Dreaming to Prune Image Deraining Networks Supplementary Material

Weiqi Zou[1,*] Yang Wang[1,*] Xueyang Fu[1] Yang Cao[1,2,†]
[1] University of Science and Technology of China
[2] Institute of Artificial Intelligence, Hefei Comprehensive National Science Center
artisan@mail.ustc.edu.cn, {ywang120, xyfu, forrest}@ustc.edu.cn

In this supplementary material, we provide more detailed information to further explain our proposed method. **First**, we introduce the details about our framework and hyper-parameters. **Next**, we present the intermediate results throughout our optimizing iterations to facilitate illustrating the reasons why our method works, i.e., why we can preserve the deraining capabilities of the compressed model without original data. **Then**, we display more experimental results to demonstrate the superiority of our method. **Finally**, we extend our framework to image dehazing tasks, and we are looking forward to bringing more inspiration for future studies.

1. Implementation Details

1.1. Framework Details

For a fair comparison, we follow the official open source implementation of the state-of-the-art methods and adopt their released pre-trained models, including HINet [1][1], MPRNet [17][2], AGANet [11][3], DuRNet [10][4]. For each model, we hook the deep features from the penultimate layer to calculate the batch diversity loss ($\mathcal{L}_{orth}$) mentioned in the paper. And we implement the similarity measure for the network outputs utilizing the most commonly used *L1* regularization. We collect clean images by random sampling in the Place365 dataset [20] which contains a rich set of scenes. It is worth clarifying that these clean images do not appear in the training set or test set of the pre-trained networks.

1.2. Hyper-parameters

We jointly optimize random input noise to images and distill the pruned model, the learning rates of these two branches are set to $5 \cdot e^{-2}$ and $1 \cdot e^{-4}$, respectively. Their magnitude differences are attributed to the different value levels of image pixels and model weights. The distillation loss factor (λ_{KD}) is set to 1.0, which is required relatively loosely. We set the batch diversity loss factor (λ_{orth}) to 0.05 and explore its sensitivity below.

2. Optimization Illustration

We display the intermediate results of our framework throughout the convergence process. As shown in Fig. 1, the performance of the pruned model (HINet [1]) drops without fine-tuning. To preserve the performance of the pruned model, we jointly reconstruct diverse and in-distribution degraded images using dream loss ($\mathcal{L}_{dream}$) and leverage these images to distill the pruned model using distillation loss ($\mathcal{L}_{KD}$). As shown in Fig. 1, with the total loss converges, on the one hand, we achieve to reconstruct diverse rainy images which resemble the original distribution, on the other hand, we distill the pruned model to refine its deraining performance. Thus, we can compress the pre-trained deraining model while preserving its capabilities of handling various degradation characteristics.

*Equal contribution. † Corresponding author.
[1] https://github.com/megvii-model/HINet
[2] https://github.com/swz30/MPRNet
[3] https://github.com/rui1996/DeRaindrop
[4] https://github.com/liu-vis/DualResidualNetworks

		Test2800 [4]		Test1200 [18]		Test100 [19]		Rain100H [15]		Rain100L [15]		Average	
Deraining Models		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DerainNet [3]		24.31	0.861	23.38	0.835	22.77	0.810	14.92	0.592	27.03	0.884	22.48	0.796
SEMI [14]		24.43	0.782	26.05	0.822	22.35	0.788	16.56	0.486	25.03	0.842	22.88	0.744
DIDMDN [18]		28.13	0.867	29.65	0.901	22.56	0.818	17.35	0.524	25.23	0.741	24.58	0.770
UMRL [16]		29.97	0.905	30.55	0.910	24.41	0.829	26.01	0.832	29.18	0.923	28.02	0.880
RESCAN [9]		31.29	0.904	30.51	0.882	25.00	0.835	26.36	0.786	29.80	0.881	28.59	0.857
PReNet [13]		31.75	0.916	31.36	0.911	24.81	0.851	26.77	0.858	32.44	0.950	29.42	0.897
MSPFN [6]		32.82	0.930	32.39	0.916	27.50	0.876	28.66	0.860	32.40	0.933	30.75	0.903
MPRNet [17]	$\ell 1$ [5] (-37.7%)	27.81	0.844	26.32	0.781	24.41	0.782	16.07	0.459	27.72	0.831	24.47	0.739
	erk [2] (-37.7%)	32.84	0.931	31.82	0.888	28.44	0.868	27.76	0.844	34.32	0.946	31.04	0.895
	lamp [7] (-37.7%)	33.07	0.934	32.38	0.899	29.09	0.879	28.82	0.864	35.59	0.957	31.79	0.907
	Ours (-37.7%)	**33.40**	**0.938**	**32.70**	**0.912**	**30.07**	**0.894**	**29.19**	**0.874**	**36.56**	**0.965**	**32.38**	**0.917**
	original	33.64	0.938	32.91	0.916	30.27	0.897	30.41	0.890	36.40	0.965	32.73	0.921
HINet [1]	$\ell 1$ [5] (-41.3%)	29.34	0.887	27.49	0.831	24.92	0.816	18.66	0.599	29.49	0.888	25.98	0.804
	erk [2] (-41.3%)	32.37	0.929	31.09	0.895	25.30	0.835	23.46	0.783	28.74	0.880	28.19	0.864
	lamp [7] (-41.3%)	33.23	0.936	32.52	0.912	27.58	0.872	27.21	0.862	30.98	0.919	30.30	0.900
	Ours (-41.3%)	**33.79**	**0.940**	**32.95**	**0.919**	**30.12**	**0.906**	**29.54**	**0.890**	**36.94**	**0.969**	**32.67**	**0.925**
	original	33.91	0.941	33.05	0.919	30.29	0.906	30.65	0.894	37.28	0.970	33.03	0.926

Table 1. Data-free pruning results on Test2800 [4], Test1200 [18], Test100 [19], Rain100H [15], and Rain100L [15]. We compress the pre-trained state-of-the-art deraining models HINet [1] and MPRNet [17]. Our method can reduce about 40% of the FLOPs while maintaining comparable performance with these original models, which outperforms other modern pruning methods. And the deraining performance of our pruned models still outperforms many other deraining models by a large margin. Best scores of original model and pruned model are underlined and **highlighted**.

3. More Results

3.1. Quantitative Results on Rain13k

Rain13k includes five test datasets, which are Test2800 [4], Test1200 [18], Test100 [19], Rain100H [15], and Rain100L [15], respectively. We compress state-of-the-art deraining networks HINet [1] and MPRNet [17] on Rain13k, since these two methods outperform other methods by a margin, as shown in Tab. 1. We compress about 40% of the FLOPs while maintaining comparable performance with the state-of-the-art methods in the absence of the original data.

3.2. Qualitative Results

In this section, we compare our proposed method with LAMP [7] on Rain13k and RainDrops [11] datasets, and more results are shown in Fig. 2, Fig. 3, and Fig. 4. For a fair comparison, we ensure that all pruned models have the same FLOPs. We can observe that our proposed method can remove the rain more completely than LAMP, which demonstrates the effectiveness or our proposed method. Furthermore, our method can remove many types of rain, from rain streaks to rain drops in various directions and densities, which demonstrates the generalization of our proposed method.

3.3. Dreamed Images Visualization

We display the reconstructed images with dreaming approach. Our dreamed degraded images based on the deraining model pre-trained on Rain13k and RainDrops datasets, as shown in Fig. 5. It can be seen that with the introduced batch diversity loss $\mathcal{L}_{orth}$, these dreamed images appear various types of rain, including different orientations and densities. In contrast, without $\mathcal{L}_{orth}$, those images tend to appear a similar rain style. This demonstrates that we can improve the diversity of our reconstructed rainy images by increasing the distance between content-agnostic degradation representations.

4. Extension to Image Dehazing

To explore the potential of our data-free compression framework for other low-level tasks, we extend our method to the image dehazing task, as shown in Tab. 2. It can be seen that, we compress about 41.5% FLOPs the pre-trained state-of-the-art dehazing model FFA-Net [12] while maintaining its original performance, which outperforms other modern pruning methods

	$\ell 1$ [5]	erk [2]	lamp [7]	**Ours**	original [12]
FLOPs	168 G				287 G
PSNR	20.93	27.40	28.60	**32.28**	33.57
SSIM	0.861	0.915	0.939	**0.976**	0.984

Table 2. Data-free pruning results on SOTS Outdoor dataset [8]. We compress about 41.5% FLOPs the pre-trained state-of-the-art dehazing model FFA-Net [12] while maintaining its original performance, which outperforms other modern pruning methods by a large margin.

by a large margin. And the qualitative results of our reconstructed hazy images are shown in Fig. 7. With batch diversity loss $\mathcal{L}_{orth}$, we can reconstruct a variety of haze types within a batch. Thus, we can provide sufficient supervision for distilling the pruned model and preserving its dehazing performance without original data. This demonstrates the potential to extend our approach to other low-level tasks.

References

[1] Liangyu Chen, Xin Lu, Jie Zhang, Xiaojie Chu, and Chengpeng Chen. Hinet: Half instance normalization network for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 182–192, June 2021. 1, 2, 6

[2] Utku Evci, Trevor Gale, Jacob Menick, Pablo Samuel Castro, and Erich Elsen. Rigging the lottery: Making all tickets winners. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 2943–2952. PMLR, 13–18 Jul 2020. 2, 3

[3] Xueyang Fu, Jiabin Huang, Xinghao Ding, Yinghao Liao, and John Paisley. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Transactions on Image Processing*, 26(6):2944–2956, 2017. 2

[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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3855–3863, 2017. 2

[5] Song Han, Jeff Pool, John Tran, and William Dally. Learning both weights and connections for efficient neural network. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. 2, 3

[6] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8346–8355, 2020. 2

[7] Jaeho Lee, Sejun Park, Sangwoo Mo, Sungsoo Ahn, and Jinwoo Shin. Layer-adaptive sparsity for the magnitude-based pruning. In *International Conference on Learning Representations*, 2021. 2, 3, 6, 7, 8

[8] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018. 3

[9] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 254–269, 2018. 2

[10] 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. 1

[11] Rui Qian, Robby T. Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for raindrop removal from a single image. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 1, 2, 8, 10

[12] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11908–11915, 2020. 2, 3, 10

[13] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3937–3946, 2019. 2

[14] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. Semi-supervised transfer learning for image rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3877–3886, 2019. 2

[15] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1357–1366, 2017. 2

[16] Rajeev Yasarla and Vishal M Patel. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8405–8414, 2019. 2

[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*, 2021. 1, 2, 7

[18] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 695–704, 2018. 2

[19] He Zhang, Vishwanath Sindagi, and Vishal M Patel. Image de-raining using a conditional generative adversarial network. *IEEE transactions on circuits and systems for video technology*, 30(11):3943–3956, 2019. 2

[20] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017. 1

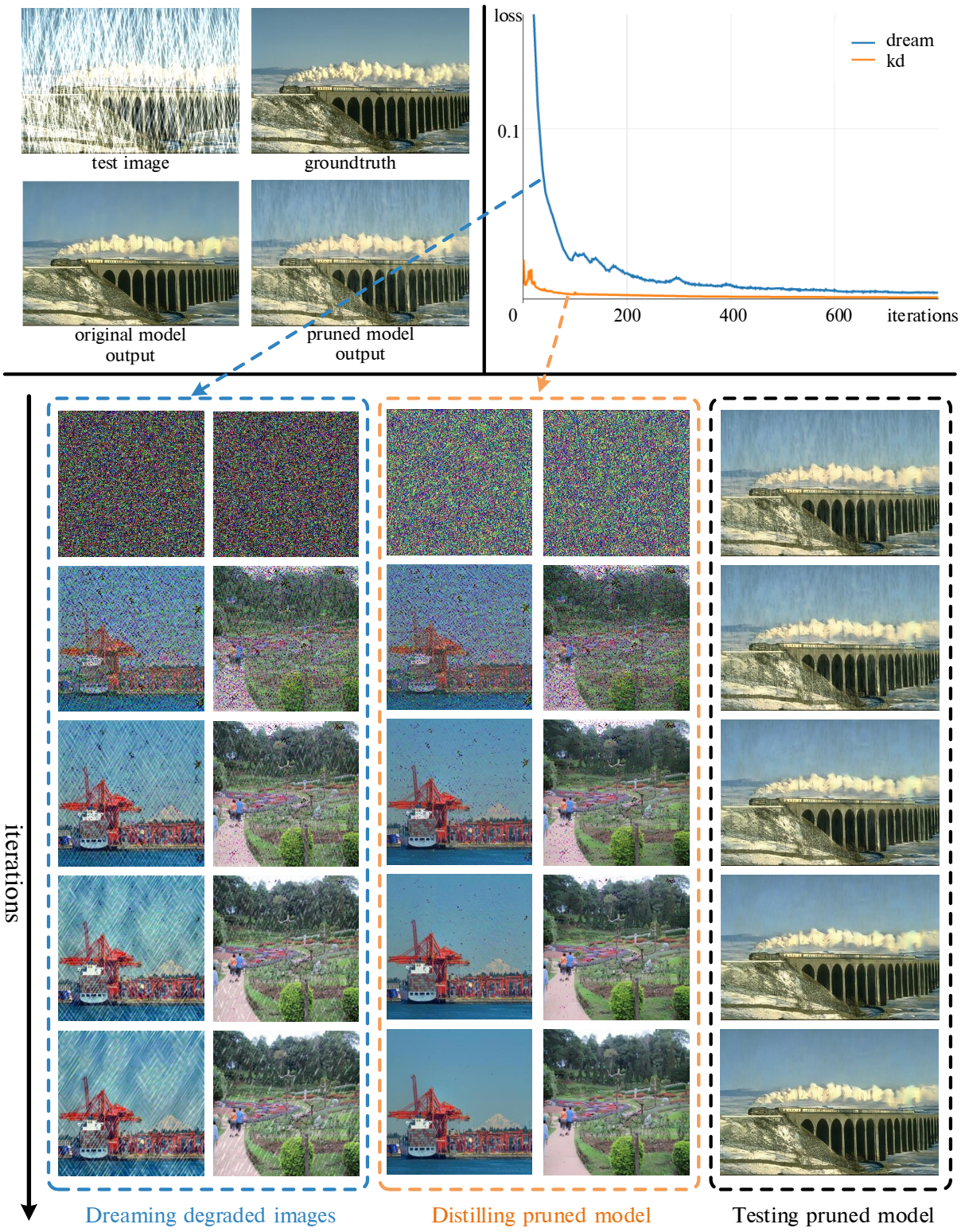
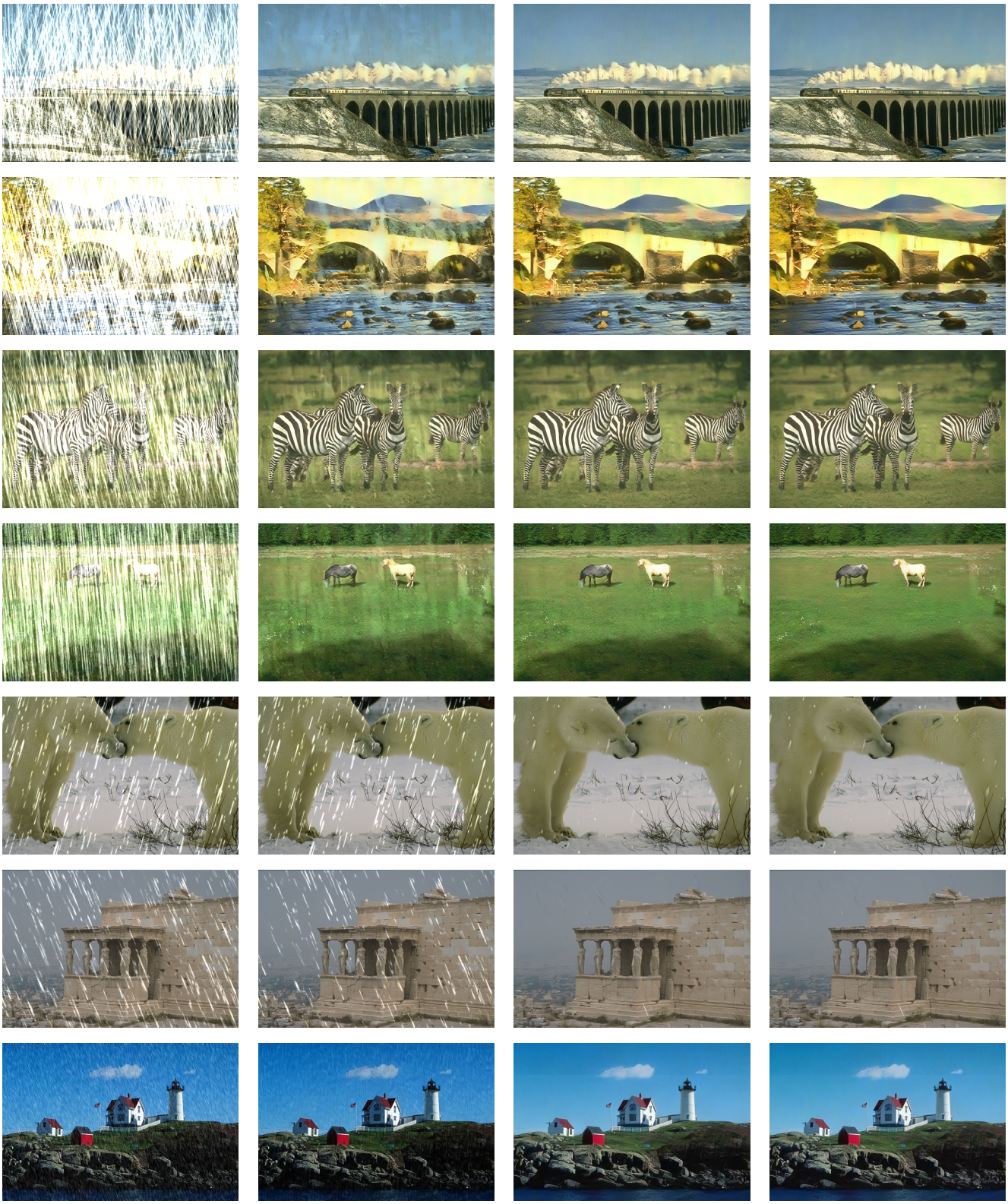
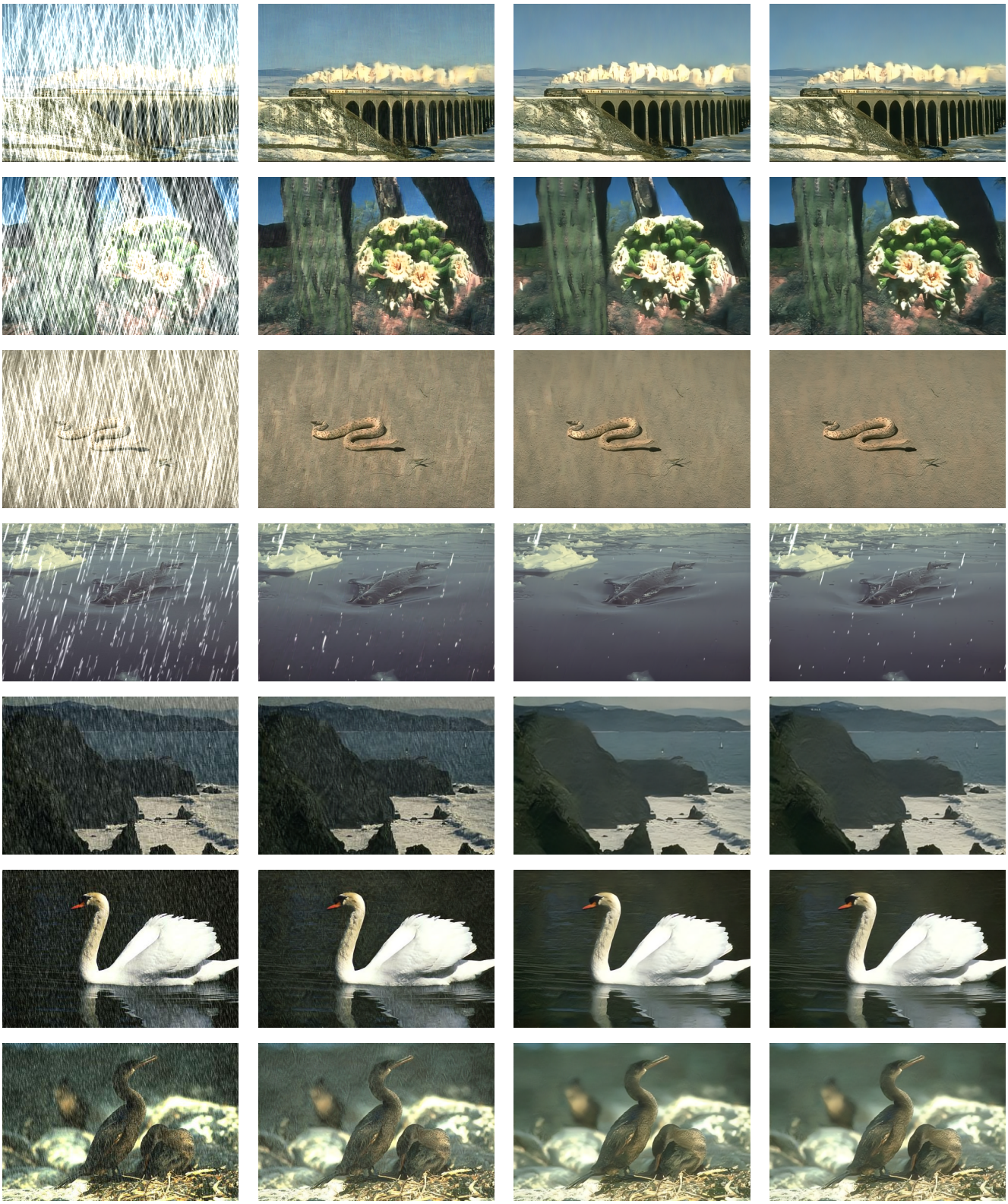


Figure 1. The performance of the deraining model drops after direct pruning. Throughout the optimization process, we jointly optimize random noise to diverse rainy images and distill the pruned model. As the total loss converges, the pruned model gets more and more improved until achieving a comparable performance with the original model. We sample two images within a batch for a better view.



(a) Input (b) LAMP [7] (c) **Ours** (d) original

Figure 2. Qualitative results of pruning image deraining model HINet [1] on Rain13k. Our method preserve the performance of the pruned model on handling various degradation characteristics, and outperforms modern pruning method LAMP [7] by a margin.

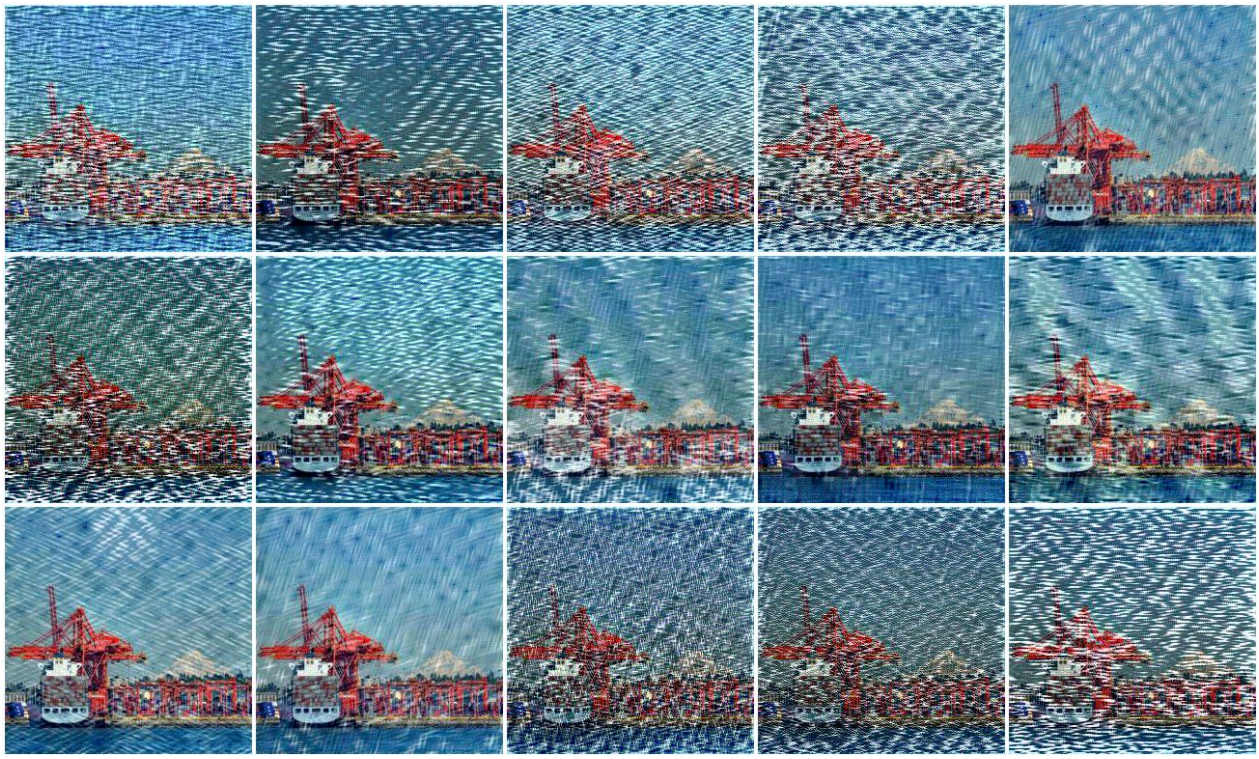


(a) Input (b) LAMP [7] (c) **Ours** (d) original

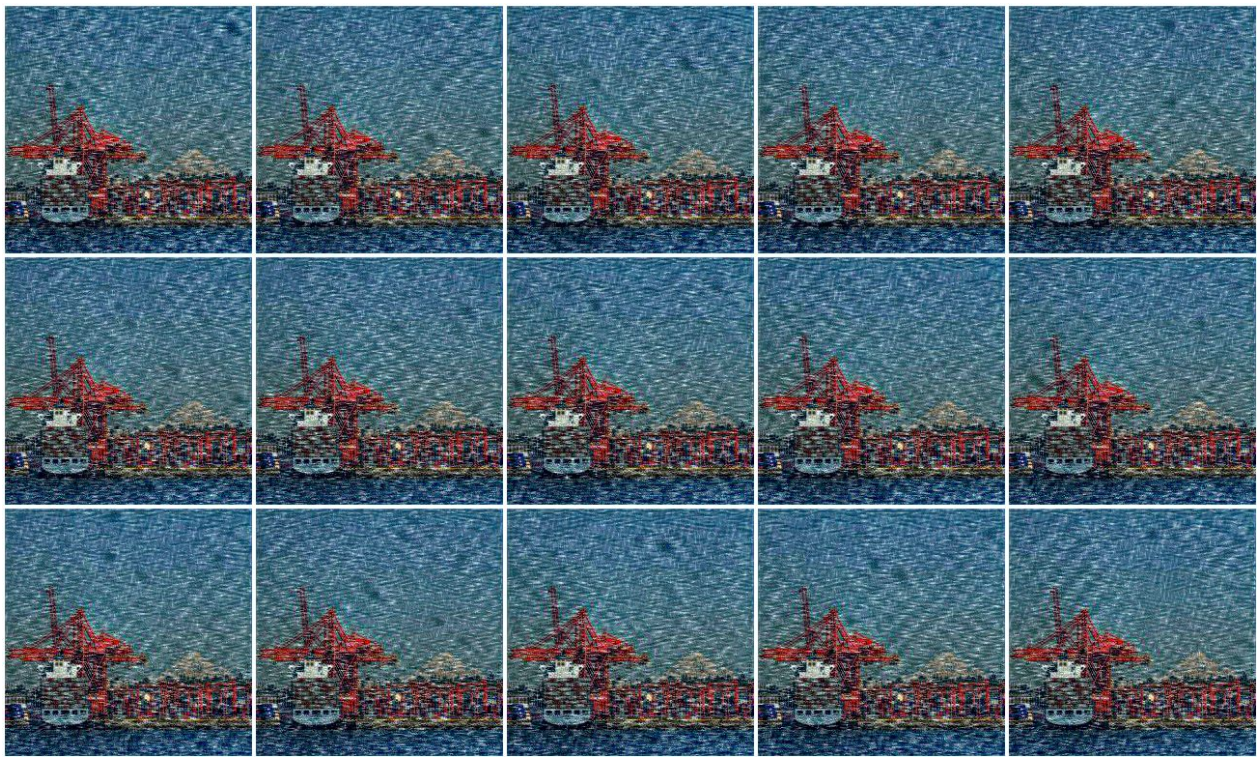
Figure 3. Qualitative results of pruning image deraining model MPRNet [17] on Rain13k. Our method preserve the performance of the pruned model on handling various degradation characteristics, and outperforms modern pruning method LAMP [7].

(a) Input (b) LAMP [7] (c) **Ours** (d) original

Figure 4. Qualitative results of pruning image deraining model AGANet [11] on Raindrops dataset [11]. Our method preserve the performance of the pruned model on handling various degradation characteristics, and outperforms modern pruning method LAMP [7] by a margin.

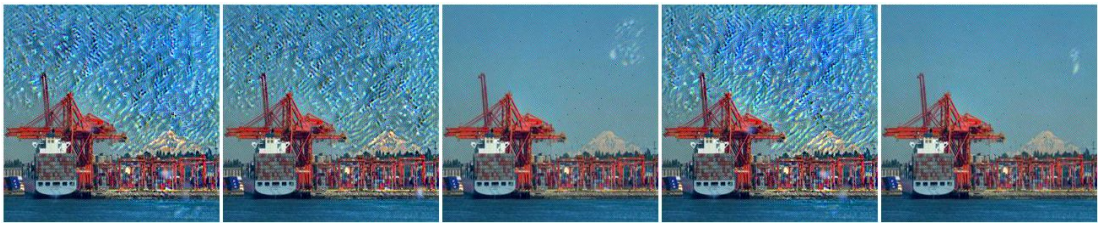


(a) w $\mathcal{L}_{orth}$



(b) w/o $\mathcal{L}_{orth}$

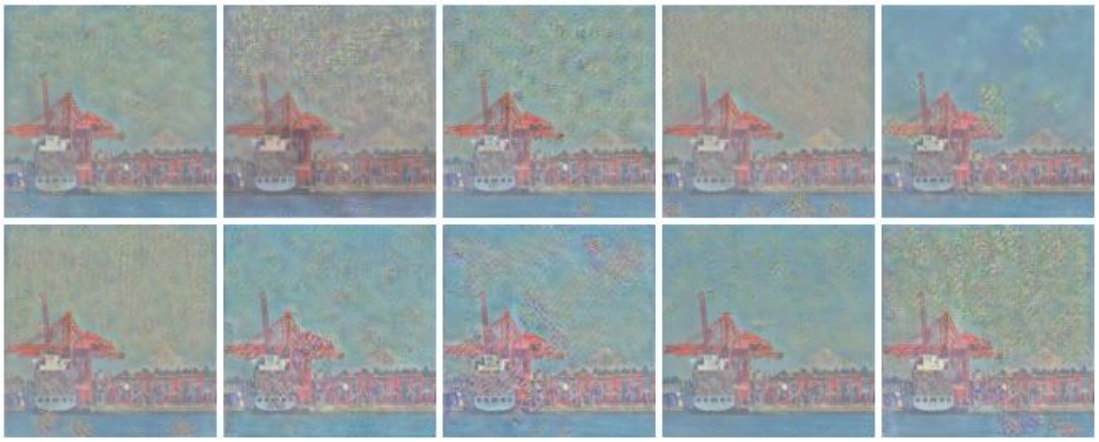
Figure 5. Our dreamed degraded images by inverting the deraining model pre-trained on Rain13k datasets. With batch diversity loss $\mathcal{L}_{orth}$, we can reconstruct a variety of rain types within a batch.



(a) w $\mathcal{L}_{orth}$

(b) w/o $\mathcal{L}_{orth}$

Figure 6. Our dreamed degraded images by inverting the deraining model (AGANet [11]) pre-trained on RainDrops dataset [11]. With batch diversity loss $\mathcal{L}_{orth}$, we can reconstruct various raindrops types within a batch.



(a) w $\mathcal{L}_{orth}$

(b) w/o $\mathcal{L}_{orth}$

Figure 7. Our dreamed degraded images by inverting the pre-trained dehazing model (FFA-Net [12]). With batch diversity loss $\mathcal{L}_{orth}$, we can reconstruct a variety of haze types within a batch. This demonstrates the potential to extend our approach to other low-level tasks.