Computer Vision and Pattern Recognition

Giacomo Boracchi

giacomo.boracchi@polimi.it

February 18th, 2020

USI, Lugano

The Team

Who I am

Giacomo Boracchi (giacomo.boracchi@polimi.it)

Mathematician (Università Statale degli Studi di Milano 2004),

PhD in Information Technology (DEIB, Politecnico di Milano 2008)

Associate Professor since 2019 at DEIB, Polimi (Computer Science)

My research interests are mathematical and statistical methods for:

- Image analysis and processing
- Machine Learning and in particular unsupervised learning, change and anomaly detection

… and the two combined

Who I am

Pietro Verzelli (pietro.verzelli@usi.ch)

Bachelor in Physics (University of Bologna, 2015)

Master in Physics of Complex Systems (University of Turin, 2017)

Currently a PhD student here at USI

In my research I study the relationship between dynamical systems and machine learning, in particular recurrent neural networks.

What about you?

Let’s start

Photo Credits: Andrea Sanfilippo

RGB Images

$I \in \mathbb{R}^{R \times C \times 3}$

$R \in \mathbb{R}^{R \times C}$

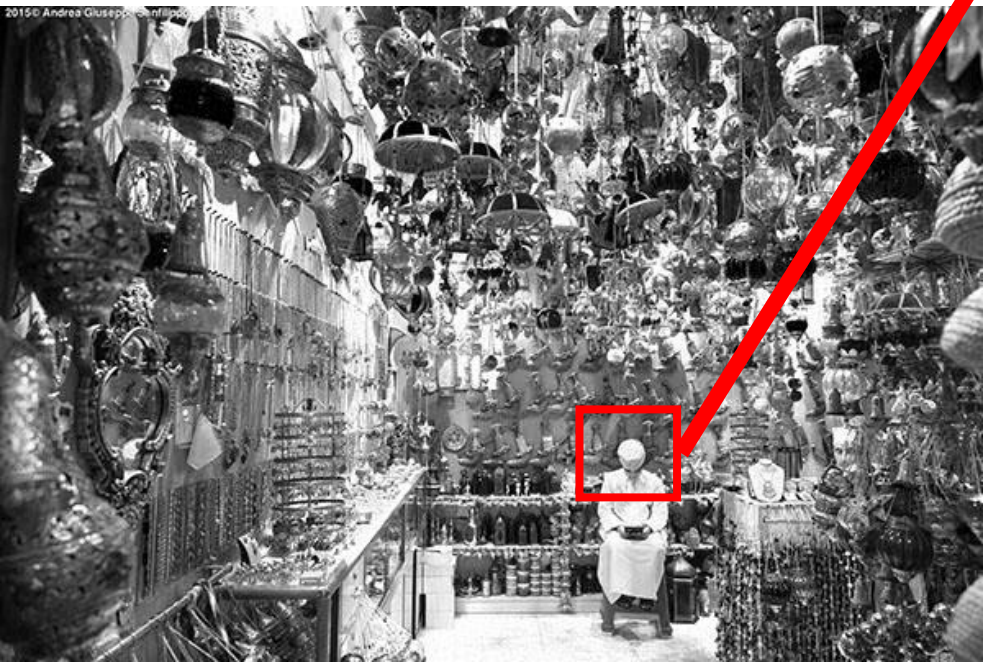
$G \in \mathbb{R}^{R \times C}$

$B \in \mathbb{R}^{R \times C}$

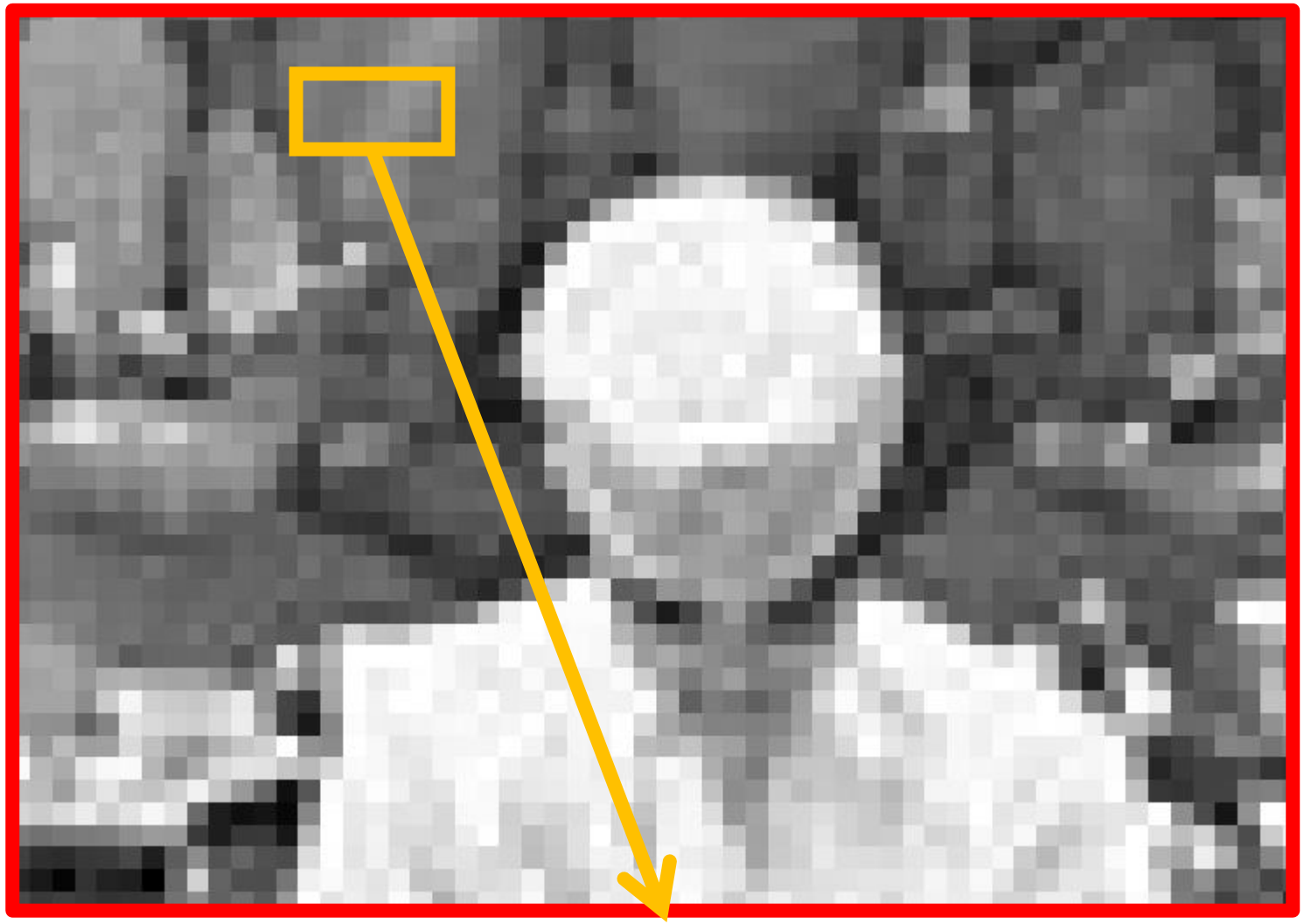
Giacomo Boracchi

RGB Images

Images are saved by encoding each color information in 8 bits. So images are rescaled and casted in [0,255]



$$R \in \mathbb{R}^{R \times C}$$



123	122	134	121	132
122	121	125	132	124
119	127	137	119	139

RGB Images

[0, 205, 155]

[15, 17, 19]

[230,234, 233]

[253, 5, 6]

[106,124,138]

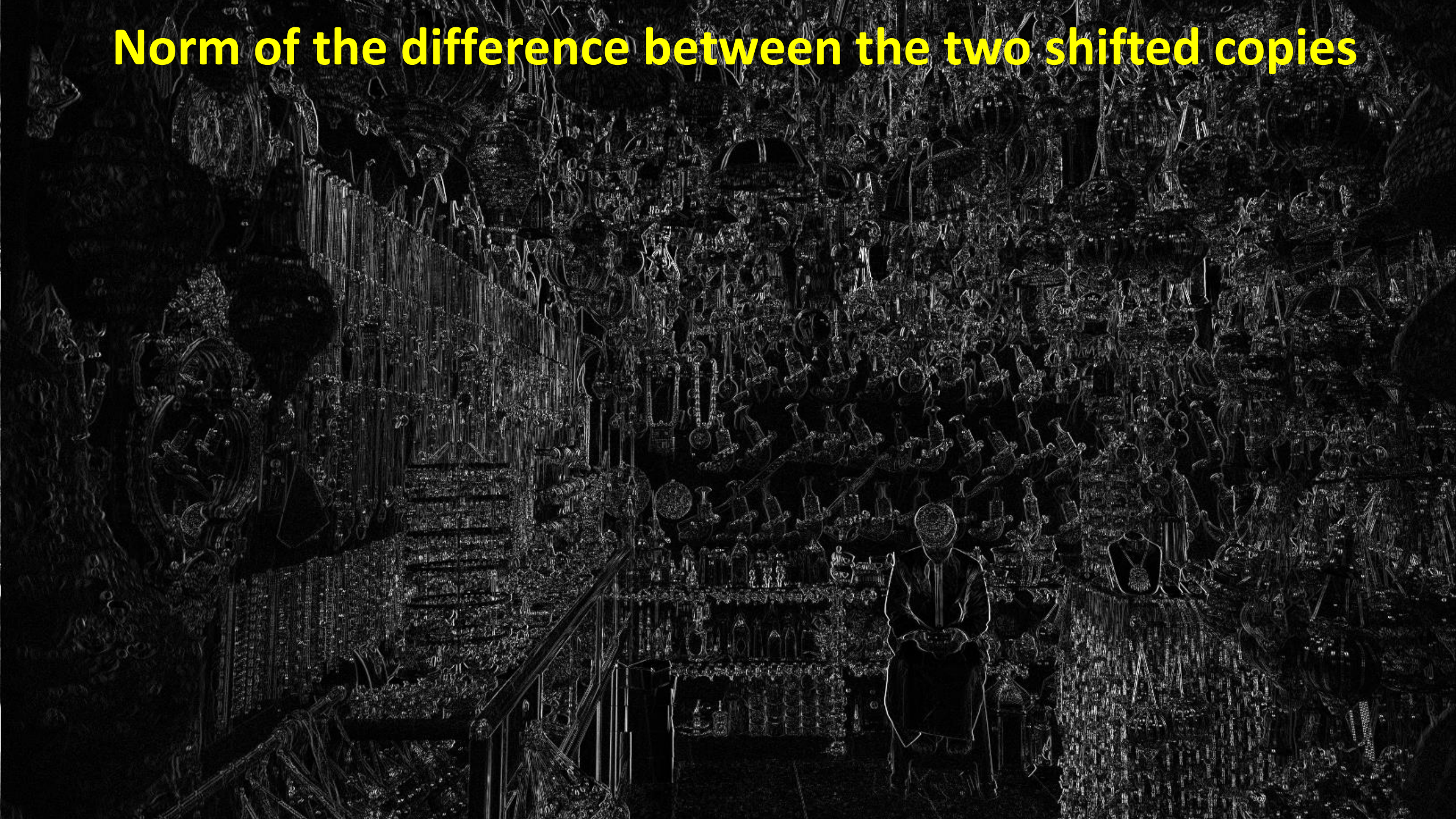
Giacomo Boracchi

Information is difficult to grasp when we look at images as matrices

Photo Credits: Andrea Sanfilippo

That was a shift of a single pixel.
It does not seem to change the information
in this image

Norm of the difference between the two shifted copies




Original Range [0-380]

Can you see only pixels?


Photo Credits: Andrea Sanfilippo

So why Computer Vision and Pattern Recognition?

Computer Vision

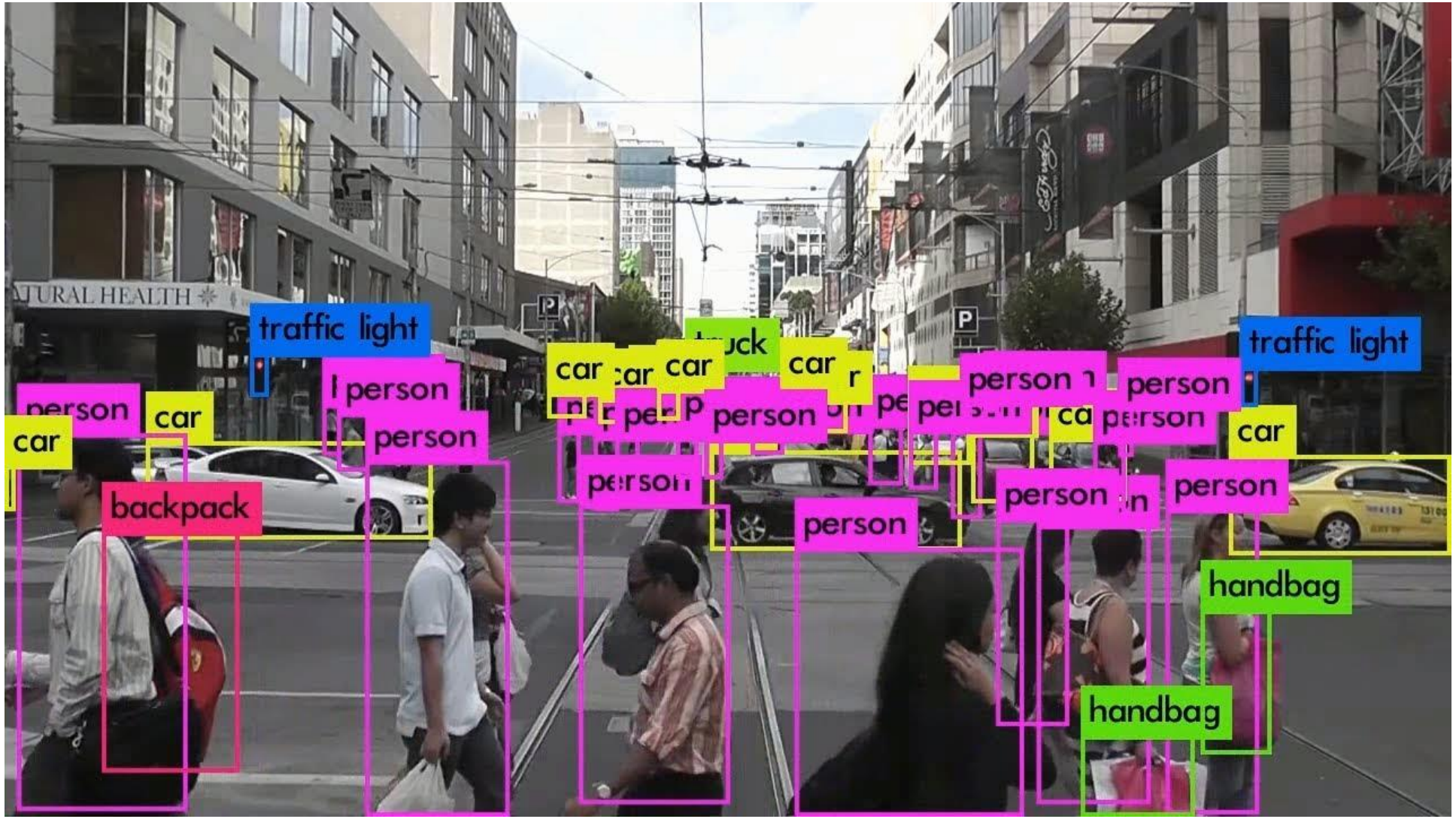
An interdisciplinary scientific field that deals with how computers can be made to **gain high-level understanding from digital images or videos**

Computer Vision

An interdisciplinary scientific field that deals with how computers can be made to **gain high-level understanding from digital images or videos**

… which has grown incredibly fast in the last years

Object Detection



Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

Pose Estimation

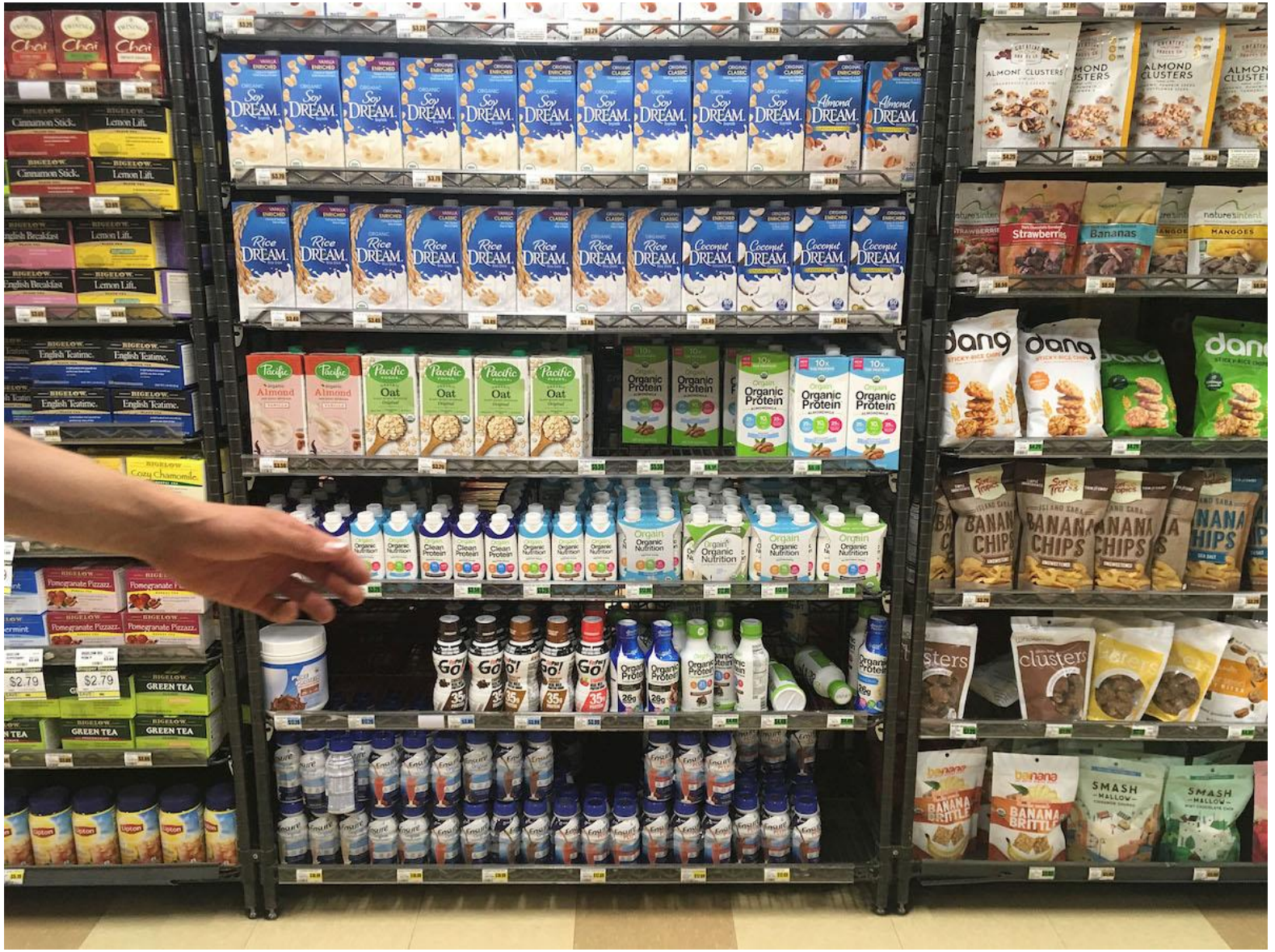
Cao, Z., Hidalgo, G., Simon, T., Wei, S.E. and Sheikh, Y.. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. CVPR 2017 arXiv preprint arXiv:1812.08008.

Autonomous Driving

By Dllu - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=64517567

Ian Maddox [CC BY-SA (https://creativecommons.org/licenses/by-sa/4.0)]

Automatic Shelf Analysis

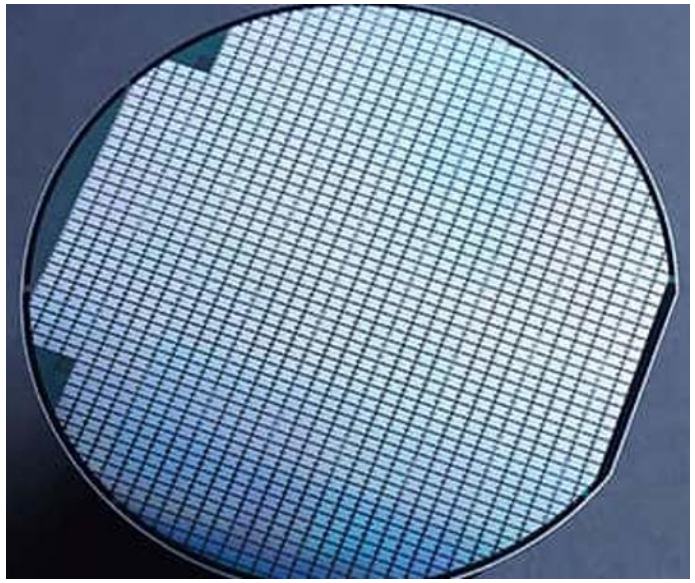


http:focal.systems

Automatic Shelf Analysis

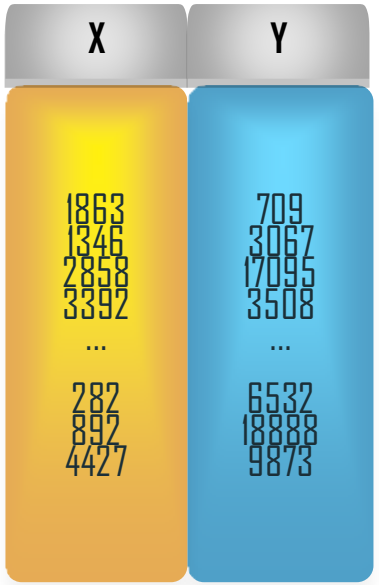
http:focal.systems

Quality Inspection

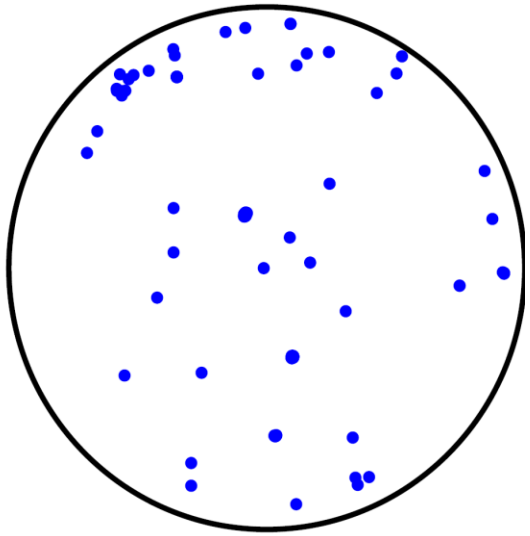


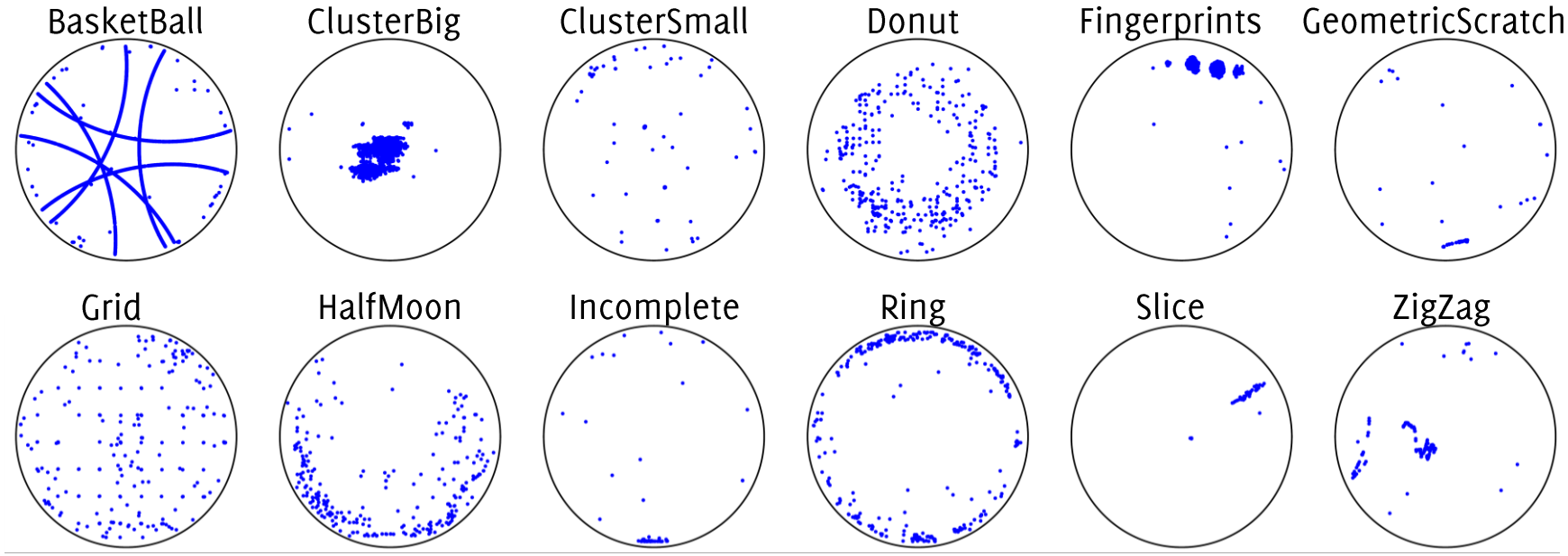
Silicon Wafer

Inspection Tool



Wafer Defect Map (WDM)





Di Bella, Carrera, Rossi, Fragneto, Boracchi *Wafer Defect Map Classification Using Sparse Convolutional Neural Networks* ICIAP09

Image Searches

Inbox (39) - giacomo79@gmail.c

rabbit - Google Photos

photos.google.com/search/rabbit

rabbit

Sat, Apr 6

Thu, Apr 4

Sat, Apr 9, 2016

Image / Video Restoration

Noisy 16.10 dB

Image / Video Restoration

Restored 28.49 dB

Maggioni, M., Boracchi, G., Foi, A., & Egiazarian, K. (2012). Video denoising, deblocking, and enhancement through separable 4-D nonlocal spatiotemporal transforms. IEEE *TIP 2012*

Inpainting

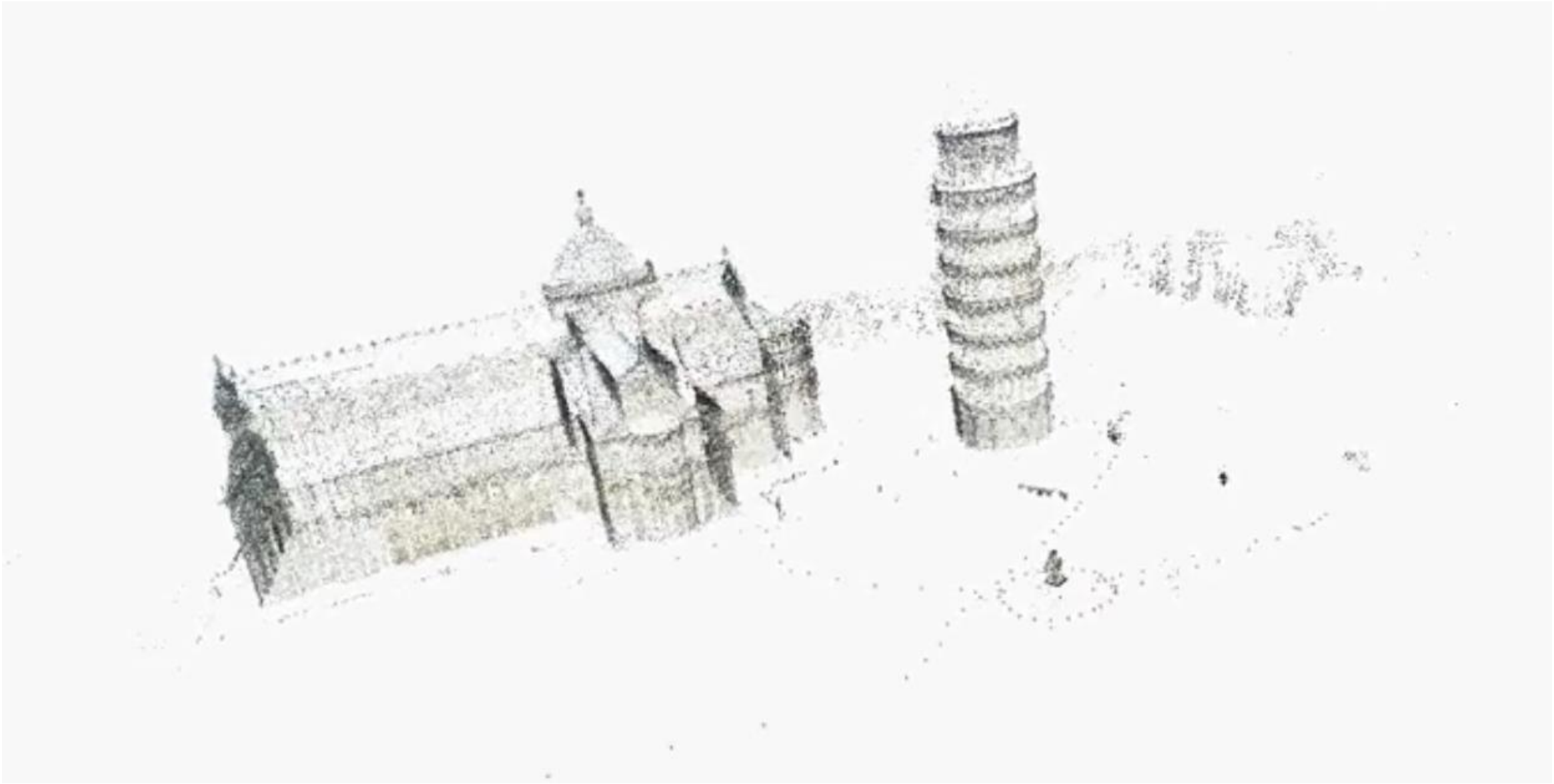
Original damaged photo

Restored photo

Bertalmio, M., Sapiro, G., Caselles, V., & Ballester, C. (2000, July). Image inpainting. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (pp. 417-424).

Enhancement (Instance Segmentation)

3D Reconstruction



… and Pattern Recognition

The field of pattern recognition is concerned with

- **the automatic discovery of regularities in data** through the use of computer algorithms and with
- **the use of these regularities to take actions** such as classifying the data into different categories

CV and PR are very related, since regularities is what we use to understand images

… and Pattern Recognition

The field of pattern recognition is concerned with

- **the automatic discovery of regularities in data** through the use of computer algorithms and with
- **the use of these regularities to take actions** such as classifying the data into different categories

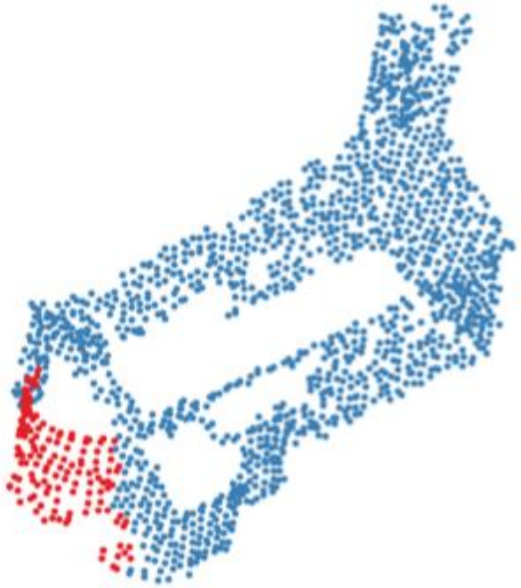
CV and PR are very related, since regularities is what we use to understand images

In this course, **data will be images** where the extraction of patterns is a primary concern.

Pattern Recognition scope is however wider (e.g. biomedical signals, vectors, graphs, etc…)

Model Fitting

Image of the scene



colour coded class

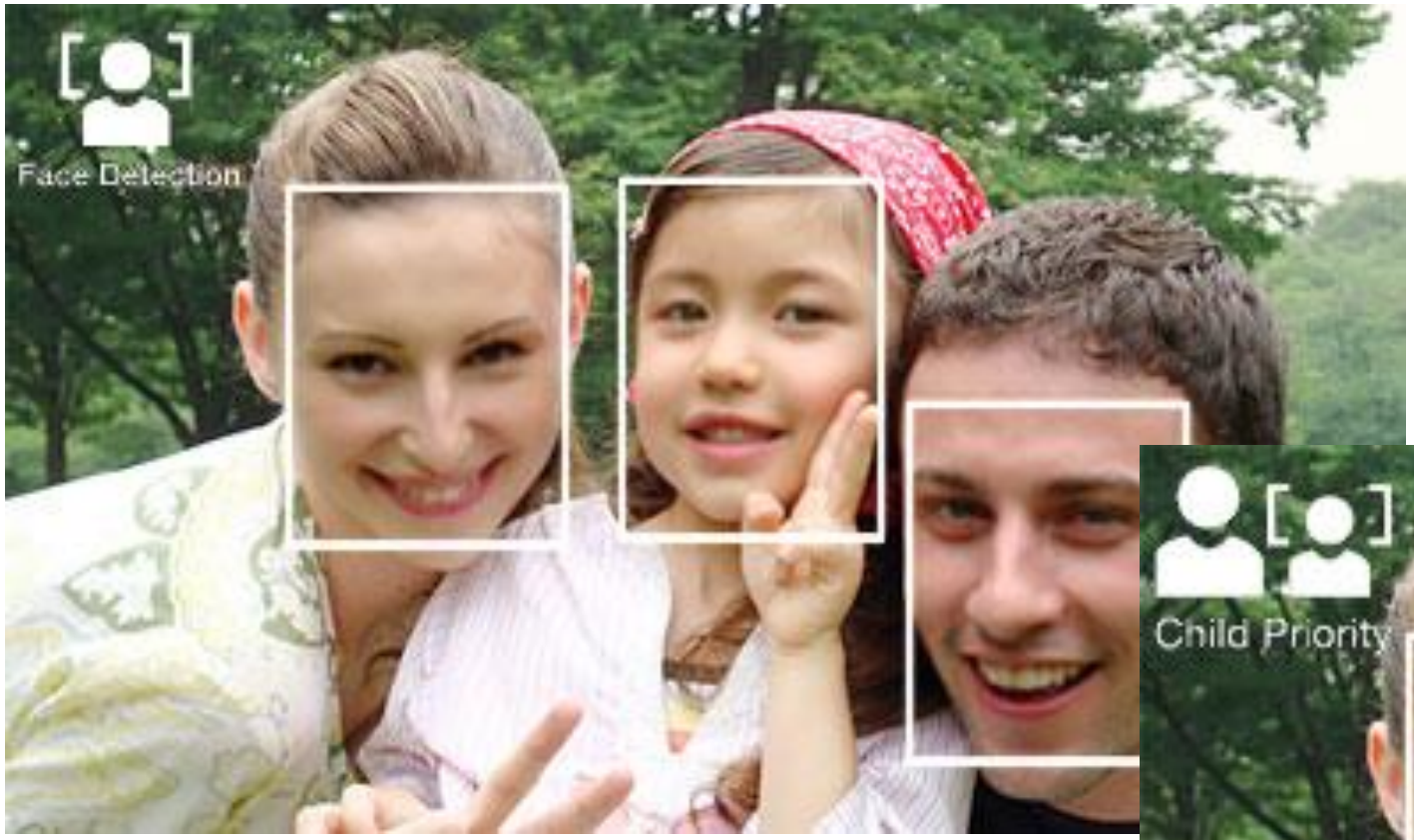
colour coded model

Magri, Luca, and Andrea Fusiello. "Fitting Multiple Heterogeneous Models by Multi-Class Cascaded T-Linkage." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.

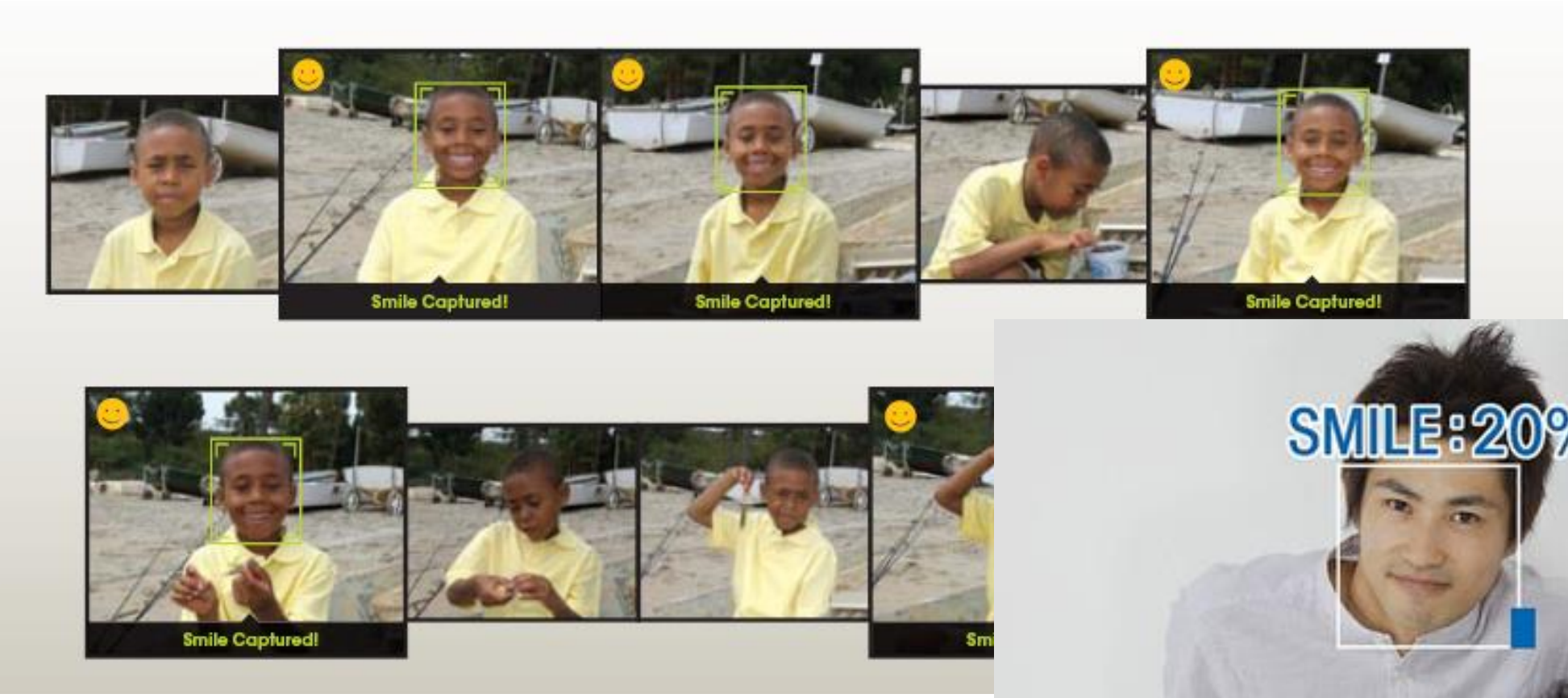
Image Segmentation

Zheng et al. “Conditional Random Fields as Recurrent Neural Networks”, ICCV 2015

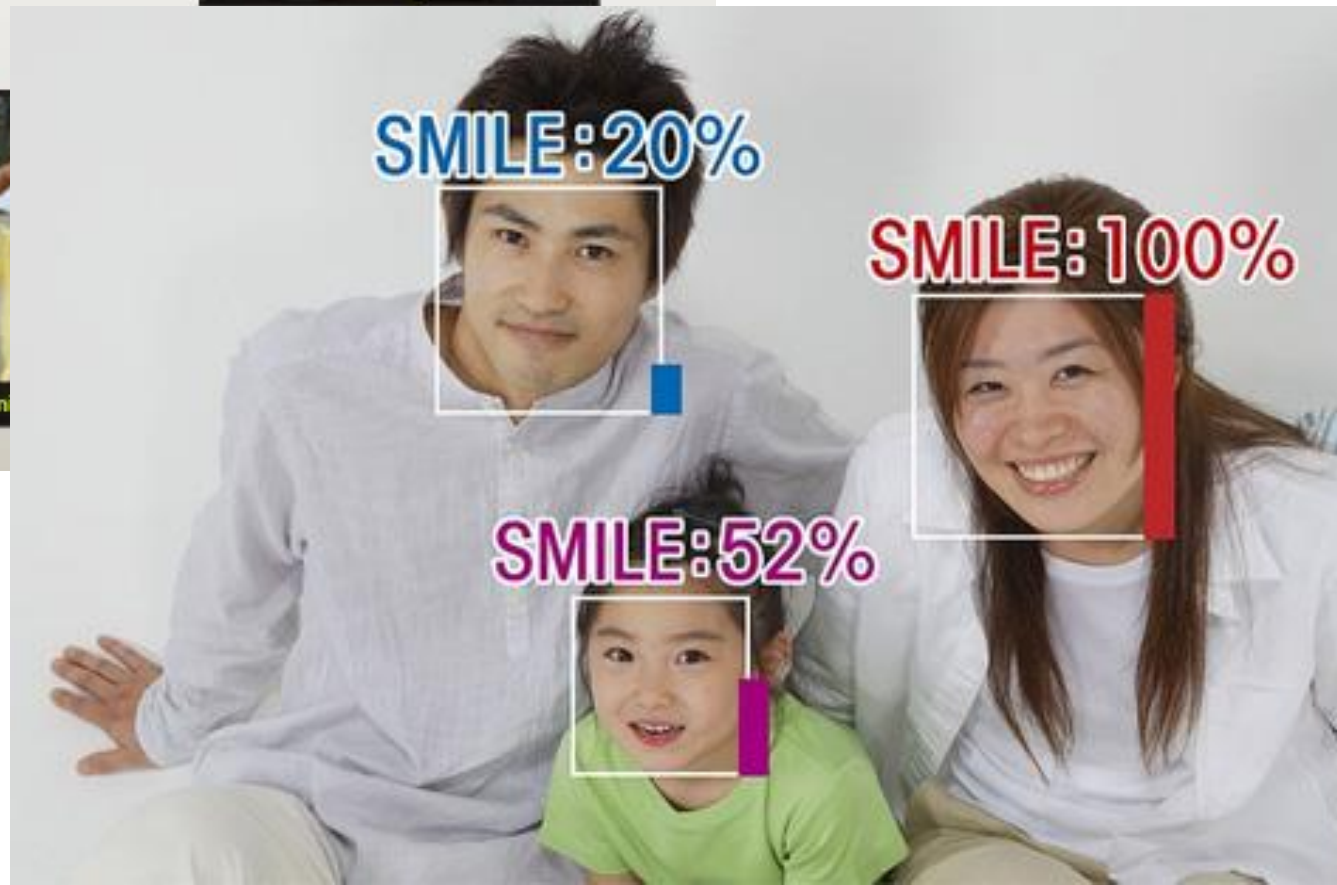
Face Detection



Smile Detection



©Akibahara News

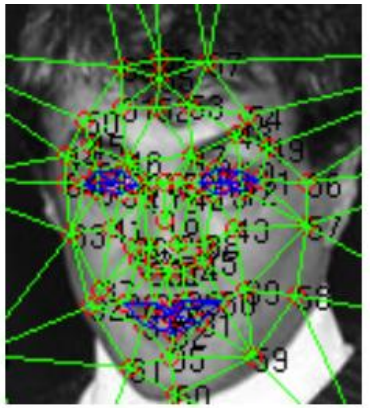


© Sony

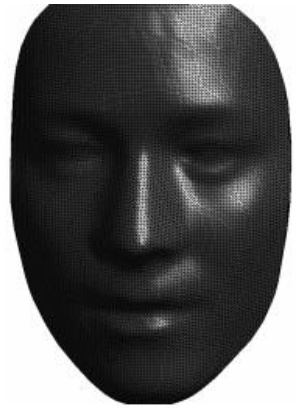
Face Identification

(a)

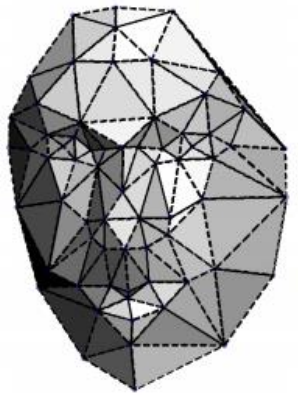
(b)



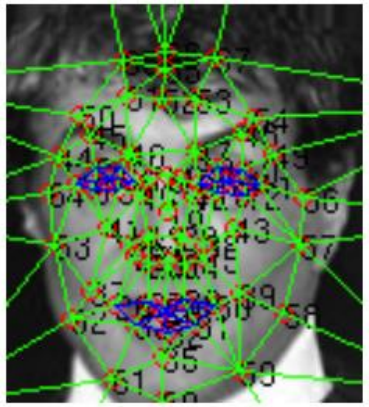
(c)



(d)



(e)



(f)

(g)

(h)

Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. Deepface: Closing the gap to human-level performance in face verification. CVPR 2014

Are these Challenging Problems?

Yes… for several reasons:
Images are very high-dimensional data

CIFAR-10 dataset

The CIFAR-10 dataset contains 60000 images:

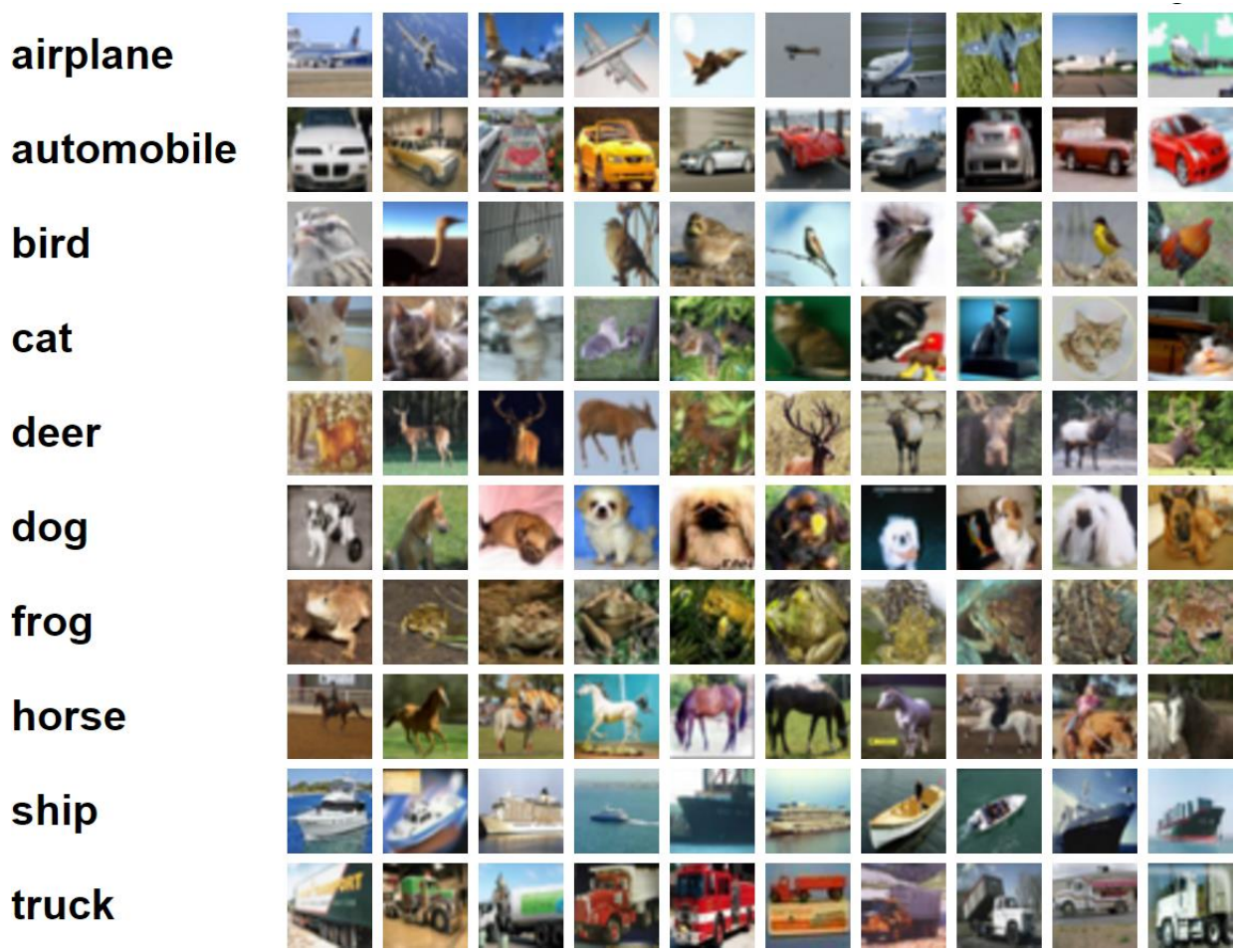
Each image is 32x32 RGB

Images are in 10 classes

6000 images per class

Extremely small images, but high-dimensional:

$$d = 32 \times 32 \times 3 = 3072$$



Learning Multiple Layers of Features from Tiny Images, Alex Krizhevsky, 2009.


This resolution is by far smaller than what we are used to
$d = 3072$

Former standard repository for ML research

$d = 3072$

UCI Machine Learning Repository
Center for Machine Learning and Intelligent Systems

Attributes
Less than 10 (116)
10 to 100 (218)
Greater than 100 (86)

- 88% < 500 attributes
- 92% < 3.2K attributes

It is even difficult to describe unambiguously image content

Label Ambiguity

Man?

Beer?

Dinner?

Restaurant?

Sausages?

….

There are many **transformations** that change the image matrix drammaticaly, while not its semantic content

Illumination conditions changes

CS231n: Convolutional Neural Networks for Visual Recognition http://cs231n.github.io/

Deformations

Copyright Christine Matthews

© Copyright Patrick Roper

View Point Change

… and many othres

Occlusion

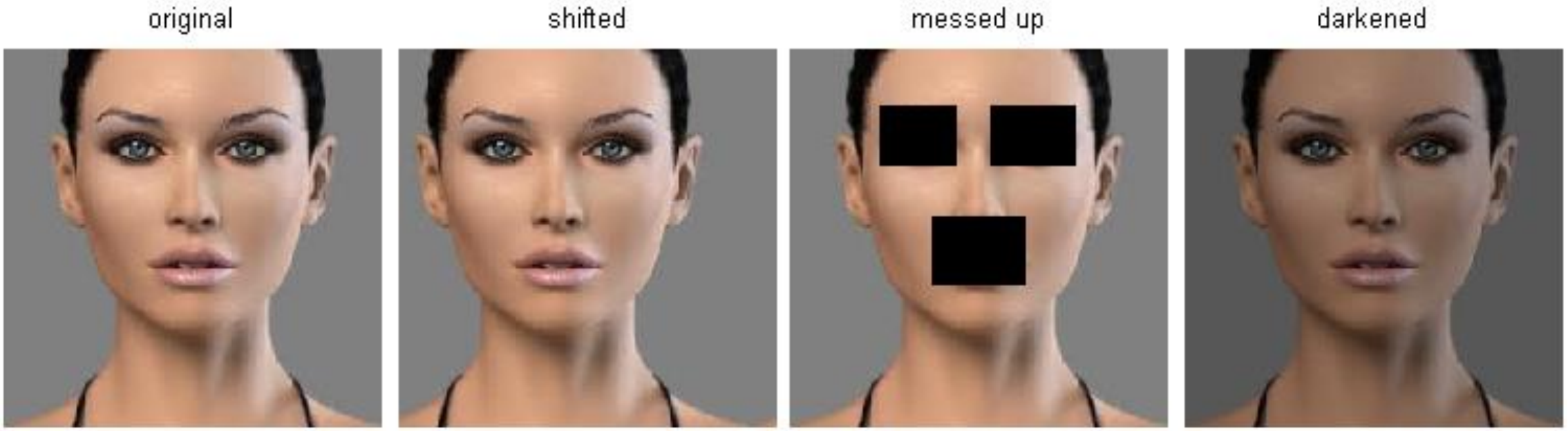
Background clutter

Scale variation

Perceptual similarity in images is not related to pixel-similarity

Giacomo Boracchi

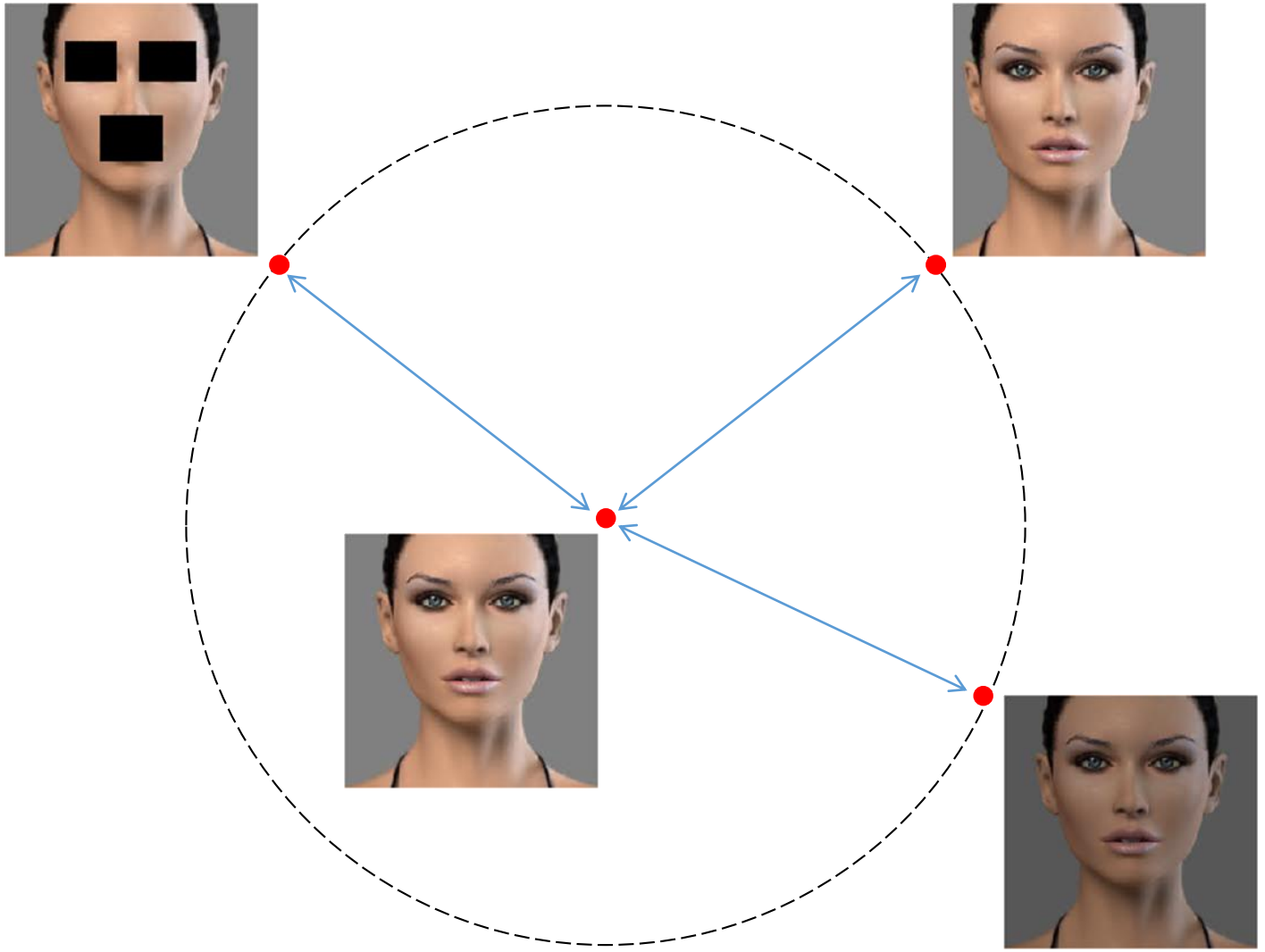
Perceptual Similarity vs Pixel Similarity



The three images have the same pixel-wise distance from the original one

…but perceptually they are very different

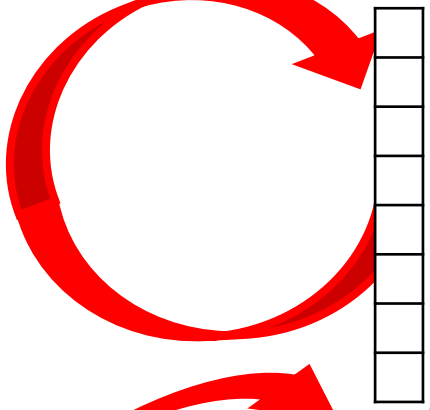
Perceptual Similarity vs Pixel Similarity

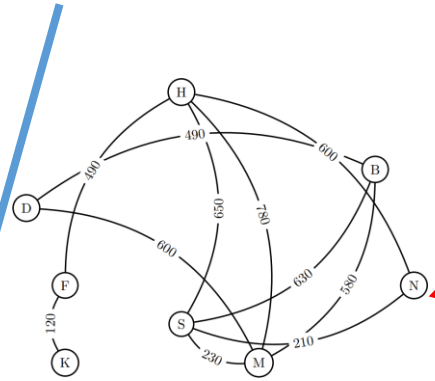


The Course Outline and the Broad Landscape of CVPR

Giacomo Boracchi

Machine Learning
Pattern Analysis



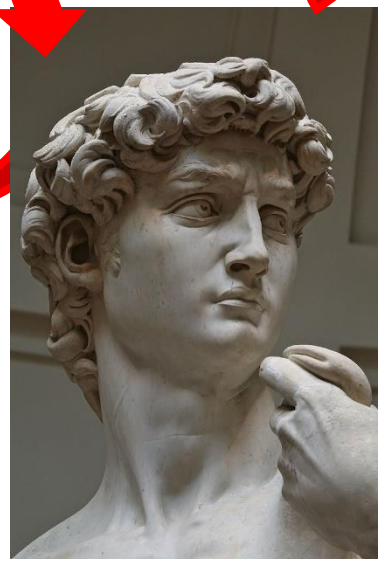


nD

Image Analysis

Shape Analysis

Image
Processing



2D

Computer Graphics

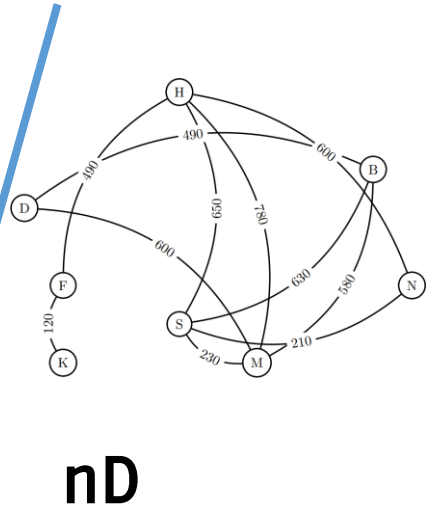
3D

Geomery
Processing

Computer Vision

from M. Bronstein

Machine Learning
Pattern Analysis



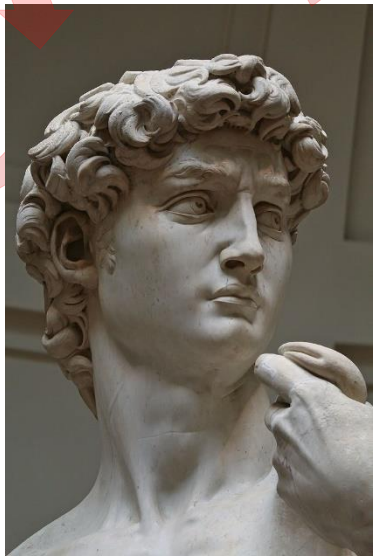
nD

Image Analysis

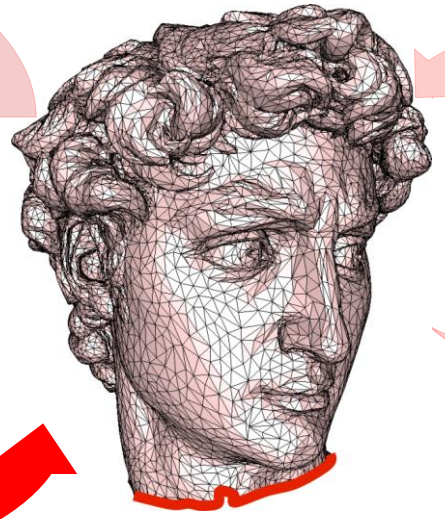
Shape Analysis

Computer Graphics

Image
Processing



2D

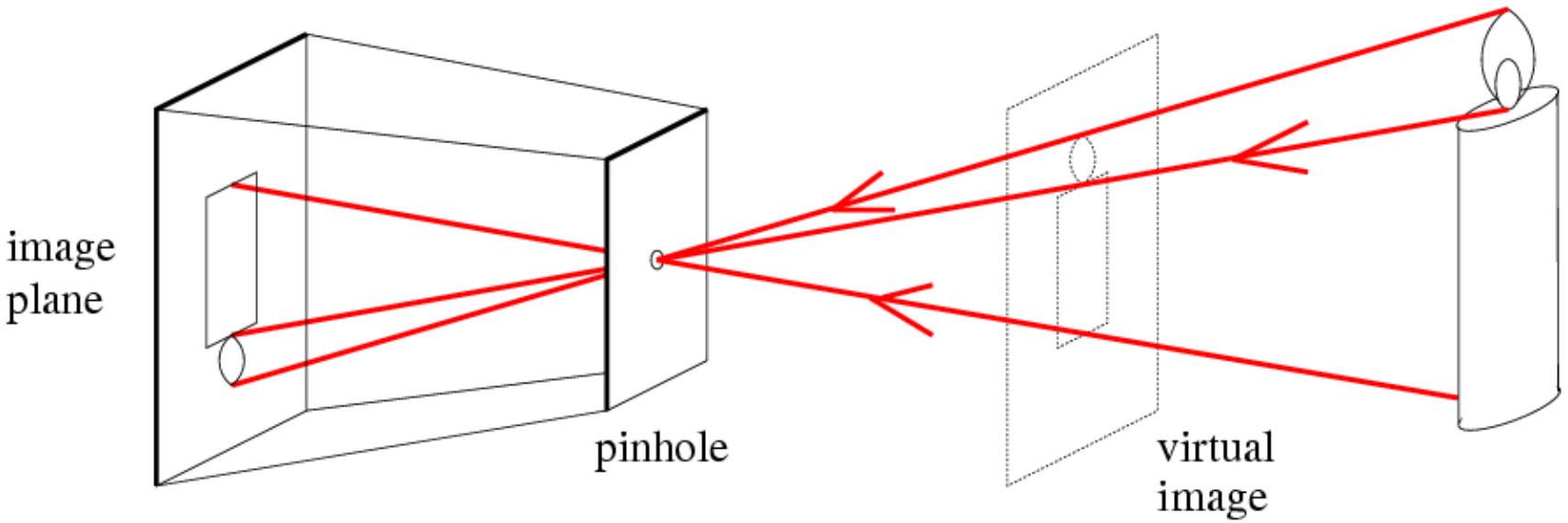


3D

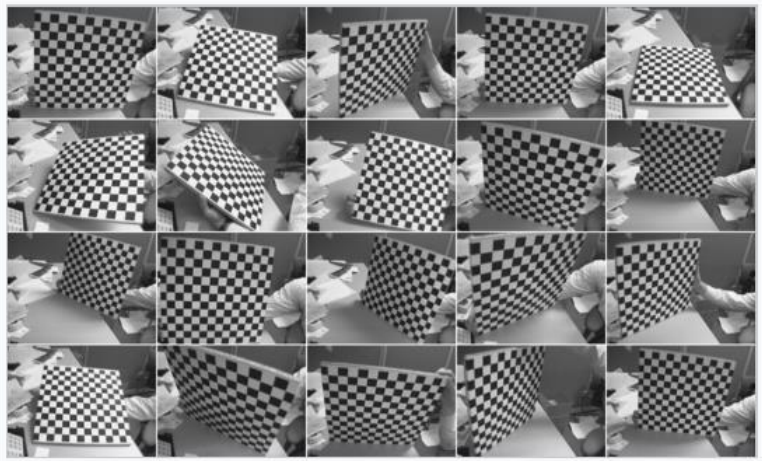
Geomery
Processing

Computer Vision

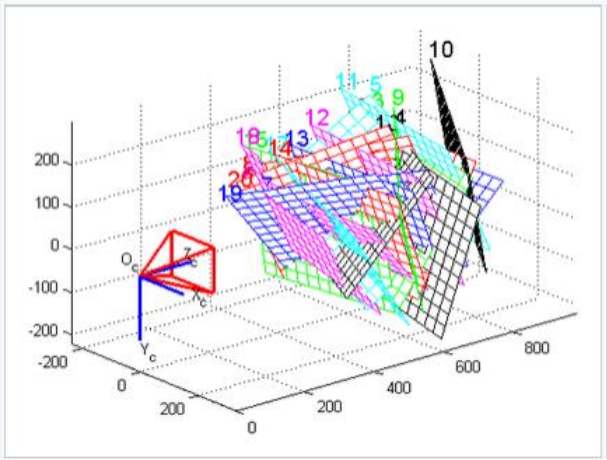
Camera Models and Calibration



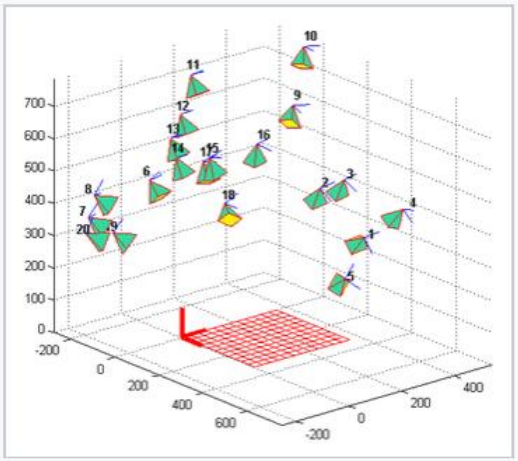
Example: multiplane calibration



Multiple views of a chessboard for multiplane calibration



Reconstructed orientations
(camera-centric coordinates)



Reconstructed orientations
(world-centric coordinates)

Single View Geometry

HZ Fig.2.18

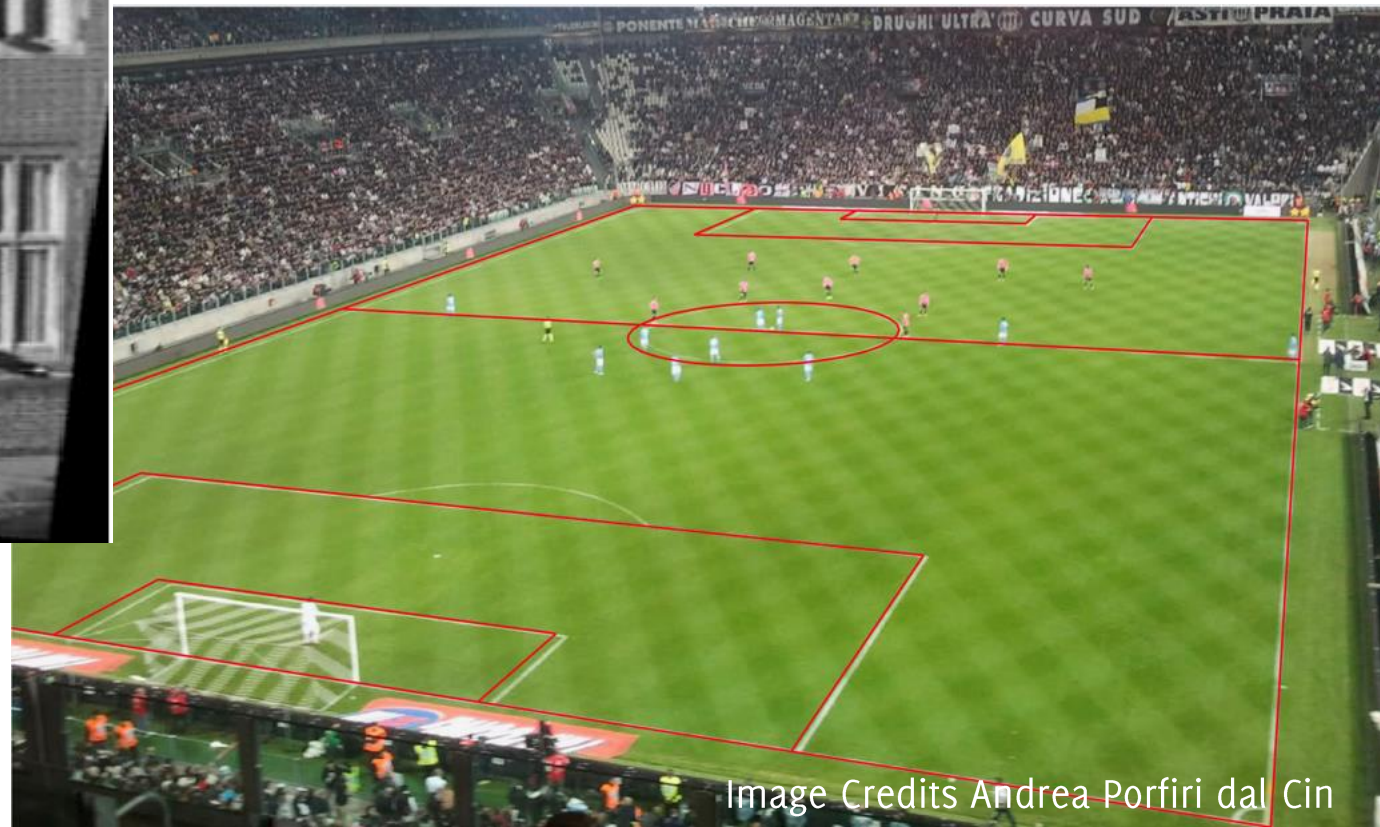
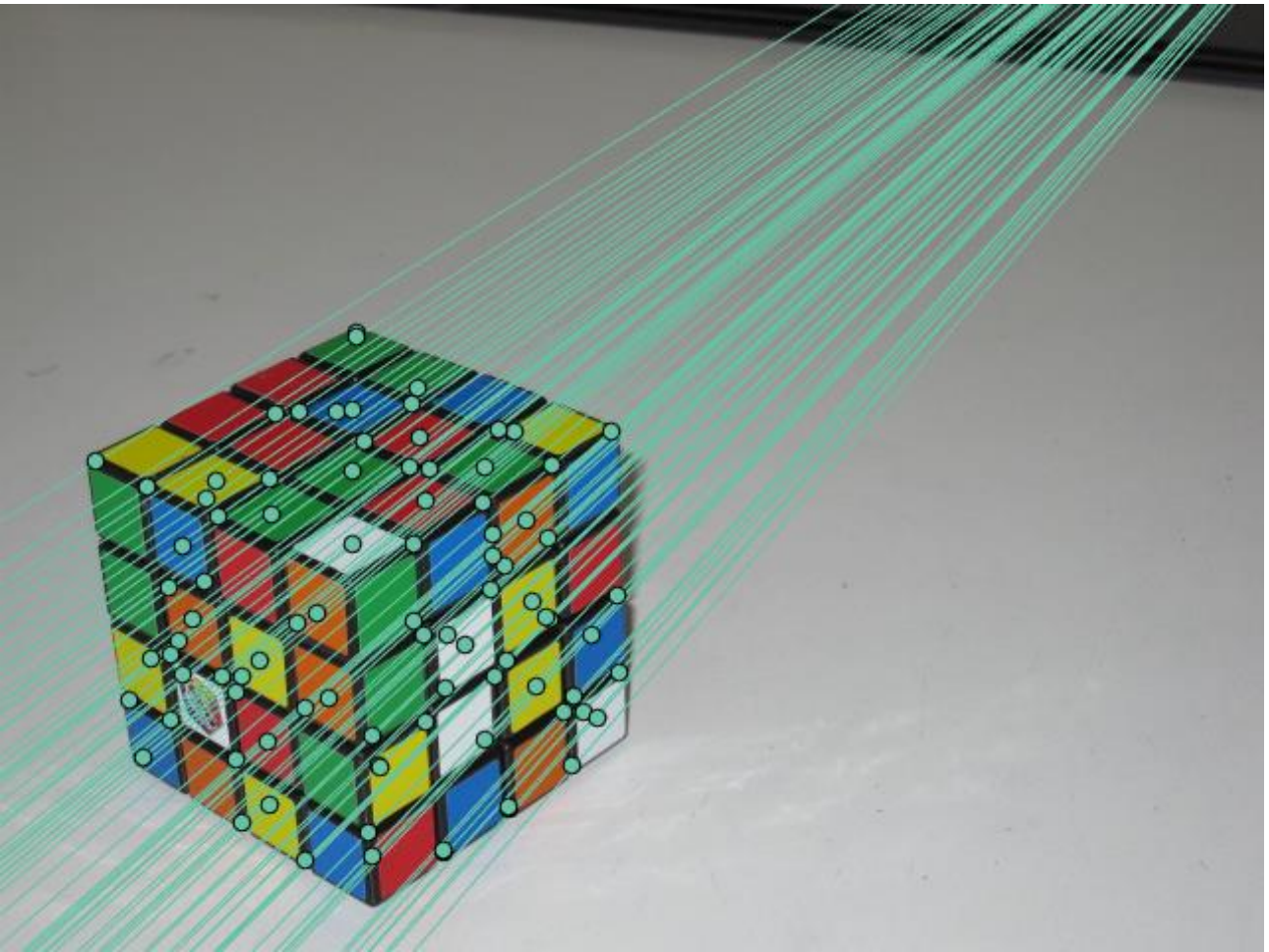
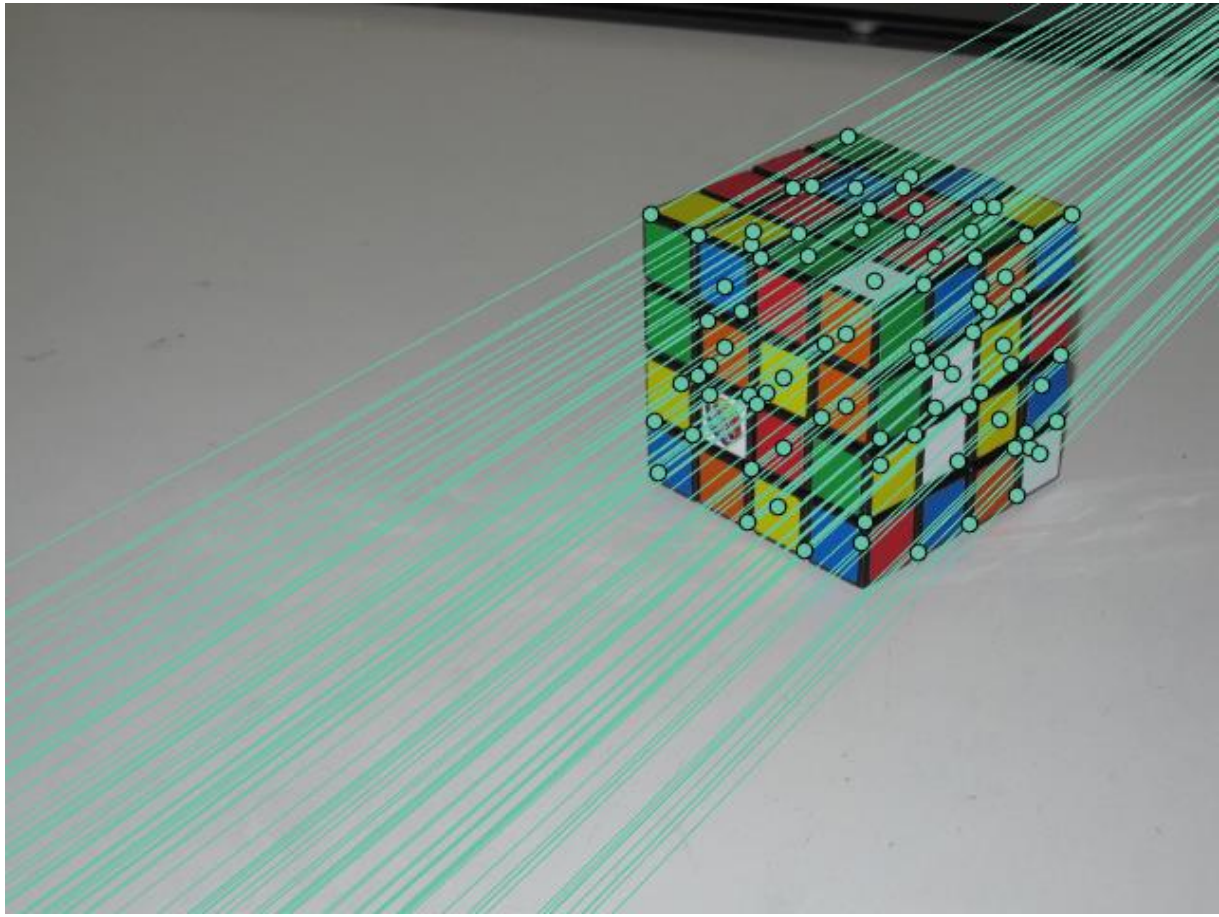


Image Credits Andrea Porfiri dal Cin

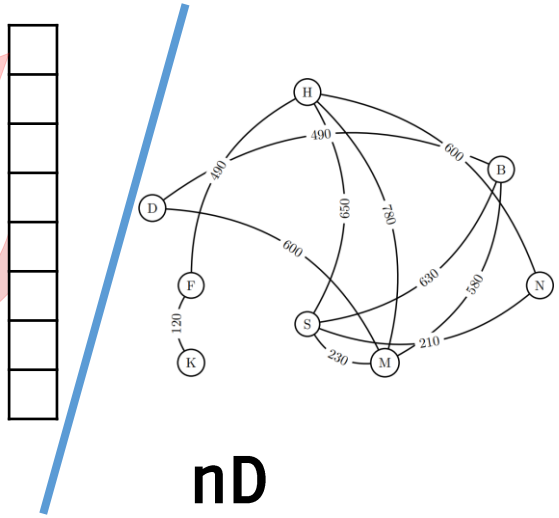
Multi-View Geometry (Stereo System, 3D reconstruction)





Giacomo Boracchi

Machine Learning
Pattern Analysis

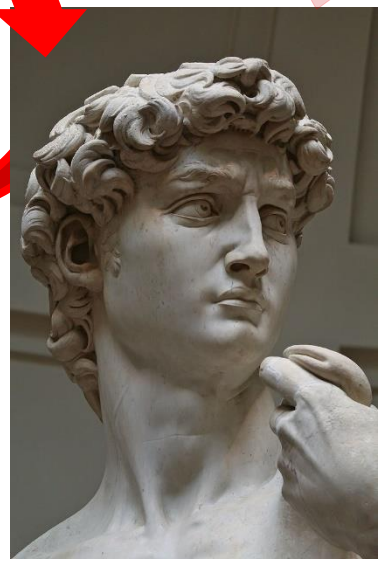


nD

Image Analysis

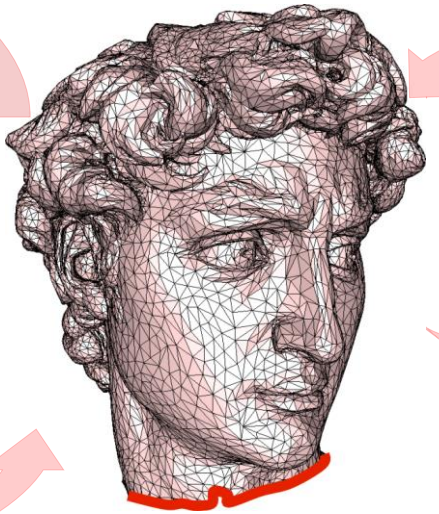
Shape Analysis

Image
Processing



2D

Computer Graphics

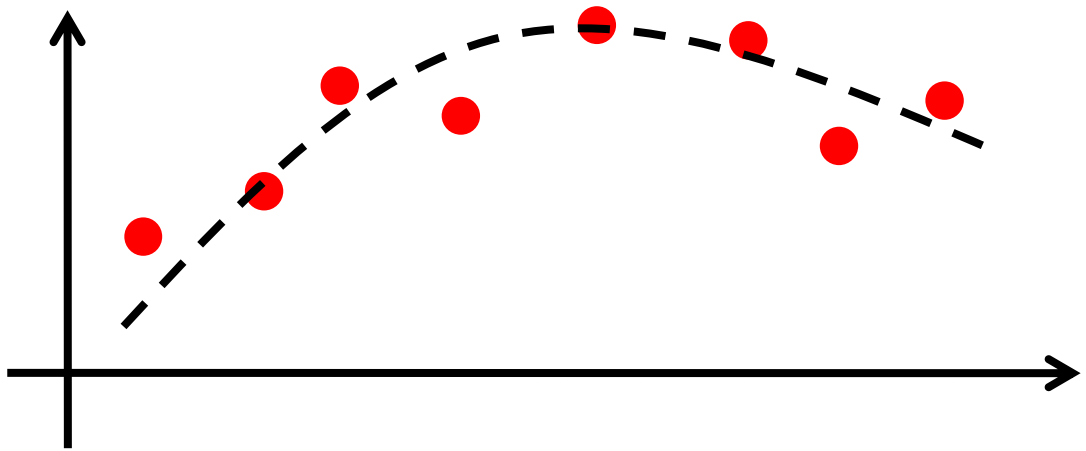


3D

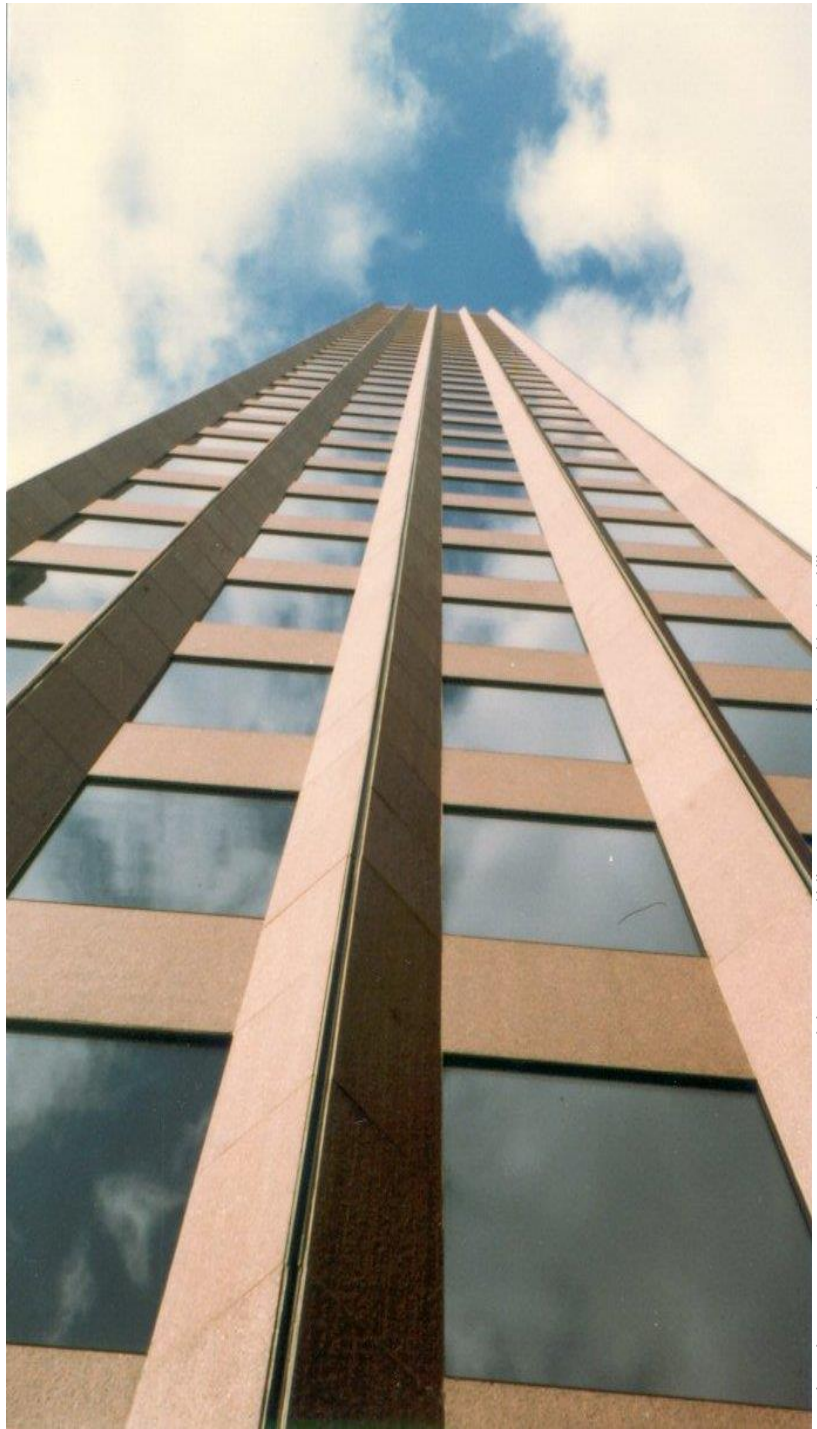
Geomery
Processing

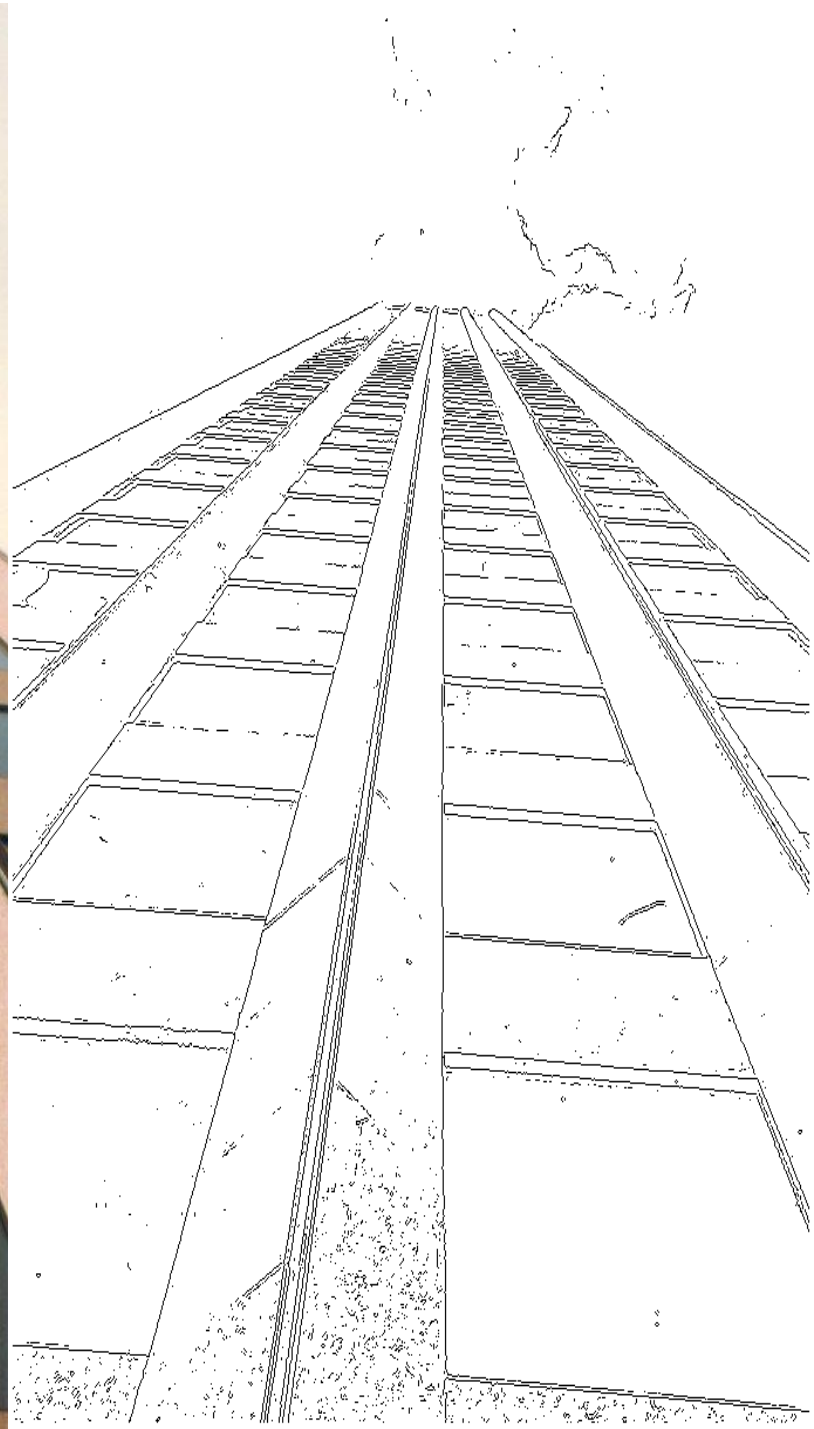
Computer Vision

Filtering

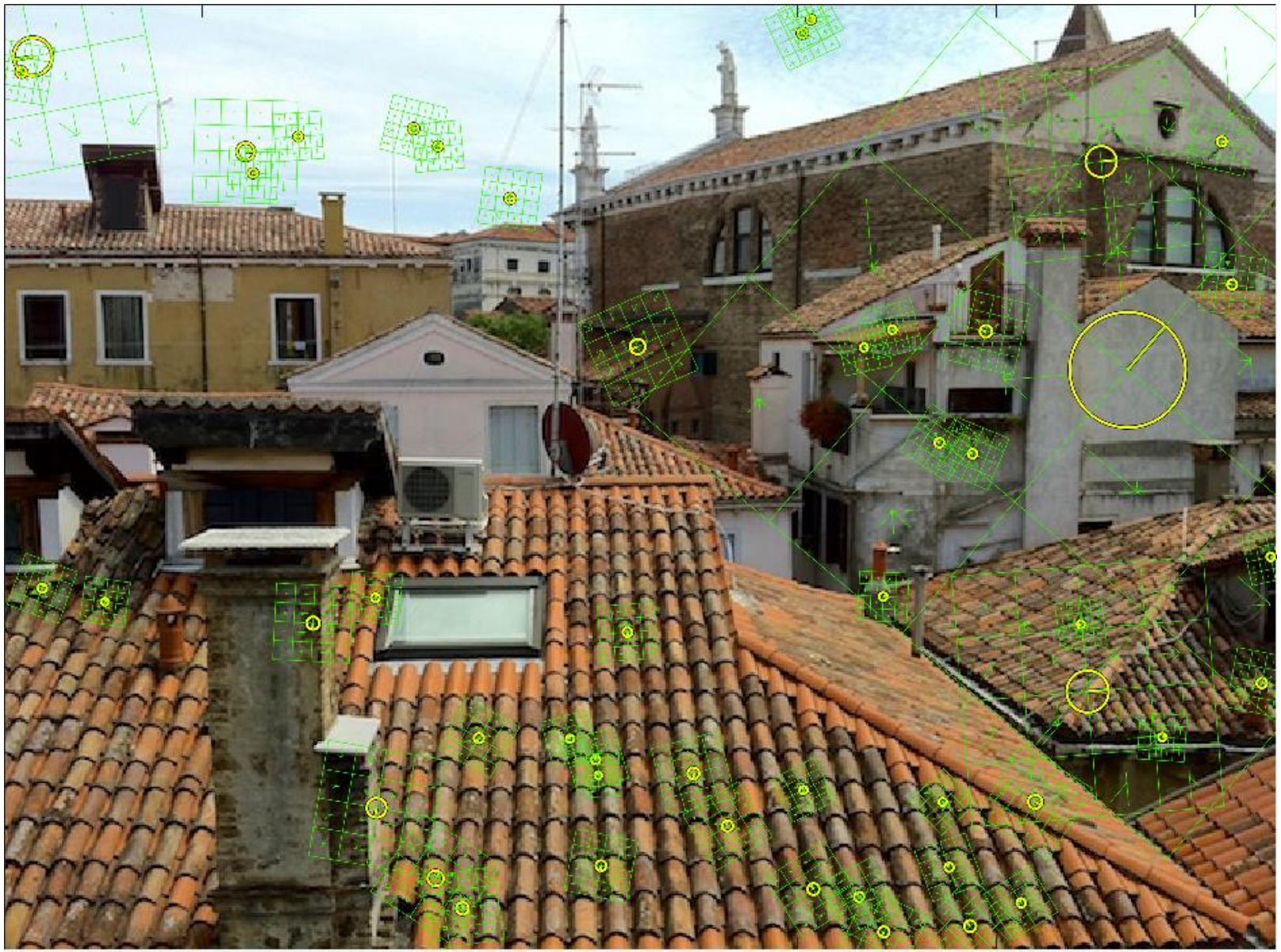


Edge detection





Salient Point / Feature Extraction



Giacomo Boracchi

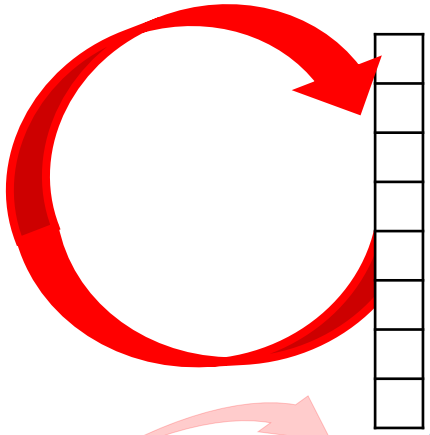
Restoration & Inverse Problems

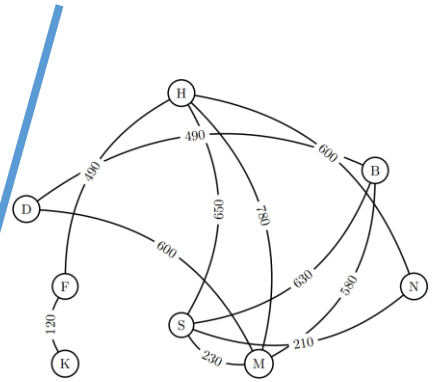
Dabov, K., Foi, A., Katkovnik, V., & Egiazarian, K. Image denoising by sparse 3-D transform-domain collaborative filtering. *TIP 2007*

Restoration & Inverse Problems

Dabov, K., Foi, A., Katkovnik, V., & Egiazarian, K. Image denoising by sparse 3-D transform-domain collaborative filtering. *TIP 2007*

Machine Learning
Pattern Analysis



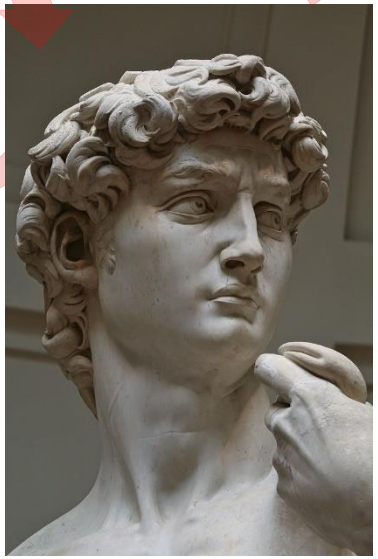


nD

Image Analysis

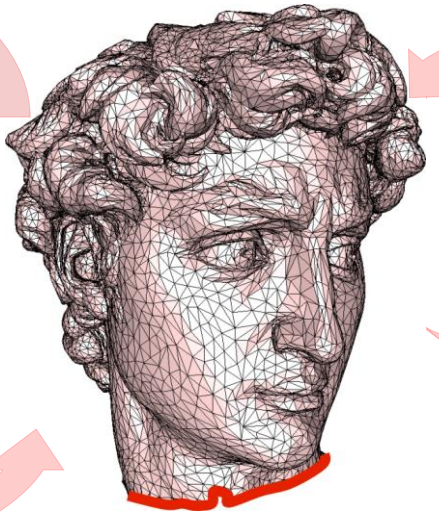
Shape Analysis

Image
Processing



2D

Computer Graphics

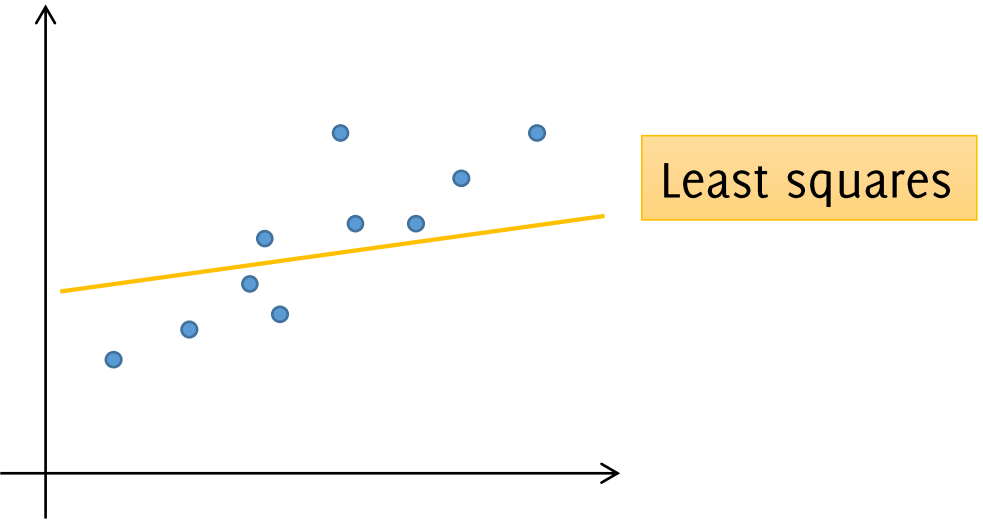


3D

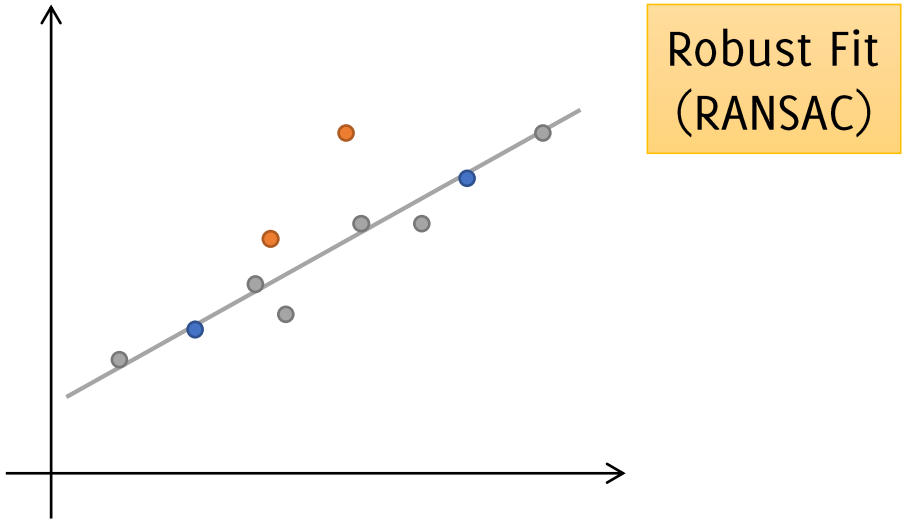
Geomery
Processing

Computer Vision

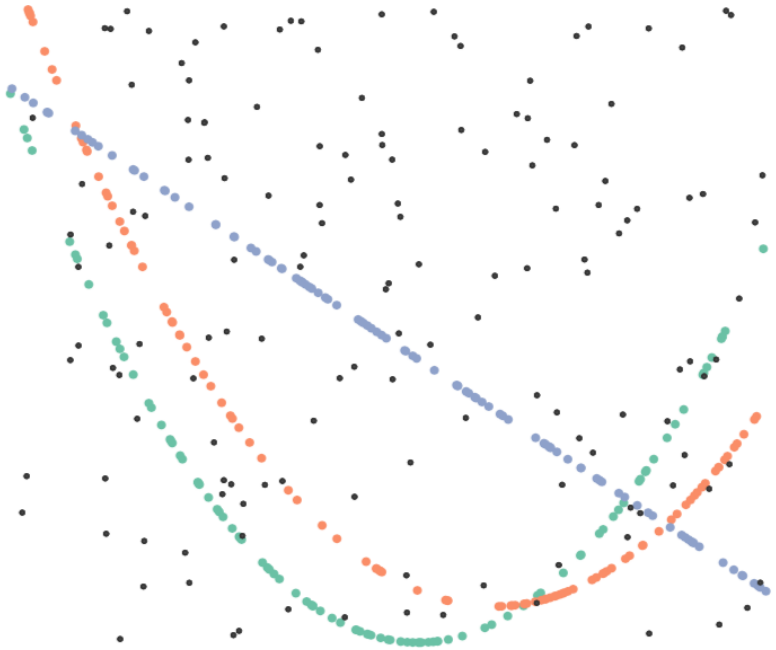
Model Fitting

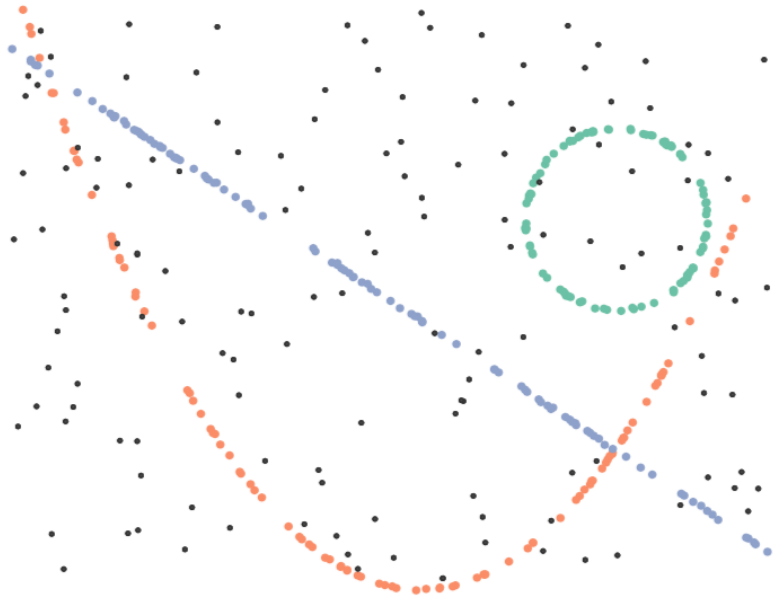


Least squares



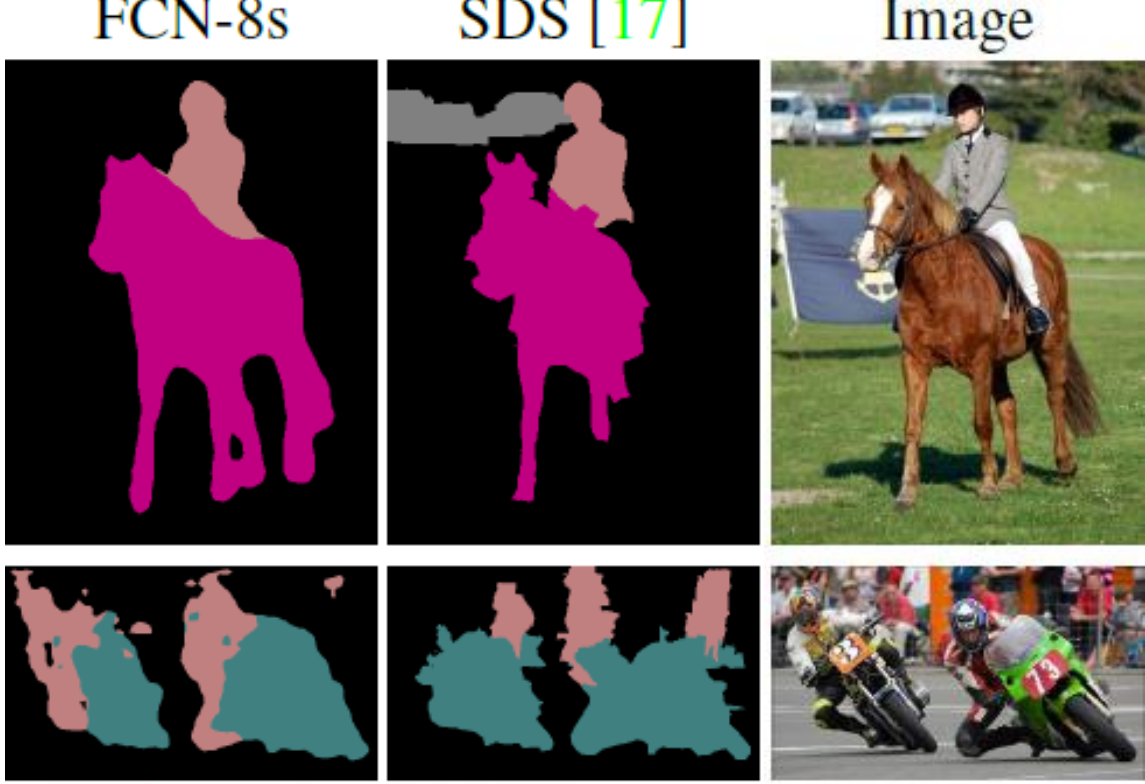
Robust Fit
(RANSAC)

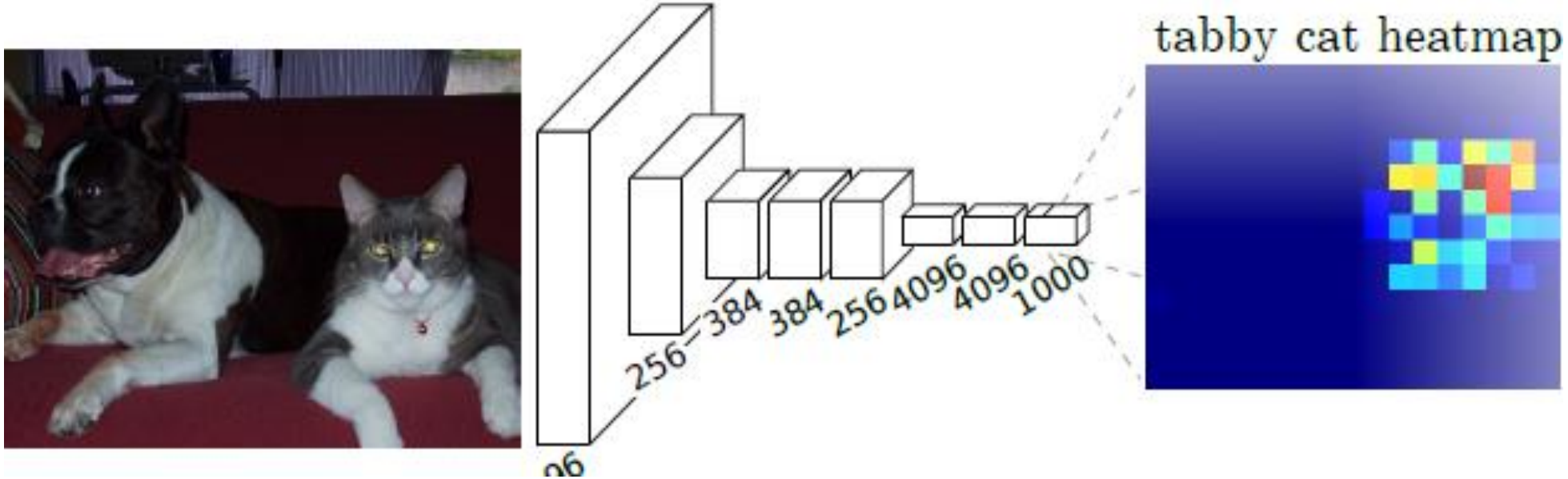




Model Fitting

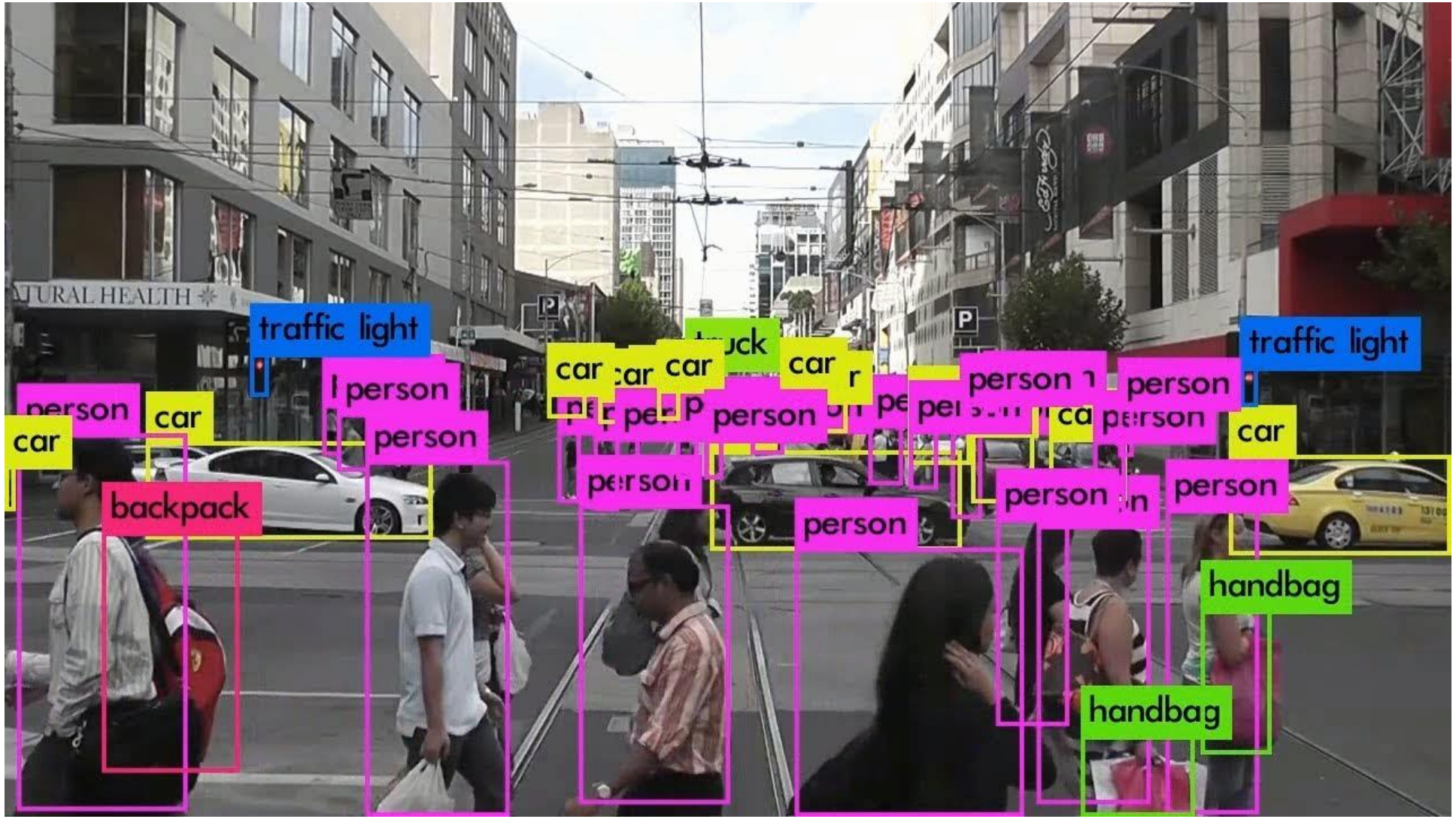
Segmentation





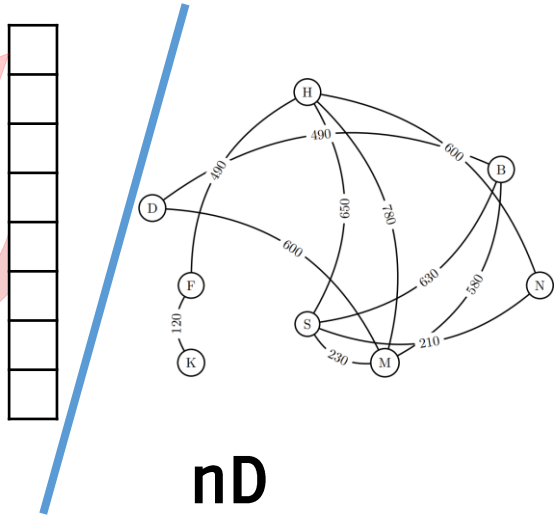
Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." CVPR 2015

Object Detection



Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

Machine Learning
Pattern Analysis



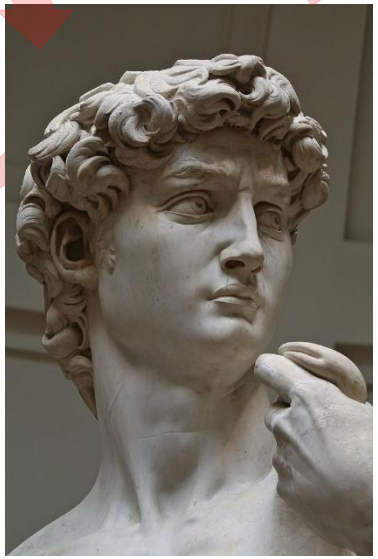
nD

Image Analysis

Shape Analysis

Computer Graphics

Image
Processing



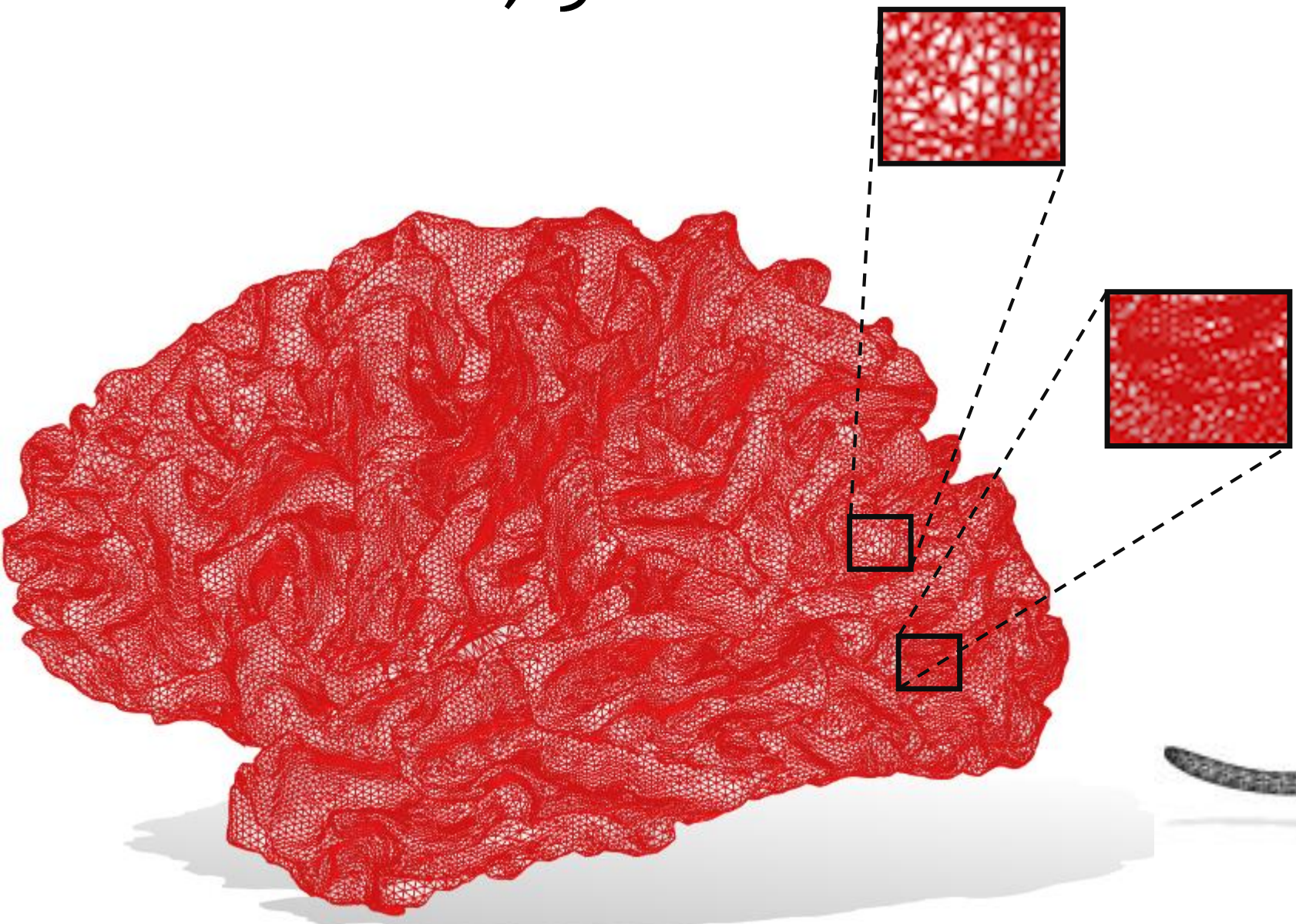
2D

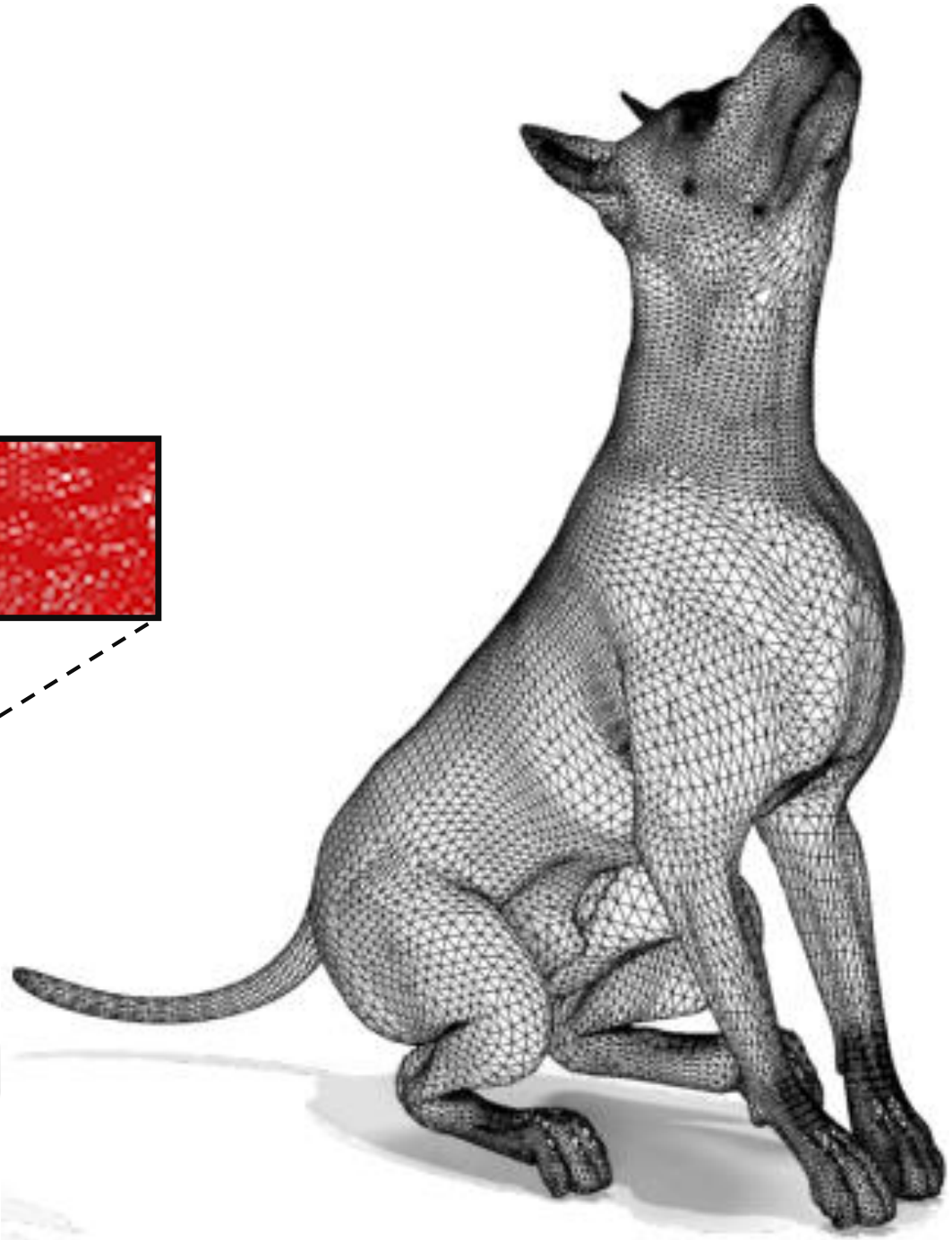
3D

Geomery
Processing

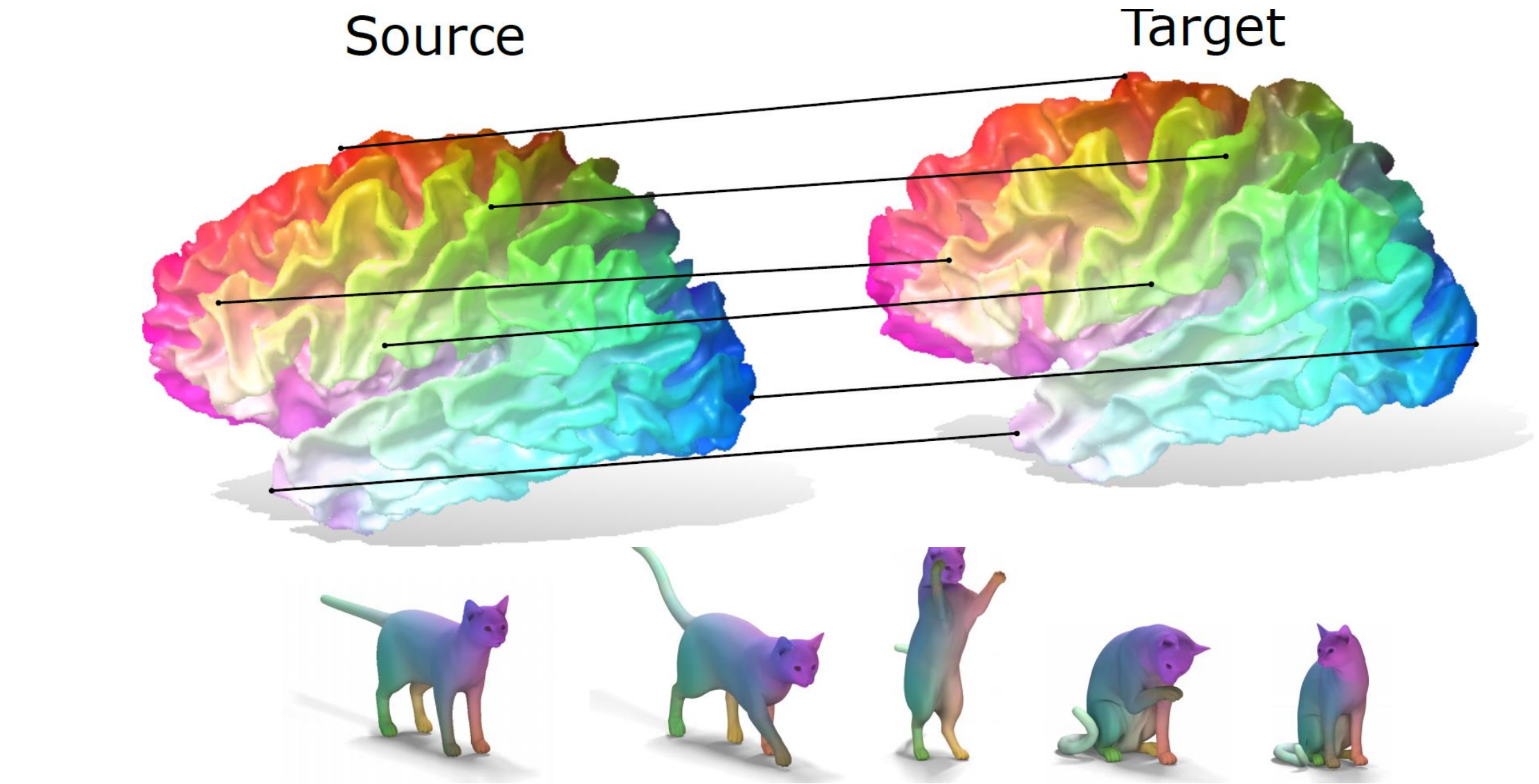
Computer Vision

Point Clouds, 3D data





3D Data Registration



Ovsjanikov, M., Ben-Chen, M., Solomon, J., Butscher, A., & Guibas, L. (2012). Functional maps: a flexible representation of maps between shapes. *ACM Transactions on Graphics (TOG)*, *31*(4), 1-11.

Course Schedule

Important Disclaimer:
this is the first edition of this course

Schedule might change
Some parts might not be sufficiently covered
Materials might require fine tuning
…

Course Schedule

Computer Vision

- Image Formation Model
- Single View Geometry
- Multi-View Geometry

Image Processing / Analisys

- Filtering
- Edges/Corner/Feature Detection
- Image Restoration / Inverse Problems

Pattern Recognition

- Model Fitting and Segmentation (Clustering)
- Face Detection / Identification

Geometry Processing

- Point Clouds
- Registration

Deep Learning

- Object Detection
- Semantic Segmentation
- Image Restoration

Guest Seminars

- Advanced Model Fitting
- Manifold Registration

Course Schedule

Feb-March

Computer Vision

- Image Formation Model
- Single View Geometry
- Multi-View Geometry

March

Image Processing / Analisys

- Filtering
- Edges/Corner/Feature Detection
- Image Restoration / Inverse Problems

April/May

Pattern Recognition

- Model Fitting and Segmentation (Clustering)
- Face Detection / Identification

May

Geometry Processing

- Point Clouds
- Registration

Deep Learning

- Object Detection
- Semantic Segmentation
- Image Restoration

Guest Seminars

- Advanced Model Fitting
- Manifold Registration

Course Schedule

Computer Vision

- Image Formation Model
- Single View Geometry
- Multi-View Geometry

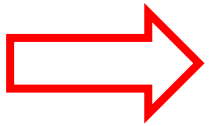
Feb-March

Image Processing / Analisys

- Filtering
- Edges/Corner/Feature Detection
- Image Restoration / Inverse Problems

Pattern Recognition

- Model Fitting and Segmentation (Clustering)
- Face Detection / Identification



First Homework:
25% + bonus
Assigned around mid March,
due after 3 weeks

Course Schedule

Computer Vision

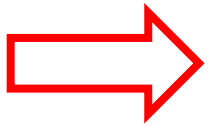
- Image Formation Model
- Single View Geometry
- Multi-View Geometry

Image Processing / Analisys

- Filtering
- Edges/Corner/Feature Detection
- Image Restoration / Inverse Problems

Pattern Recognition

- Model Fitting and Segmentation (Clustering)
- Face Detection / Identification



Second Homework:
25% + bonus
Assigned around end of April
March, due at the end of May

Course Schedule

Computer Vision

- Image Formation Model
- Single View Geometry
- Multi-View Geometry

Image Processing / Analisys

- Filtering
- Edges/Corner/Feature Detection
- Image Restoration / Inverse Problems

Pattern Recognition

- Model Fitting and Segmentation (Clustering)
- Face Detection / Identification

Geometry Processing

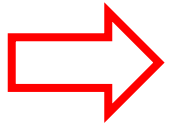
- Point Clouds
- Registration

Deep Learning

- Object Detection
- Semantic Segmentation
- Image Restoration

Guest Seminars

- Advanced Model Fitting
- Manifold Registration



Final exam: 50%

Course Organization

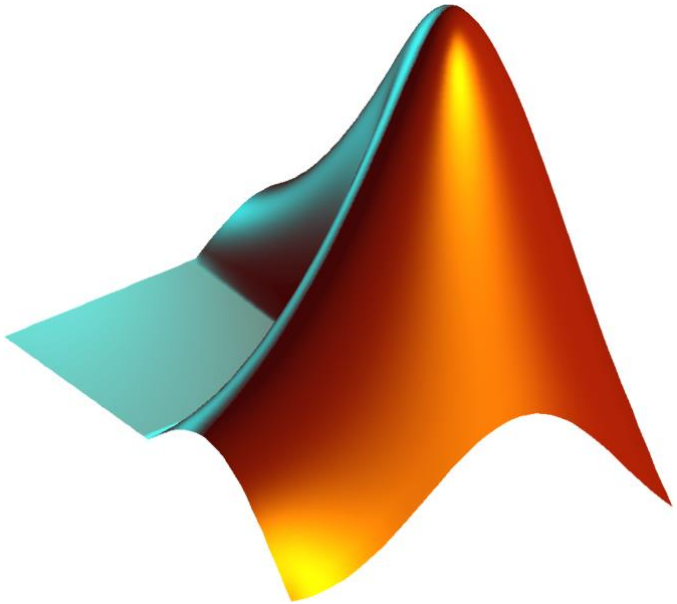
Lectures

Lectures will be on Tuesday and Fridays 15.30 – 17.15, with a short break in between.

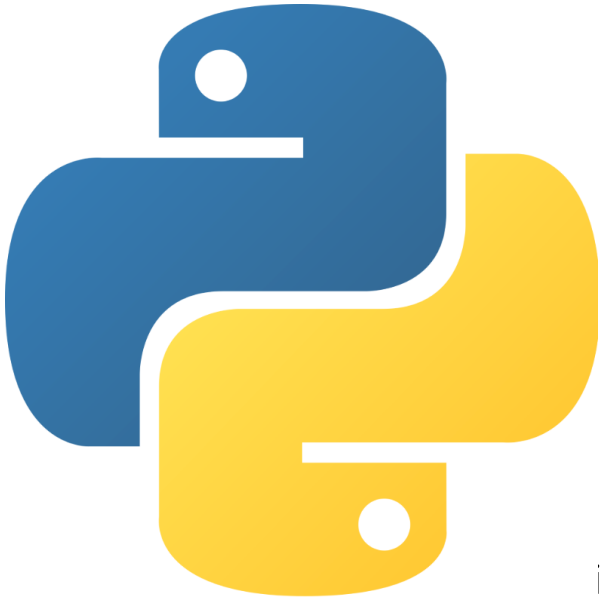
Breaks are meant for everybody to recover, and promote useful interaction among us

During Lectures I will give some illustrative example through coding

Bad news: I am a Matlab user



Good news: you are allowed to work (and prepare homework) in Python as well



Teaching Materials

Slides will be provided before lecture starts in *draft mode*.

Slides & Codes will be uploaded in a final (fixed) version after lectures

The reference book for Computer Vision part is:

[**HZ**] Hartley, Richard, and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003 (second edition)

The reference book for Image Processing part is:

[**GW**] Gonzalez, Rafael C., Richard E. Woods. *Digital image processing*. Prentice Hall, 2007 (third edition)

… papers will be also provided as references

Labs

Three lab sessions have been planned so far.

These are meant to recap the lecture materials and make some direct practice under the supervision of Pietro Verzelli.

During Labs a laptop with Python or Matlab installed is required

Requirements and Tools

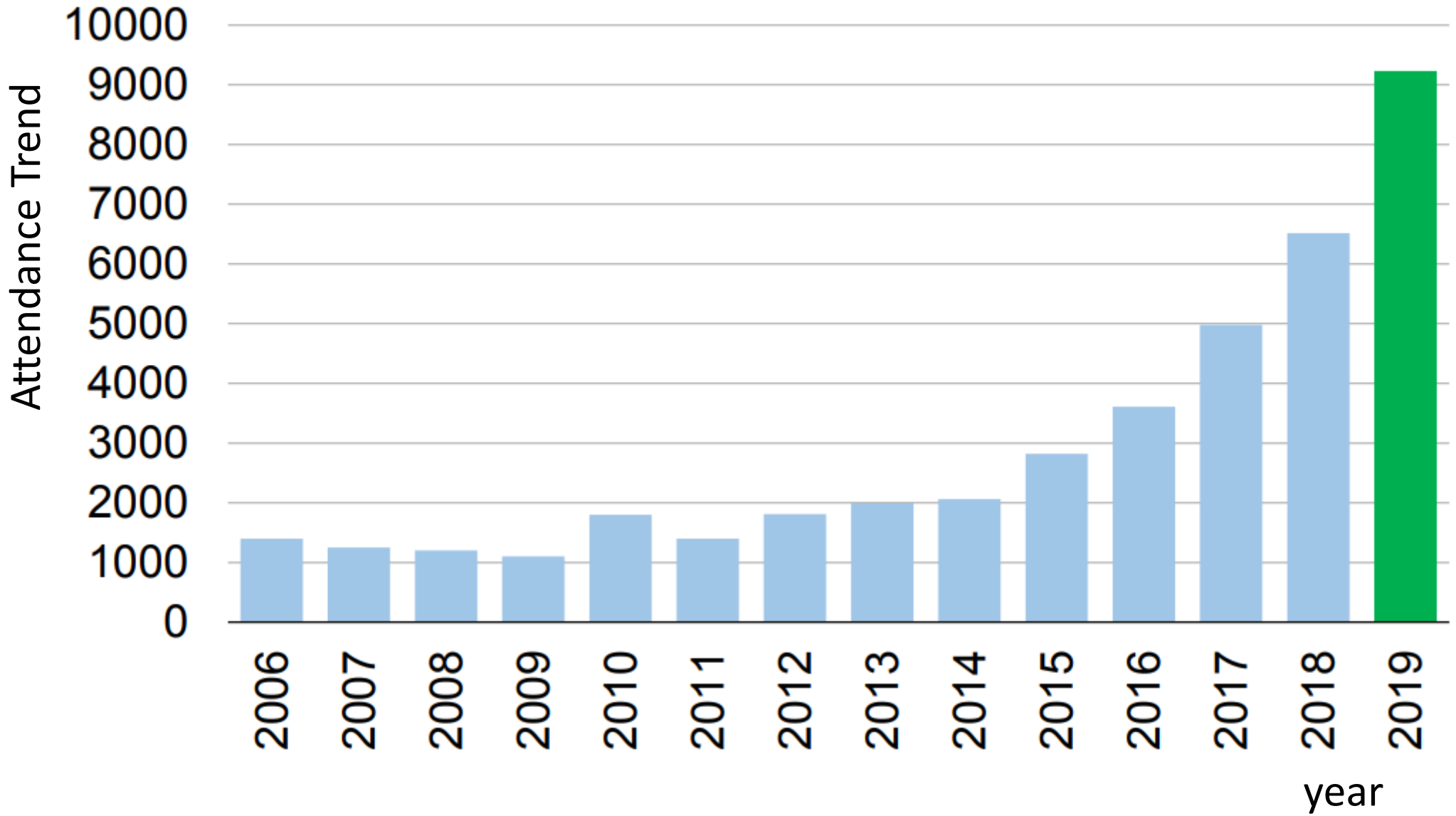
Giacomo Boracchi

Course Requirement and Tools

- Linear algebra (including solution of linear systems, basis expansions…)
- Programming Skills (Matlab or Python)
- Willingness to learn
- Being familiar with core machine learning practices and methods
    - How about Neural Networks?

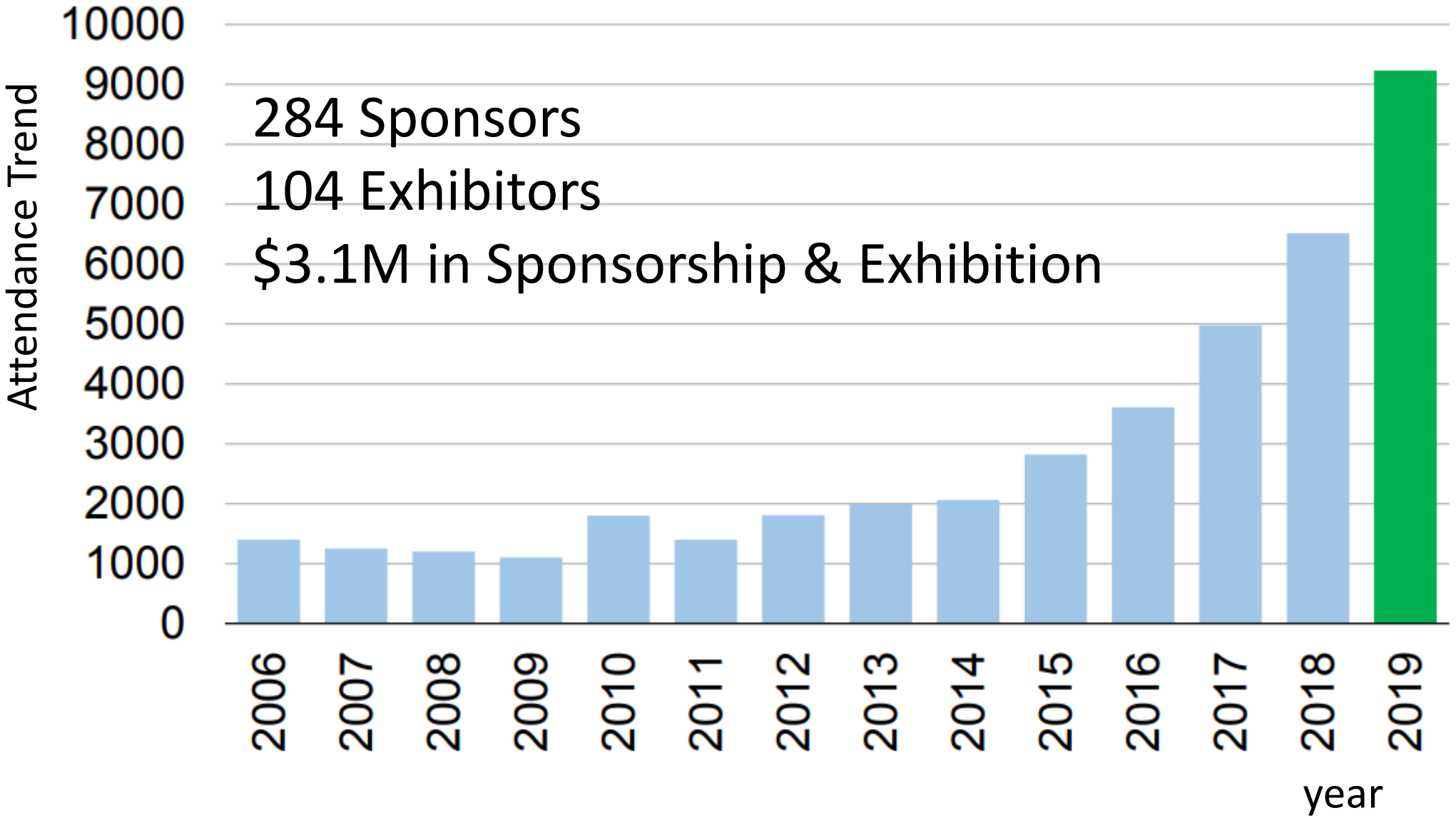
Is it worth to take this course?

CVPR



source: CVPR 2019 Welcome slides

CVPR



source: CVPR 2019 Welcome slides

Lately, connection with ML

There has seen a dramatic change in CV:

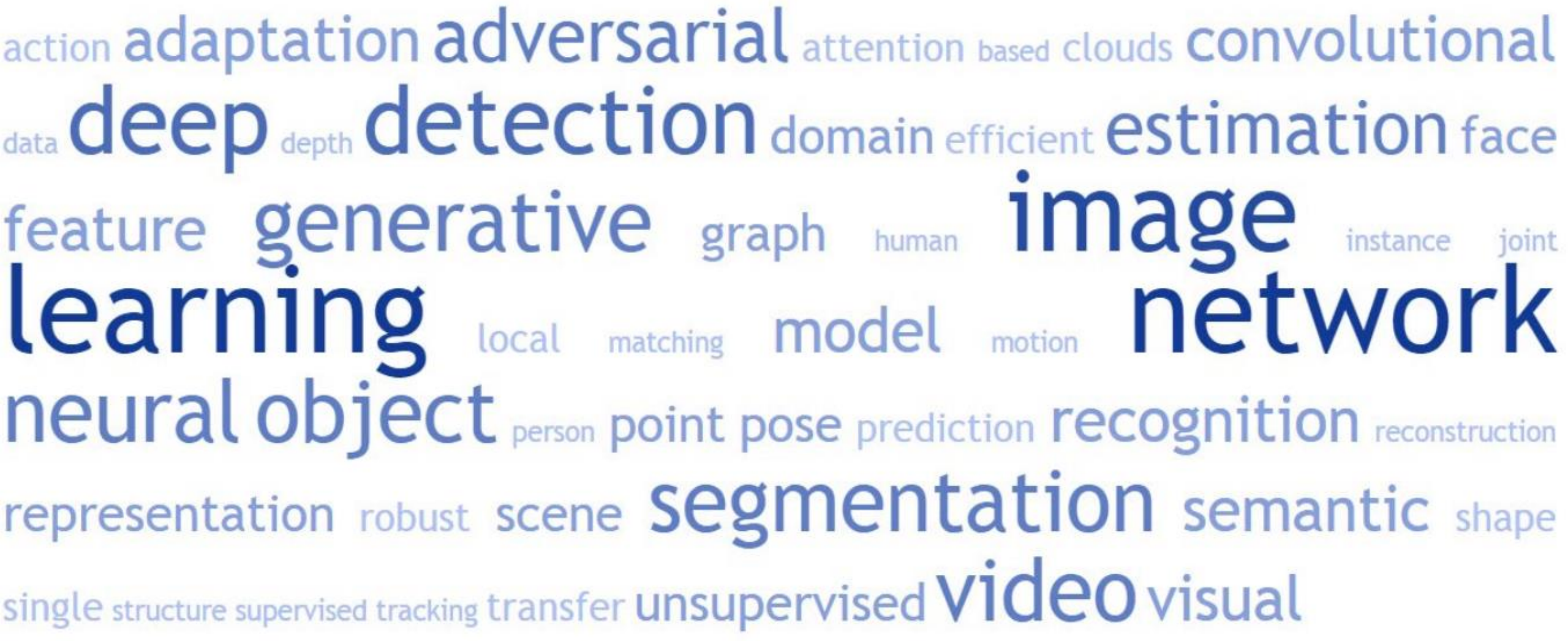
Once, most of techniques and algorithms build upon a mathematical/statistical description of images

Nowadays, machine-learning methods are much more popular

Giacomo Boracchi

1,294 papers in CVPR'19 (25.2% acceptance rate)

Acceptance rate is roughly even across topics

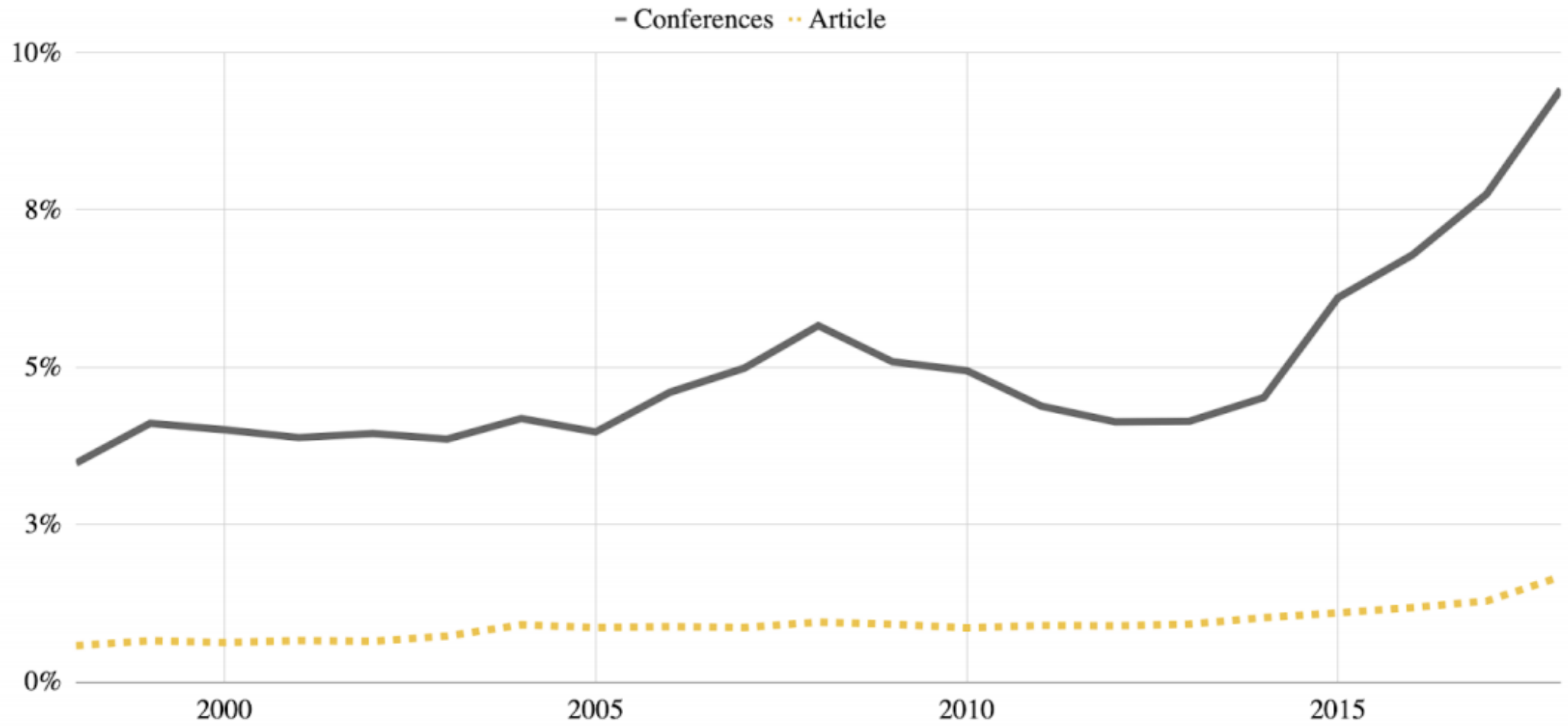


All in all…

If you plan pursuing a research-oriented career:

- If you go towards AI and ML the two fields are increasingly contaminated
- In any case a strong background in Computer Vision and Pattern Recognition is definitively a plus considering the widespread use of imaging data and the need of automation
- This course is a good way to practice more fundamental theory aspects

AI publications over all the publications



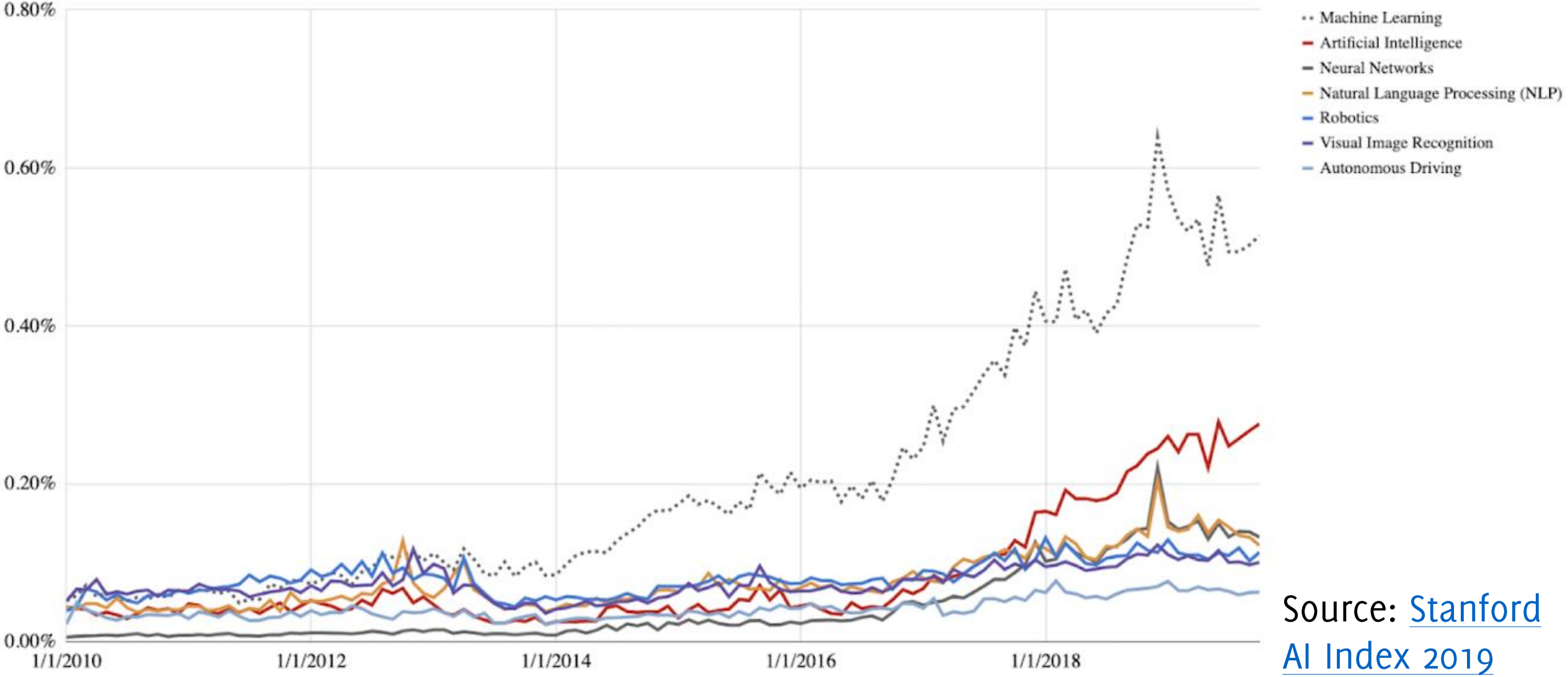
Source: Stanford AI Index 2019.

All in all...

Or if you plan pursuing a engineering / computer science job:

- Companies are increasinly interested in Computer Vision and Pattern Recogntion appliction

Percentages of AI posted job positions



Source: Stanford AI Index 2019