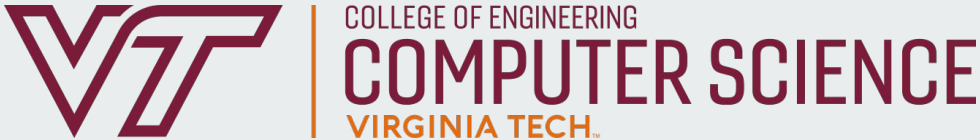
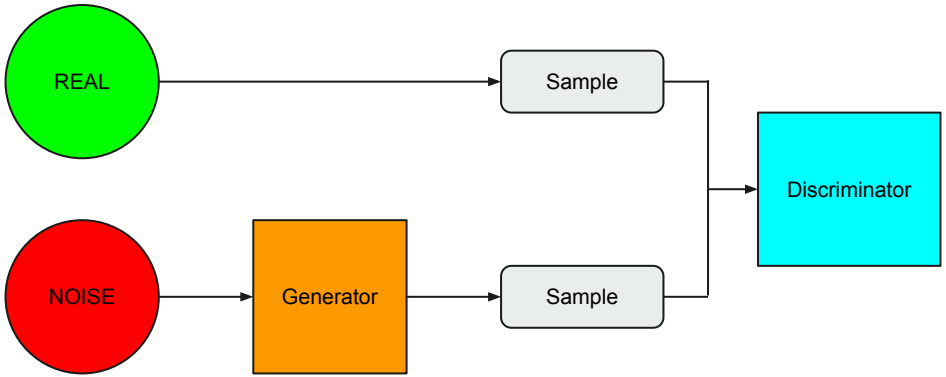
On the Evaluation of Generative Adversarial Networks By Discriminative Models

Amirsina Torfi, Mohammadreza Beyki, Edward A. Fox



Evaluation of GANs

A novel approach to evaluate performance of GANs



Contributions

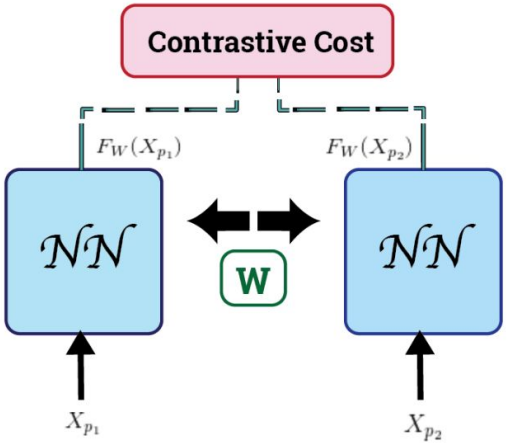
- Use Siamese Neural Networks (SNN) to evaluate GANs
- Proposed approach is **Domain Agnostic**
  - Electronic Healthcare Domain
  - Computer Vision Domain
- Comparable to FID Score and Inception Score
  - Aligned with statistical methods
  - Aligned with human evaluations

Characteristics of a good GAN

- Quality
  - Visual quality of the generated dataset
- Variety in the generated dataset
  - Diversity in generated samples
  - Covers all modes
  - Does not invent modes

Approach

- Leverage Siamese Architecture
- Train discriminator network on the real dataset
- Distance between same category of samples should be small
- Distance between different category of samples should be big



Training

- Definition
  - Genuine Pair -> both samples from the same category
  - Impostor Pair -> samples from different categories
- Contrastive Cost

$$D_W(X_{p_1}, X_{p_1}) = ||F_W(X_{p_1}) - F_W(X_{p_2})||_2.$$

- Goal is to minimize this cost

Training (Cont.)

$$\mathcal{L}_W(Y_i, (X_{p_1}, X_{p_2})_i) = Y * \mathcal{L}_{gen}(D_W(X_{p_1}, X_{p_2})_i) + (1 - Y) * \mathcal{L}_{imp}(D_W(X_{p_1}, X_{p_2})_i$$

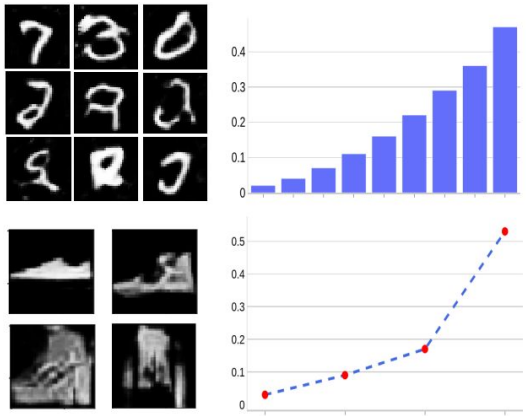
- Different losses for genuine and impostor pairs

$$\begin{cases} \mathcal{L}_{gen}(D_W) = \frac{1}{2}(D_W)^2 \\ \mathcal{L}_{imp}(D_W) = \frac{1}{2}(max\{0, M - D_W\})^2 \end{cases}$$

Siamese Distance Score

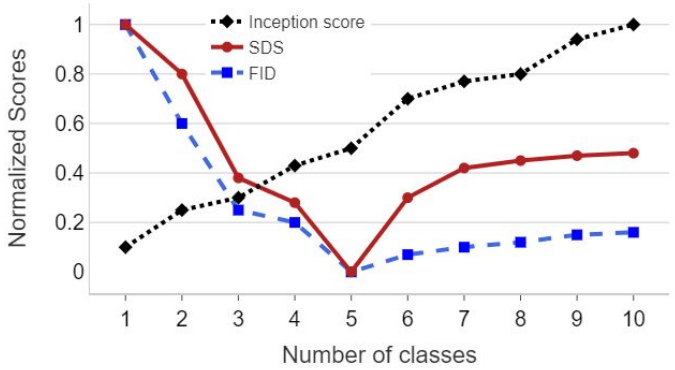
- Feed real and fake samples to the network and calculate their features
- Find all real samples that closely resemble fake samples
- Classify fake samples
- Calculate SDS Score for all samples

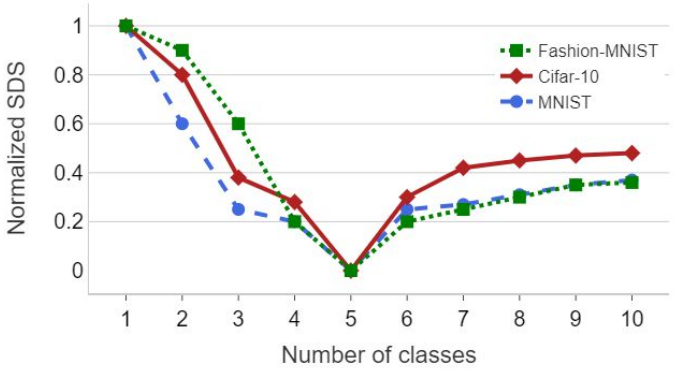
Comparison



Graphs represent images from left to right and then top to bottom

Comparison (Cont')





Summary

- We have proposed a novel approach to evaluate GANs using Siamese Neural Networks
- Our method is domain agnostic
- It can be used on other generative models as well as GANs
- It is sensitive to common GAN failures
- It is capable of capturing visual imperfections similar to human evaluation