Semi-Supervised State-Space Model with Dynamic Stacking Filter for Real-World Video Deraining

Shangquan Sun[1,2] **Wenqi Ren**[3,4,5] **Juxiang Zhou**[6]
Shu Wang[7] **Jianhou Gan**[6] **Xiaochun Cao**[3†]
[1]Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
[2]School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
[3]School of Cyber Science and Technology, Shenzhen Campus of Sun Yat-sen University
[4]MoE Key Laboratory of Information Technology
[5]Guangdong Provincial Key Laboratory of Information Security Technology
[6]Key Laboratory of Educational Information for Nationalities, Yunnan Normal University
[7]School of Mechanical Engineering and Automation, Fuzhou University
shangquansun@gmail.com, {renwq3,caoxiaochun}@mail.sysu.edu.cn

Overview

In this supplementary material, we first provide additional quantitative results, specifically the outcomes of non-reference image quality assessments, detailed in Section 1. We then present the details of our proposed RVDT dataset in Section 2 and the proof of Theorem 1 in Section 3. Moreover, we present extensive visual comparisons among existing video deraining methods in Section 5.

1. Non-reference Quality Assessment on Real-world Unpaired Data

In the absence of ground truths for the rainy videos in the real-world test set of NTURain [1], we evaluate the qualitative performance using non-reference image quality assessment metrics, namely NIQE [7] and BRISQUE [6], computed on a randomly selected subset of 20 frames.

The results, presented in Table 1, clearly demonstrate that our method achieves superior image quality across both metrics. Additional visual examples are provided in Section 5.

Table 1. The metrics comparison of real-world rain streak removal for non-reference quality assessment on the real videos of NTURain [1].

	Input	FastDerain [5]	S2VD [14]	ESTINet [15]	MFGAN [13]	RainMamba [11]	Ours
NIQE↓	3.32	4.47	4.51	4.49	4.33	4.45	**2.94**
BRISQUE↓	20.21	21.49	22.87	23.43	22.39	22.69	**16.81**

2. Details of Our Proposed RVDT Dataset

Our RVDT is annotated using the open annotation tool, LabelMe [9]. Sample frames and statistics are presented in Fig. 1 and Tab. 2.

3. Proof of Theorem 1

Proof. We compute the sub-differential of the mean absolute deviation for the set as

$$\frac{\partial}{\partial x}\left(\frac{1}{n}\sum_{i=1}^{n}|x-x_i|\right)=\frac{1}{n}\sum_{i=1}^{n}\operatorname{sign}(x-x_i).$$

Table 2. The statistics of the proposed RVDT.

Class	Count	Unique	Class	Count	Unique
car	11180	167	handbag	350	1
person	9505	99	bus	283	5
motorbike	5931	76	aeroplane	229	1
umbrella	2840	24	traffic light	148	4
truck	1269	14	boat	130	1
cow	573	8	backpack	122	2
bicycle	402	5	train	115	1

Figure 1. 4 sample frames of real-world rainy sequences in RVDT.

It equals zero when $x = \text{median}(G)$. □

4. Single Image Deraining

To evaluate VDMamba's capability for single-image deraining, we further apply the spatial-branch-only variant, VDMamba-Single, to a real-world rainy image dataset, Internet-Data [10].

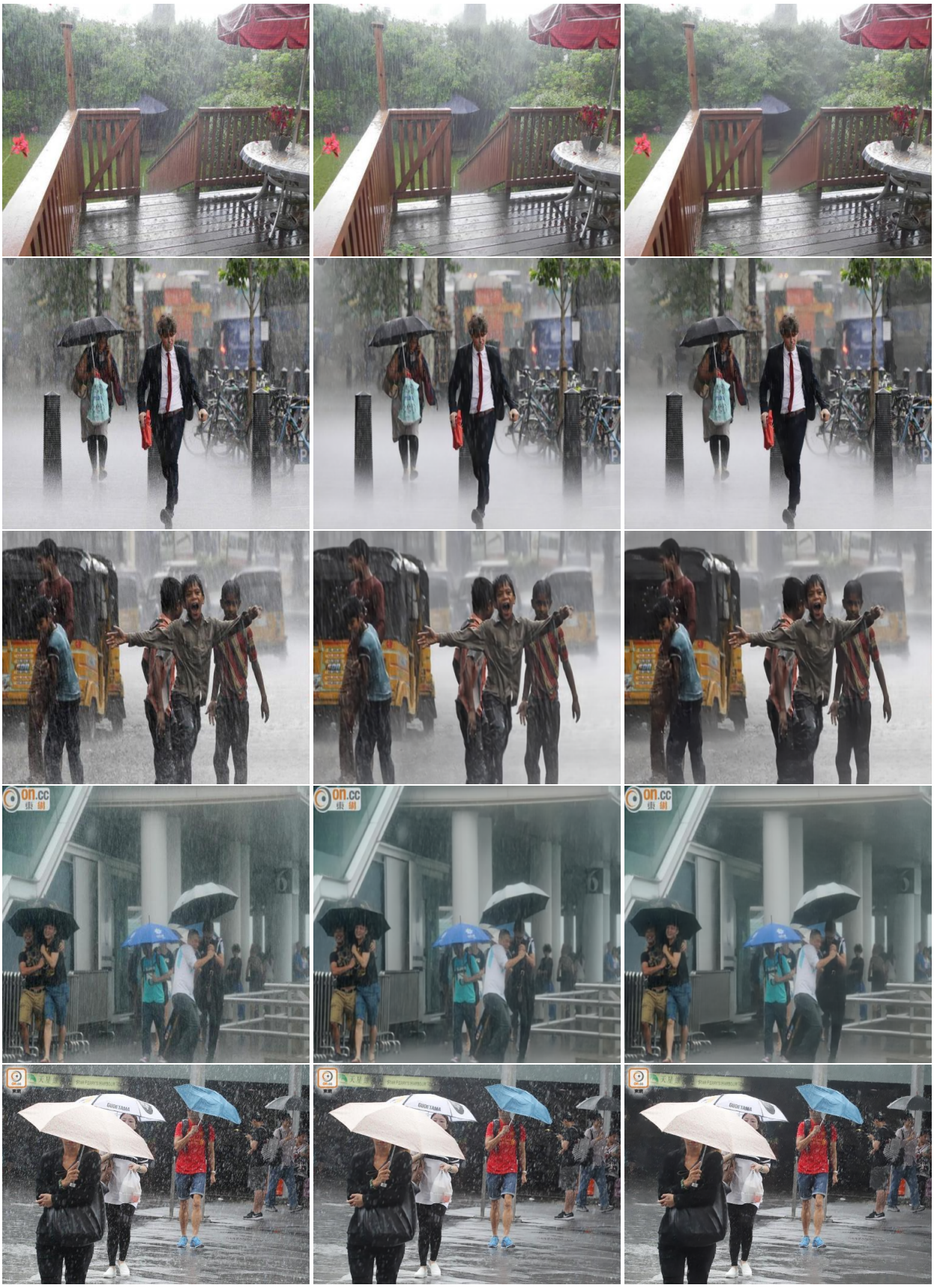
The visual comparisons between our model and the recent single-image deraining method, DRSformer [2], are presented in Figure 2. As depicted, our method demonstrates superior visual quality in real-world deraining scenarios.

5. More Visual Comparisons

In this section, we provide additional samples for visual comparison on rain-degraded video datasets. These include Figures 3 to 9 from NTURain [1], Figures 10 to 13 from our RVDT dataset, and Figures 14 to 15 sourced from the Internet.

For the evaluation of object detection and tracking in real-world rainy scenes, we assess various rain streak removal methods on real rainy videos from our RVDT dataset and present visual comparison results in Figures 16 to 22.

As illustrated, our VDMamba consistently achieves superior visual quality on real-world rainy video datasets and demonstrates the most significant improvements for downstream tasks.



(a) Input (b) DRSformer [2] (c) Ours

Figure 2. Visual comparisons of real-world single image deraining on Internet-Data [10]. The results of ours are generated by our VDMamba-single.

(a) Input (b) MSCSC [4]

(c) FastDerain [5] (d) DRSformer [2]

(e) ESTINet [15] (f) MFGAN [13]

(g) RainMamba [11] (h) Ours

Figure 3. A visual comparison of real-world video deraining on NTURain [1].

(a) Input (b) MSCSC [4]

(c) FastDerain [5] (d) DRSformer [2]

(e) ESTINet [15] (f) MFGAN [13]

(g) RainMamba [11] (h) Ours

Figure 4. A visual comparison of real-world video deraining on NTURain [1].

(a) Input (b) MSCSC [4]

(c) FastDerain [5] (d) DRSformer [2]

(e) ESTINet [15] (f) MFGAN [13]

(g) RainMamba [11] (h) Ours

Figure 5. A visual comparison of real-world video deraining on NTURain [1].

(a) Input (b) MSCSC [4]

(c) FastDerain [5] (d) DRSformer [2]

(e) ESTINet [15] (f) MFGAN [13]

(g) RainMamba [11] (h) Ours

Figure 6. A visual comparison of real-world video deraining on NTURain [1].

(a) Input (b) MSCSC [4]

(c) FastDerain [5] (d) DRSformer [2]

(e) ESTINet [15] (f) MFGAN [13]

(g) RainMamba [11] (h) Ours

Figure 7. A visual comparison of real-world video deraining on NTURain [1].

(a) Input (b) MSCSC [4]

(c) FastDerain [5] (d) DRSformer [2]

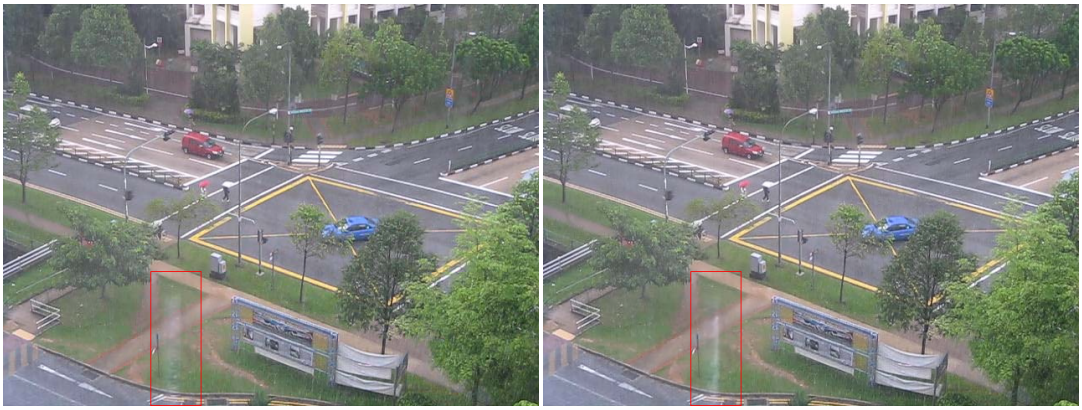
(e) ESTINet [15] (f) MFGAN [13]

(g) RainMamba [11] (h) Ours

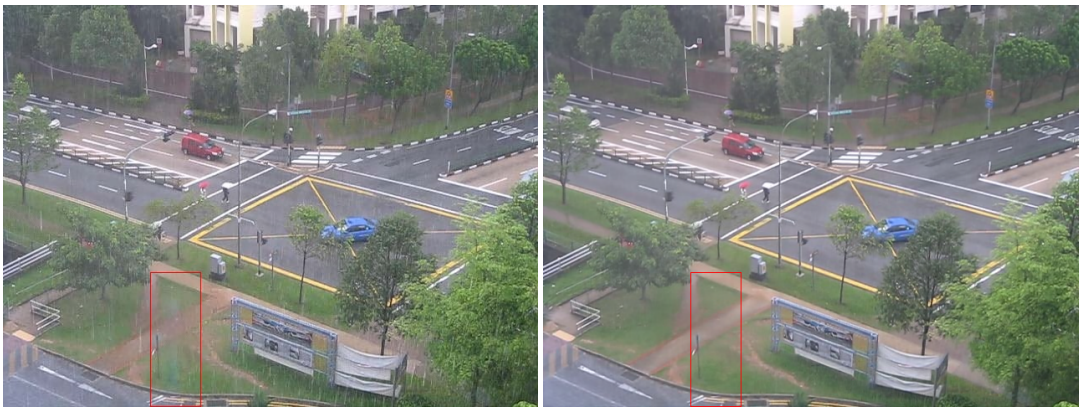
Figure 8. A visual comparison of real-world video deraining on NTURain [1].

(a) Input (b) MSCSC [4]

(c) FastDerain [5] (d) DRSformer [2]

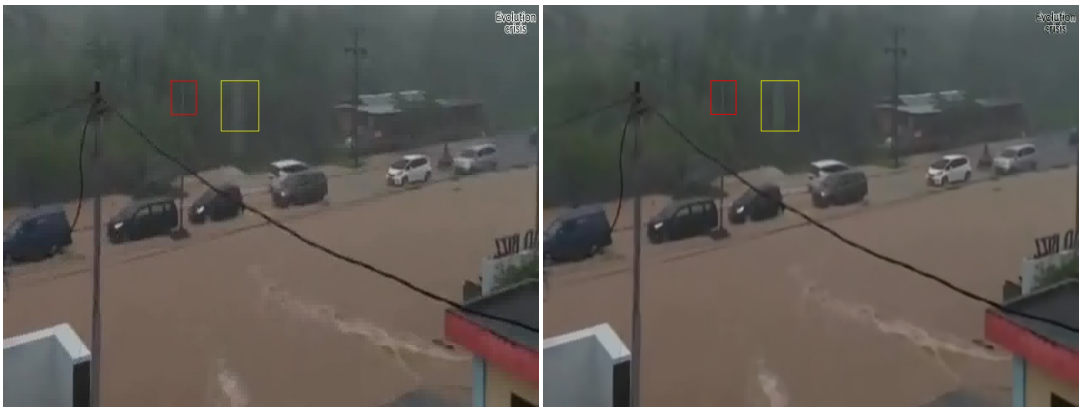


(e) ESTINet [15] (f) MFGAN [13]



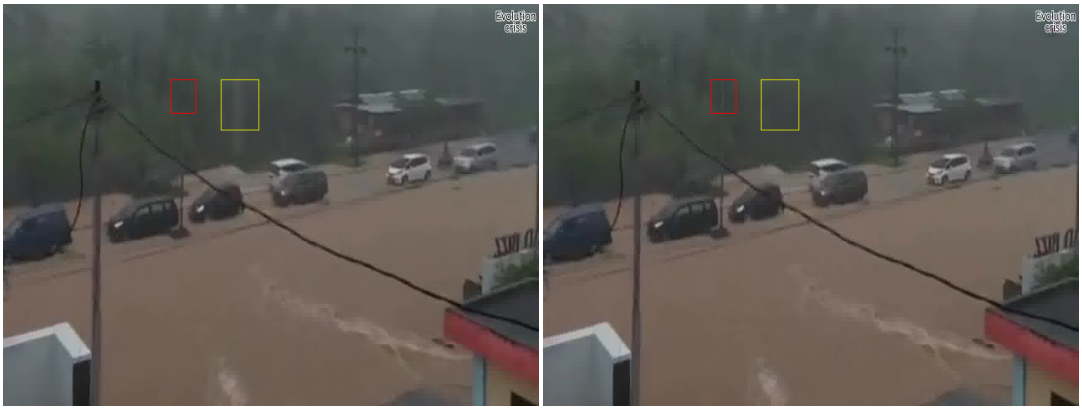
(g) RainMamba [11] (h) Ours

Figure 9. A visual comparison of real-world video deraining on NTURain [1].



(a) Input (b) SLDNet [12]

(c) FastDerain [5] (d) DRSformer [2]



(e) ESTINet [15] (f) MFGAN [13]

(g) RainMamba [11] (h) Ours

Figure 10. A visual comparison of real-world video deraining on RVDT.

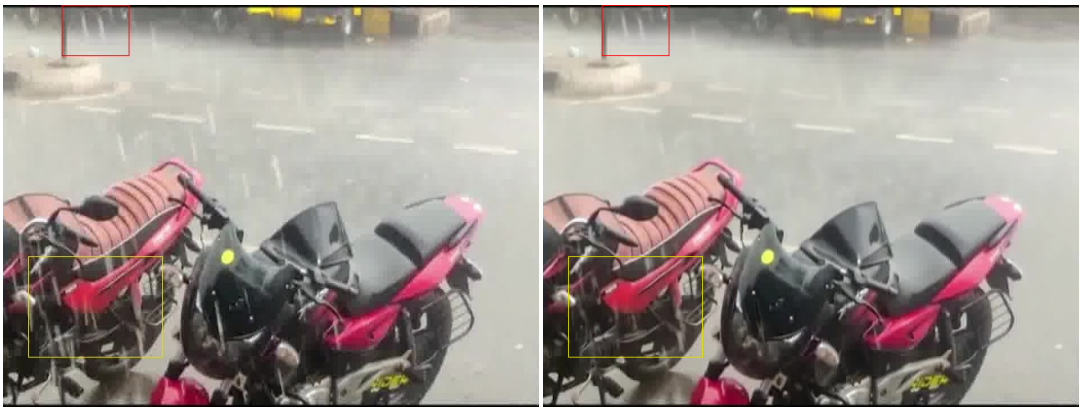
(a) Input (b) SLDNet [12]

(c) FastDerain [5] (d) DRSformer [2]

(e) ESTINet [15] (f) MFGAN [13]

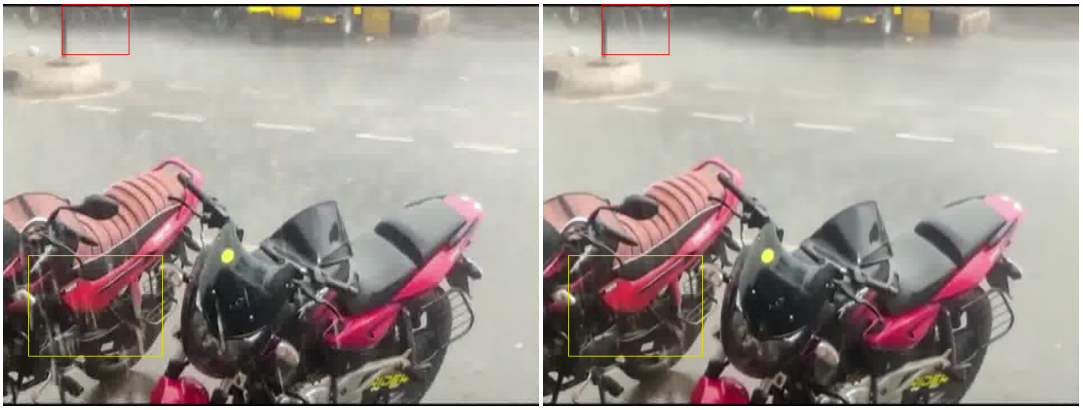
(g) RainMamba [11] (h) Ours

Figure 11. A visual comparison of real-world video deraining on RVDT.

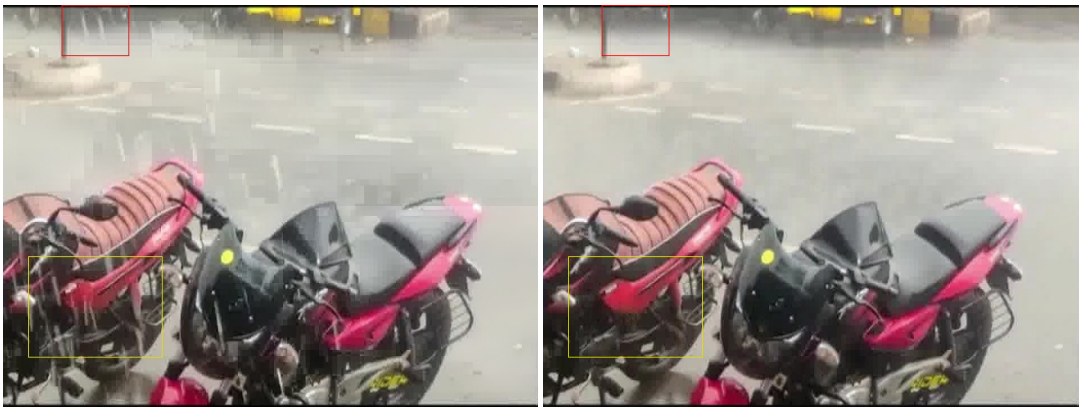


(a) Input (b) SLDNet [12]

(c) FastDerain [5] (d) DRSformer [2]

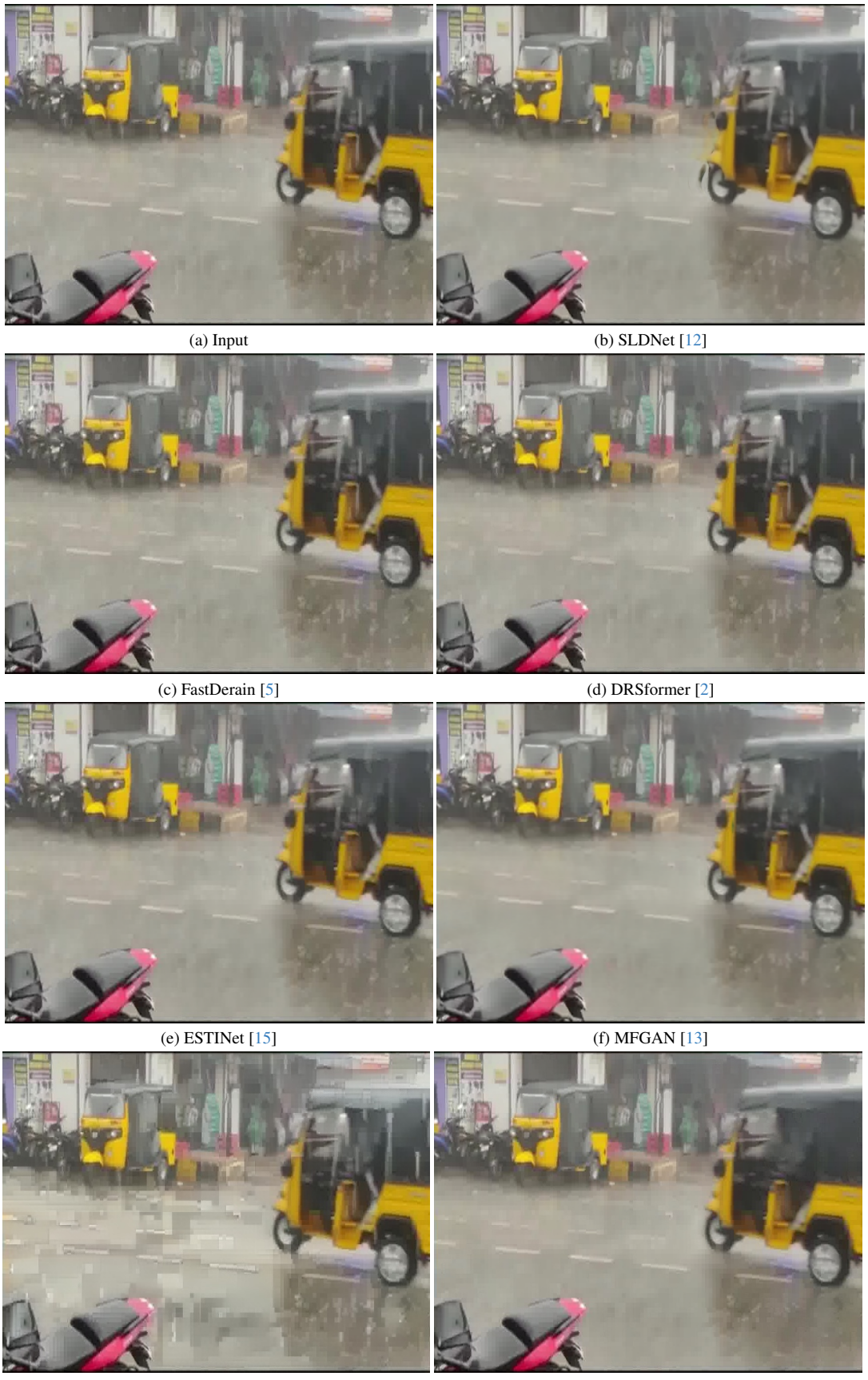


(e) ESTINet [15] (f) MFGAN [13]



(g) RainMamba [11] (h) Ours

Figure 12. A visual comparison of real-world video deraining on RVDT.



(a) Input (b) SLDNet [12]

(c) FastDerain [5] (d) DRSformer [2]

(e) ESTINet [15] (f) MFGAN [13]

(g) RainMamba [11] (h) Ours

Figure 13. A visual comparison of real-world video deraining on RVDT.

(a) Input (b) MSCSC [4]

(c) FastDerain [5] (d) DRSformer [2]

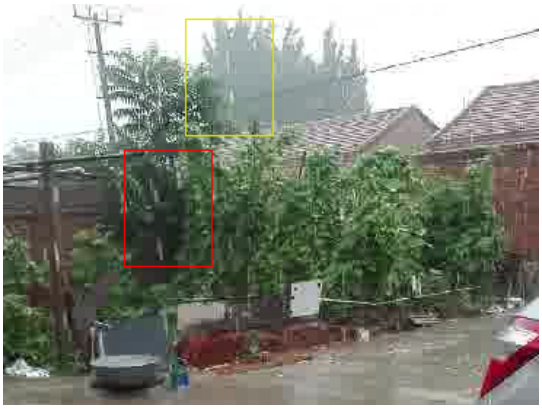
(e) ESTINet [15] (f) MFGAN [13]

(g) RainMamba [11] (h) Ours

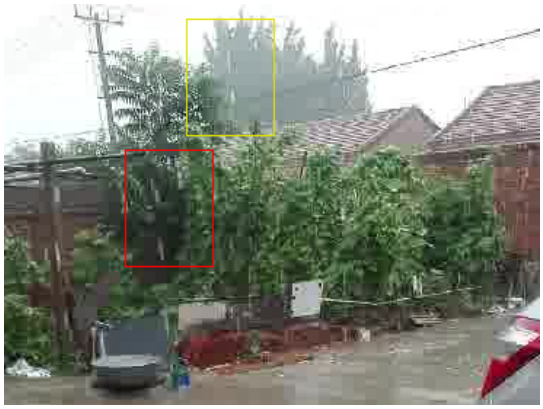
Figure 14. A visual comparison of real-world video deraining on a real-world rainy video from Internet.

(a) Input

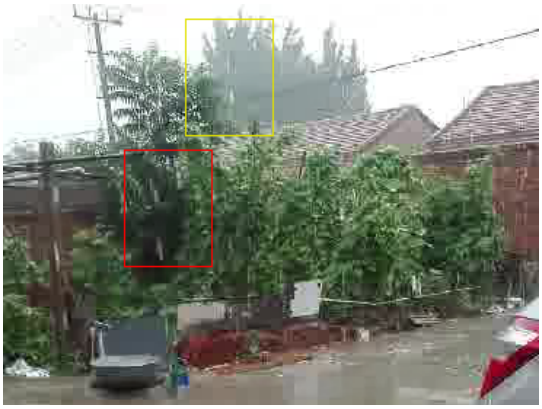
(b) MSCSC [4]



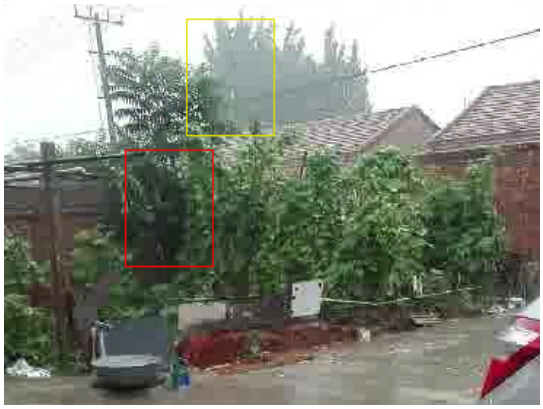
(c) FastDerain [5]



(d) DRSformer [2]

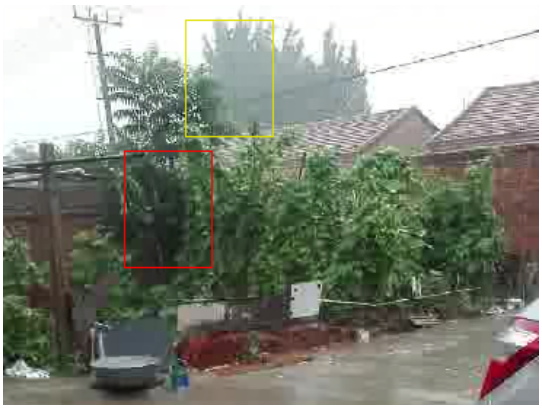


(e) ESTINet [15]



(f) MFGAN [13]

(g) RainMamba [11]



(h) Ours

Figure 15. A visual comparison of real-world video deraining on a real-world rainy video from Internet.

(a) Input (b) SLDNet [12]

(c) FastDerain [5] (d) DRSformer [2]

(e) ESTINet [15] (f) MFGAN [13]

(g) RainMamba [11] (h) Ours

Figure 16. A visual comparison of object detection on RVDT using YOLOv3 [8].

(a) Input (b) SLDNet [12]

(c) FastDerain [5] (d) DRSformer [2]

(e) ESTINet [15] (f) MFGAN [13]

(g) RainMamba [11] (h) Ours

Figure 17. A visual comparison of object detection on RVDT using YOLOv3 [8].

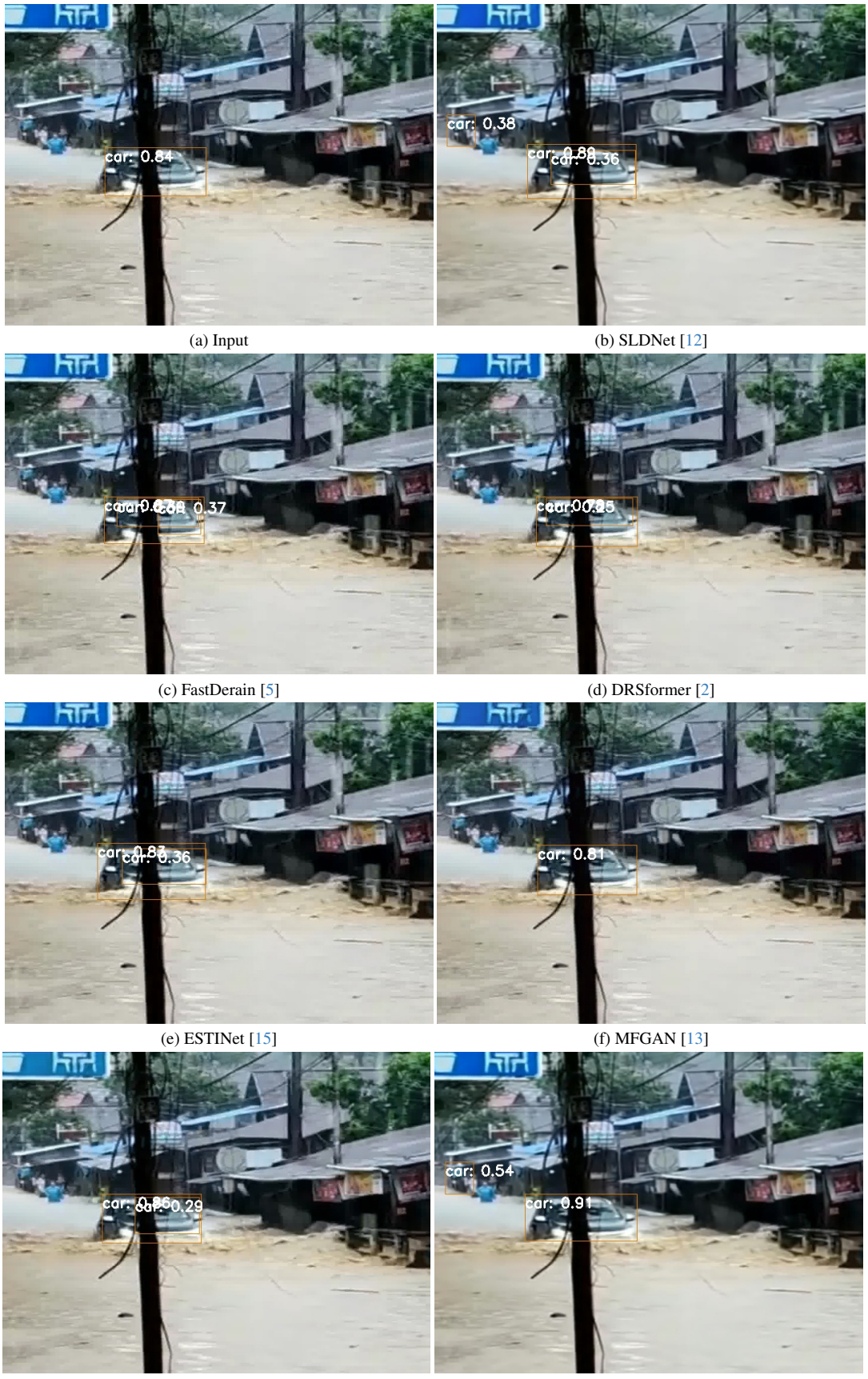
(a) Input (b) SLDNet [12]

(c) FastDerain [5] (d) DRSformer [2]

(e) ESTINet [15] (f) MFGAN [13]

(g) RainMamba [11] (h) Ours

Figure 18. A visual comparison of object detection on RVDT using mega [3].



(a) Input (b) SLDNet [12]

(c) FastDerain [5] (d) DRSformer [2]

(e) ESTINet [15] (f) MFGAN [13]

(g) RainMamba [11] (h) Ours

Figure 19. A visual comparison of object detection on RVDT using mega [3].

(a) Input (b) SLDNet [12]

(c) FastDerain [5] (d) DRSformer [2]

(e) ESTINet [15] (f) MFGAN [13]

(g) RainMamba [11] (h) Ours

Figure 20. A visual comparison of object tracking on RVDT using GTR [16].

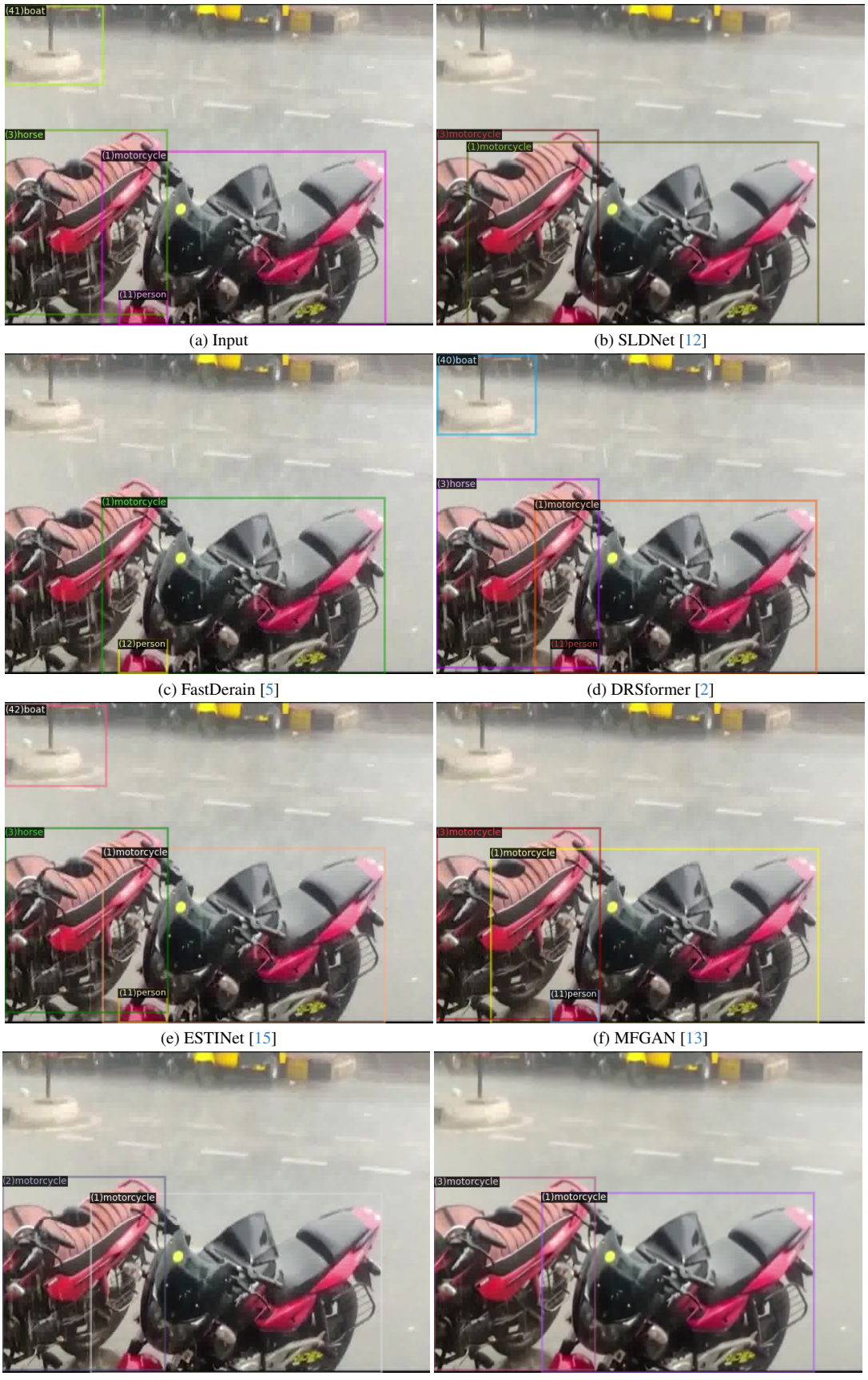
(a) Input (b) SLDNet [12]

(c) FastDerain [5] (d) DRSformer [2]

(e) ESTINet [15] (f) MFGAN [13]

(g) RainMamba [11] (h) Ours

Figure 21. A visual comparison of object tracking on RVDT using GTR [16].



(a) Input (b) SLDNet [12]

(c) FastDerain [5] (d) DRSformer [2]

(e) ESTINet [15] (f) MFGAN [13]

(g) RainMamba [11] (h) Ours

Figure 22. A visual comparison of object tracking on RVDT using GTR [16].

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 *CVPR*, 2018. 1, 2, 4, 5, 6, 7, 8, 9, 10

[2] Xiang Chen, Hao Li, Mingqiang Li, and Jinshan Pan. Learning a sparse transformer network for effective image deraining. In *CVPR*, 2023. 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23

[3] Yihong Chen, Yue Cao, Han Hu, and Liwei Wang. Memory enhanced global-local aggregation for video object detection. In *CVPR*, 2020. 19, 20

[4] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. In *CVPR*, 2020. 4, 5, 6, 7, 8, 9, 10, 15, 16

[5] Tai-Xiang Jiang, Ting-Zhu Huang, Xi-Le Zhao, Liang-Jian Deng, and Yao Wang. Fastderain: A novel video rain streak removal method using directional gradient priors. *IEEE TIP*, 2019. 1, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23

[6] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE TIP*, 2012. 1

[7] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 2012. 1

[8] Joseph Redmon. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. 17, 18

[9] Bryan C Russell, Antonio Torralba, Kevin P Murphy, and William T Freeman. Labelme: a database and web-based tool for image annotation. *International journal of computer vision*, 77:157–173, 2008. 1

[10] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson W.H. Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In *CVPR*, 2019. 2, 3

[11] Hongtao Wu, Yijun Yang, Huihui Xu, Weiming Wang, Jinni Zhou, and Lei Zhu. Rainmamba: Enhanced locality learning with state space models for video deraining. In *ACM MM*, 2024. 1, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23

[12] Wenhan Yang, Robby T. Tan, Shiqi Wang, and Jiaying Liu. Self-learning video rain streak removal: When cyclic consistency meets temporal correspondence. In *CVPR*, 2020. 11, 12, 13, 14, 17, 18, 19, 20, 21, 22, 23

[13] Wenhan Yang, Robby T. Tan, Jiashi Feng, Shiqi Wang, Bin Cheng, and Jiaying Liu. Recurrent multi-frame deraining: Combining physics guidance and adversarial learning. *IEEE TPAMI*, 2022. 1, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23

[14] Zongsheng Yue, Jianwen Xie, Qian Zhao, and Deyu Meng. Semi-supervised video deraining with dynamical rain generator. In *CVPR*, 2021. 1

[15] Kaihao Zhang, Dongxu Li, Wenhan Luo, Wenqi Ren, and Wei Liu. Enhanced spatio-temporal interaction learning for video deraining: A faster and better framework. *IEEE TPAMI*, 2022. 1, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23

[16] Xingyi Zhou, Tianwei Yin, Vladlen Koltun, and Philipp Krähenbühl. Global tracking transformers. In *CVPR*, 2022. 21, 22, 23