Video Restoration Framework and its Meta-adaptations to Data-poor Conditions

Prashant W Patil, Sunil Gupta, Santu Rana, Svetha Venkatesh
A^2I^2, Deakin University, Geelong Warun Ponds, VIC, Australia
Supplementary Material

Overview

The supplementary material contains:

1. Database Generation Procedure
2. Computational Complexity Analysis
3. Qualitative Results on RainSynAll100 Database
4. Ablation Study on Losses
5. Qualitative Results on Meta-adaptation

1 Database Generation Procedure

Hazy Video Generation: In general, the synthetic haze formation [1] is modelled as:

$$f_h^t(x,y) = f_c^t(x,y)f_{tr}^t(x,y) + A(1 - f_{tr}^t(x,y)) \tag{1}$$

where, A represents global atmospheric light, f_h^t, f_c^t and f_{tr}^t represent hazy, clean frames and respective transmission map at time instance t respectively. The transmission map is defined as $f_{tr}^t = e^{-\beta d(t)}$, where $d(t)$ and β denote the scene depth map and the atmosphere scattering parameter, respectively. As depth maps are required for haze modelling, the respective depth map of each frame in a video is estimated using [2]. The remaining parameters for hazing process *i.e.* attenuation coefficient and atmospheric light are set as β=2 and $A = (1,1,1)$ respectively. With these parameters and estimated depth maps, 30 (**20 training** and **10 validation**) videos for day and 30 (**10 training** and **20 testing**) night hazy videos are generated.

Rainy Video Generation: From rain generation model [3], the rainy video generation process is formulated as:

$$f_r^t(x,y) = f_c^t(x,y) + f_{rs}^t(x,y), \qquad m \in 1,...,t \tag{2}$$

where, f_r^t, f_c^t and f_{rs}^t represent a rainy, rain-free and rain-streaks frames respectively. The rain streaks for each frame are generated with [4]. With these parameters and generated rain streaks, 30 (**20 training** and **10 validation**) videos for day and 30 (**10 training** and **20 testing**) night rainy videos are generated.

Rain with Veiling effect Video Generation: The model from [4] is used for rain with veiling effect frames generation as:

$$f_{r_v}^t(x,y) = f_{tr}^t(x,y) \odot \left(f_c^t(x,y) + f_{rs}^t(x,y)\right) + A \odot \left(1 - f_{tr}^t(x,y)\right) \tag{3}$$

where, $f_{r_v}^t$, f_{tr}^t, f_{rs}^t and f_c^t, represents a rain with veiling effect frame, transmission map of rain free frame, rain-streaks frame and rain-free frame respectively, attenuation coefficient and atmospheric light are set as β=2 and $A = (1,1,1)$ respectively. All these parameters are taken from [4]. With these settings, we have

(a) de-hazing (b) de-raining (c) de-raining with veiling effect

Synthetically generated samples frames for (a) de-hazing, (b) de-raining and (c) de-raining with veiling effect.

generated 30 (**20 training** and **10 validation**) videos for day and 30 (**10 training** and **20 testing**) videos for night time rain with veiling effects.

As the number of video frames per video in DAVIS-2016 [5] database are less, we have performed the data augmentation, which includes horizontal flipping. With this, 60 videos (40 training and 20 validation) are considered for each day-time task (haze, rain and rain with veiling effect). In night-time videos, 10 and 20 videos for each adapted task (haze, rain and rain with veiling effect) are used for training and testing purpose. Each training video comprises of 10 video frames and 200 frames in each testing video. Few samples from night-time de-raining, de-hazing and de-raining with veiling effect video are depicted in Figure shown above. These synthetically generated night-time weather degraded, haze, rain and rain with veiling effect datasets are abbreviated as **C**omprehensive **N**ight-**H**aze **D**ataset (CNHD), **C**omprehensive **N**ight-**R**ain **D**ataset (CNRD) and **C**omprehensive **N**ight-**R**ain with **V**eiling **D**ataset (CNRVD) respectively. *To the best of authors' knowledge, this is the first approach which generates synthetic night-time weather degraded videos.*

2 Computational Complexity Analysis

Any restoration framework is used as pre-processing module for high-level application like depth estimation, object detection, *etc*. Thus, computational complexity of the proposed network is compared with existing state-of-the-art methods in terms of number of learning parameters, number of GFLOPs and inference time to restore the one frame is provided in Table S1. From Table S1, it is clear that the proposed network is efficient in terms of computational complexity as compared to the existing state-of-the-art methods.

Table S 1: Computational complexity analysis of the proposed network with SOTA methods.

Factors	RDNet [6]	DLF [7]	RMFD [8]	**Proposed**
Para. (M)	$\sim$ 65	$\sim$ 4.6	$\sim$ 29	$\sim$ **10**
FLOPs (G)	227	300	416	**128**
Time (Sec)	0.169	0.211	0.238	**0.137**

3 Qualitative Results on RainSynAll100 Database

The number of frames in each video of RainSynAll100 database are very less. Also, the state-of-the-art approaches [7], [8] provide the results by skipping the first and last two frames in each video. Thus, we have provided the collage for each video qualitative results. Figure S1 to S5 depicts the visual results on RainSynAll100 database with proposed architecture and DLF [7], RMFD [8].

Figure S 1: Qualitative result analysis on RainSynAll100 for video de-raining with veiling effect (DLF [7] and RMFD [8])

Figure S 2: Qualitative result analysis on RainSynAll100 for video de-raining with veiling effect (DLF [7] and RMFD [8])

Figure S 3: Qualitative result analysis on RainSynAll100 for video de-raining with veiling effect (DLF [7] and RMFD [8])

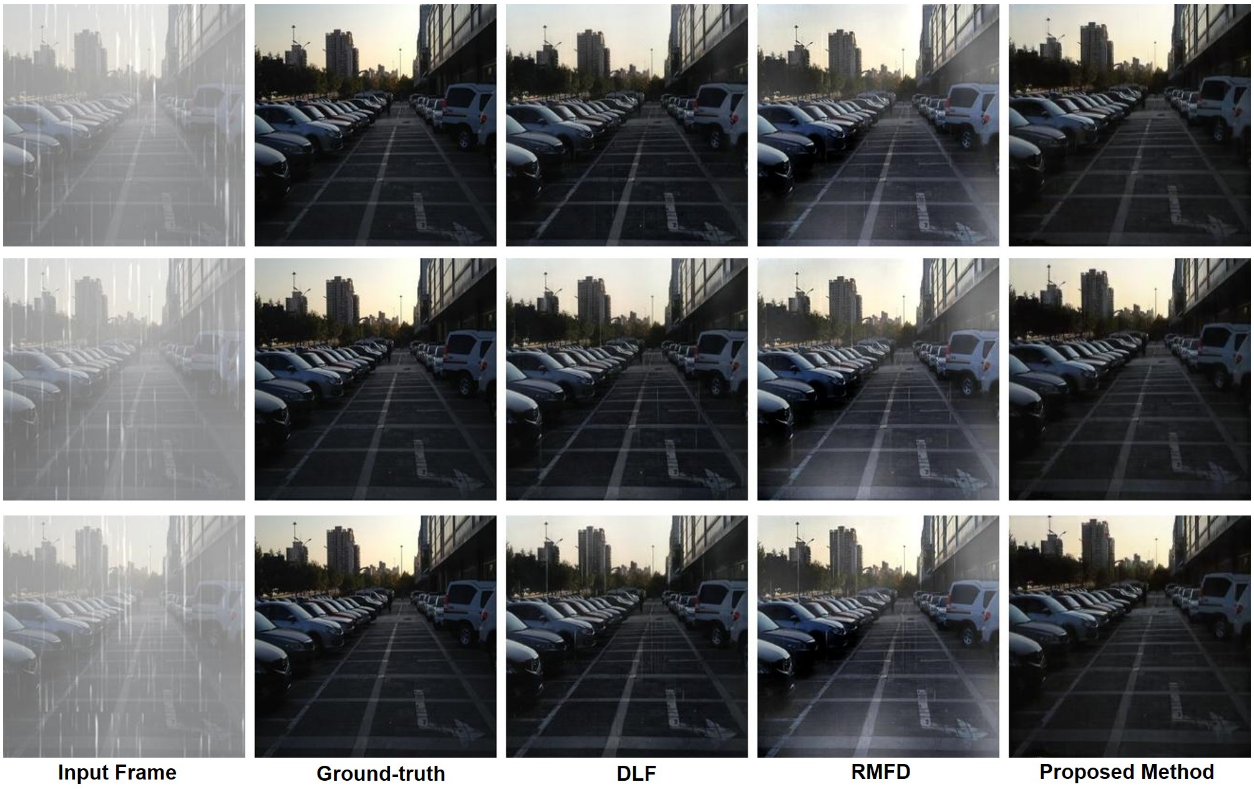

Figure S 4: Qualitative result analysis on RainSynAll100 for video de-raining with veiling effect (DLF [7] and RMFD [8])

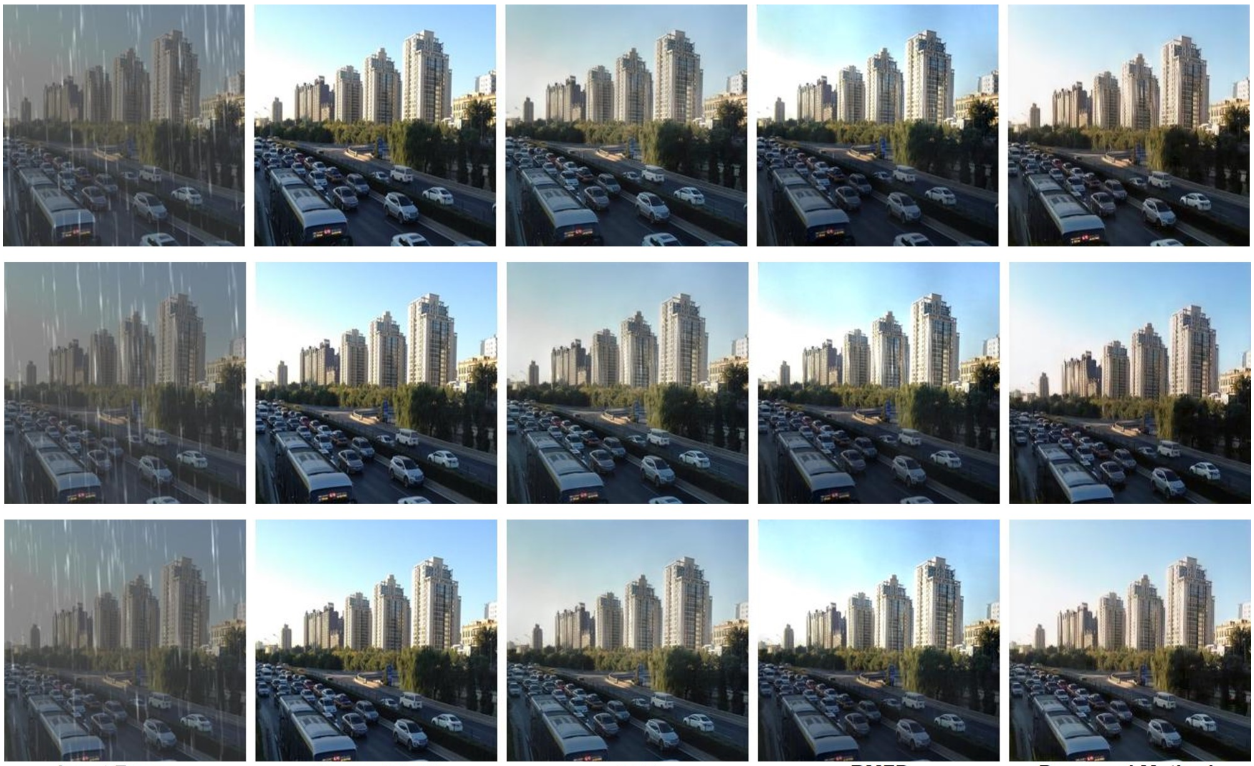

Figure S 5: Qualitative result analysis on RainSynAll100 for video de-raining with veiling effect (DLF [7] and RMFD [8])

4 Ablation Study on Losses

In the proposed framework, we have integrated the edge ($\mathbb{L}_{ed}$), perceptual ($\mathbb{L}_{P}$) and SSIM ($\mathbb{L}_{S}$) losses with $\mathbb{L}_1$ loss for weight optimization of the proposed network. Here, we have formulated four different combination of all losses by keeping the $\mathbb{L}_1$ loss common for ablation study. The quantitative evaluation for different combinations of losses is depicted in Table S2. From Table S2, it is evident that the combinations of $\mathbb{L}_{ed}$, $\mathbb{L}_{P}$ and $\mathbb{L}_{S}$ losses with $\mathbb{L}_1$ *i.e.* $\mathbb{L}_{Total}$ is giving superior performance compared to other combinations.

Table S 2: Effect of different training losses.

Losses	$\mathbb{L}_1$	$\mathbb{L}_1 + \mathbb{L}_{ed}$	$\mathbb{L}_1 + \mathbb{L}_{ed} + \mathbb{L}_P$	$\mathbb{L}_{Total}$
PSNR	25.21	25.59	25.97	**26.36**
SSIM	0.8898	0.8911	0.8954	**0.9034**

5 Qualitative Results on Meta-adaptation

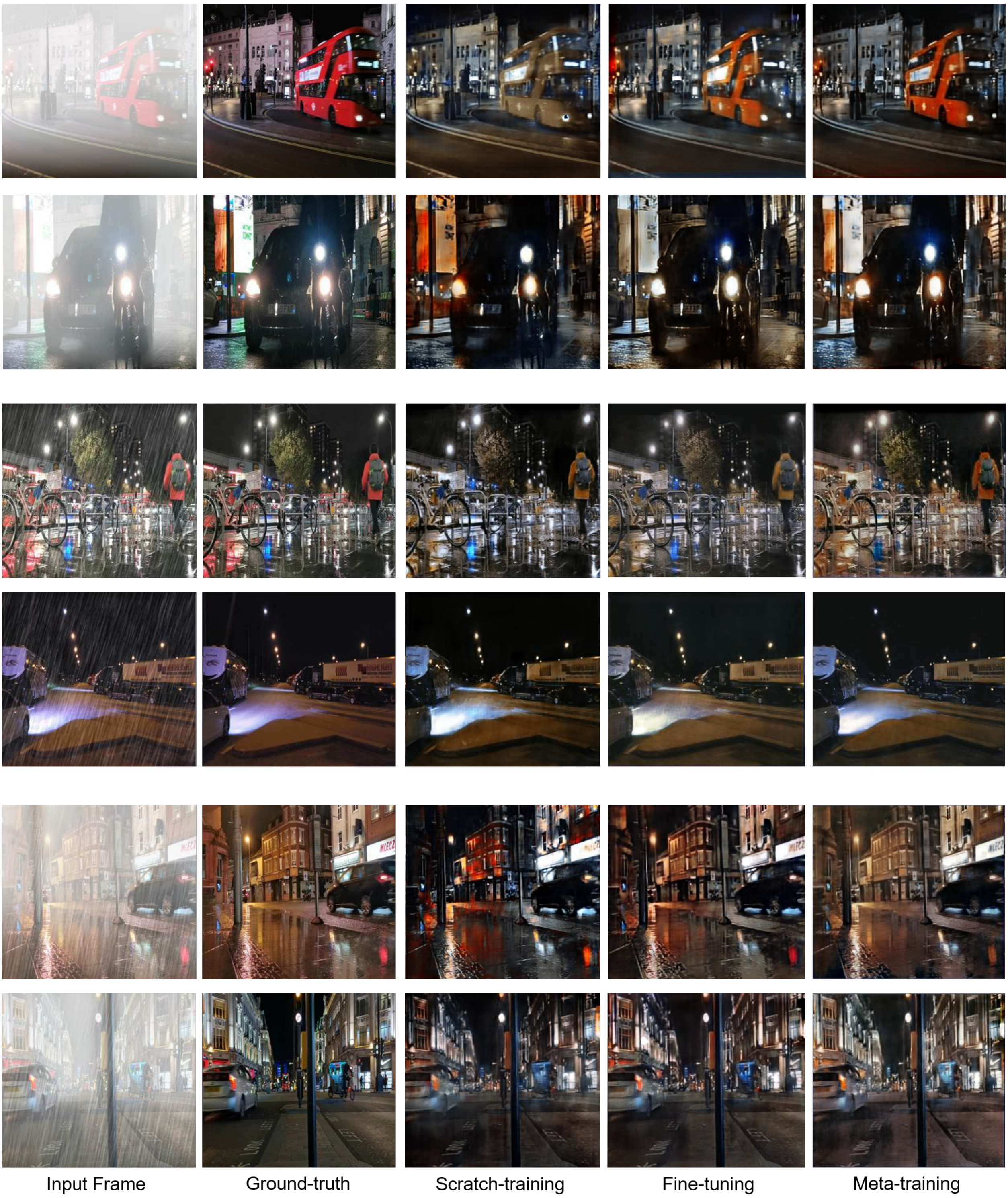

Figure S 6: Qualitative analysis with scratch training, fine-tuning and meta-training for night-time video restoration (*first two rows: video de-hazing, middle two rows: video de-raining and last two rows: video de-raining with veiling effect*).

References

[1] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.

[2] R. Ranftl, K. Lasinger, D. Hafner, K. Schindler, and V. Koltun, "Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[3] K. Garg and S. K. Nayar, "Photorealistic rendering of rain streaks," *ACM Transactions on Graphics (TOG)*, vol. 25, no. 3, pp. 996–1002, 2006.

[4] R. Li, L.-F. Cheong, and R. T. Tan, "Heavy rain image restoration: Integrating physics model and conditional adversarial learning," in *Proceedings of the IEEE/CVF Conference on CVPR*, 2019, pp. 1633–1642.

[5] F. Perazzi, J. Pont-Tuset, B. McWilliams, L. Van Gool, M. Gross, and A. Sorkine-Hornung, "A benchmark dataset and evaluation methodology for video object segmentation," in *Proceedings of the IEEE conference on CVPR*, 2016, pp. 724–732.

[6] S. Zhao, L. Zhang, Y. Shen, and Y. Zhou, "Refinednet: A weakly supervised refinement framework for single image dehazing," *IEEE Transactions on Image Processing*, vol. 30, pp. 3391–3404, 2021.

[7] W. Yang, J. Liu, and J. Feng, "Frame-consistent recurrent video deraining with dual-level flow," in *Proceedings of the IEEE/CVF Conference on CVPR*, 2019, pp. 1661–1670.

[8] W. Yang, R. T. Tan, J. Feng, S. Wang, B. Cheng, and J. Liu, "Recurrent multi-frame deraining: Combining physics guidance and adversarial learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.