Bidirectional Multi-Scale Implicit Neural Representations for Image Deraining - Supplemental Material -

Xiang Chen Jinshan Pan* Jiangxin Dong
Nanjing University of Science and Technology

Overview

In this document, we first visualize the rain effect at different scales in Section 1. Then, we present the dataset description, network architecture of UNet, and loss function in Sections 2-4. In Section 5, we conduct the comparison between INR and other low-pass filters. Finally, we show more visual comparisons on both synthetic and real-world images in Section 6.

1. Visualization of Rain Effect at Different Image Scales

We visualize the rain effect at different image scales in Figure 1. We observe that the rain effect decreases significantly at coarser image scales, which also naturally enables the emergence of underlying clean representations in some coarser scale of the rainy image. This motivates us to explore the multi-scale representation to facilitate the rain removal. Considering that the rain effect varies at different image scales, we propose multiple unequal Transformer branches to handle spatially-varying rain streaks by incorporating the networks at finer scales with deeper architectures.

2. Dataset Description

To evaluate the effectiveness of our method in complex rain scenes, we adopt multiple benchmark datasets. Rain200L and Rain200H [14] datasets contain 1,800 synthetic rainy images for training and 200 ones for testing. DID-Data [18] and DDN-Data [7] consist of 12,000 and 12,600 synthetic images with different rain directions and density levels. There are 1,200 and 1,400 rainy images for testing. SPA-Data [11] is the paired real-world dataset which utilizes the human-supervised percentile video filtering to obtain the ground turths. It contains 638,492 rainy/clear image patches for training and 1,000 testing ones. GT-RAIN [1] is the recent large-scale dataset with real paired data captured diverse rain effects. It contains 31,524 rainy and clear frame pairings, which are divided into 26,124 training frames, 3,300 validation frames, and 2,100 test frames. The scenes also include varying degrees of illumination from different times of day and rain of varying densities, streak lengths, shapes, and sizes. RE-RAIN [5] contains 300 real rainy images without ground truths which are elaborately selected from the Internet and related works [9]. The detailed descriptions of the used training and testing datasets are tabulated in Table 1.

Table 1. Descriptions of different image deraining benchmark datasets.

Dataset	Rain200L [14]	Rain200H [14]	DID-Data [18]	DDN-Data [7]	SPA-Data [11]	GT-RAIN [1]	RE-RAIN [5]
Train Samples	1,800	1,800	12,000	12,600	638,492	26,124	0
Test Samples	200	200	1,200	1,400	1,000	2,100	300
Syn/Real	Syn	Syn	Syn	Syn	Real	Real	Real

3. Network Architecture of UNet

Each UNet comprises a sequence of Transformer blocks [16] that calculate self-attention across channels, rather than the spatial dimension, in order to reduce time and memory complexity. The number of Transformer blocks $\{N_1, N_2, N_3\}$ are set to $\{2, 3, 3\}$. The expansion ratio of the number of feature channels is set to 2. We also add skip-connections to bridge across continuous intermediate features for stable training. We show the network architecture of UNet in Figure 2.

*Corresponding author.

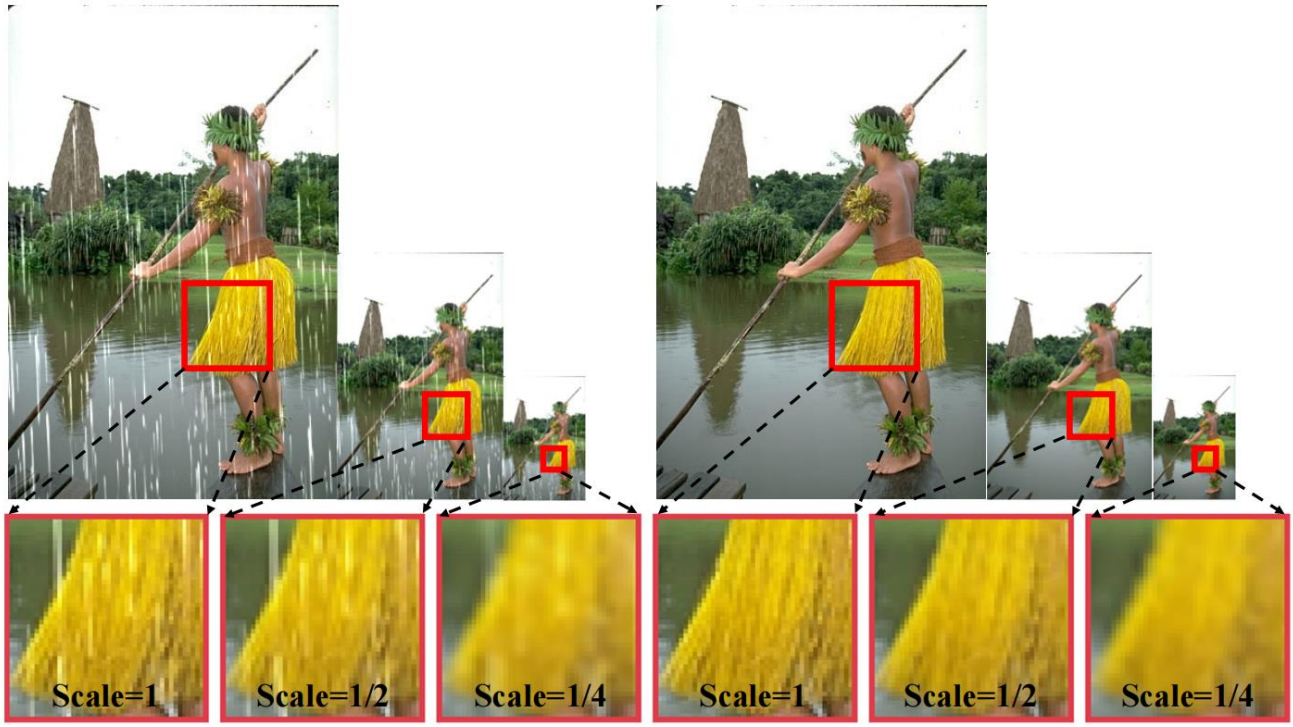


Figure 1. A rainy/rain-free image and its coarser versions at $1/2$ and $1/4$ image scales. Note that the rain effect decreases significantly at coarser image scales.

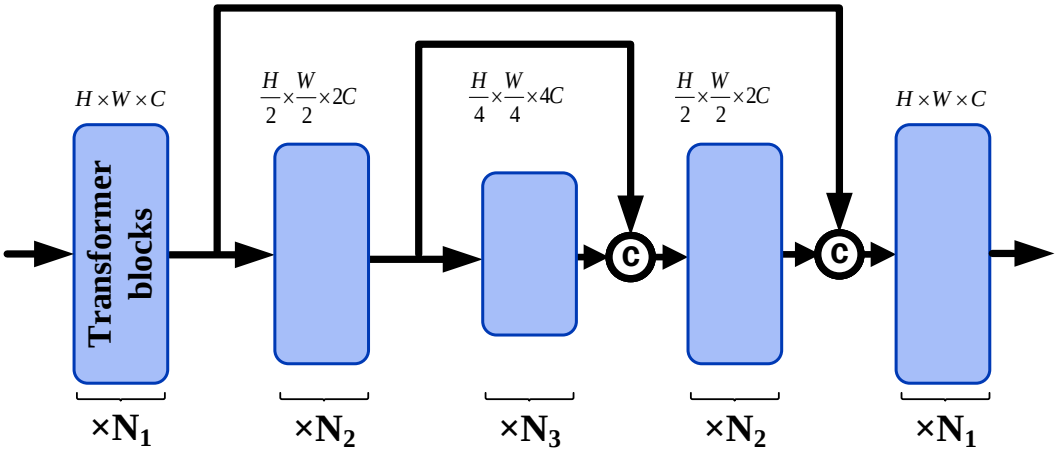


Figure 2. Network architecture of UNet.

4. Loss Function

In the main manuscript, we introduce the INR-related loss. In this document, we introduce three additional loss functions. We utilize Charbonnier [17], frequency and edge [6] loss to constrain the network training. The Charbonnier loss is defined as follows:

$$\mathcal{L}_{char} = \sum_{s=1}^{3} \sqrt{\|\mathbf{B}_s - \mathbf{T}_s\|^2 + \varepsilon^2}, \tag{1}$$

where $\mathbf{B}_s$ and $\mathbf{T}_s$ denote the s-th scale reconstructed image and ground-truth image, respectively. Here, the penalty coefficient ε is set to 10^{-3}. The frequency loss is defined as follows:

$$\mathcal{L}_{freq} = \sum_{s=1}^{3} \left\| \mathcal{FT}\left(\mathbf{B}_s\right) - \mathcal{FT}(\mathbf{T}_s) \right\|_1, \tag{2}$$

where $\mathcal{FT}$ represents the Fourier transform operator to obtain the frequency domain of the original image. The edge loss is defined as follows:

$$\mathcal{L}_{edge} = \sum_{s=1}^{3} \sqrt{\left\| \Delta\left(\mathbf{B}_s\right) - \Delta\left(\mathbf{T}_s\right) \right\|^2 + \varepsilon^2}, \tag{3}$$

where Δ represents the Laplacian operator.

5. Comparison with Low-pass Filters

As mentioned in the main manuscript, recent studies point out the low-pass filtering characteristics in the implicit neural representation (INR). We attribute the deraining ability of INR to a basic fact that the intensity values of rain-affected pixels tend to surpass those of their neighboring non-rain pixels. Similar to [2], we replace INR with some classical low-pass filters including uniform and median filters. We visualize some comparison results in Figure 3. Obviously, other methods are not able to effectively remove rain streaks in complex scenes and are prone to losing high-frequency image details. Thanks to the combined effect of implicit interpolation and spatial encoding operations, our proposed INR branch generates clearer results with better detail and texture recovery.

(a) Rainy input (b) Uniform filter (c) Median filter (d) INR

Figure 3. Visual comparisons with other low-pass filters.

6. More Experimental Results

In this section, we show more experimental results to demonstrate the effectiveness of our method. Firstly, we evaluate our approach on recent GT-RAIN dataset [1]. Note that we retrain the learning-based methods that are not trained on this dataset using the same protocols for fairness. Table 2 shows that our method performs well on the challenging GT-RAIN dataset.

Table 2. Quantitative evaluations of the proposed approach against state-of-the-art image deraining methods on the GT-RAIN dataset.

Methods		RCDNet [10]	SPDNet [15]	Uformer [12]	Restormer [16]	IDT [13]	DRSformer [4]	NeRD-Rain
GT-RAIN [1]	PSNR	22.61	22.98	22.74	23.38	22.51	22.56	**23.50**
	SSIM	0.6363	0.6387	0.6389	0.6401	**0.6502**	0.6290	0.6459

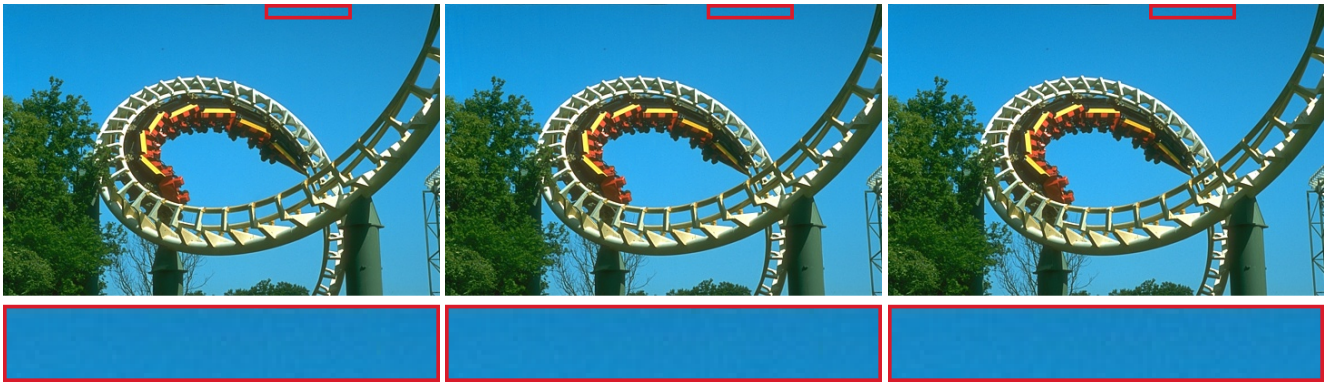
We further compare with more recent Transformer-based image deraining approach, UDR-S2Former [3]. Table 3 reports quantitative results trained on the Rain200L benchmark [14]. Our method still achieves the highest PSNR and SSIM values.

Table 3. Quantitative evaluations of the proposed approach against state-of-the-art image deraining methods on the Rain200L dataset.

Methods		SPDNet [15]	Uformer [12]	Restormer [16]	IDT [13]	DRSformer [4]	UDR-S2Former [3]	NeRD-Rain
Rain200L [14]	PSNR	40.50	40.20	40.99	40.74	41.23	40.96	**41.71**
	SSIM	0.987	0.986	0.989	0.988	0.989	0.989	**0.990**

(a) Rainy input (b) DualGCN [8] (c) SPDNet [15]

(d) Restormer [16] (e) IDT [13] (f) DRSformer [4]



(g) NeRD-Rain-S (Ours-S) (h) NeRD-Rain (Ours) (i) Ground truth

Figure 4. Visual comparison results on the Rain200L dataset [14]. The results shown in (b)-(f) still contain significant white rain streaks. In contrast, our models generate much clearer images.

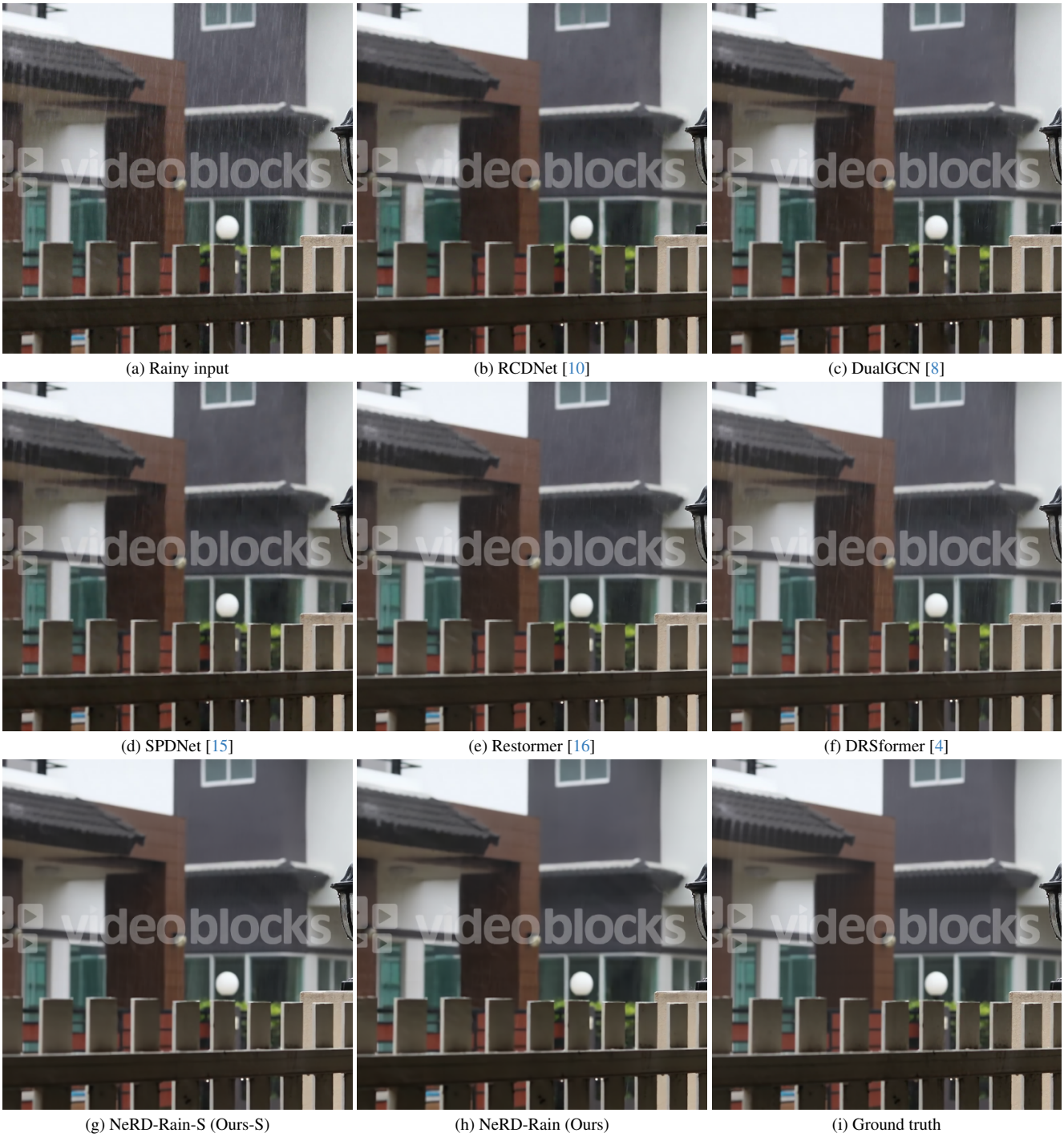
Finally, Figures 4-5 show the visual comparison results on the synthetic datasets (*i.e.*, Rain200L [14] and Rain200H [14]). Compared to other methods, our NeRD-Rain can generate high-quality deraining results with more accurate detail and texture recovery. Figures 6-9 also present the visual comparison results on the real-world datasets (*e.g.*, SPA-Data [11] and RE-RAIN [5]). Our method can successfully remove complex and random rain streaks and own visual pleasant recovery results.

(a) Rainy input (b) RCDNet [10] (c) DualGCN [8]

(d) SPDNet [15] (e) Restormer [16] (f) IDT [13]

(g) DRSformer [4] (h) NeRD-Rain (Ours) (i) Ground truth

Figure 5. Visual comparison results on the Rain200H dataset [14]. The evaluated approaches do not generate clear images, where some structural details are not restored well. In contrast, our method generates a better-derained image with finer structural details.

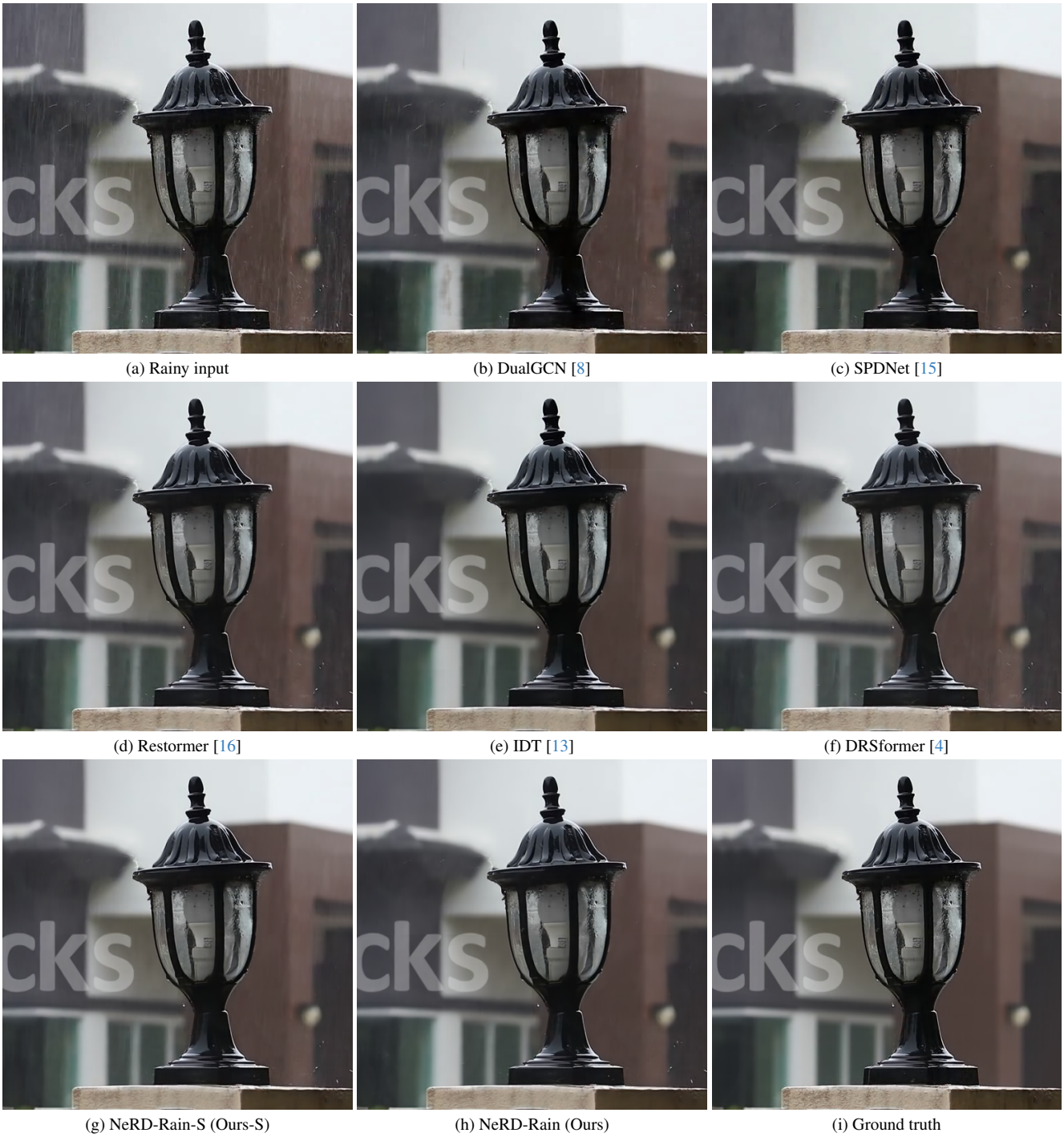


(a) Rainy input (b) RCDNet [10] (c) DualGCN [8]

(d) SPDNet [15] (e) Restormer [16] (f) DRSformer [4]

(g) NeRD-Rain-S (Ours-S) (h) NeRD-Rain (Ours) (i) Ground truth

Figure 6. Visual comparison results on the SPA-Data dataset [11]. The results shown in (b)-(f) still contain significant rain streaks. In contrast, our models generate much clearer images.



(a) Rainy input (b) DualGCN [8] (c) SPDNet [15]

(d) Restormer [16] (e) IDT [13] (f) DRSformer [4]

(g) NeRD-Rain-S (Ours-S) (h) NeRD-Rain (Ours) (i) Ground truth

Figure 7. Visual comparison results on the SPA-Data dataset [11]. The results shown in (b)-(f) still contain significant rain streaks. In contrast, our models generate much clearer images.

(a) Rainy input

(b) Restormer [16]

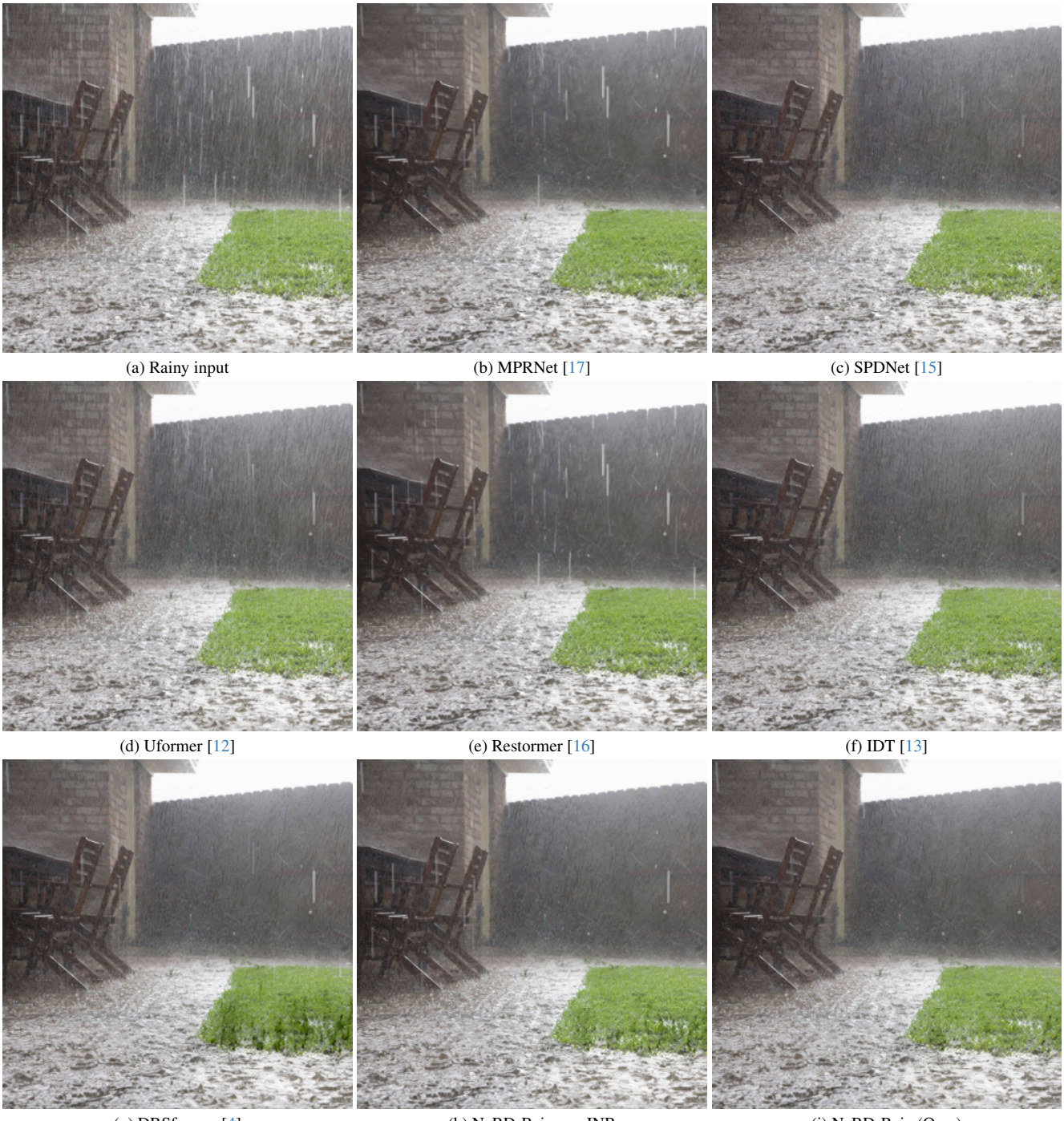
(c) IDT [13]

(d) DRSformer [4]

(e) NeRD-Rain$_{w/o}$ INR

(f) NeRD-Rain (Ours)

Figure 8. Visual comparison results on the RE-RAIN dataset [5]. The results shown in (b)-(d) still contain significant rain streaks. The approach (e) without INR is sensitive to complex rain streaks. In contrast, our method generates a much clearer image.



(a) Rainy input (b) MPRNet [17] (c) SPDNet [15]

(d) Uformer [12] (e) Restormer [16] (f) IDT [13]

(g) DRSformer [4] (h) NeRD-Rain$_{w/o}$ INR (i) NeRD-Rain (Ours)

Figure 9. Visual comparison results on the RE-RAIN dataset [5]. The results shown in (b)-(g) still contain significant rain streaks. The approach (h) without INR is sensitive to complex rain streaks. In contrast, our method generates a much clearer image.

References

[1] Yunhao Ba, Howard Zhang, Ethan Yang, Akira Suzuki, Arnold Pfahnl, Chethan Chinder Chandrappa, Celso M de Melo, Suya You, Stefano Soatto, Alex Wong, et al. Not just streaks: Towards ground truth for single image deraining. In *ECCV*, pages 723–740, 2022. 1, 3

[2] Hao Chen, Bo He, Hanyu Wang, Yixuan Ren, Ser Nam Lim, and Abhinav Shrivastava. Nerv: Neural representations for videos. *NeurIPS*, 34:21557–21568, 2021. 3

[3] Sixiang Chen, Tian Ye, Jinbin Bai, Erkang Chen, Jun Shi, and Lei Zhu. Sparse sampling transformer with uncertainty-driven ranking for unified removal of raindrops and rain streaks. In *ICCV*, pages 13106–13117, 2023. 3, 4

[4] Xiang Chen, Hao Li, Mingqiang Li, and Jinshan Pan. Learning a sparse transformer network for effective image deraining. In *CVPR*, pages 5896–5905, 2023. 3, 4, 5, 6, 7, 8, 9

[5] Xiang Chen, Jinshan Pan, Jiangxin Dong, and Jinhui Tang. Towards unified deep image deraining: A survey and a new benchmark. *arXiv preprint arXiv:2310.03535*, 2023. 1, 4, 8, 9

[6] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *ICCV*, pages 4641–4650, 2021. 2

[7] 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*, pages 3855–3863, 2017. 1

[8] Xueyang Fu, Qi Qi, Zheng-Jun Zha, Yurui Zhu, and Xinghao Ding. Rain streak removal via dual graph convolutional network. In *AAAI*, volume 35, pages 1352–1360, 2021. 4, 5, 6, 7

[9] Yang Liu, Ziyu Yue, Jinshan Pan, and Zhixun Su. Unpaired learning for deep image deraining with rain direction regularizer. In *ICCV*, pages 4753–4761, 2021. 1

[10] Hong Wang, Qi Xie, Qian Zhao, and Deyu Meng. A model-driven deep neural network for single image rain removal. In *CVPR*, pages 3103–3112, 2020. 3, 5, 6

[11] 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 *CVPR*, pages 12270–12279, 2019. 1, 4, 6, 7

[12] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *CVPR*, pages 17683–17693, 2022. 3, 4, 9

[13] Jie Xiao, Xueyang Fu, Aiping Liu, Feng Wu, and Zheng-Jun Zha. Image de-raining transformer. *IEEE TPAMI*, 2022. 3, 4, 5, 7, 8, 9

[14] 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*, pages 1357–1366, 2017. 1, 3, 4, 5

[15] Qiaosi Yi, Juncheng Li, Qinyan Dai, Faming Fang, Guixu Zhang, and Tieyong Zeng. Structure-preserving deraining with residue channel prior guidance. In *ICCV*, pages 4238–4247, 2021. 3, 4, 5, 6, 7, 9

[16] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, pages 5728–5739, 2022. 1, 3, 4, 5, 6, 7, 8, 9

[17] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, pages 14821–14831, 2021. 2, 9

[18] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *CVPR*, pages 695–704, 2018. 1