Research Paper using GANs for Video Prediction :

1) FutureGAN: Anticipating the Future Frames of Video Sequences using Spatio-Temporal 3d Convolutions in Progressively Growing GANs
https://arxiv.org/pdf/1810.01325.pdf

Github Code : https://github.com/TUM-LMF/FutureGAN

Encoder-decoder GAN model, FutureGAN, that predicts future frames of a video sequence conditioned on a sequence of past frames. During training, the networks solely receive the raw pixel values as an input, without relying on additional constraints or dataset specific conditions. To capture both the spatial and temporal components of a video sequence, spatio-temporal 3d convolutions are used in all encoder and decoder modules. PGGAN is used to achieve high-quality results on generating high-resolution single images

2) Stochastic Adversarial Video Prediction (savp)

https://arxiv.org/pdf/1804.01523.pdf

Github Code: https://alexlee-gk.github.io/video_prediction/

Combining these two methods: (a) latent variational variable models that explicitly model underlying stochasticity and (b) adversarially-trained models that aim to produce naturalistic images

3) Videoflow: a conditional flow-based model for stochastic video generation (videoflow)

https://arxiv.org/pdf/1903.01434v3.pdf

Github Code : https://github.com/fatemehazimi990/Pytorch-VideoFlow

They propose multi-frame video prediction with normalizing flows, which allows for direct optimization of the data likelihood, and produces high-quality stochastic predictions. They describe an approach for modeling the latent space dynamics, and demonstrate that flow-based generative models.

4) Generating Videos with Scene Dynamics

https://dl.acm.org/doi/pdf/10.5555/3157096.3157165

Github Code: https://github.com/GV1028/videogan

They propose a generative adversarial network for video with a spatio-temporal convolutional architecture that untangles the scene's foreground from the background.

5) Improving video generation for multi-functional applications:

https://github.com/bernhard2202/improved-video-gan

https://arxiv.org/pdf/1711.11453.pdf

6) MoCoGAN: Decomposing Motion and Content for Video Generation

https://arxiv.org/pdf/1707.04993.pdf

https://github.com/sergeytulyakov/mocogan

Side Note : (The Code is not available explicitly but you can run it) (Simplified Network)

From here to there: Video in between using direct 3d convolutions.
https://arxiv.org/pdf/1905.10240.pdf

Github Code
https://github.com/tensorflow/hub/blob/master/examples/colab/tweening_conv3d.ipynb

We consider the problem of generating plausible and diverse video sequences, when we are only given a start and an end frame. In this paper, we propose instead a fully convolutional model to generate video sequences directly in the pixel domain. We first obtain a latent video representation using a stochastic fusion mechanism that learns how to incorporate information from the start and end frames. Our model learns to produce such latent representation by progressively increasing the temporal resolution, and then decode in the spatiotemporal domain using 3D convolutions. The model is trained end-to-end by minimizing an adversarial loss.

Research Papers related to Audio Generation using Gans

1) Audio inpainting with generative adversarial network

https://arxiv.org/pdf/2003.07704.pdf
Github Code: https://github.com/nperraud/gan_audio_inpainting

This focuses on audio inpainting in general using GANs, having validated their results on three different datasets of different musical instruments. They base their architecture on the Wasserstein GAN

2) Adversarial generation of time-frequency features with application in audio synthesis

https://arxiv.org/abs/1902.04072

https://github.com/tifgan/stftGAN

They demonstrate the potential of deliberate generative TF modeling by training a generative adversarial network (GAN) on short-time Fourier features (STFT) . They show that by applying their guidelines, their TF-based network was able to outperform a state-of-the-art GAN generating waveforms directly, despite the similar architecture in the two networks

3) Adversarial audio synthesis

https://arxiv.org/pdf/1802.04208.pdf

Github Code: https://github.com/chrisdonahue/wavegan

WaveGAN, an attempt at applying GANs to unsupervised synthesis of raw-waveform audio.

4) Wave Glow: A flow-based generative network for speech synthesis. *CoRR*, abs/1811.00002, 2018

https://arxiv.org/abs/1811.00002

https://github.com/NVIDIA/waveglow

In this paper we propose WaveGlow: a flow-based network capable of generating high quality speech from mel spectrograms. WaveGlow is implemented using only a single network, trained using only a single cost function: maximizing the likelihood of the training data, which makes the training procedure simple and stable. .

5) SEGAN: speech enhancement generative adversarial network :

https://arxiv.org/abs/1703.09452

Github Code : https://github.com/santi-pdp/segan

This is based on a GAN variant called Speech Enhancement GAN (SEGAN) that operates in the time domain by producing raw audio signals directly. The SEGAN generator is constructed as an Autoencoder, where the audio is encoded by using successive convolutional layers into a vector. This is concatenated with a vector of random noise and together, they are passed to the decoder which has a mirrored structure to that of the encoder. The decoder learns to recreate an enhanced version of the audio input to the encoder. In order to not lose low-level details of the input audio, the authors use skip connections between the corresponding layers of the encoder and the decoder to allow information such as phase or alignment to pass. On the other hand, given a pair of an impaired speech and its enhanced version, the discriminator D is trained to classify if the enhanced speech is real (actually from the dataset) or fake (bad imitation of the dataset).

6) A context encoder for audio inpainting (Does not use GANs)

https://arxiv.org/pdf/1810.12138.pdf

Github Code : https://github.com/andimarafioti/audioContextEncoder

They propose a DNN structure that is provided with the signal surrounding the gap in the form of time-frequency (TF) coefficients. Two DNNs with either complex-valued TF coefficient output or magnitude TF coefficient output were studied by separately training them on inpainting two types of audio signals (music and musical instruments) having 64-ms long gaps. (works especially for music instruments)

7) Audio super-resolution using neural nets

https://arxiv.org/pdf/1708.00853.pdf
Github Code : https://github.com/kuleshov/audio-super-res

They introduce a new audio processing technique that increases the sampling rate of signals such as speech or music using deep convolutional neural networks. Their model is trained on pairs of low and high-quality audio examples; at test-time, it predicts missing samples within a low-resolution signal in an interpolation process similar to image super-resolution