Machine Art

TIN0255/09P9M80/D82 - Aprendizagem Profunda

2026.2 - Prof. Pedro Moura

Machine Art

- We are now going to see some of the concepts that enable deep learning models to seemingly **create art**.
  - An idea that may be paradoxical at first.
- Alva Nöe, a philosopher from the University of California, said that “Art can help us frame a better picture of our human nature”.
- How can then machines create art?
- Are the creations that emerge from these machines, in fact, art?

Machine Art

- Another interpretation is that these creations are indeed art and that programmers are artists wielding deep learning models as brushes.

- **Generative Adversarial Network** (**GAN**) produced paitings have been snapped up to the tune of $400,000 a pop:
  - https://www.nytimes.com/2018/10/25/arts/design/ai-art-sold-christies.html

- We will see the high-level concepts behind GANs and examples of the novel visual works they can produce

- We will see what is the latent spaces associated with GANs and the word-vector spaces.

A Boozy All-Nighter

- Below Google's offices in Montreal sits a bar called *Les 3 Brasseurs*.
- It was at at this place in 2014, that Ian Goodfellow conceived an algorithm for fabricating **realistic-looking images**.
  - He was a PhD student in Yoshua Bengio's renowned lab.
- Yann LeCun has hailed this as the "most important" recent breakthrough in deep learning.
  - bit.ly/DLbreakthru

A Boozy All-Nighter

- His friends described to him a **generative model** they were working on → a computational model that aims to produce something novel.
  - a quote in the style of Shakespeare
  - a musical melody
  - a work of abstract art

- They were trying to design a model that could generate photorealistic images such as portraits of human faces.

- **Traditional machine learning:**
  - **engineers would catalog and approximate the critical individual features of faces like eyes, noses and mouths;**
  - **but also accurately estimate how these features should be arranged relative to each other.**

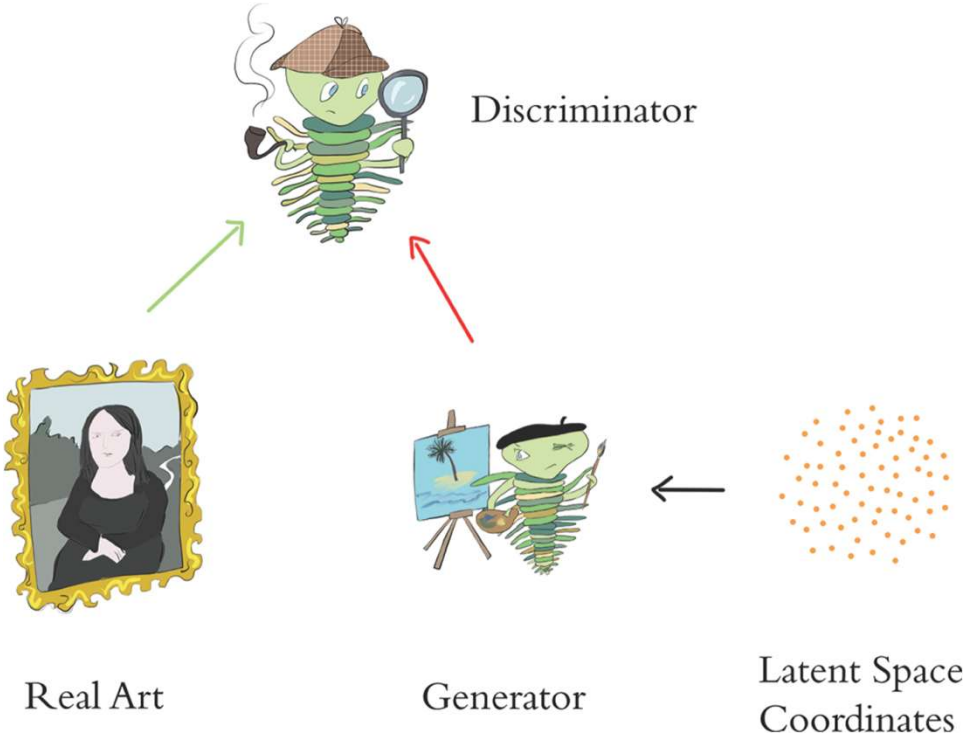
A Boozy All-Nighter

- Their results had been underwhelming.
  - The generated faces tended to be **excessively blurry**, or they tended to be **missing essential elements** like the nose or the ears.

- Goodfellow proposed a revolutionary idea:
  - **A deep learning model in which two ANNs act against each other competitively as adversaries.**

  - **One would be programmed to produce forgeries X the other would be programmed to act as a detective and distinguish the fakes from real images (which would be provided separately).**

A Boozy All-Nighter



High-level schematic of a generative adversarial network (GAN).

A Boozy All-Nighter

- These adversarial deep learning networks would play off one another:
  - As the **generator** became better at producing fakes, the **discriminator** would need to become better at identifying them.
  - So the generator would need to produce even more compelling counterfeits, and so on.

- **This virtuous cycle would eventually lead to convincing novel images of the real images.**
  - Be they of faces or otherwise.

- Best of all, this approach would approach would circumnavigate **the need to program features into the generative model manually.**

A Boozy All-Nighter

- As we have seen, deep learning would sort out the model's features automatically.
- Goodfellow's friends were doubtful his imaginative approach would work.
- He arrived home and found his girlfriend asleep, he worked late to architect his dual-ANN design.
- **It worked the first time, and the astounding deep learning family of Generative Adversarial Networks was born!**

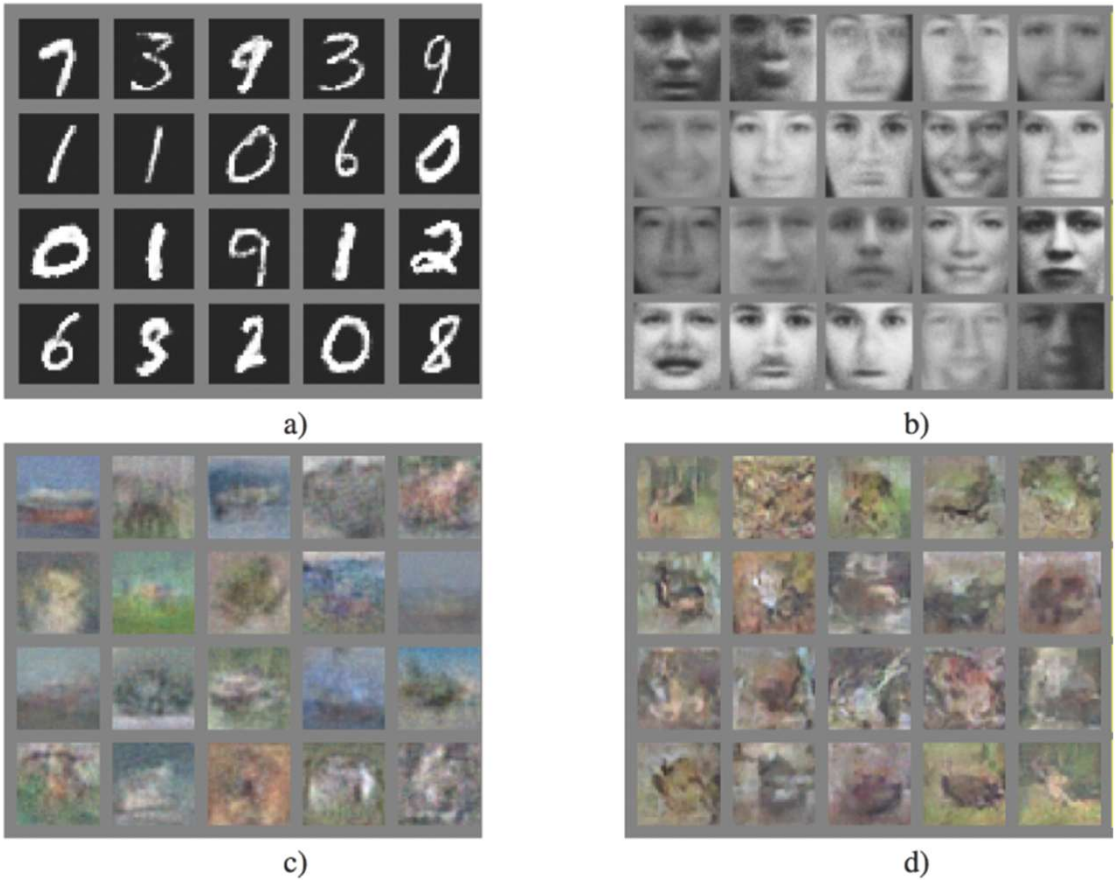
A Boozy All-Nighter

- That same year, Goodfellow and his colleagues revealed GANs to the world at the prestigious Neural Information Processing Systems (NIPS) conference.

- Their GAN produced novel images by being trained on:
  - (a) handwritten digits (from LeCun's MNIST dataset)
  - (b) photos of human faces (from the Hinton re.search group's Toronto Face database)
  - (c) and (d) photos from across ten diverse classes, e.g., planes, cars, dogs (CIFAR-10 dataset)

- The results in (c) are markedly less crisp than in (d).
  - The GAN that generated produced (c) featured **convolutional layers**, whereas the GAN that produced (d) used a more general layer type only.

A Boozy All-Nighter



Results presented in Goodfellow and colleagues' 2014 GAN paper

Arithmetic on Fake Human Faces

- After this, a research team led by the American machine learning engineer Alec Radford determined architectural constraints for GANs that guide considerably more realistic image creation.
- Their **deep convolutional GANs** produced portraits of fake humans.
- Radford and colleagues cleverly demonstrated interpolation through, and arithmetic with, the **latent space** associated with GANs.
- **But what is a latent space?**

Arithmetic on Fake Human Faces

- The latent-space cartoon may be reminiscent of the word-vector space cartoon which we saw for Word Embeddings.

- There are three major similarities between latent spaces and vector spaces:
  - Latent spaces are *n*-dimensional spaces, usually in the order of hundreds of dimensions (for example, we will later architect a GAN with *n* = *100* dimensions).

  - **The closer two points are in the lantent space, the more similar the images that those points represent.**

  - **Movement through the latent space in any particular direction can correspond to a gradual change in a concept being represented (e.g., age or gender).**

Arithmetic on Fake Human Faces

- Let's perform the following steps:
  - Pick two points far away from each other along some *n*-dimensional axis representing **age.**
  - Interpolate between them.
  - Sample points from the interpolated line.

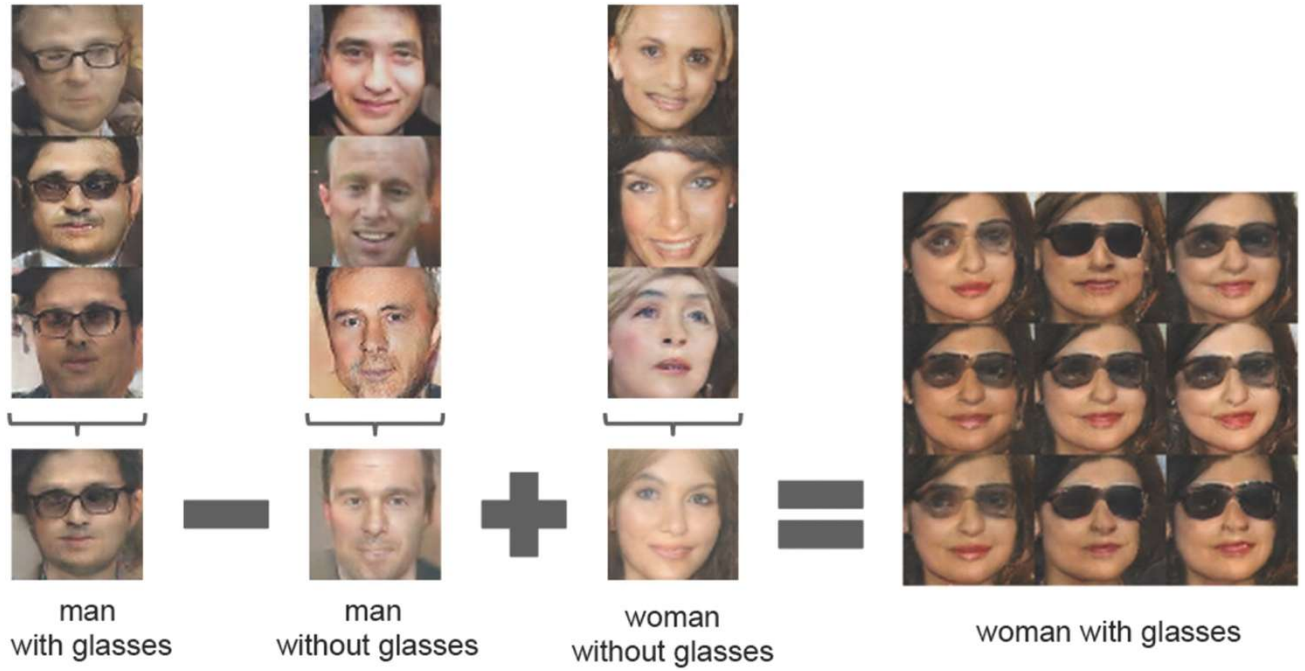
- **We could find what appears to be the same (fabricated) man gradually appearing older and older.**

- In our latent-space cartoon, we represent such an "age" axis in purple.

Arithmetic on Fake Human Faces

- Let's check this video of one of the main papers about GANs.
  - This video, produced by researchers at the graphics-card manufacturer Nvidia, provides a breathtaking interpolation through high-quality portrait "photographs" of celebrities.
  - Let's play with whichfaceisreal.com.
- We could now perform **arithmetic** with images sampled from a GAN's latent space.
- Point within the latent space → **it can be represented by the coordinates of its location**.
  - Resulting vector is analogous to the word vectors we saw previously.
- **We can perform arithmetic with these vectors and move through the latent space in a semantic way.**

Arithmetic on Fake Human Faces



An example of latent-space arithmetic from Radford et al. (2016)

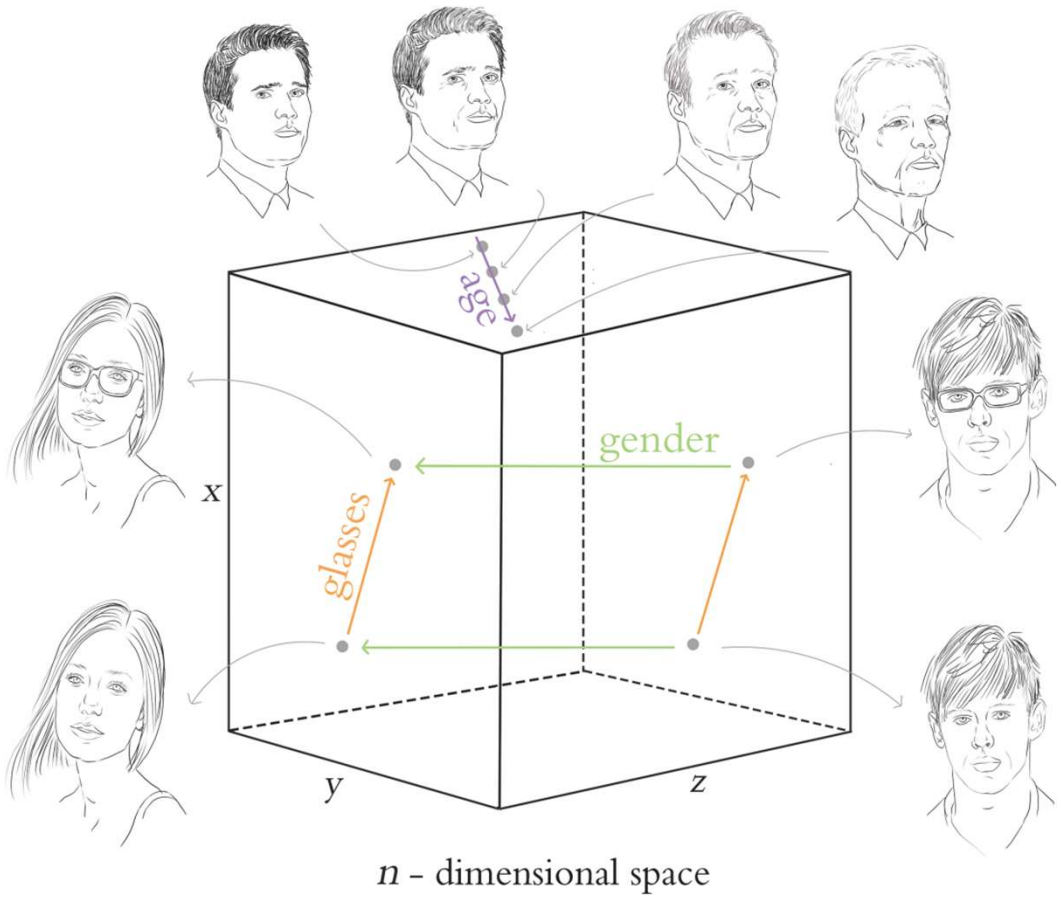
Arithmetic on Fake Human Faces

- This figure showcases an instance of latent-space arithmetic from Radford and his coworkers.

- Starting with a point in their GAN's latent space that represents a man with glasses:
  - **(-)** subtracting a point that represents a man **without** glasses
  - **(+)** adding a point representing a **woman** without glasses
  - **=** the resulting point exists in the latent space near to images represent **women with glasses**.

- Let's see how the **relationships between meaning** in latent space are stored (akin to word-vector space), thereby facilitating arithmetic on points in latent space.

Arithmetic on Fake Human Faces

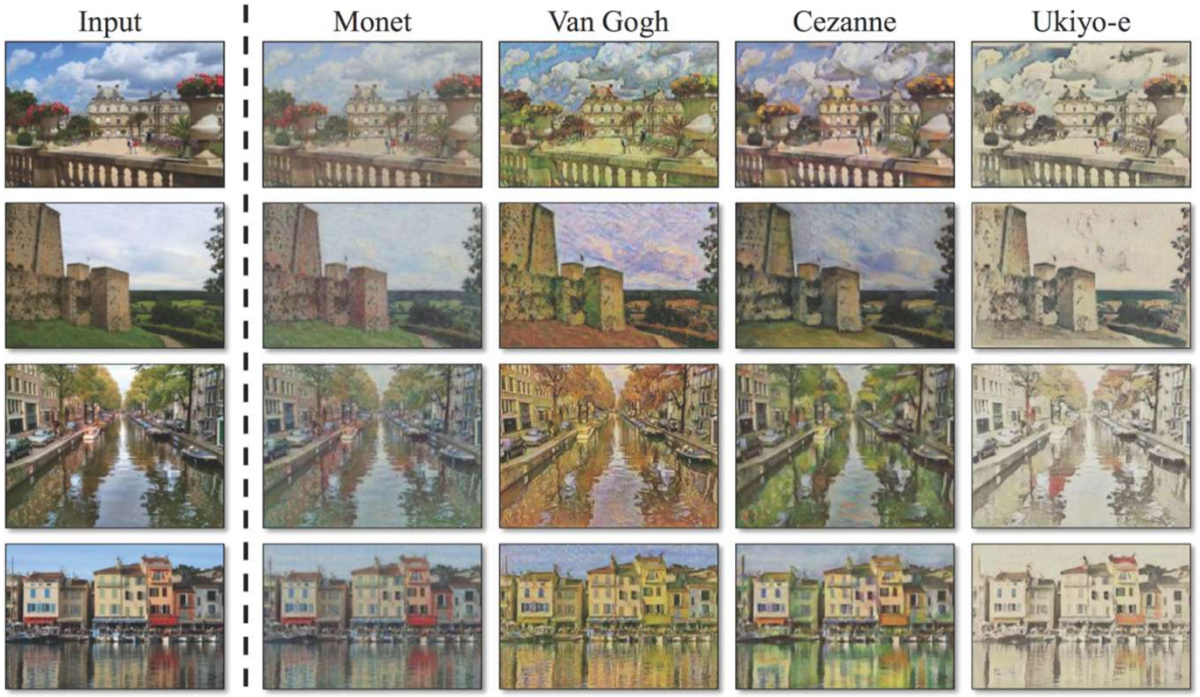


A cartoon of the latent space associated with generative adversarial networks (GANs).

Style Transfer: Converting Photos into Monet (and Vice Versa)

- One of the more magical applications of GANs is **style transfer**.
- Zhu, Park, and their coworkers from the Berkeley Artificial Intelligence Research (BAIR) Lab introduced a new flavor of GAN.
- Alexei Efros took photos while on holiday in France and the researchers employed their **CycleGAN** to transfer these photos into the style of Claude Monet, Vincent Van Gogh, and Ukiyo-e genre, among others.

Style Transfer: Converting Photos into Monet (and Vice Versa)



Photos converted into the styles of well-known painters by CycleGANs

Style Transfer: Converting Photos into Monet (and Vice Versa)

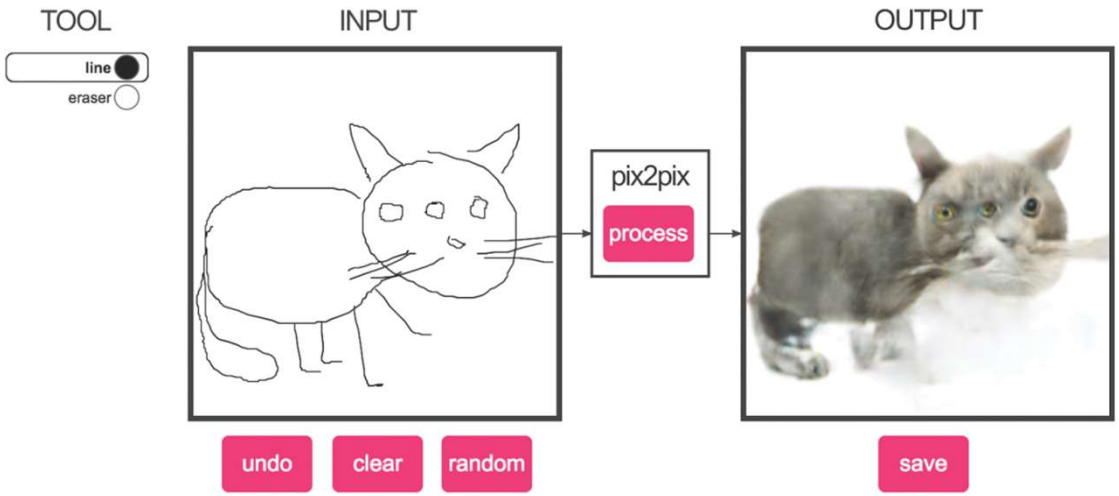
- Let's navigate to bit.ly/cycleGAN.

- There are instances of the inverse (Monet paintings converted into photorealistic images), as well as:
  - Summer scenes converted into wintry ones, and vice versa.
  - Baskets of apples converted into baskets of oranges, and vice versa.
  - Flat, low-quality photos converted into what appear to be ones captured by highend (single-lens reflex) cameras.
  - A video of a horse running in a field converted into a zebra.
  - A video of a drive taken during the day converted into a nighttime one.

Make Your Own Sketches Photorealistic

- Another GAN application out of Alexei Efros's BAIR lab is **pix2pix**: bit.ly/pix2pixDemo.
- It interactively translate images from one type to another.
- The authors of the pix2pix paper call their approach a **conditional GAN** (**cGAN** for short) because the generative adversarial network produces an output that is conditional on the particular input provided to it.

Make Your Own Sketches Photorealistic

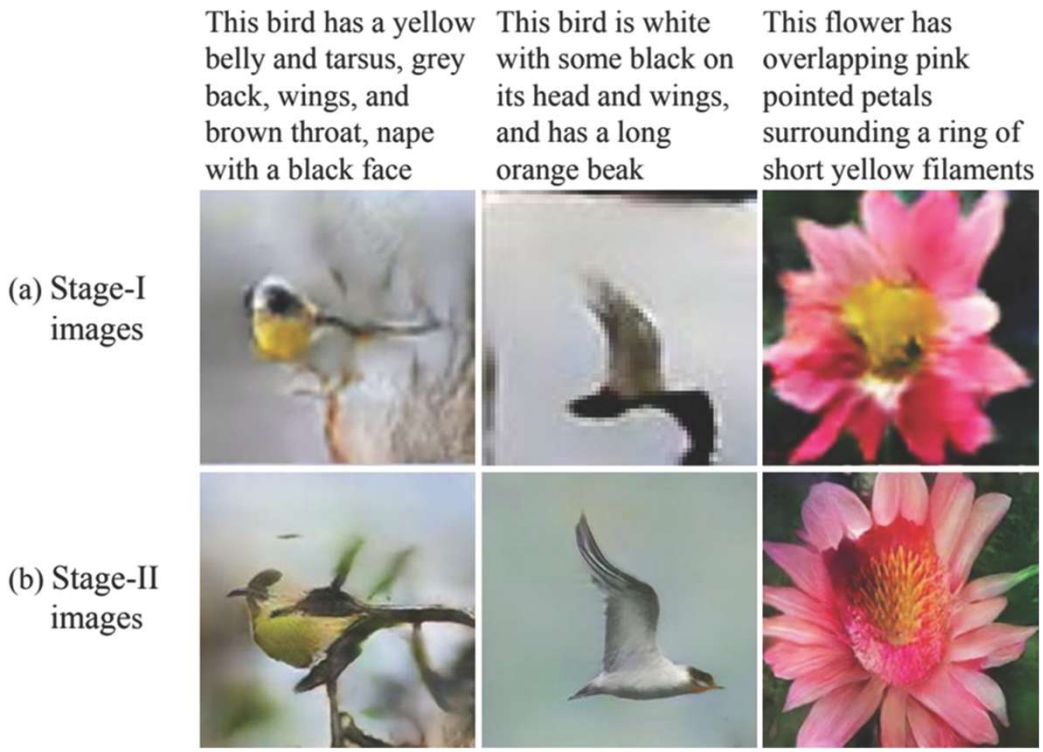


A mutant three-eyed cat (right-hand panel) synthesized via the pix2pix web application.

Creating Photorealistic Images from Text

- Let's take a gander at these truly photorealistic high-resolution images.
- These images were generated by StackGAN, an approach that **stacks two GANs on top of each other**:
  - First GAN → produces a rough, low-resolution image with the general shape and colors of the relevant objects.
  - This is supplied to the second Gan.
  - Second GAN → forged "photo" is refined by fixing up imperfections and adding considerable detail.
- The StackGAN is a **conditional GAN** like the pix2pix network.
- **The image output is conditioned on text input instead of an image.**

Creating Photorealistic Images from Text



Photorealistic high-resolution images output by StackGAN, which involves two GANs stacked upon each other

Image Processing Using Deep Learning

- Since the advent of digital camera technology, **image processing** (both on-device and postprocessing) has become a staple in most photographers' workflows.
  - From simple on-device processing (increasing saturation and sharpness immediately after capture)
  - To complex editing of raw image files in software applications like Adobe Photoshop and Lightroom.

- Machine learning → present in on-device processing, where the camera manufacturer would like the image that the consumer sees to be vibrant and pleasing to the eye with minimal user effort.

Image Processing Using Deep Learning

- Examples:
  - Early face-recognition algorithms (they selectively fire the shutter when they recognize that the subject is smiling).
  - Scene-detection algorithms that adjust the exposure settings to capture the whiteness of snow or activate the flash for nighttime photos.

- Postprocessing arena:
  - A variety of automatic tools exists.
  - Photographers who are taking the time to postprocess images are investing considerable time and domain-specific knowledge into color and exposure correction, denoising, sharpening, tone mapping, and touching up (to name just a few of the corrections that may be applied).

Image Processing Using Deep Learning

- Chen Chen and his collaborators at Intel Labs (2018) -> deep learning was applied to the **enhancement of images** that were taken in near total darkness, with astonishing results.
- Their model involves convolutional layers organized into the **innovative U-Net architecture** (we will see later).
- The authors generated a custom dataset for training this model:
  - the See-in-the-Dark dataset consists of 5,094 raw images of very dark scenes.
  - Short-exposure image + long-exposure image (using a tripod for stability) of the same scene.

Image Processing Using Deep Learning

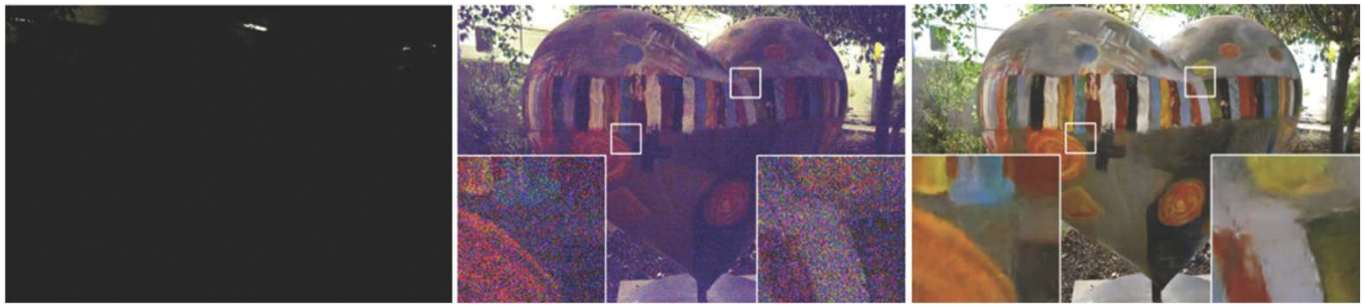
- Exposure times:
  - Long-exposure images were 100 to 300 times those of the short-exposure training images.
  - Actual exposure times → 10 to 30 seconds.

- **The proposal image-processing pipeline of U-Net far outperforms the results of the traditional pipeline.**

Image Processing Using Deep Learning

- Limitations:
  - The model is not fast enough to perform this correction in real time (and certainly not on-device).
  - A dedicated network must be trained for different camera-models and sensors → a more generalized and camera-model-agnostic approach would be favorable.
  - It's better than traditional pipelines, **but there are still some artifacts** present in the enhanced photos that could stand to be **improved**.
  - The dataset is limited to **selected static scenes and needs to be expanded to other subjects** (most notably, humans).

Creating Photorealistic Images from Text



A sample image (left) processed using a traditional pipeline (center) and the deep learning pipeline by Chen et al. (right)

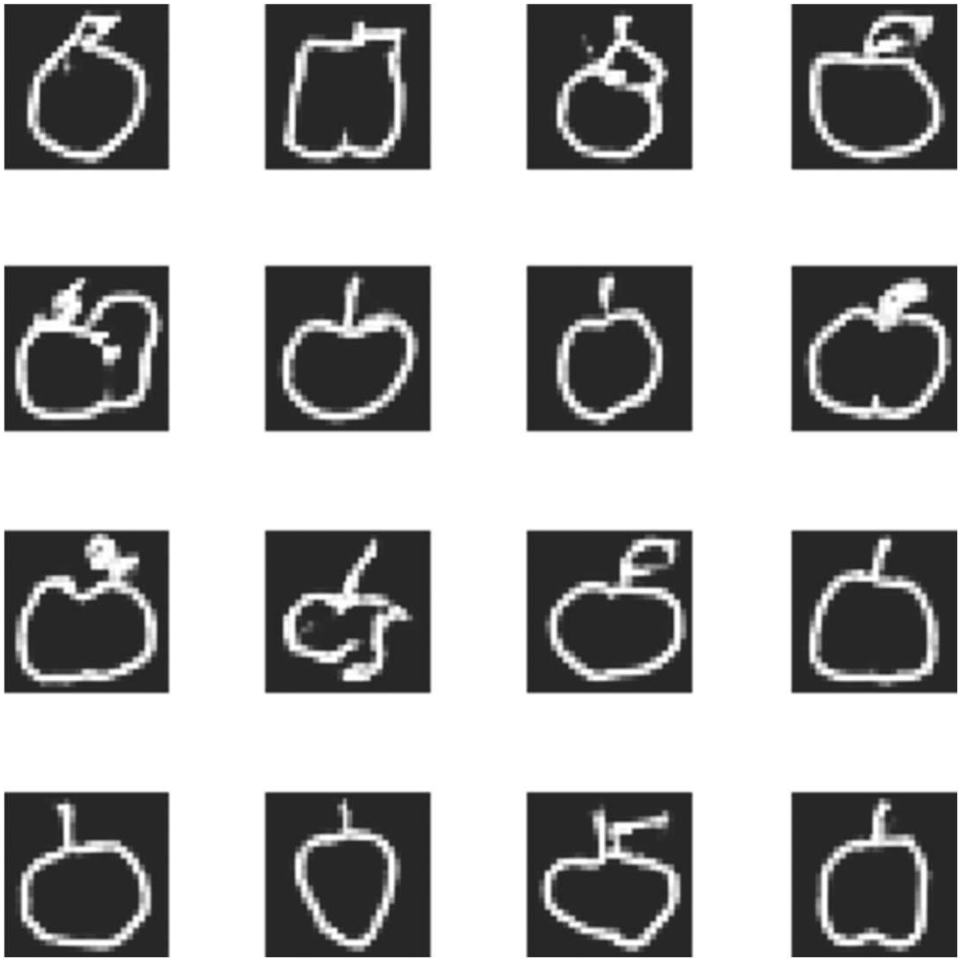
Summary

- We introduced GAN and conveyed that this deep learning approach encodes **exceptionally sophisticated representations within their latent spaces**.

- These rich visual representations enable GANs to create novel images with particular, granular artistic styles.
  - The outputs of GANs aren't purely aesthetic; they can be practical, too.

- They can:
  - Simulate data for training autonomous vehicles.
  - Accelerate the pace of prototyping in fashion and architecture.
  - **Substantially augment the capacities of creative humans.**

Summary



Novel “hand drawings” of apples produced by the GAN architecture we will develop later in the class.

Referências

1. Capítulo 3 do livro *“Deep Learning Illustrated: A Visual, Interactive Guide to Artificial Intelligence”* de Jon Krohn, Grant Beyleveld e Aglaé Bassens.