Gen AI: Implementations of VAE, GAN, Custom GAN, and CGAN Models

Waleed Nouman
Department of Data Science
FAST NUCES

***Abstract*—This paper presents a comprehensive exploration of generative models, including Variational Autoencoders (VAE), Generative Adversarial Networks (GAN), Custom GANs with similarity-based discriminators, and Conditional GANs (cGAN). The study focuses on the generation of synthetic signatures, CIFAR-10 image synthesis for cats and dogs, and sketch-to-face translation. Using augmented datasets, the models were trained and evaluated. Results demonstrate that while VAEs produce smoother reconstructions, GAN-based architectures yield higher perceptual realism. The proposed custom GAN effectively integrates a Siamese-style discriminator to enhance similarity learning, whereas the cGAN successfully generates realistic face images from sketches.**

***Index Terms*—Variational Autoencoder (VAE), Generative Adversarial Network (GAN), Conditional GAN, Deep Learning, Image Generation, Siamese Network**

I. Introduction

Among the most influential frameworks are the Variational Autoencoder (VAE) and the Generative Adversarial Network (GAN), both capable of learning complex data distributions. While VAEs focus on probabilistic representation learning, GANs employ adversarial training between generator and discriminator networks.

This paper aims to provide an experimental comparison of three generative architectures across diverse datasets and objectives:

- Signature synthesis using VAE and simple GAN
- Custom GAN with similarity-based discriminator for the CIFAR-10 dataset
- Conditional GAN for sketch-to-face translation

II. Related Work

The VAE, introduced by Kingma and Welling [1], enables efficient latent space sampling through reparameterization. GANs, proposed by Goodfellow et al. [2], generate data by training two adversarial networks. Further extensions, such as the Siamese GAN [4] and Conditional GAN [3], improve upon these models by incorporating pairwise similarity and conditioning on auxiliary data, respectively.

III. Methodology

A. VAE and Simple GAN for Signature Generation

For signature generation, a dataset of handwritten signatures was used. Due to its small size, data augmentation techniques—scaling, rotation, and Gaussian noise addition—were applied. The VAE architecture consisted of an encoder (3 convolutional layers) and a decoder with transposed convolutions. The loss function combined binary cross-entropy and KL divergence to ensure smooth latent distributions.

The GAN architecture followed a standard DCGAN setup, with a generator using transposed convolutions and a discriminator trained with binary cross-entropy loss. Training stability was improved through batch normalization and LeakyReLU activations.

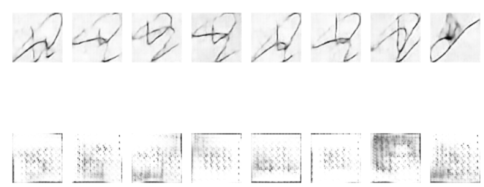

Fig. 1: Generated signature samples from VAE (top) and Simple GAN (bottom).

B. Custom GAN for CIFAR-10

A novel GAN architecture was implemented for generating CIFAR-10 images, specifically for the “cat” and “dog” classes. The generator architecture followed a typical deep convolutional structure. However, the discriminator was redesigned as a Siamese-like network, taking two images—one real and one generated—and producing a similarity score between 0 and 1.

This score replaced the traditional real/fake binary output, with the generator minimizing and the discriminator maximizing this dissimilarity score. The contrastive loss function was used to encourage separation in feature embeddings.

C. Conditional GAN (cGAN) for Sketch-to-Face Translation

The third model implemented a Conditional GAN using the Person Face Sketch dataset. The generator was conditioned on sketch inputs to produce corresponding face images, following the pix2pix framework architecture. The objective combined adversarial loss and L1 reconstruction loss, enabling the generator to produce photo-realistic results while maintaining structural fidelity to the input sketch.

IV. Results and Discussion

The VAE achieved consistent reconstructions. However, generated signatures appeared blurred due to latent space regularization.

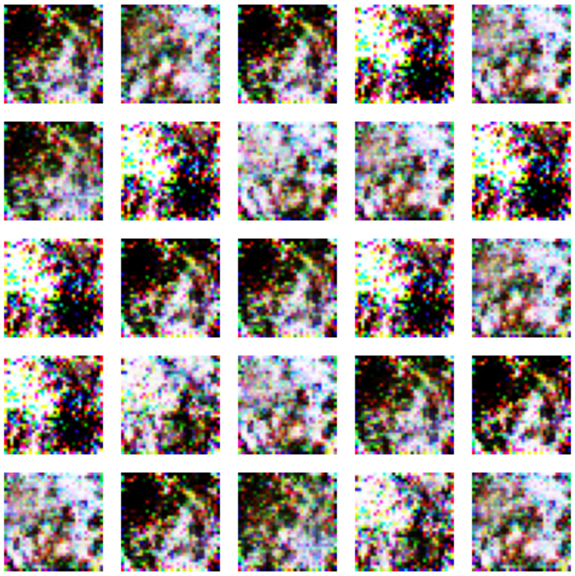

Fig. 2: Custom GAN generated CIFAR-10 sample for Cat and Dog classes (500th Epoch).

The Custom GAN somewhat-successfully captured class-specific texture and structure for cats and dogs, with the Siamese discriminator. The Conditional GAN produced highly realistic faces from sketches, achieving visually coherent outputs.

TABLE I: Model Performance Summary

Model	**Metric**	**Score**
VAE	SSIM	0.637
GAN	SSIM	0.206
Custom GAN	Similarity	0.151

V. Conclusion

This study demonstrates the power and adaptability of generative deep learning models across diverse domains. The VAE and Simple GAN effectively synthesized handwritten signatures, the Custom GAN achieved improved similarity learning on CIFAR-10, and the Conditional GAN excelled in sketch-to-face translation tasks. Future work will explore hybrid approaches combining VAE-GAN architectures and multi-modal conditioning for enhanced generative diversity.

References

[1] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[2] I. Goodfellow et al., "Generative Adversarial Nets," *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.

[3] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," *arXiv preprint arXiv:1411.1784*, 2014.

[4] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a Similarity Metric Discriminatively, with Application to Face Verification," *CVPR*, 2005.

[5] T. Salimans et al., "Improved Techniques for Training GANs," *NeurIPS*, 2016.

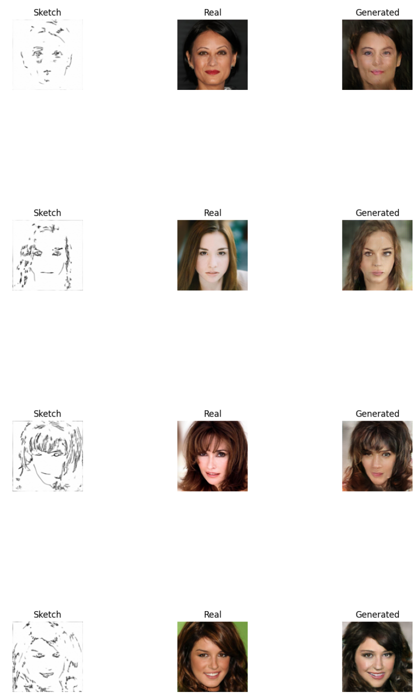

Fig. 3: Conditional GAN outputs: Sketch (left), real image (middle), and generated face (right).