

Generative Models for Signature Synthesis and Face Translation: A Comparative Study of VAE, GAN, Conditional GAN, and CycleGAN

Laiba Ameer

laibaameerofficial@gmail.com

National University of Computer and Emerging Sciences

Research Article

Keywords: Variational Autoencoder, Generative Adversarial Network, Conditional GAN, CycleGAN, Signature Generation, Image-to-Image Translation, Deep Learning

Posted Date: August 11th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10638512/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Generative Models for Signature Synthesis and Face Translation: A Comparative Study of VAE, GAN, Conditional GAN, and CycleGAN

Abstract

This paper presents a comparative implementation of four generative modeling approaches: a Variational Autoencoder (VAE), a standard Generative Adversarial Network (GAN), a Conditional GAN (cGAN), and CycleGAN. The experiments address three image-generation settings: signature synthesis, cat-and-dog image generation using a CIFAR-10 subset, and bidirectional translation between facial sketches and face photographs. The study describes dataset preparation, preprocessing, model design, training objectives, and evaluation procedures. In the reported experiments, the VAE obtained a test reconstruction loss of 0.015, the signature GAN discriminator achieved 85% classification accuracy, the custom CIFAR-10 GAN obtained a similarity score of 0.72, and the cGAN achieved an average pixel-similarity score of 0.80. For CycleGAN, the cycle-consistency loss stabilized after approximately 50 epochs and the generated translations became visually more coherent. The experiments highlight the strengths of latent-variable, adversarial, conditional, and cycle-consistent learning, while also illustrating common challenges such as adversarial instability, sensitivity to hyperparameters, and the limitations of pixel-level evaluation for perceptual image quality.

Variational Autoencoder, Generative Adversarial Network, Conditional GAN, CycleGAN, Signature Generation, Image-to-Image Translation, Deep Learning.

1 Introduction

Generative models learn the underlying structure of observed data and use that structure to synthesize new samples. Among the most influential approaches are Variational Autoencoders (VAEs) ([kingma2014vae](#)) and Generative Adversarial Networks (GANs) ([goodfellow2014gan](#)). VAEs formulate generation as probabilistic latent-variable learning, whereas GANs train a generator and discriminator in an adversarial game. Conditional GANs extend the adversarial framework by introducing side information into both networks ([mirza2014cgan](#)), enabling controlled generation. CycleGAN further extends adversarial learning to unpaired image-to-image translation through cycle-consistency constraints ([zhu2017cyclegan](#)).

This work compares these four approaches across multiple visual-generation tasks. First, VAE and GAN models are used to synthesize signature-like images. Second, a custom GAN is trained on cat and dog classes selected from CIFAR-10. Third, a cGAN is used for sketch-to-face translation, where the sketch explicitly conditions the generated output. Finally, CycleGAN is applied to bidirectional translation between sketches and face photographs.

The main objectives are to: (1) implement representative generative architectures; (2) preprocess and adapt datasets to the requirements of each model; (3) compare model behavior using the reported experimental metrics and qualitative observations; and (4) identify practical difficulties encountered during generative-model training.

2 Background

2.1 Variational Autoencoder

A VAE contains an encoder that approximates the posterior distribution over a latent variable z and a decoder that reconstructs an observation x from a sampled latent representation (**kingma2014vae?**). Its objective can be written as

$$\mathcal{L}_{\text{VAE}} = \mathcal{L}_{\text{rec}} + \beta D_{\text{KL}}(q_{\phi}(z|x) \parallel p(z)),$$

where $\mathcal{L}_{\text{rec}}$ measures reconstruction error and the Kullback–Leibler term regularizes the approximate posterior toward a prior distribution. In the standard formulation, $\beta = 1$.

2.2 Generative Adversarial Network

A GAN consists of a generator G and discriminator D trained competitively (**goodfellow2014gan?**). The classical minimax objective is

$$\min_G \max_D \mathbb{E}_{x \sim p_{\text{data}}} [\log D(x)] + \mathbb{E}_{z \sim p_z} [\log(1 - D(G(z)))].$$

The generator learns to produce samples that resemble the training distribution, while the discriminator learns to distinguish real samples from generated ones.

2.3 Conditional GAN

A cGAN introduces conditioning information c into the generator and discriminator (**mirza2014cgan?**). For sketch-to-face synthesis, the sketch acts as the condition, giving an objective of the form

$$\min_G \max_D \mathbb{E}_{x,c} [\log D(x, c)] + \mathbb{E}_{z,c} [\log(1 - D(G(z, c), c))].$$

This conditioning encourages the generated face to remain structurally consistent with the corresponding input sketch.

2.4 CycleGAN

CycleGAN learns mappings between two visual domains without requiring one-to-one paired training samples (**zhu2017cyclegan?**). Let $G: X \rightarrow Y$ and $F: Y \rightarrow X$ denote the two generators. In addition to adversarial losses, CycleGAN uses the cycle-consistency term

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_X} [\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_Y} [\|G(F(y)) - y\|_1].$$

The constraint encourages a sample translated to the opposite domain and back again to retain its original content.

3 Methodology

3.1 Datasets

Three data configurations were used in the experiments.

- **Signature dataset:** Signature images were used to train the VAE and standard GAN for synthetic-signature generation. Augmentation was applied to increase visual diversity and reduce overfitting.
- **CIFAR-10 subset:** Images belonging to the cat and dog classes were selected for the custom GAN experiment. CIFAR-10 images have a native resolution of 32×32 pixels.
- **Face-sketch dataset:** Face photographs and corresponding or domain-related sketches were used for sketch-to-face translation with cGAN and for bidirectional sketch/face translation with CycleGAN.

3.2 Preprocessing

The preprocessing pipeline was adapted to each architecture. Images were resized to the model input resolution and converted to tensors. Pixel values were normalized to a numerically convenient range consistent with the output activation of the relevant generator or decoder. Data augmentation included small rotations, scaling transformations, and noise perturbations where appropriate. For the CIFAR-10 experiment, the cat and dog classes were filtered before training.

3.3 VAE for Signature Generation

The VAE encoder transformed each signature image into the parameters of a latent probability distribution, represented by a mean vector μ and log-variance vector $\log \sigma^2$. Latent vectors were sampled using the reparameterization trick

$$z = \mu + \sigma \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I),$$

and passed to the decoder to reconstruct the signature. Training jointly minimized reconstruction error and KL divergence. After training, new signatures were synthesized by sampling from the latent prior and decoding the resulting vectors.

3.4 Standard GAN for Signature Generation

The signature GAN contained a generator that mapped random latent vectors to synthetic signature images and a discriminator that classified images as real or generated. Training alternated between discriminator updates and generator updates. The discriminator was optimized to correctly classify real and generated samples, while the generator was optimized to make generated signatures increasingly difficult to distinguish from real ones.

3.5 Custom GAN for CIFAR-10

The custom GAN targeted cat and dog image generation from the selected CIFAR-10 classes. In addition to adversarial learning, the discriminator-side design incorporated a Siamese-style comparison mechanism to estimate similarity between real and generated representations. This formulation aimed to encourage generated samples to resemble the target image distribution while providing an additional similarity-oriented training signal.

3.6 Conditional GAN for Sketch-to-Face Translation

For the cGAN experiment, the input sketch served as conditioning information. The generator learned a direct mapping from sketch structure to a synthesized face, while the discriminator evaluated whether a candidate face was realistic in the context of the supplied sketch. This conditional setup is more appropriate than unconditional generation when correspondence between an input structure and output image is required.

3.7 CycleGAN for Bidirectional Translation

CycleGAN was trained with two generators and two discriminators: one generator translated sketches to face images and the other translated face images to sketches. Each discriminator evaluated realism in its target domain. The total objective combined adversarial and cycle-consistency components, enabling bidirectional translation even when perfectly aligned image pairs were not required.

4 Evaluation Strategy

Generative modeling is difficult to evaluate with a single scalar metric; therefore, the experiments combined task-specific numerical measurements with qualitative inspection. The VAE was assessed using reconstruction loss. The signature GAN discriminator accuracy was recorded as a measure of discriminator behavior, although discriminator accuracy alone does not directly quantify generator perceptual quality. The custom GAN was evaluated using the reported similarity score, while the cGAN used an average pixel-similarity score between generated and reference images. CycleGAN behavior was examined through cycle-consistency loss and visual coherence of translated outputs.

Table 1 summarizes the reported measurements.

Summary of Reported Experimental Results

Model	Task	Reported Result
VAE	Signature generation	Test reconstruction loss: 0.015
GAN	Signature generation	Discriminator accuracy: 85%
Custom GAN	CIFAR-10 cats/dogs	Similarity score: 0.72
cGAN	Sketch-to-face	Mean pixel similarity: 0.80
CycleGAN	Sketch $\leftrightarrow$ face	Cycle loss stabilized after $\sim$ 50 epochs

5 Results

5.1 VAE and Signature GAN

The VAE produced recognizable signature-like reconstructions and achieved a test reconstruction loss of 0.015. The relatively low reported reconstruction loss indicates that the decoder was able to reproduce major structures present in the input signatures. Sampling from the learned latent space also produced new signature-like outputs, demonstrating the generative role of the latent representation.

The standard GAN generated synthetic signatures that became visually more plausible as training progressed. The discriminator achieved a reported accuracy of 85% when distinguishing real and generated signatures. This value should be interpreted primarily as a training diagnostic rather than a standalone measure of generator quality, because a strong or weak discriminator can reflect the current balance of the adversarial game rather than perceptual realism alone.

5.2 Custom GAN on CIFAR-10

The custom GAN generated cat- and dog-like images from the CIFAR-10 subset and achieved a reported similarity score of 0.72. Training was more difficult than in the reconstruction-based VAE because the generator and discriminator had to improve simultaneously. Incorporating a Siamese-style similarity mechanism introduced an additional optimization constraint and made the balance between adversarial and similarity objectives an important practical consideration.

5.3 Conditional GAN

The cGAN generated face images conditioned on input sketches and achieved an average pixel-similarity score of 0.80 in the reported evaluation. Conditioning improved structural correspondence because the generator was not required to synthesize an arbitrary face; instead, it learned to produce an output consistent with the provided sketch. Qualitative inspection indicated improved facial structure and visual plausibility over the course of training.

5.4 CycleGAN

CycleGAN learned transformations in both sketch-to-face and face-to-sketch directions. The cycle-consistency loss stabilized after approximately 50 epochs, and the translated outputs became progressively more coherent. The model was particularly useful for bidirectional domain translation because cycle consistency constrained the learned mappings even when direct sample-level correspondence was not assumed.

6 Discussion

The experiments demonstrate important differences among the four generative approaches. The VAE offered comparatively stable optimization because its training objective was directly differentiable and did not involve an adversarial opponent. Its latent space was useful for reconstruction and sampling, although VAE outputs can be visually smoother than adversarially generated images.

GAN-based models produced sharper and more visually expressive samples but were substantially more sensitive to optimization settings. The main challenge was maintaining an appropriate balance between generator and discriminator learning. If the discriminator became too strong, the generator received weak or uninformative gradients; if the discriminator was too weak, the generator could exploit it without learning a meaningful approximation of the target distribution.

The cGAN benefited from explicit conditioning because sketch information provided structural guidance to the generator. In contrast, CycleGAN addressed a different scenario: it used adversarial and cycle-consistency losses to learn translations between two domains without requiring strict paired supervision. This distinction is important when selecting a model for an image-to-image translation problem.

A further observation concerns evaluation. Reconstruction error and pixel similarity are straightforward to compute but do not fully capture perceptual realism. Likewise, discriminator accuracy is not a direct measure of generated-image quality. For a more rigorous future study, perceptual or distribution-level measures such as Fréchet Inception Distance (FID), Kernel Inception Distance (KID), or human evaluation could complement the metrics reported here.

7 Limitations

Several limitations should be considered when interpreting the results. First, the reported metrics are task-specific and therefore are not directly comparable across all four models. Second, pixel-level similarity can penalize perceptually valid outputs that differ slightly from a reference image. Third, adversarial models can exhibit instability, mode collapse, or sensitivity to learning rate, initialization, batch size, and loss weighting. Finally, stronger

conclusions would require repeated runs, explicit train/validation/test split statistics, hardware and training-time reporting, and standardized perceptual metrics.

8 Conclusion

This study implemented and compared VAE, GAN, cGAN, and CycleGAN architectures for signature synthesis, CIFAR-10 cat-and-dog generation, and face-sketch translation. The VAE demonstrated stable latent-variable learning and achieved a reconstruction loss of 0.015. The signature GAN produced increasingly realistic synthetic samples while its discriminator reached 85% reported accuracy. The custom CIFAR-10 GAN obtained a similarity score of 0.72, and the cGAN achieved a mean pixel-similarity score of 0.80 for sketch-conditioned face synthesis. CycleGAN produced coherent bidirectional translations, with its cycle-consistency loss stabilizing after approximately 50 epochs. Overall, the experiments show that the suitability of a generative model depends on the task: VAEs are effective for structured latent representations, GANs for adversarial synthesis, cGANs for condition-guided generation, and CycleGAN for unpaired domain translation.

9 Experimental Task Definitions

The implemented models were trained according to the following task definitions rather than natural-language prompts:

- **VAE and signature GAN:** learn the distribution of real signature images and synthesize new signature-like samples.
- **Custom CIFAR-10 GAN:** learn from cat and dog image classes and generate new samples that resemble those categories.
- **Conditional GAN:** transform an input facial sketch into a corresponding realistic face image.
- **CycleGAN:** learn bidirectional mappings between the facial-sketch and face-photo domains.

00

D. P. Kingma and M. Welling, “Auto-Encoding Variational Bayes,” in *Proc. 2nd Int. Conf. Learning Representations (ICLR)*, 2014.

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Nets,” in *Advances in Neural Information Processing Systems*, vol. 27, pp. 2672–2680, 2014.

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

J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, pp. 2223–2232, 2017.