

Text to Image/Video Synthesis Using GANs

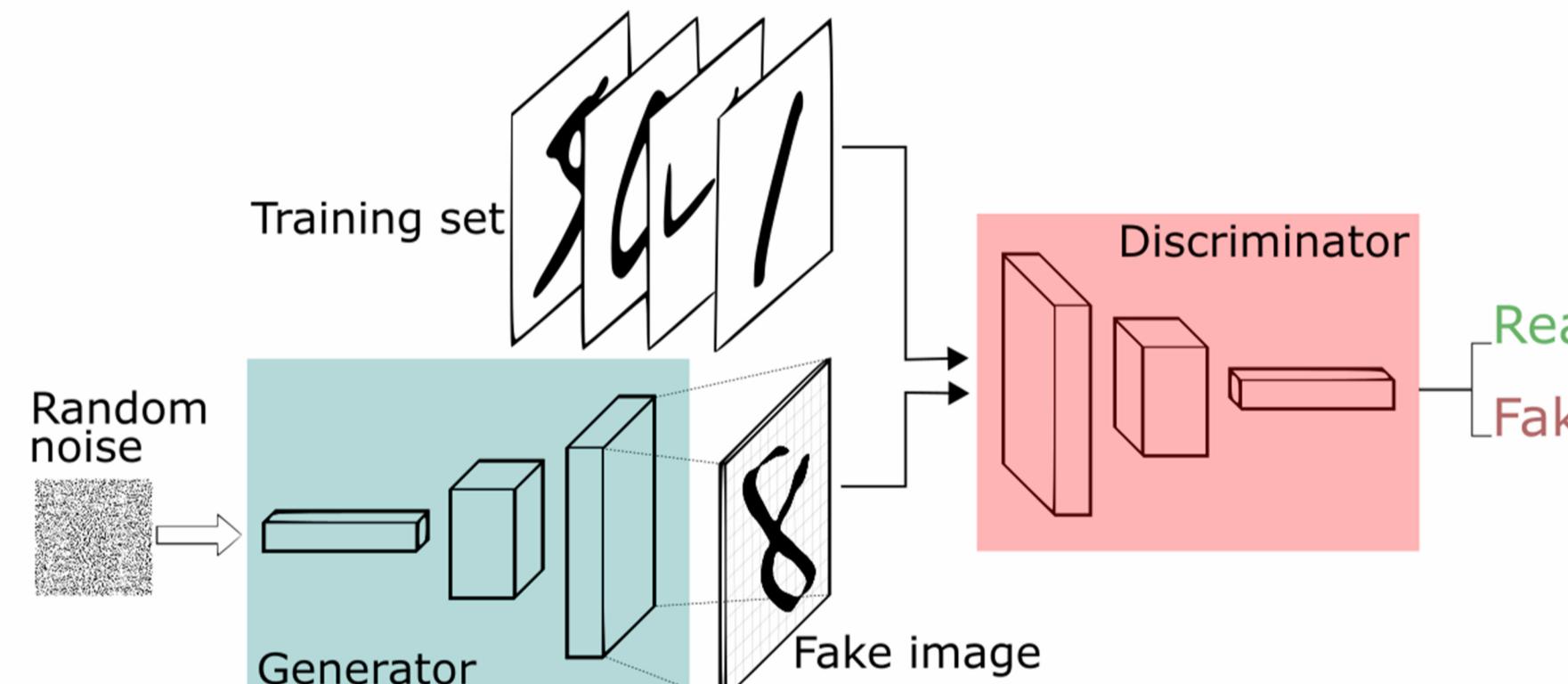
Deval, Aishwarya Mentor: Kalind Karia, Aditya Panwar

Objective:

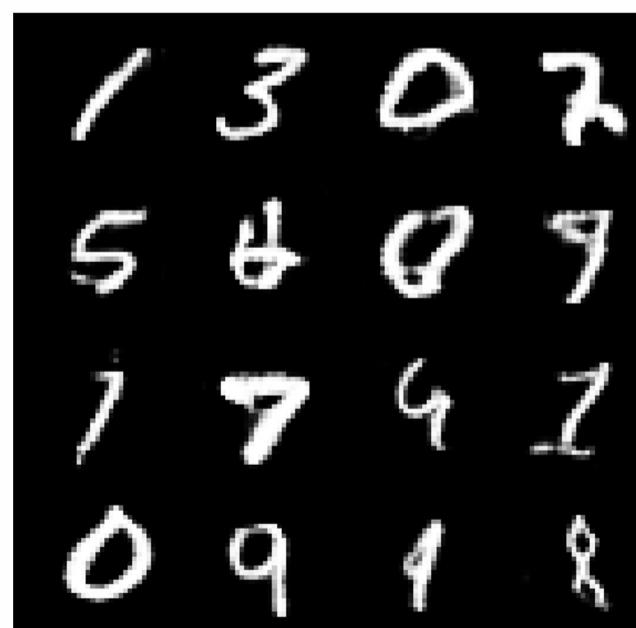
To generate images or video from a given text caption using Generative adversarial networks.

What are GANs?

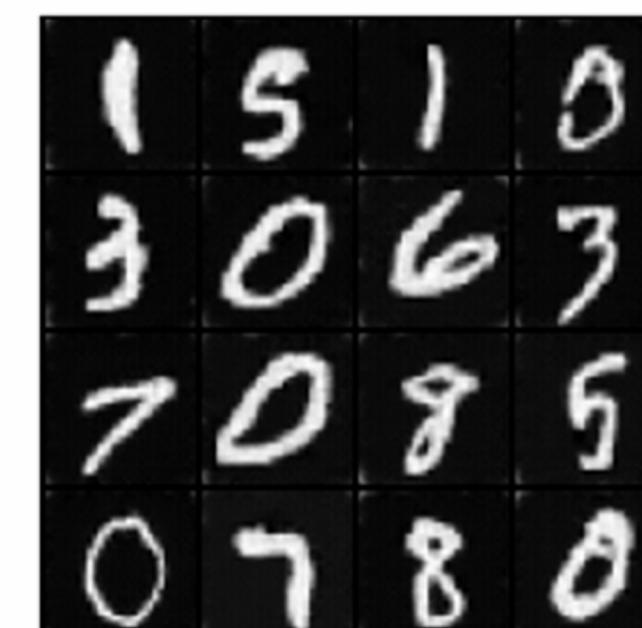
Generative adversarial networks (GANs) are deep neural net architectures comprised of two nets pitting one against the other while both of them learn.



Some of our generated examples on MNIST dataset using a DCGAN model which is essentially a GAN with a convolutional network.



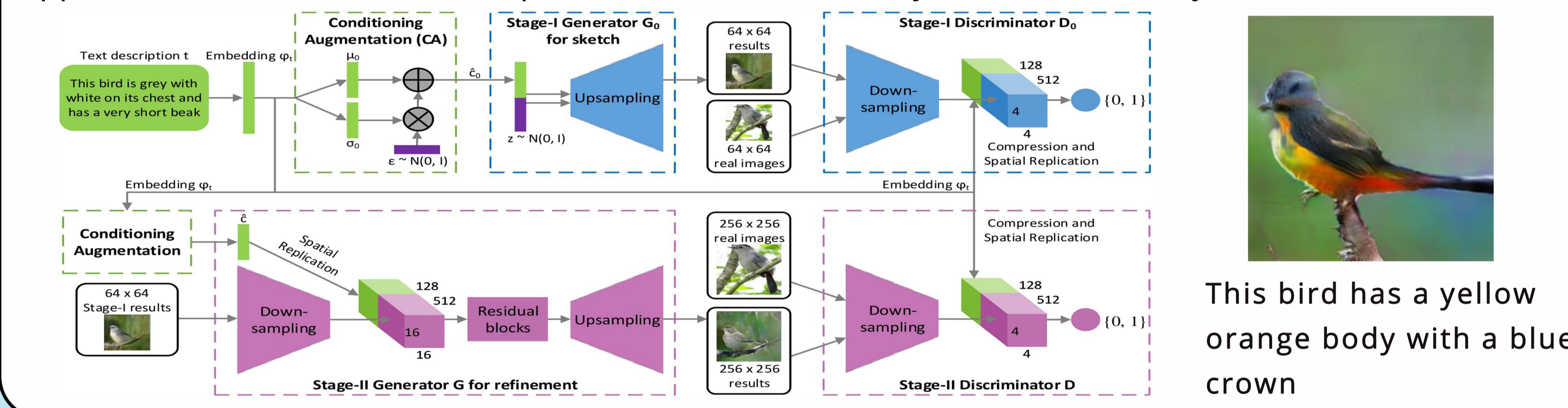
Actual



Generated

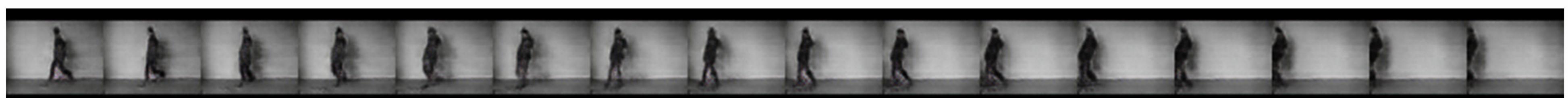
Text to Image Generation

We have used a stacked GAN model for image generation where the first GAN generates a low resolution image from noise and a text encoding then the second GAN produces higher resolution images. This approach allows the model to capture text details that may have been lost by the first GAN.



Text to Video Generation

We have used MoCoGAN(motion decomposition GAN) to generate videos based on categories which can be actions or shapes. MoCoGan uses a RNN to generate the latent variables, each of these variables correspond to a frame in a video and are fed to an image generator which creates the video.



The Category here was walking



The Category here was handwaving

The images above are a concatenation of video frames