Introduction

Generative models are designed to create new data that closely resembles data from a dataset. By using the patterns in the data they understand and generate realistic outputs. The three types of generative models we are looking at are Conditional Generative Adversarial networks, Variational Autoencoders, and Generative Adversarial networks. We used the MNIST dataset, which contains images of grayscale handwritten digits ranging from 0-9. It's ideal for exploring and comparing these models due to its simplicity and structure.

Differences

Conditional Generative Adversarial Networks (CGANs)

CGANs extend GAN by including a condition which allows it to generate specific types of images. To generate an image of the digit “9” you would input a label “9” along with random noise. CGANs are ideal for controlled generation however it limits the variety of the outputs given the condition. Training involves two parts: a generator (which produces images) and a discriminator (which distinguishes real images from generated ones). The generator and discriminator train in a loop to improve each other.

Variational Autoencoders (VAEs)

VAEs encode images into a compressed latent space which captures the important features of the data and then decodes it back into images. The structure enables VAE to explore smooth transitions between different images in the latent space (like blending). It focuses on structural coherence which can output a blurry image. VAE training has two objectives: to reconstruct images and organize the latent space.

Generative Adversarial Networks (GANs)

GANs generate images by transforming random noise into realistic outputs. Unlike the CGANs they do not have control so images are determined by the random noise as input. GANs have a generator and discriminator and they train adversarially (generator tries to generate the most realistic fake data while the discriminator tries to discriminate between fake and real data). GANs are very good at producing sharp images however they do not generate specific image types due the absence of a condition.

Comparison of Outputs

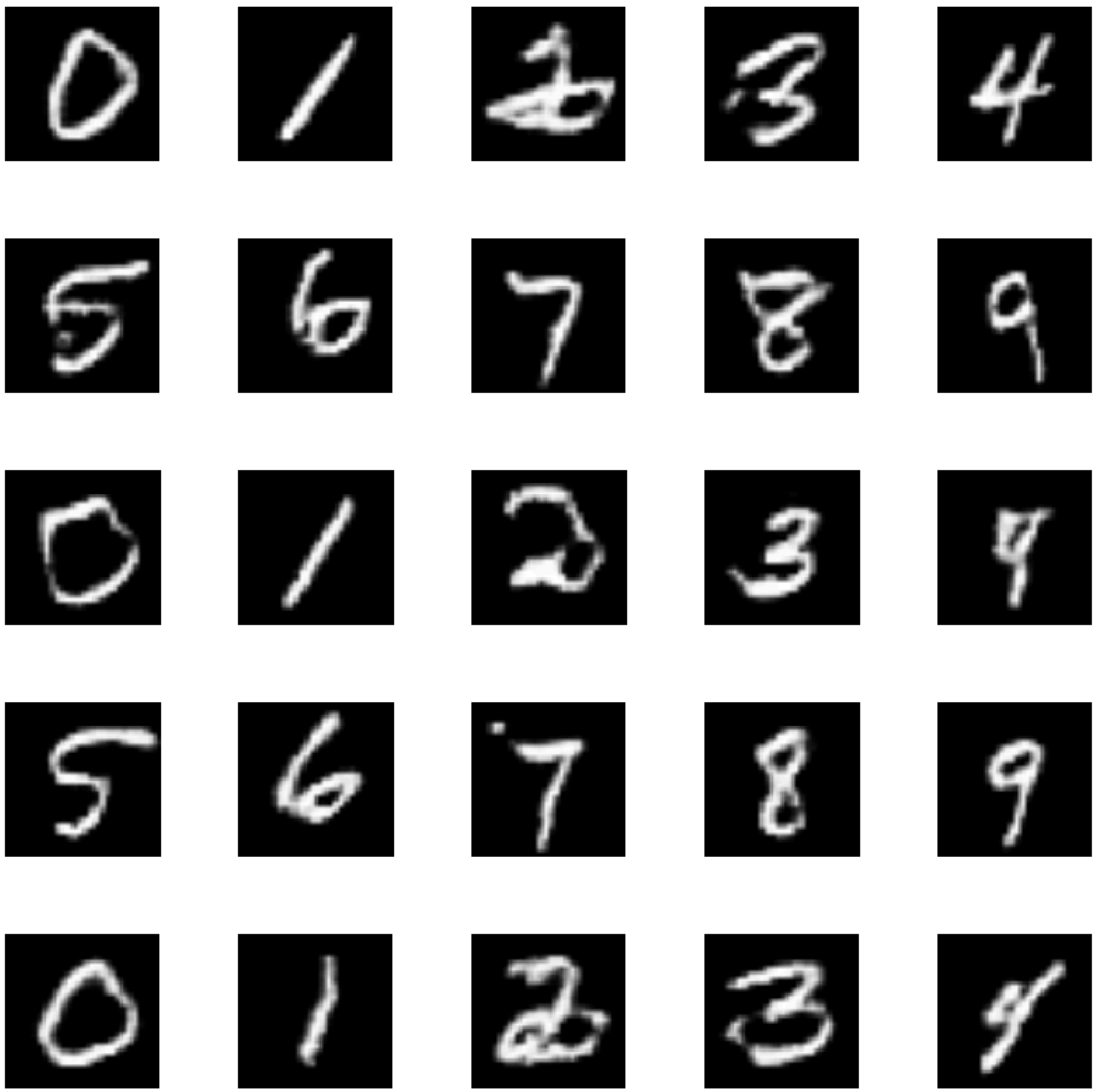
CGAN
The digits are fairly crisp and well defined. Most digits closely resemble their counterparts, except for 2 which it seemed to struggle a lot with. Each set of 2 rows has the digits from 0-9 which show that they followed the labels correctly. This demonstrates the strength in conditional generation. The twos and one of the fours have distortion however only the 2s have enough distortion to cause issue. The output is quite sharp and has minimal blurring when compared to the VAEs but not as sharp as the GAN. The digits within the labels look fairly consistent however they do not have a ton of different stylistic choices, like all the 3s look roughly the same. CGAN excels in generating specific digits and images are fairly sharp however they lack stylistic diversity and some digits have irregularities.

VAE
The digits exhibit significant blurriness, which is characteristic of VAEs. This occurs because VAEs balance two competing objectives: reconstruction quality and latent space regularization. The blurriness suggests the model prioritizes smoothness in the latent space at the cost of sharpness in the images. The images are moderately realistic. While most digits are identifiable, some digits, such as “2" appear distorted. There is noticeable variability within each row. For example, some digits like "8" are well-defined, while others like "7" seem less distinct. VAEs are excellent for exploring smooth transitions between digits. The latent space's continuous nature allows gradual morphing between different digit types.

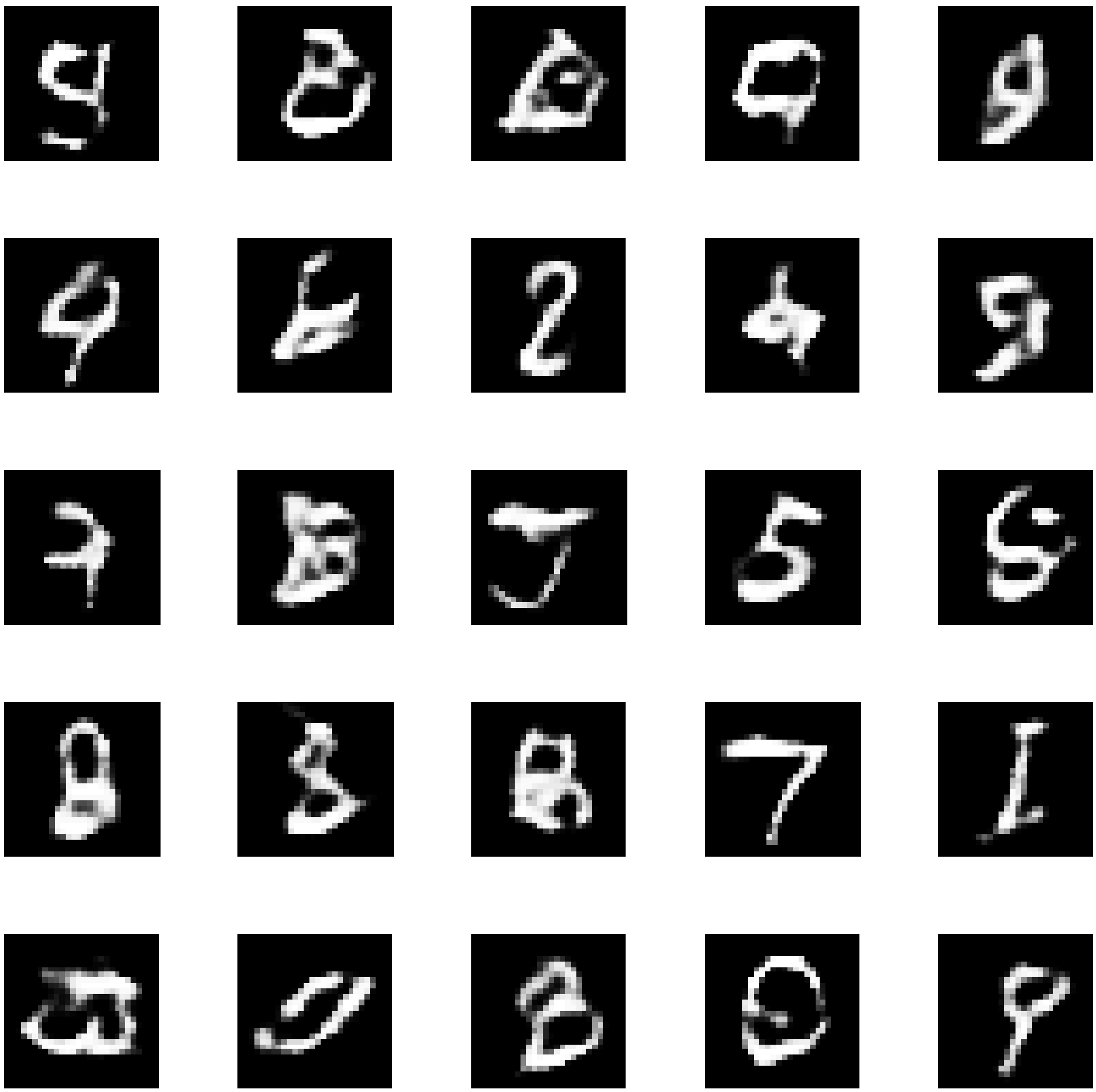
GAN
The images are significantly sharper compared to the VAE outputs but slightly less consistent than CGAN outputs. Almost all the digits are malformed and suffer from extreme distortion making them unrecognizable. The digits “5” and “8” seem to be some of the only recognizable digits. The lack of a conditional input means the generated digits are influence purely by the noise. The GAN produces sharp images however most of the digits are do not look at all like their counterpart.

Generated Images

CGAN



GAN



VAE