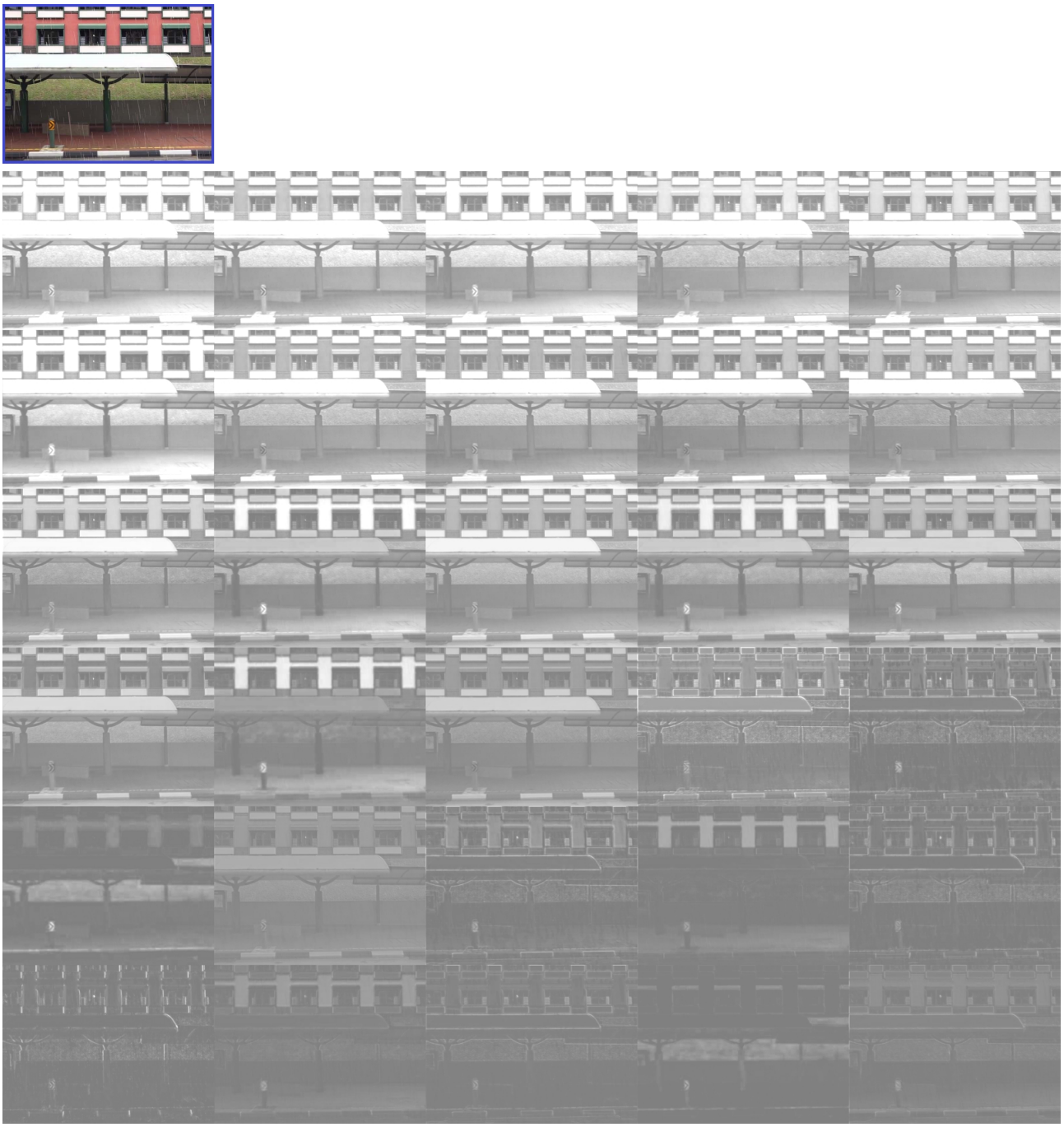


Figure 11. Visualization of the 1*st*-32*th* feature maps in the last layer by a fully-supervised network on *b2* sequence in NTURain dataset. The feature maps are ranked by their variations.

4. Visual Comparisons

We provide more visual comparisons in Figs. 12-16. 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.

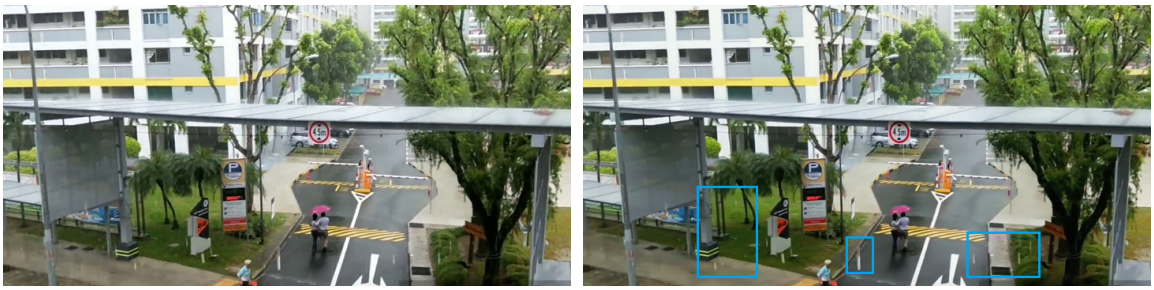
(a) Input (b) UMRL

(c) DIP (d) FastDeRain

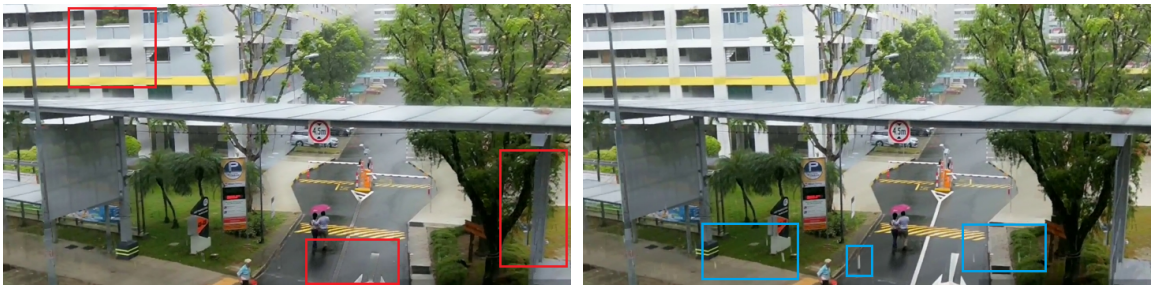
(e) MSCSC (f) J4R

(g) SE (h) SLDNet

Figure 12. 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 13. 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.

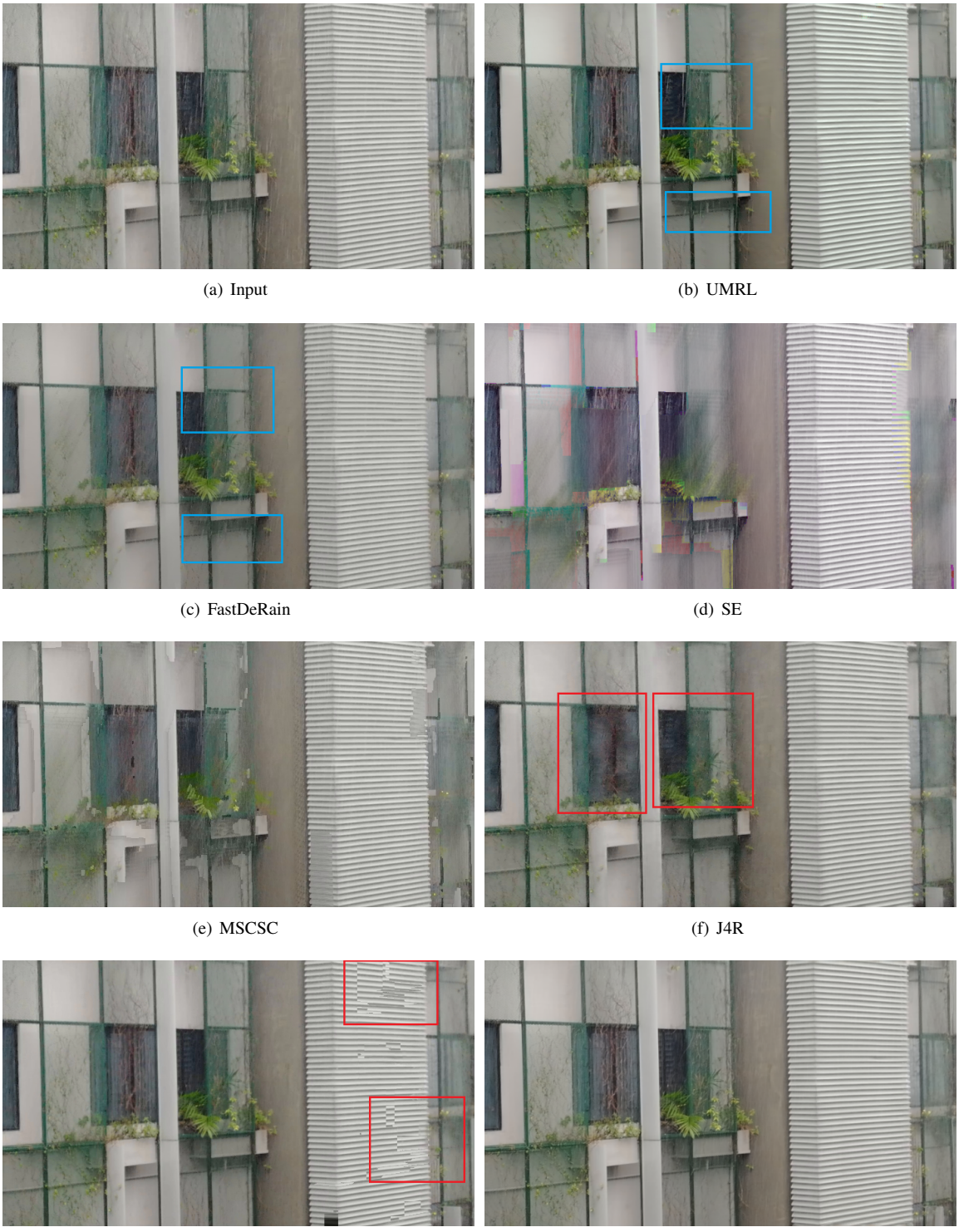
(a) Input (b) UMRL

(c) DIP (d) SE

(e) MSCSC (f) J4R

(g) SpacCNN (h) SLDNet

Figure 14. 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

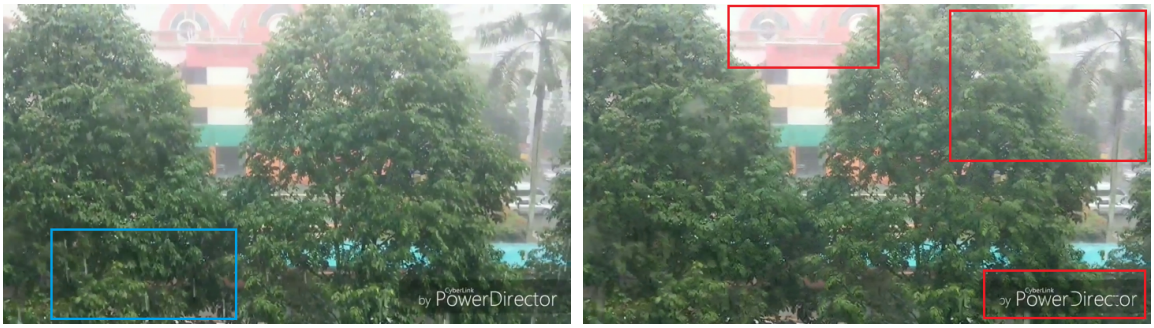
(c) FastDeRain (d) SE

(e) MSCSC (f) J4R

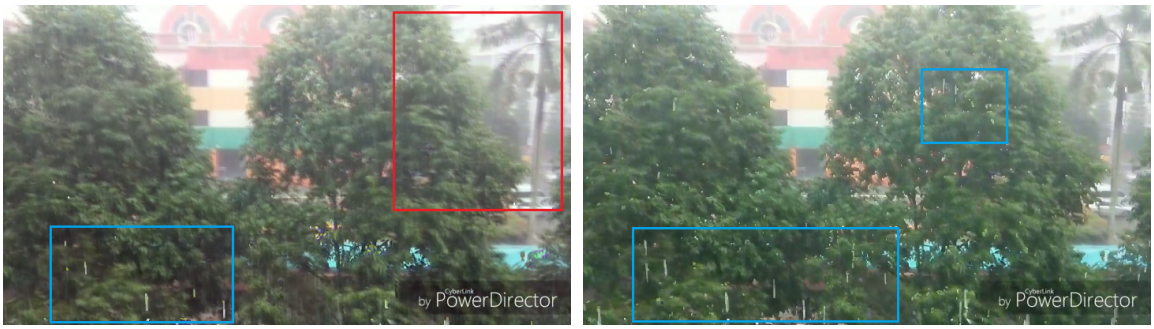
(g) SpacCNN (h) SLDNet

Figure 15. 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) JCAS



(c) FastDeRain (d) PReNet



(e) SE (f) MSCSC

(g) J4R (h) SLDNet

Figure 16. 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] 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

[4] 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

[5] 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

[6] 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

[7] 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

[8] 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

[9] 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