

Image Processing and Computer Vision

Prof. Giuseppe Lisanti
giuseppe.lisanti@unibo.it

Info

- The course is divided in two parts:
 - First part, Prof. Giuseppe Lisanti (giuseppe.lisanti@unibo.it)
 - Second Part, Prof. Samuele Salti (samuele.salti@unibo.it)
- Lectures will be:
 - Monday 11.00 -- 12.45 (2.6) with a break
 - Thursday 10.00 – 11.45 (0.5) with a break
 - If necessary, we will exploit the additional hour and start at 9.30am instead of 10am.
- For this semester, we will keep all office hours online
 - Write to us and we will schedule an appointment

Lab sessions

- The laboratory this year will follow the ***bring your own device*** modality (classroom 0.5 on Thursday)
- Tools:
 - Jupyter: web application to create and share documents that contain live code
 - OpenCV: python/c++ library containing several computer vision algorithms implementation
 - PyTorch: deep learning library developed by Facebook's AI Research lab (FAIR)
 - Google colab: hosted Jupyter notebook environment providing free access to resources such as GPUs
 - It is free but with limitations
- What you will need: Google account



<https://jupyter.org/>



<https://opencv.org>



<https://pytorch.org/>



<https://colab.research.google.com>

Overview

Prof. Giuseppe Lisanti (First Part)

- Image formation and acquisition process
- Spatial Filtering
- Edge detection
- Local Invariant Features
- Instance Detection

- OpenCV + spatial filtering/edge Laboratory
- Local Invariant Features Laboratory

Prof. Samuele Salti (Second Part)

- Camera Calibration
- Shallow Classifiers for images
- Convolutional Neural Networks
- Successful architectures

- Camera Calibration Laboratory
- Pytorch + Vanilla CNN
- Transfer learning

IPCV Exam

The exam consists in two parts:

1. a **written** exam in which the student will answer two questions on the theory presented in the whole course (max grade 24, min grade 15). An example will be available on Virtuale.
 - The grade for the written exam can be rejected by the student only once.
 - The student MUST communicate that he/she does not accept the grade 5 days after checking the errors.
 - The accepted grade is valid for three years.
 - Four exam dates:
 - 25th of June 2024
 - 17th of July 2024
 - 13th of September 2024
 - ...yet to be defined by the AI course in back office...should be in January or February 2025

IPCV Exam

2. A practical part consisting of two assignments, one for each part of the course (max grade 7).
 - The assignment should be completed in groups of 2 (maximum 3) students - you must decide the members of your group yourselves.
 - The assignments can be submitted ***only once, and it cannot be resubmitted.***
 - If the assignment is not submitted in the respective academic year, the following year it will be necessary to complete the assignment for that new academic year.
 - Cut-off dates to upload the assignments on Virtuale (related to scholarship deadlines and final exam requisite deadlines):
 - 15th of July 2024 (for scholarship deadline in august)
 - 10th of September 2024
 - 10th of November 2024
 - 7th of January 2025
 - 10th of February 2025

IPCV Exam

- One bonus point will be given if the exam is completed in the academic year of enrolment
- The exam is considered passed if the sum of the two grades is greater than or equal to 18.
- When the student has passed both parts of the exam, we will register the final grade on AlmaEsami.
- Questions?

IPCV Project Work

- Students are required to carry out and present a software project related to solving a computer vision problem.
- The student should:
 1. Send us a list of three articles of her/his interest, preferably with the available code.
 - The choice should be based on articles published in the conferences or journals discussed in the next slide
 - We will analyze these three articles and propose one from the three.
 2. Thoroughly study the chosen article and its accompanying code to replicate at least one experiment mentioned in the article or perform a test by modifying the studied technique.
 3. Prepare a 20-minute presentation that describes the technique proposed in the selected article and the experiment conducted.
- During and/or at the end of the presentation, we will ask questions about the project work.

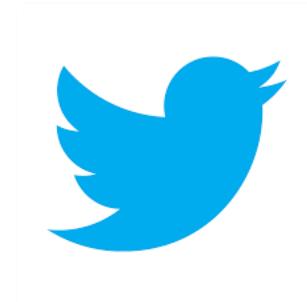
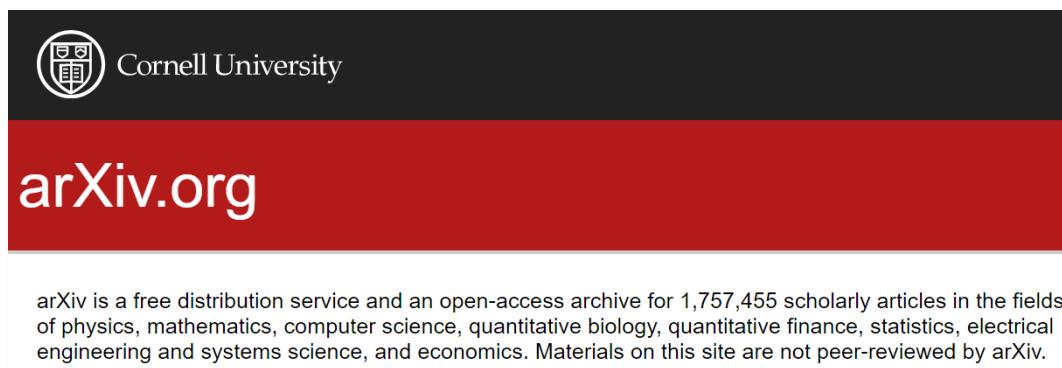
Where to look?

Conferences

- Computer Vision and Pattern Recognition (CVPR)
- International Conference on Computer Vision (ICCV)
- European Conference on Computer Vision (ECCV)
- International Conference on Learning Representations (ICLR)
- Neural Information Processing Systems (NeurIPS)
- International Conference on Machine Learning (ICML)

Journals

- IEEE Trans. On Pattern Analysis and Machine Intelligence (PAMI)
- International Journal of Computer Vision (IJCV)
- Journal of Computer Vision and Image Understanding (CVIU)



Computer Vision research is highly influential...



Pubblicazione

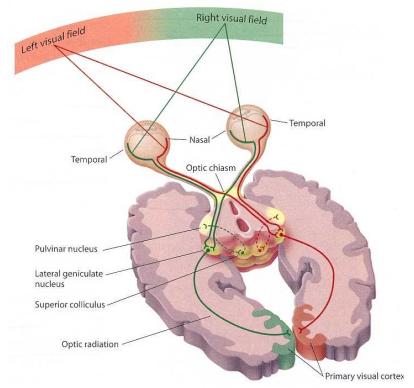
h5-index

1. Nature	467
2. The New England Journal of Medicine	439
3. Science	424
4. IEEE/CVF Conference on Computer Vision and Pattern Recognition	422
5. The Lancet	368
6. Nature Communications	349
7. Advanced Materials	326
8. Cell	316
9. Neural Information Processing Systems	309
10. International Conference on Learning Representations	303

Sponsors @CVPR

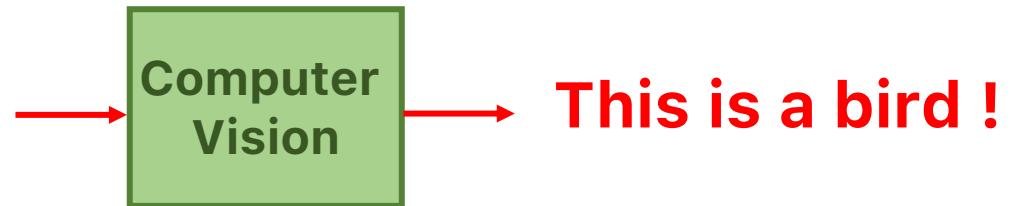


What is Computer Vision?

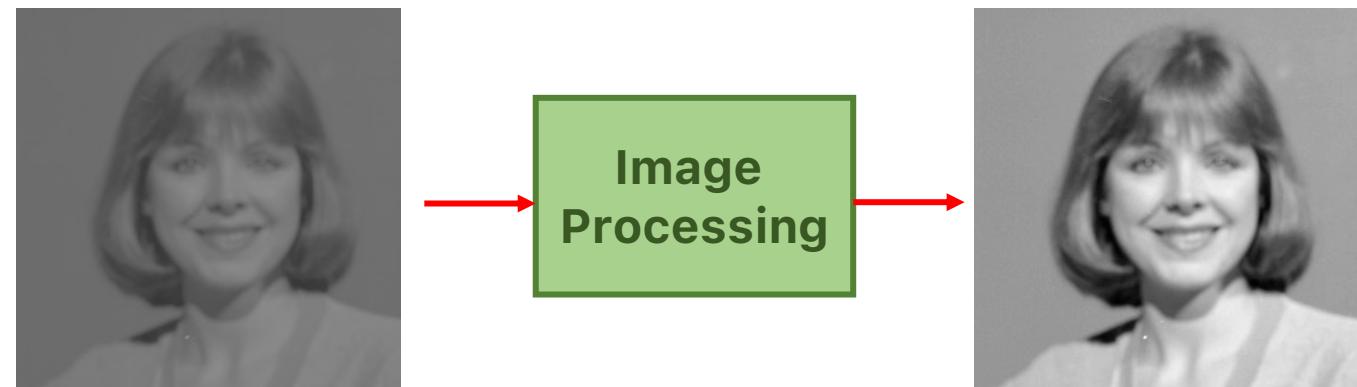


Computer Vision vs. Image Processing

- Computer Vision
deals with extraction
of information from
images



- Image Processing
aims at improving the
quality of images



Computer vision & Image Processing

Quite often Image Processing helps Computer Vision

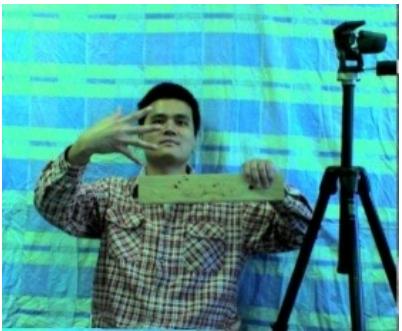


What kind of information?

- Automatic understanding of images and video
 - Computing properties of the 3D world from visual data (measurement)

What kind of information?

Real-time stereo



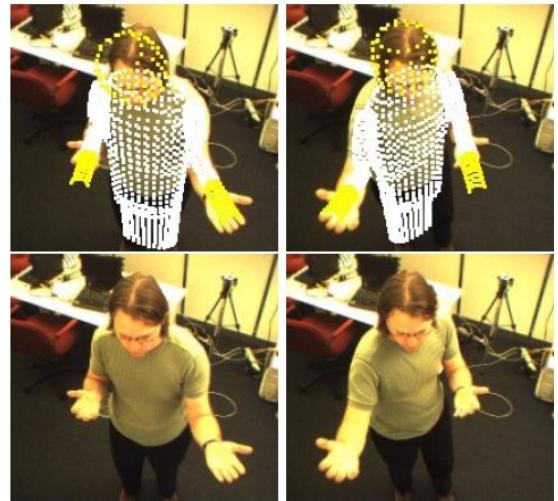
Wang et al.

Structure from motion



Snavely et al.

Tracking



Demirdjian et al.

What kind of information?

- Automatic understanding of images and video
 - Computing properties of the 3D world from visual data (measurement)
 - Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities (perception and interpretation)

What kind of information?



Objects
Activities
Scenes
Locations
Text / writing
Faces
Gestures
Motions
Emotions...

What kind of information?



Objects
Activities
Scenes
Locations
Text / writing
Faces
Gestures
Motions
Emotions...

What kind of information?



Objects
Activities
Scenes
Locations
Text / writing
Faces
Gestures
Motions
Emotions...

What kind of information?



Objects
Activities
Scenes
Locations
Text / writing
Faces
Gestures
Motions
Emotions...

What kind of information?

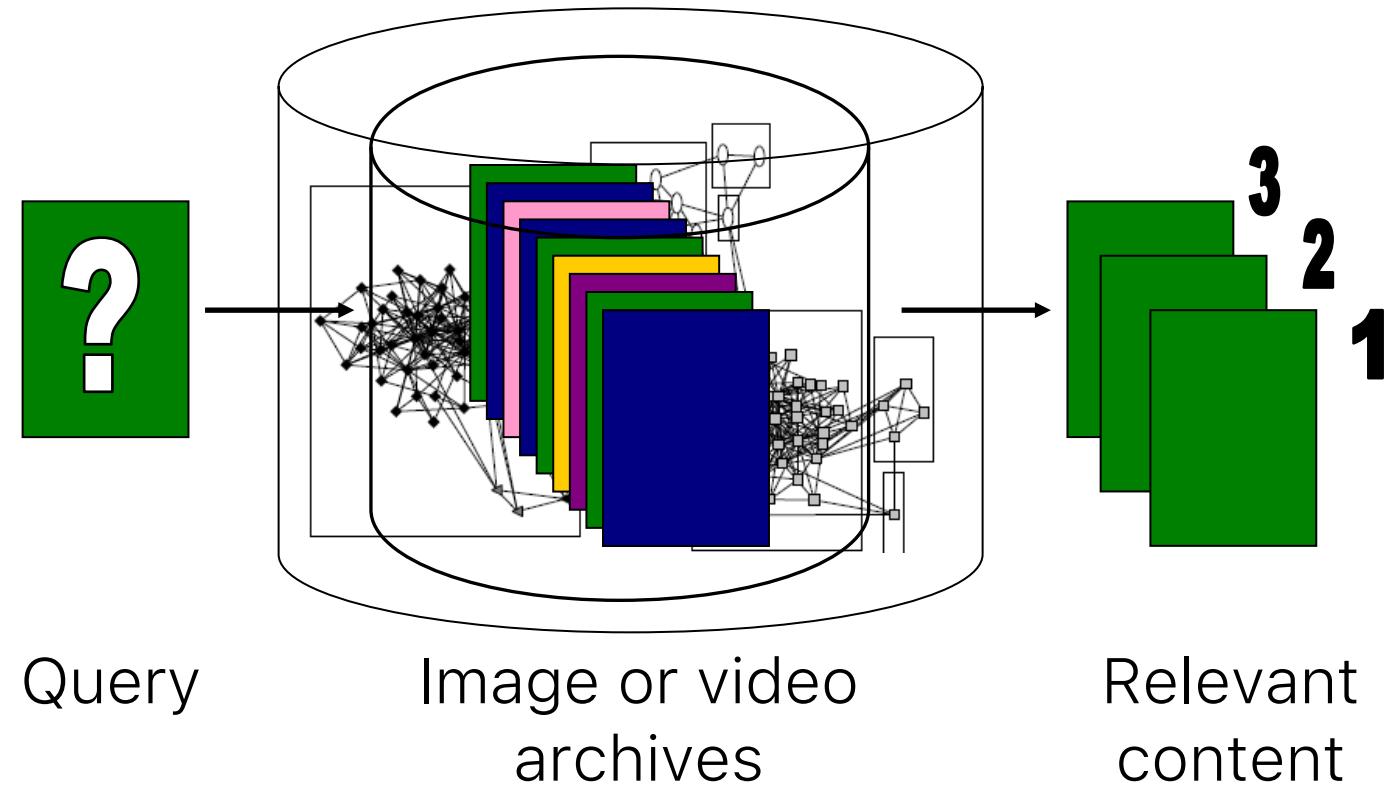


Objects
Activities
Scenes
Locations
Text / writing
Faces
Gestures
Motions
Emotions...

What kind of information?

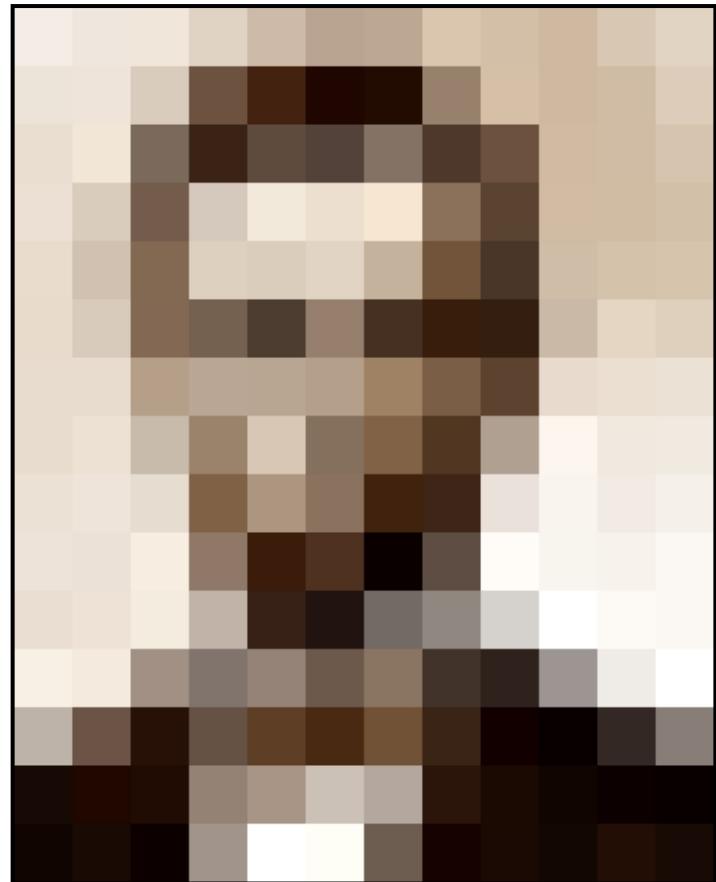
- Automatic understanding of images and video
 - Computing properties of the 3D world from visual data (measurement)
 - Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities. (perception and interpretation)
 - Algorithms to mine, search, and interact with visual data (*search and organization*)

What kind of information?



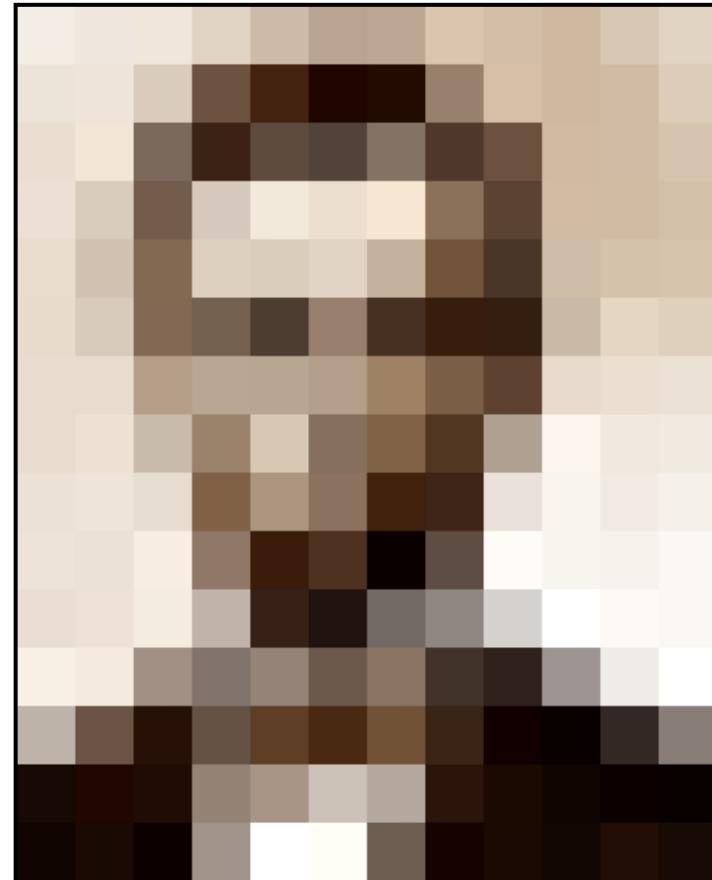
Computer Vision Challenges

- What the computer sees?



Computer Vision Challenges

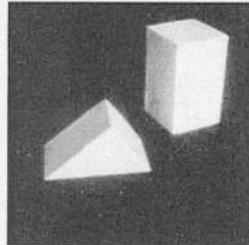
- What the computer sees?
- Loss of depth
- Varying scale
- Varying instances
- Varying illumination
- Varying shape (articulated objects)
- Occlusions
- ...



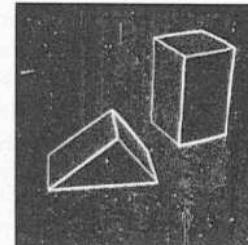
Quite a successful journey !



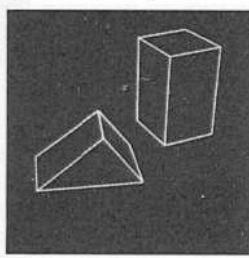
a)



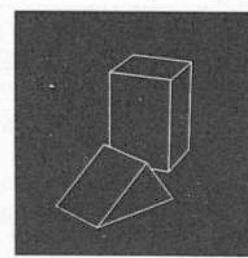
b)



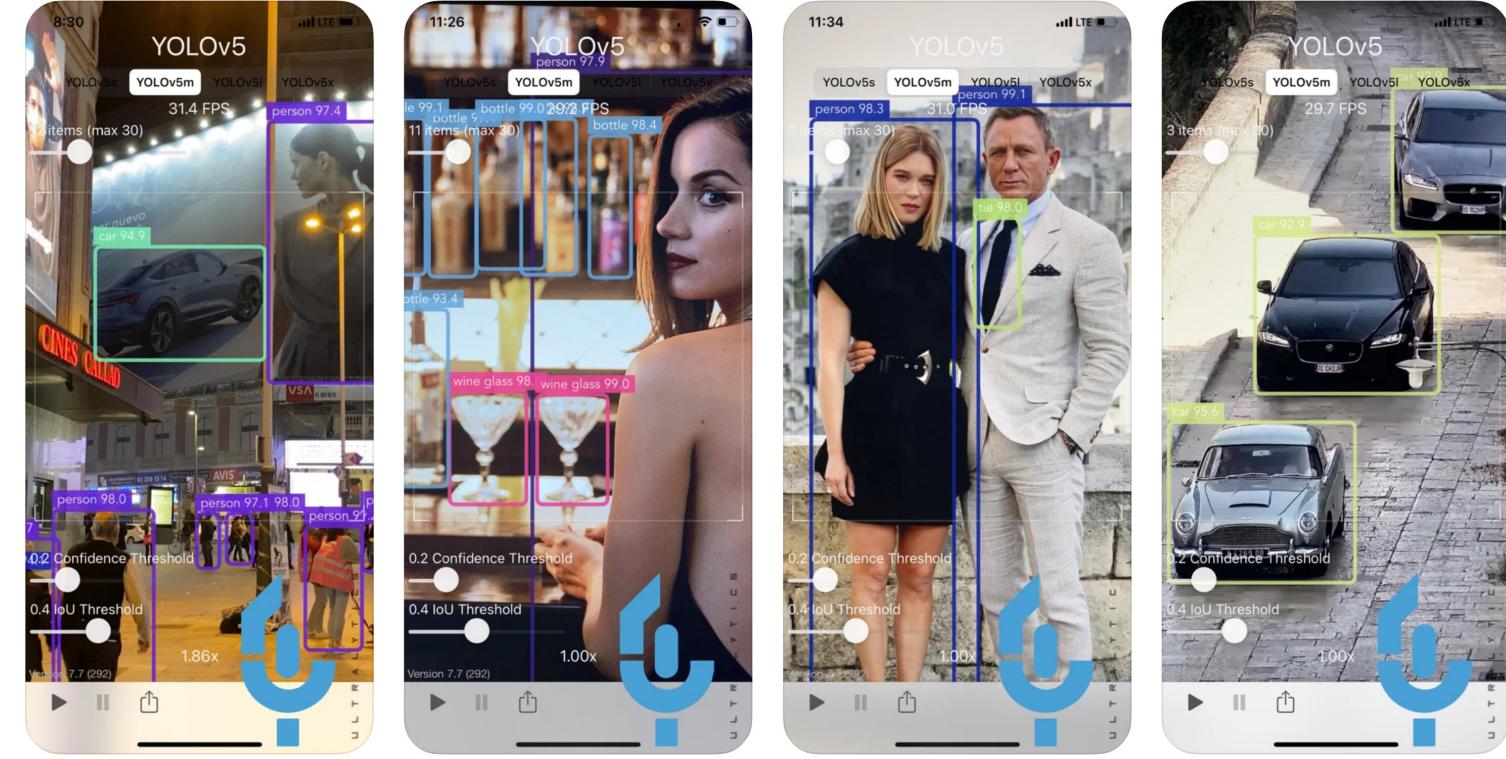
c)



d)



e)

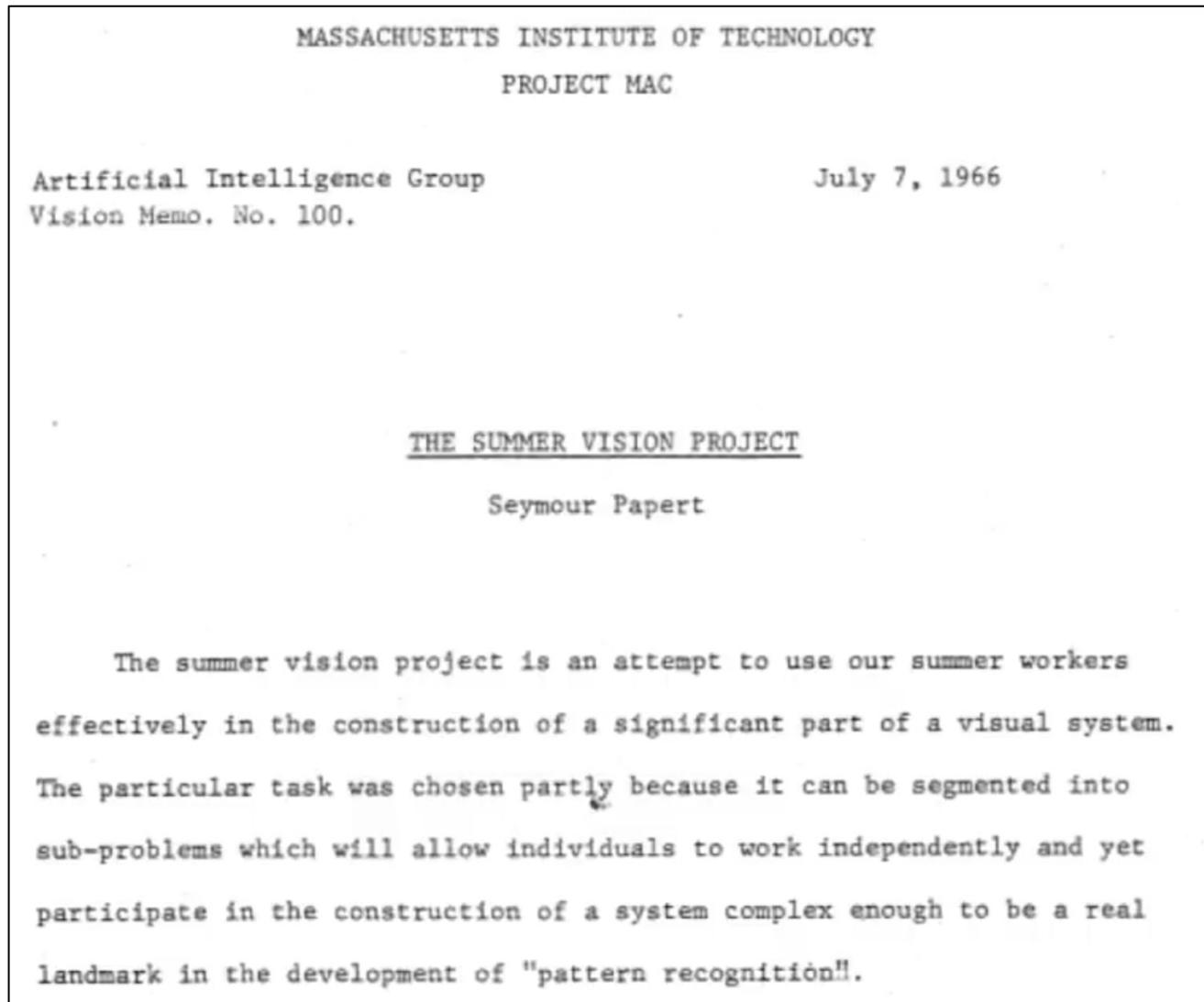


Larry Roberts, PhD thesis, Machine perception of three-dimensional solids, MIT, 1963.

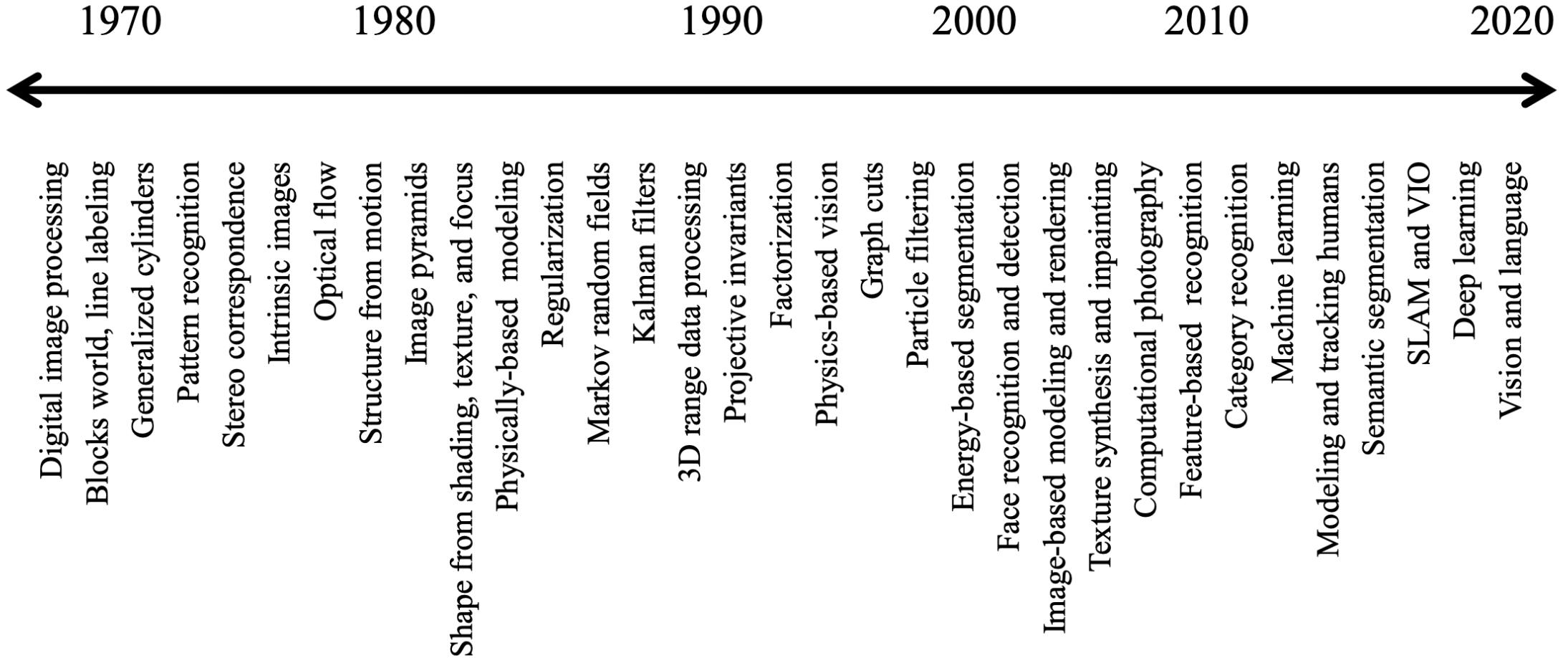
iDetection App (Ultralytics), YOLO v5 (2020)

Some History...

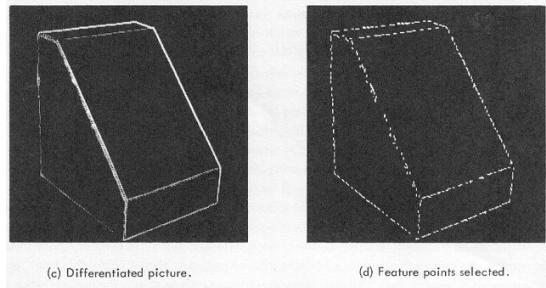
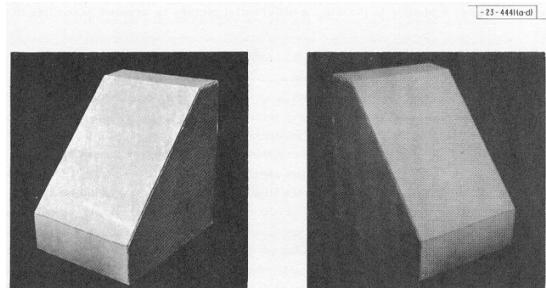
- From this “summer vision project” to...
- Visual data is very complex...
- We still did not solve the vision problem :)



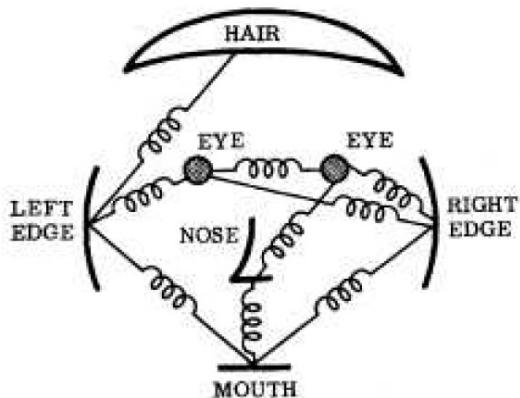
Some History...



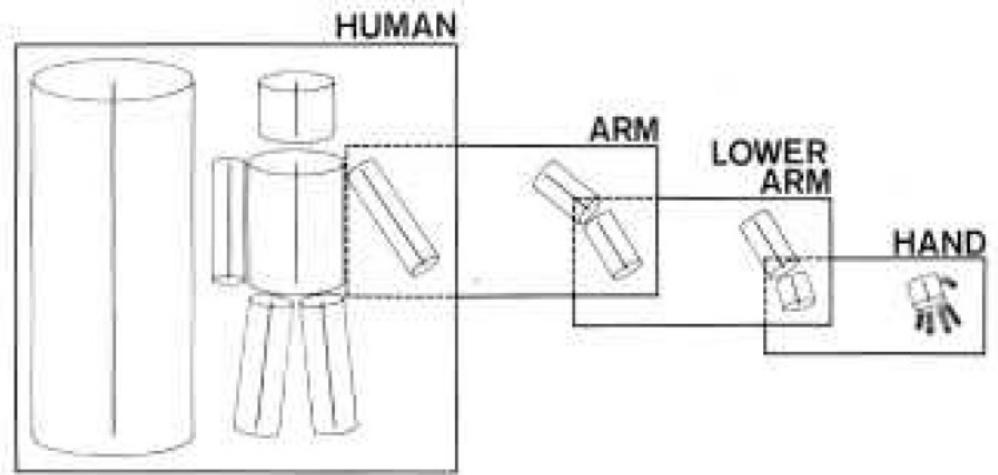
Some History...



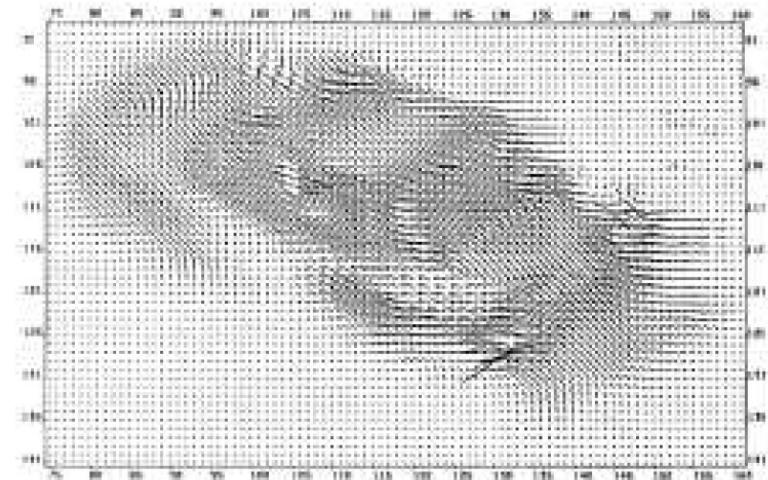
Blocks world
(Roberts 1963)



Pictorial structure
(Fischler and Elschlager 1973)

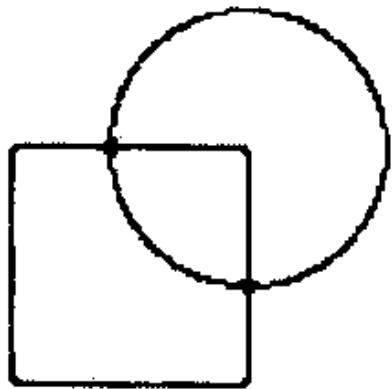


Articulated body model (Marr 1982)



Optical flow (Nagel and Enkelmann 1986)

Some History...



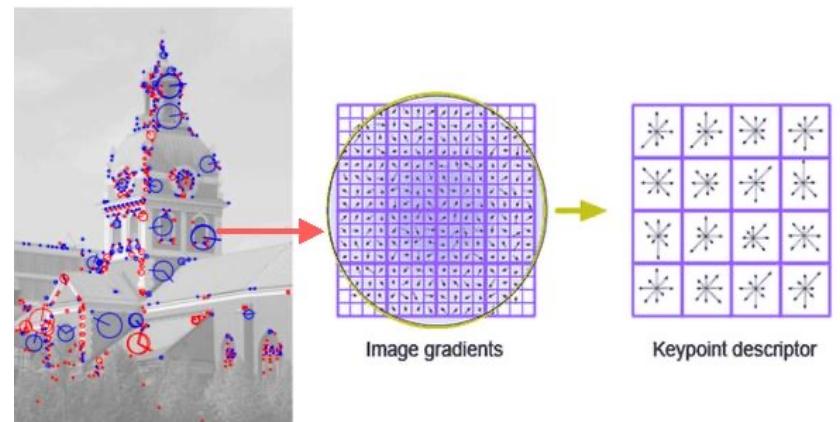
Edge detection
(Freeman and Adelson 1991)



Multi-view reconstruction
(Seitz and Dyer 1999)



Dense stereo matching
(Boykov, Veksler, and Zabih 2001)



Scale Invariant Feature
(Lowe 2001)

Panorama stitching



Key Process Technology in all Industries

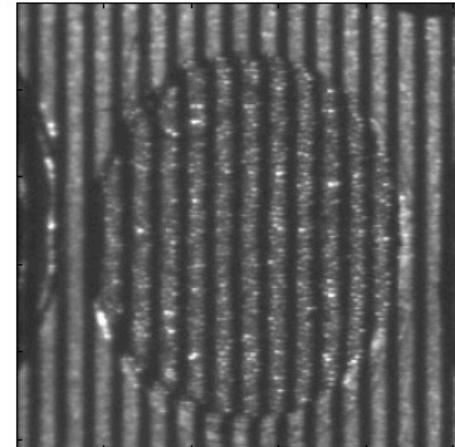


Cognex DataMan (1982)

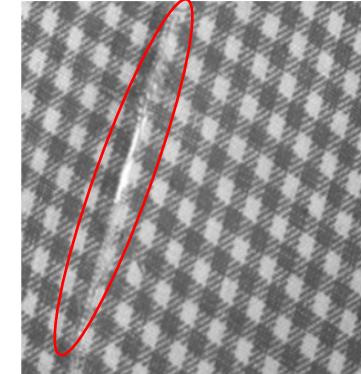
Pharma & Tobacco



Electronics



Textile



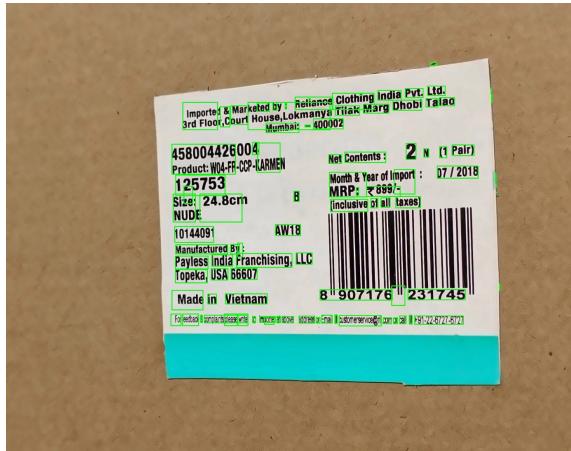
Food&Beverage



Automotive



Transportation & Logistics

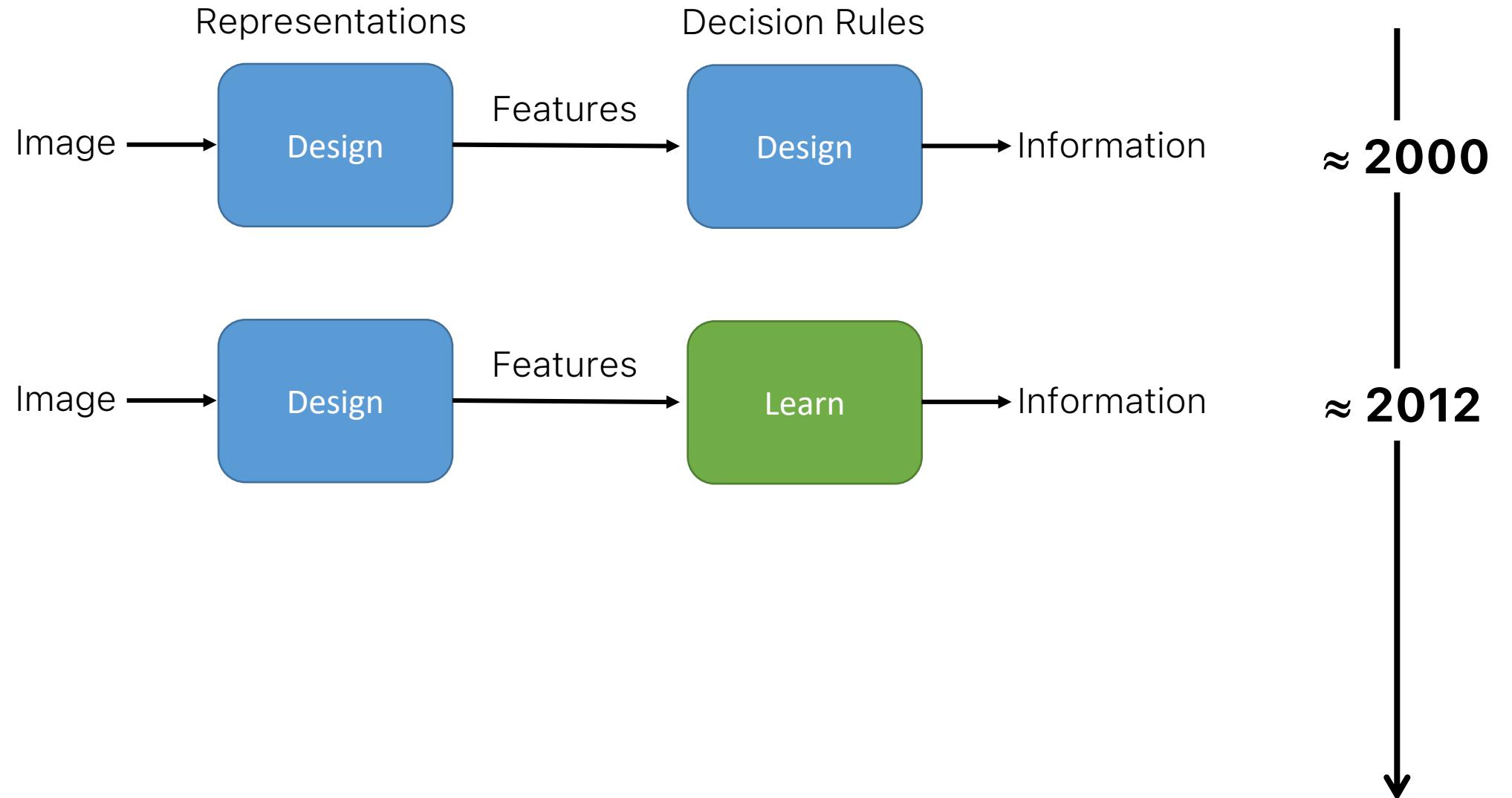


Inspection, Gauging,
Guidance, Tracing,
Compliance, Monitoring

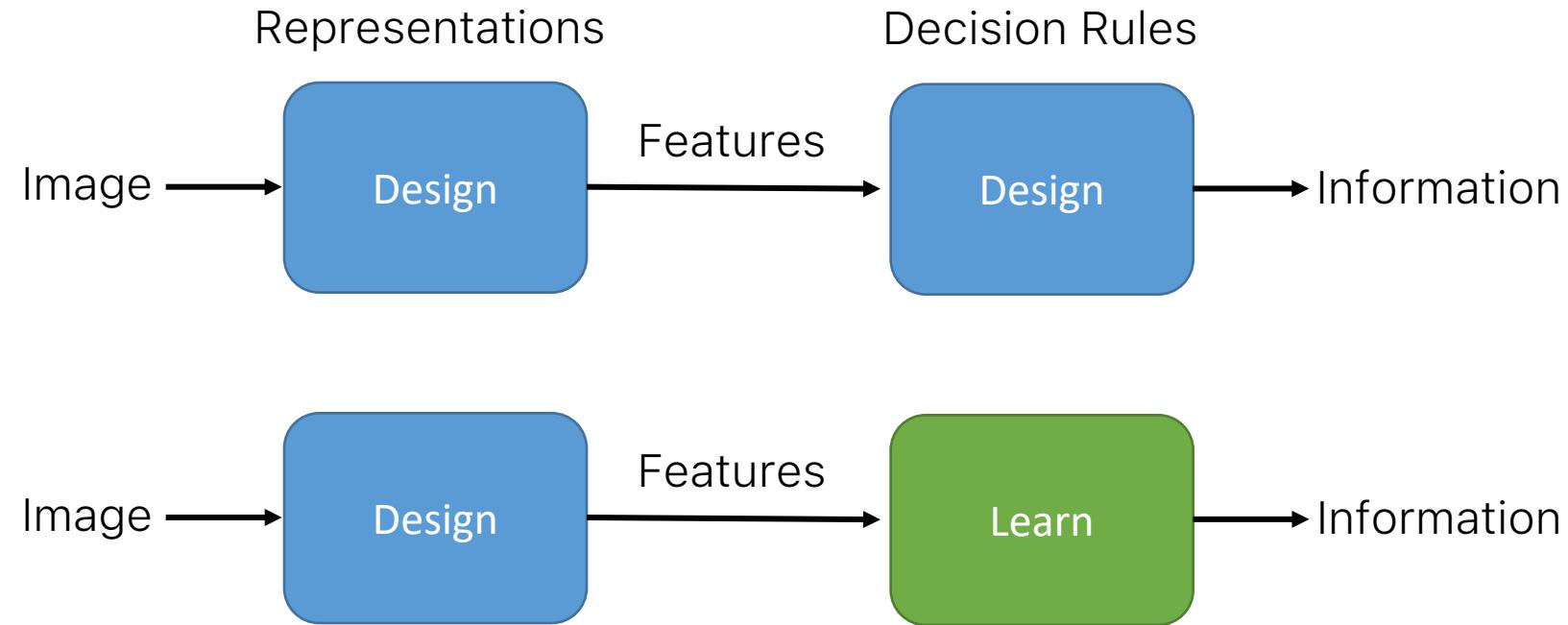


Process Optimization

Paradigm Shift



Paradigm Shift (what about data?)

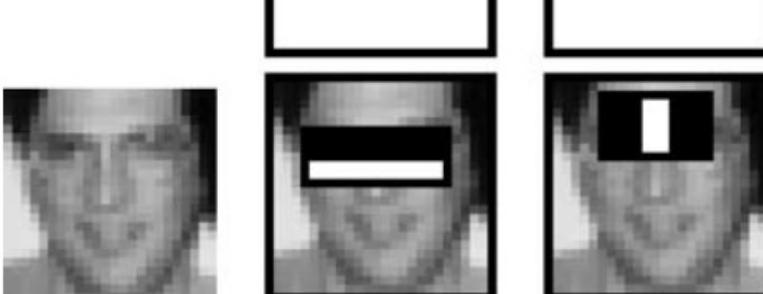


Some History...

Object → Bag of ‘words’



Bag of visual words
(Sivic and Zisserman 2003)

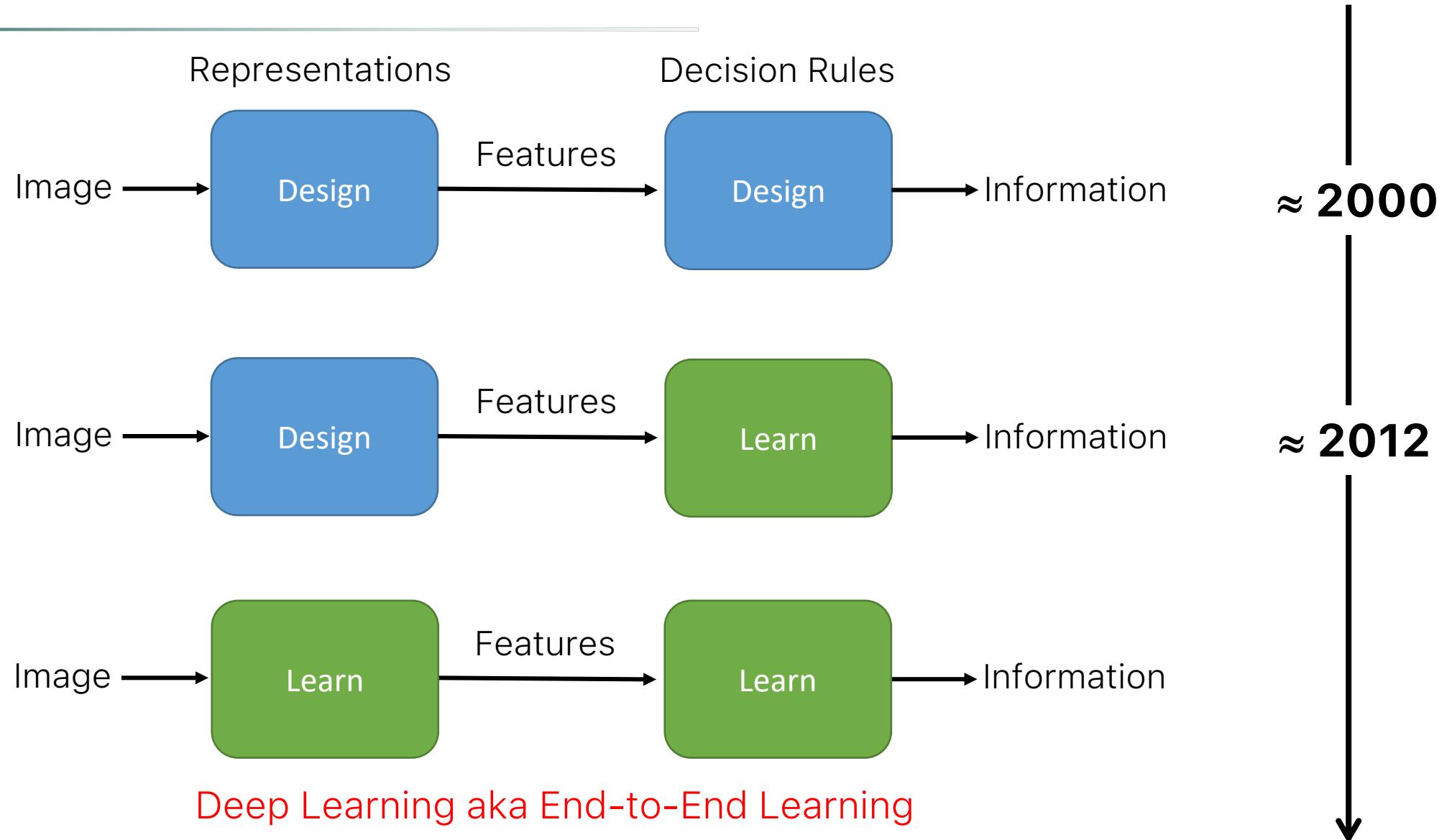


Robust object detection
(Viola and Jones 2003)

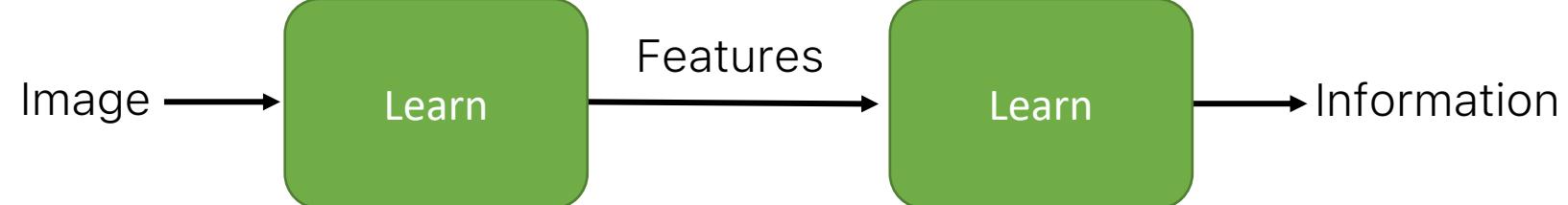
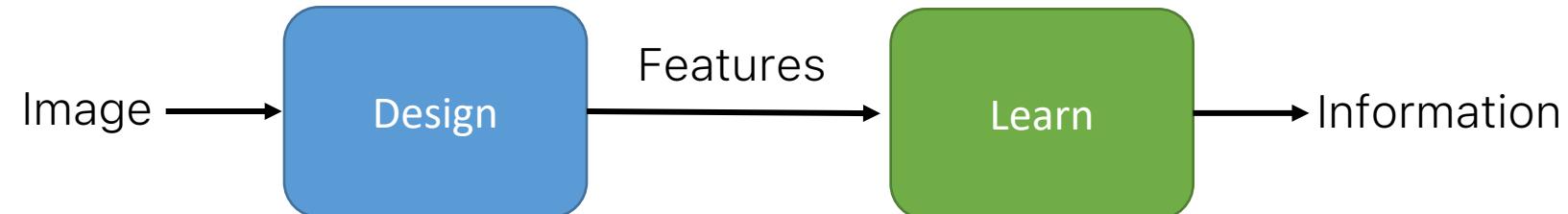
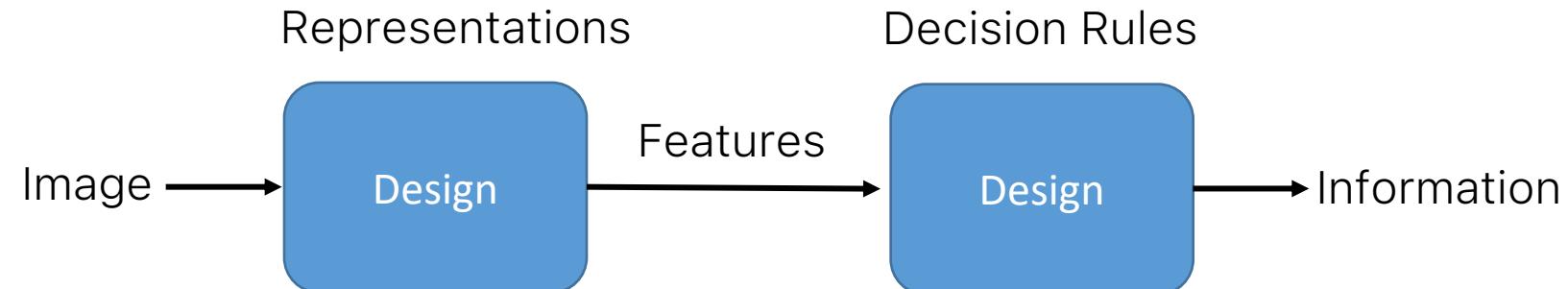


Pedestrian detection
(Dalal and Triggs 2005)

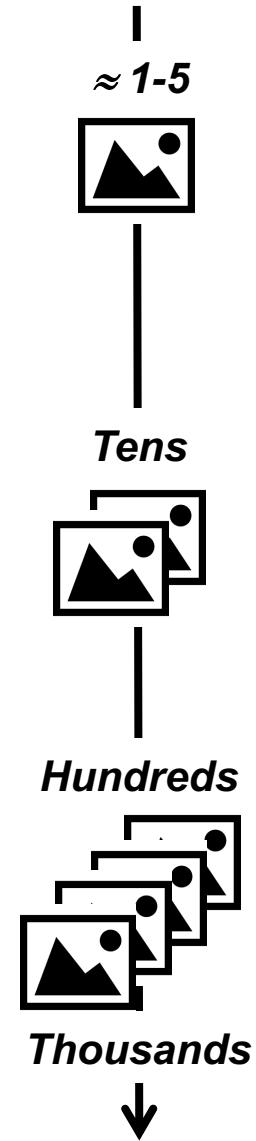
Paradigm Shift



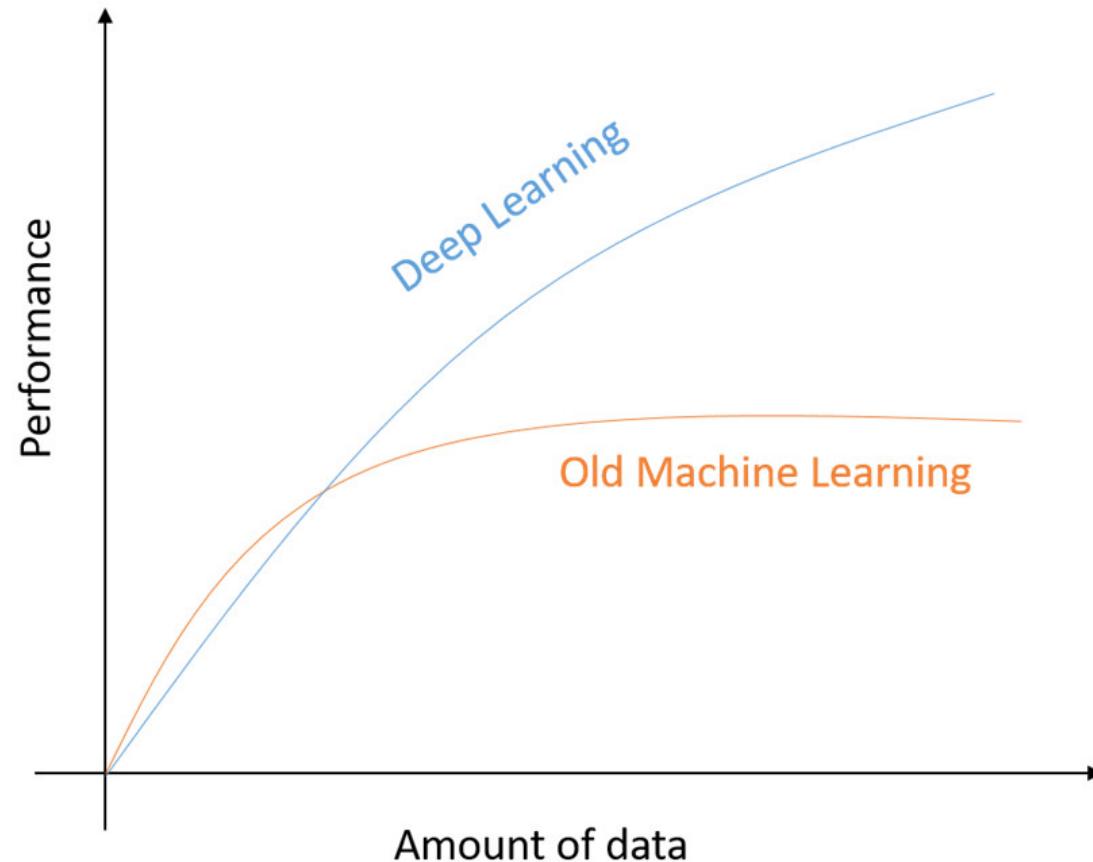
Paradigm Shift (what about data?)



Deep Learning aka End-to-End Learning

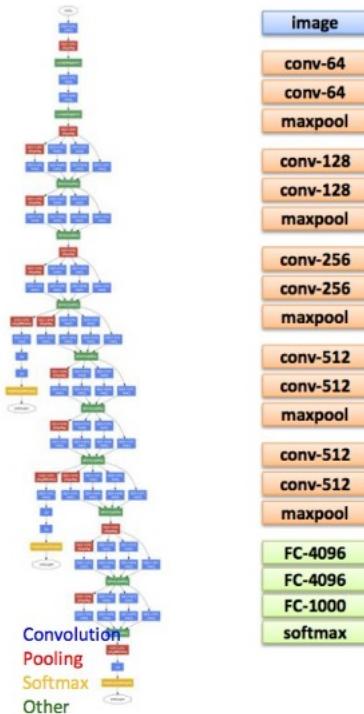


Deep & Machine Learning vs. Data

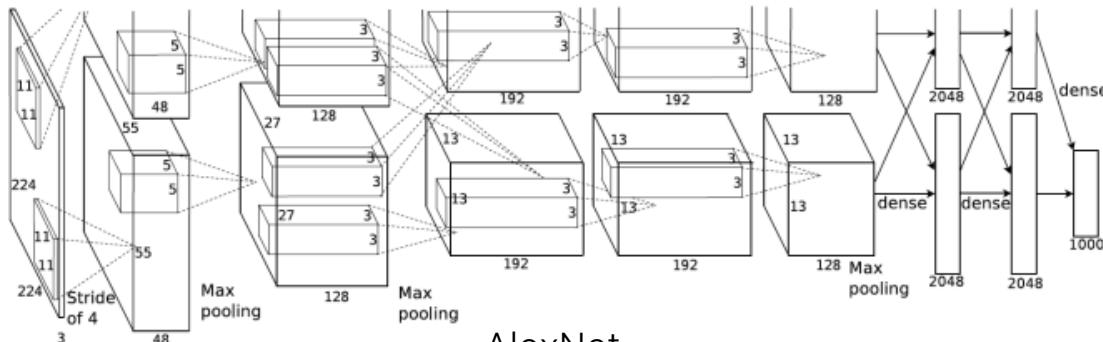


Some History...

GoogLeNet VGG



[Szegedy arxiv 2014] [Simonyan arxiv 2014]



AlexNet
(A. Krizhevsky et al 2012)

Turing award 2018



Image-Net and ILSVRC

- ImageNet
 - 22K Categories
 - 14M Images
- Image-Net Large Scale Visual Recognition Challenge
 - 1000 object classes
 - 1431167 images

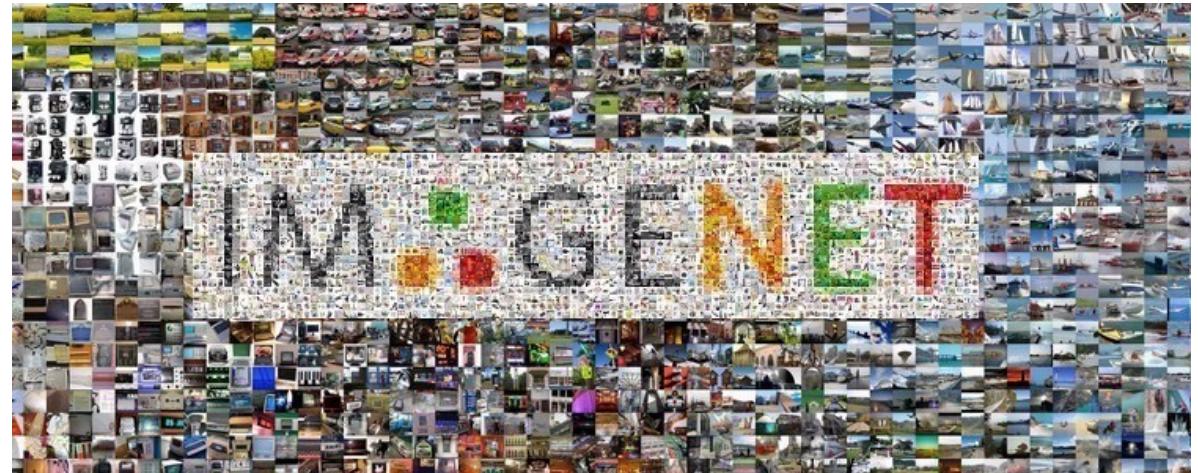
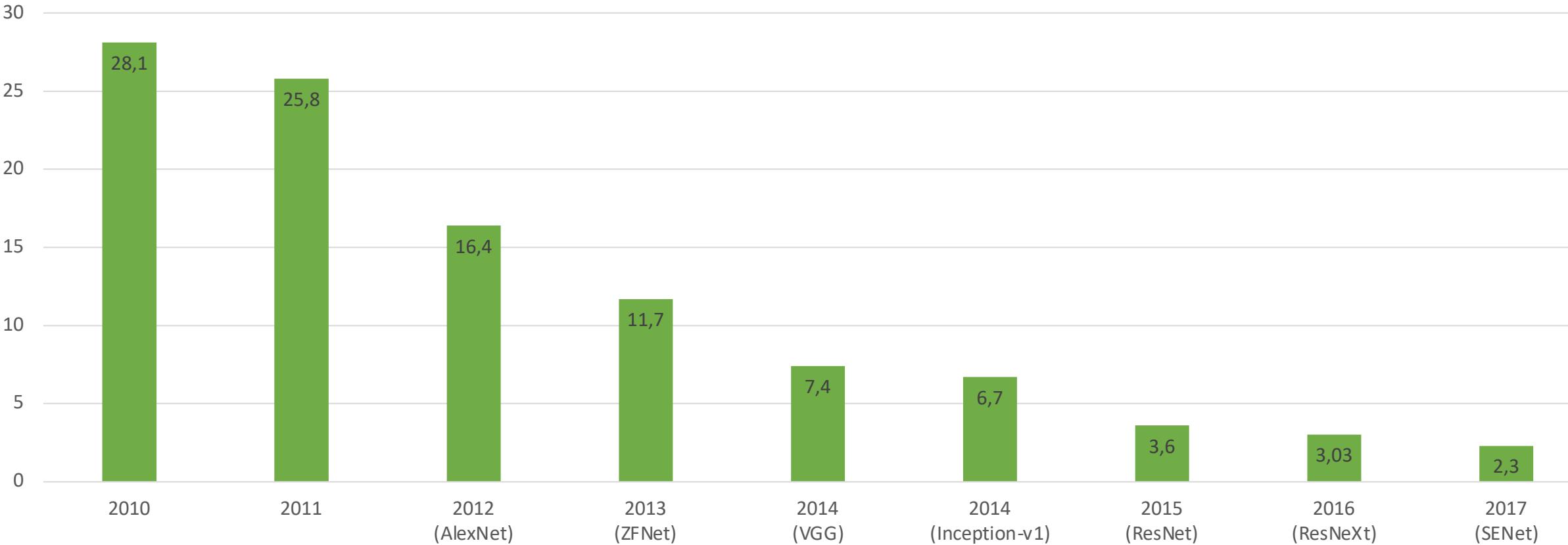


Image-Net and ILSVRC

ILSVRC Top-5 error rate

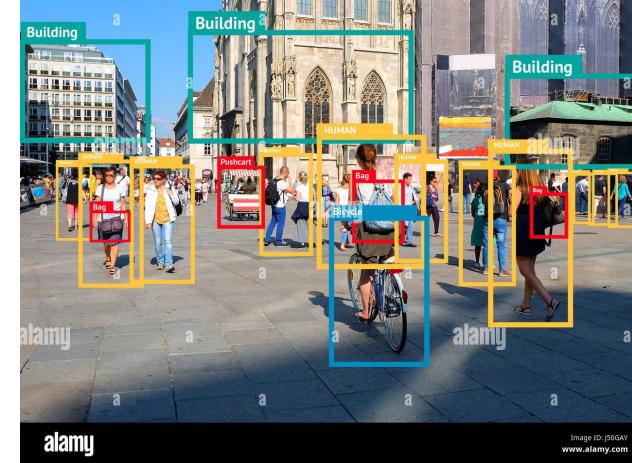


Deep Learning for Computer Vision

Image Classification



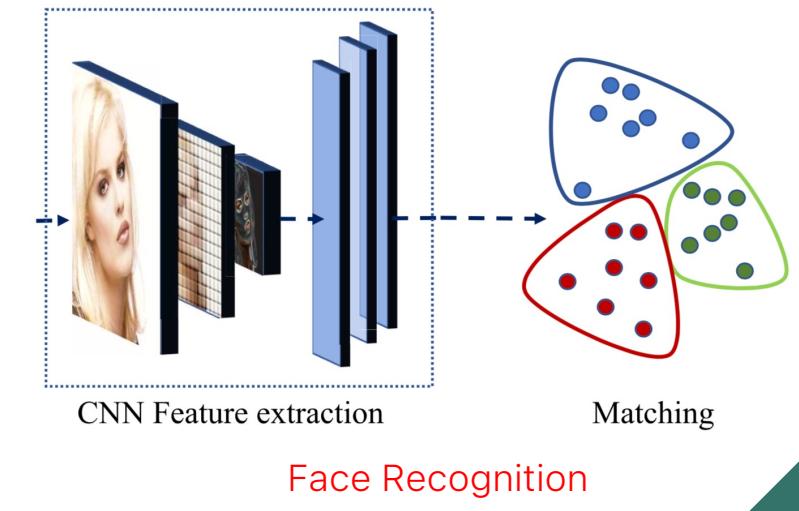
Object Detection and Instance Segmentation



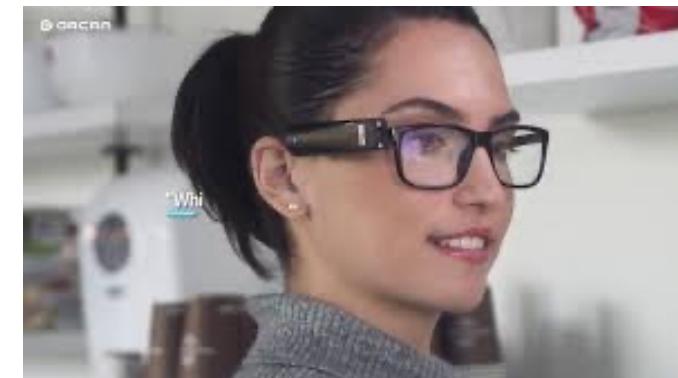
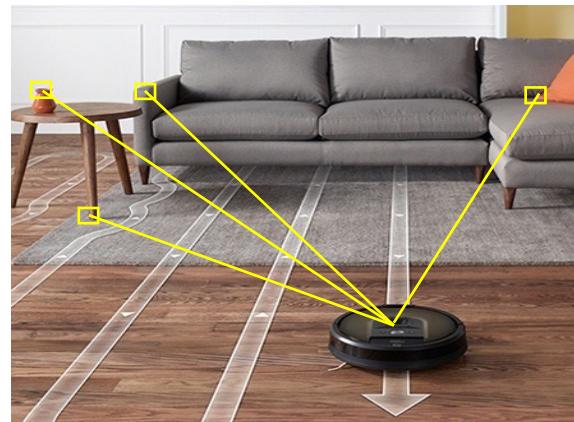
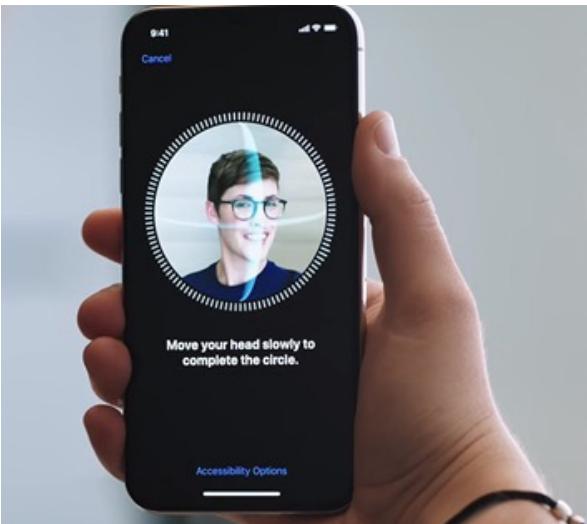
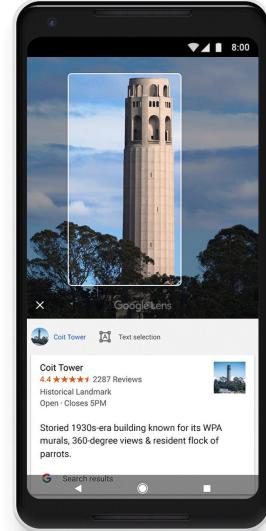
Semantic Segmentation



Traffic Sign Recognition



Mass-Market Consumer Products



New -generative- tasks

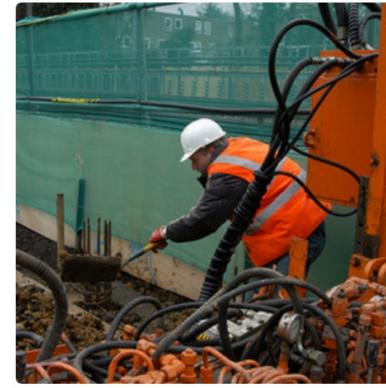
Image Synthesis (StyleGan3, NVIDIA)



Image Captioning



"man in black shirt is playing guitar."



"construction worker in orange safety vest is working on road."

<https://towardsdatascience.com>

Neural Graphics (NERF, 2020)



Text-to-Image (Google, May 2022)

«A bucket bag made of blue suede. The bag is decorated with intricate golden paisley patterns. The handle of the bag is made of rubies and pearls»



Where are we?

Browse State-of-the-Art

12,401 benchmarks 4,682 tasks 117,781 papers with code

Computer Vision



Semantic
Segmentation

281 benchmarks

4906 papers with code

▶ See all 1666 tasks



Image
Classification

455 benchmarks

3594 papers with code



Object
Detection

332 benchmarks

3510 papers with code



Contrastive
Learning

1 benchmark

1977 papers with code



Image
Generation

432 benchmarks

1775 papers with code



Some are duplicates

Also for datasets...



5499 dataset results

Search for datasets

Best match

Filter by Modality

Images	1688
Texts	1529
Videos	542
Audio	256
Medical	199

CIFAR-10
The CIFAR-10 dataset (Canadian Institute for Advanced Research, 10 classes) is a subset of the Tiny Images dataset and consists of 60000 32x32 color images. The images are labelled...
8,207 PAPERS • 54 BENCHMARKS

ImageNet
The ImageNet dataset contains 14,197,122 annotated images according to the WordNet hierarchy. Since 2010 the dataset is used in the ImageNet Large Scale Visual Recognition Chal...
7,760 PAPERS • 80 BENCHMARKS

COCO (Microsoft Common Objects in Context)
The MS COCO (Microsoft Common Objects in Context) dataset is a large-scale object detection, segmentation, key-point detection, and captioning dataset. The dataset consists of...
5,223 PAPERS • 66 BENCHMARKS

What's next?