Lab CudaVision
Learning Vision Systems on Graphics Cards (MA-INF 4308)

Generative Adversarial Networks

13.01.2023

PROF. SVEN BEHNKE, ANGEL VILLAR-CORRALES

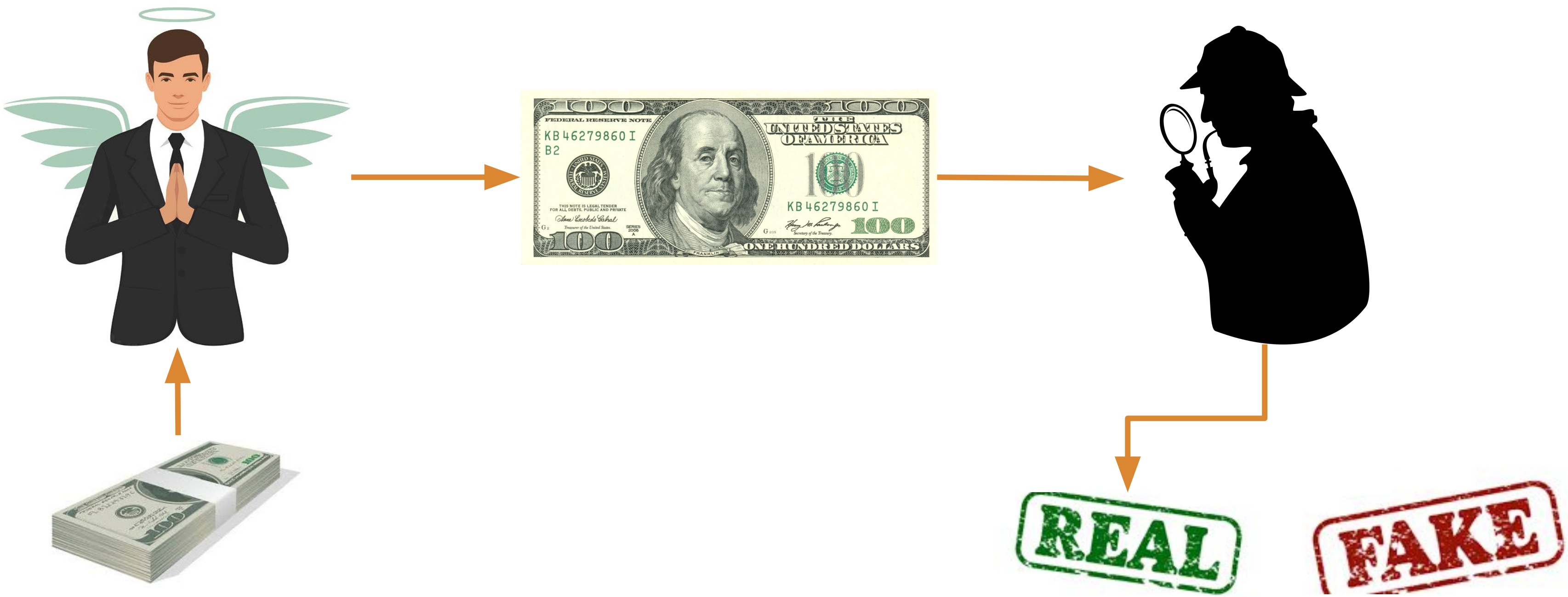
Contact: villar@ais.uni-bonn.de

GANs

Let’s Play a Game...

Generator

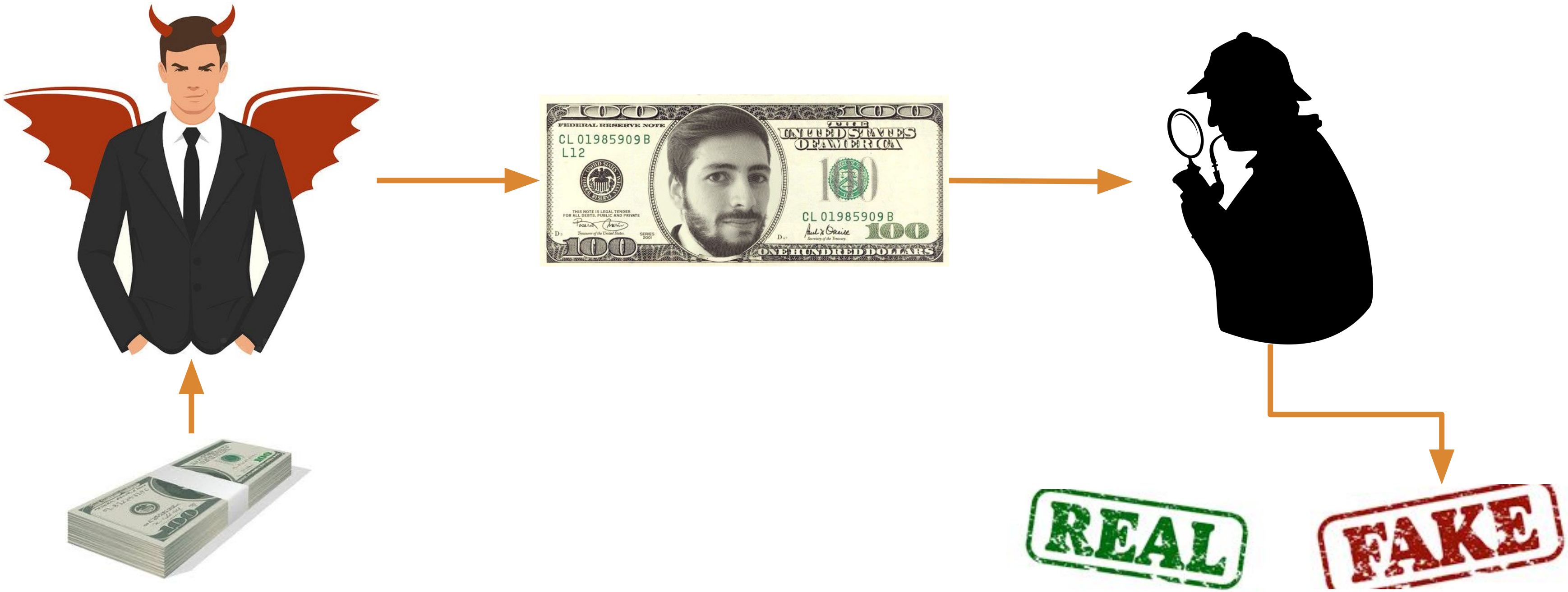
Discriminator



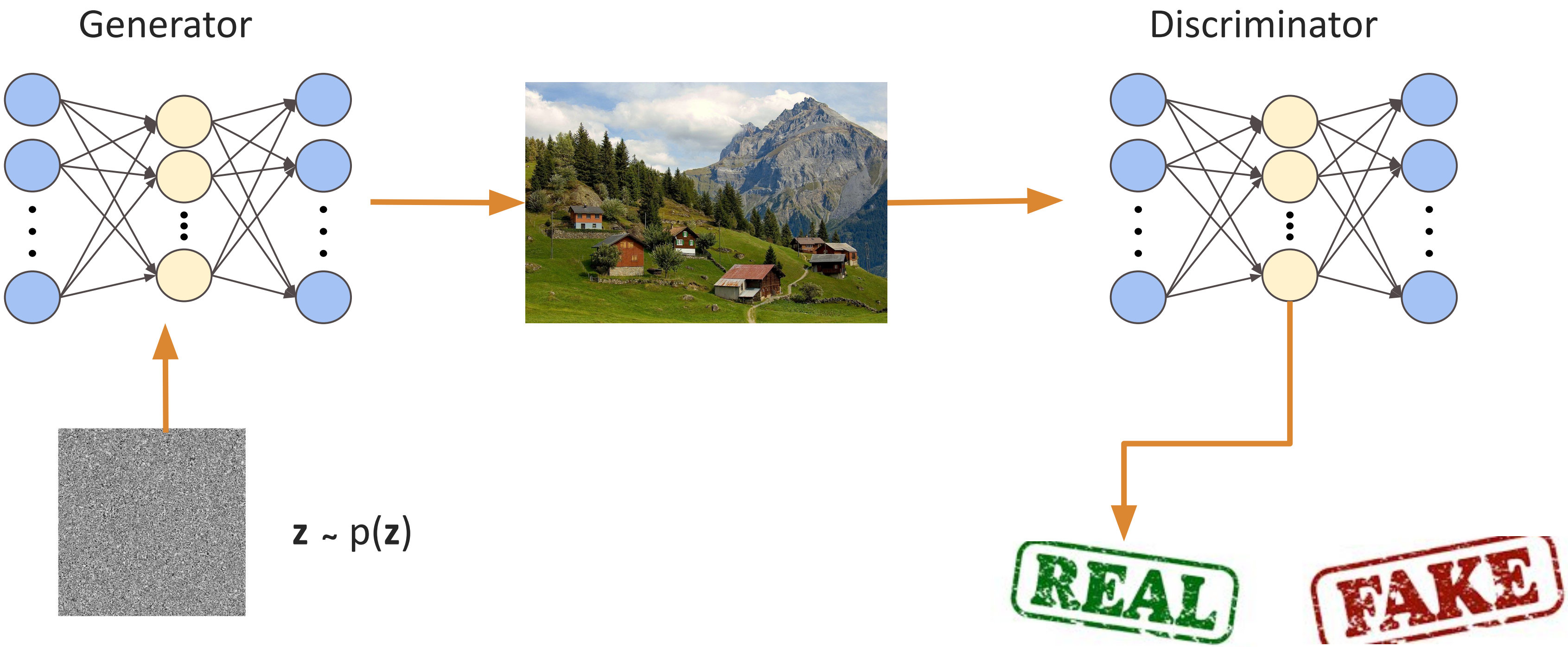
Let’s Play a Game...

Generator

Discriminator



Principle of GANs



Generative Adversarial Networks

Generative Adversarial Nets

Ian J. Goodfellow*, Jean Pouget-Abadie†, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair‡, Aaron Courville, Yoshua Bengio§
Département d'informatique et de recherche opérationnelle
Université de Montréal
Montréal, QC H3C 3J7

Abstract

We propose a new framework for estimating generative models via an adversarial process, in which we simultaneously train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player game. In the space of arbitrary functions G and D, a unique solution exists, with G recovering the training data distribution and D equal to $\frac{1}{2}$ everywhere. In the case where G and D are defined by multilayer perceptrons, the entire system can be trained with backpropagation. There is no need for any Markov chains or unrolled approximate inference networks during either training or generation of samples. Experiments demonstrate the potential of the framework through qualitative and quantitative evaluation of the generated samples.

1 Introduction

The promise of deep learning is to discover rich, hierarchical models [2] that represent probability distributions over the kinds of data encountered in artificial intelligence applications, such as natural images, audio waveforms containing speech, and symbols in natural language corpora. So far, the most striking successes in deep learning have involved discriminative models, usually those that map a high-dimensional, rich sensory input to a class label [14, 20]. These striking successes have primarily been based on the backpropagation and dropout algorithms, using piecewise linear units [17, 8, 9] which have a particularly well-behaved gradient . Deep *generative* models have had less of an impact, due to the difficulty of approximating many intractable probabilistic computations that arise in maximum likelihood estimation and related strategies, and due to difficulty of leveraging the benefits of piecewise linear units in the generative context. We propose a new generative model estimation procedure that sidesteps these difficulties. [1]

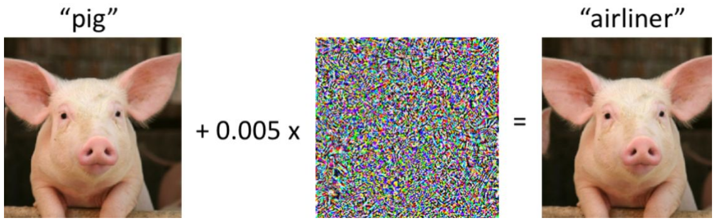
We propose a new framework for estimating generative models via an adversarial process, in which we simultaneously train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player game. In the space of arbitrary functions G and D, a unique solution exists, with G recovering the training data distribution and D equal to $\frac{1}{2}$ everywhere. In the case where G and D are defined

Why GANs?

- State-of-the-art* models in:
  - Image generation: **BigGan**
  - Text-to-Speech: **GAN-TTS**
  - Instrument score synthesis: **GANSynth**

- Understanding of adversarial attacks

- Wide use of adversarial training



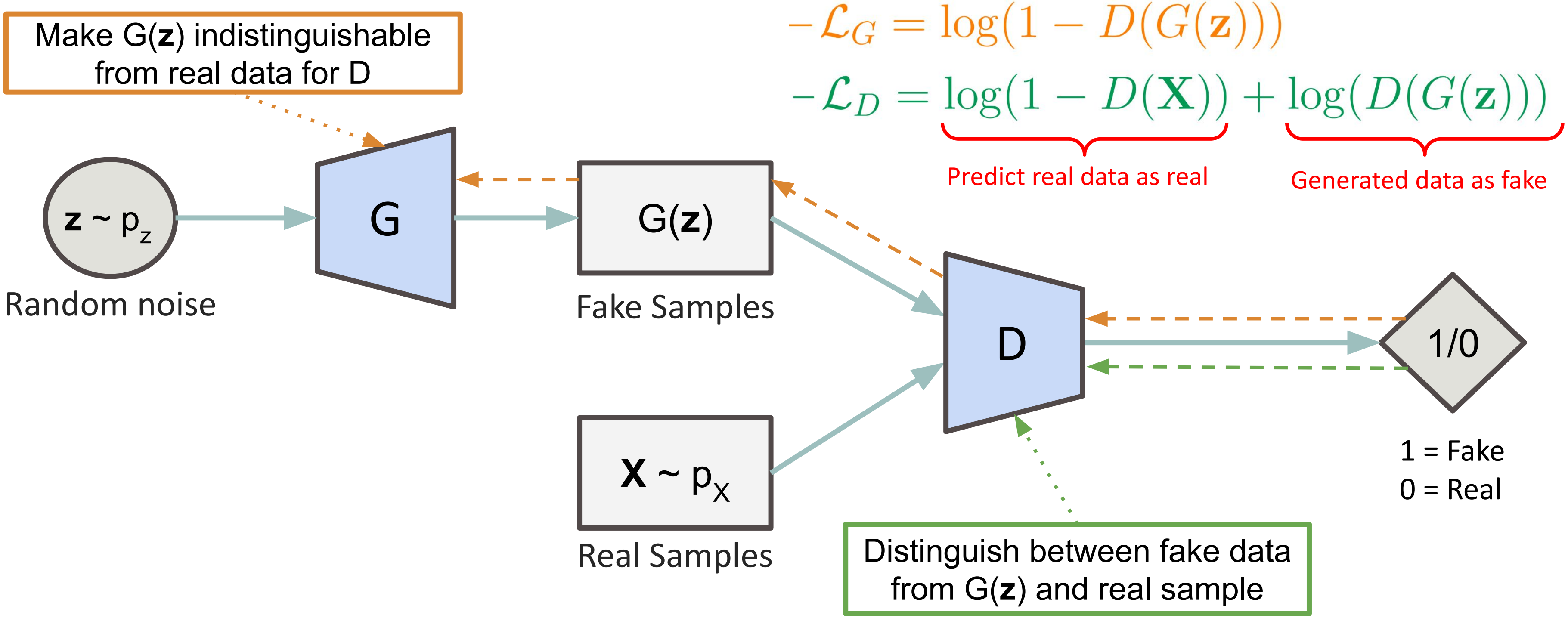
* State-of-the-art in 2020

Training GANs

Training GANs

$$-\mathcal{L}_G = \log(1 - D(G(\mathbf{z})))$$

$$-\mathcal{L}_D = \underbrace{\log(1 - D(\mathbf{X}))}_{\text{Predict real data as real}} + \underbrace{\log(D(G(\mathbf{z})))}_{\text{Generated data as fake}}$$



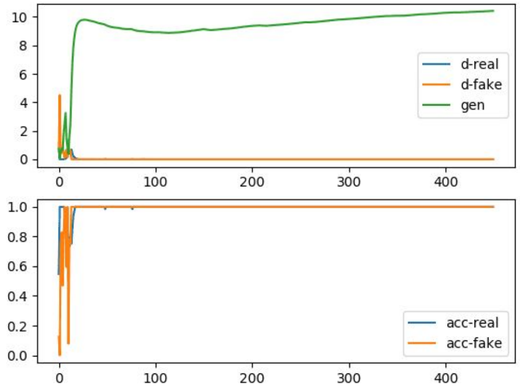
Training GANs

- Alternate between:
  1. Train Discriminator:
     - Minimize $-\mathcal{L}_D = \log(1 - D(\mathbf{X})) + \log(D(G(\mathbf{z})))$
  2. Train Generator:
     - Minimize $-\mathcal{L}_G = \log(1 - D(G(\mathbf{z})))$

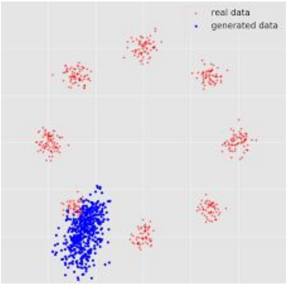
- Optionally run *k* steps for each model
- Trying to find an equilibrium between generator and discriminator

Training Difficulties

- Failure to converge
  - Discrimination is easier than generation
  - Unstable training of generator



- Mode collapse:
  - mapping several inputs to the same output
  - *“when something works, why change it?”*

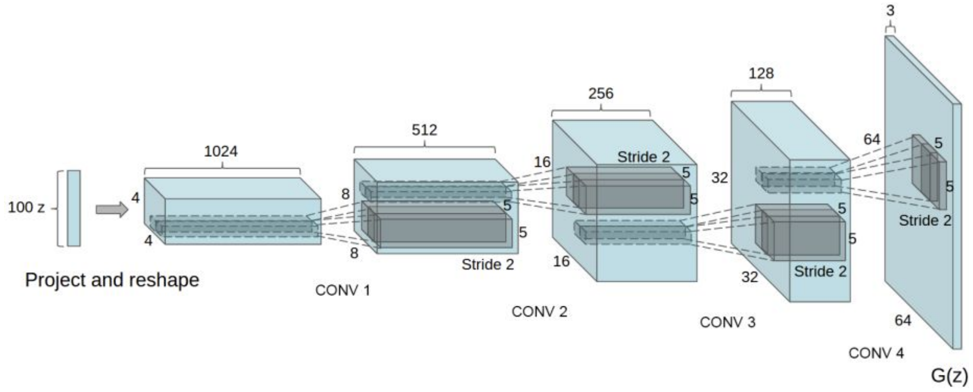


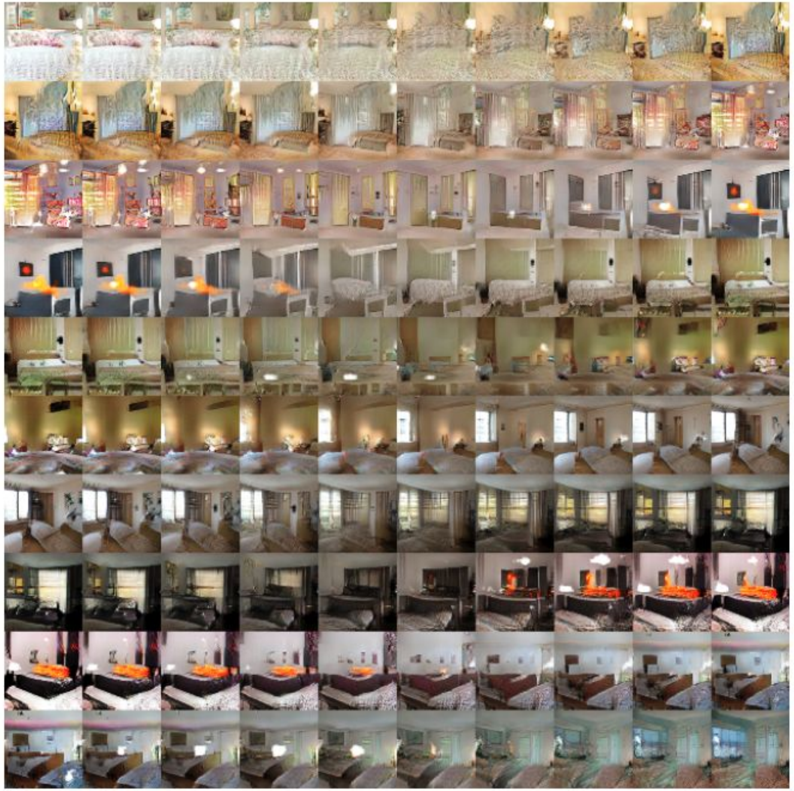
Popular GANs

Deep Convolutional GAN (DCGAN)

- Fully convolutional generator and discriminator
  - Strided convolutions instead of pooling
  - G: ReLU activation and TanH in output
  - D: Leaky ReLU activation
  - Batch normalization

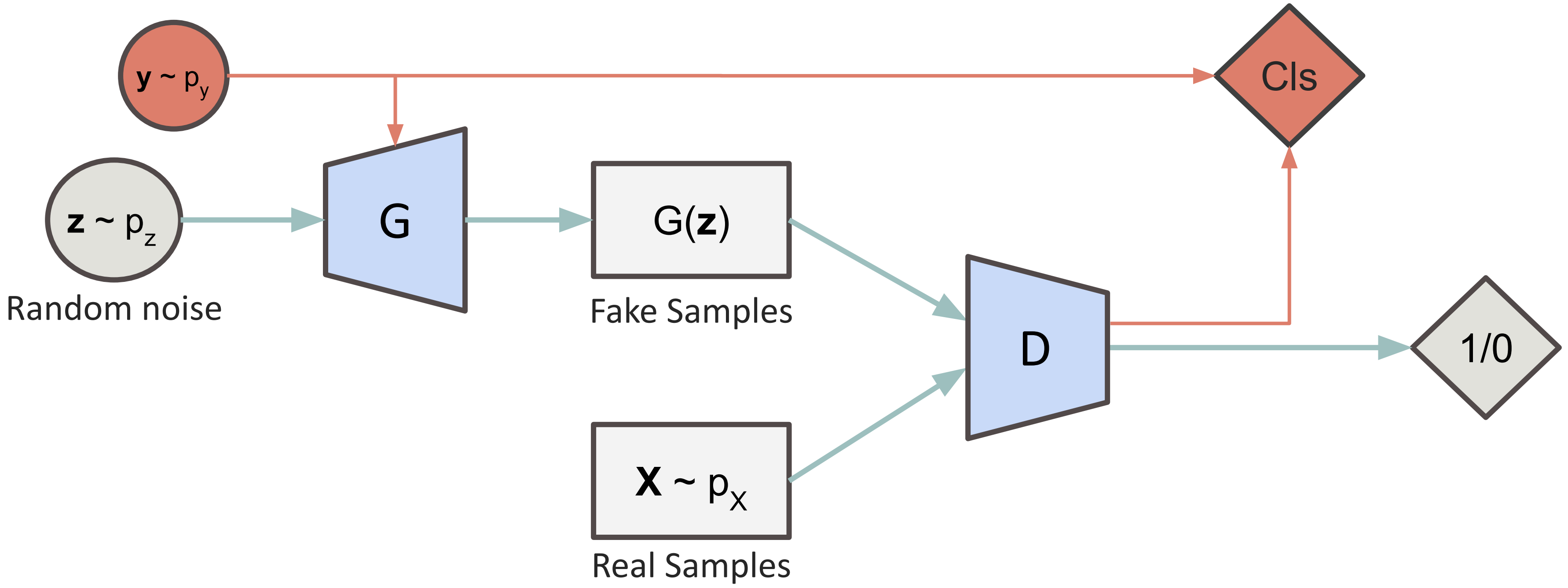
- Extremely popular architecture
  - DCGAN-like autoencoders
  - >11,000 citations





Epoch 1

Conditional GAN (CGAN)



Dog

Cat

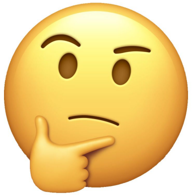
Tiger

Cycle Consistent GANs

- Model that performs image-to-image translation
- Paired images are expensive/impossible to obtain
- **Cycle consistency loss**: trainable inverse mapping F such that:

$$F(G(\mathbf{x})) \approx \mathbf{x} \quad \text{and} \quad G(F(\mathbf{y})) \approx \mathbf{y}$$

zebra ⟶ horse

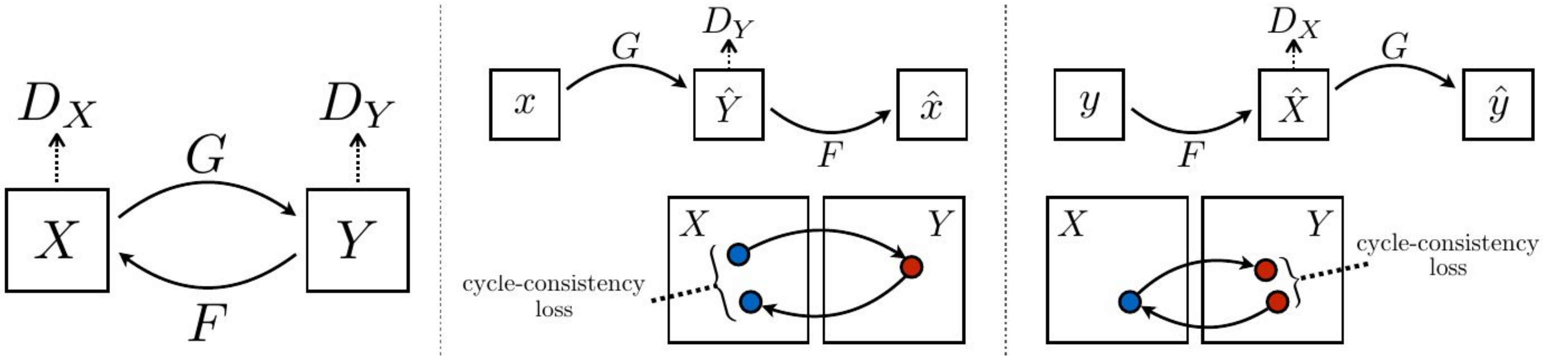


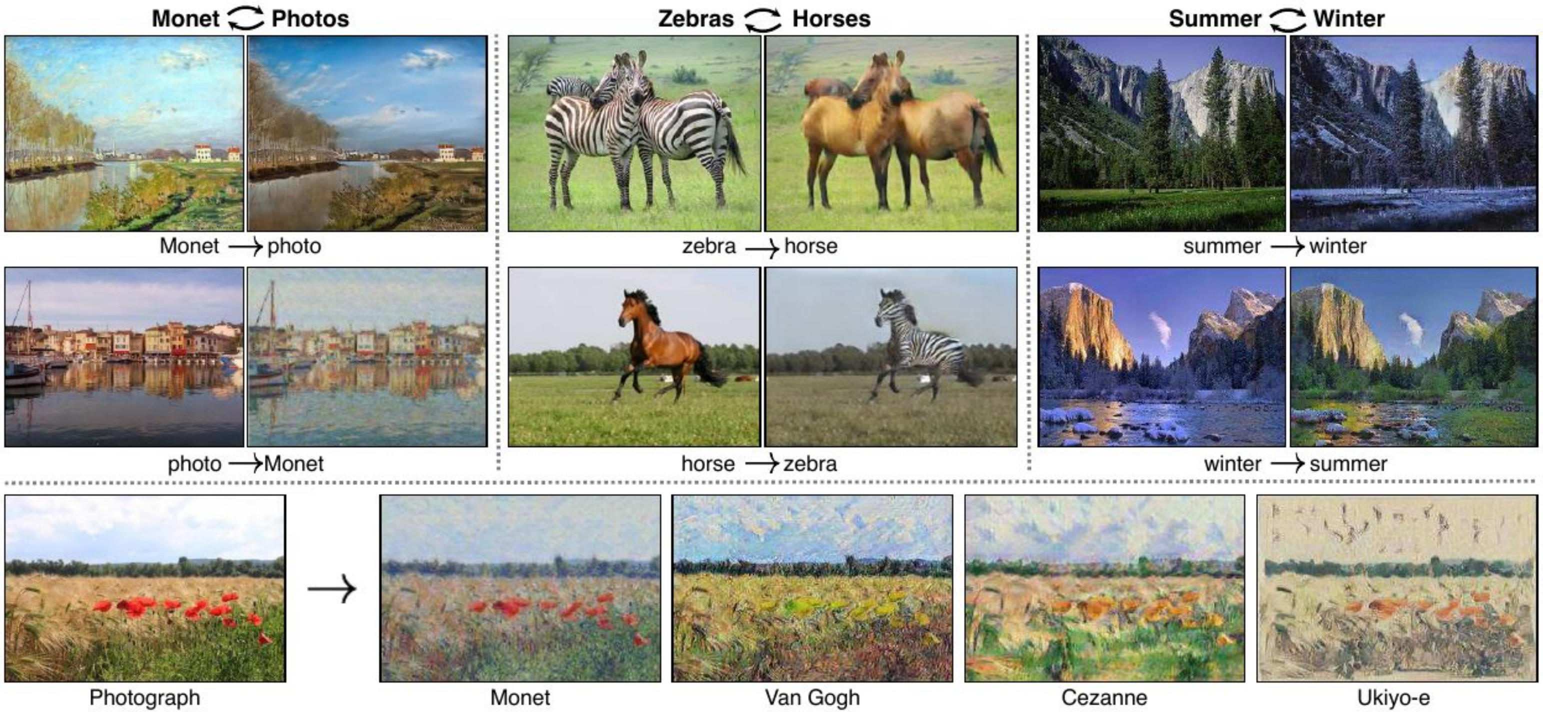
Cycle Consistent GANs

- Use two generators (G and F) and two discriminators (D_X and D_Y)

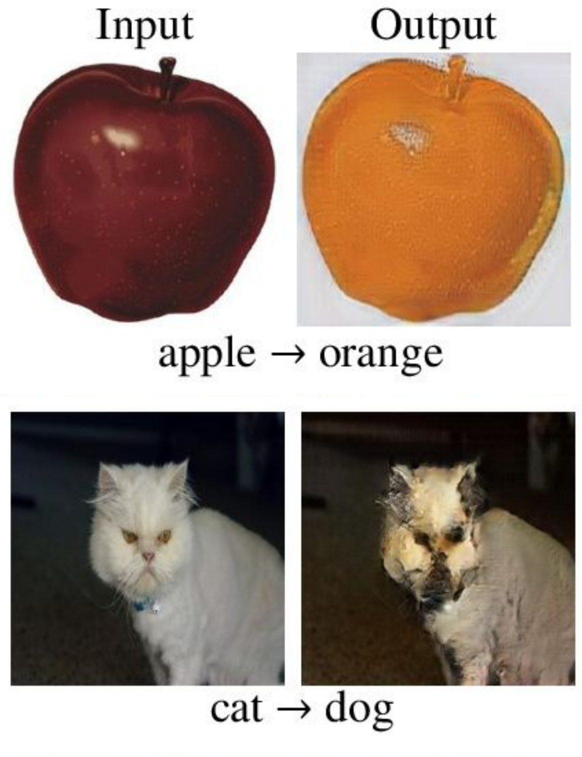
$$\mathcal{L}_{Cyc}(G, F) = ||F(G(\mathbf{x})) - \mathbf{x}||_1 + ||G(F(\mathbf{y})) - \mathbf{y}||_1$$

$$\mathcal{L} = \mathcal{L}_{GAN}(G, D_X) + \mathcal{L}_{GAN}(F, D_Y) + \lambda \mathcal{L}_{Cyc}(G, F)$$





Monet ⇄ Photos
Zebras ⇄ Horses
Summer ⇄ Winter
Monet → photo
zebra → horse
summer → winter
photo → Monet
horse → zebra
winter → summer
Photograph
→
Monet
Van Gogh
Cezanne
Ukiyo-e


Input
Output
apple → orange
cat → dog

References

1. https://towardsdatascience.com/all-you-want-to-know-about-deep-learning-8d68dcffc258
2. Goodfellow, Ian, et al. Deep learning. Vol. 1. No. 2. Cambridge: MIT press, 2016.
3. Brock, Andrew, Jeff Donahue, and Karen Simonyan. "Large scale GAN training for high fidelity natural image synthesis." arXiv preprint arXiv:1809.11096 (2018).
4. Bińkowski, Mikołaj, et al. "High fidelity speech synthesis with adversarial networks." arXiv preprint arXiv:1909.11646 (2019).
5. Engel, Jesse, et al. "Gansynth: Adversarial neural audio synthesis." arXiv preprint arXiv:1902.08710 (2019).
6. https://salu133445.github.io/ismir2019tutorial/
7. https://machinelearningmastery.com/practical-guide-to-gan-failure-modes/
8. Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434 (2015).

References

9. Mirza, Mehdi, and Simon Osindero. "Conditional generative adversarial nets." arXiv preprint arXiv:1411.1784 (2014).
10. Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." Proceedings of the IEEE international conference on computer vision. 2017.