Supplementary Material: Align-and-Attend Network for Globally and Locally Coherent Video Inpainting

Sanghyun Woo[1]
shwoo93@kaist.ac.kr
Dahun Kim[1]
mcahny@kaist.ac.kr
Kwanyong Park[1]
pkyong7@kaist.ac.kr
Joon-Young Lee[2]
jolee@adobe.com
In So Kweon[1]
iskweon77@kaist.ac.kr

[1] Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea
[2] Adobe Research, San Jose, CA, USA

A Appendix

A.1 Network Details

The network consists of three main parts: homography encoder, image encoder/decoder, and flow encoder/decoder. Our network design of the homography encoder follows [3]. For the image and the flow encoder/decoder design, we follow [2].

A.2 Implementation Details

Our model is implemented using Pytorch v0.4, CUDNN v7.0, CUDA v9.0. It runs on the hardware with Intel(R) Xeon(R) (2.10GHz) CPU and NVIDIA GTX 1080 Ti GPU. The model runs at 15 fps on a GPU for frames of 256×256 pixels. We use Adam optimizer with $\beta = (0.9, 0.999)$. The learning rate starts with 2e-4 and divided by 10 every 5 epochs. We train our model from scratch. The homography training and video inpainting training take about 3 days, each using eight NVIDIA GTX 1080 Ti GPUs.

A.3 Video Inpainting Results

We provide video inpainting results on DAVIS dataset. We compare our method with two strong baselines [1, 2]. The video is encoded with H264 codec into MP4 format. All the videos are played at the original speed. Please use 'pause' or adjust the speed if needed. Please refer to the attached video file.

References

[1] Jia-Bin Huang, Sing Bing Kang, Narendra Ahuja, and Johannes Kopf. Temporally coherent completion of dynamic video. *ACM Transactions on Graphics (TOG)*, 35(6):196, 2016.

[2] Dahun Kim, Sanghyun Woo, Joon-Young Lee, and In So Kweon. Deep video inpainting. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2019.

[3] Ignacio Rocco, Relja Arandjelovic, and Josef Sivic. Convolutional neural network architecture for geometric matching. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2017.

[4] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Guilin Liu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. Video-to-video synthesis. In *Proc. of Neural Information Processing Systems (NeurIPS)*, 2018.