TPSeNCE: Towards Artifact-Free Realistic Rain Generation for Deraining and Object Detection in Rain
Supplementary Material

Shen Zheng[1], Changjie Lu[2], Srinivasa G. Narasimhan[1]
[1]Carnegie Mellon University, [2]University of Illinois Urbana-Champaign
{shenzhen, srinivas}@andrew.cmu.edu, cl140@illinois.edu

1. Supplementary Material

1.1. Overview

We begin by showcasing the selection of the most challenging 100 images from the BDD rainy test dataset. This emphasizes the efficacy of our method in enhancing object detection within these intense rainy conditions. Subsequently, we provide further insights into the implementation details of segmentation, deraining, and detection. Following this, we offer additional comparisons involving day2night and clear2snowy image-to-image translation, clear2rainy video-to-video translation, and object detection on heavy rain videos. Lastly, we feature additional rainy images generated by our method and display the failure cases.

1.2. Most Challenging 100 Images

Most images in the BDD test dataset with rain show light rain, which poses no substantial challenges for advanced deraining and detection algorithms. To evaluate the efficacy of the proposed approach on challenging rainy images, we selected 100 images from the BDD test set that contain heavy rain drops, streaks, mists, road reflections, and sub-optimal lighting, as demonstrated in Fig. 1 and 2. These factors result in blurring, occlusion, reflections, and distortions that pose significant challenges for deraining and detection algorithms.

Our work is one of the pioneering studies to examine diverse and demanding real heavy rain images that have been largely ignored by the research community of rain generation, deraining, and detection. While our method significantly improve the detection on heavy rains, it is vital that the research community draw attention to these challenges and drive continual performance improvements in the future.

1.3. Implementation Details

Source Code We include a PyTorch [8] implementation in the folder named *'TPSeNCE'* to enhance the reproducibility and transparency of this work

SeNCE Optimal Transport Inspired by [17, 18], SeNCE uses optimal transport [9] to collectively optimize patch weights for collaborative evaluation of contrastive learning objects. The optimal transport objective is:

$$\begin{aligned} &\min_{w_{ij}, i,j\in[1,N]} \left[\sum_{i=1}^{N} \sum_{\substack{j=1 \\ j\neq i}}^{N} w_{ij} \cdot e^{\frac{x_i \cdot y_j}{\beta}} \right] \\ &\text{s.t.} \sum_{i=1}^{N} w_{ij} = 1, \sum_{j=1}^{N} w_{ij} = 1 \end{aligned} \tag{1}$$

SeNCE Semantic Segmentation We utilize UperNet [14] with the ConvNeXt-B [6] backbone from the MMsegmentation [3] toolbox. The segmentation model was pretrained with semantic segmentation annotations on the entire BDD100k dataset (including both clear and real rainy images). Before training the rain generation model, we precompute and save the semantic segmentation map of all clear and rainy images, so we do not need to go through the computationally expensive semantic segmentation process during training. During each training iteration, we calculate the mPA between the precomputed semantic segmentation map of the clear and the rainy image and use that mPA to guide the weighting procedure of contrastive learning.

Deraining We employ state-of-the-art deraining methods, including EffDerain [4], VRGNet [13], PreNet [11] and SAPNet [19] to aid the comparison for rain generation methods. Due to the data-driven nature, deraining algorithms trained on better generated rainy images deliver better deraining performances. For the Rain100H dataset [15], we utilize the pretrained models trained on the Rain100H training set. For others, we train it from scratch using generated rainy images from various rain generation methods.

Detection We adopt Yolov3 [10] from the MMdetection [2] toolbox as the object detection model to compare the performances of rain generation methods. Similar to deraining, detection algorithms trained on better generated rainy images should deliver better detection results. For all detection experiments, the batch size is 32, the model backbone is DarkNet-53, and the training takes 273 epochs on 4 RTX 3090 GPUs.

1.4. Dataset Details[1]

BDD100K The BDD100K [16] dataset contains 100,000 720p images captured in different weather conditions and time-of-day divided into train:val:test=8:1:1. We first exclude unlabelled images, images with undefined categories, images captured on foggy days, based on the annotation files. Images not in the 'Clear' category are then labeled as 'Rainy' for rainy days, 'Snowy' for snowy days, and 'Night' for nighttime. We finally merge the validation and testing images to create the test set.

INIT The INIT [12] dataset contains over 50,000 images captured in four conditions: sunny, night, cloudy, and rainy, divided into train:test=17:3. Batch1 and Batch2 have images of resolution 1208×1920 and 3000×4000, respectively. Based on their training and testing images, we select the sunny subset as 'Clear' and the raining subset as 'Rainy'.

Boreas The Boreas [1] dataset contains 44 720p videos captured in various weather conditions and time-of-day. We extract 2 frames per second, define the train:test ratio as 4:1, categorize 'Sunny' videos as 'Clear', and classify 'Snowing' videos as 'Snowy'.

1.5. More Image Translation and Video Analysis

Day2Night Translation We show the qualitative comparison for day2night translation on BDD100K dataset in Fig. 3. The proposed model clearly outperforms other baselines in terms of quality, demonstrating lower contrast and illumination, accurate simulation of artificial light sources, and realistic shadow generation.

Clear2Snowy Image Translation In Tab. 1, we quantitatively analyze clear-to-snowy translation on the BDD100K dataset. The BDD100K dataset, with its minimal snow presence, doesn't distinguish snow generation models as effectively as the snow-rich Boreas dataset. Therefore, we exclude BDD100K from the visual comparison.

Clear2Rainy Video Translation As shown in this video, we qualitatively compare our method against several image-to-image translation methods on a BDD100K video regarding rain generation. While video-to-video translation is more challenging than image-to-image translation, the proposed method achieves excellent temporal consistency and image quality with minimal artifacts and distortions.

Object Detection on Heavy Rain Videos As shown in this video, we apply Yolov3 on a video captured in heavy rain. Even in heavy rains, Yolov3 finetuned on the generated rainy images from our method delivers strong performance on detecting various objects, such as cars, trucks, traffic lights, and stop signs.

[1]During training, all images are reshaped to a height of 360 with the height-width ratio unchanged to reduce computational costs.

Methods	BDD100K Dataset (clear $\rightarrow$ snowy)			
	Content↑	Style↑	MMD↓	ED↓
CUT [7]	2.87	2.60	473.456	47.759
QS-Attn [5]	2.90	2.90	473.289	47.728
MoNCE [18]	3.07	2.97	473.460	47.847
Ours	**3.53**	**3.37**	**473.069**	**47.461**

Table 1: **Quantitative comparison of snow generation on the BDD100K dataset.**

1.6. More Generated Images

We show additional clear2rainy and day2night image-to-image translation of our method in Fig. 4. In brief, our method is able to generate realistic rainy and night images with minimal artifacts and distortions.

1.7. Failure Cases

As shown in Fig. 5, the object detection model trained on rainy images generated by the contemporary proposed method struggles to handle extremely heavy rain or strong light sources. We have discussed the potential remedies for these cases in the main manuscript.

References

[1] Keenan Burnett, David J Yoon, Yuchen Wu, Andrew Z Li, Haowei Zhang, Shichen Lu, Jingxing Qian, Wei-Kang Tseng, Andrew Lambert, Keith YK Leung, et al. Boreas: A multi-season autonomous driving dataset. *The International Journal of Robotics Research*, 42(1-2):33–42, 2023. 2

[2] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, et al. Mmdetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*, 2019. 1

[3] MMSegmentation Contributors. MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark. https://github.com/open-mmlab/mmsegmentation, 2020. 1

[4] Qing Guo, Jingyang Sun, Felix Juefei-Xu, Lei Ma, Xiaofei Xie, Wei Feng, Yang Liu, and Jianjun Zhao. Efficientderain: Learning pixel-wise dilation filtering for high-efficiency single-image deraining. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 1487–1495, 2021. 1

[5] Xueqi Hu, Xinyue Zhou, Qiusheng Huang, Zhengyi Shi, Li Sun, and Qingli Li. Qs-attn: Query-selected attention for contrastive learning in i2i translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18291–18300, 2022. 2, 5

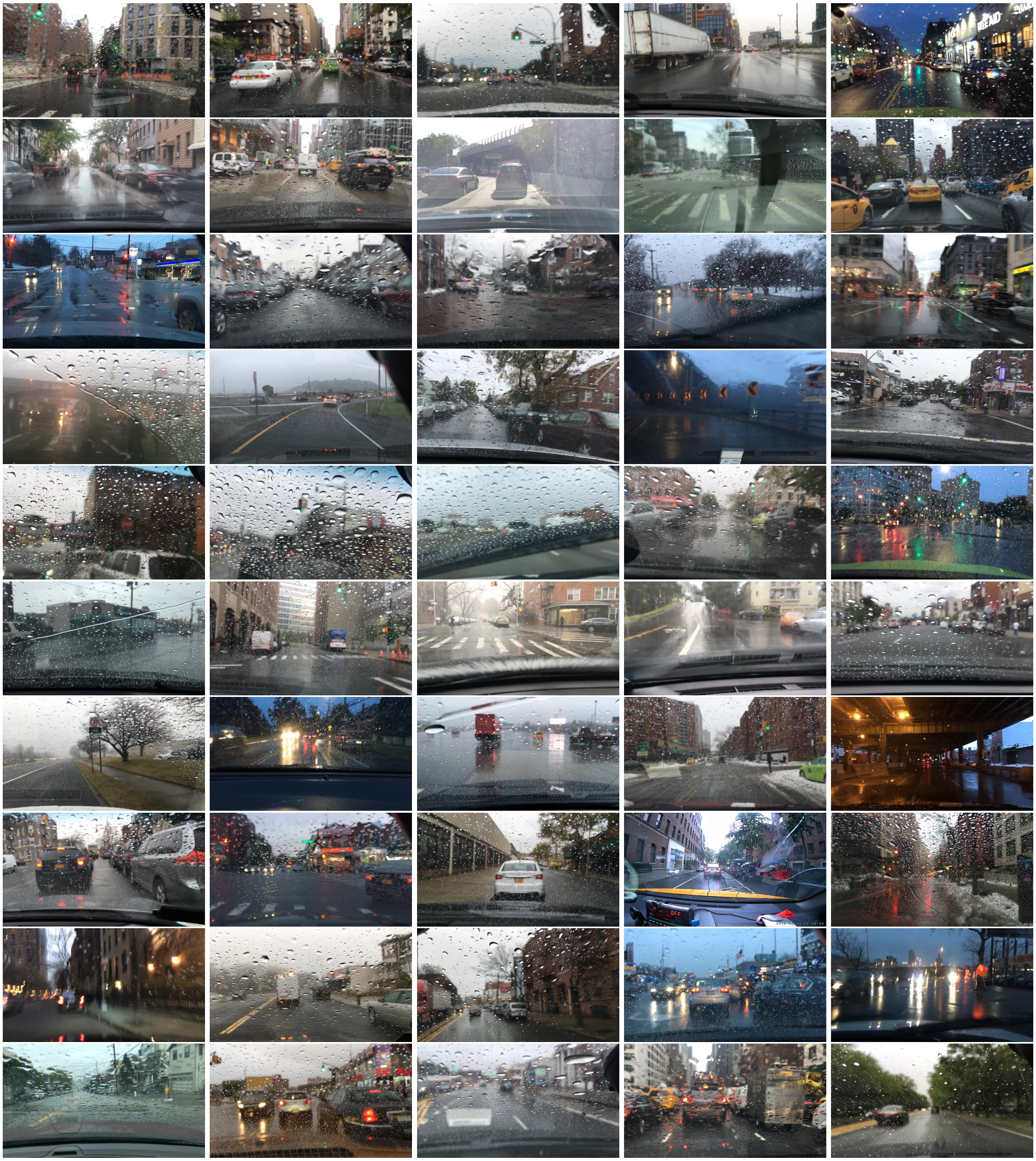

Figure 1: **Sample challenging rainy images we select from BDD test rainy.** These images contain heavy rain drops, streaks, mists, road reflections, and sub-optimal lighting, which post significant challenges to deraining and detection algorithms.

[6] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11976–11986, 2022. 1

[7] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-

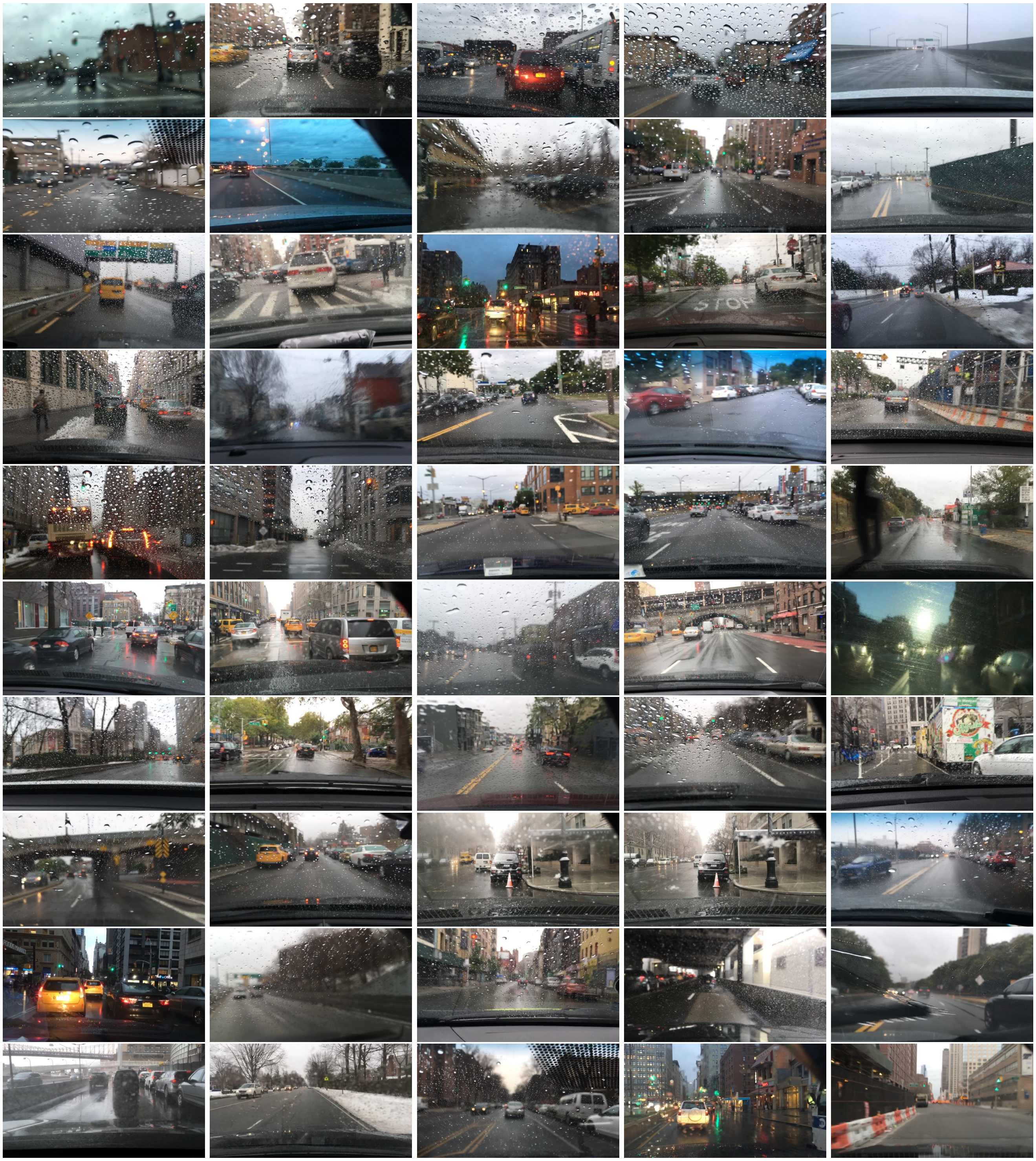

Figure 2: **Sample challenging rainy images we select from BDD test rainy.** These images contain heavy rain drops, streaks, mists, road reflections, and sub-optimal lighting, which post significant challenges to deraining and detection algorithms.

Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*, pages 319–345. Springer, 2020. 2, 5

[8] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming

Day Ours CUT [7] QS-Attn [5] MoNCE [18]

Figure 3: **Qualitative comparison for day2night translation on the BDD100K dataset.**

Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 1

[9] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019. 1

[10] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. 1

[11] 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. 1

[12] Zhiqiang Shen, Mingyang Huang, Jianping Shi, Xiangyang Xue, and Thomas S Huang. Towards instance-level image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3683–3692, 2019. 2

[13] Hong Wang, Zongsheng Yue, Qi Xie, Qian Zhao, Yefeng Zheng, and Deyu Meng. From rain generation to rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14791–14801, 2021. 1

[14] Tete Xiao, Yingcheng Liu, Bolei Zhou, Yuning Jiang, and Jian Sun. Unified perceptual parsing for scene understanding. In *Proceedings of the European conference on computer vision (ECCV)*, pages 418–434, 2018. 1

[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. 1

[16] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2636–2645, 2020. 2

[17] Fangneng Zhan, Yingchen Yu, Kaiwen Cui, Gongjie Zhang, Shijian Lu, Jianxiong Pan, Changgong Zhang, Feiying Ma, Xuansong Xie, and Chunyan Miao. Unbalanced feature transport for exemplar-based image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15028–15038, 2021. 1

[18] Fangneng Zhan, Jiahui Zhang, Yingchen Yu, Rongliang Wu, and Shijian Lu. Modulated contrast for versatile image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18280–18290, 2022. 1, 2, 5

[19] Shen Zheng, Changjie Lu, Yuxiong Wu, and Gaurav Gupta. Sapnet: Segmentation-aware progressive network for perceptual contrastive deraining. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 52–62, 2022. 1

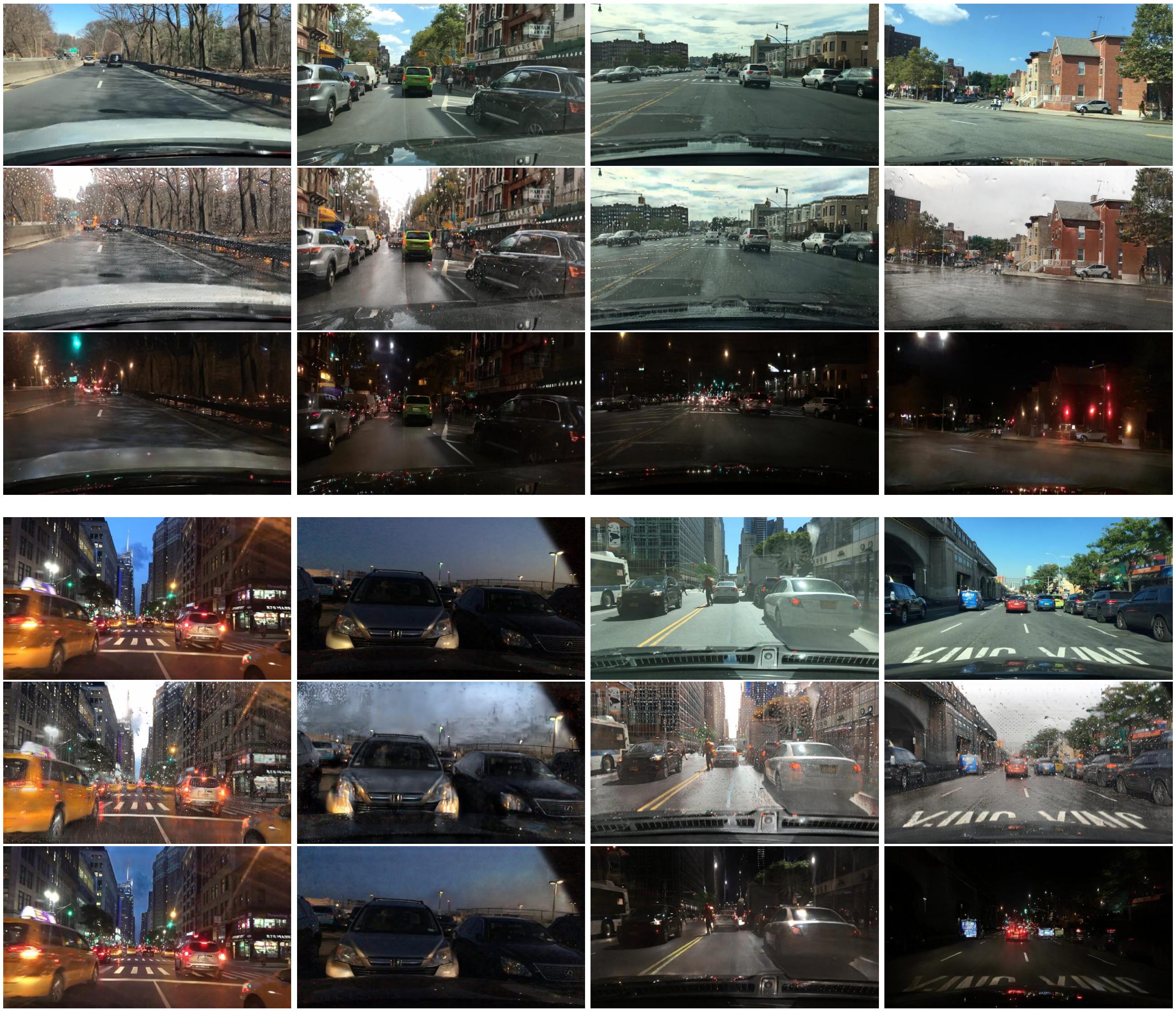

Figure 4: **Additional generated rainy images from the proposed TPSeNCE**. The images are categorized as follows: the 1st and 4th rows are clear, the 2nd and 5th rows are generated rainy, and the 3rd and 6th rows are generated night.

Figure 5: **Failure cases for the proposed method.** Red: Yolov3 pretrained on clear weather images. Green: Yolov3 finetuned on our generated rainy images.