

Real-world Video Super-resolution: A Benchmark Dataset and A Decomposition based Learning Scheme - Supplementary Material -

Xi Yang^{1,2,*}, Wangmeng Xiang^{1,2,*}, Hui Zeng^{1,2}, Lei Zhang^{1,2,†}

¹The Hong Kong Polytechnic University, ²DAMO Academy, Alibaba Group

xxxxi.yang@connect.polyu.hk, {cswxiang, cshzeng, cslzhang}@comp.polyu.edu.hk

1. More example sequences and content analysis of RealVSR dataset

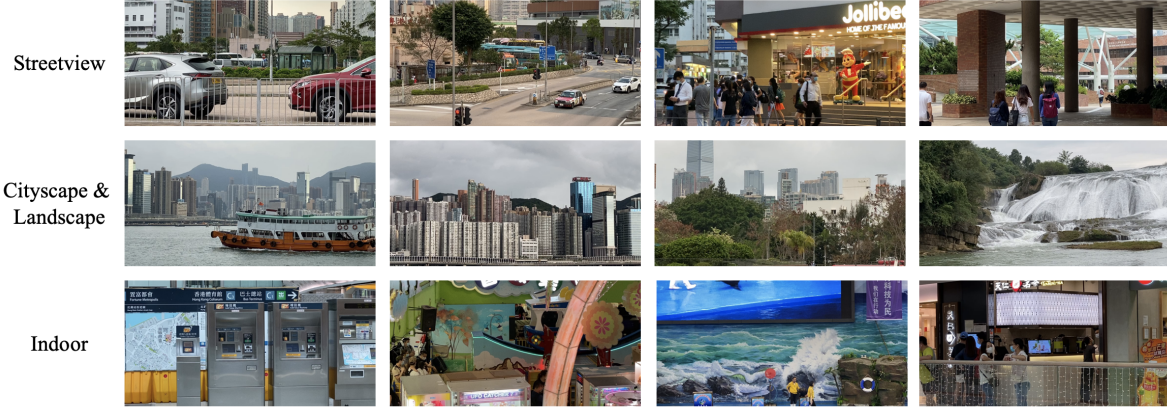


Figure 1. More example video sequences from the constructed RealVSR dataset.

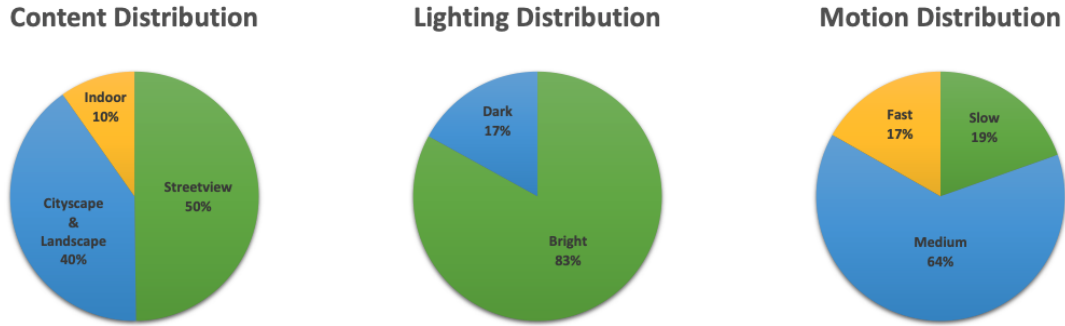


Figure 2. More content analysis of the constructed RealVSR dataset.

*Equal contribution.

†Corresponding author. This work is supported by the Hong Kong RGC RIF grant (R5001-18).

2. More visual results by VSR models trained on synthetic dataset and our RealVSR dataset

In this section, we provide more visual results by VSR models trained on the synthetic Vimeo-90k [4] dataset and our RealVSR dataset. The results on RealVSR testing set are shown in Fig. 3. Again, models trained on the RealVSR dataset produce sharper edges and generate much less artifacts than models trained on the synthetic Vimeo-90k dataset.



Figure 3. $\times 2$ VSR results on our RealVSR testing set by different models.

3. More visual results by VSR models trained with different losses

In this section, we provide more visual results by VSR models trained with different losses and the results are shown in Fig. 4.

Compared to the baseline \mathcal{L}_{CB}^{YCbCr} , decomposition-based losses $\mathcal{L}_s + \mathcal{L}_d + \mathcal{L}_{CB}^{CbCr}$ and \mathcal{L}_{v1} could reconstruct finer details and shaper edges. The proposed \mathcal{L}_{v2} enables VSR models to generate sharper details compared with the baseline $\mathcal{L}_{v1} + \text{RaGAN}$, and also improves the visual quality of the VSR models trained by \mathcal{L}_{v1} .

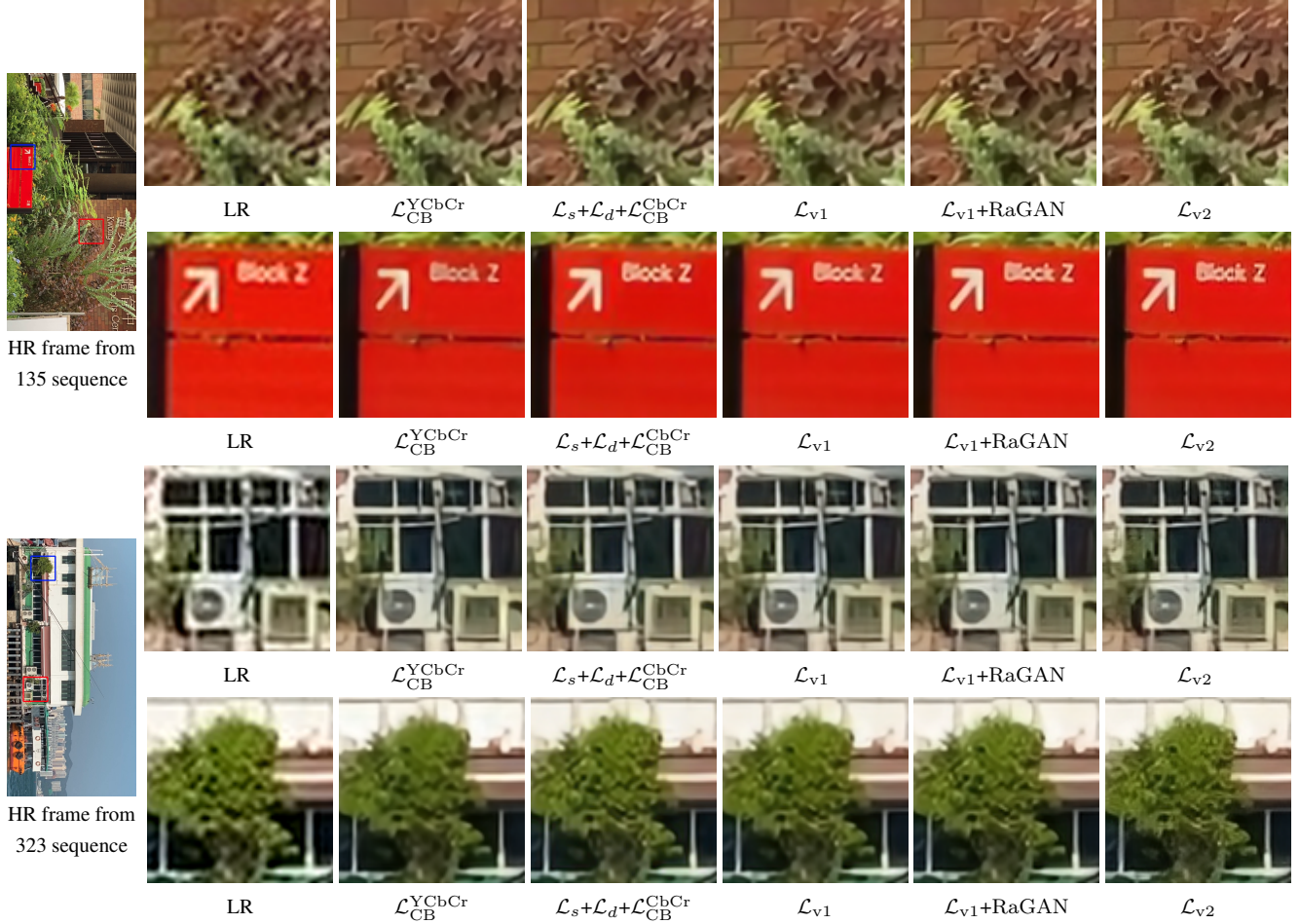


Figure 4. $\times 2$ VSR results on videos from the RealVSR testing set by the EDVR [3] models trained with different losses.

4. More visual results on real-world videos

In this section, we provide more visual VSR results on real-world videos outside the RealVSR dataset. The results are shown in Fig. 5. Compared with the model trained on the synthetic Vimeo-90k dataset, the model trained on our RealVSR dataset with loss \mathcal{L}_{v1} reconstructs clearer edges with less artifacts. Moreover, the model trained with loss \mathcal{L}_{v2} enriches the details and textures, which further improves the visual quality.



Figure 5. $\times 2$ VSR results on real-world videos outside RealVSR dataset by the EDVR [3] models trained on synthetic Vimeo-90k [4] and our RealVSR.

5. RealSR dataset vs. RealVSR dataset

In this section, we compare models trained on real-world SISR dataset (RealSR [1]) and those trained on our real-world VSR dataset (RealVSR). Two representative SISR models, SRResNet [2] and RCAN [5] are selected in this experiment. For fair comparison, we train all models with the baseline Charbonnier (CB) loss in YCbCr space. As shown in Fig. 6, models trained on the RealSR dataset are less effective in reconstructing sharp details on real-world distorted videos captured by several models of mobile phone. In contrast, models trained on the RealVSR dataset produce much better results.

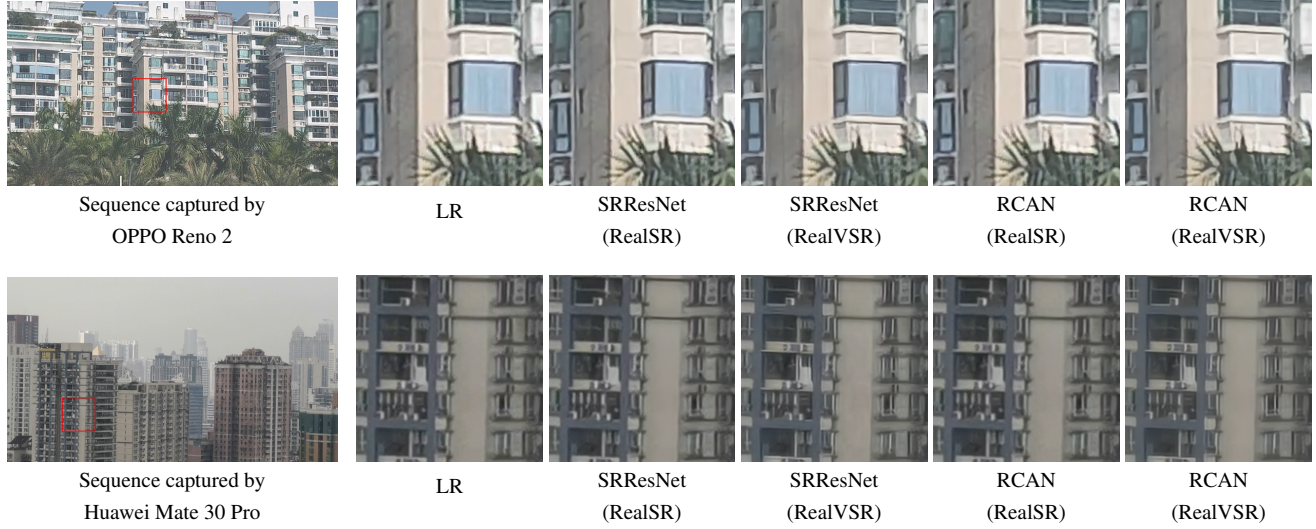


Figure 6. $\times 2$ VSR results on real-world videos outside RealVSR dataset by the different models trained on RealSR [1] and our RealVSR.

6. Video demonstration

In this section, we provide some video demonstrations. The videos can be found in the same folder. To better compare the results, we crop the center location of each video.

References

- [1] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3086–3095, 2019. 5
- [2] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 5
- [3] Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 3, 4
- [4] Tianfan Xue, Baian Chen, Jiajun Wu, Donglai Wei, and William T Freeman. Video enhancement with task-oriented flow. *International Journal of Computer Vision*, 127(8):1106–1125, 2019. 2, 4
- [5] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 286–301, 2018. 5