

Supplementary 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 the details of the process used for latent space exploration on MUG dataset in case of class conditional video generation. A [MP4 video file](#) has also been uploaded separately, for visualization of the results.

2. Latent Space Exploration

To understand the latent space, we study the effect of merging two facial expressions or two people’s faces.

2.1. Average of Two Expressions

We show the video samples generated by merging two expressions. To do this, we choose a noise vector z , and separately concatenate one hot vector embedding corresponding to class 1 (h_1) and class 2 (h_2), to generate the base videos. Then we take the average of h_1 and h_2 and concatenate it with z to generate the new video. This new video yields an expression as that between classes 1 and 2. Figure 1 shows that our model successfully merges two different expressions. For example, combining disgust and surprise produces an appalling expression, and so on.

2.2. Average of Two Faces

For merging two people’s faces with same facial expression, we randomly sample two noise vectors z_1 and z_2 and generate base videos that corresponds to z_1 and z_2 . We take the average of z_1 and z_2 and use it to generate the third video. It can be clearly seen in Figure 2 that the new video (in the third row) contains facial features of both the base videos (first and second rows). These experiments were carried out on the trained model of conditional video generation.

3. Visual Results



Figure 1. **Latent space exploration combining expressions:** Expression classes are given above the corresponding video sequence. Third row in each example represents combined expression of first and second row. Each row represents a video. Continuous frames are picked for visualization purpose.

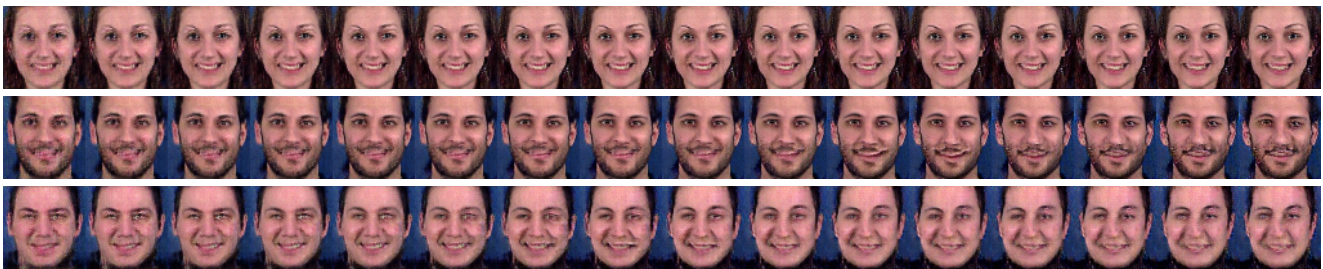


Figure 2. **Latent space exploration combining faces:** First and second row video sequence are generated by z_1 and z_2 , the third row video sequence is generated by taking average of z_1 and z_2 . Above video sequences are generated for happiness class. Each row represents a video. Continuous frames are picked for visualization purpose.