

# Supplementary-1 RV-GAN: Recurrent GAN for Unconditional Video Generation

Sonam Gupta      Arti Keshari      Sukhendu Das  
Visualization and Perception Lab, Department of Computer Science  
Engineering, Indian Institute of Technology, Madras, India

`cs18d005@cse.iitm.ac.in, cs19s008@cse.iitm.ac.in, sdas@iitm.ac.in`

## 1. Appendix

In this appendix, we provide more examples of the videos generated by our method for longer video sequence on Weizmann action dataset and MUG dataset. We have shown linear interpolation of video samples in latent space in figure 3

### 1.1. Longer Video Sequence Generation

For longer video generation, we trained TransGAN for 16 frames but at the inference time unrolled the RNN to generate 32 frames. The results are presented in Figures 1 and 2. We can see that our model is able to capture the spatio-temporal context well, giving temporally consistent videos.

### 1.2. Linear Interpolation of noise

Since the Weizmann Action dataset is very small, to verify that our model is not over-fitting and not remembering the data sample. In Figure 3, we have shown linear interpolation between data points in latent space to prove that our model is able to generate new data samples. Middle sequence of frames are generated using intermediate noise vector between top and bottom sequence of frames.

## 2. Visual Results

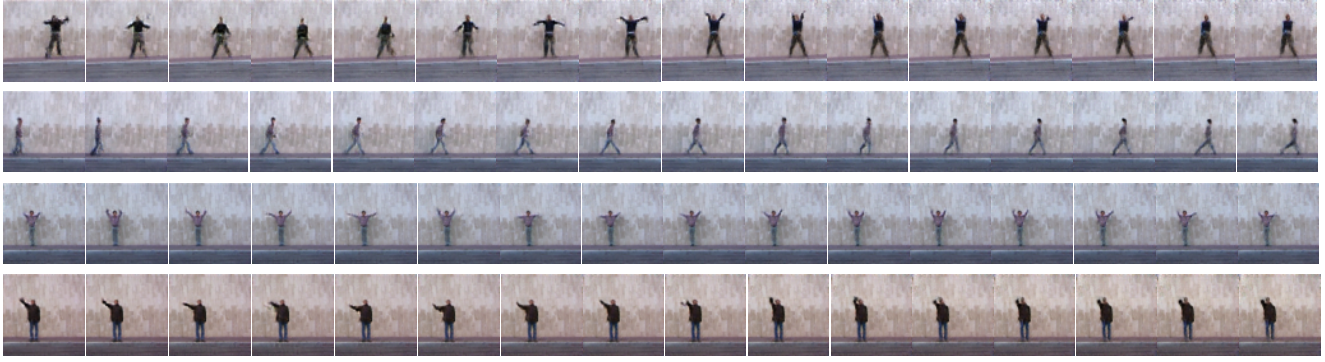


Figure 1. **Longer Video Sequence Generation:** Video Sequences of length 32 on Weizmann Dataset. Frames are sampled at equal interval for the purpose of illustration. Each row represents a video.

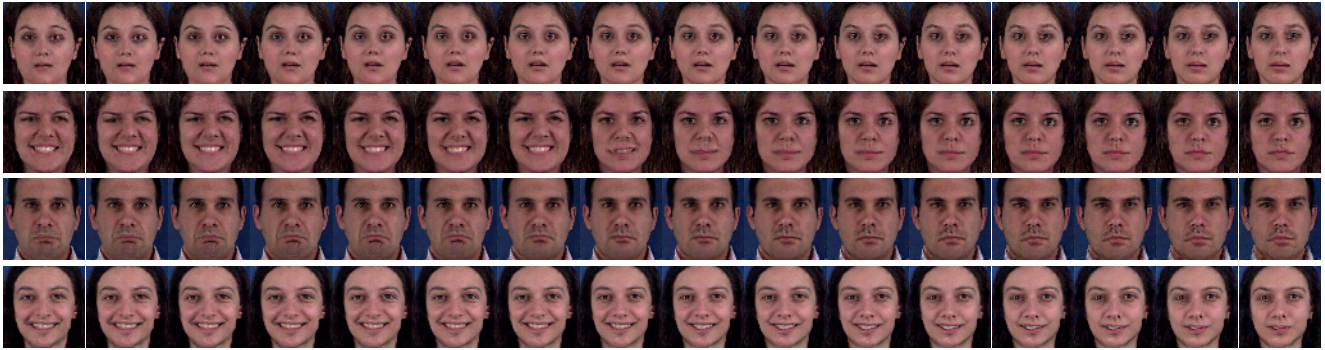


Figure 2. **Longer Video Sequence Generation:** Video Sequences of length 32 on MUG Dataset. Frames are sampled at equal interval for the purpose of illustration. Each row represents a video.



Figure 3. **Linear interpolation of input noise vectors:** First and last video sequence are generated by  $z_1$  and  $z_2$ , the middle row video sequence is generated by linearly interpolating  $z_1$  and  $z_2$ . Each row represents a video. Continuous frames are picked for visualization purpose.