From art to deep fakes: an introduction to Generative Adversarial Networks

Machine Learning course 2019/2020

Gabriele Graffieti

gabriele.graffieti@unibo.it

PhD student in Data Science and Computation @ University of Bologna

December 12, 2019

About me

- PhD Student in DS&C @ University of Bologna.
- Head of AI research @ AIforPeople.
- Teaching Assistant of the course Algorithms and Data Structures @ UniBo.
- Board Member of Data Science Bologna.
- Proud Member of *ContinualAI*.

Summary

1. Introduction
2. Generative models
3. Generative Adversarial Networks
4. Research Directions

The limits of AI I

AI already outperforms human abilities in many tasks

- Computer Vision (classification, segmentation, medical image analysis, image reconstruction, super-resolution . . .)
- Forecasting (stock market prediction, traffic, people you may know, advertising, . . .)
- Fraud detection, marketing, health monitoring, drug discovery, DNA sequence classification, robot locomotion, play games (chess, go, starcraft) and MANY more.

And we are still far from general AI!

The limits of AI II

There is something that AI can never be able to do?

The limits of AI III

My answer: be creative

- Create music.
- Create art.

Generative Models

(a misleading name)

Discriminative vs Generative models I

Discriminative Models

- The majority of models used in machine learning.
- They model the conditional probability $P(Y|X)$.
- They learn the boundaries between classes.
- The conditional probability is directly estimated from data.
- More effective on classification tasks.
- Most discriminative models are inherently supervised.

Discriminative vs Generative models II

Generative Models

- They model the join probability $P(X, Y)$.
- They learn the distribution of the data.
- More general than discriminative models (a DM can always be derived from a GM).
- Since they learn a distribution as similar as possible to the real distribution of data, it is possible to sample artificial data from them.
- They can handle multi-modal output, where more than one Y is the correct prediction for a single X (e.g. X is a frame in a video, and Y is the predicted next frame).
- Usually unsupervised.

Deriving a Discriminative Model from a Generative Model

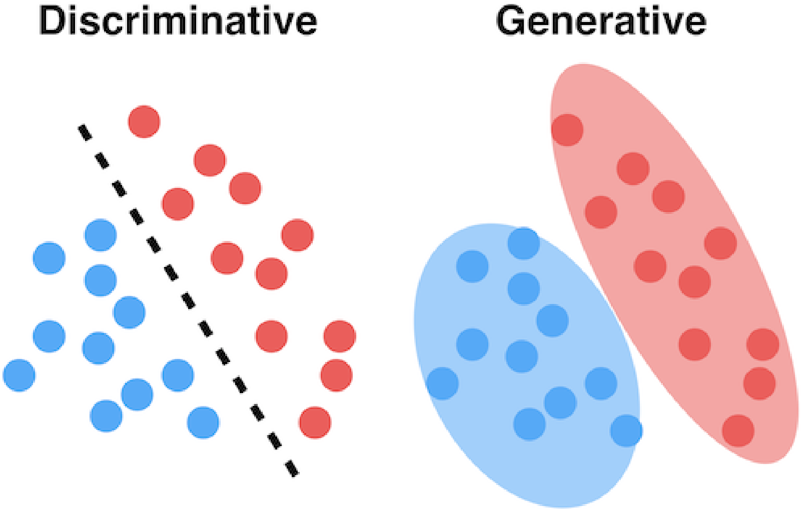
Bayes Rule:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

We have that:

$$P(X,Y) = P(X|Y)P(Y) \quad \text{and} \quad P(X) = \sum_y P(X, Y = y)$$

Discriminative vs Generative models III



Discriminative vs Generative models, a different perspective

- Suppose only 2 classes of animals exists in the world: cats and rabbits, and we want classify them.
- Discriminative model → tries to find a boundary between the two classes.
    - When new images are presented to the DM, it simply check which side of the decision boundary the animal should go.
- Generative model → develops a model of what a cat and rabbit should look like.
    - New images are then classified based on whether they look more like the cat model the GM developed or the rabbit model the GM developed during training.

Generative Adversarial Networks

Generative Models (recap) I

They model the join probability $P(X, Y) = P(X|Y)P(Y)$

Explicit Models

- Tractable explicit models → the density function of the model is chosen to be computationally tractable.
  - Fully visible belief nets, nonlinear ICA, pixelRNN.
- Approximated explicit models → Often the real distribution of the data is too complex to deal with, thus a simpler and tractable function is used as a lower bound.
  - Variational autoencoders, Boltzmann machines.

Generative Models (recap) II

Too many dimensions!

- What if we want to model the distribution of *cat* images in order to generate new super-cute imaginary cats?
  - The space of 64×64 RGB images have $12,288$ dimensions.
  - The space of 1920×1080 (full HD) RGB images have $6,220,800$ dimensions!
- It's impossible to deal with functions in such high-dimensional spaces.

Implicit Models

- We don't care about the actual distribution of data, we just want to produce images as similar as possible to the real ones.
- The model of the PDF exists, but it's hidden (e.g. in the weights of a neural network).

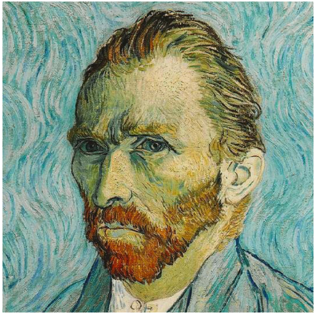
Let's do it!

Just take a super-pro neural network, fed it with some noise, use the magic backpropagation algorithm and the network will learn to produce hyper-realistic cat images!

Yeah boy! But we need a loss!

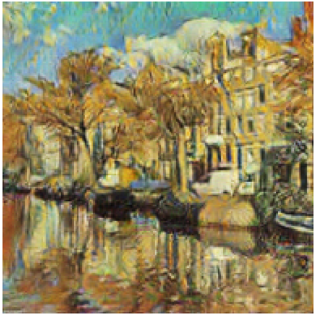
What is reality? I

How can we mathematically define how much an image of a cat is real? How can we define the concept of *catness*? And the concept of *van goghness*? Or the concept of *Trumpness*?

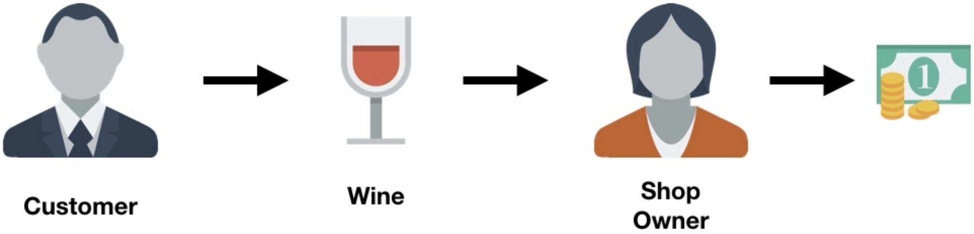


What is reality? II

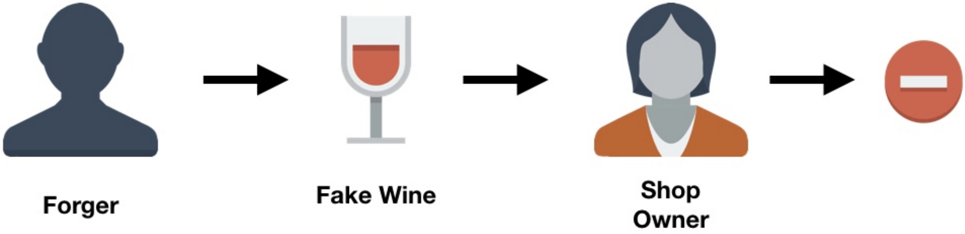
We have seen a lot of cats in our life, so our brain have modeled some sort of statistical model that describes *how a cat looks like*. The same for Van Gogh portraits or images of Donald Trump.



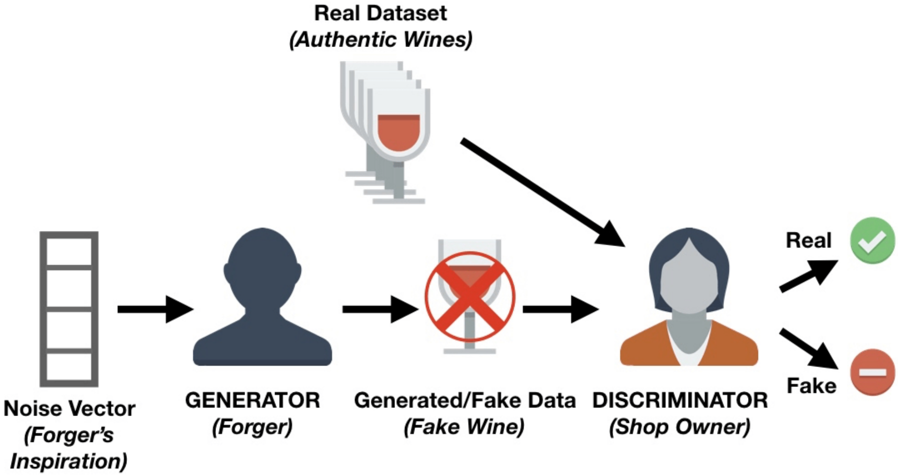
Proto-GAN example I



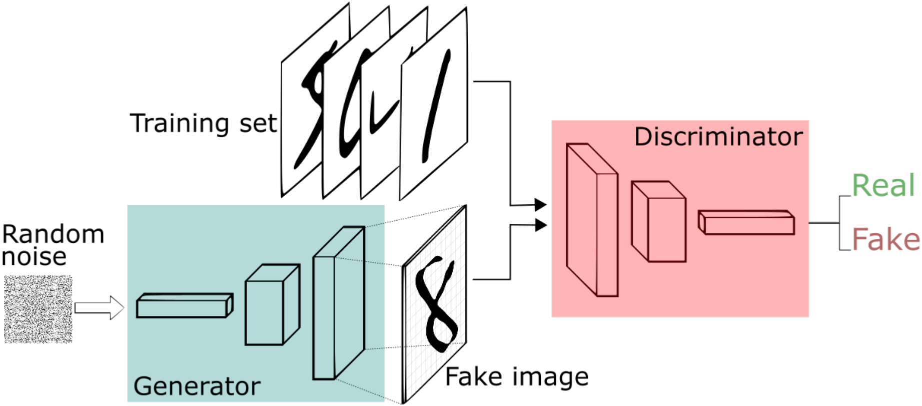
Proto-GAN example II



Proto-GAN example III



GAN model I


Training set
Random noise
Generator
Fake image
Discriminator
Real
Fake

GAN model II

Generator

- Takes as input a noise vector, and output a sample as similar as possible to real data.
- Unsupervised learning.
- (Implicit) Generative model.

Discriminator

- Takes as input real and fake data, and tries to discriminate between them.
- Supervised learning.
- Discriminative model.

Your greatest enemy is your best friend I

Let x = real sample, z = noise vector.

- Suppose the discriminator outputs a probability of an image to be real. Let's say that $1 = 100\%$ real; $0 = 100\%$ fake; $0.5 =$ I don't know.
- The discriminator want to minimize the function $(D(x) - 1) + D(G(z))$.
- The generator want to maximize the function $D(G(z))$

Your greatest enemy is your best friend II

$$\mathcal{L}_{GAN}(D, G) = (1 - D(x)) + D(G(z))$$

The GAN game:

$$\min_{D} \max_{G} \mathcal{L}_{GAN}(D, G)$$

- The goal of a minimax game is to find an equilibrium point.
- The two components pull the equilibrium in two opposed directions.
- The solution of this game is the equilibrium point (Nash equilibrium), where the discriminator outputs 0.5 for every input.

Your greatest enemy is your best friend III

- The goal of the generator is to maximally deceive the discriminator → modeling a distribution identical to the real data distribution.
- The goal of the discriminator is to accurately distinguish between real and fake data → find the best boundary to discriminate the real and fake distributions of data.

- The generator don't see real data! The only feedback comes from the discriminator.
- The discriminator share its gradient with the generator.

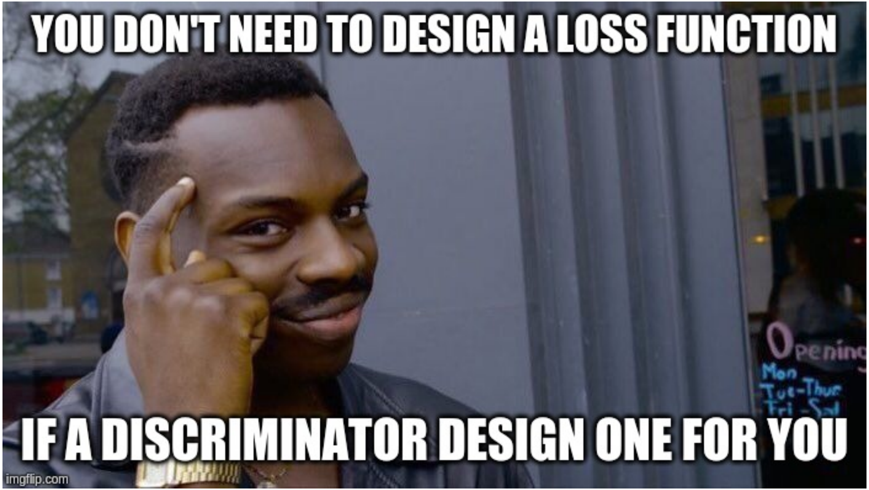
Why GANs are everywhere?

There is no hand-crafted loss!

The generator can potentially learn to mimic every type of data using the same game!

Just change the training data and the same network can generate cats or Van Gogh portraits!

Why GANs are everywhere?


YOU DON'T NEED TO DESIGN A LOSS FUNCTION
IF A DISCRIMINATOR DESIGN ONE FOR YOU
imgflip.com
Opening
Mon
Tue-Thur
Fri - Sat

Let's make it real!

https://github.com/ggraffieti/GAN-tutorial-DSBo

Research Directions

Problems

- Instability during training → intrinsic in the problem definition (find an equilibrium point).
- Non-convergence → the discriminator task is usually more simple than the generator's one.
- Mode collapse → the generator produce almost the same output for every input.
- Evaluation of results → how we can assess the grade of reality of the output data?
  - Human evaluation is still preferred (Amazon Mechanical Turk).

Image-to-image translation I

- Like a language translation, we want to translate one image from one domain to another, maintaining unchanged the semantic content.
- The ground truth result of the translation is not known, and more than one result might be correct.

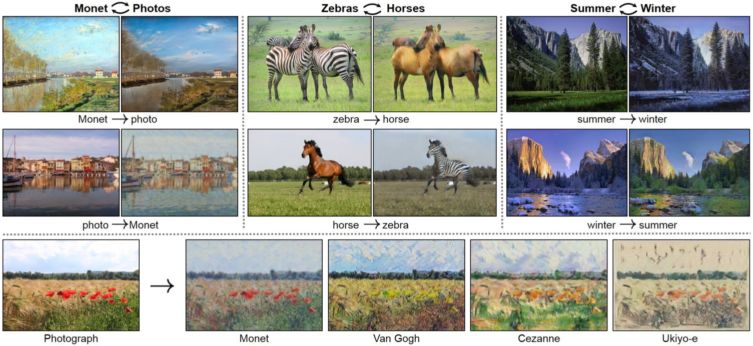
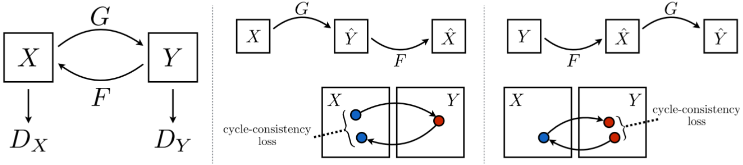


Image-to-image translation II

- The problem is similar to the problem of learning to translate between 2 languages without a dictionary.
- We could exploit the *cycle consistency* properties of mappings:

$$F(G(x)) \approx x \quad and \quad G(F(y)) \approx y$$

- We force the model to learn the *inverse mapping*, and the composition of the 2 translation have to be similar to the input.



Photoshop 2.0?

- `http://nvidia-research-mingyuliu.com/gaugan`
- `http://gandissect.res.ibm.com/ganpaint.html`

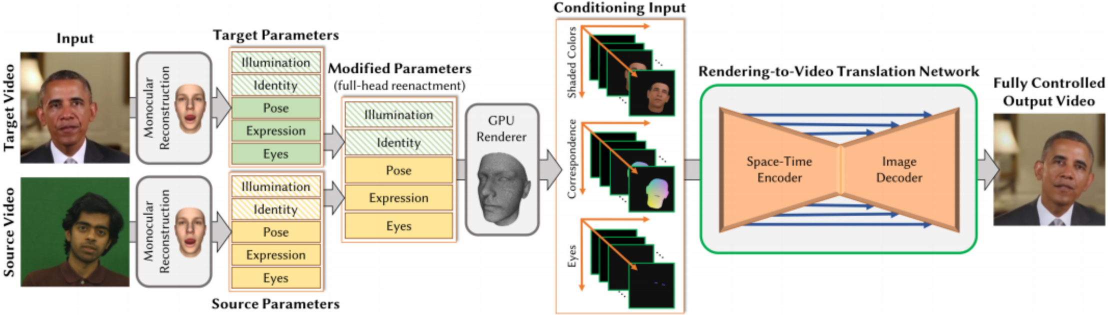
Deepfakes I

Similar to an image-to-image (or a video-to-video) translation task, but with more constraints:

- We want to maintain the realism of the scene → GAN
- We want to maintain the semantic content of the scene → cycleGAN
- We want to maintain the identity of the person in the target scene → identity recognition network.
- We want to maintain the facial expression of the person in the source scene → facial landmarks comparison.

Deepfakes II

(Fortunately) Models that produces deepfakes are fairly complicated.



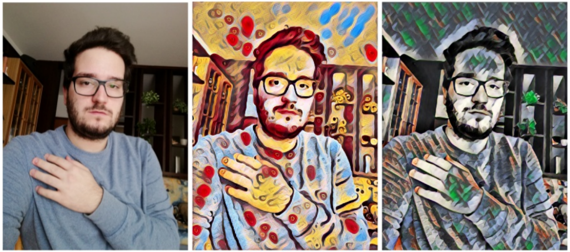
The results are not perfect yet, but these models are continuously improving.

Deepfakes III

- `https://www.youtube.com/watch?v=qc5P2bvfl44` (deep video portraits)
- `https://www.youtube.com/watch?v=KHKVTDbR5Ig` (few shot adversarial training)

Art and creativity

- GANs are really good at imitating real data, so they can produce piece of art that resemble real art produced by humans.
- As an example, they can imitate pretty well the style of a painter, or the genre of music of a track.
- This is creativity or imitation?



Creative Adversarial Networks (CANs)

- We want the generator to be creative:

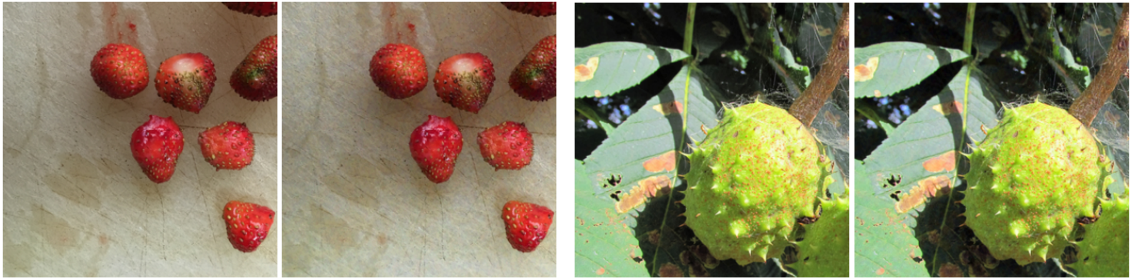
> *"The use of imagination or original ideas to create something."*
>
> *Oxford dictionary*

- The discriminator discriminate between real and generated art.
- To include creativity in the process, the discriminator also classify artworks into a style (cubism, impressionism, . . .).
- If a generated image is classified as art but the discriminator is not able to classify it in a known style, the generated image can be considered a new and creative work.

Creative Adversarial Networks (CANs)

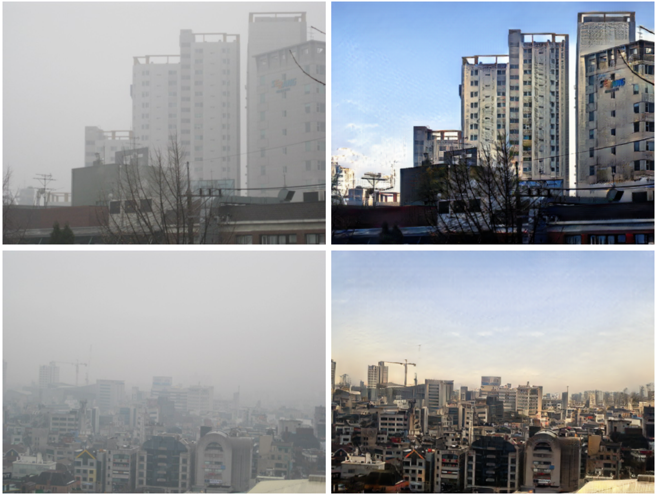
Adversarial Examples

Find and exploit weaknesses of classifiers, in order to produce example that humans are able to confidently classify in the right class, but are extremely confusing for ML models.

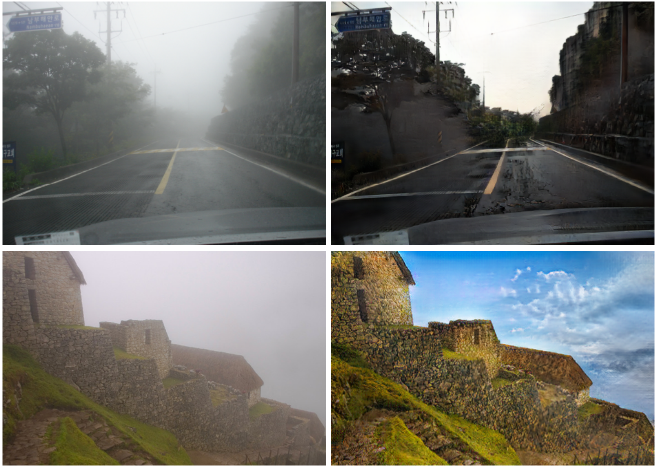


(a) Strawberry (b) Toy poodle (c) Buckeye (d) Toy poodle

Defogging I



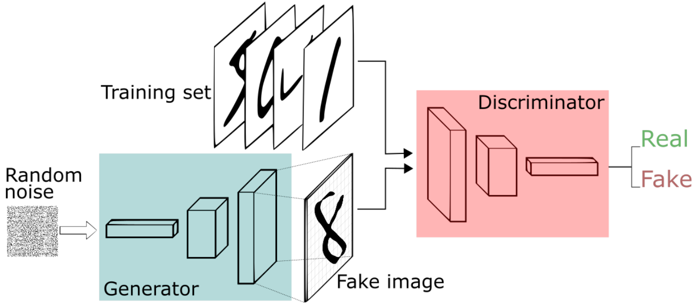
Defogging II



That's all Folks!

Summary

$$\min_{D} \max_{G} (1 - D(x)) + D(G(z))$$



What I cannot create I do not understand.

Richard Feynman

From art to deep fakes: an introduction to Generative Adversarial Networks

Machine Learning course 2019/2020

Gabriele Graffieti

gabriele.graffieti@unibo.it

PhD student in Data Science and Computation @ University of Bologna

December 12, 2019