Generative Adversarial Network

- Characteristics of good Evaluation Metrics –
  - Detect mode collapse – helps in evaluating models' ability to produce wide variety of realistic images
  - Align with human visual perception – high scoring images should appear realistic and diverse to human observers
  - Capture both intra- and inter-class variability
  - Less sensitive to (pre-trained) model parameters
  - Not dependent on distributions
- Popular metrics – Inception Score (IS), Fréchet Inception Distance (FID), Learned Perceptual Image Patch Similarity (LPIPS), Structural Similarity Index (SSIM) and Multi-Scale Structural Similarity Index (MS-SSIM)

Generative Adversaria

The Inception v3 Network (Szegedy et al., 2016) is a deep convolutional architecture designed for classification tasks on ImageNet (Deng et al., 2009), a dataset consisting of 1.2 million RGB images from 1000 classes. Given an image $\mathbf{x}$, the task of the network is to output a class label y in the form of a vector of probabilities $p(y|\mathbf{x}) \in [0,1]^{1000}$,

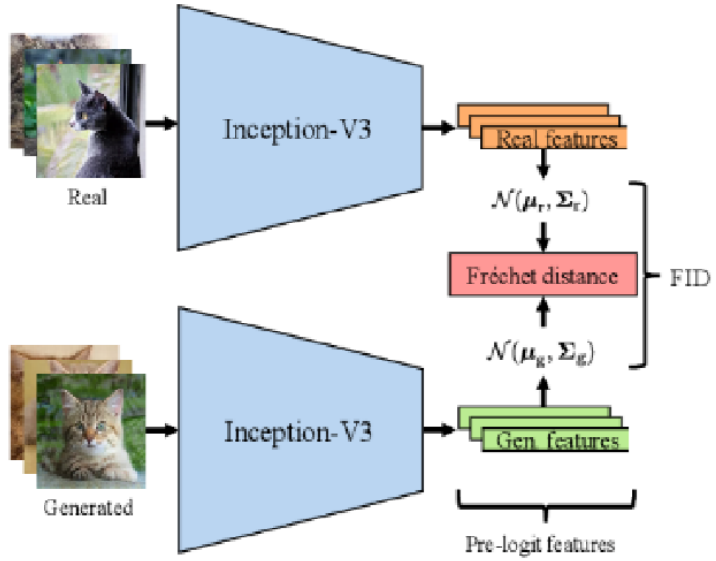
- **Inception Score** – quantifies quality and diversity of generated images using a pre-trained CNN
  - Evaluates only the distribution of generated images
  - Apply pre-trained Inception V3 on generated images to get conditional class distribution $\boldsymbol{p(y|x)}$
    - Images that contain meaningful objects should have a conditional class distribution with low entropy – meaning the classification model confidently predicts a single class
  - GAN to generate diverse images, marginal distribution $\boldsymbol{p(y)}$ should have high entropy, average distribution over all generated images
  - Finally, combining these two requirements calculate KL-divergence between conditional class distribution and marginal class distribution
  - Uses a large number of images, 30k -50k

$$\mathrm{IS}(G) \approx \exp\left(\frac{1}{N}\sum_{i=1}^{N} D_{KL}(p(y|\mathbf{x}^{(i)}) \parallel \hat{p}(y))\right).$$

https://arxiv.org/pdf/1801.01973

Generative Adversarial Network

- **FID Score –** quantifies quality and diversity of generated images using feature embeddings of a pre-trained CNN
  - FID compares the distributions of generated images with the distribution of a set of real images
- Extract feature embeddings from intermediate layer of a pre-trained model for both real and fake images
  - Calculate mean and covariance of these features for each distribution
  - Use Fréchet distance (aka Wasserstein-2 distance) to measures difference between two distributions
- Lower FID score implies better quality



$$\text{FID} = \|\mu_R - \mu_G\|^2 + \text{Tr}\left(\Sigma_R + \Sigma_G - 2(\Sigma_R\Sigma_G)^{1/2}\right)$$

https://arxiv.org/pdf/2402.03654
https://arxiv.org/pdf/2402.03654

Challenges in Training GANs – Mode Collapse

- GANs aim to generate diverse, realistic synthetic images based on random inputs.
- The generator's goal is to produce images that fool the discriminator into classifying them as real.
- Once the generator succeeds, it may focus on producing similar images repeatedly.
- This repetitive process leads to ***mode collapse*** – where the generator creates a limited variety of images, trapping the discriminator.
- As a result, the generator fails to maintain diversity in its output, producing images in the same style.

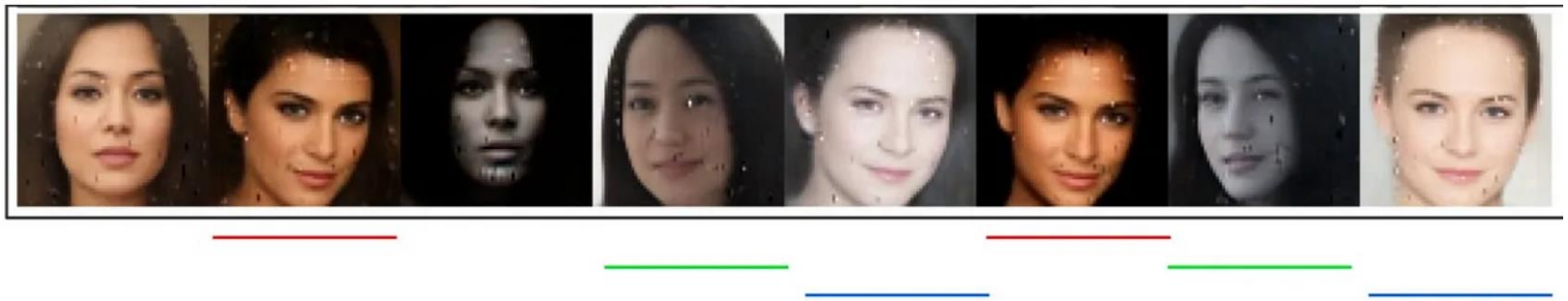


Image credit - https://jonathan-hui.medium.com/gan-why-it-is-so-hard-to-train-generative-advisory-networks-819a86b3750b

Challenges in Training GANs – Mode Collapse

- Mode collapse can be categorized into two types:
  - inter-class – affecting diversity across multiple classes
  - intra-class – impacting diversity within a single class
- Identification – observing **nature of generated images & generator's loss curves**

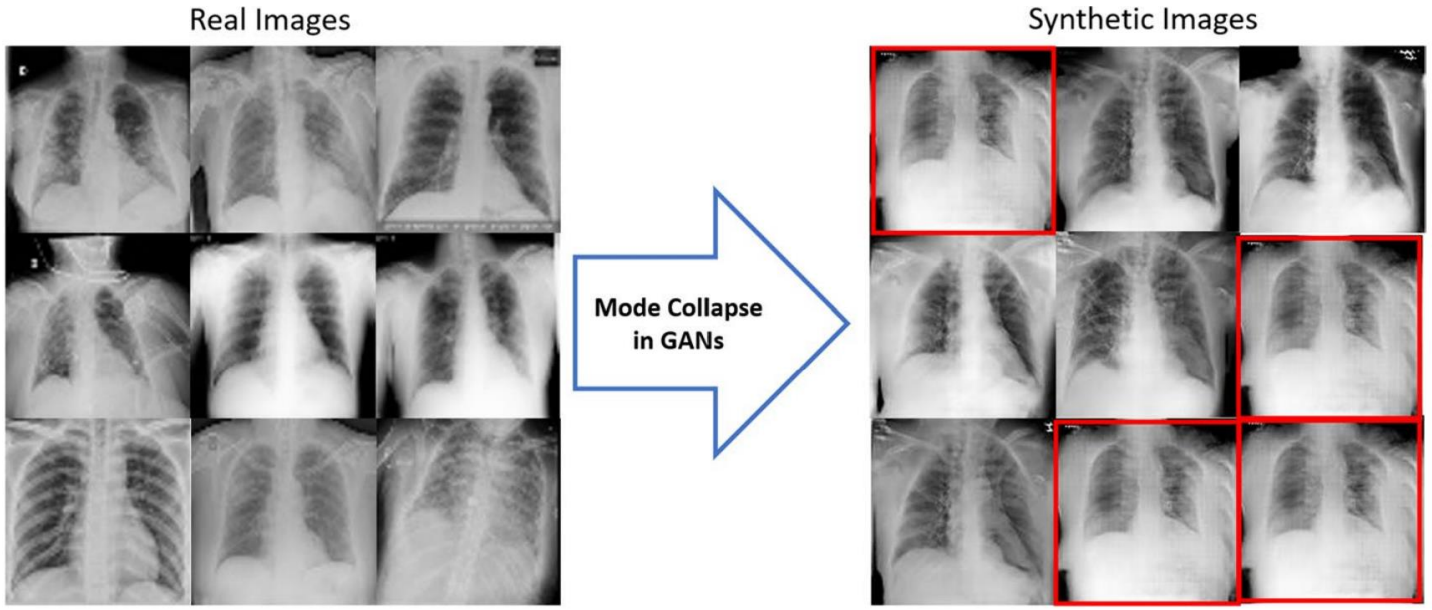


Fig. 9 Identification of mode collapse in GANs for X-ray image synthesis. The red areas highlighted illustrate the repetition of synthetic X-ray images with a similar distribution of features such as lungs. The chest bones are also suppressed indicating the occurrence of a mode collapse problem in GANs

Examples of Mode Collapse

Image source- Saad, M. M., O'Reilly, R., & Rehmani, M. H. (2024). A survey on training challenges in generative adversarial networks for biomedical image analysis. *Artificial Intelligence Review*, *57*(2), 19.
https://wandb.ai/authors/DCGAN-ndb-test/reports/Measuring-Mode-Collapse-in-GANs-Using-Weights-Biases--VmlldzoxNzg5MDk

Challenges in Training GANs – Mode Collapse

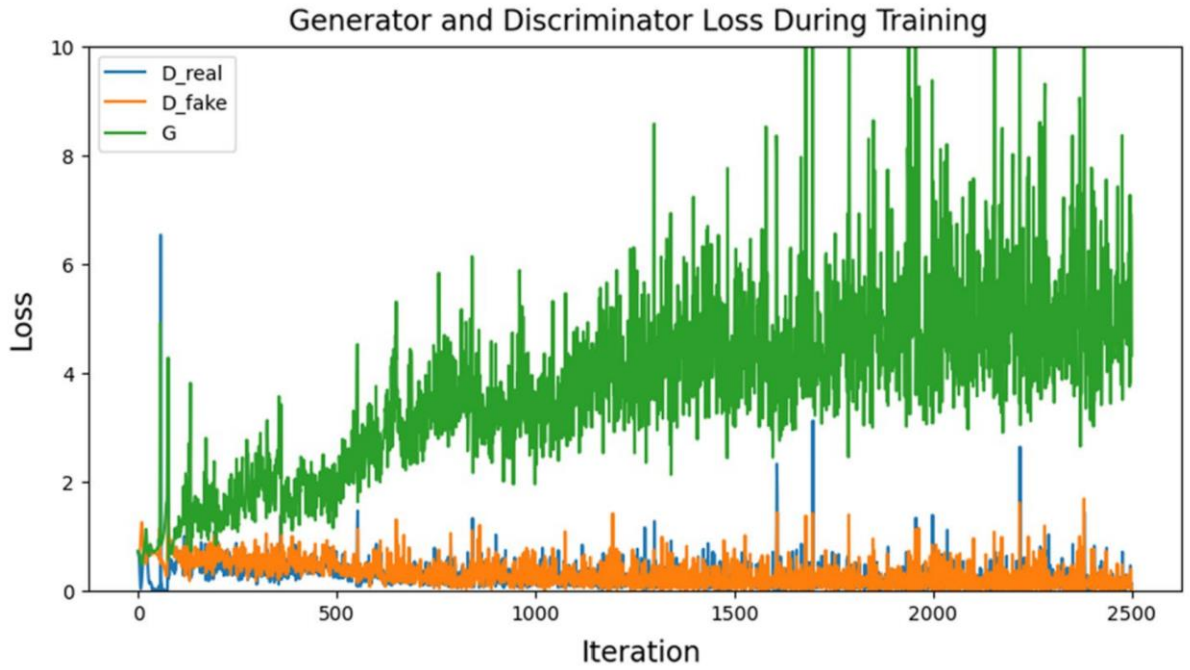


Fig. 10 Identification of the mode collapse problem using the non-converging generator loss of GANs for X-ray image synthesis. The generator loss depicted by label (*G*) illustrates the non-converging behavior as compared to the discriminator losses (*D_real* and *D_fake*)

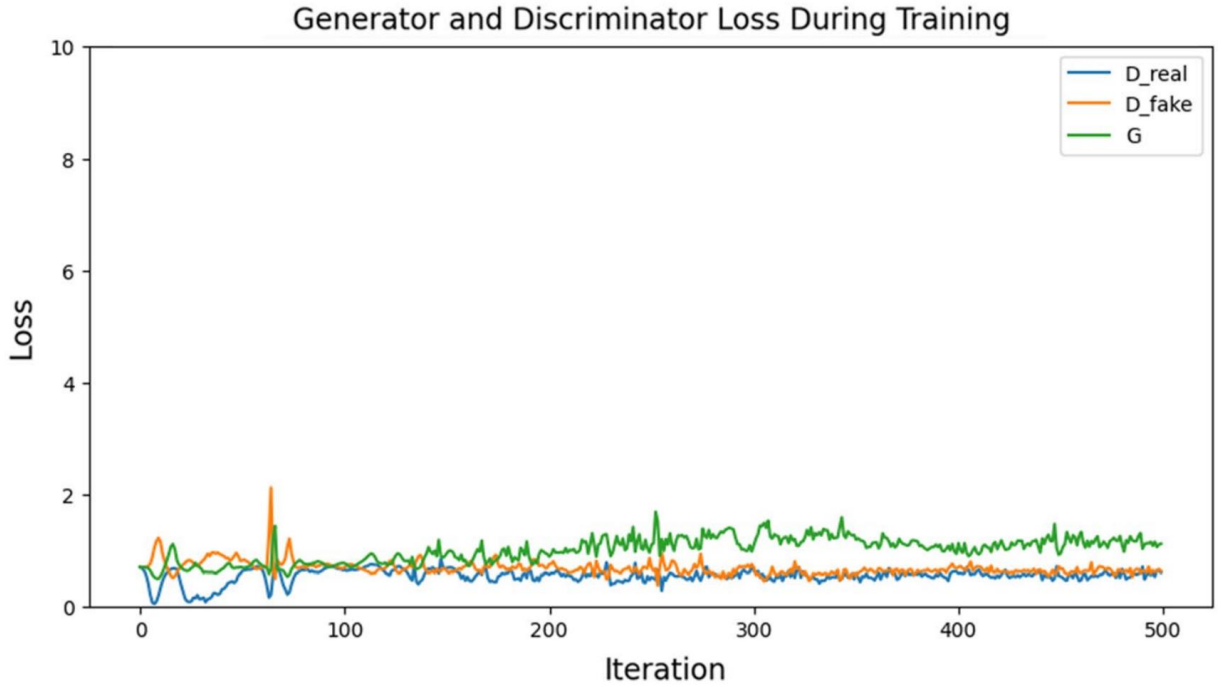


Fig. 11 Identification of the no-mode collapse in GANs for X-ray image synthesis. The generator loss depicted by label (*G*) illustrates the converging and balanced behavior as compared to the discriminator losses (*D_real* and *D_fake*)

Saad, M. M., O'Reilly, R., & Rehmani, M. H. (2024). A survey on training challenges in generative adversarial networks for biomedical image analysis. *Artificial Intelligence Review*, *57*(2), 19.

Challenges in Training GANs – Mode Collapse

- Solution to Mode collapse –
  - Regularization –
    - Weight Normalization, Spectral Normalization, and so on.
    - Input Normalization
      - DCGAN with Adaptive Input-Image Normalization (AIIN) –
        - Uses contrast based equalization to enhance X-ray image features
        - Highlight diverse features in X-ray images
        - Enhanced feature learning and improved, diverse image generation

Saad, M. M., O'Reilly, R., & Rehmani, M. H. (2024). A survey on training challenges in generative adversarial networks for biomedical image analysis. *Artificial Intelligence Review*, *57*(2), 19.

Challenges in Training GANs – Mode Collapse

- Solution to Mode collapse –
  - Modified architecture for generator – mode collapse can be reduced by using multiple generators rather than a single one, but this approach is computationally intensive.
  - For instance: Wu et al. (2018) proposed using multiple distributions with a Gaussian Mixture Model (GMM)-based generator to handle diverse data distributions in latent space, generating varied image samples without multiple generators.

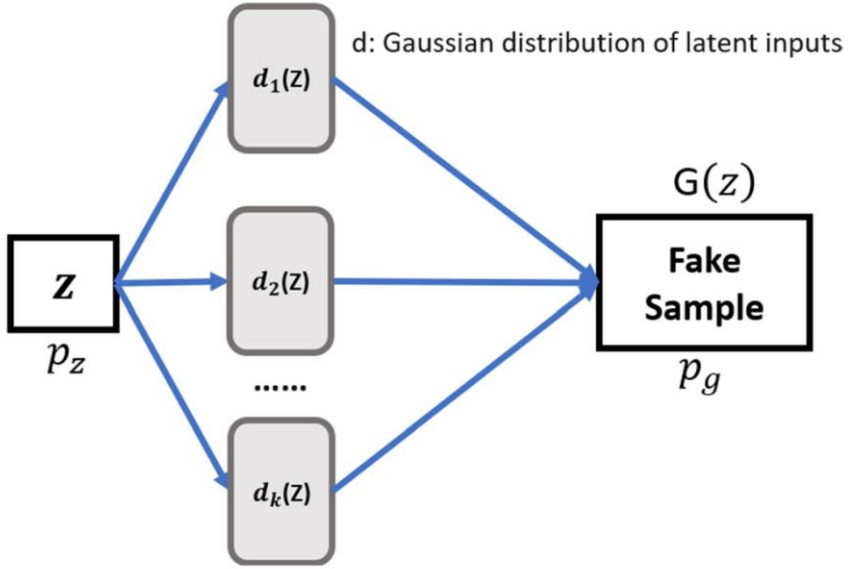


Fig. 14 The generator architecture of MDGAN. The figure is redesigned from Wu et al. (2018b)

Saad, M. M., O'Reilly, R., & Rehmani, M. H. (2024). A survey on training challenges in generative adversarial networks for biomedical image analysis. *Artificial Intelligence Review*, *57*(2), 19.

Challenges in Training GANs – Mode Collapse

- Some famous GANs to overcome Mode collapse –
  - **Wasserstein GAN (WGAN) –** uses Wasserstein distance as loss function instead of traditional cross entropy
  - Wasserstein distance, aka Earth Mover's Distance (EMD), quantifies two distributions (fake image vs. real image)
  - The loss is the difference between the expected value of the discriminator output for real images vs. expected value of the discriminator for fake images $\max_{\|f\|_L \leq 1} \mathbb{E}_{x \sim \mathbb{P}_r}[f(x)] - \mathbb{E}_{x \sim \mathbb{P}_\theta}[f(x)]$
  - Discriminator (aka ***critic)*** must satisfy Lipschitz constraints
    - Weight Clipping – Critic weights are clipped to a small range to satisfy the Lipschitz constraint, enforcing smoother and stable training
    - Gradient Penalty

WGAN Paper- https://arxiv.org/abs/1701.07875

Challenges in Training GANs – Mode Collapse

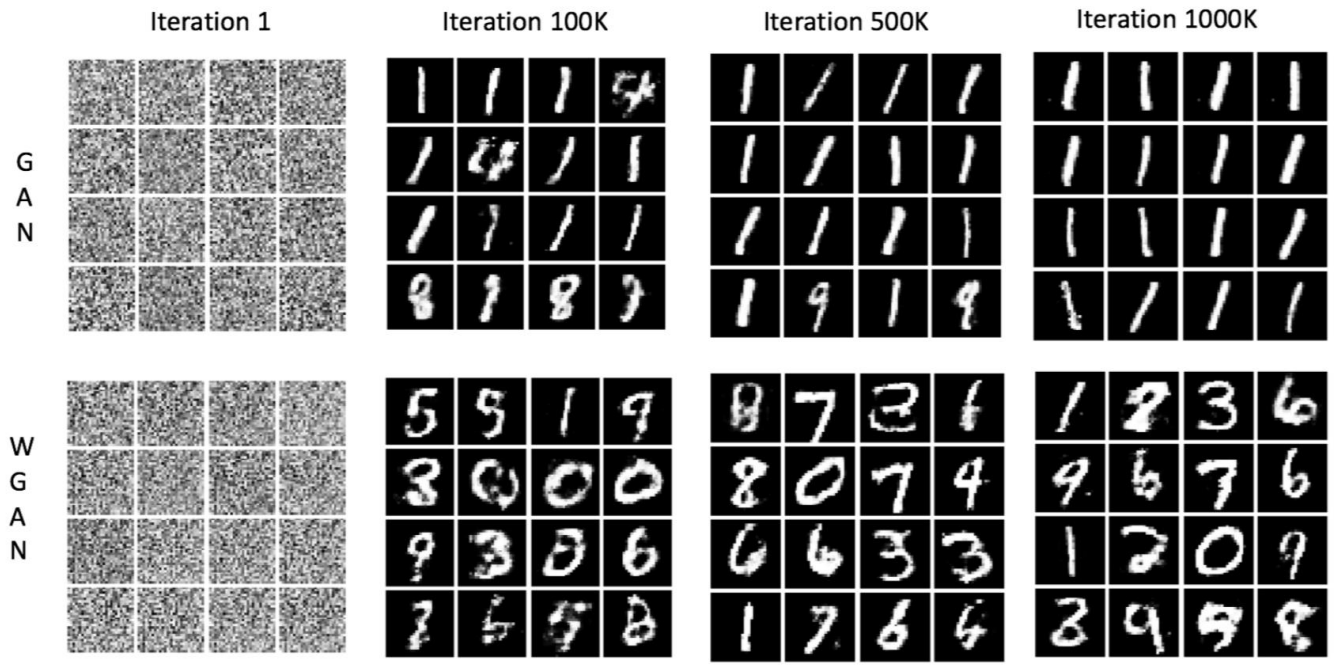
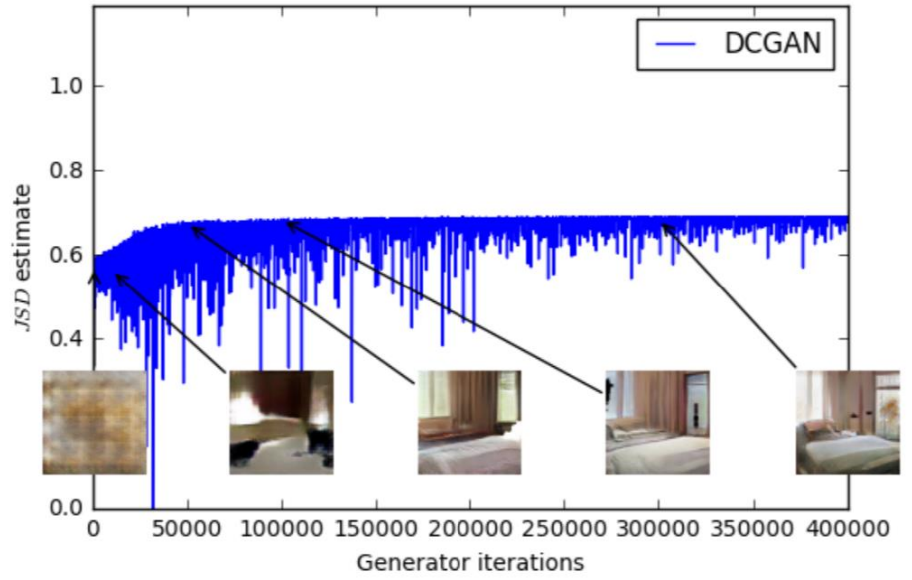
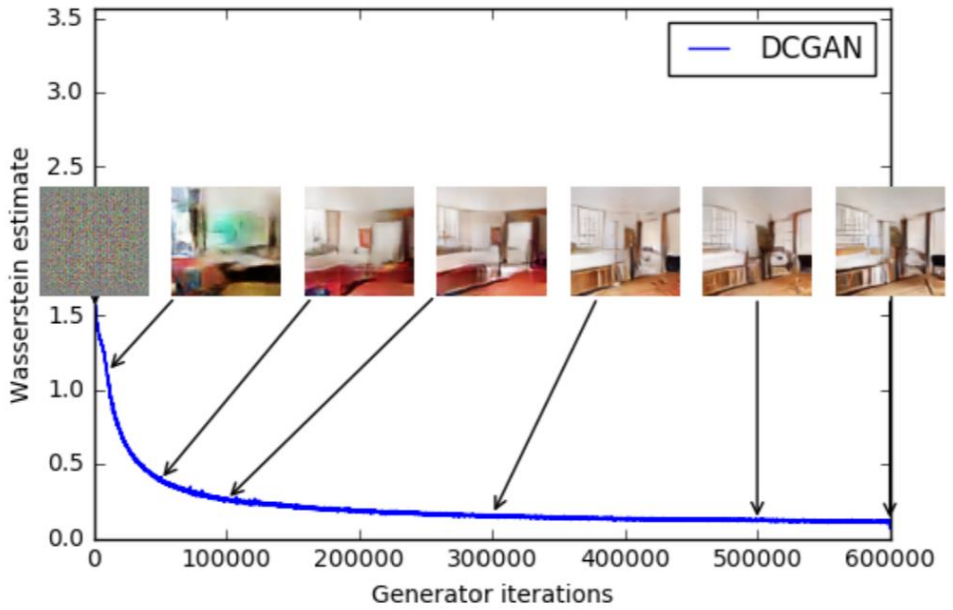


Figure 2. Generated images by GAN and WGAN models trained on MNIST after $1{,}100k$,$500k$,$1000k$ iterations.





WGAN Paper- https://arxiv.org/abs/1701.07875

Challenges in Training GANs – Mode Collapse

- **Unrolled GANs –** extended current loss function:
  - Instead of updating the generator right after one discriminator step, Unrolled GANs let the discriminator take several steps (or "unrolls") to better evaluate the generator's output.
  - It lowers the chance that the generator is overfitted for a specific discriminator

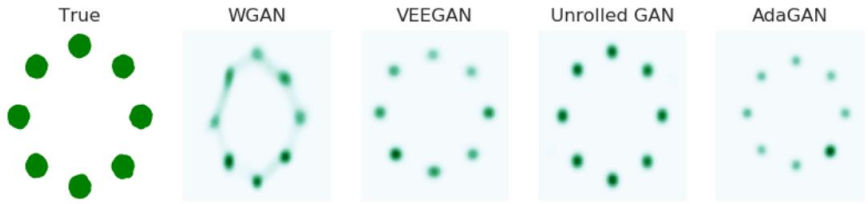


Fig. 1: True distribution and corresponding kernel density estimate of the learned distribution based on the generated samples of WGAN, VEEGAN, Unrolled GAN and AdaGAN from training on 2D ring dataset.

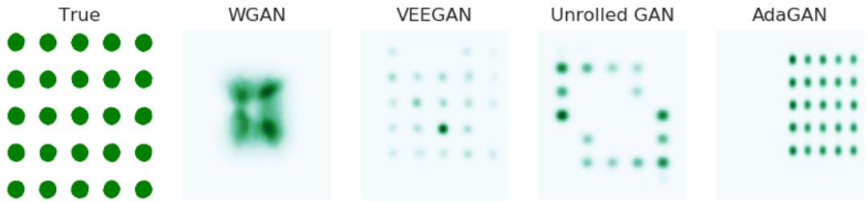


Fig. 2: True distribution and corresponding kernel density estimate of the learned distribution based on the generated samples of WGAN, VEEGAN, Unrolled GAN and AdaGAN from training on 2D grid dataset.

Image credit - A Probe Towards Understanding GAN and VAE Models

Challenges in Training GANs and Stabilization Techniques

- Non-Convergence – GAN fails to reach a stable equilibrium during training
  - The Generator and Discriminator enter a loop of endless competition without significant improvement
  - Simultaneous Update – updating G and D simultaneously without proper coordination can lead to instability
  - Saddle point optimization – Standard GD methods are not well suited for saddle points problems
  - Learning Rate Imbalance - Inappropriate learning rate for G and D can cause oscillations

https://jonathan-hui.medium.com/gan-why-it-is-so-hard-to-train-generative-advisory-networks-819a86b3750b