

Face Generation with Conditional Generative Adversarial Networks

Xuwen Cao, Subramanya Rao Dulloor, Marcella Cindy Prasetio

Motivation

Objective: Generate face images conditioned on identity. Quantitatively measure the quality of generated samples.

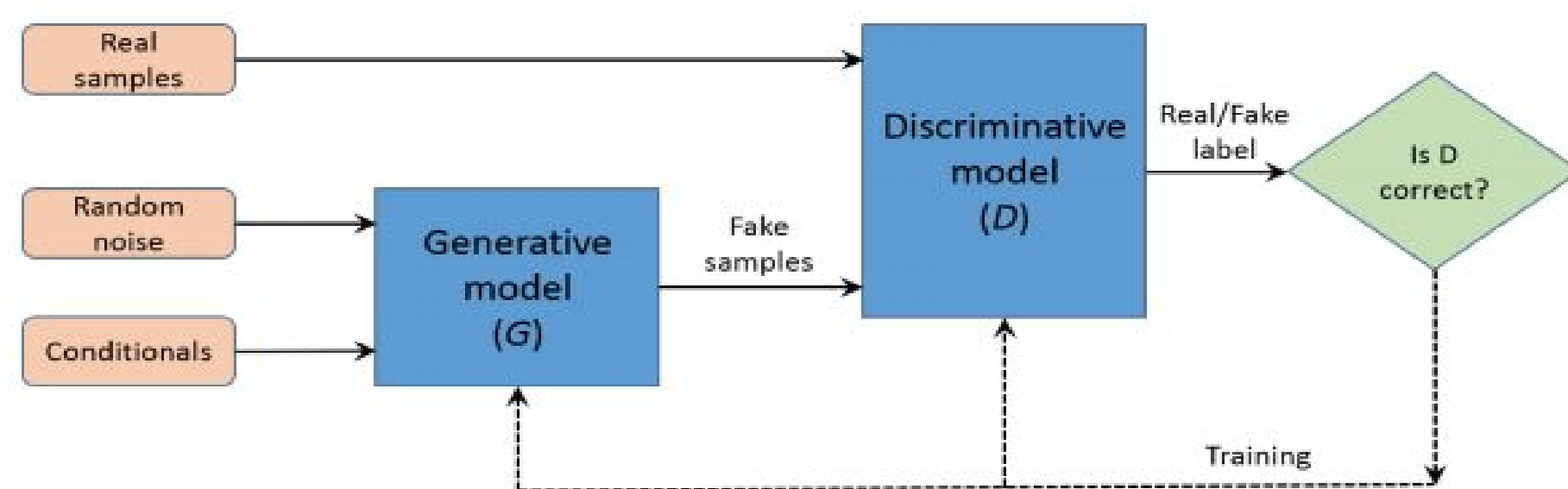
Use cases:

- Security - Strengthening face authentication systems
- Retail - Trialing style accessories under simulated conditions
- Entertainment - Social sharing of realistic face alterations

Approach:

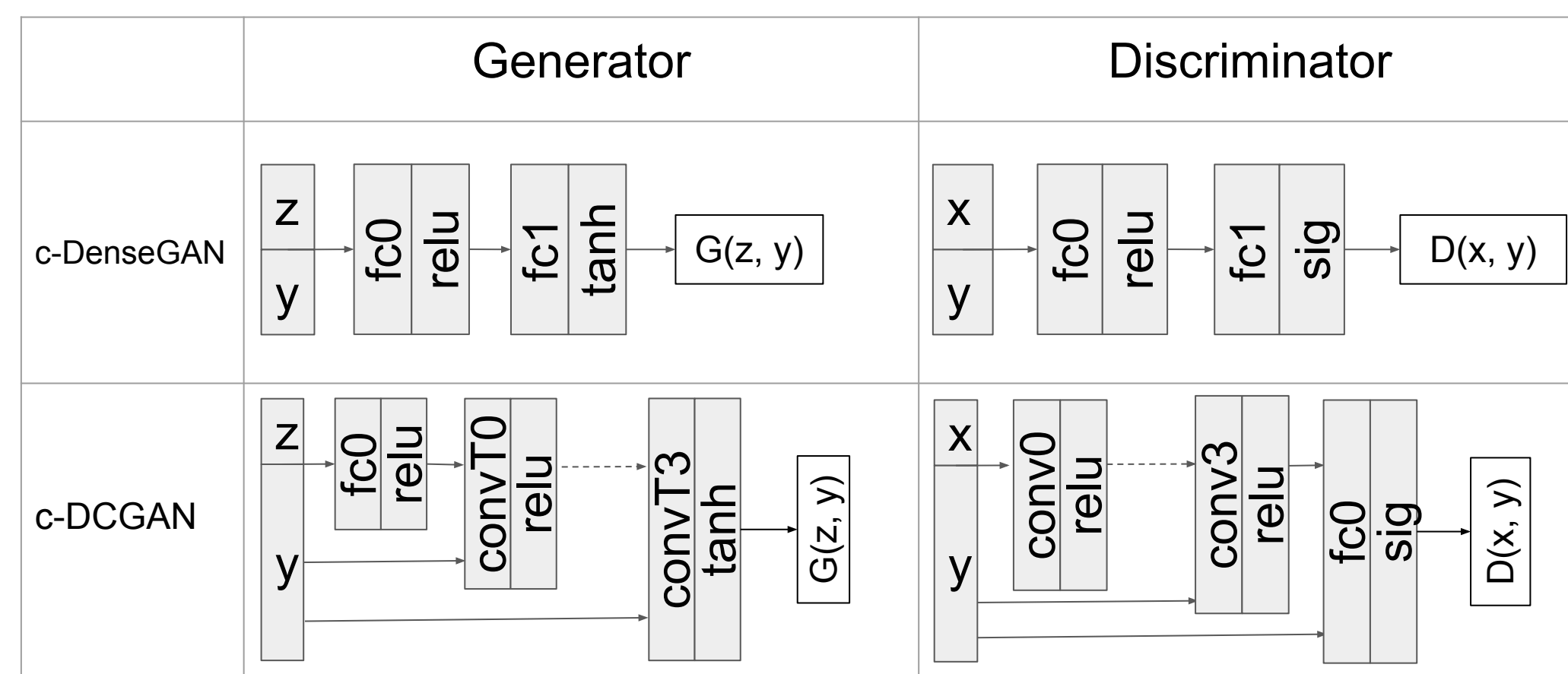
- Traditional: MLE on observed data (GMM, KDE). Not a good fit for complex data. Results in noisy output.
- Generative Adversarial Networks (GANs)

GAN Models



x: 160x160 input images; **y:** One-hot encoded conditionals; **z:** Noise
We experimented with variants of the following architectures.

Network architecture



Loss functions

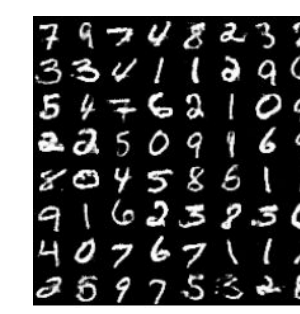
Log loss/CE loss
$$\min_G \max_D V(D, G) = E[\log(D(x))] + E[\log(1 - D(x))]$$

Wasserstein distance
$$\min_G \max_D V(D, G) = E[D(x)] + E[-D(G(z))]$$

Data and Evaluation

MNIST

♦ 60K images of handwritten digits. Good for basic evaluation.
Evaluation: Generated samples are evaluated using a deep MNIST classifier with 99.2% accuracy.



MNIST dataset

CelebA

♦ 10K identities. 200K images. 40 binary attributes per image.
♦ Noisy identity labels. Only ~80 images per identity.
♦ Not a good fit for face generation conditioned on identity.



Custom dataset collected from online sources

Custom

♦ Face images collected using Google's image search.
♦ 10 identities. 3K images. Identity/gender labels per image.
♦ Images post-processed via cropping and alignment.
Evaluation: Generated faces are evaluated using MT-CNN face detector and using an identity classifier trained on Facenet embeddings extracted from the same dataset.

Results

MNIST

Evaluation accuracy:

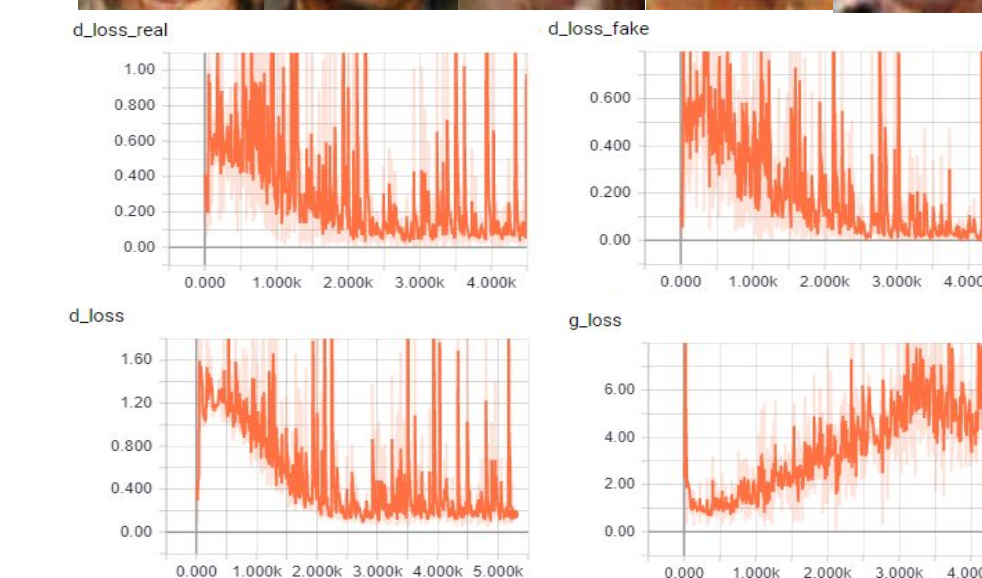
Noisy	88.28%	99.22%	98.44%	97.66%
GMM	c-DenseGAN#1	c-DenseGAN#2	c-DCGAN#2 CE loss	c-DCGAN#2 Wasserstein loss

Custom celebrity face dataset



c-DCGAN#3

ID	Acc.	ID	Acc.
Trump	42.2%	Messi	18.5%
Leonardo	39.4%	Will Smith	14.8%
Jennifer Aniston	34.4%	Obama	14.8%
Brad Pitt	22.7%	Hillary	13.5%
Zhang Ziyi	20.8%	Beyonce	7.9%



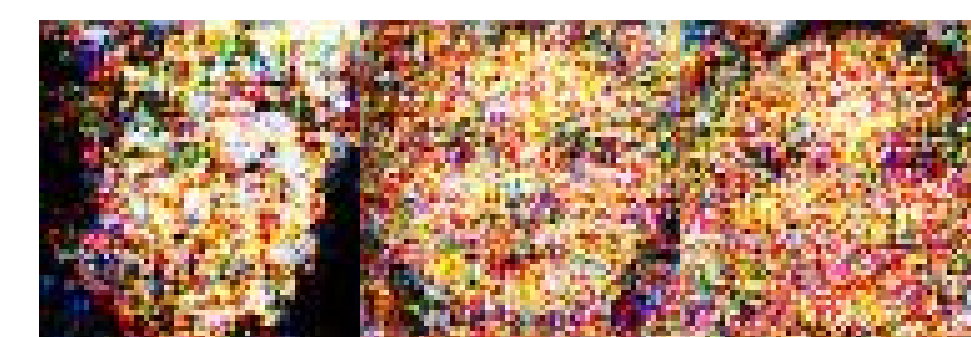
Error Analysis

Data processing related errors



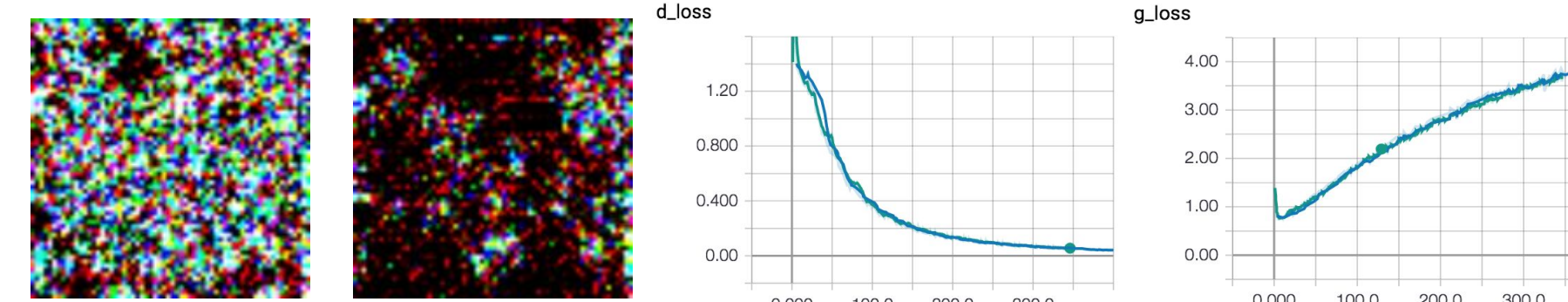
Noisy data input (unaligned)

Loss related errors

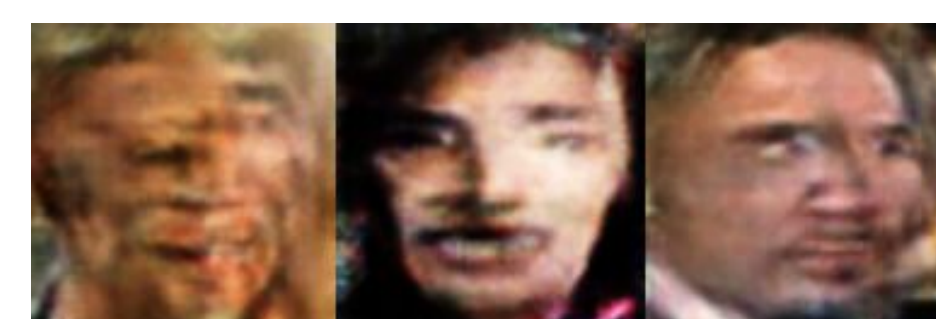


Wasserstein loss

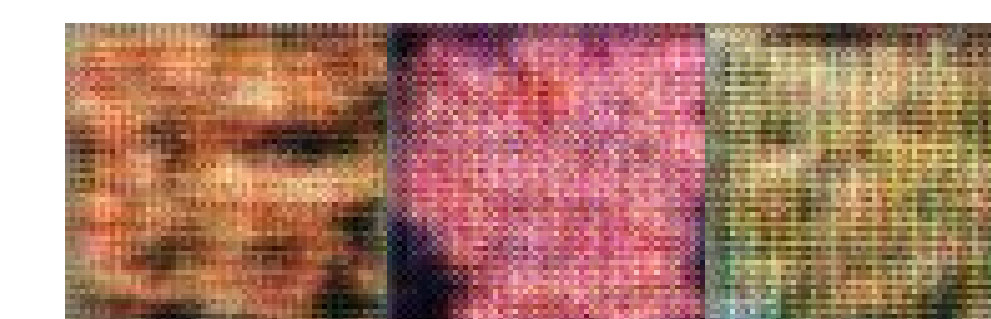
Network architecture related errors



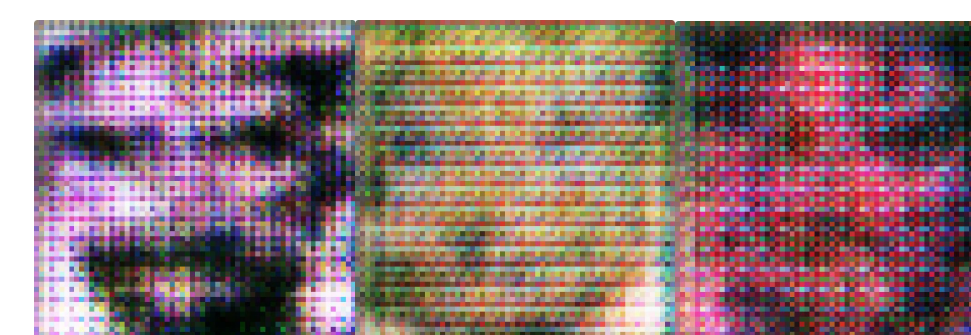
Generator activation function (sigmoid instead of tanh)



An extra discriminator stacked on top of D(x, y)

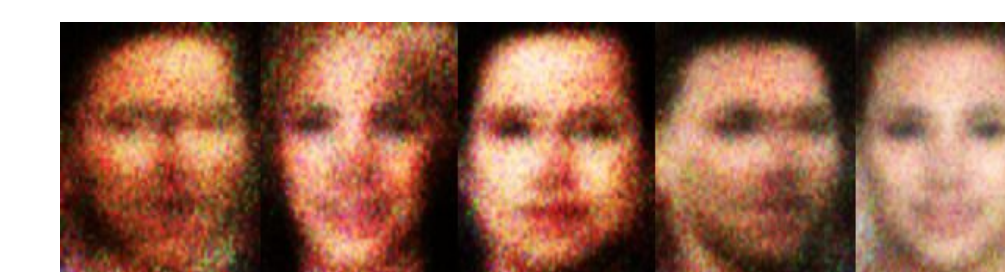


Leaky RELU in generator's hidden layers

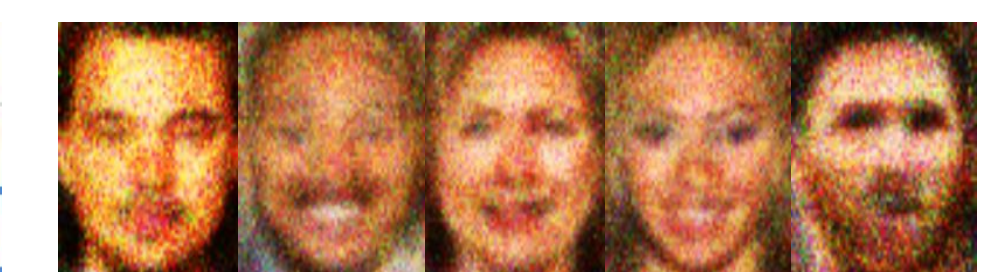


Mixing images and conditionals directly in every layer

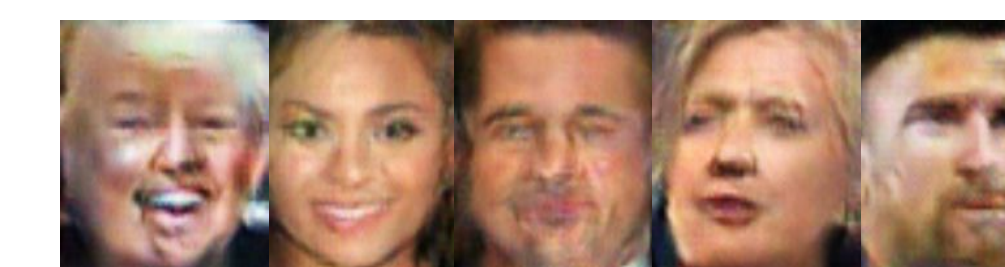
Evaluation	c-DenseGAN#1	c-DenseGAN#2	c-DCGAN#1	c-DCGAN#2	c-DCGAN#3
Successful face detection	89.5%	96.1%	99.5%	98.4%	99.1%
ID match on full dataset	10.1%	15.7%	5.6%	6.0%	11.4%
ID match on successful detection	10.5%	16.1%	5.7%	6.0%	11.5%



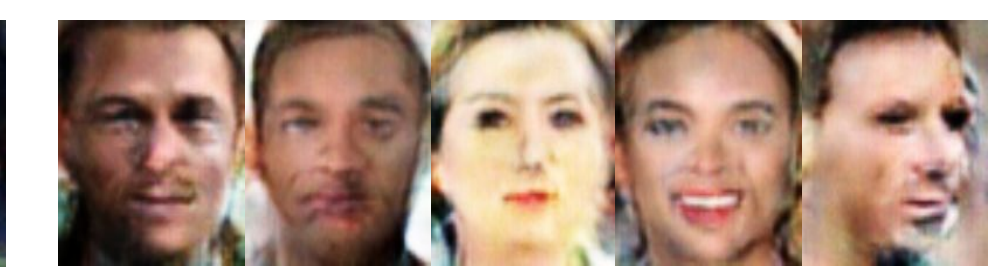
c-DenseGAN#1



c-DenseGAN#2



c-DCGAN#1



c-DCGAN#2