Generative Adversarial Networks

Rafael Lima Anh Tong

UNIST

June 20, 2017

Overview

Adversarial Training

- Current Definition: "Training a model in a worst-case scenario, with inputs chosen by an adversary"
- Example: An agent playing against a copy of itself in a board game

Generative Adversarial Networks

- Both players are Neural Networks
- Goal: Learn to generate data resembling the training set
- worst-case input for one network is produced by another network

Generative Modeling

Density Estimation

Given a lot of examples, find a probability distribution that describes these examples

Sample Generation

Learn a function or program which can generate more samples from that underlying distribution

Adversarial Nets Framework

Two different agents playing a game against each other:

Generative Network (G)

Tries to generate data

Discriminative Network (D)

Examines data and try to say if it is real or fake

Adversarial Nets Framework

$\rightarrow$ The goal of the generator is to fool the discriminator
$\rightarrow$ As both players get better and better over time, the G is forced to create data that is as realistic as possible (same distribution as the training data)

Adversarial Nets Framework

Example: Police and Money Counterfeiter

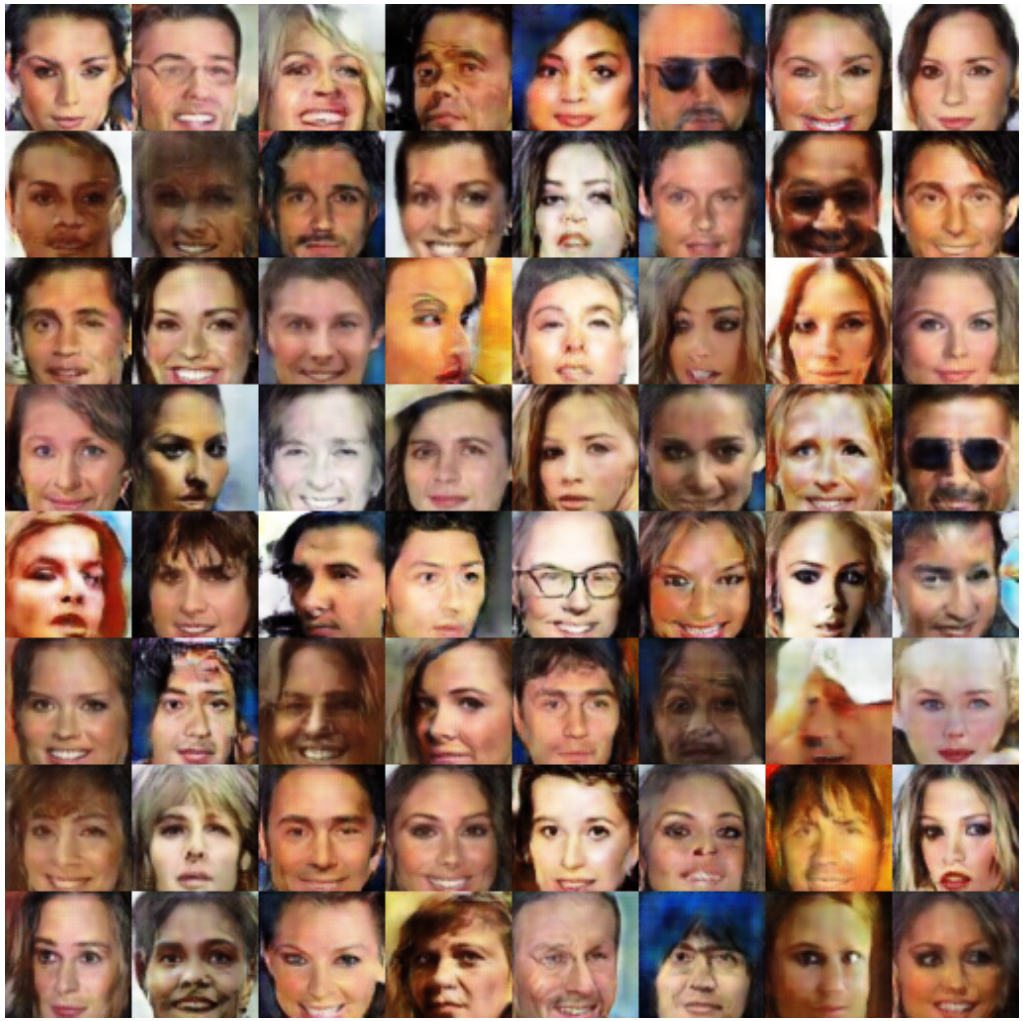
Nash equilibrium is perfectly fake money: 50% fake and 50% true choice for police

CelebA Dataset

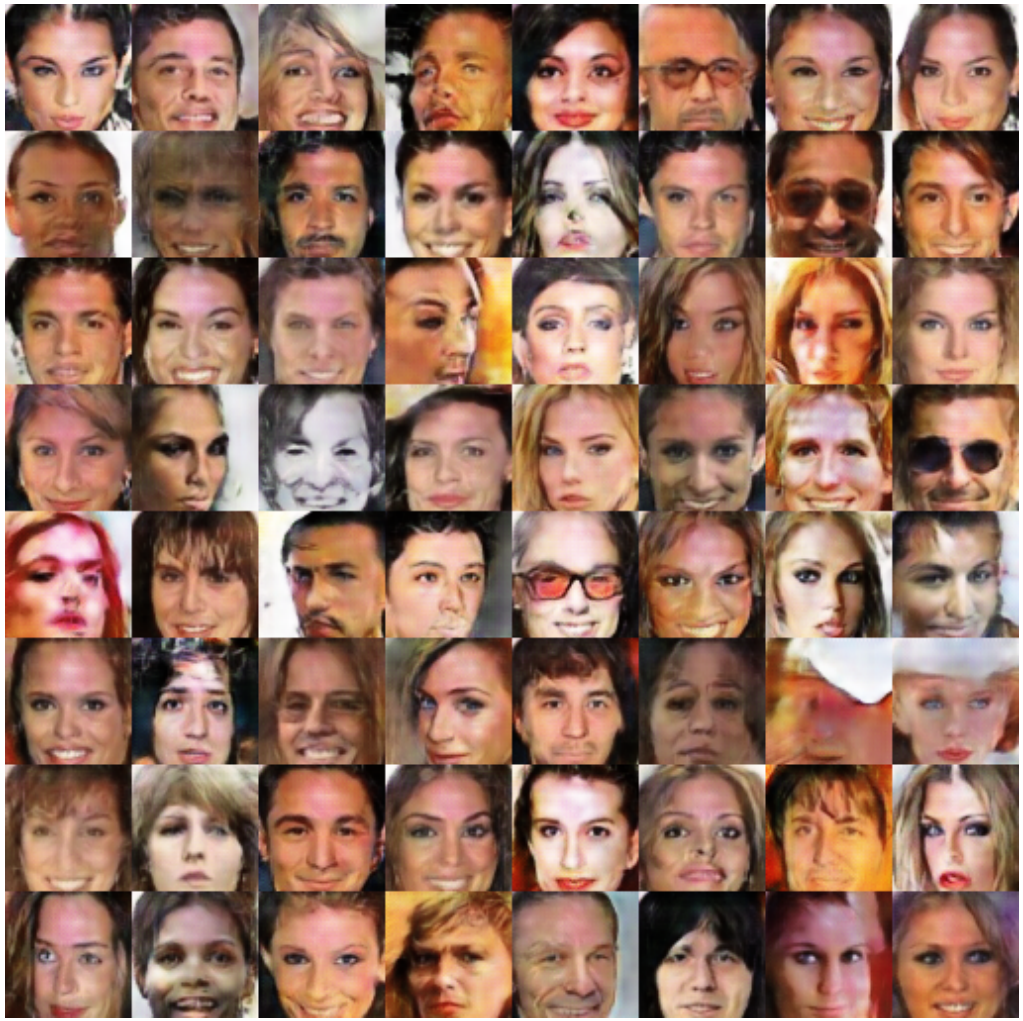
- Large-scale face attributes dataset with more than 200K celebrity images, each with 40 attribute annotations.
- Can be employed as the training and test sets for computer vision tasks:
  - Face attribute recognition
  - Face detection
  - Landmark (or facial part) localization

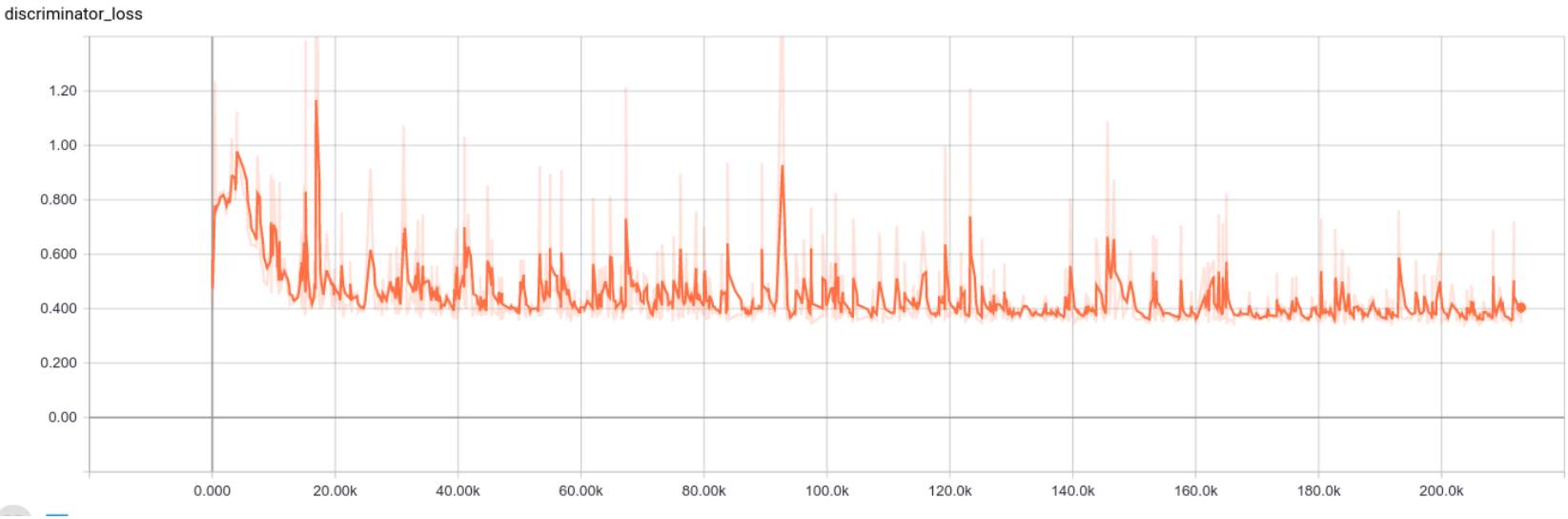
Samples: 10 Epochs

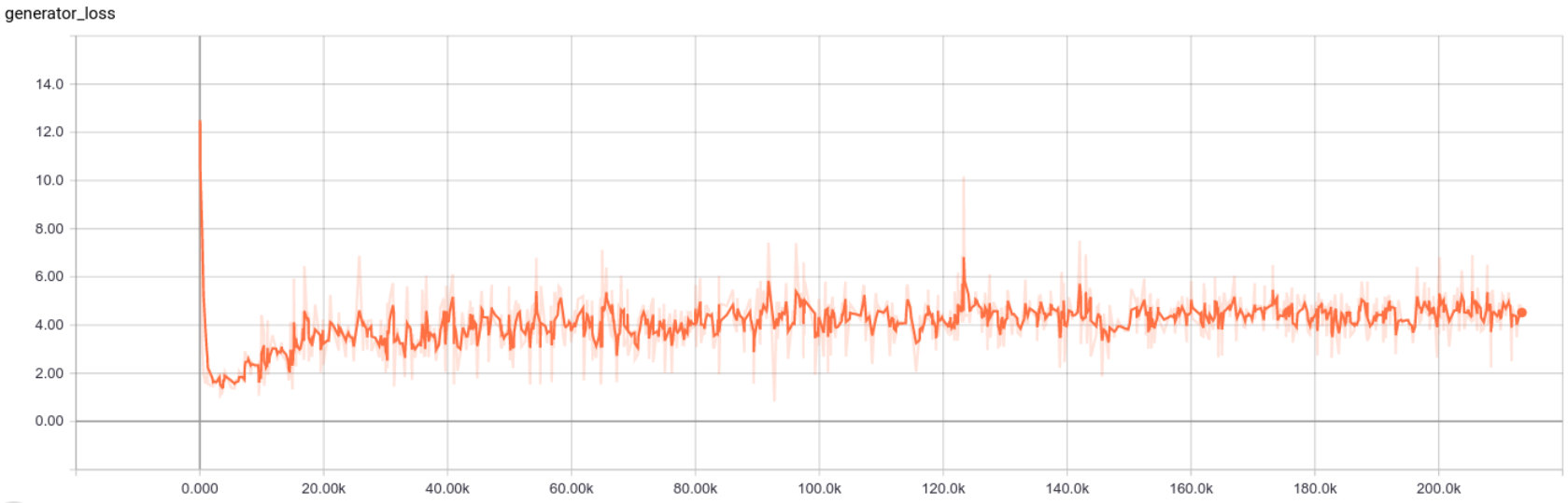
Samples: 15 Epochs



Samples: 20 Epochs




discriminator_loss
1.20
1.00
0.800
0.600
0.400
0.200
0.00
0.000
20.00k
40.00k
60.00k
80.00k
100.0k
120.0k
140.0k
160.0k
180.0k
200.0k


generator_loss
14.0
12.0
10.0
8.00
6.00
4.00
2.00
0.00
0.000
20.00k
40.00k
60.00k
80.00k
100.0k
120.0k
140.0k
160.0k
180.0k
200.0k

References

Ian Goodfellow (2016)

NIPS 2016 Workshop on Adversarial Training - Introduction to GANs

Neural Information Processing Systems (2016)

Soumith Chintala (2016)

NIPS 2016 Workshop on Adversarial Training - How to train a GAN

Neural Information Processing Systems (2016)

Thank You