Supplementary Material for DTVNet: Dynamic Time-lapse Video Generation via Single Still Image

Anonymous ECCV submission for

Paper ID 3335

1 Experiments on the Beach Dataset

We further conduct qualitative and quantitative experiments compared with MoCoGAN [2] and MDGAN [4] on the Beach dataset [3], which contains different video contexts compared to the Sky Time-lapse dataset [4].

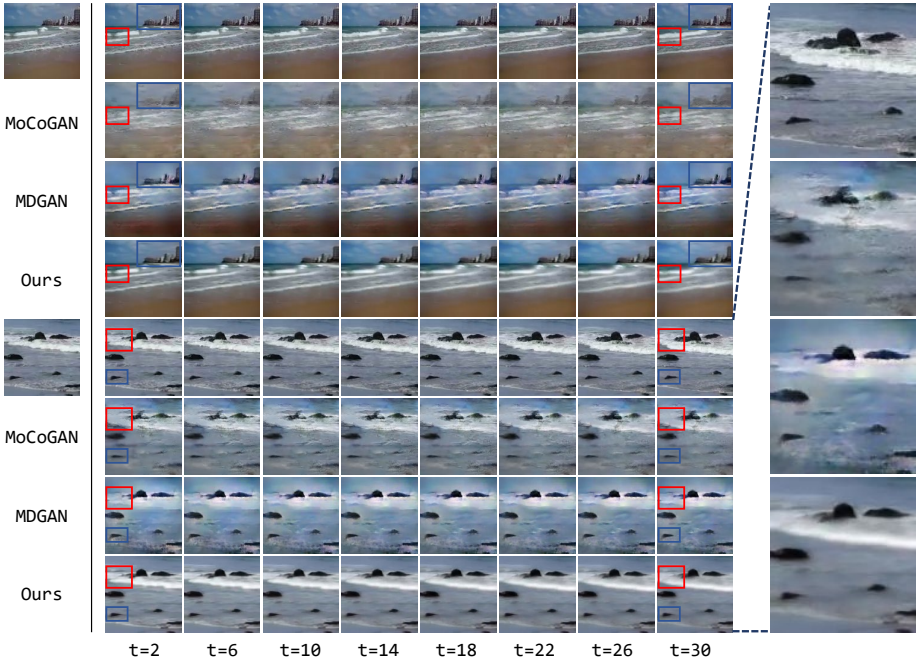

Fig. 1. Qualitative experimental results compared with MoCoGAN [2] and MDGAN [4] on the Beach dataset. The first column lists two different landscape images as the start frames, and the middle eight columns are generated video frames by different methods at different times. Right four enlarged images are long-term results at t=30 for a better visual comparison. Please zoom in red and blue rectangles for a more clear comparison.

Qualitative Results. We conduct and discuss qualitative experiments, compared with MoCoGAN and MDGAN, on the Beach dataset. As shown in Fig-

ure 1, we randomly sample two videos from the test dataset, and the first column shows the start frames of two videos while the second to ninth columns are generated video frames by different methods at different times. Note that the first and fifth rows are ground truth frames, and the SOTA models of other methods are supplied by official codes. Results similarly show that our method can not only keep better content information than other SOTA methods (Comparing the generated results in column at a special time), but also capture the dynamic motion (Viewing the generated results in row). In detail, the generated sequences produced by MoCoGAN (second and sixth rows) and MDGAN (third to seventh rows) become more and more distorted over time, thus the quality and motion can not be well identified. Specifically, we mark some dynamic and still details in red and blue rectangles respectively, and results show that our method can well keep the content of the still objects while generate reasonable dynamic details, which obviously outperforms all other state-of-the-art methods. *More qualitative experiments and generated results can be viewed in supplementary demo video.*

Table 1. Quantitative experimental results compared with MoCoGAN [2] and MDGAN [4] on the Beach dataset. The up arrow indicates that the larger the value, the better the model performance, and vice versa.

Method	PSNR ↑	SSIM ↑	Flow-MSE ↓
MDGAN [4]	16.195	0.802	1.046
MoCoGAN [2]	21.413	0.826	0.822
Ours	**26.228**	**0.879**	**0.764**

Quantitative Results. We similarly choose PSNR, SSIM, and Flow-MSE metrics to quantitatively evaluate the effectiveness of our proposed method on the Beach dataset. As shown in Table 1, our approach gains +10.033 and +4.815 improvements for PSNR as well as +0.077 and +0.053 for SSIM compared to MDGAN and MoCoGAN, respectively. For the Flow-MSE metric, our method achieves the lowest value, *i.e.* 0.764, which means that our generated video sequences are the closest to the ground truth videos in terms of the motion. On the whole, evaluation results indicate that our approach outperforms other two baselines in all three metrics on the Beach dataset, which also illustrates that our model can generate more high-quality and dynamic videos than other methods.

2 Detailed Structure and Parameter of DTVNet

The network structures and parameters of the proposed *Optical Flow Encoder* (OFE, ψ), *Dynamic Video Generator* (DVG, ϕ), and the discriminator (D) are detailedly shown in the supplementary material by the Py-Torch framework [1], *c.f.* **DTVNet.py**. Complete codes for training and testing will be made available upon publication.

References

1. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch (2017)
2. Tulyakov, S., Liu, M.Y., Yang, X., Kautz, J.: Mocogan: Decomposing motion and content for video generation. In: CVPR. pp. 1526–1535 (2018)
3. Vondrick, C., Pirsiavash, H., Torralba, A.: Generating videos with scene dynamics. In: NeurIPS. pp. 613–621 (2016)
4. Xiong, W., Luo, W., Ma, L., Liu, W., Luo, J.: Learning to generate time-lapse videos using multi-stage dynamic generative adversarial networks. In: CVPR. pp. 2364–2373 (2018)