

# RAVEN: RETHINKING ADVERSARIAL VIDEO GENERATION WITH EFFICIENT TRI-PLANE NETWORKS

## Supplementary Material

### 1. REAL DATA

Along with this PDF in our supplementary material, you will find a .zip archive that holds sample videos from our training dataset. Note that we face similar problems finding suitable training data as noted in [1, 2]. The problems namely are — modern datasets of videos are either too small (only a few 100s videos (Fashion videos, FaceForensics, CelebV-HQ)), too simple ((Synthetically generated videos of NeRF objects)No object articulation, only camera motion, etc.), or too complex (very wide collection of videos of random scenes (Kinetics-600, UCF-101)). Therefore, we use clever ‘animation’ techniques to increase data volume, as described in the datasets subsection of the experiments section. However, the methods we use are not perfect and produce artifacts depending upon the source-appearance and motion. Furthermore, we found with a manual inspection that these methods [3, 4], although state-of-the-art, have the effect of generating videos with limited motion. This is then carried over to the final result of our method. Therefore, for manual judgment of generation quality, we highly recommend an inspection of the training data. In the directory named real\_samples, there are 3, videos corresponding to each used datasets. Every video is a  $10 \times 6$  grid showing 60 different samples.

### 2. GENERATED DATA

In the directory named generated\_samples, there are 3 subdirectories corresponding to each dataset. Each of these subdirectories in turn contains 4 videos generated using MoCoGAN, StyleGAN-V, Stable Video Diffusion and Ourmethod. Each video is a grid of  $10 \times 6$ , showcasing 60 random samples.

### 3.

#### EXTRAPOLATION WITH TRIPLANE REPRESENTATION

Here in table 1 and table 2 for completeness we show the efficacy of our proposed triplane+flow mechanism additionally on the UCF101 dataset.

### 4. REFERENCES

- [1] T. Brooks, J. Hellsten, M. Aittala, T. c. Wang, T. Aila, J. Lehtinen, M.-Y. Liu, A. A. Efros, and T. Karras, “Generating long videos of dynamic scenes,” in *NIPS*, 2022. 1
- [2] I. Skorokhodov, S. Tulyakov, and M. Elhoseiny, “Stylegan-v:

Method	UCF 101			
	3 frame intrp.		8 frame intrp.	
	SSIM	PSNR	SSIM	PSNR
Pos. emb	0.19	14.58	0.18	14.49
voxels	0.53	20.61	0.53	20.31
triplane	0.58	20.99	0.57	20.65
triplane+flow	<b>0.59</b>	<b>21.33</b>	<b>0.58</b>	<b>20.81</b>

**Table 1: Tri-plane video representation for interpolation:** As in table 1 of the main paper, we show here the interpolation performance of our triplane+flow framework on UCF dataset.

Method	UCF 101			
	3 frame intrp.		8 frame intrp.	
	SSIM	PSNR	SSIM	PSNR
Pos. emb	0.17	14.29	0.21	14.39
voxels	0.46	19.54	0.46	19.57
triplane	0.50	19.87	0.50	19.88
triplane+flow	<b>0.51</b>	<b>20.25</b>	<b>0.52</b>	<b>20.11</b>

**Table 2: Tri-plane video representation for extrapolation:** As in table 2 of the main paper, we show here the extrapolation performance of our triplane+flow framework on UCF dataset.

A continuous video generator with the price, image quality and perks of stylegan2,” in *CVPR*, 2022. 1

- [3] J. Zhao and H. Zhang, “Thin-plate spline motion model for image animation,” in *CVPR*, 2022. 1
- [4] W.-Y. Yu, L.-M. Po, R. C. Cheung, Y. Zhao, Y. Xue, and K. Li, “Bidirectionally deformable motion modulation for video-based human pose transfer,” in *ICCV*, 2023. 1