Learning Temporal Consistency for Low Light Video Enhancement from Single Images (Supplementary Material)

Fan Zhang[1] Yu Li[2] Shaodi You[3] Ying Fu[1*]
[1]Beijing Institute of Technology [2]Applied Research Center (ARC), Tencent PCG
[3]University of Amsterdam

In this supplementary material, we first provide more details on data preparation stage including results of instance segmentation step and optical flow prediction step. Then we provide more experimental results on synthetic data and real data. In addition, we provide results of ablation study on low light levels. Finally, we simply investigate inference speed of all methods.

1. Data Preparation

Our method utilizes optical flow to represent motions occurred in video sequences. While most existing optical flow estimation methods focus on correctly estimating optical flow between image pairs, we generate non-existed optical flow for single images whose corresponding consecutive frames are not recorded. Conditional Motion Propagation Network (CMP) [8] is an unsupervised method which predicts optical flow based on guidance vectors which provide direction and velocity information of image pixels. It is suitable for our problem and works well with manual operations. To further improve efficiency and reduce labour cost, we use instance segmentation [7] to separate objects from background and randomly sample guidance vectors to be fed into optical flow predicting network.

1.1. Instance Segmentation

We utilize Detectron2 toolkit [7] to separate objects from background. The model we choose is the pretrained R50-FPN on COCO Instance Segmentation Baselines. We use all the predefined object classes in COCO and keep all the results with confidence higher than 85%. Some visual results of masks produced by this model are shown in Figure 1.

1.2. Optical Flow Prediction

Conditional Motion Propagation Network (CMP) [8] needs guidance vectors on object area. With those segmentation masks acquired, we randomly sample 10 vectors with the same direction and velocity in each object area. Fed with image and corresponding guidance vectors, CMP predicts optical flow accordingly. While CMP mainly focuses on optical flow on moving objects, we also combine global affine transformation with predicted optical flow to get final results. Visual results of our optical flow and actual estimated one from ground truth sequences are shown in Figure 2. We can see that the ever non-existed optical flow looks quite similar to actual one containing local and global motions.

2. Results on Synthetic and Real Data

In this section, we provide more qualitative results of all compared methods including LIME [3], MBLLEN/MBLLVEN [5], RetinexNet [6], SID [1], SFR [2], SMOID [4] and our method for further comparison.

2.1. Results on Synthetic Data

More results on both synthetic noise-free data and noisy data are shown in Figures 3 and 4, respectively. We can see that results of our method are comparable to those of video-based methods including MBLLVEN [5] and SMOID [4]. LIME [3] suffers from over-saturation and under-exposure. MBLLEN [5] suffers from under-exposure. RetinexNet [6] enhances dark images with unreal color. Our baseline SID [1] suffers from severe artifacts which lead to flickering. SFR [2] alleviates these artifacts a lot but still fails in some area.

2.2. Results on Real Data

More results on our collected real data are shown in Figure 5. LIME [3] is over-saturated and RetinexNet [6] gets unreal color. LIME [3] and MBLLEN [5] are over-exposed. SID [1] still has artifacts and we can also find some defects around the area of building glasses in results of SFR [2]. Our method are comparable to video-based methods

*Corresponding author: fuying@bit.edu.cn

MBLLVEN [5] and SMOID [4] in naturalness and brightness level.

3. Other Comparisons

3.1. Ablation Study on Low Light Levels

We conduct a simple ablation study on different light levels. We split test data evenly into two groups based on PSNR, representing lower and higher light levels. The mean input PSNR in the two groups are 7.6 dB and 10.6 dB while our enhanced results get 24.3 dB and 24.4 dB accordingly (Table 1). This shows our method is robust to different light levels.

Table 1. Ablation of light levels.

Mean Input PSNR (dB)	7.6	10.6
Mean Output PSNR (dB)	24.3	24.4

3.2. Speed Comparison

We try to help image-based network be more stable on videos for its performance and relatively easy-to-collect datasets. Another reason leads us to this is the efficiency of image-based models. With no computationally costly modules like 3D convolution, image-based networks can work much faster than video-based ones. Here we conduct a simple speed test on all compared methods. Experimental results are shown in Table 2. All methods are run on platform of Intel(R) Xeon(R) CPU E5-2650 v4 @ 2.20GHz and single GTX 1080Ti. As can be seen in Table 2, SFR [2] and our method are as fast as the baseline SID [1] and much faster than other methods.

References

[1] Chen Chen, Qifeng Chen Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2018. 1, 2, 3

[2] Gabriel Eilertsen, Rafal K Mantiuk, and Jonas Unger. Single-frame regularization for temporally stable cnns. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2019. 1, 2, 3

[3] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2017. 1, 3

[4] Haiyang Jiang and Yinqiang Zheng. Learning to see moving objects in the dark. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2019. 1, 2, 3

[5] Feifan Lv, Feng Lu, Jianhua Wu, and Chongsoon Lim. Mbllen: Low-light image/video enhancement using cnns. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2018. 1, 2, 3

[6] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2018. 1, 3

[7] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. https://github.com/facebookresearch/detectron2, 2019. 1

[8] Xiaohang Zhan, Xingang Pan, Ziwei Liu, Dahua Lin, and Chen Change Loy. Self-supervised learning via conditional motion propagation. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2019. 1

Figure 1. Instance segmentation results. We only keep masks objects with confidence retio higher than 85%.

Figure 2. Comparisons of our final optical flow results with actual ones. We combine outputs of CMP with global affine transformation for our final results which look quite similar to actual optical flow between ground truth sequences.

Table 2. Speed test. Time for each method to process one image of size 1920×1080 on average.

Method	LIME [3]	MBLLEN [5]	RetinexNet [6]	SID [1]	MBLLVEN [5]	SMOID [4]	SFR [2]	Ours
Time(s)	2.384	0.640	0.234	0.006	0.913	0.353	0.006	0.006

Figure 3. Results for clean case.

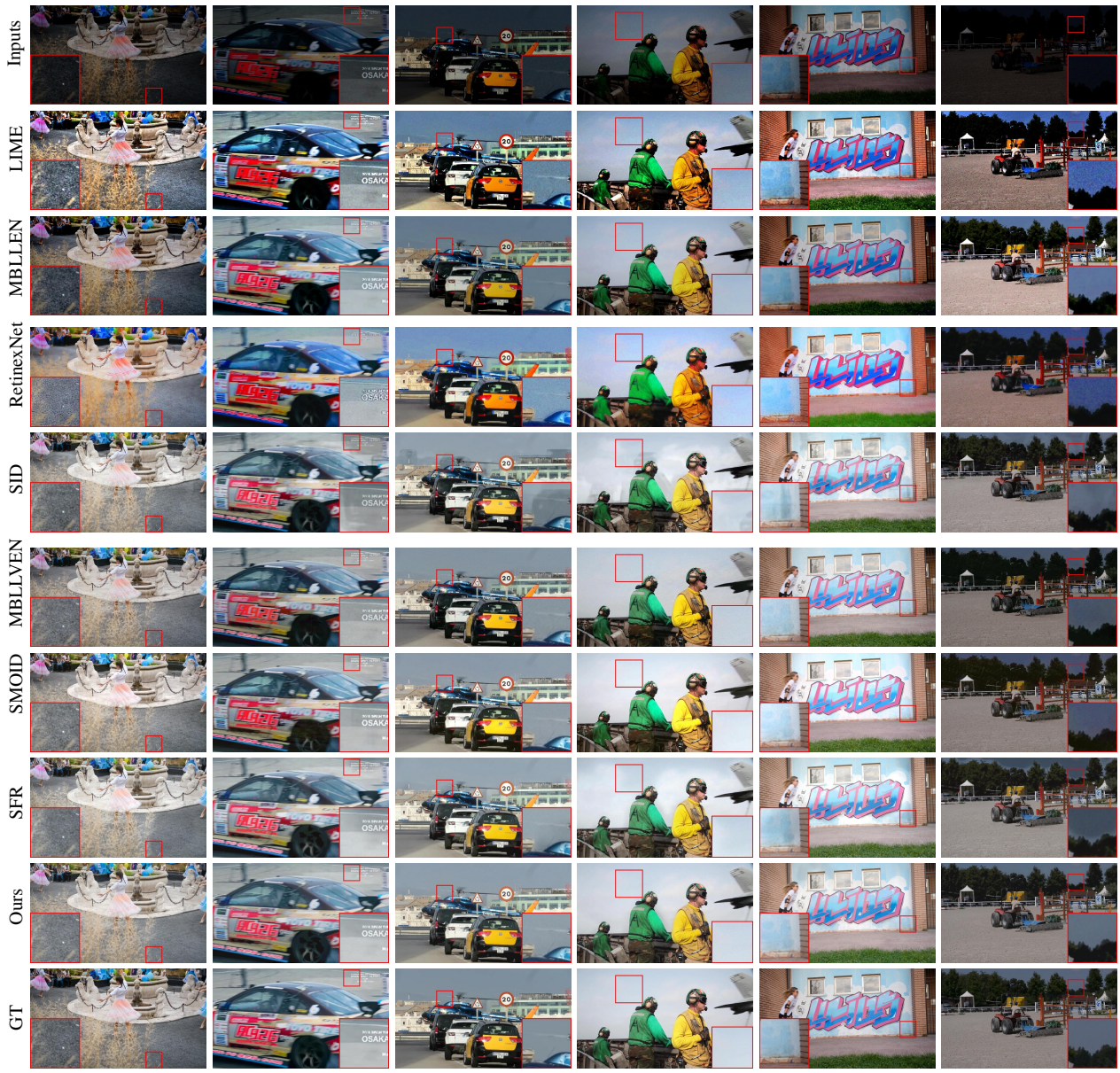

Figure 4. Results for noisy case.

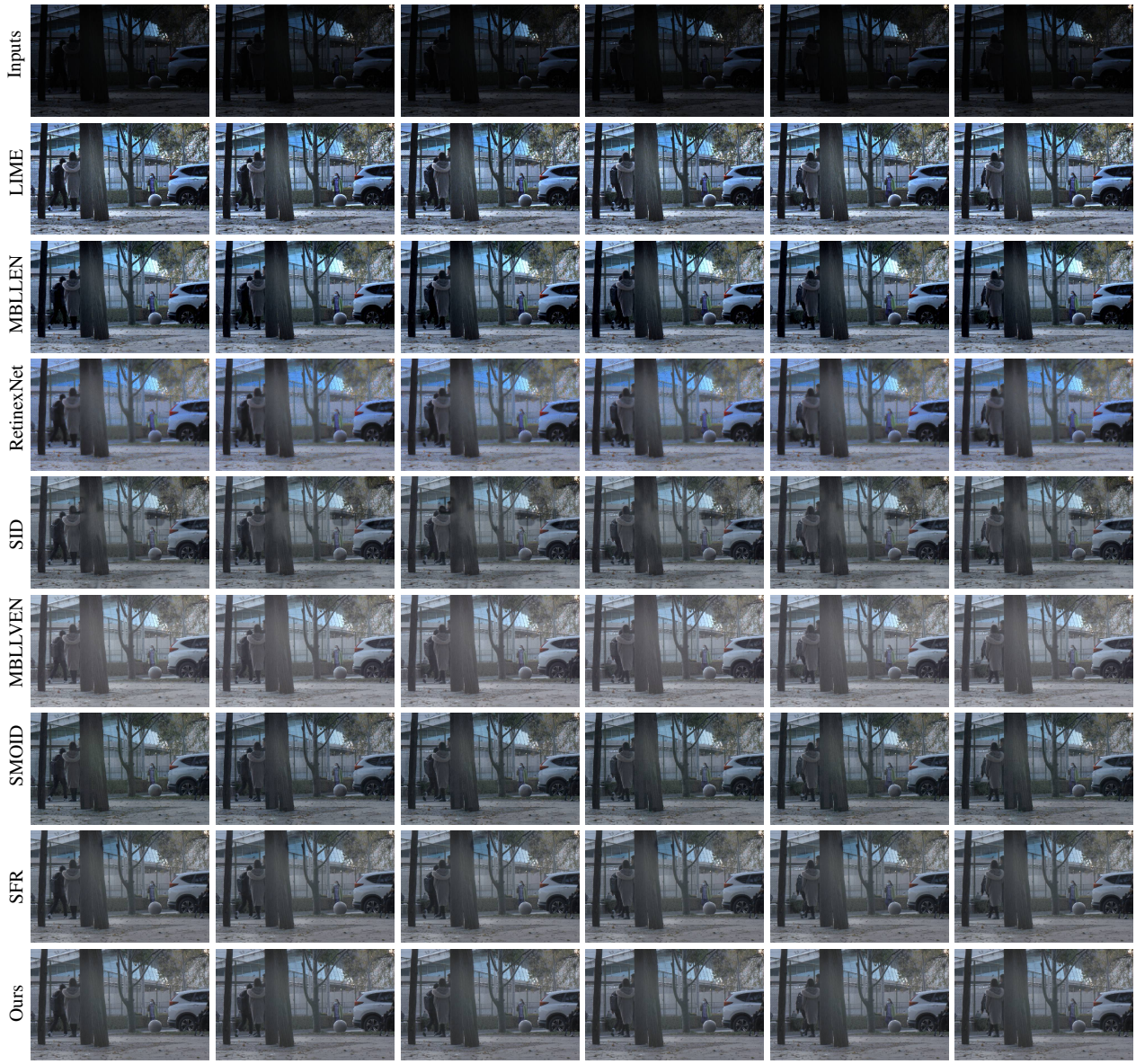

Figure 5. Results for real data.