Disclaimer

If a student turns on their microphone or camera or uses the public chat feature, this constitutes consent for the student's video image or sound audio to be uploaded with the office hour or tutorial on university approved platforms such as D2L. If the student wishes to ensure that their questions/faces/voices are not recorded in the video, they should instead use private chat to ask questions.

Generative Adversarial Networks

A gentle introduction to GANs and other uses of adversarial components

Roberto Souza
Assistant Professor
Electrical and Computer Engineering
Schulich School of Engineering

March 2022

UNIVERSITY OF CALGARY

Keep Up with the Presentation!!!

- We are going to have questions along the slides :O
  - Please answer them on the chat or unmute: let's discuss
- Please count how many times this pic appears on the slides

- At the end of the presentation, please share:
  - Something you learn
  - Share your knowledge (teach us something)

Outline

- Generative Adversarial Networks (GANs)
  - What are GANs?
  - How GANs work?
  - Types of GANs and applications
- Adversarial Attacks
- Summary

Learning Goals

- Introduce generative adversarial networks
  - What they do
  - How they work

- Illustrate how adversarial attacks work

Introduction

- Generative adversarial networks (GANs) are unsupervised deep learning methods

- There are many types of GANs
  - Wasserstein GAN
  - Cycle-GAN
  - Conditional GAN
  - Info GAN

- They all operate under the same principle of having modules with competing objectives

GANs is the most interesting idea in the last ten years in machine learning.

Yan LeCun, Chief AI Scientist at Meta

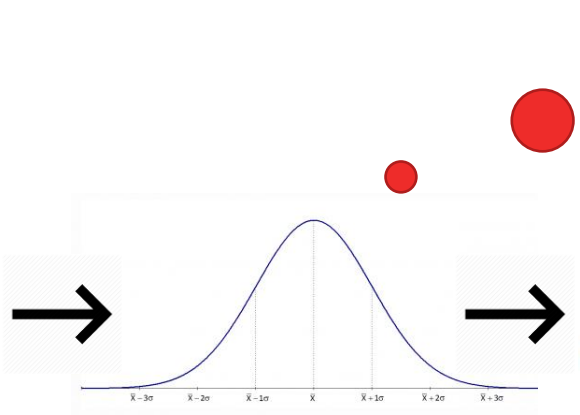
Generative vs Discriminate Models

1) Discriminative models

GANs are generative models where the data distribution is learned implicitly

2) Generative models

- Dataset

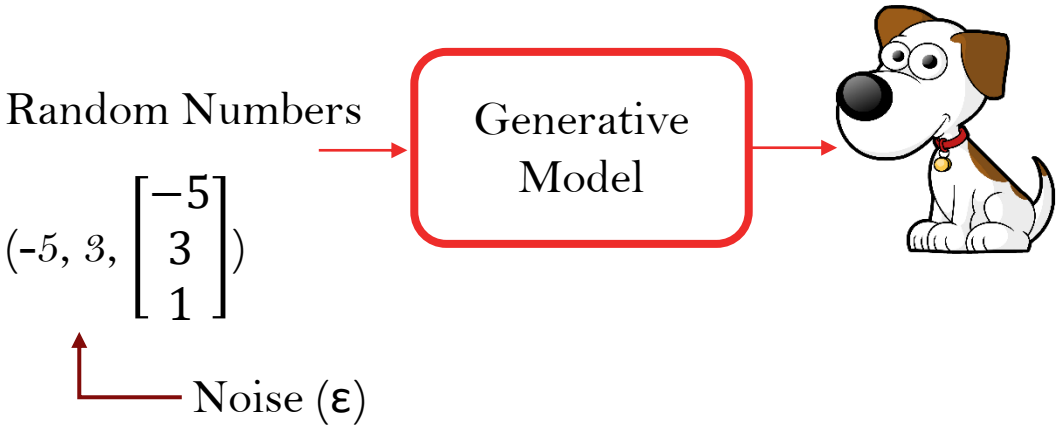


- Learning distribution

- New data

Machine Learning Models

- Generative Models
  - Generate realistic representation for each class.



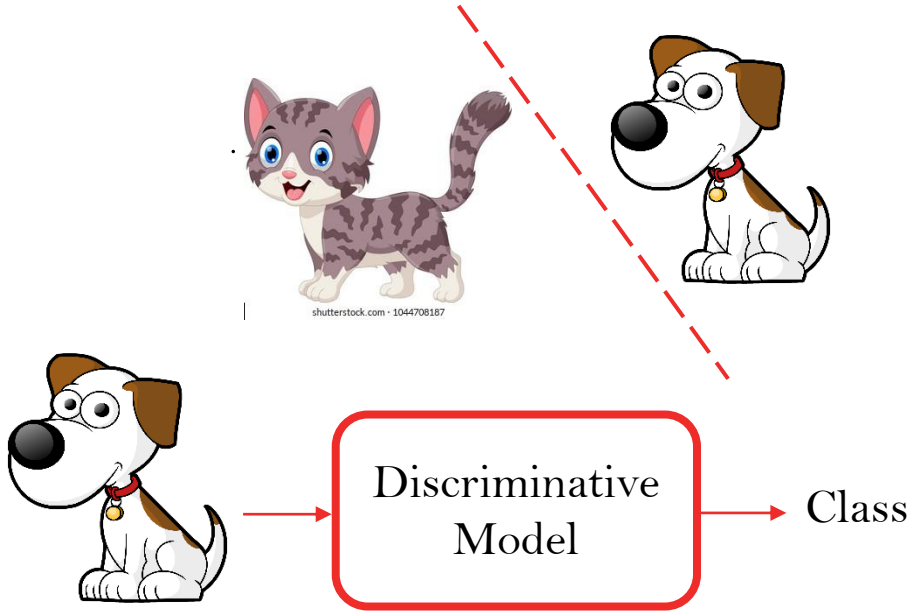
Y, ε → X (features)

$$P(X|Y)$$

or

$$P(X)$$

- Discriminative Models
  - Used for classification problem



X (features) → Y

$$P(Y|X)$$

9

Generative Adversarial Networks

To produce Realistic Presentation of different classes

To distinguish real images from fake ones (produced by generator)

$$\begin{bmatrix} -5 \\ 3 \\ 1 \end{bmatrix}$$ → Generative Model →  → Discriminative Model → Fake/ Real

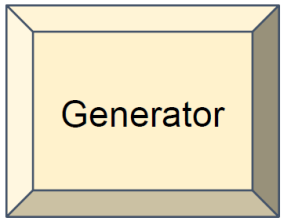
There is a competition here !

- Generator tries to make fakes that look real and fool the discriminator
- Discriminator learns how to distinguish reals from fakes

What are GANs?

- Two separate networks compete against each other: **generator** and **discriminator**

- This can be thought of as a game between a **counterfeiter** and an **art curator**

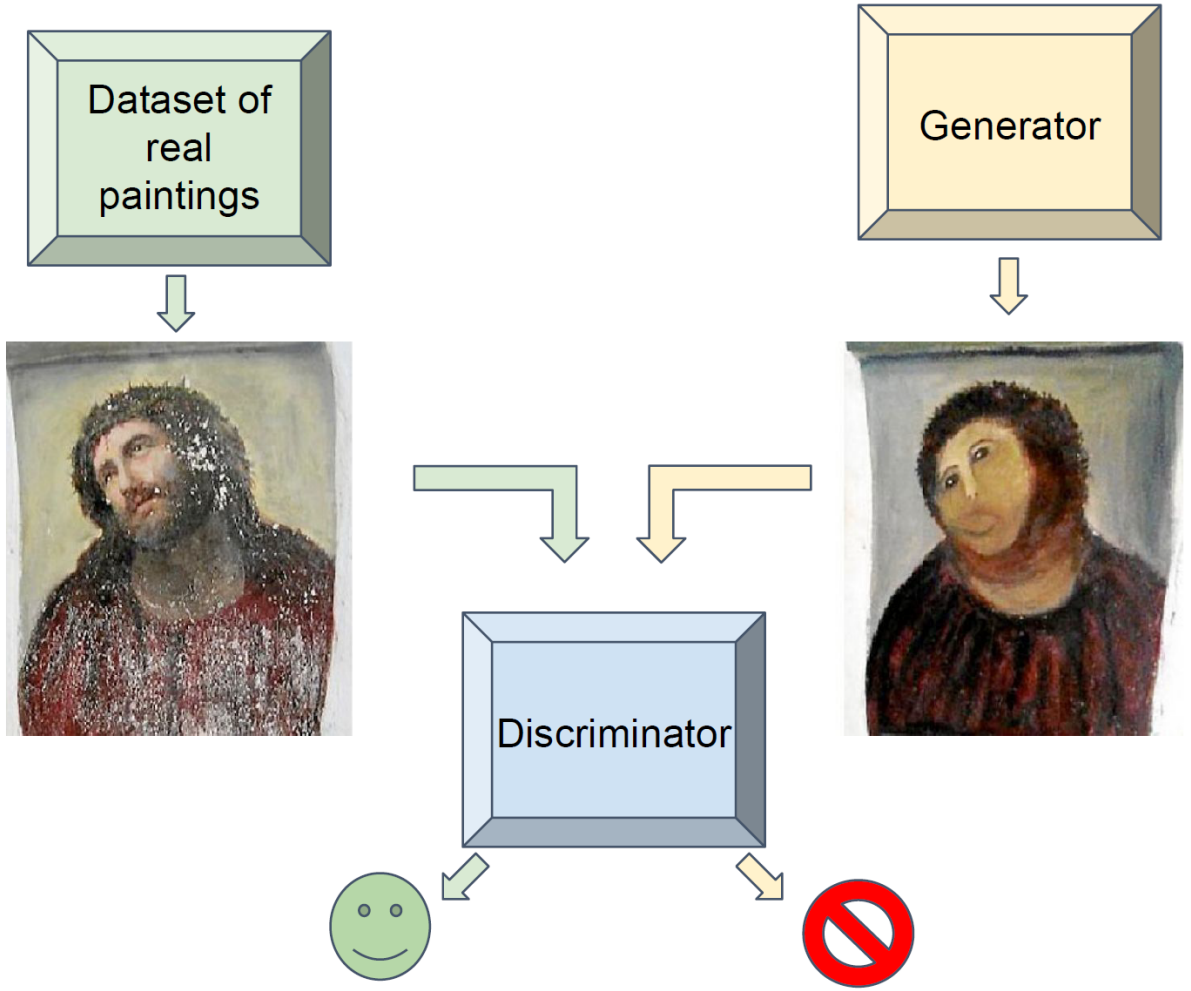


Generates samples that look convincingly real

Determines whether a sample is real or generated

What are GANs?

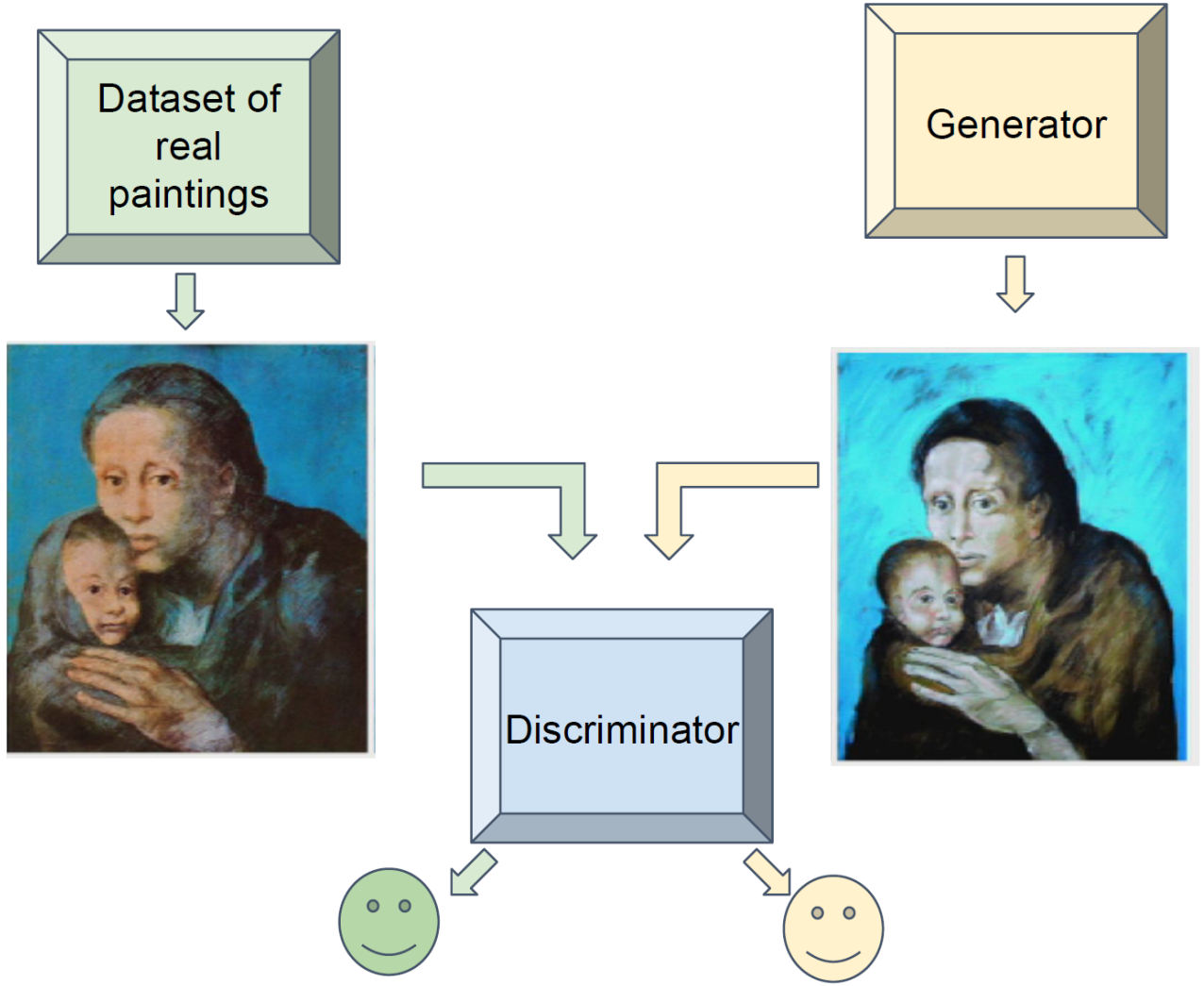
- The counterfeiter (generator) attempts to generate real looking images
  - At first the generator is pretty bad



https://www.pri.org/stories/2012-08-25/amateur-restoration-botches-jesus-painting-spain

What are GANs?

- Eventually, the generator learns to fool the discriminator and can generate convincing images



https://www.channel4.com/news/art-forgery-beltracchi-wolfgang-ernst-picasso-paraic-obrien

Question #1

- You have IQ scores for 1000 people. You model the distribution of IQ scores with the following procedure:
  - Roll three six-sided dices
  - Multiply the roll by a constant w
  - Repeat 100 times and take the average of all the results
  - You try different values for w until the result of your procedure equals the average of the real IQ scores. Is your model a generative model or a discriminative model?

- (a) Generative model
- (b) Discriminative model
- (c) Not enough information to tell

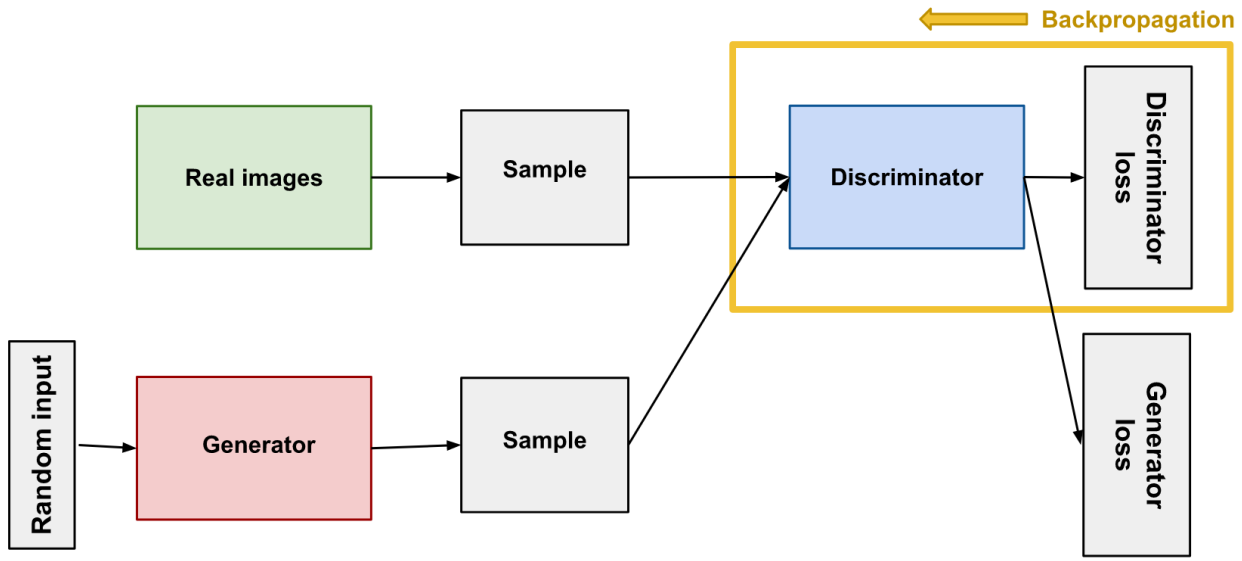
Question #1

- You have IQ scores for 1000 people. You model the distribution of IQ scores with the following procedure

- **(a) Generative model**: with every roll you are effectively generating the IQ of an imaginary person.

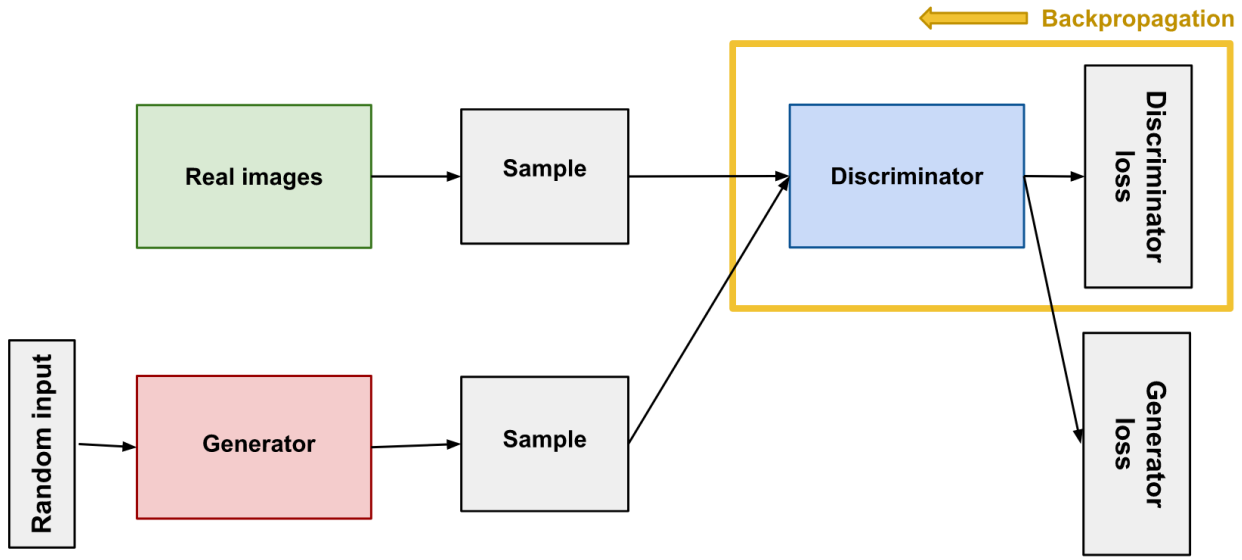
Discriminative Models

- The discriminator in a GAN is simply a classifier
- During discriminator training:
  - Discriminator classifies real data and fake data from the generator
  - The discriminator loss penalizes it for misclassification
  - Updates its weights through backpropagation



Discriminative Models

- The discriminator connects to two loss functions:
  - During discriminator training, the discriminator ignores the generator loss and just uses the discriminator loss
  - The generator loss is used during generator training



Generative Models

- Learns to create fake data by incorporating feedback from the discriminator
  - Learns to make the discriminator classify its output as real

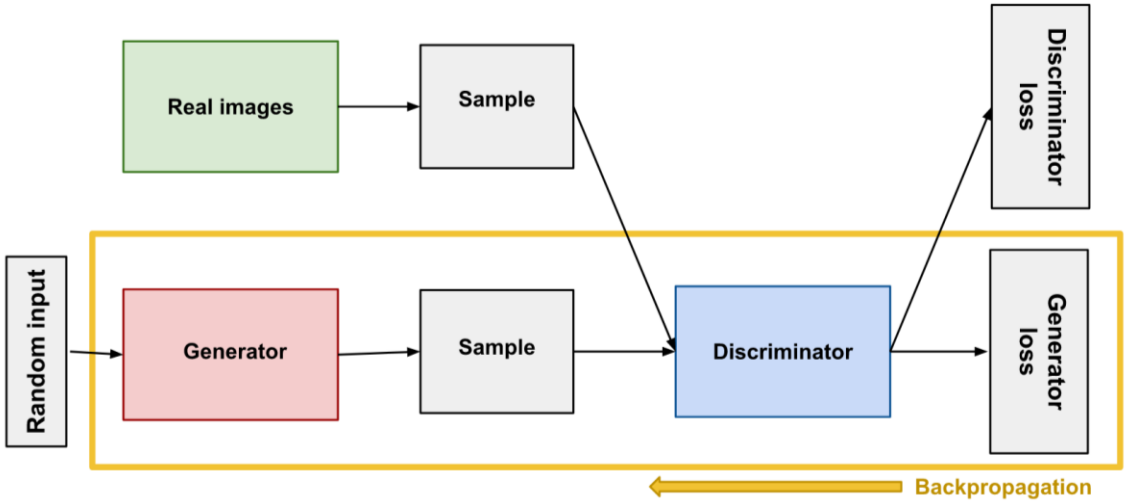
- Generator training:
  - random input
  - generator network, which transforms the random input into a data instance
  - discriminator network, which classifies the generated data
  - discriminator output
  - generator loss, which penalizes the generator for failing to fool the discriminator

Generative Models – Random Input

- GANs take random noise as its input (not always)
- The generator transforms this noise into a meaningful output
- By introducing noise, the GANs can produce a wide variety of data, sampling from different places in the target distribution
- Experiments suggest that the distribution of the noise doesn't matter much, so we can choose something that's easy to sample from, like a uniform distribution

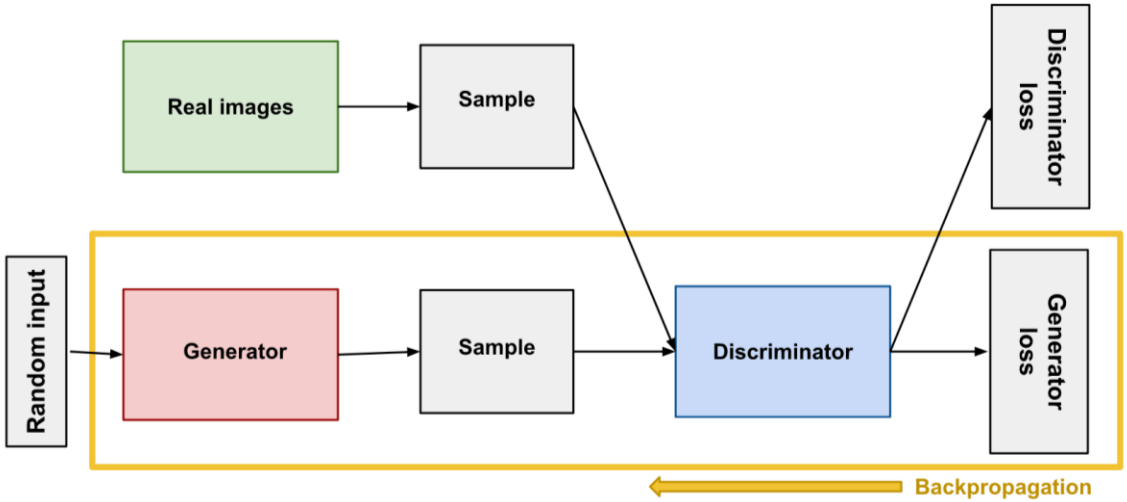
Generative Models – Discriminator to train the generator

- The generator feeds into the discriminator
- D*iscriminator* produces the output
- The generator loss penalizes the generator for producing a sample that the discriminator network classifies as fake
- **The discriminator does not change during generator training**
  - Trying to hit a moving target would be a harder problem for the generator

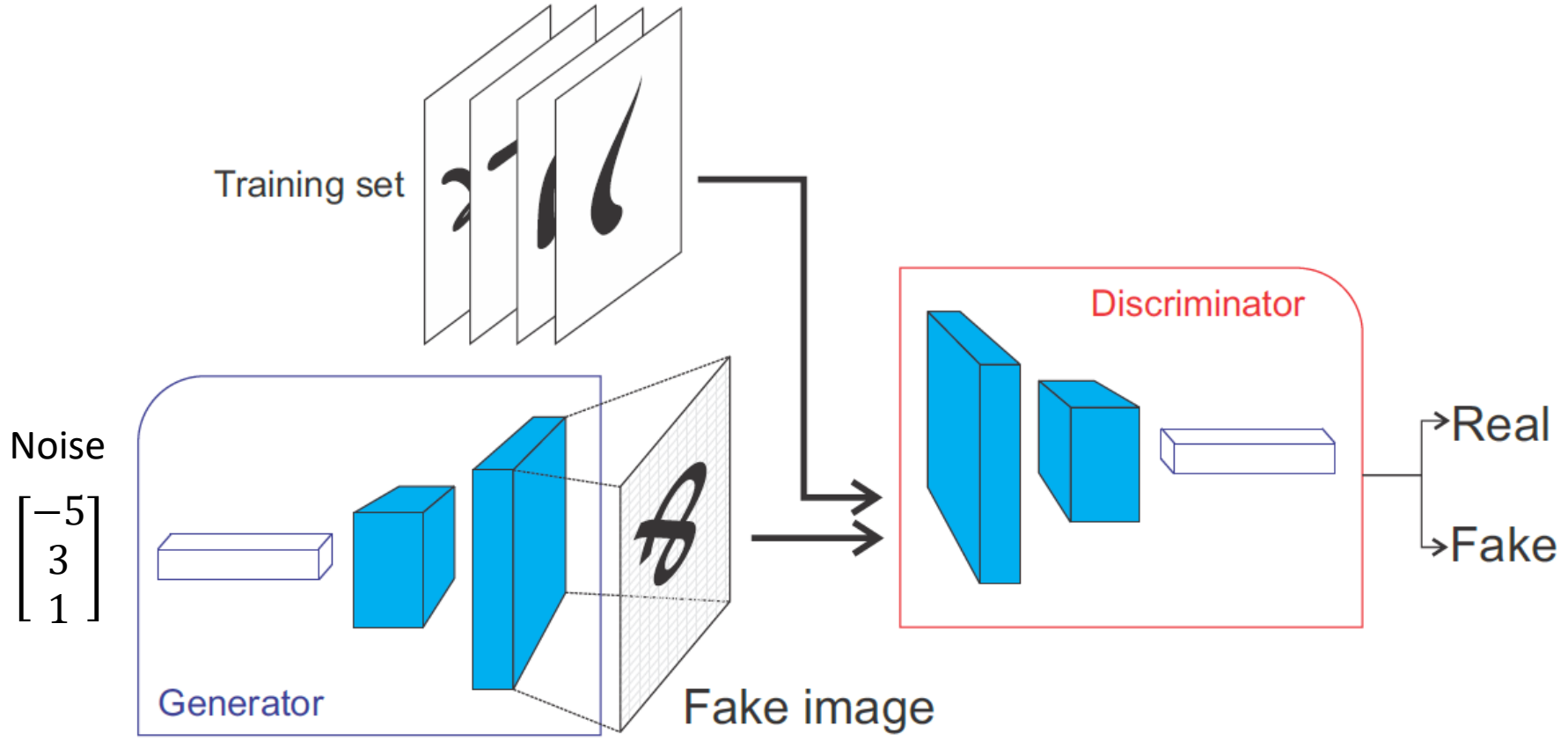


Generative Models – Training

- Sample random noise
- Produce output from sampled random noise
- Get discriminator "Real" or "Fake" classification for generator output
- Calculate loss from discriminator classification
- Backpropagate through discriminator and generator to obtain gradients
  - Change only the generator weights



GANs training summary



Question #2

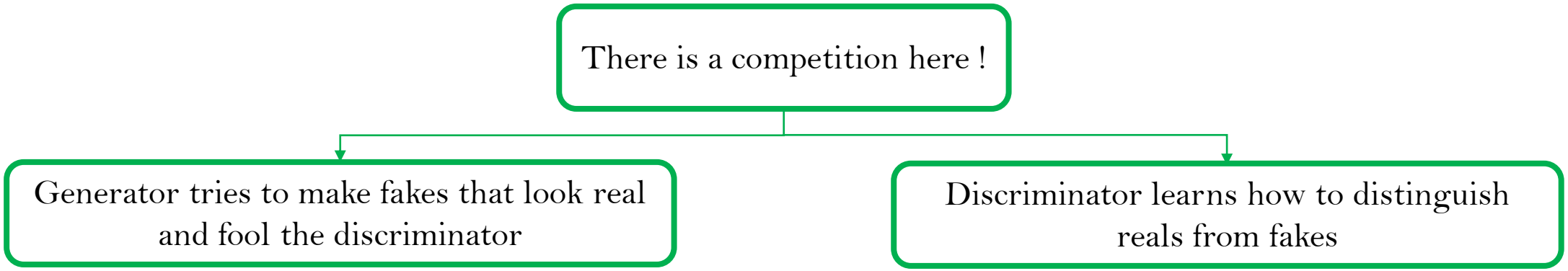
- GANs are trained by having a _____ and a _____, two separate networks that _____ against each other:

- (a) discriminator, generator, compete
- (b) discriminator, generator, collaborate
- (c) discriminator in the source data, discriminator in the target data, compete

Question #2

- GANs are trained by having a _____ and a _____, two separate networks that _____ each other:

- **(a) discriminator, generator, compete**



How do GANs Work?

- Mathematically, this can be expressed as a **MinMax** game, where the generator tries to **minimize** the objective function *V* while the discriminator tries to **maximize** it. Both networks are trained successively

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

$D(x)$ is a neural network that predicts how confident it is that the input image ***x*** is real. "The **probability** that the sample is real"

$G(z)$ is a neural network that generates an image given a noise signal *z*

How do GANs Work?

- When training the **generator**, we want to **minimize** the term in blue, i.e. we want to maximize the error that *D* will make on a generated image *G(z)*

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

- When training the **discriminator**, we want to **maximize** the terms in red, i.e. we want to minimize the error that *D* will make on a generated image *G(z).*

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

Question #3

- Select the true statement about GANs

  - (a) When training the generator, we want to minimize the error that the discriminator will make on a generated image
  - (b) When training the discriminator, we want to maximize the error that the discriminator will make on a generated image
  - (c) When training the discriminator, we want to minimize the error that the discriminator will make on a generated example

Question #3

- Select the true statement about GANs

  - (a) When training the generator, we want to minimize the error that the discriminator will make on a generated image
  - (b) When training the discriminator, we want to maximize the error that the discriminator will make on a generated image
  - **(c) When training the discriminator, we want to minimize the error that the discriminator will make on a generated example**

Question #4

- A typical GAN trains the generator and the discriminator simultaneously
  - True or False

Question #4

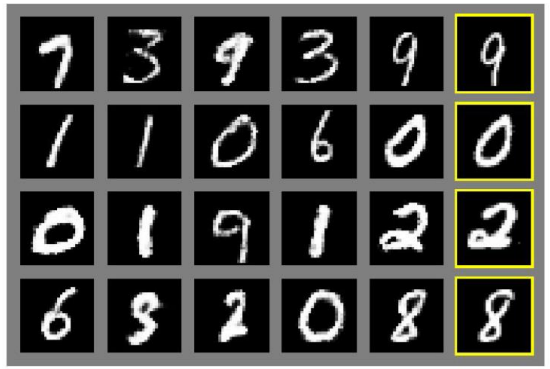
- A typical GAN trains the generator and the discriminator simultaneously

- True or **False:** A typical GAN alternates between training the discriminator and training the generator

- **The discriminator does not change during generator training**
  - Trying to hit a moving target would be a harder problem for the generator

Types of GANs and applications

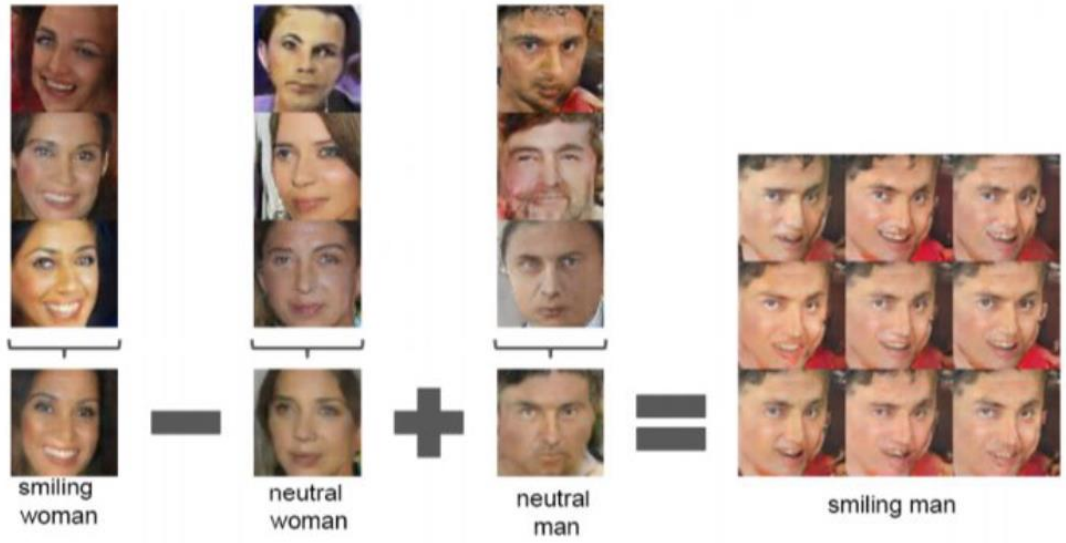
- In the original paper, both digits and faces were generated by the network and look convincingly real. The yellow boxes show the closest match to its generated neighbors in the training dataset



https://arxiv.org/pdf/1406.2661.pdf

DCGAN

- In a follow up paper, CNNs are used to generate the images



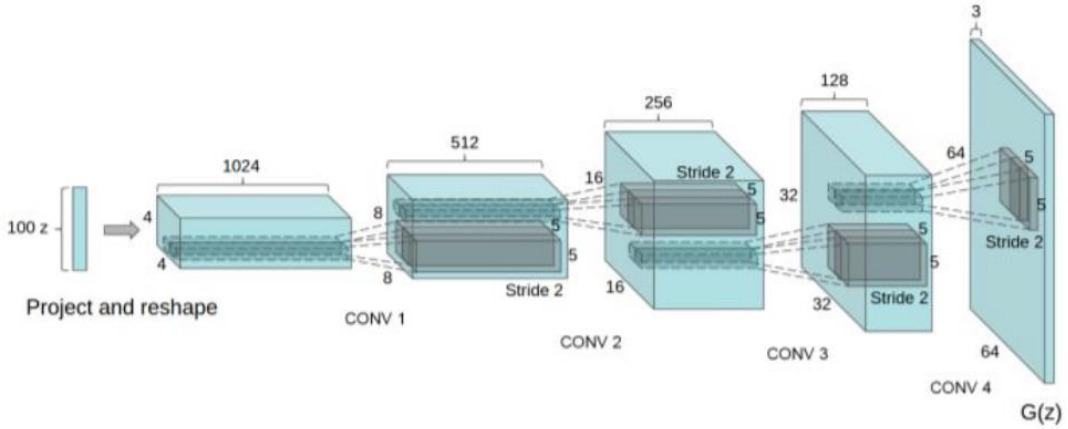


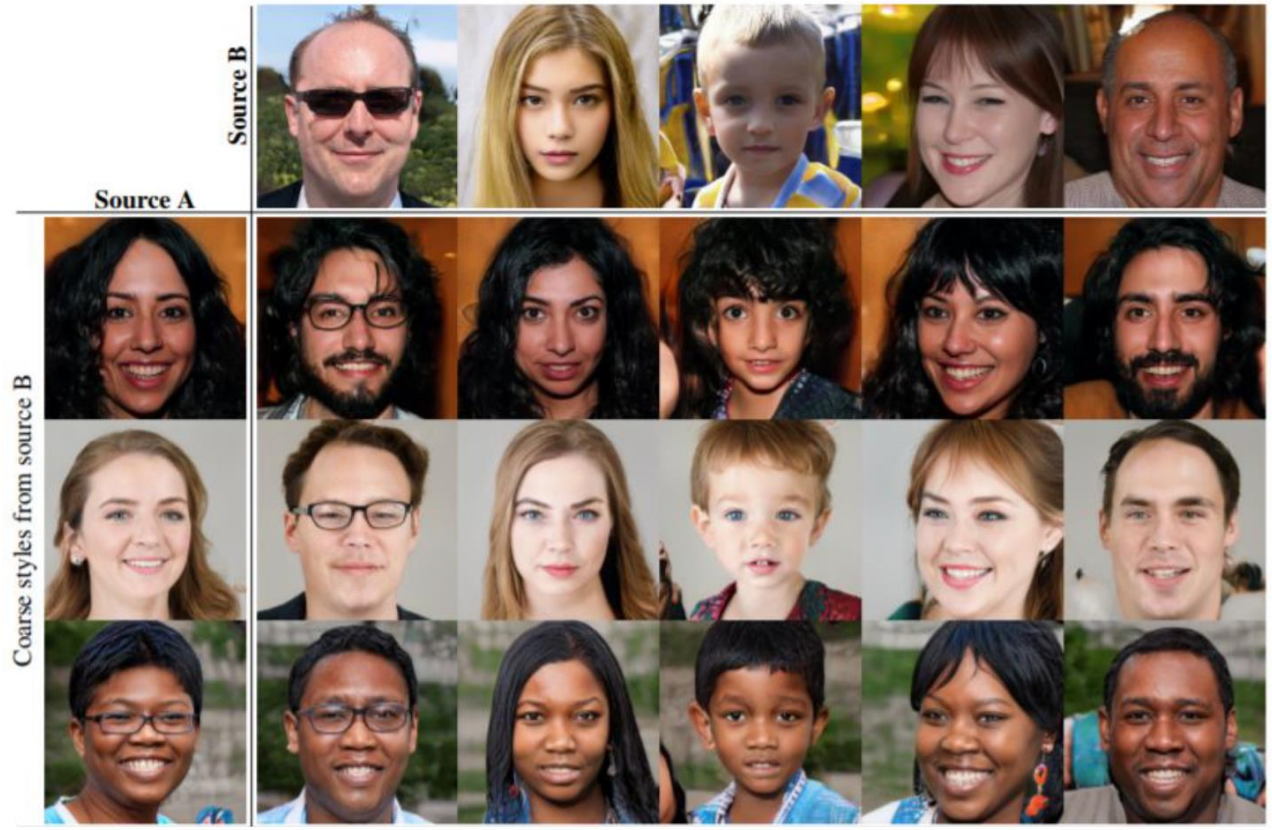
Figure 1: DCGAN generator used for LSUN scene modeling. A 100 dimensional uniform distribution Z is projected to a small spatial extent convolutional representation with many feature maps. A series of four fractionally-strided convolutions (in some recent papers, these are wrongly called deconvolutions) then convert this high level representation into a 64×64 pixel image. Notably, no fully connected or pooling layers are used.

https://arxiv.org/pdf/1511.06434.pdf

Image Synthesis

- We can teach networks to generate realistic images that have never been seen before by the network. Which face is fake?

Style Transfer



https://arxiv.org/pdf/1812.04948.pdf

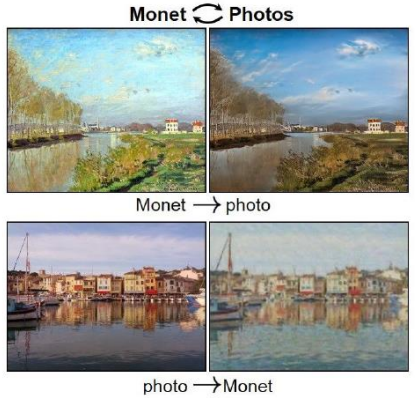
Face Inpainting

- Chunks of an image are blacked out, and the system tries to fill in the missing chunks
- GAN outperformed other techniques for inpainting images of faces

CycleGAN

- CycleGan introduced the concept of **cycle-consistency** loss to their GANs. This allowed training generators that would translate from domain X to domain Y
- This allows map between arbitrary domains with realistic results

$horse \rightarrow zebra$

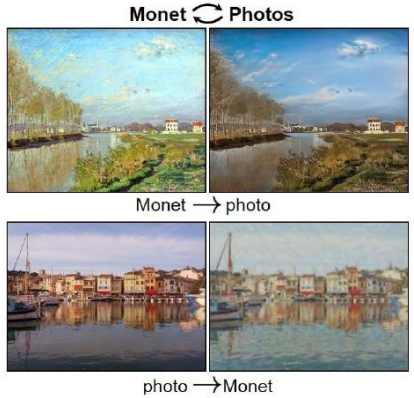


https://junyanz.github.io/CycleGAN/
https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix

CycleGAN

- The training data for the CycleGAN is simply two sets of images (a set of horse images and a set of zebra images). The system requires no labels or pairwise correspondences between images

$horse \rightarrow zebra$



https://junyanz.github.io/CycleGAN/
https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Challenge Report

Learning to synthesise the ageing brain without longitudinal data

Tian Xia[a,*], Agisilaos Chartsias[a], Chengjia Wang[b], Sotirios A. Tsaftaris[a,c], for the Alzheimer's Disease Neuroimaging Initiative

[a] Institute for Digital Communications, School of Engineering, University of Edinburgh, West Mains Rd, Edinburgh EH9 3FB, UK
[b] The BHF Centre for Cardiovascular Science, Edinburgh EH16 4TJ, UK
[c] The Alan Turing Institute, London NW1 2DB, UK

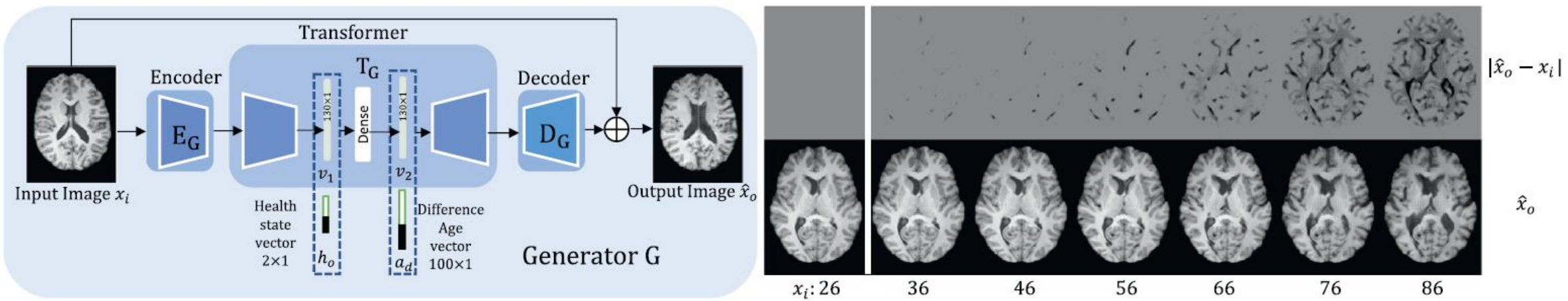
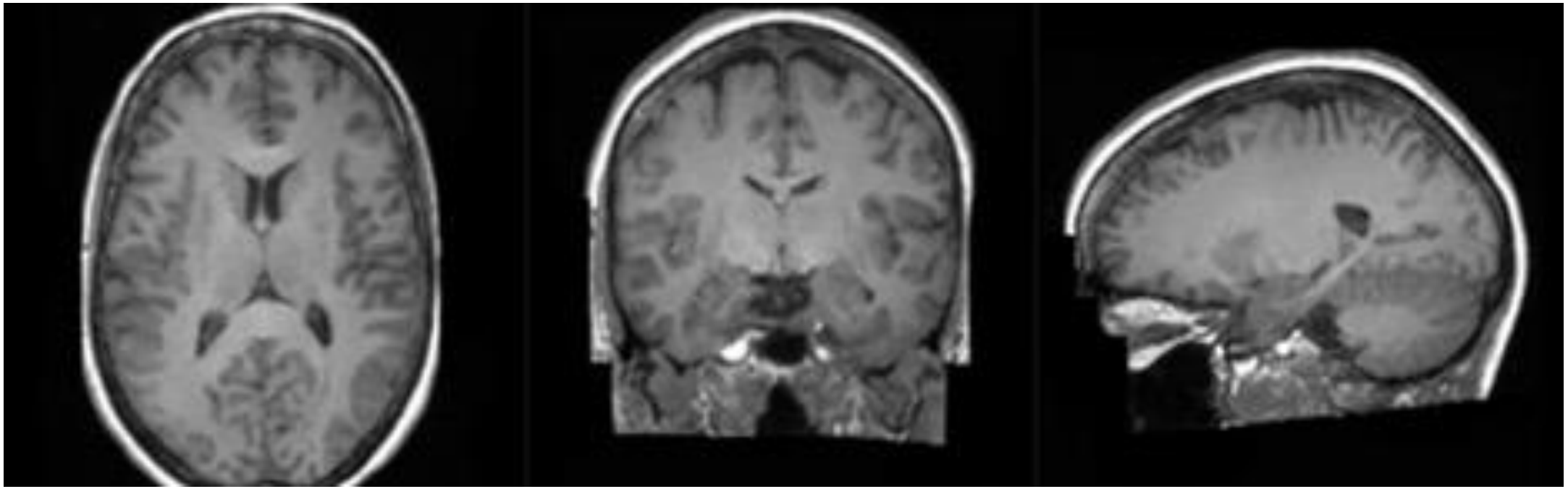


Fig. 1. Left: The input is a brain image x_i, and the network synthesises an aged brain image $\hat{x}_o$ from x_i, conditioned on the target health state vector h_o and target age difference $a_d = a_o - a_i$ between input a_i and target a_o ages, respectively. **Right:** For an image x_i of a 26 year old subject, bottom row shows outputs $\hat{x}_o$ given different target age. The top row shows the corresponding image differences $|\hat{x}_o - x_i|$ to highlight progressive changes.

Synthetic 3D medical images



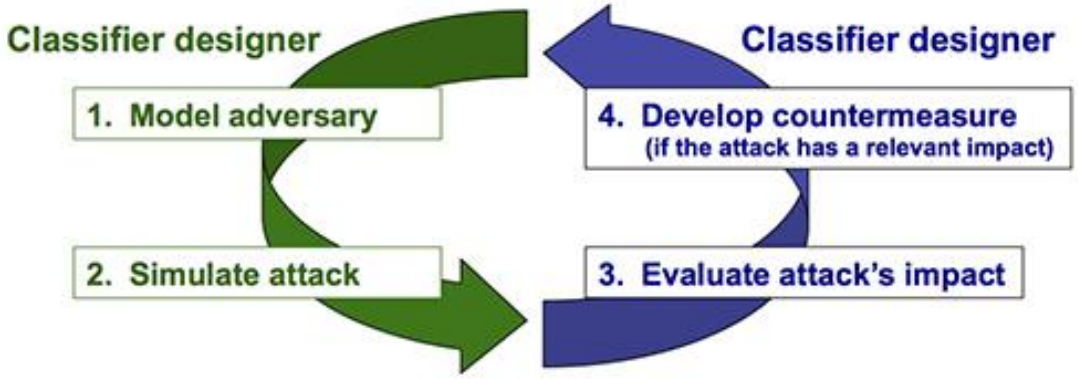
Cambridge-1 supercomputer

GANs Limitations

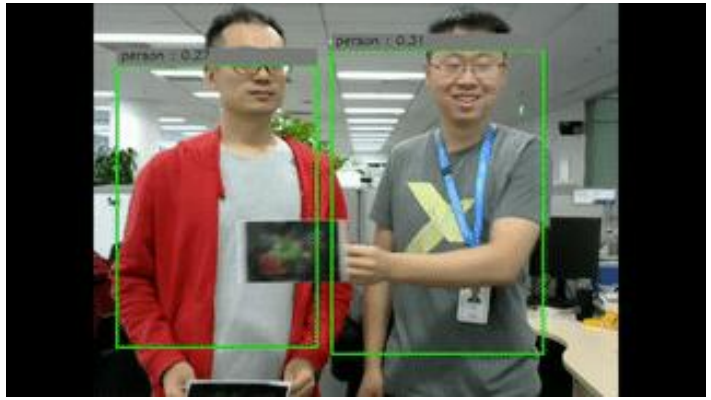
- **Non-convergence**: the model parameters oscillate, and the model does not converge
- **Mode collapse**: the generator collapses and produces a limited number of different samples
- **Diminished gradient**: the discriminator is too good that the generator gradient vanishes and learns nothing
- **Highly sensitive** to the hyperparameter selections

Adversarial Attacks

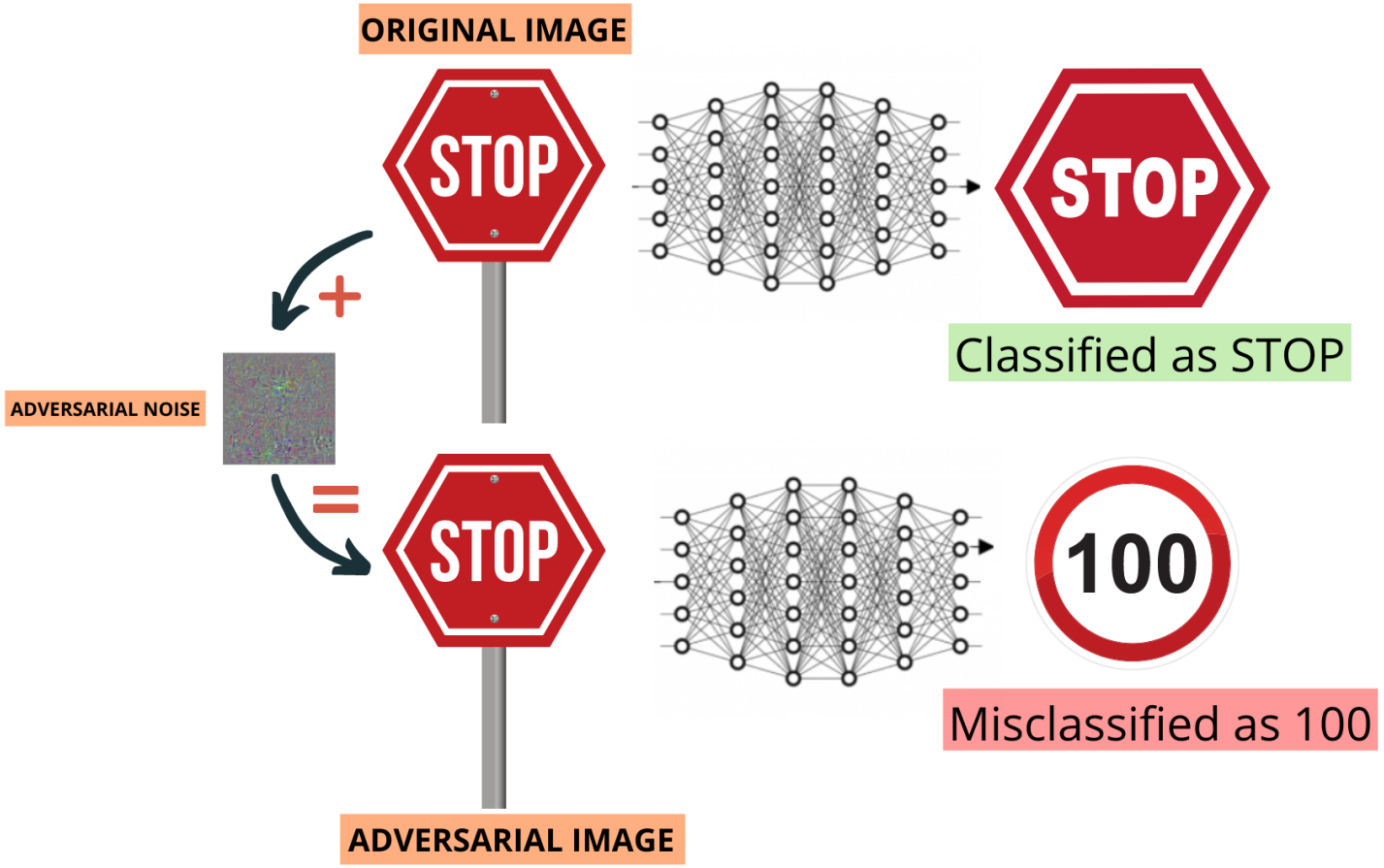
- Techniques that attempt to fool models with deceptive data
- It is a growing threat in the AI and machine learning research community.
- Goal: cause malfunctions in machine learning models
- Two types of attacks:
  - Feeding inaccurate or misrepresentative during training
  - Maliciously designed data to deceive an already trained model.



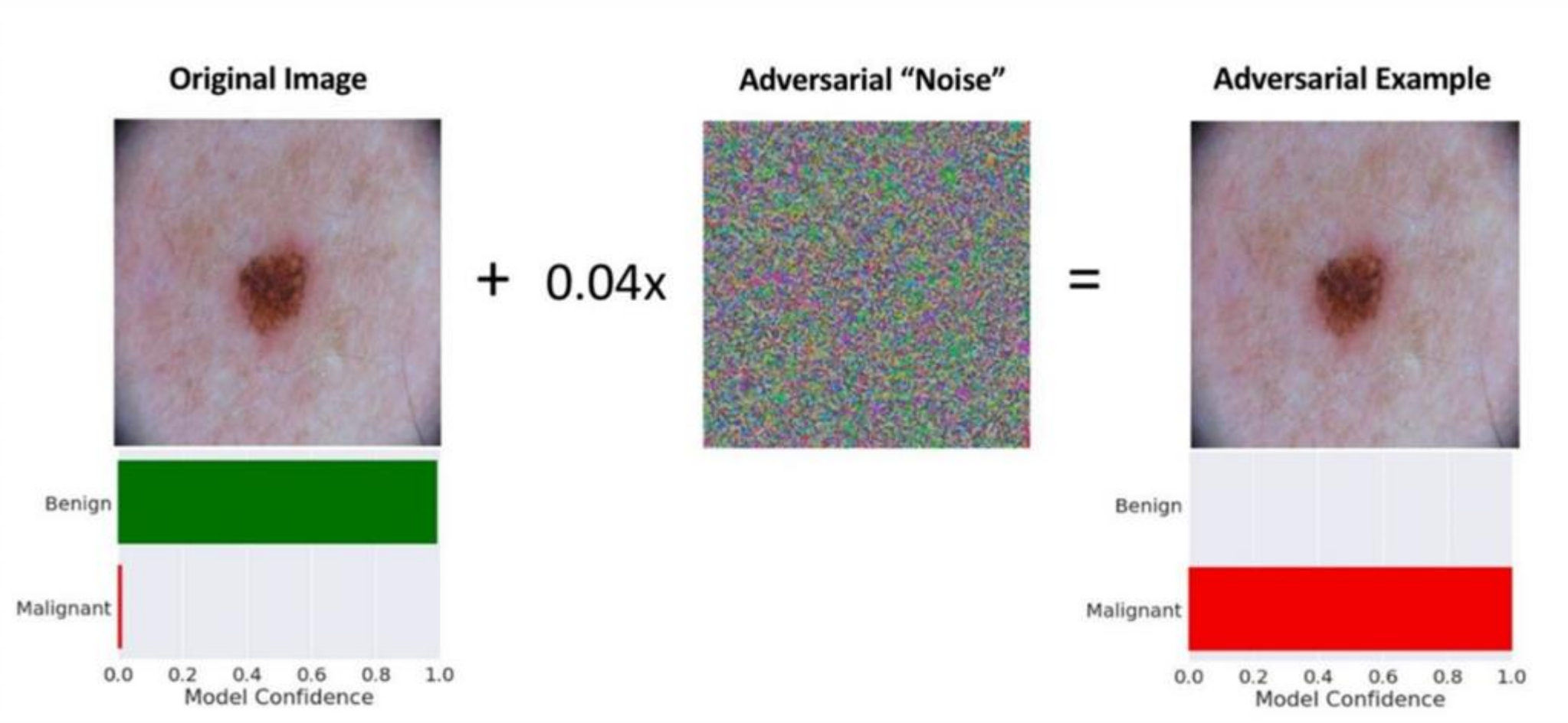
Adversarial Attacks



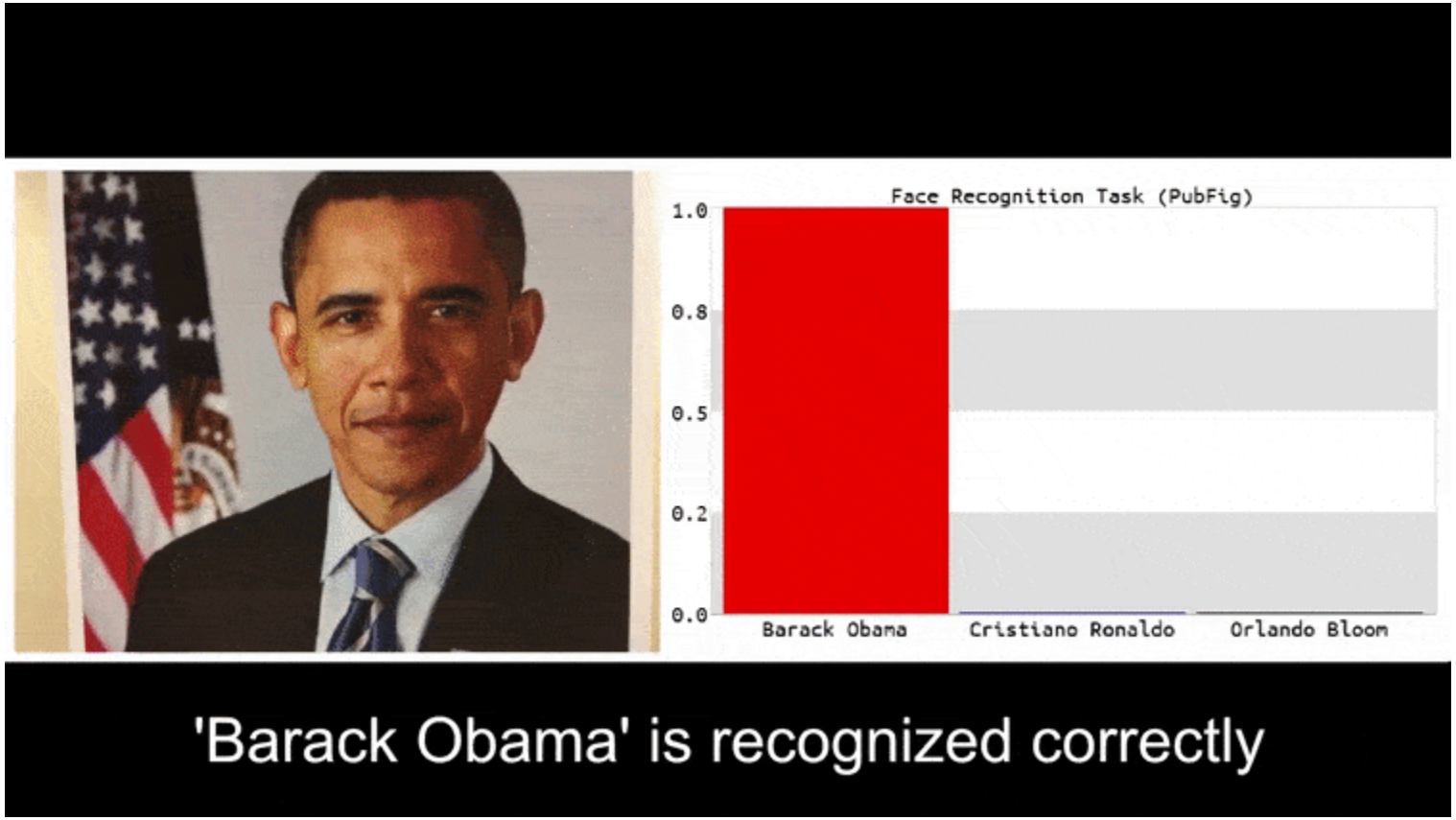
Adversarial Attacks



Adversarial Atatcks



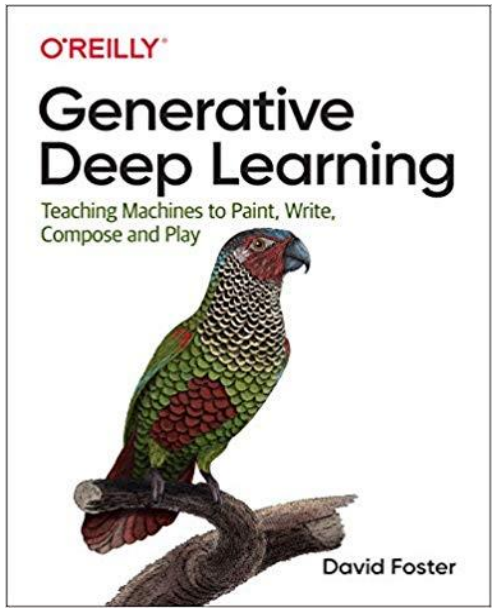
Adversarial Attacks



Summary

- GANs are unsupervised techniques
- They can be used to generate synthetic data that can potentially be used to train other deep learning models
- There are different GAN types - all based on the principle of having competing objectives
- GANs often face instabilities during training
- Adversarial attacks can be used fool machine learning systems

References



Adversarial Discriminative Domain Adaptation

Eric Tzeng
University of California, Berkeley
etzeng@eecs.berkeley.edu

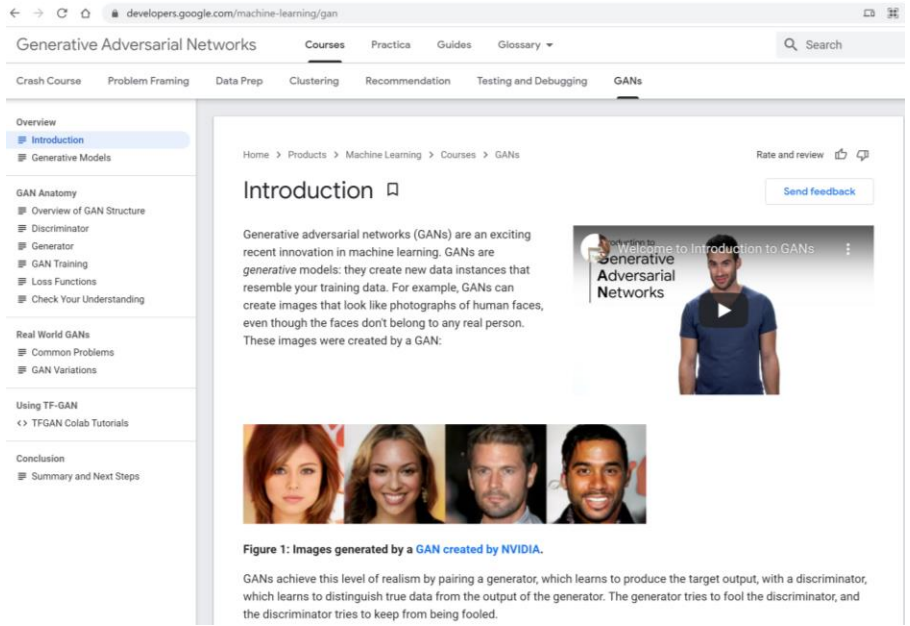
Judy Hoffman
Stanford University
jhoffman@cs.stanford.edu

Kate Saenko
Boston University
saenko@bu.edu

Trevor Darrell
University of California, Berkeley
trevor@eecs.berkeley.edu

Quebec Artificial Intelligence Institute

Mila



UNIVERSITY OF CALGARY

Thank you!

4 JORGES TODAY!