Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

Sixth International School on Open Science Cloud – SOSC 2024
Bologna, December 4th, 2024

A minimal introduction to **Computer Vision**

Lucio Anderlini

Istituto Nazionale di Fisica Nucleare
SEZIONE DI FIRENZE

ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC

What is Computer Vision

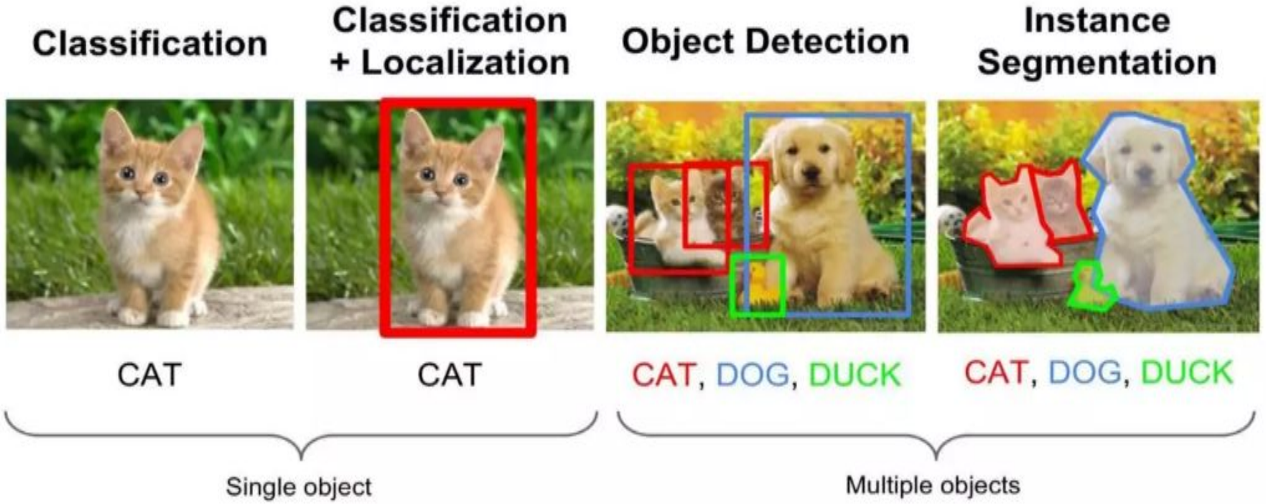
- **Analysing** visual images and videos emulating human ***"intelligence"***

- Getting computers to **handle** and **interpret** visual data, usually based on **statistics**

Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani | ICSC

Tasks specific to Computer Vision

The *Classification* and *Regression* tasks discussed in general for Machine Learning application, are extended to more specific tasks.

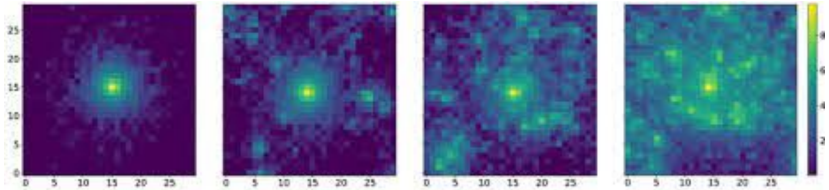


Tasks can become arbitrarily complicated, e.g. “emotion detection” in a crowd.

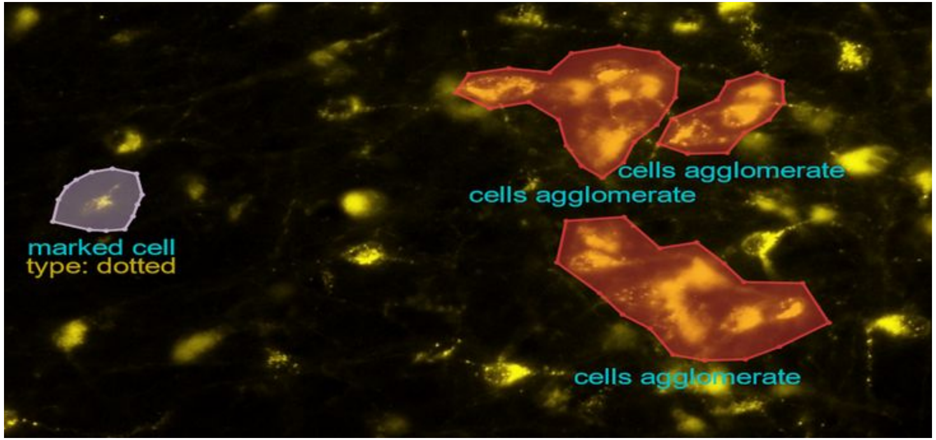
Applications of Computer Vision

Computer Vision is being applied in a huge variety of fields, including research in Physics and applications.

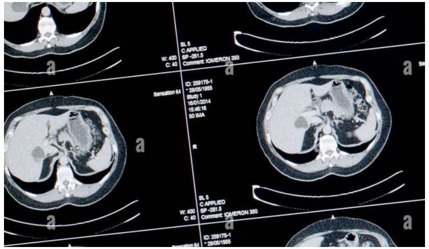
Calorimetric clusters at the LHC



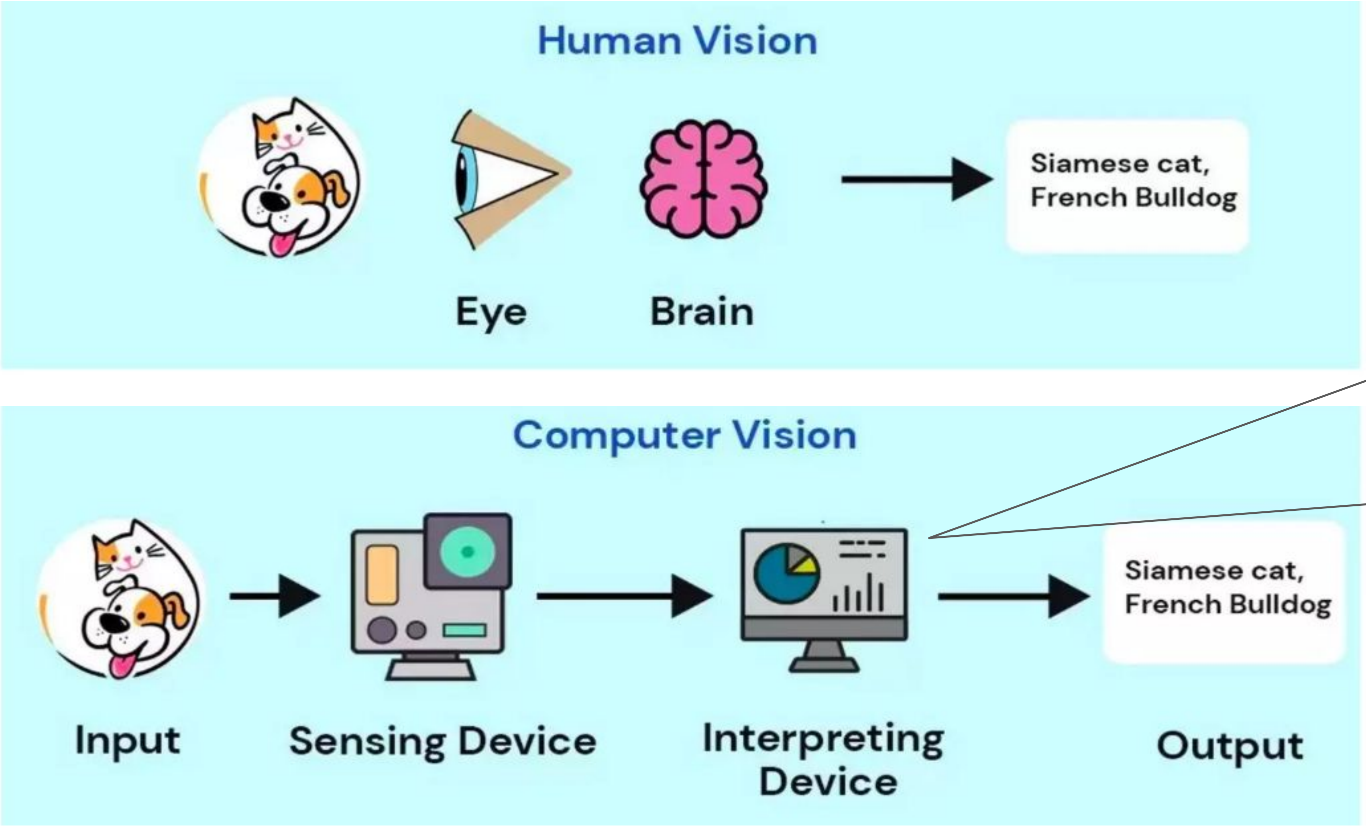
Microscopy



Medical imaging



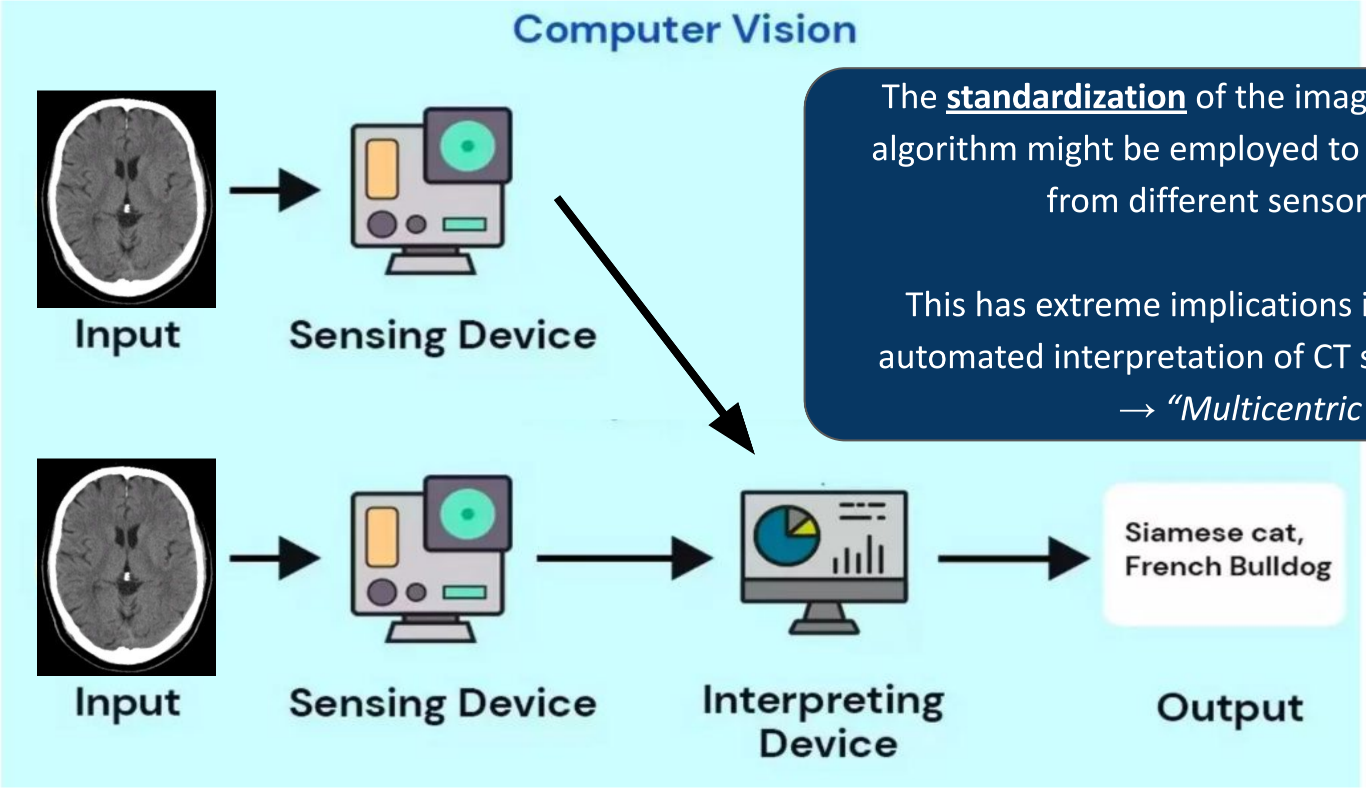
Comparing Human and Computer Vision



Interpretation is usually split in multiple steps, for example:

- Standardization of the format
- Preprocessing of the pixel values
- Feature detection and extraction
- Statistical interpretation

Comparing Human and Computer Vision

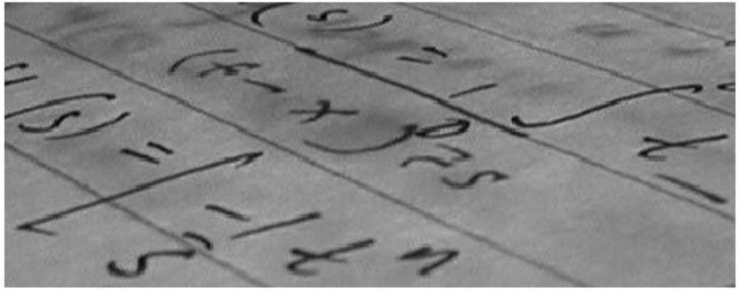


The **standardization** of the images is crucial as the same algorithm might be employed to process images acquired from different sensors and setups.

This has extreme implications in some fields, think to automated interpretation of CT scans in *Medical Physics.*
→ *"Multicentric studies"*

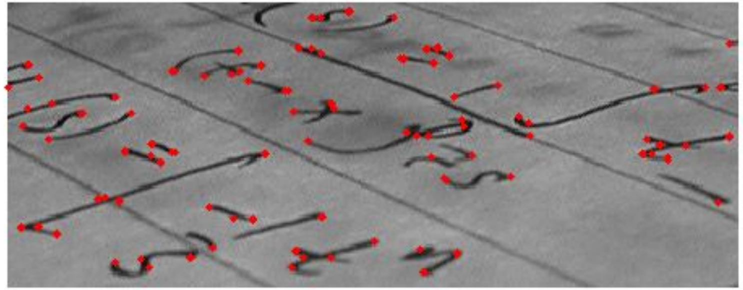
Feature engineering

Before attempting a statistical treatment of the image, it is usually necessary to define numerical features.



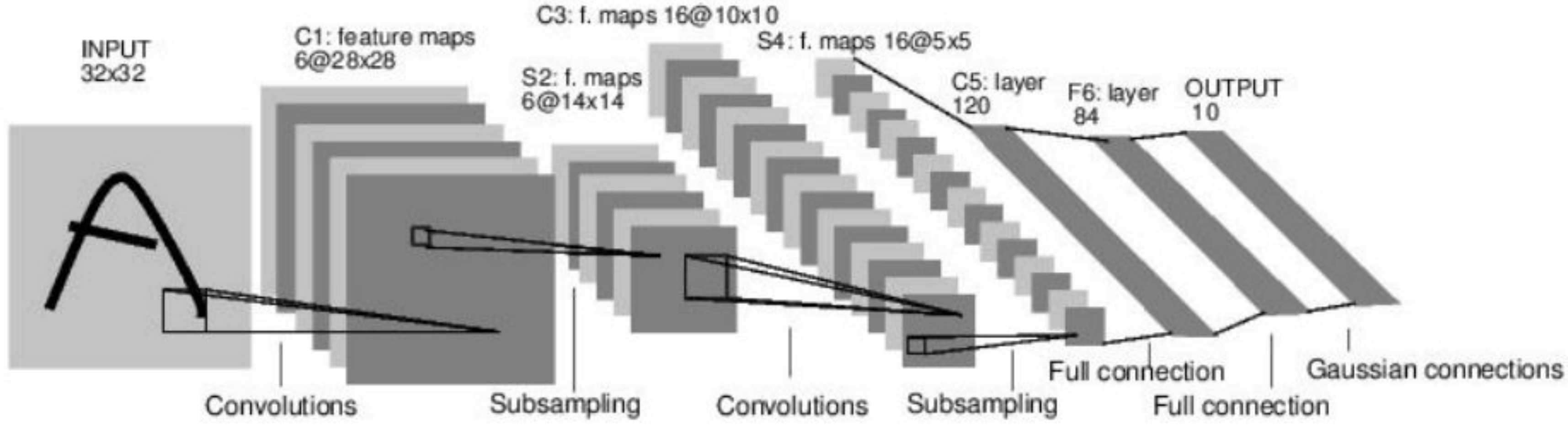
Numerical features can be point, positions, edges, …

The literature on how to detect and extract relevant feature from an image is infinite, and often strictly related to the specific domain of application.



Ok but… can we define "optimal features" as part of the statistical treatment?

Convolutional Neural Networks



[LeNet-5, LeCun 1980]

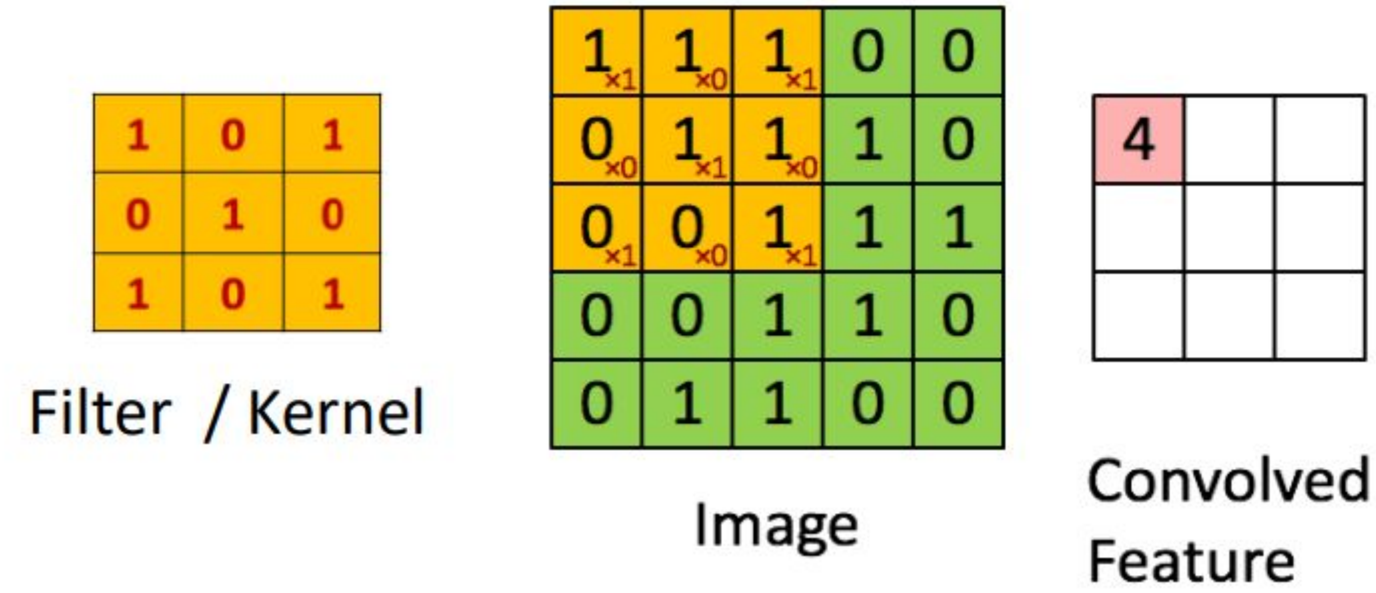
… “optimal features” as part of the statistical treatment?

Of course yes!!! *

* provided you can supply sufficient computing power

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Convolution as application of a filter



Examples of 3x3 filters

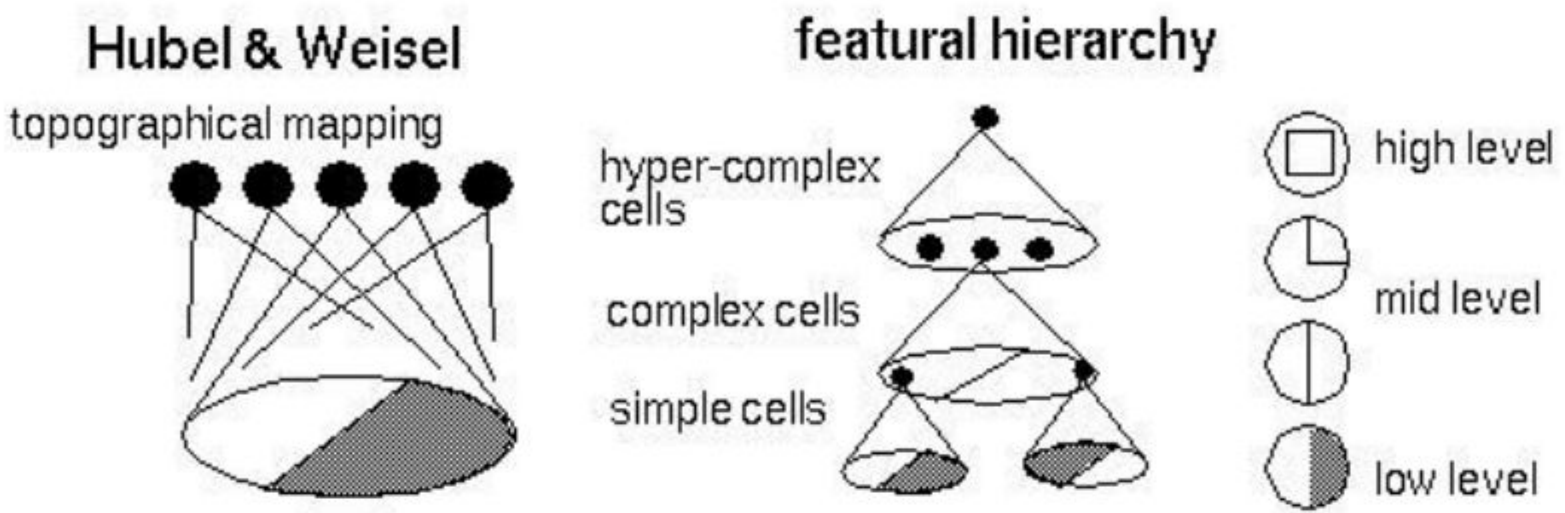
Original

Sharpen

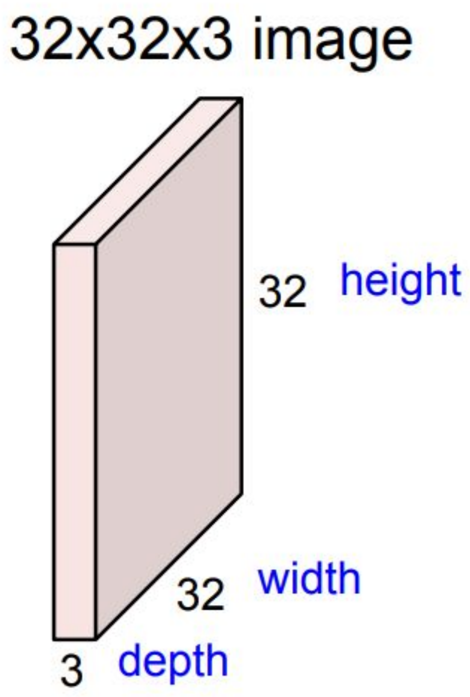
Edge detect

Strong edge detect

Hierarchical organization



Convolutional layer: *the input image*



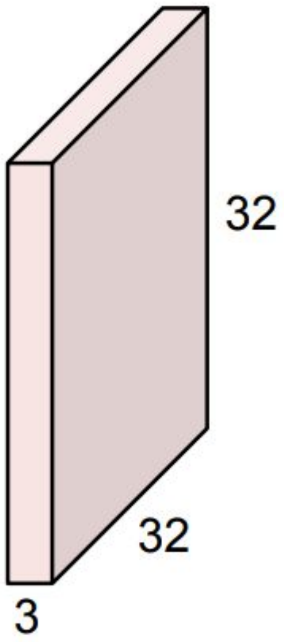
Consider a 32x32 image with three channels:

- red
- green
- blue

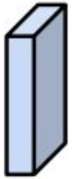
32 x 32 x 3 pixel values

Convolutional layer

32x32x3 image



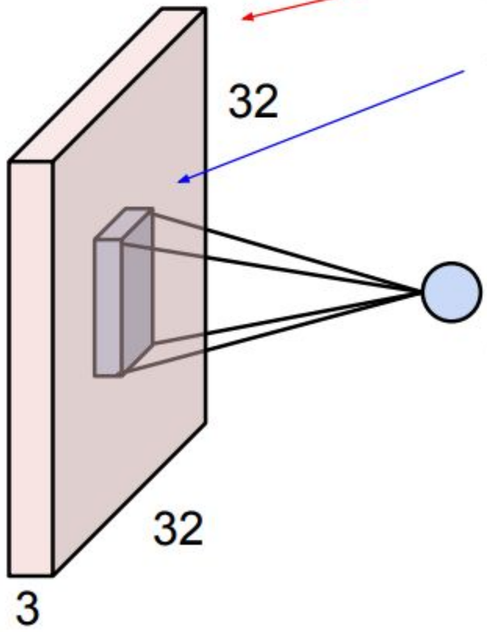
5x5x3 filter



Convolve the filter with the image
i.e. “slide over the image spatially,
computing dot products”

Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani | ICSC

Convolutional layer



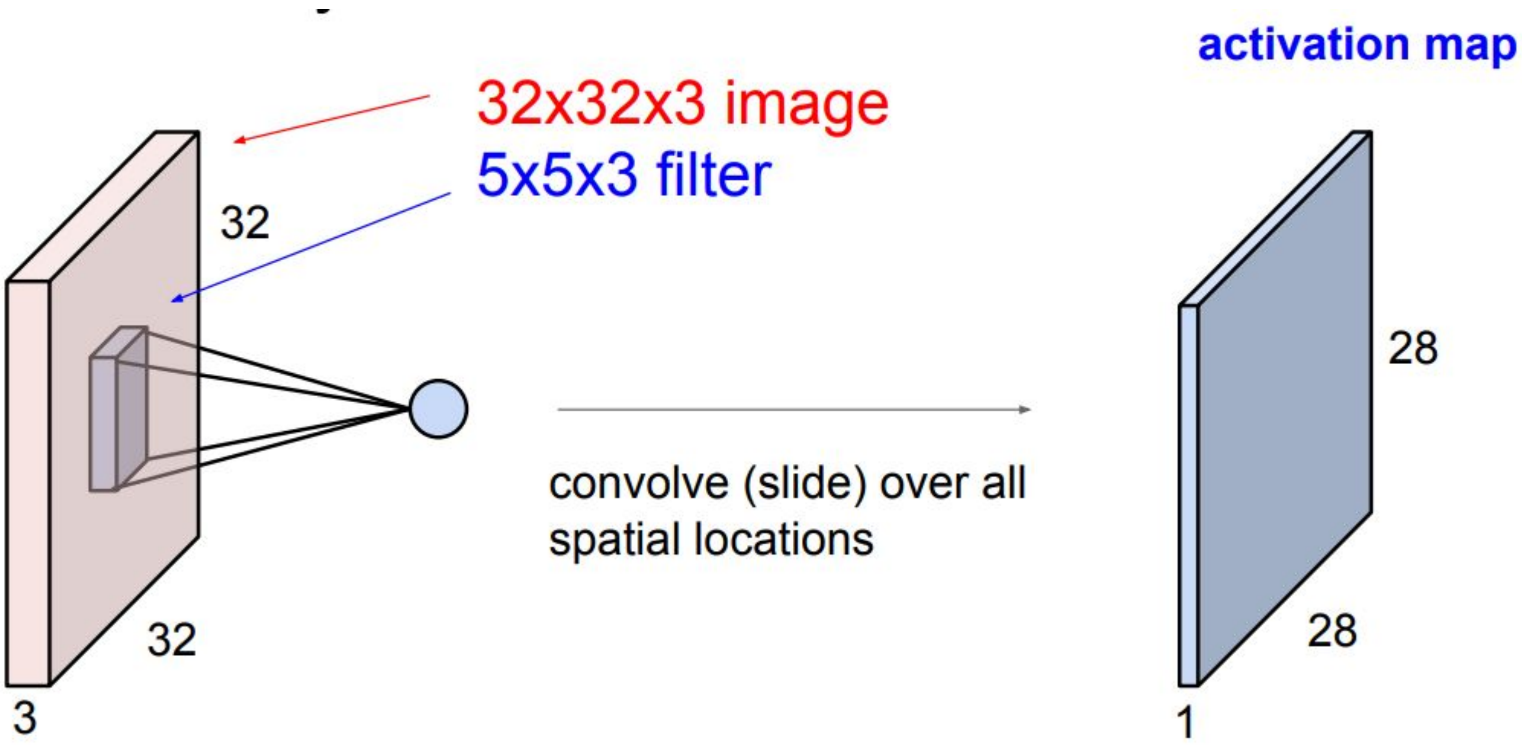
32x32x3 image

5x5x3 filter w

1 number:
the result of taking a dot product between the filter and a small 5x5x3 chunk of the image (i.e. 5*5*3 = 75-dimensional dot product + bias)

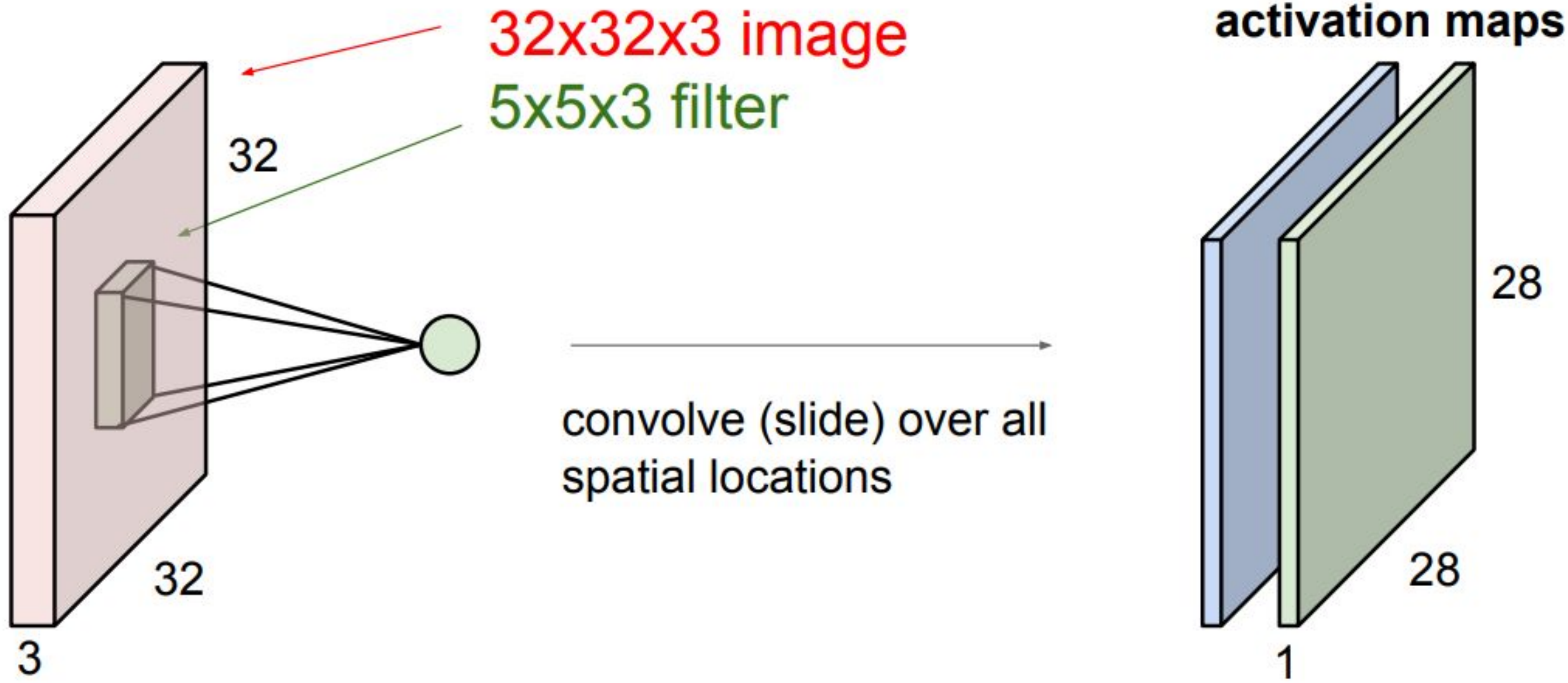
$$w^T x + b$$

Convolutional layer



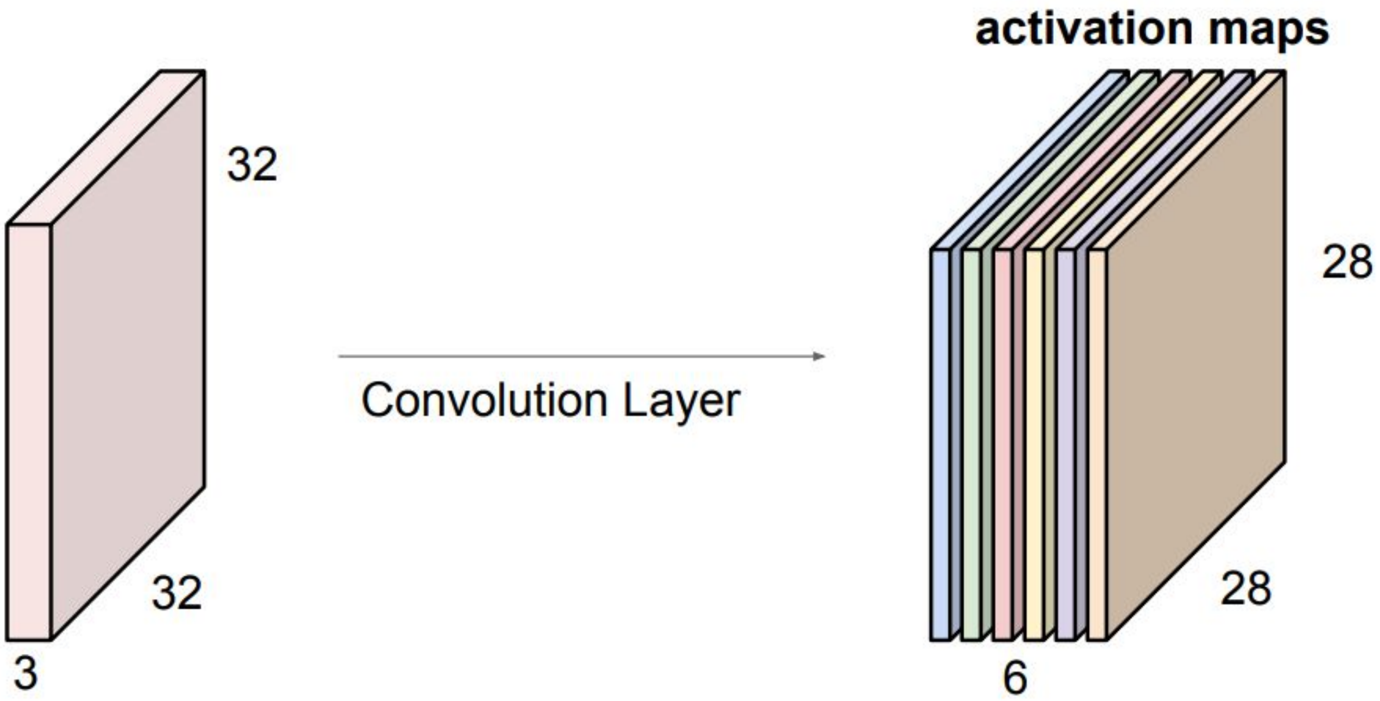
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
ICSC

Convolutional layer: *a second filter*



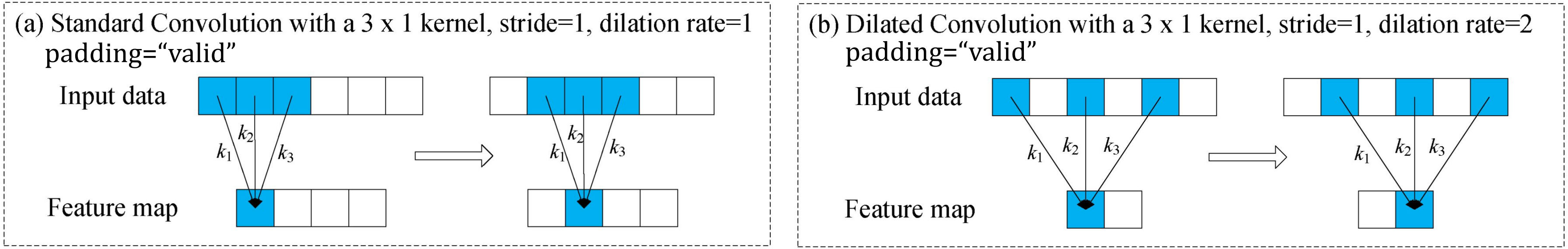
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC

Convolutional layer: *with 6 filters*

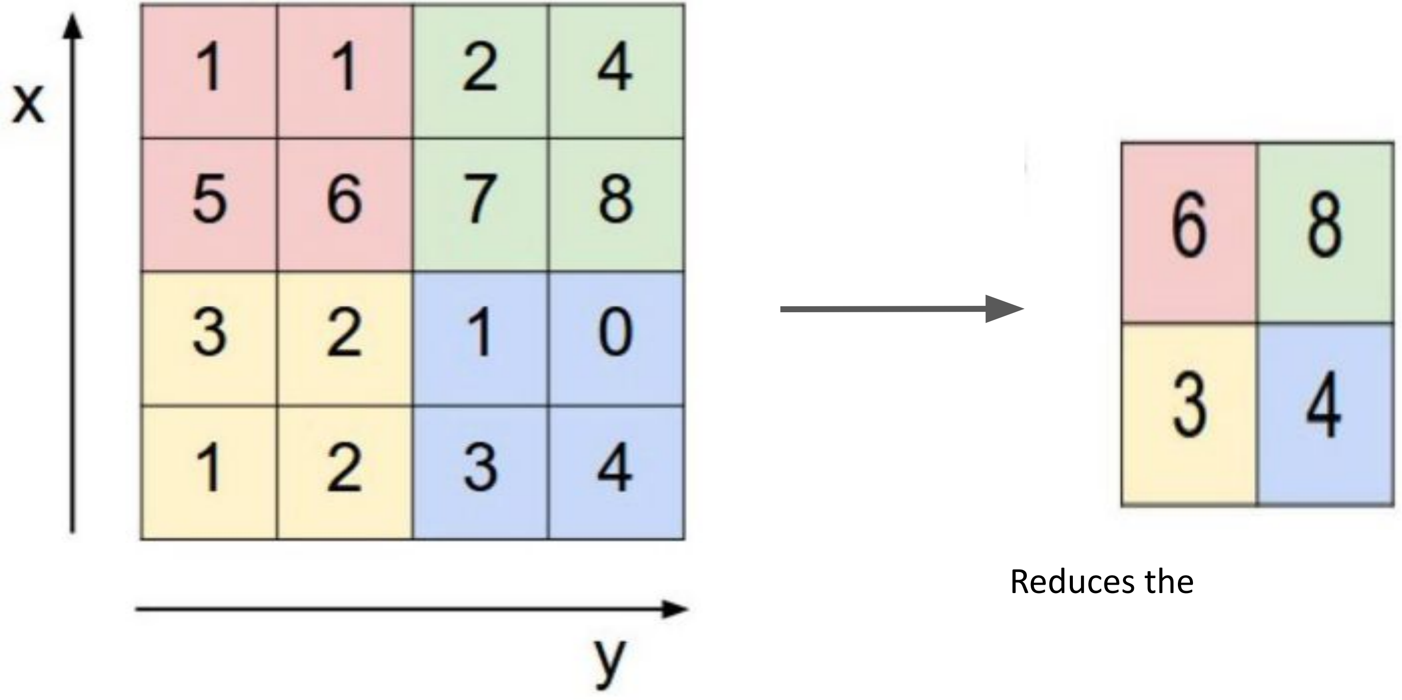


Convolutional layer: *(not really) details*

- How to deal with the borders → ***padding***
- How to choose the initial values of the kernels → ***initializers***
- How to prevent a single feature (or few of them) determine the outcome of the whole network → ***regularizers***
- In the "scan" of the filter, can one skip some steps to reduce the image? → ***stride***
- Can filter be applied to non-contiguos chunks of the image? → ***dilation***



Pooling layers: *example of max pooling*



Reduces the

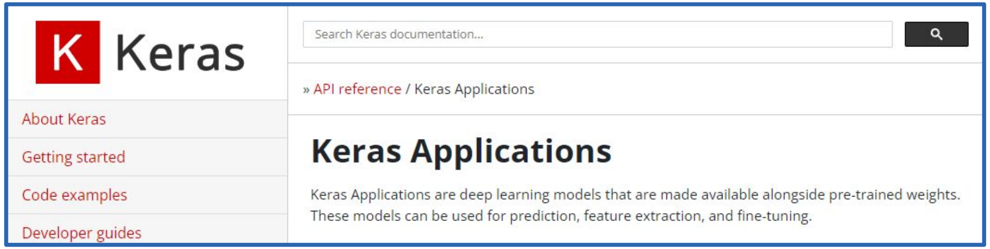
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Transfer learning: *(not) all images are created equal*

You are able to use the same eye, retina and brain to say a cat from a dog, to guess the outdoor weather from the color of the sky and to read a book.

Even if the features involved are drastically different.

Hence, you can train a neural network di distinguish dogs from cats (using images from the social media) and reuse the first layers for an OCR algorithm.



For a database of public pretrained model you can grab layers from, check ***Keras Applications***.

Standard CV datasets

To compare the performance of algorithms or simply to verify if an idea is applicable to CV, researchers use “standard datasets”.

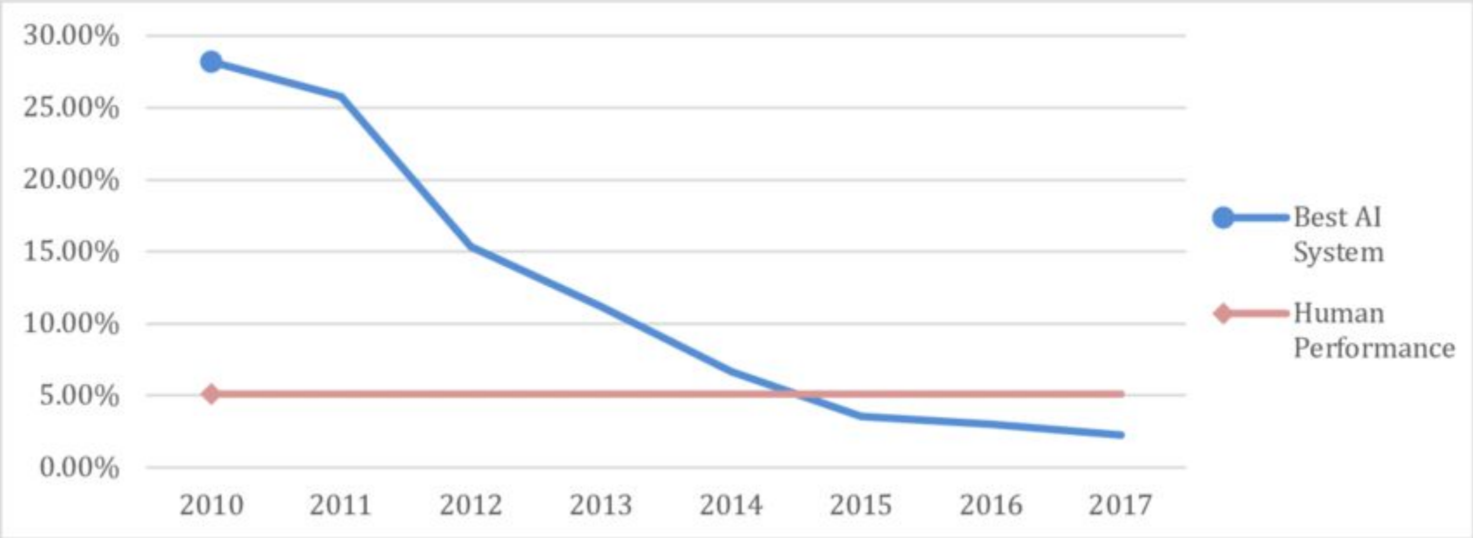
The simplest one is MNIST, a labeled set of handwritten digits with standard and common preprocessing.

For modern ML, MNIST is too simple, though and not representative for modern computer vision.

The standard dataset today is probably ImageNet.

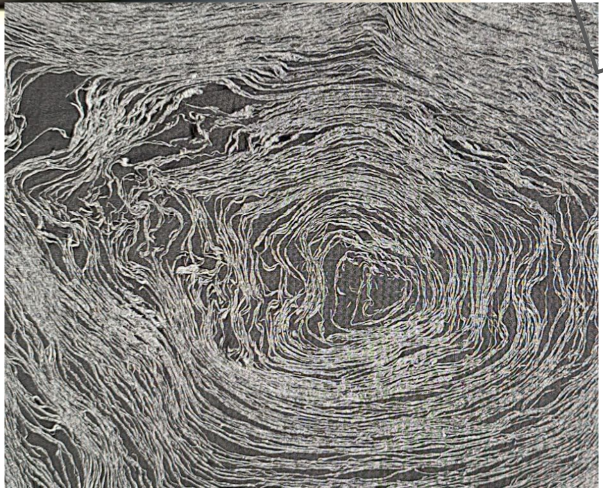
Challenges then…

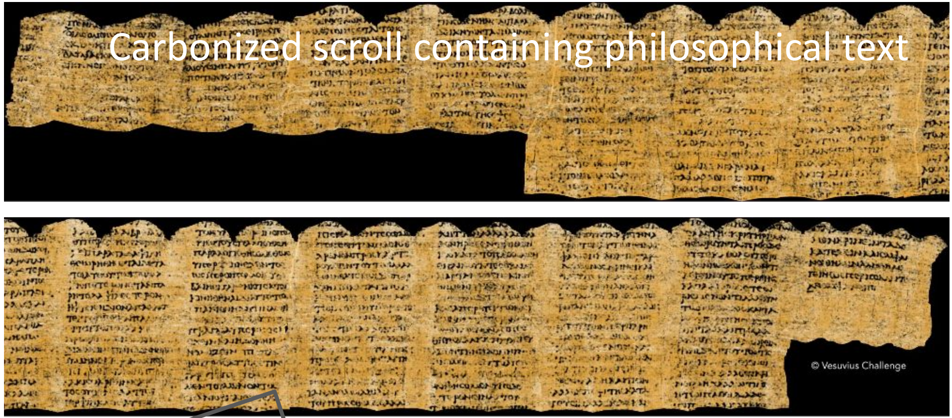
Few years after the introduction of Convolutional Neural Networks, Computer Vision started beating humans at classification tasks on ImageNet.



Challenges then and now

CT SCAN

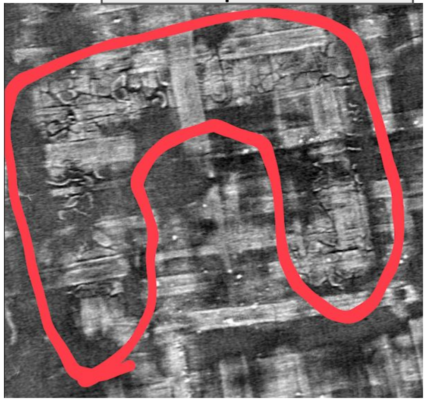




Full story: scrollprize.org

Feature detection

Feature interpretation



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Conclusion

Computer Vision is a branch of Computer Science dedicated to the automated processing of images.

It combines software challenges and advanced statistics for interpretation.

GPUs and Convolutional Neural Networks were game changers.

Today, new field of applications even extending the “vision” to probes the human eye cannot process (non-visible light, tomographies…).

References (I deliberately copied from)

- https://www.slideshare.net/knoldus/introduction-to-computer-visionpdf
- http://cs231n.stanford.edu/slides/2016/winter1516_lecture7.pdf
- https://mit6874.github.io/assets/sp2021/slides/l03.pdf