Supplementary Meterails for Fast-Vid2Vid: Spatial-Temporal Compression for Video-to-Video Synthesis

Long Zhuo[1], Guangcong Wang[2], Shikai Li[3], Wayne Wu[1,3], and Ziwei Liu[2⋆]

[1] Shanghai AI Laboratory
[2] S-Lab, Nanyang Technological University
[3] SenseTime Research
zhuolong@pjlab.org.cn {guangcong.wang, ziwei.liu}@ntu.edu.sg
lishikai@sensetime.com wuwenyan0503@gmail.com

A. More comparison results

We further provide more comparison results in Figs. 1-3. As shown, our Fast-Vid2Vid generates more realistic sequences compared with other compression methods [1,2,3] and original vid2vid [4] in Sketch2Face, Segmentation2City and Pose2Body. In particular, as the time stamps move on, the sequences generated by other compression methods are with more and more artifacts while ours are with the stronger temporal coherence. The possible reason is that our method preserves the well-designed structures that can generate realistic sequence in the time dimension.

References

1. Jin, Q., Ren, J., Woodford, O.J., Wang, J., Yuan, G., Wang, Y., Tulyakov, S.: Teachers do more than teach: Compressing image-to-image models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13600–13611 (2021) 1
2. Li, M., Lin, J., Ding, Y., Liu, Z., Zhu, J.Y., Han, S.: Gan compression: Efficient architectures for interactive conditional gans. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5284–5294 (2020) 1
3. Liu, Y., Shu, Z., Li, Y., Lin, Z., Perazzi, F., Kung, S.Y.: Content-aware gan compression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12156–12166 (2021) 1
4. Wang, T.C., Liu, M.Y., Zhu, J.Y., Liu, G., Tao, A., Kautz, J., Catanzaro, B.: Video-to-video synthesis. arXiv preprint arXiv:1808.06601 (2018) 1

⋆ Corresponding author

Fig. 1: Qualitative results of the testing data compared with the advanced GAN compression methods in the task of Sketch2Face. From top to the bottom, rows are semantic maps, CA's results, CAT's results, NAS's results, Vid2Vid's results, Fast-Vid2Vid's results and GT.

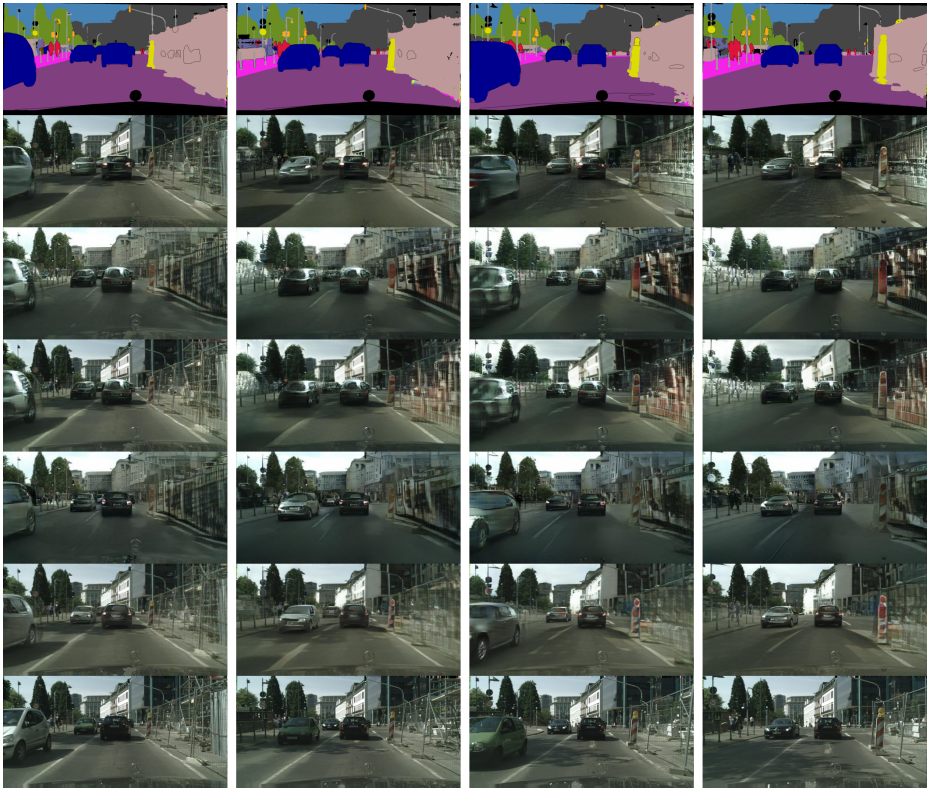

Fig. 2: Qualitative results of the testing data compared with the advanced GAN compression methods in the task of Segmentation2City. From top to the bottom, rows are semantic maps, CA's results, CAT's results, NAS's results, Vid2Vid's results, Fast-Vid2Vid's results and GT.

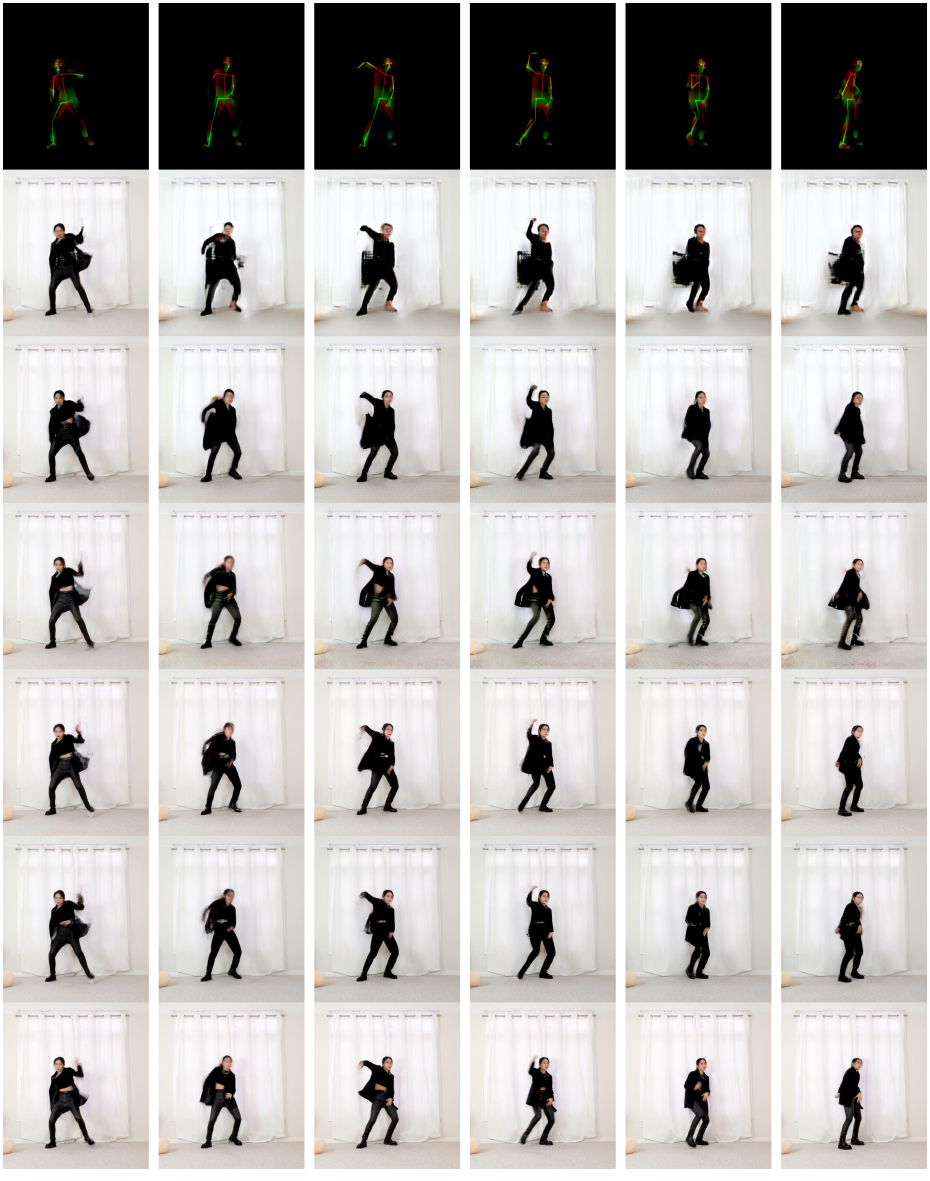

Fig. 3: Qualitative results of the testing data compared with the advanced GAN compression methods in the task of Pose2Body. From top to the bottom, rows are semantic maps, CA's results, CAT's results, NAS's results, Vid2Vid's results, Fast-Vid2Vid's results and GT.