

Deep Learning and Data exploitation

Angelo Mariano
@angelinux74



centro interdipartimentale di ricerca
di eccellenza sulle tecnologie innovative
per la rivelazione e l'elaborazione
del segnale" (TIRES)



AVVISO DI SEMINARIO

Martedì 16 Maggio 2017, alle ore 15:30
in Aula A



Angelo MARIANO

ENEA, Bari
terrà un seminario dal titolo

Big Data & Machine Learning

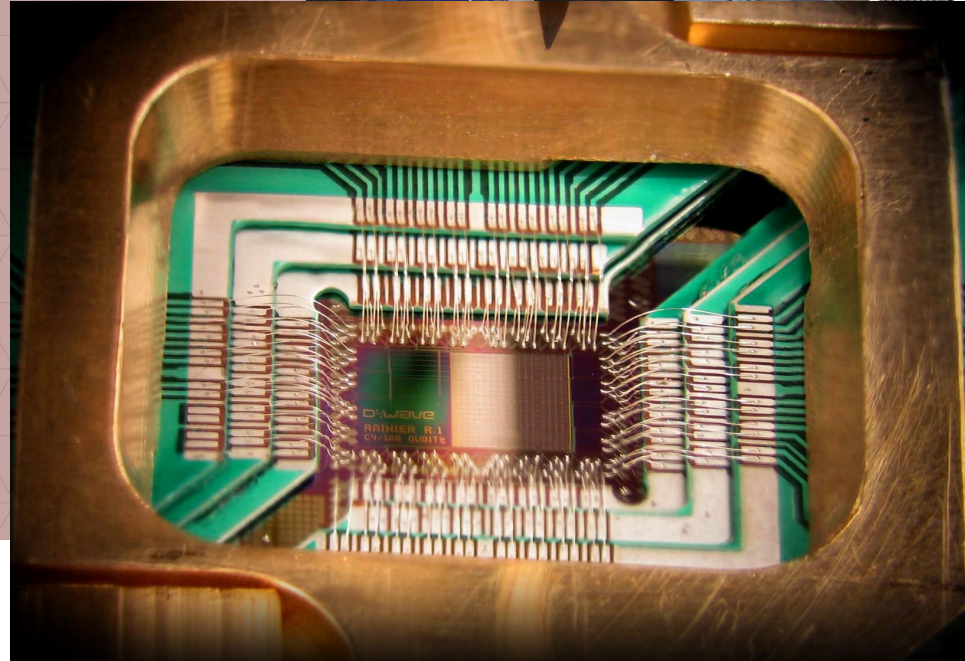
L'esplosione e la rapida diffusione di dati digitali legati a tutte le attività umane pone interessanti questioni in merito alle modalità di



What's going on?

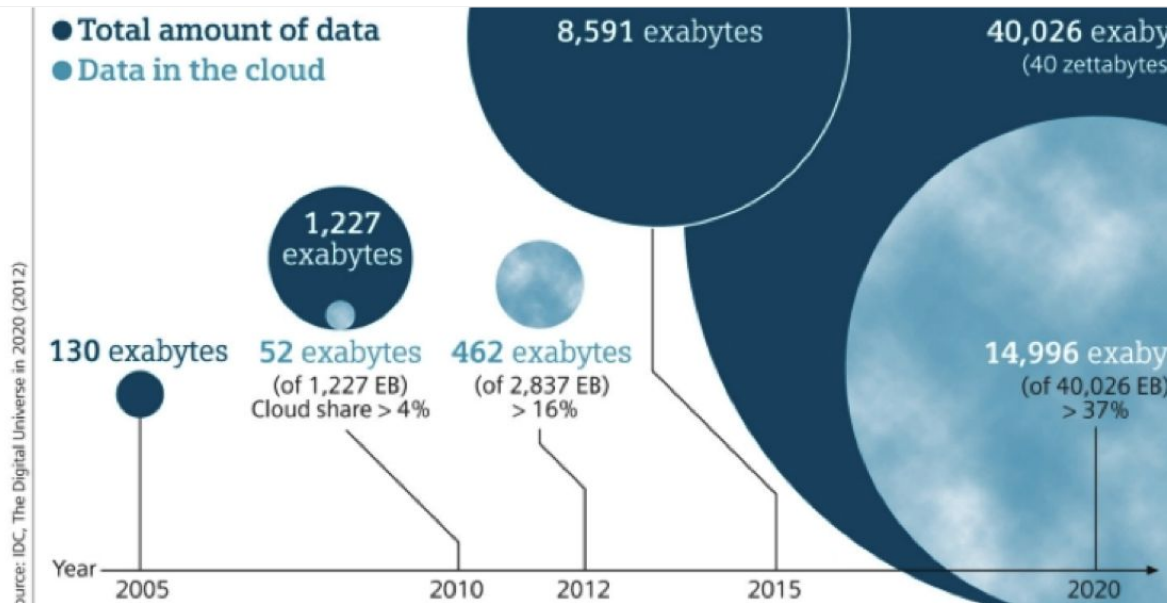


blockchain
shared
single
source of
truth

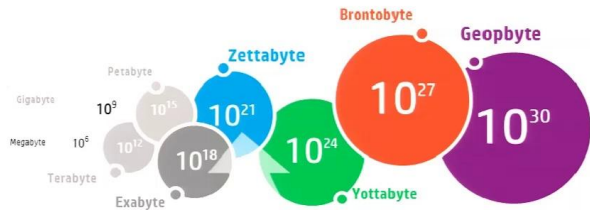


EXPONENTIAL EXPANSION OF THE DIGITAL UNIVERSE

50x Fold Growth from 2010 to 2020



Data explosion pushing limits of today's IT

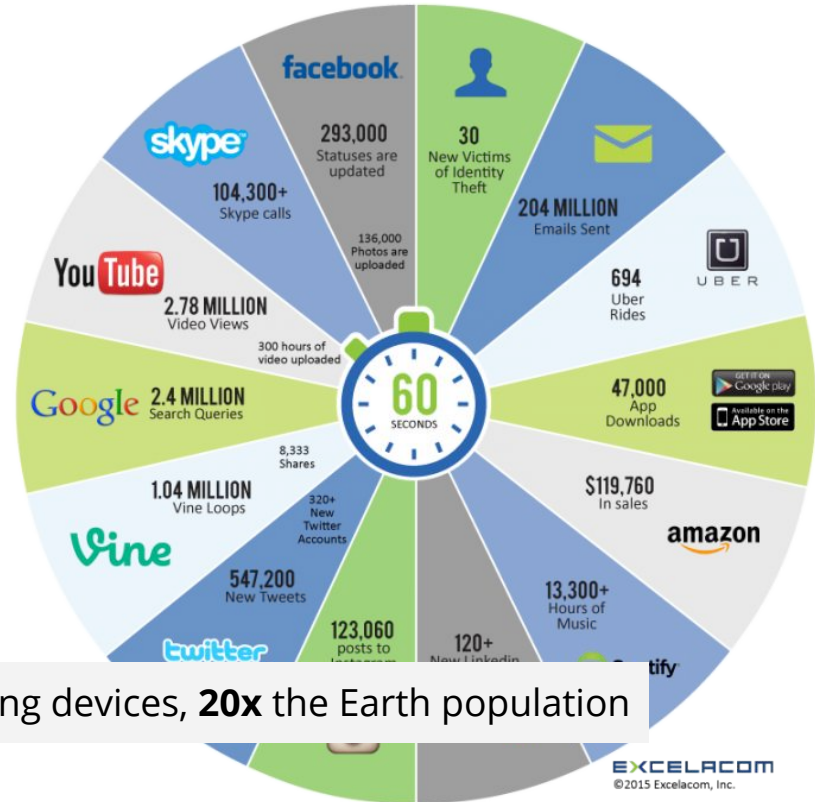


Data explosion

Data sources:

- GPS
- Internet
- Smartphones
- Wearable devices
- Sensors

What happens in an INTERNET MINUTE?



In 2020, there will be **150 billions** of recording devices, **20x** the Earth population

Daniel Yergin, The Prize ('91)

“Today, we are so dependent on oil, and oil is so embedded in our daily doings, that we hardly stop to comprehend its pervasive significance. It is oil that makes possible where we live, how we live, how we commute to work, how we travel – even where we conduct our courtships. It is the lifeblood of suburban communities.”

"DATA IS THE NEW OIL."

From the beginning of recorded time until 2000, we created **5 exabytes** of data.

In 2011 the same amount was created every two days.

By 2013, it's expected that the time will shrink to 10 minutes.

Every hour, we create enough Internet traffic to fill **7 billion** DVDs.

Side by side, that's that's seven times the height of Everest.

Coined in 2006 by Clive Humby, a British data commercialization entrepreneur, this now famous phrase was embraced by the World Economic Forum in a 2011 report, which considered data to be an economic asset, like oil.

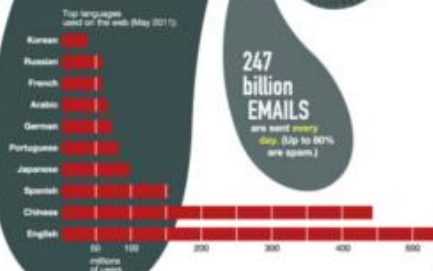
There are nearly as many bits of information in the digital universe as there are stars in our actual universe.

As of August 2012, there were just over **4 million** articles in the English Wikipedia.

There are **133 million** BLOGS on the web.

80% of all humans own a mobile phone of some sort. Out of 2 billion mobiles, 1 billion are smartphones. (In Singapore, 54% of citizens are smartphone users.)

247 billion EMAILS are sent every day. (Up to 90% are spam.)



Just as a study of activity on Twitter gave residents, family members, and journalists advance warning of details about the devastating earthquake and tsunami in Japan, **high-frequency traders**, with the help of computer algorithms, use Big Data to follow trends and to act quickly on their findings.

These specialized algorithms make split-second decisions to buy or sell a commodity. New cable being laid under the Atlantic will shave **5 milliseconds**

from the current 65 milliseconds it takes for trading instructions to travel between New York City and London.

With new fiber-optic cable, the round-trip time between New York and London will be 98.6 milliseconds.

This 5-millisecond saving is worth many millions of dollars to the trading firms who use the cable (and who will pay millions to do so).

How they save 5 milliseconds

The depth of the Atlantic Ocean varies.

The new cable will be on areas of the ocean floor that are up to 1,000 feet shallower than the current twisted cable. By taking a different route, the new cable is shorter, meaning that the time it takes for messages to travel along it is shortened.



60% of all humans (5.4 billion people) are active texters. In 2010, 193,000 text messages were sent every second.

10% of all photos ever taken were taken in 2011.

50% of 5-year-old kids in the U.S. are given access to a smartphone.



1
**DATA
CREATION**
producer



Volume, Velocity, Variety, Value
SYSTEMS INTEGRATION



2
**DATA
ACQUISITION**
*architects/
engineers*

3
**INFO
PROCESSING**
*analysts/
scientist*

OBJECTIVES

DEEP INSIGHTS



REAL-TIME EVENTS



**BIG DATA ECOSYSTEM
FROM DATA TO DECISIONS**

DELIVERY MODELS

4
**BUSINESS
PROCESS**
end users

EMBEDDED



ON-DEMAND



PUSH



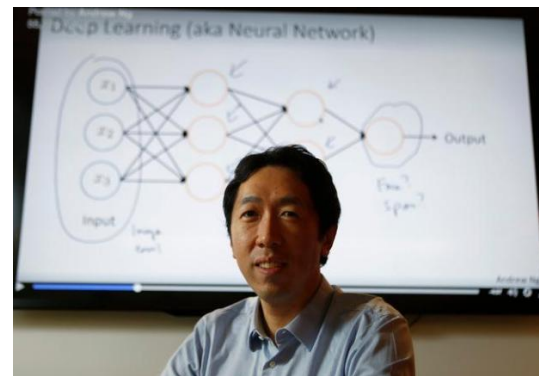
Automation

From dumb to smart objects

- Machines **capable of learning**
- Algorithms can now recognize patterns almost **as well as humans** and even complete some tasks better than them
- Today 70% of all financial transactions are performed by algorithms
- It is easier and cheaper to generate data than ever before, and the tools to turn these **data into insights** are growing exponentially in both quality and quantity.

Andrew Ng (2016), Fortune magazine

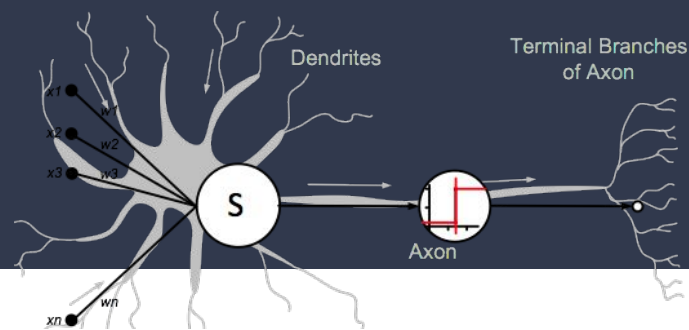
“AI is the new electricity. Just as 100 years ago electricity transformed industry after industry, AI will now do the same.”



Machine learning

- Through the machine learning technique computers are enabled to learn without having been programmed to do so, thanks to a **huge amount of data** that is provided to them as study material
- **As humans learn from experience**, these algorithms learn from the data. A child learns to recognize a cat after having seen five / six specimens, a computer needs thousands and thousands of examples, but it's amazing how many things they can do when they are able to profit from their learning
- Once trained, the machines can make **predictions**. This is the most important skill of the AI revolution: algorithms provide on Amazon books that we would like to read, on Facebook what we want to see in our feed, but may provide, along with a doctor, even a disease in the early stages a patient is developing

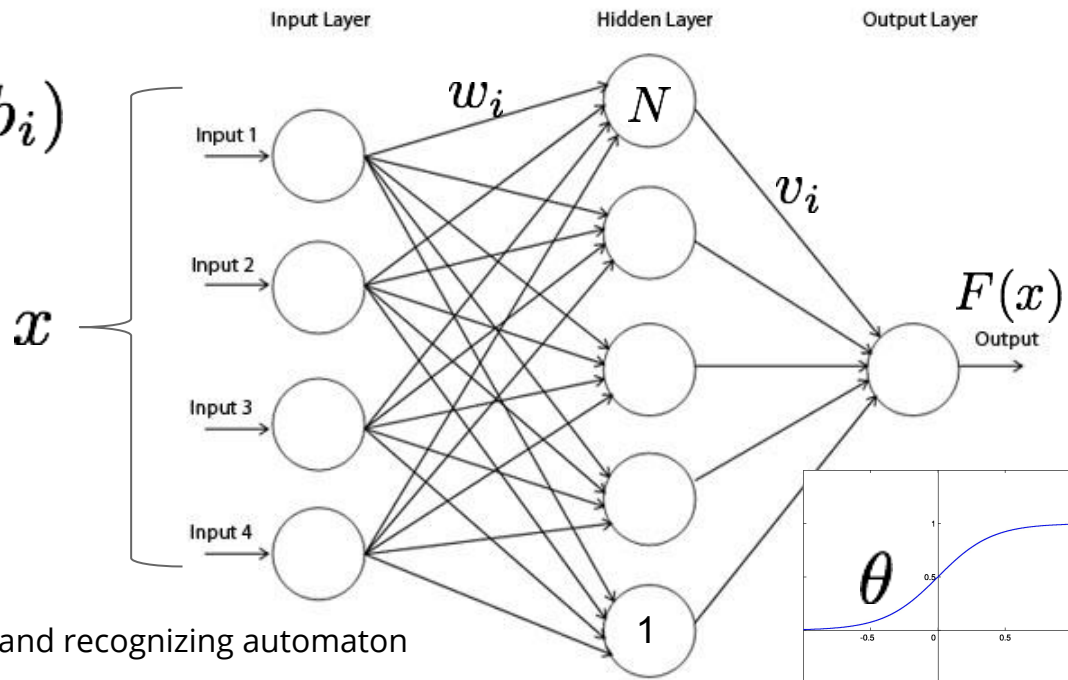
An example: Feed Forward Neural Net



$$F(x) = \sum_{i=1}^N v_i \theta(w_i^T x + b_i)$$

$$x \in \mathbb{D}^{(m)}$$

$$\mathbb{D}^{(m)} \text{compact} \subset \mathbb{R}^m$$



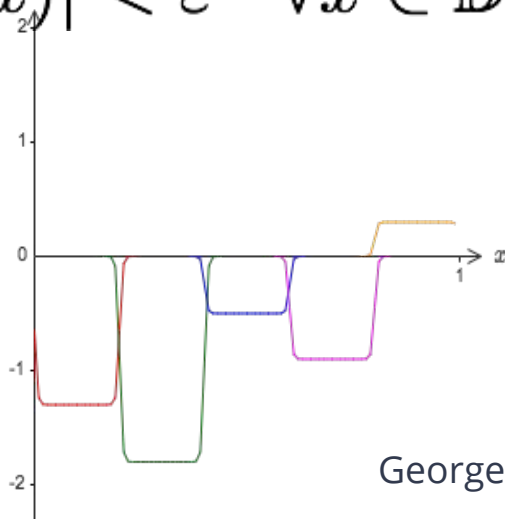
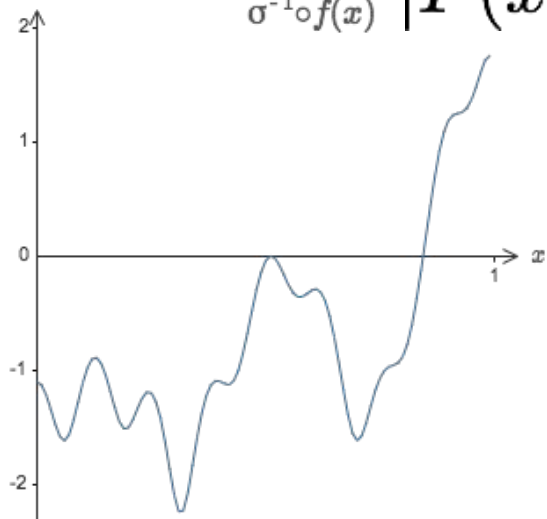
Frank Rosenblatt (1957), The Perceptron - a perceiving and recognizing automaton

Universal Approximation Theorem

$$f \in C(\mathbb{D}^{(m)}), \quad \varepsilon > 0$$

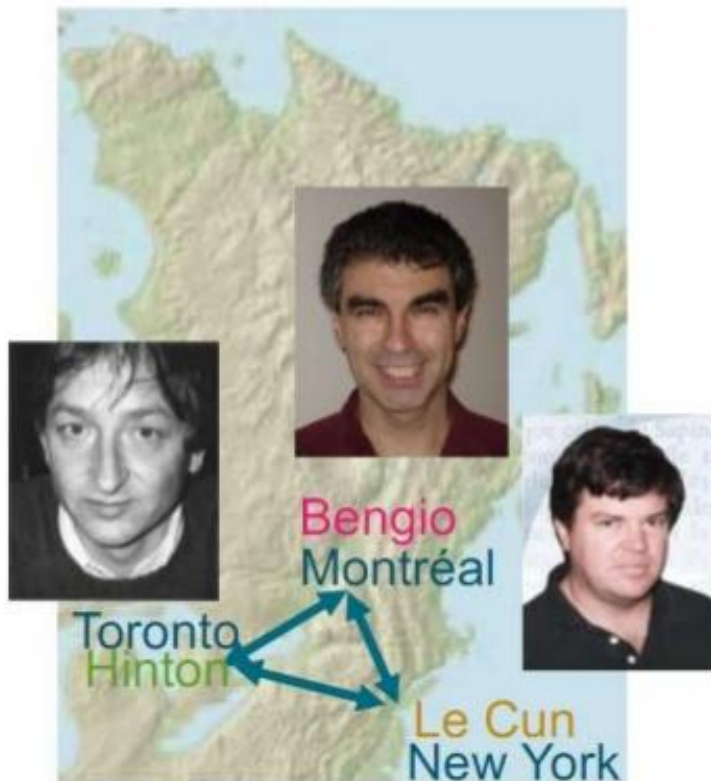
$$F(x) = \sum_{i=1}^N v_i \theta(w_i^T x + b_i)$$

$$\sigma^{-1} \circ f(x) \quad |F(x) - f(x)| < \varepsilon \quad \forall x \in \mathbb{D}^{(m)}$$



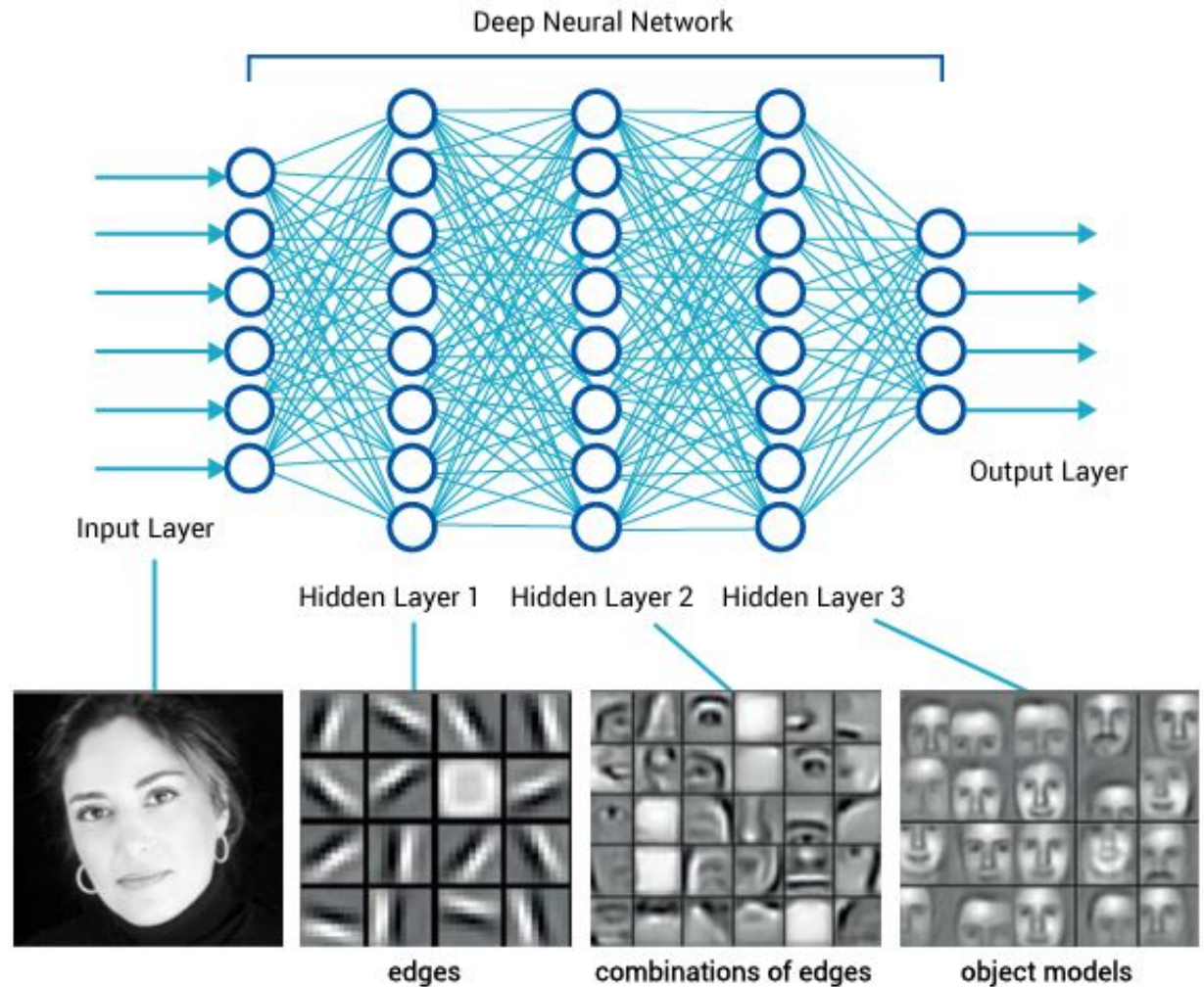
George Cybenko (1989), Kurt Hornik (1991)

2006: The Deep Breakthrough



- Hinton, Osindero & Teh
« A Fast Learning Algorithm for Deep Belief Nets », *Neural Computation*, 2006
- Bengio, Lamblin, Popovici, Larochelle
« Greedy Layer-Wise Training of Deep Networks », *NIPS'2006*
- Ranzato, Poultney, Chopra, LeCun
« Efficient Learning of Sparse Representations with an Energy-Based Model », *NIPS'2006*

Deep Neural Network





Deep Visualization Toolbox

3 big families of ML algorithms:

Supervised

Goal: predict output y for any input x , given a huge set of pairs (x,y)

Unsupervised

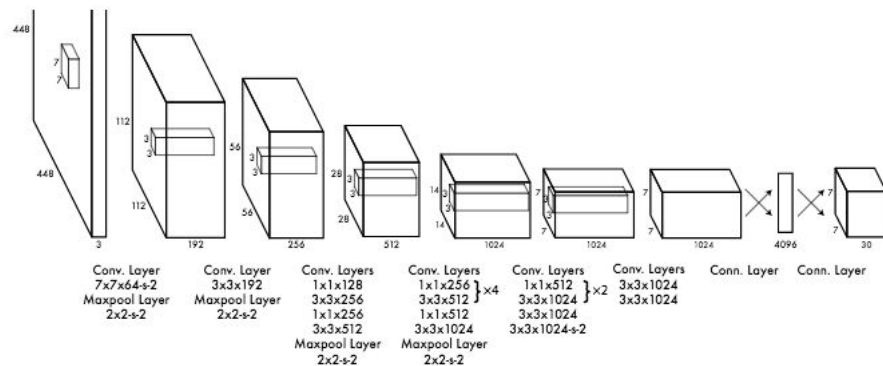
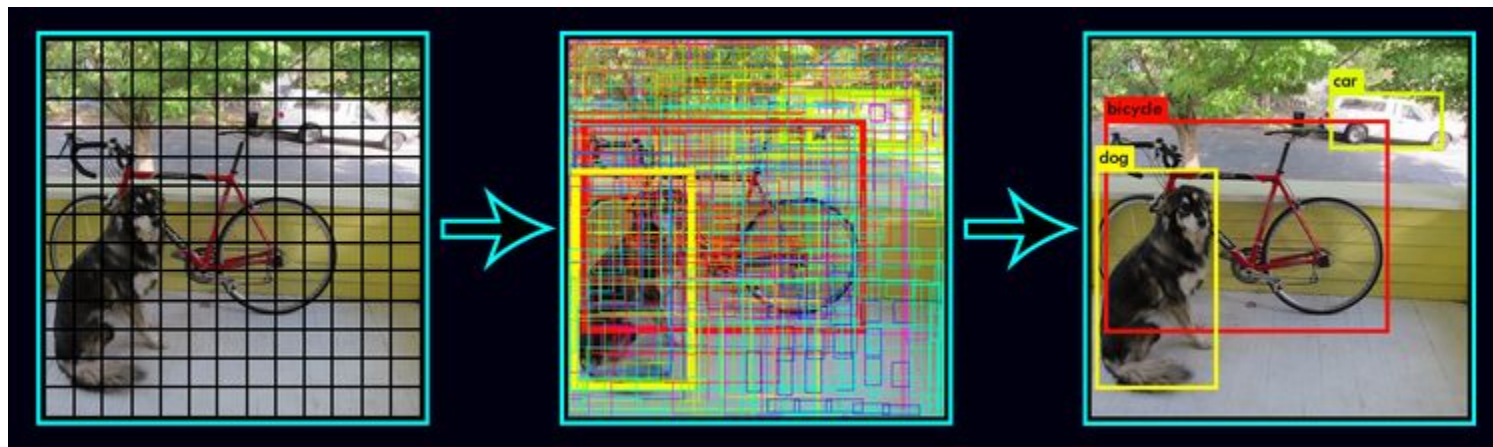
Goal: analysis of unlabeled data, e.g. clustering

Reinforcement

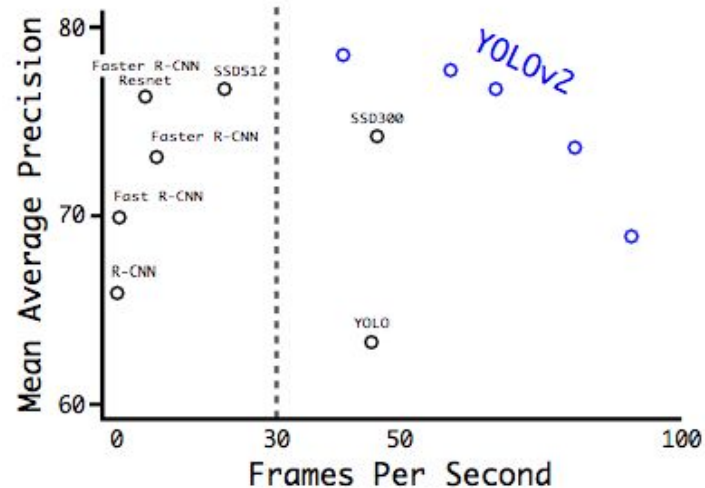
Goal: after a set of state change, obtain a good reward



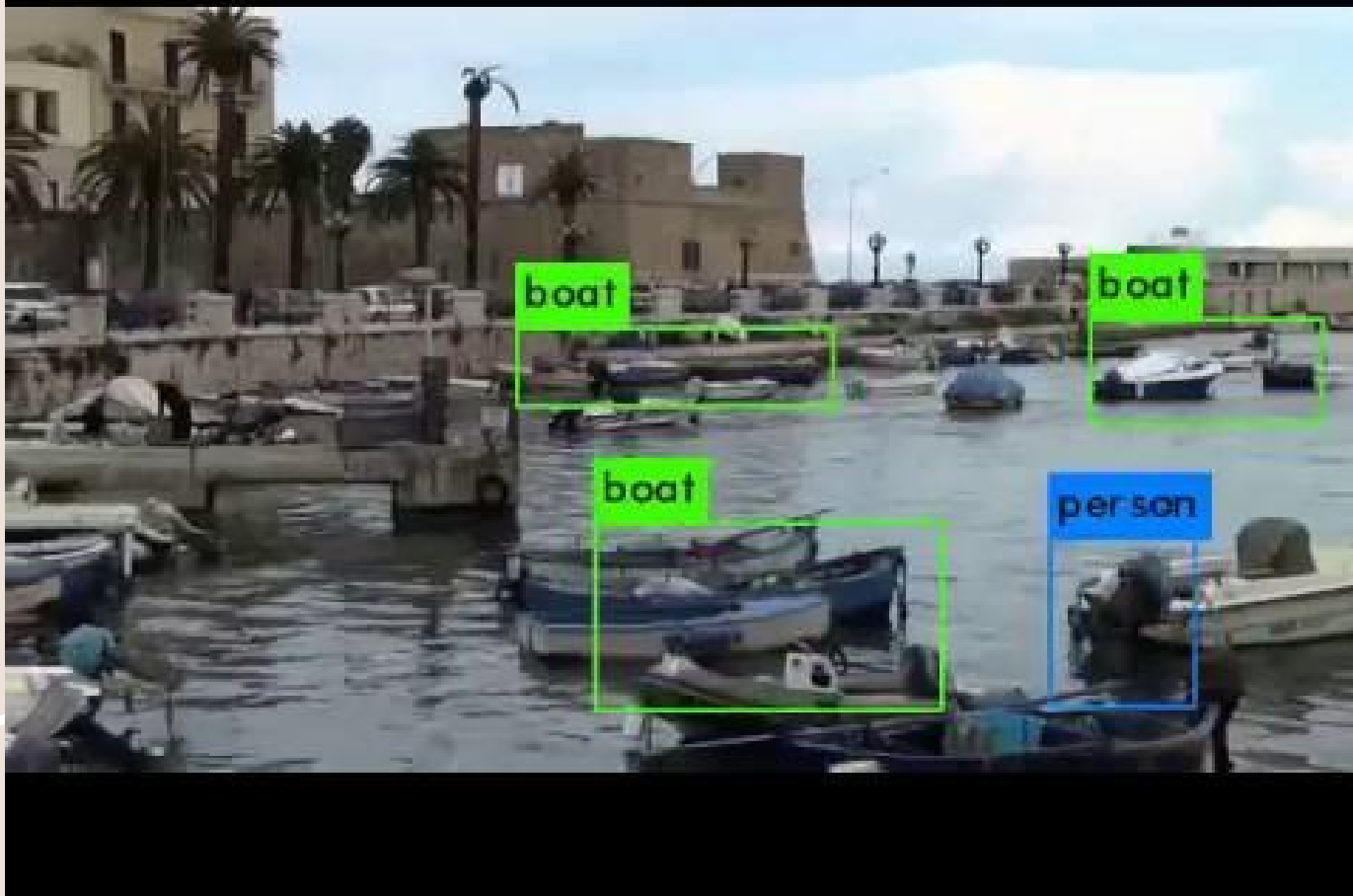
AlphaGo



You Only Look Once



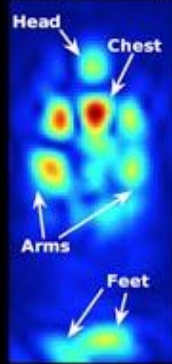
Yolo



Human



Captured Figure



Our device can also monitor
breathing and heart rate

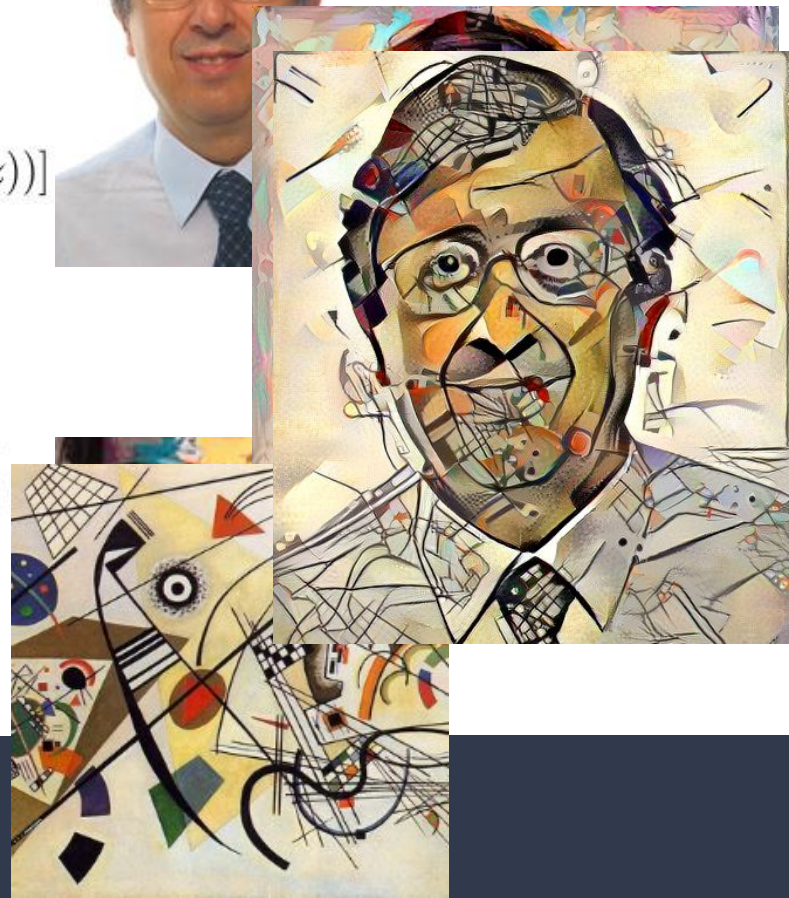
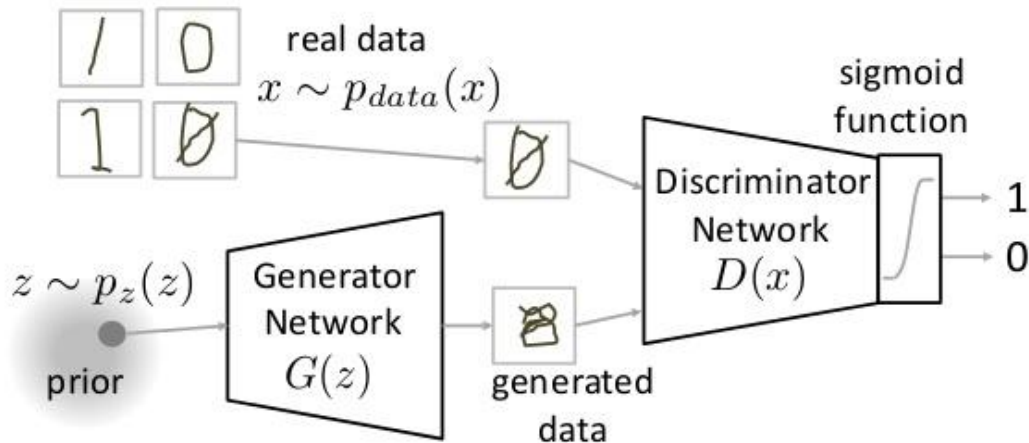


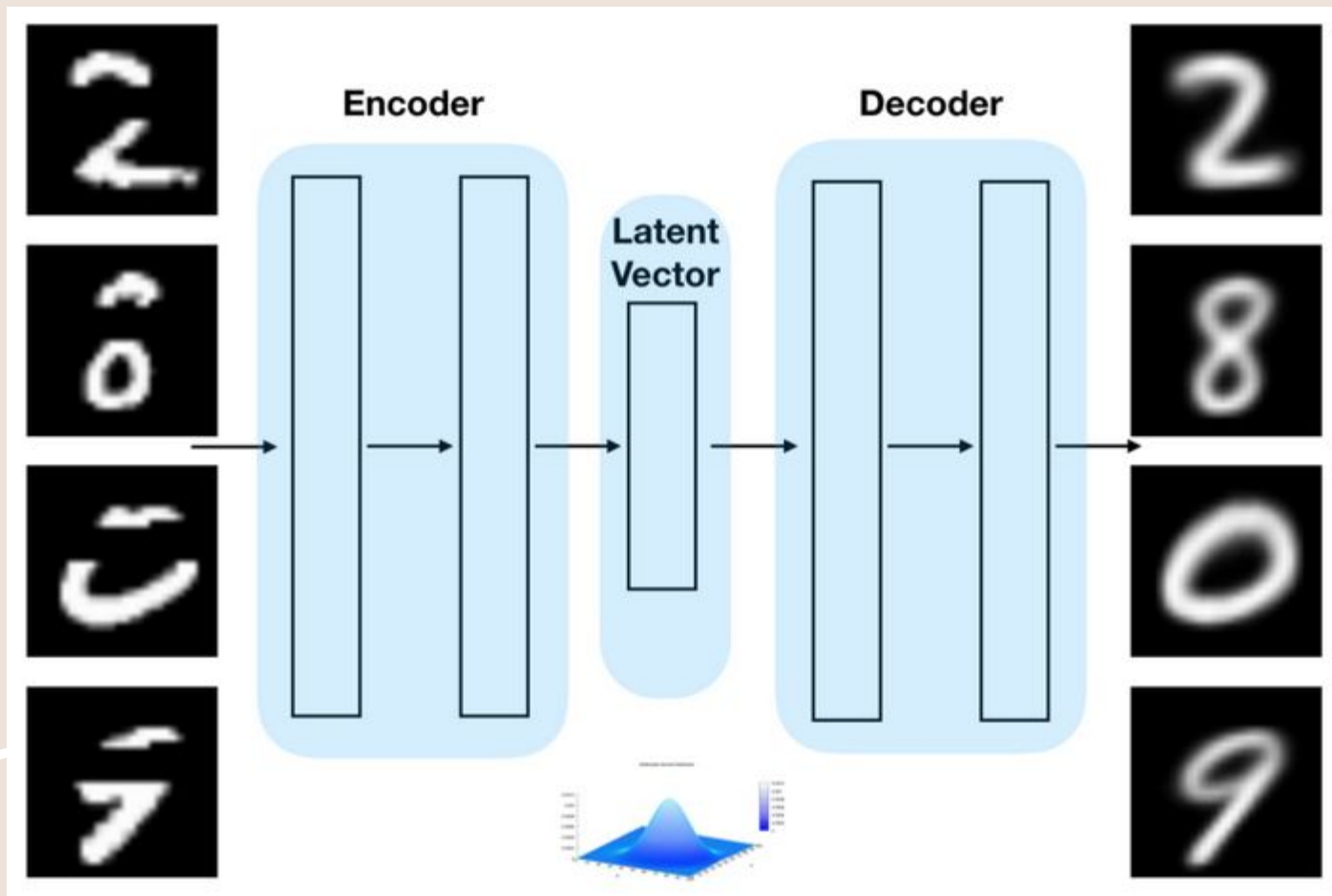
Signal Kinetics

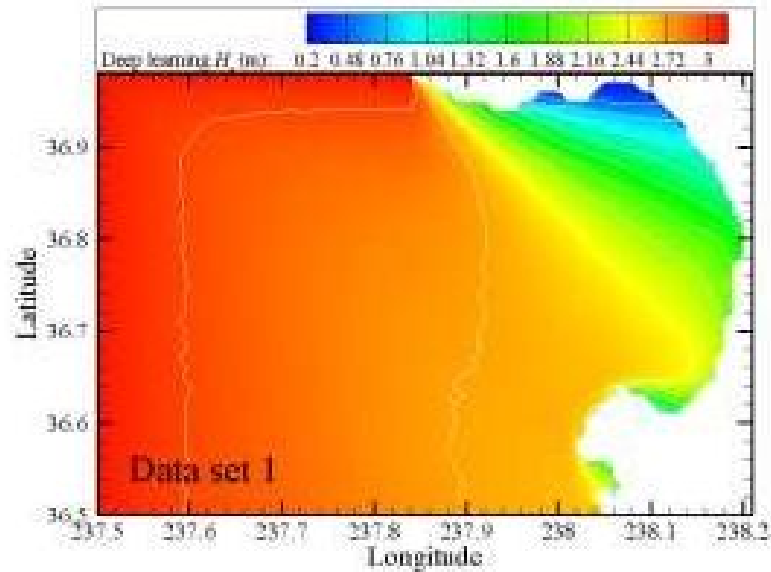
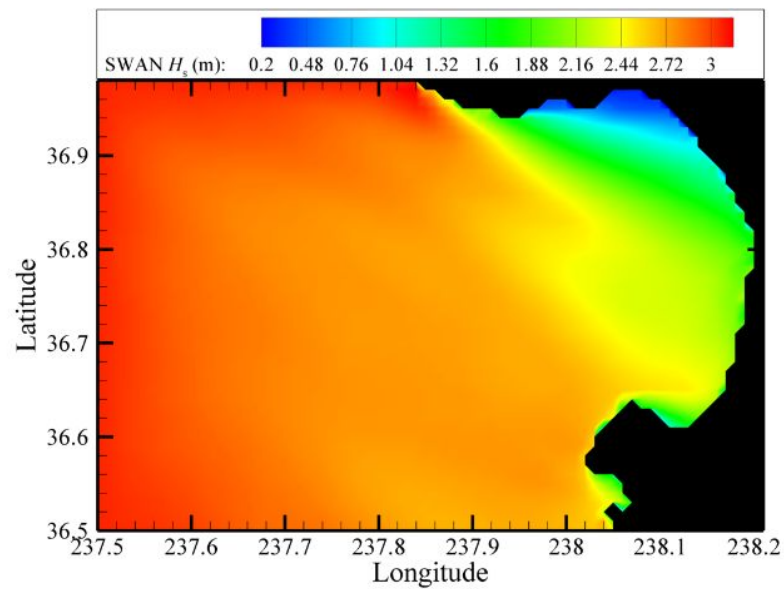
Generative Adversarial Networks

$$\min_G \max_D V(D, G)$$

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







12,000% faster

Halevy, Norvig, Pereira (2009), The Unreasonable Effectiveness of Data

“So, follow the data. Choose a representation that can use unsupervised learning on unlabeled data, which is so much more plentiful than labeled data. Represent all the data with a nonparametric model rather than trying to summarize it with a parametric model, because with very large data sources, the data holds a lot of detail.”

Do you want to bet?

References

- [Data explosion](#)
- [AI is going to change everything](#)
- [Internet minute](#)
- [Daniel Yergin, The prize](#)
- [Data is the new oil](#)
- [Big Data & Analytics](#)
- [Big Data ecosystem](#)
- [Automation and democracy](#)
- [Andrew Ng, Fortune](#)
- [Introduction to ANN](#)
- [Universal approximation theorem](#)
and [its illustration](#)
- [Deep neural networks](#)
- [Deep learning tutorial](#)
- [Multi GPU deep learning from Nvidia](#)
- [Deep Visualization Toolbox](#)
- [How Google's AlphaGo beat a Go world champion](#)
- [YOLOv2](#)
- [Signal Kinetics at MIT](#)
- [Generative Adversarial Networks](#)
- [Deep Learning Art](#)
- [Digits and sounds with Artificial Neural Nets](#)
- [Deep learning and ocean waves](#)
- [The Unreasonable Effectiveness of Data](#)



Let's keep
in touch

 @angelinux74
 angelo.mariano@gmail.com
 angelo-mariano-6348061