

Spatially-Adaptive Image Restoration using Distortion-Guided Networks

SUPPLEMENTARY DOCUMENT

Kuldeep Purohit¹ Maitreya Suin²
¹ Michigan State University

A. N. Rajagopalan² Vishnu Naresh Boddeti¹
² Indian Institute of Technology Madras

Supplementary Details: This document is organized as follows. Sections S1 and S2 contains additional analysis and visualization of various aspects of our localization and restoration networks. Experiments on performing multi-task learning using SPAIR are provided in Section S3. The document concludes with Section S4 providing additional qualitative results and comparisons on various benchmark datasets for all four tasks.

S1. Model Analysis

S1.1. Effect of Spatially-Selective Processing

We perform a few analytical experiments to verify the effectiveness of spatially-selective processing. We vary the threshold (for intensity change to be classified as significant distortion) and calculate the PSNR on thus obtained degraded and non-degraded regions separately. Owing to the adaptiveness of our approach and specialized modules for the current task, we achieve better restoration results for the regions with significant corruption (Fig. S2). This behavior is generally expected from better restoration network design. Interestingly, we observed that, for regions with negligible amount of degradation (non-degraded regions for very low threshold), the input image itself is sufficiently good (Fig. S2). Due to generic processing, most of the previous methods are unable to reconstruct those regions accurately (i.e. they corrupt the background pixels) and result in poor PSNR even in the simplest of regions.

Next, we visualize the error-maps (the difference between degraded and ground-truth image) for the non-degraded regions in Figs. S3 (RainStreak) and Figs. S4 (Raindrop). As we can observe, our method results in the least amount of reconstruction error. The improvement over existing methods in the non-degraded areas shows that SPAIR causes least amount of changes in the non-degraded pixels. Improved restoration in degraded regions is attributed to modules specifically processing the degraded regions. Our method preserves input details which are uncorrupted and is able to improve the quality of the restoration of affected regions.

S1.2. Effect of SNL

The SNL module processes only degraded pixels by non-locally aggregating information selectively from the non-degraded regions. Attention over regions filtered by mask selects only the most relevant features, improving performance. In SNL's first step, each pixel/position receives weighed aggregation of features lying only in horizontal and vertical directions. Serially performing two such steps expands the span to all pixels. For a direction d and degraded location (i,j) , Eq.3 generates an attention map using $\mathbf{F}_{i,j}^d$ (features from non-degraded locations). Next, Eq.4 performs a weighted sum to gather contextual information from uncorrupted positions based on their utility (measured using the attention maps). Instead of operating with a $N^2 \times N^2$ attention matrix (as in [59]), decomposition into processing sparse pixels in 4 directions separately reduces complexity to $O(N\sqrt{N})$, making SNL effective and memory efficient.

S2. Analysis on Distortion Localization Network (Net_L)

We have visualized predicted distortion-mask along with the ground truth distortion-mask in Fig. S1. The close resemblance of the predicted and ground truth degradation maps shows the effectiveness of Net_L .



Figure S1. Visualization of degradation mask for different tasks. First, second, third row describes input image, ground-truth mask, predicted mask respectively

S2.1. Supervised vs unsupervised learning

We choose to train Net_L in a supervised fashion, since accurate pixel-level distortion estimation is of key importance in restoration.

S2.2. Impact of Accuracy of Net_L

Note that, compared to the ground truth mask, there will inevitably be some errors in the predicted one. Although our proposed modules in the decoder leverage the extra guidance from the predicted mask, as there are standard convolution layers too in the decoder and the restoration network is trained with the predicted mask itself, errors in very few pixels of the predicted mask do not adversely affect the final restoration output. Net_L classifies some regions as non-degraded, while slight intensity distortions may well be present in those areas. Nonetheless, restoration of such distortions is easy and this is achieved through the few non-sparse layers in our network.

We conduct the following experiment to show that at convergence, final performance of Net_R is not very sensitive to small error in the prediction of Net_L . The variation of Net_R 's performance with respect to Net_L 's cross-entropy loss is shown in Table S1

Table S1. Influence of the accuracy of Net_L on performance of SPAIR for raindrop removal evaluated using AGAN Dataset [13].

Epoch	20	50	70	90
BCE Loss ($\times 10^{-2}$)	8.4	6.06	5.95	5.92
PSNR	32.08	32.67	<u>32.73</u>	32.73

S2.3. Relationship between mask and blur

We performed the suggested analysis using optical flow, which is a good approximation to motion blur (as demonstrated in [10]). Fig. S2.3 shows optical flow for two consecutive sharp frames (from dataset [35]), indicating spatial distribution of blur magnitude. The blurred image (available in [35]) corresponding to the sharp frame is passed through Net_L to obtain the distortion mask. High spatial correlation can be observed, establishing the utility of our mask. Dynamic objects or regions closer to the camera usually contain highly blurred texture, and the same regions are revealed in the mask \mathcal{M} . Note that textureless pixels (eg. in clothes, skin) incur minimal distortions from blur and hence are absent in \mathcal{M} .



S2.4. Comparison with AGAN [13]

Balance error rate [7] scores (with best threshold) for AGAN [13] and Net_L are 8.15 and 4.58, justifying the utility of Net_L .

S2.5. Choice of loss

Even through advanced loss functions/training optimizers would offer benefits complementary to our architecture, we used L1 loss for fair comparison of SPAIR with all competing restoration methods [44,51,56]. L1 error value at $\mathcal{M} = 0$ is often smaller than at $\mathcal{M} = 1$, making it predominantly a restoration loss. Nonetheless, there is potential for further performance improvements by incorporating perceptual loss and confidence-driven reconstruction loss specifically on the degraded regions.

S3. Exploring Multi-Task Learning

We also explore an additional benefit of our design. Since our architecture design does not change across restoration tasks, it opens venues for multi-task learning. We perform experiments on jointly learning two restoration tasks. Among the 4 spatially-varying degradations we consider in this paper, we choose to address Rain-Streaks and RainDrops jointly, as they are closely related to each other and generally occur in similar environments. We explore the possibility of obtaining a single trained model which can remove raindrops as well as rainstreaks from a given test image. We train SPAIR jointly on the datasets of two tasks: Mixed RainStreak Dataset [9] (used in Table 1 of main paper) and RainDrop Dataset [13] (used in Table 3 of main paper). After training jointly on the two tasks, we evaluate the model on the two benchmarks and compare against existing task specific methods. We also include the recent multi-task restoration model OWAN [17] as a baseline and train it in the same setting as SPAIR.

Specifically, we train our model on 14573 clean-degraded image pairs gathered from rain-streak datasets [4, 11, 22, 24, 25] and the raindrop dataset [13]. With this single trained model (referred to as SPAIR (Joint)), we perform evaluation on different test sets, including Rain100H [22], Rain100L [22], Test100 [25], Test2800 [4], Test1200 [24] and AGAN [13]. The results on the two tasks are reported in Tables S2 and S3. It is evident that SPAIR (Joint) is the first model in literature to achieve state-of-the-art results on Rain-Streak and RainDrop removal tasks, without additional training.

S4. Additional qualitative comparisons

Rain-Streaks: Figs. S5, S6 show additional qualitative results and comparisons state-of-the-art methods on Rain100H Dataset (Table 2 of main paper). Existing methods suffer from visible rain streaks or texture-smearing

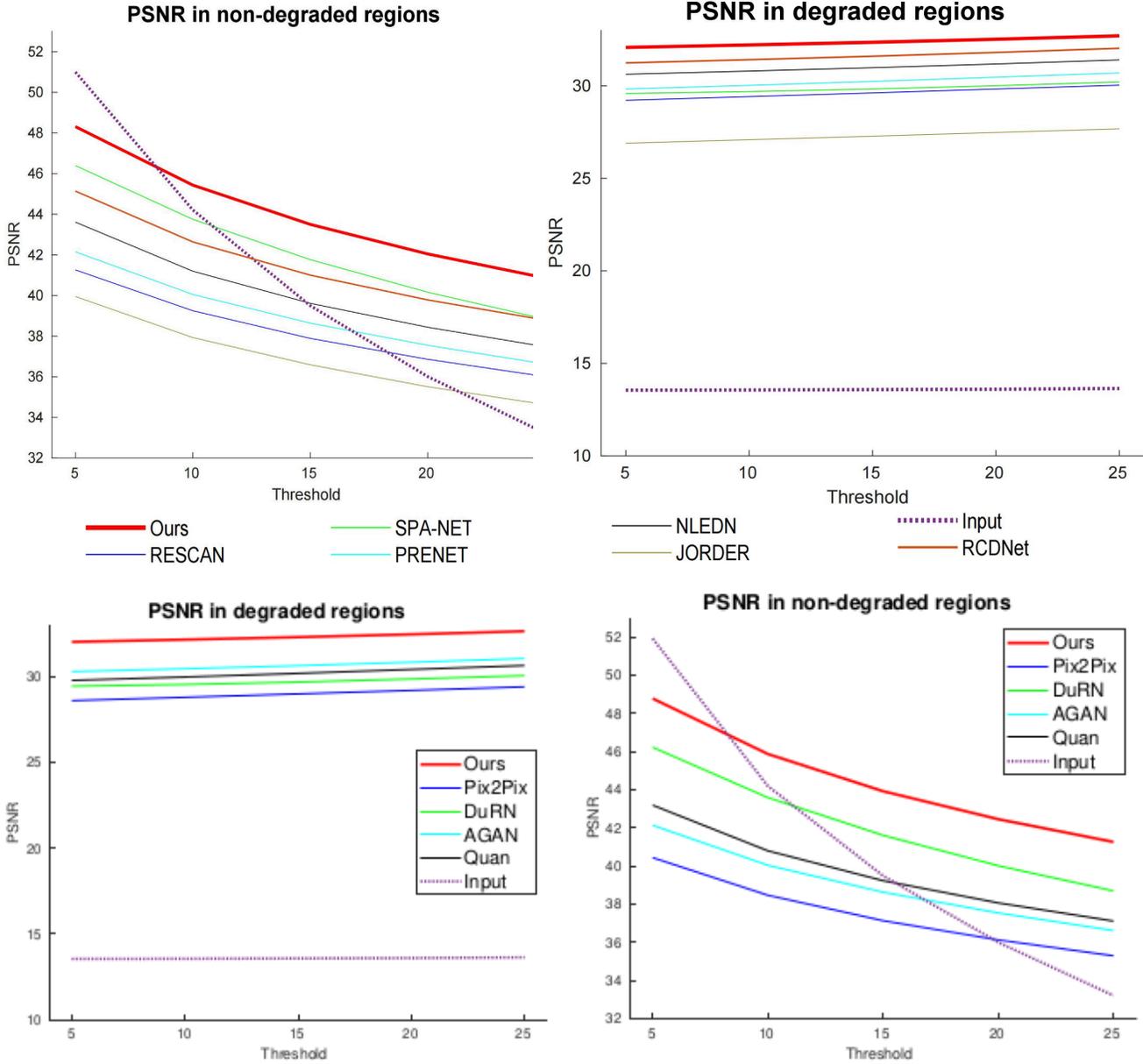


Figure S2. Comparison with baseline methods using PSNR scores in degraded and non-degraded regions for two tasks: rainstreak removal and raindrop removal. (Best viewed in color).

Table S2. Image deraining results using SPAIR trained jointly for Rain-Streak and Raindrop removal tasks. Best and second best scores are **highlighted** and underlined. SPAIR significantly outperforms baselines methods in both settings: Single task and Joint task learning.

Methods	Test100 [25]		Rain100H [22]		Rain100L [22]		Test2800 [4]		Test1200 [24]		Average	
	PSNR \uparrow	SSIM \uparrow										
MSPFN [9]	27.50	0.876	28.66	0.860	32.40	0.933	32.82	0.930	32.39	0.916	30.75	0.903
SPAIR (Single)	30.35	0.909	30.95	0.892	36.93	0.969	33.34	0.936	<u>33.04</u>	<u>0.922</u>	32.91	0.926
OWAN [17] (Joint)	23.85	0.810	24.46	0.724	28.54	0.878	30.40	0.891	30.09	0.872	27.47	0.835
SPAIR (Joint)	<u>30.33</u>	<u>0.909</u>	<u>30.81</u>	<u>0.892</u>	<u>36.39</u>	<u>0.964</u>	<u>33.34</u>	<u>0.936</u>	33.10	0.925	<u>32.79</u>	<u>0.925</u>

along rain direction. In comparison, our results are visually more pleasing, while being faithful to the ground-truth im-

age. Fig. S7 contains comparisons of all methods trained on the combined RainStreak Dataset (Table 1 of main paper).

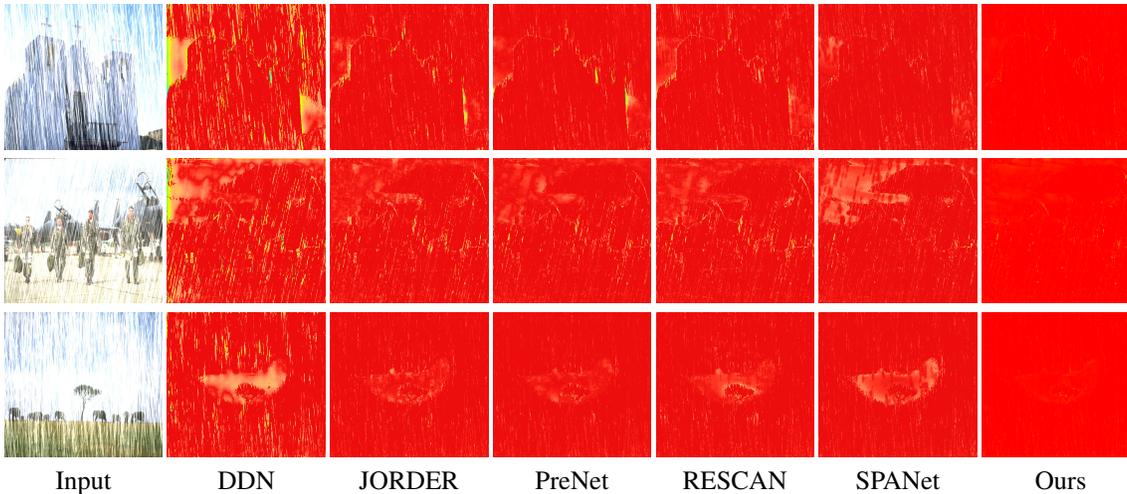


Figure S3. Visualization of errors only on the pixels belonging to non-degraded regions (on images affected with Rain-Streaks from Rain100H dataset). From left to right: Input Image, DDN, JORDER, PRENET, RESCAN, SPANET, OURS. (Best viewed in color and zoomed-in).

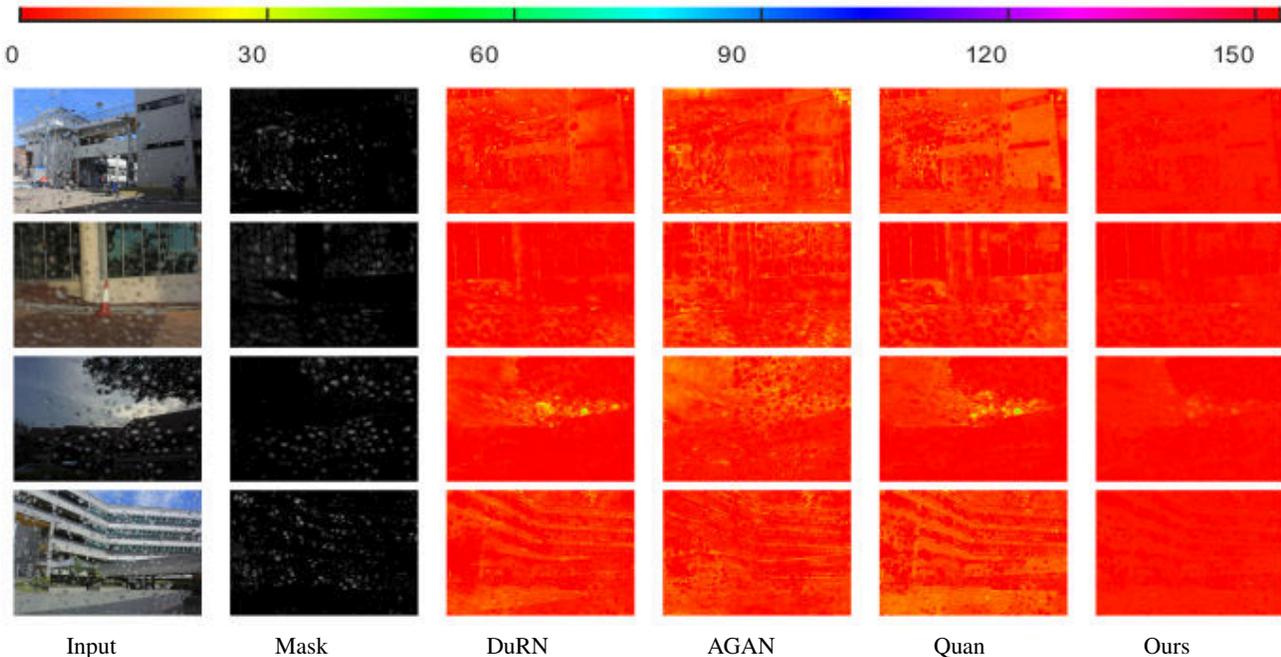


Figure S4. Comparison with baseline methods using error maps calculated only in non-degraded regions within images affected with Raindrops. (Best viewed in color).

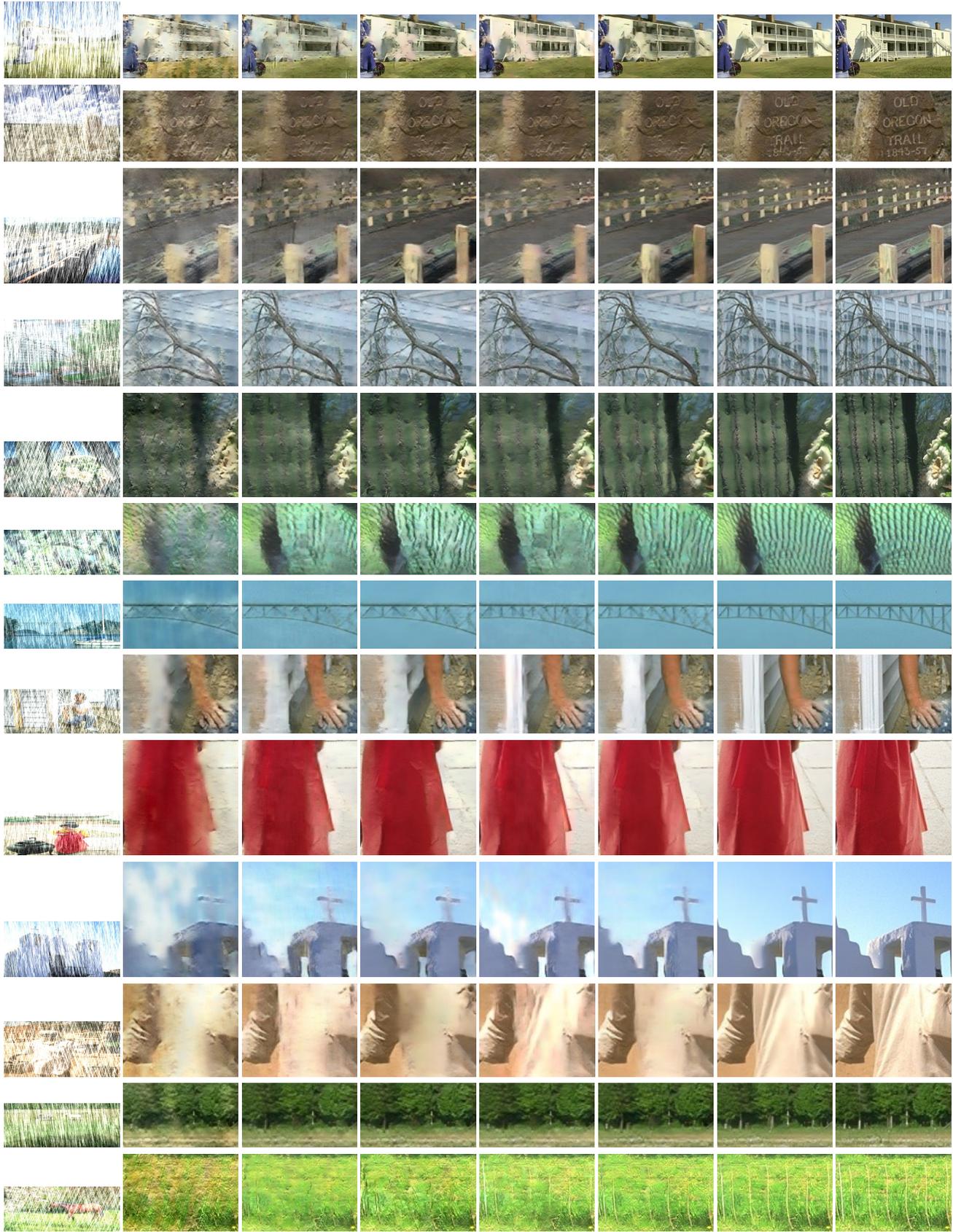
It is evident that few rain streaks remain visible and background remains unclear in the results of all existing methods while our approach generates satisfactory deraining results.

Raindrop: We show additional results on the test-set of AGAN dataset in Figs. S8,S9,S10. We also include comparisons on a real-world image in Fig. S11. Visually, we can observe significant improvement over prior works.

Shadow Removal: Fig. S12 provides additional qualitative comparisons on shadow removal showing that most existing approaches produce shadow boundaries or color incon-

sistencies. In contrast, SPAIR has minimal artifacts in the shadow boundaries, outperforming the baselines both qualitatively and quantitatively.

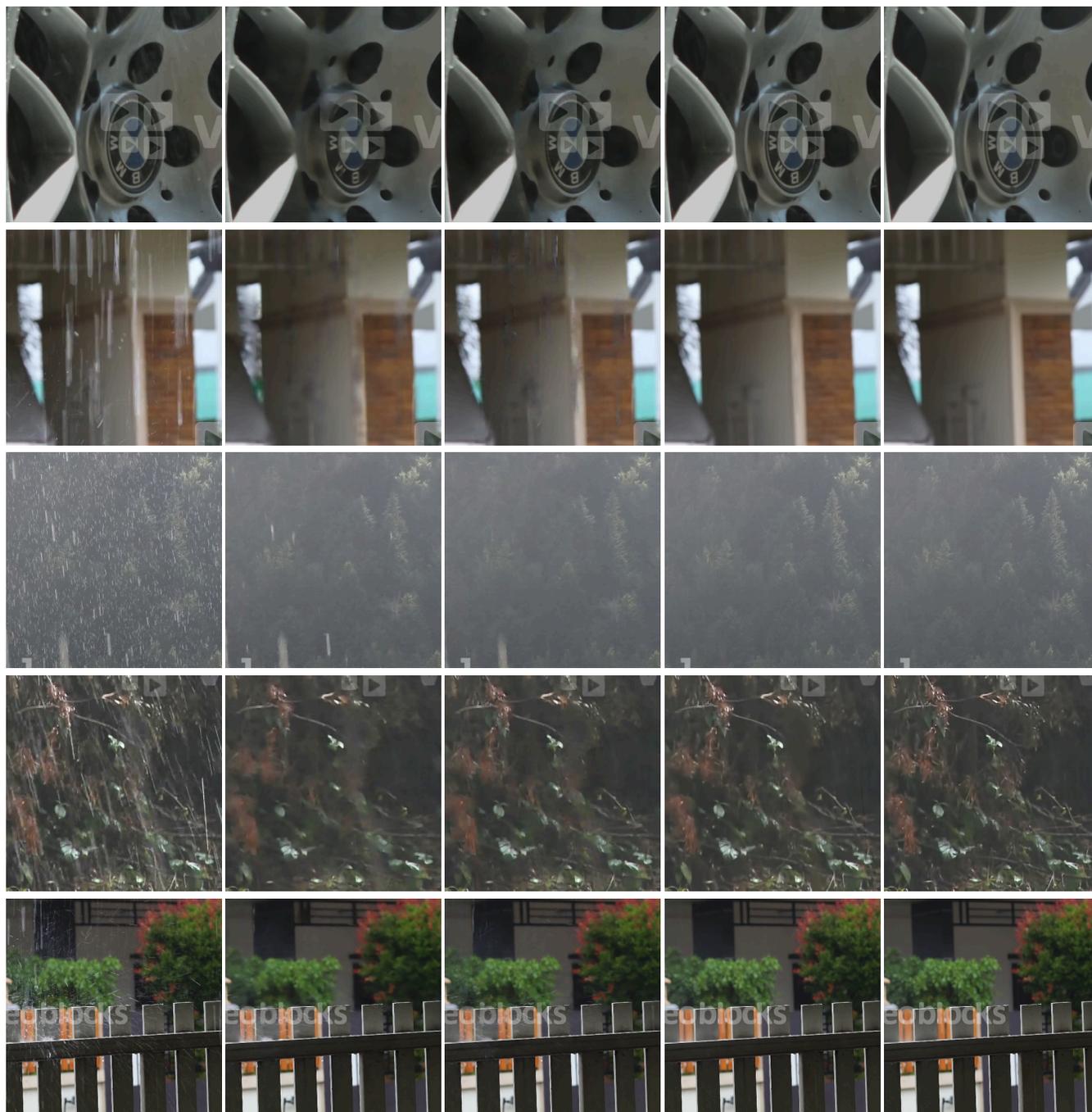
Motion Blur: While the GoPro and HIDE datasets are generated by averaging consecutive frames of real high frame-rate videos, the blurred images in RealBlur-J dataset are captured in real-world conditions. In Fig. S13 we provide comparisons of our results with the best results from [16]. In Figs. S14-S18, we provide additional results and visual comparisons of our architectures with competing meth-



Input DDN RESCAN PReNet SPA-Net RCDNet Ours GT
 Figure S5. Qualitative comparison of results on test images from the Rain100H test-set (corresponding to Table 2 of main paper).

Table S3. Raindrop removal results (AGAN Dataset [13]) using SPAIR trained jointly for Rain-Streak and Raindrop removal tasks. SPAIR yield better results than existing methods in both settings: Single task and Joint task learning.

Method	Eigen [2]	Pix2pix [8]	AGAN[13]	DuRN[12]	Quan[14]	SPAIR (Single)	OWAN [17] (Joint)	SPAIR (Joint)
PSNR	28.59	30.59	31.51	31.24	31.44	32.73	28.44	<u>32.59</u>
SSIM	0.6726	0.8075	0.9213	0.9259	0.9263	0.9410	0.841	<u>0.935</u>



(a) Input (b) SPA-Net [20] (b) RCDNet [18] (c) Ours (d) GT

Figure S6. Visual comparisons on real rain-affected images from the SPANet dataset [20] (corresponding to Table 2 of main paper).

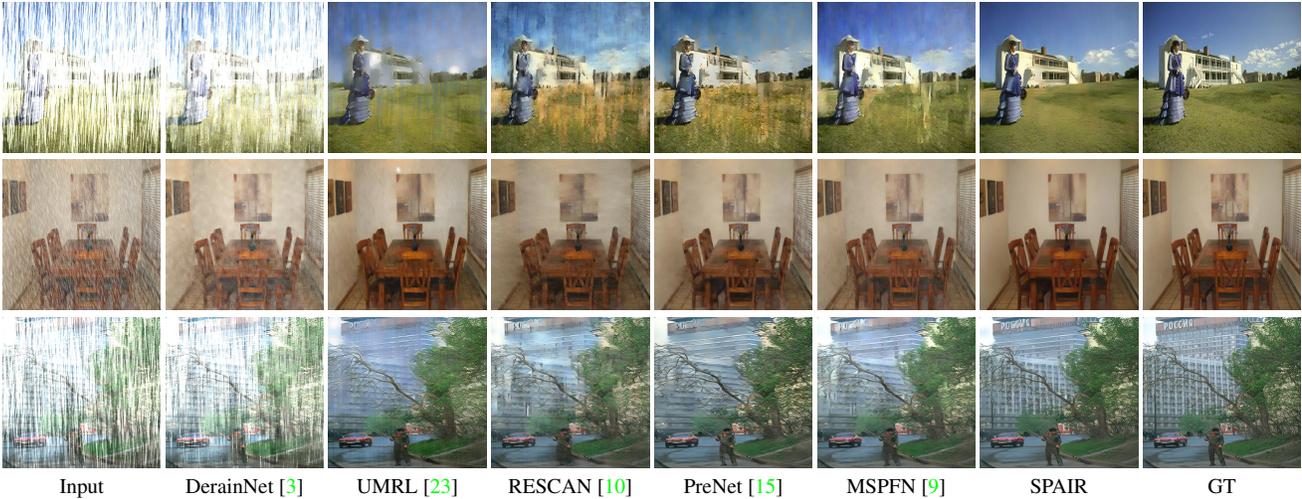


Figure S7. Qualitative comparisons on test images from various benchmarks considered in Table 1 of main paper.

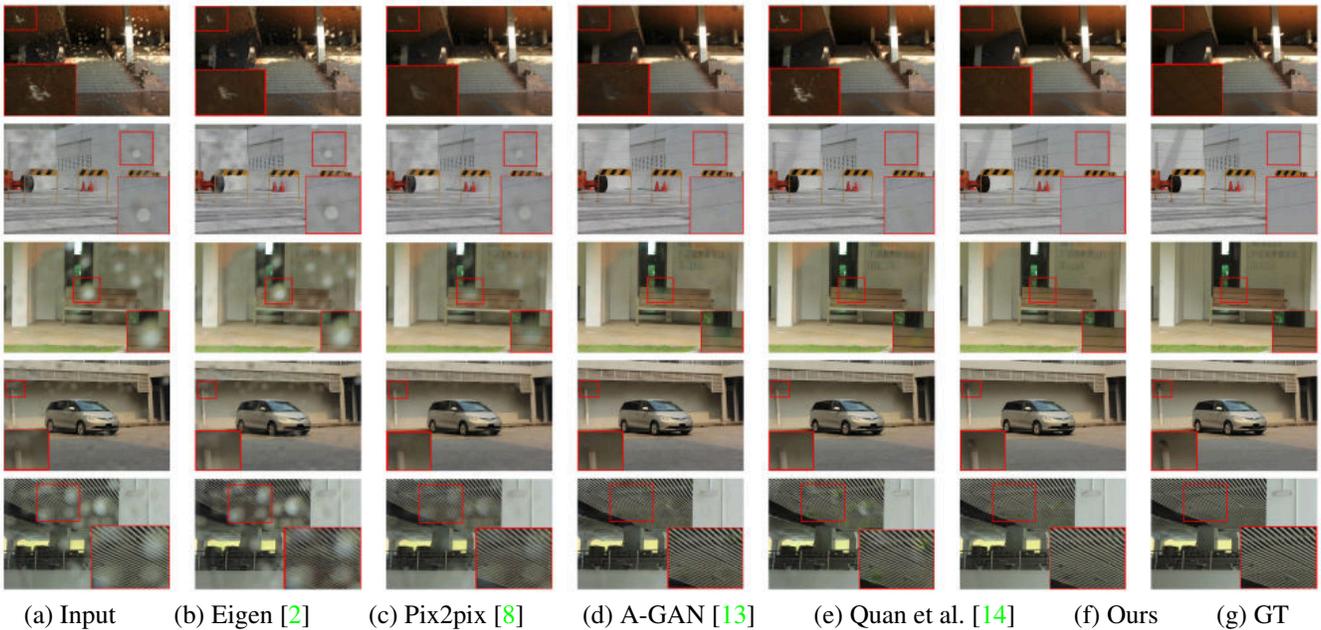
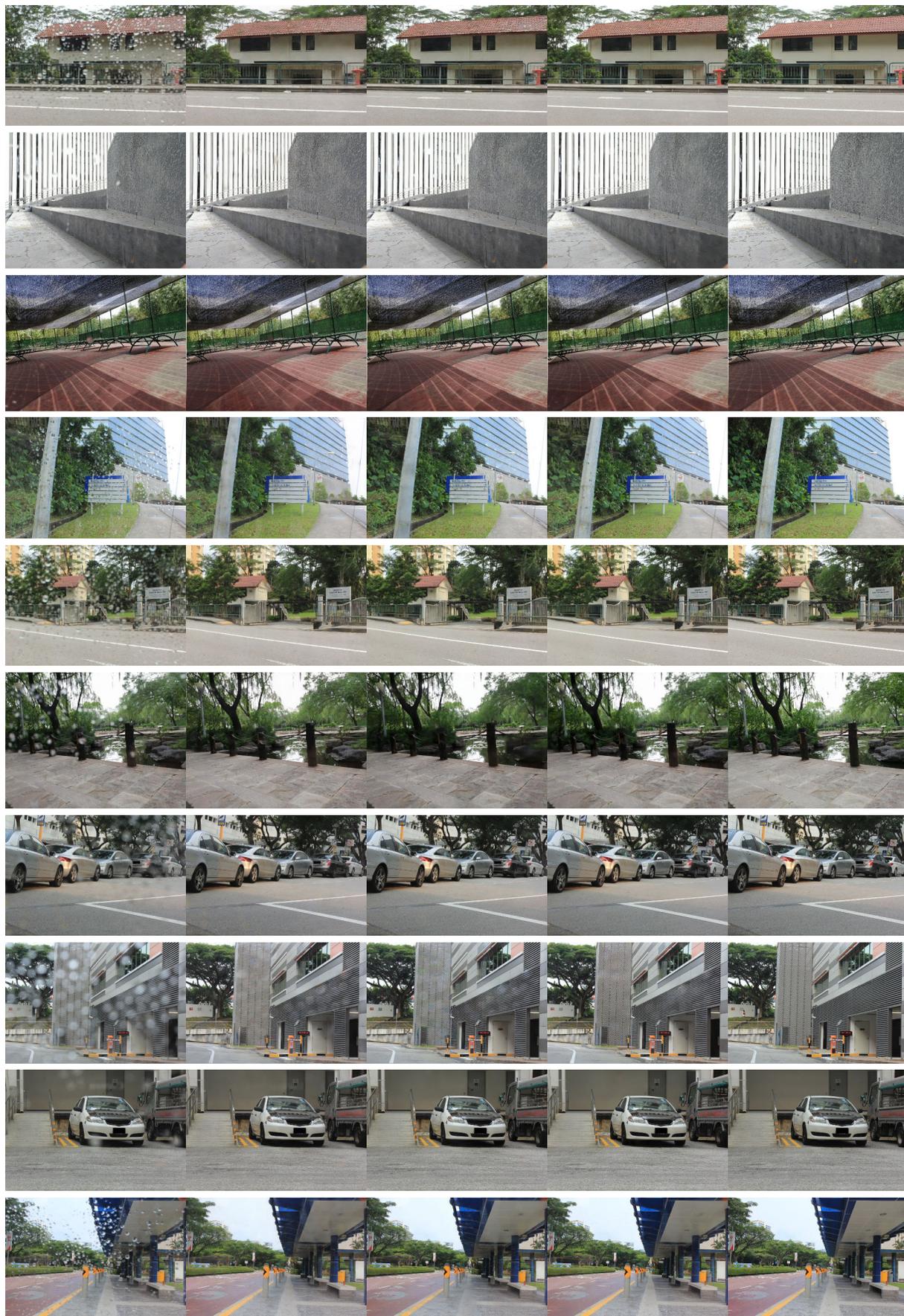


Figure S8. Qualitative comparisons of results on test images from the AGAN testset [13].

ods on the GoPro deblurring benchmark. The visual results show that our results closely mimic the ground-truth sharp images, while producing artifact-free results in regions containing challenging blur. Improvements over prior methods become more pronounced on images affected with large blur.

References

- [1] Xiaodong Cun, Chi-Man Pun, and Cheng Shi. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan. In *AAAI*, pages 10680–10687, 2020. 10
- [2] David Eigen, Dilip Krishnan, and Rob Fergus. Restoring an image taken through a window covered with dirt or rain. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 633–640, 2013. 6, 7, 10
- [3] Xueyang Fu, Jiabin Huang, Xinghao Ding, Yinghao Liao, and John Paisley. Clearing the skies: A deep network architecture for single-image rain removal. *TIP*, 2017. 7
- [4] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *CVPR*, 2017. 2, 3
- [5] Han Gong and Darren Cosker. Interactive shadow removal and ground truth for variable scene categories. In *BMVC*, pages 1–11. Citeseer, 2014. 10
- [6] Ruiqi Guo, Qieyun Dai, and Derek Hoiem. Paired regions for shadow detection and removal. *IEEE transactions on pattern*



(a) Input

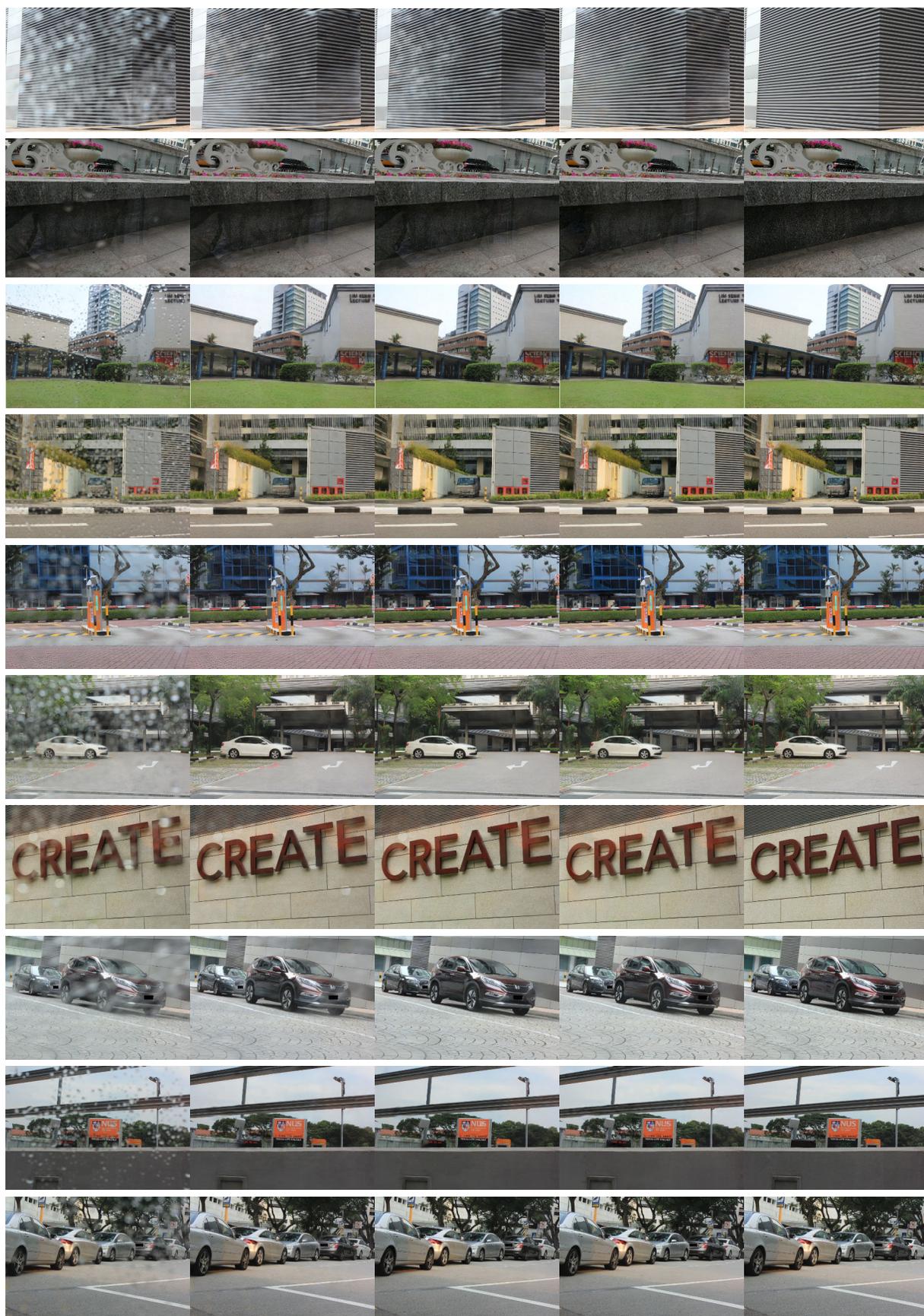
(b) A-GAN [13]

(c) DuRN [12]

(d) Ours

(e) GT

Figure S9. Qualitative comparisons of results on images from the AGAN testset [13].



(a) Input

(b) A-GAN [13]

(c) DuRN [12]

(d) Ours

(e) GT

Figure S10. Qualitative comparisons of results on images from the AGAN testset [13].

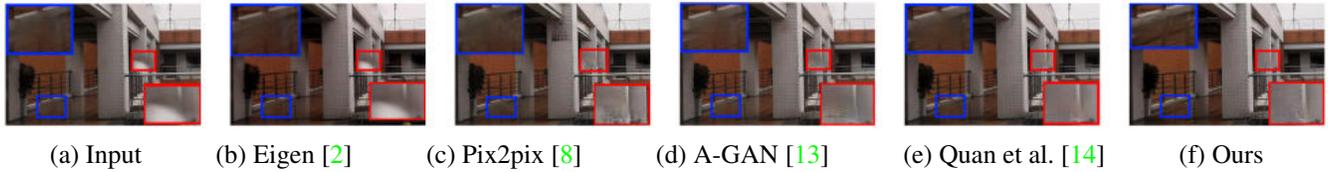


Figure S11. Qualitative comparisons of results on a real-world image from the AGAN testset [13].

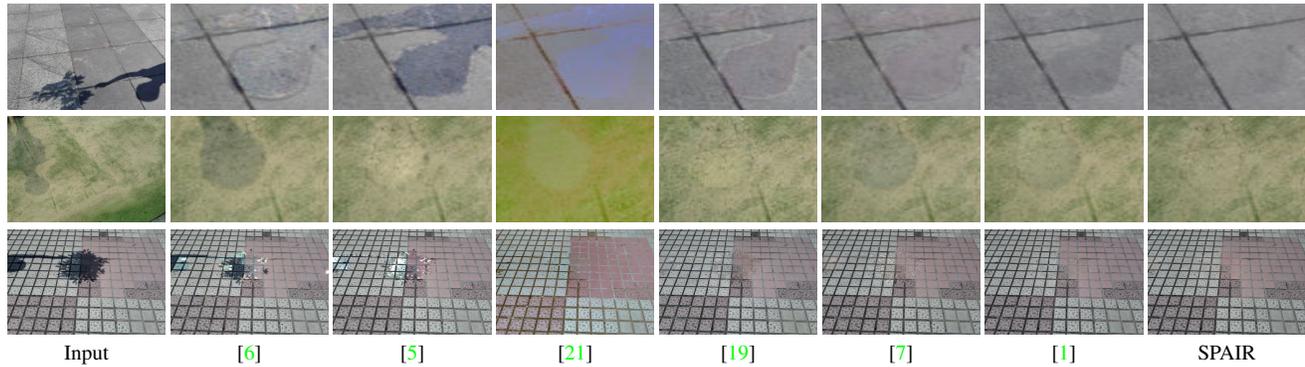
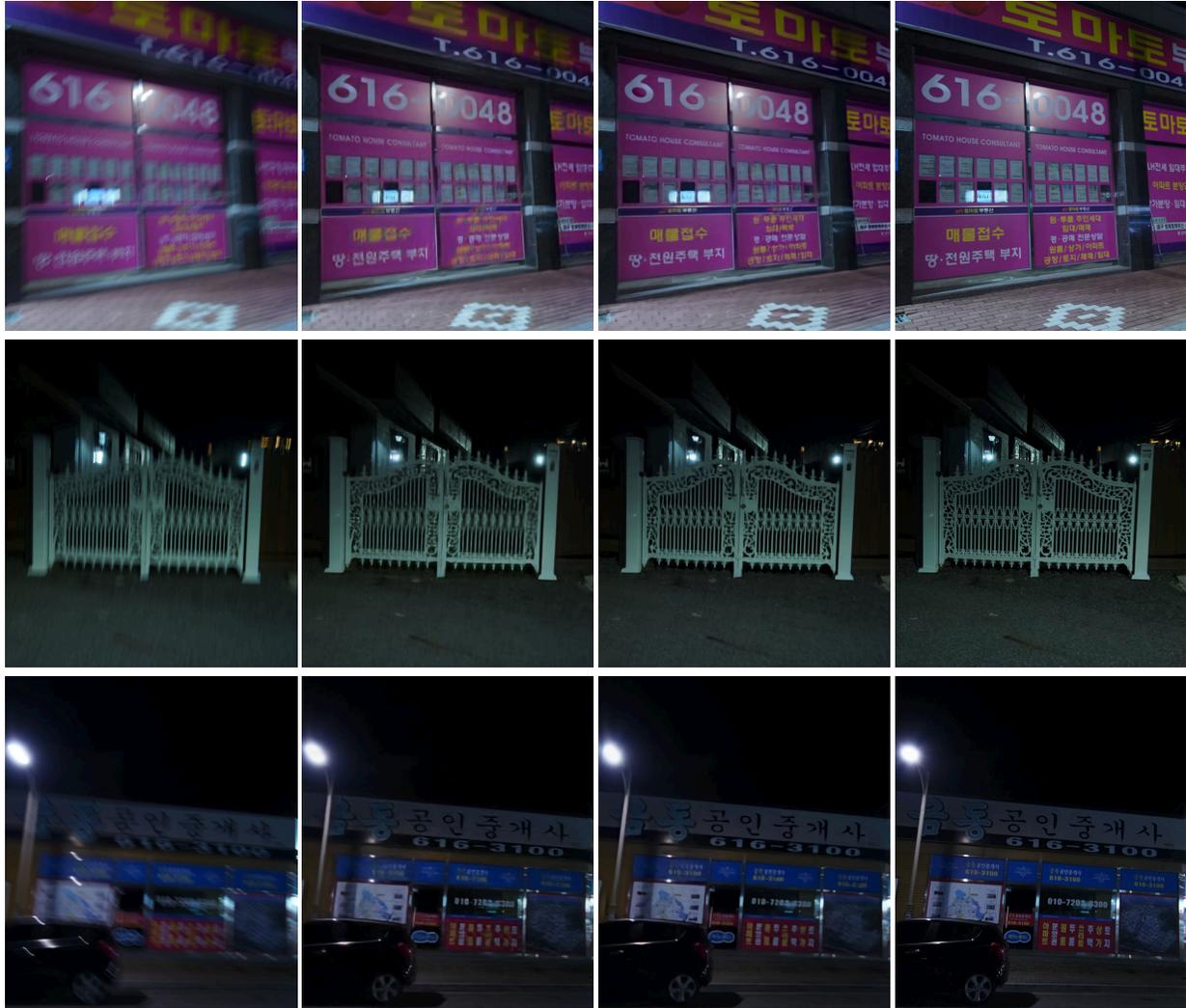


Figure S12. Comparison of shadow removal results on ISTD Dataset [19]. Shadow region and boundaries are visible in existing approaches.

- analysis and machine intelligence*, 35(12):2956–2967, 2012. 10
- [7] Xiaowei Hu, Lei Zhu, Chi-Wing Fu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7454–7462, 2018. 2, 10
- [8] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 6, 7, 10
- [9] Kui Jiang, Zhongyuan Wang, Peng Yi, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. In *CVPR*, 2020. 2, 3, 7
- [10] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *ECCV*, 2018. 7
- [11] Yu Li, Robby T Tan, Xiaojie Guo, Jiangbo Lu, and Michael S Brown. Rain streak removal using layer priors. In *CVPR*, 2016. 2
- [12] Xing Liu, Masanori Suganuma, Zhun Sun, and Takayuki Okatani. Dual residual networks leveraging the potential of paired operations for image restoration. In *Proc. Conference on Computer Vision and Pattern Recognition*, pages 7007–7016, 2019. 6, 8, 9
- [13] Rui Qian, Robby T Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for rain-drop removal from a single image. In *CVPR*, 2018. 2, 6, 7, 8, 9, 10
- [14] Yuhui Quan, Shijie Deng, Yixin Chen, and Hui Ji. Deep learning for seeing through window with raindrops. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2463–2471, 2019. 6, 7, 10
- [15] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *CVPR*, 2019. 7
- [16] Jaesung Rim, Haeyun Lee, Jucheol Won, and Sunghyun Cho. Real-world blur dataset for learning and benchmarking deblurring algorithms. In *ECCV*, 2020. 4, 11
- [17] Masanori Suganuma, Xing Liu, and Takayuki Okatani. Attention-based adaptive selection of operations for image restoration in the presence of unknown combined distortions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9039–9048, 2019. 2, 3, 6
- [18] Hong Wang, Qi Xie, Qian Zhao, and Deyu Meng. A model-driven deep neural network for single image rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3103–3112, 2020. 6
- [19] Jifeng Wang, Xiang Li, and Jian Yang. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1788–1797, 2018. 10
- [20] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson WH Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 12270–12279, 2019. 6
- [21] Qingxiong Yang, Kar-Han Tan, and Narendra Ahuja. Shadow removal using bilateral filtering. *IEEE Transactions on Image processing*, 21(10):4361–4368, 2012. 10
- [22] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *CVPR*, 2017. 2, 3
- [23] Rajeev Yasarla and Vishal M Patel. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In *CVPR*, 2019. 7



(a) Input

(b) [16]

(c) SPAIR

(d) GT

Figure S13. Visual comparisons on real-world blurred images from the RealBlurJ dataset [16] (corresponding to Table 2 of main paper).

- [24] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *CVPR*, 2018. 2, 3
- [25] He Zhang, Vishwanath Sindagi, and Vishal M Patel. Image de-raining using a conditional generative adversarial network. *TCSVT*, 2019. 2, 3



Figure S14. Visual comparison for deblurring on images from GoPro test-set. The figure shows the full sized images along with zoomed-in patches corresponding to the Blurred image, results of DeblurGANv2, MS-CNN, DeblurGAN, SRN, Stack(4)-DMPHN, Our Result, and Ground-truth, respectively.

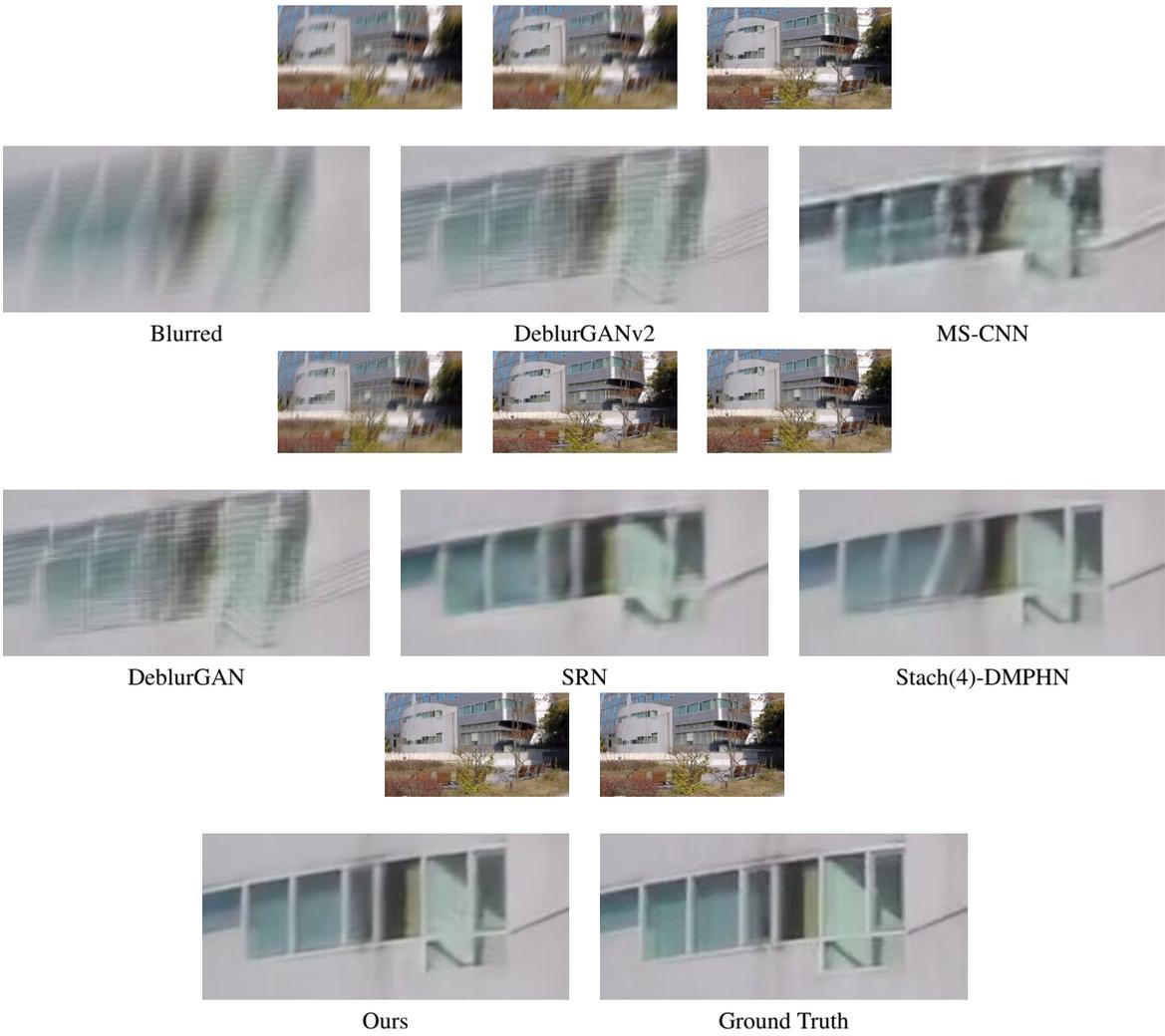


Figure S15. Visual comparison for deblurring on images from GoPro test-set. The figure shows the full sized images along with zoomed-in patches corresponding to the Blurred image, results of DeblurGANv2, MS-CNN, DeblurGAN, SRN, Stack(4)-DMPHN, Our Result, and Ground-truth, respectively.

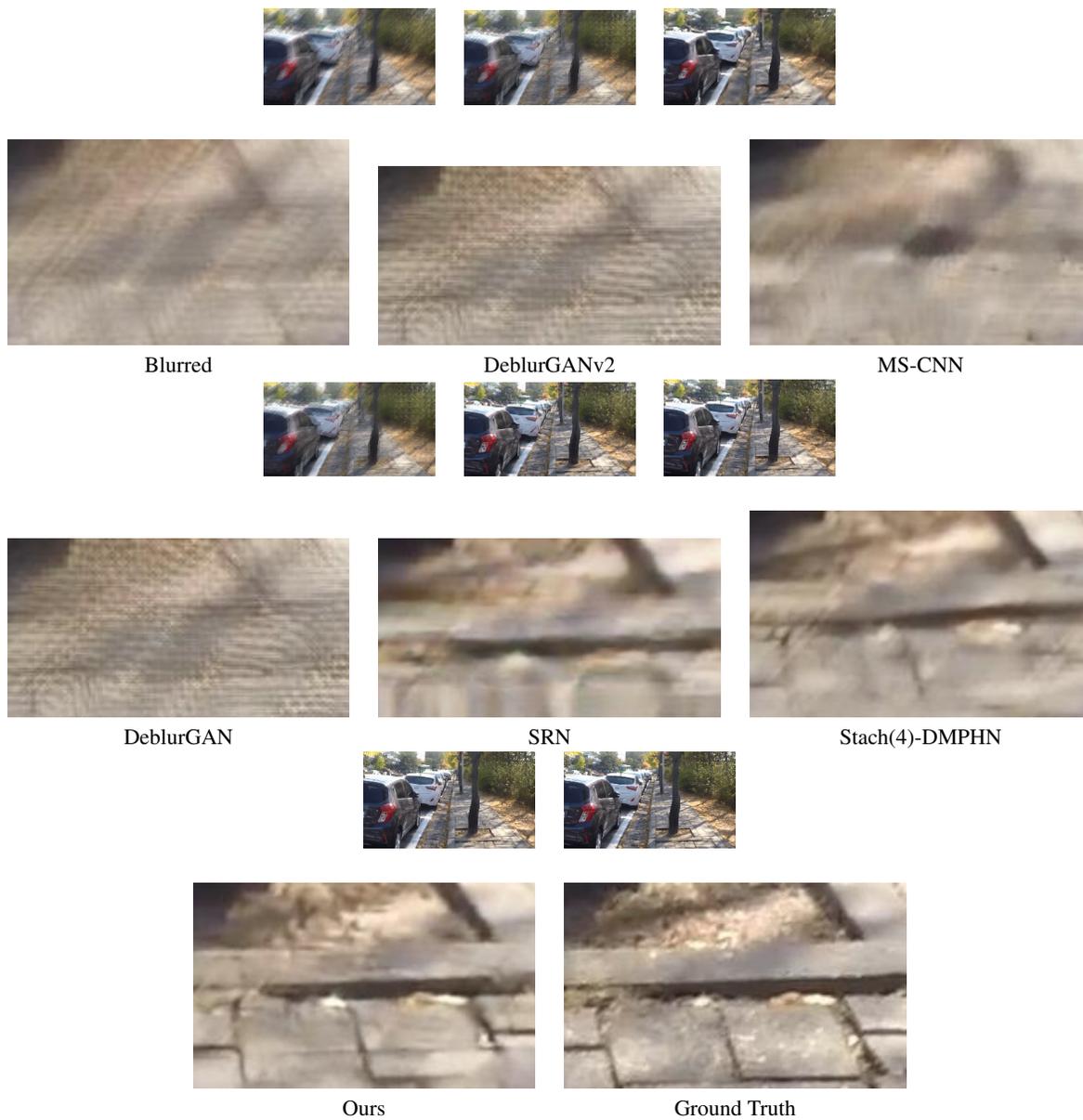


Figure S16. Visual comparison for deblurring on images from GoPro test-set. The figure shows the full sized images along with zoomed-in patches corresponding to the Blurred image, results of DeblurGANv2, MS-CNN, DeblurGAN, SRN, Stack(4)-DMPHN, Our Result, and Ground-truth, respectively.

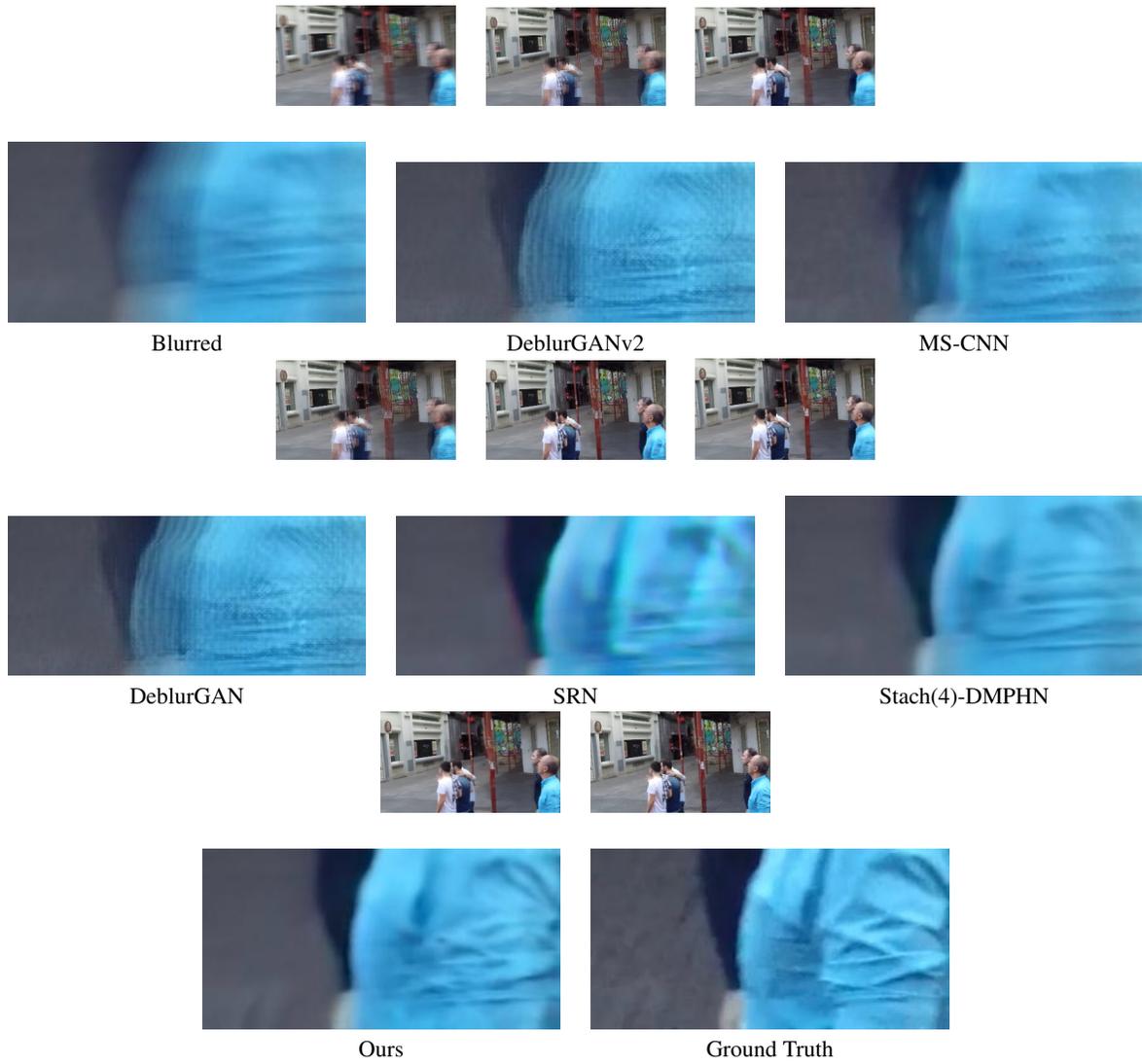


Figure S17. Visual comparison for deblurring on images from GoPro test-set. The figure shows the full sized images along with zoomed-in patches corresponding to the Blurred image, results of DeblurGANv2, MS-CNN, DeblurGAN, SRN, Stack(4)-DMPHN, Our Result, and Ground-truth, respectively.

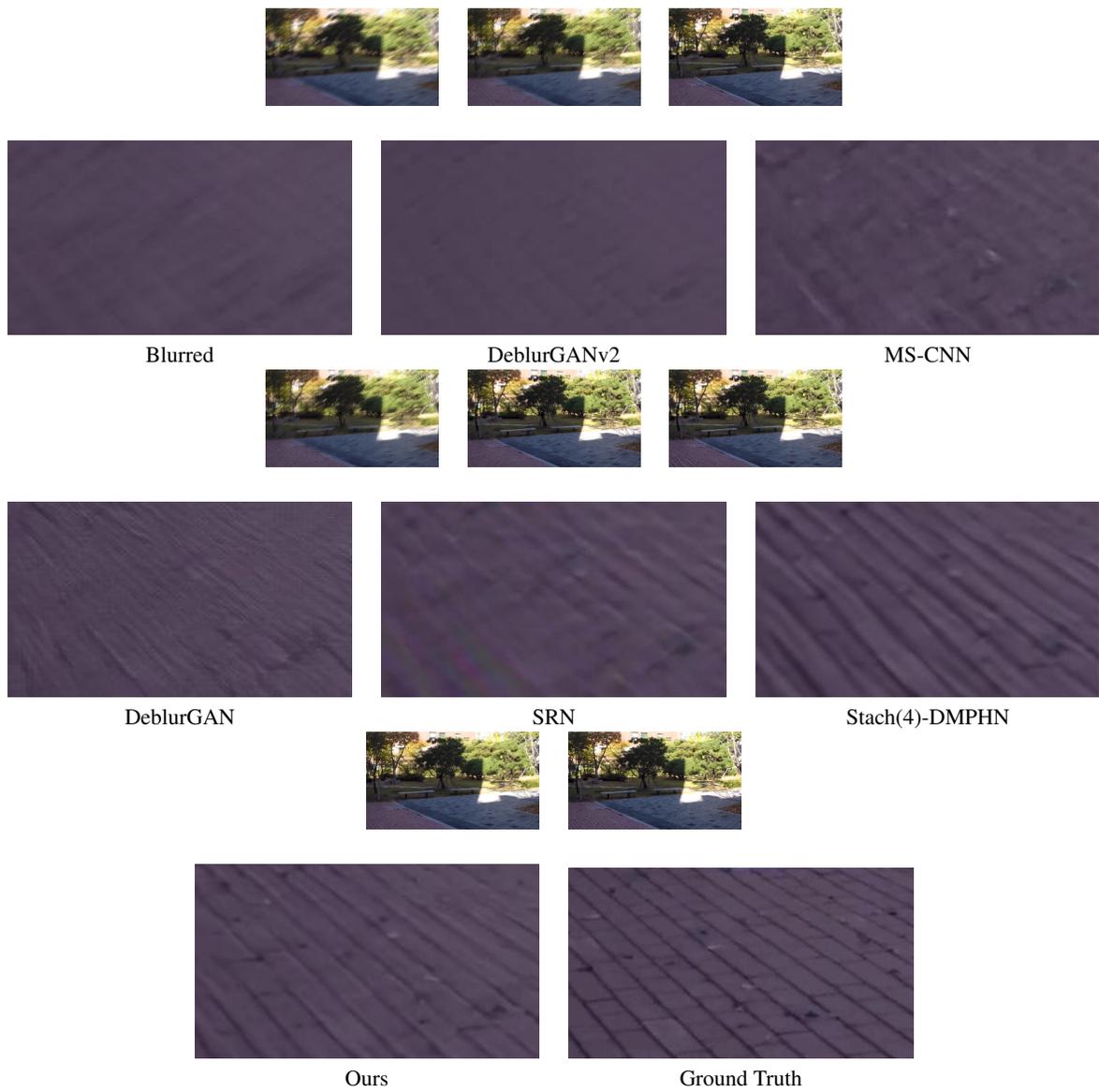


Figure S18. Visual comparison for deblurring on images from GoPro test-set. The figure shows the full sized images along with zoomed-in patches corresponding to the Blurred image, results of DeblurGANv2, MS-CNN, DeblurGAN, SRN, Stack(4)-DMPHN, Our Result, and Ground-truth, respectively.