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 (room 2.8) with a 10 minutes break, If necessary, we will exploit the additional 30 minutes
  - Thursday 11.00 – 12.45 (room 0.5) with a 10 minutes break, If necessary, we will exploit the additional hour (usually when we have a laboratory)

- For this semester, we will keep all office hours online
  - Write to us and we will schedule an appointment

Resources

- The lecture’s slides will be available on Virtuale
- The lecture’s registration will be available on Virtuale
- Several resources are freely available online:
  - Gonzalez, R. C., "Digital Image Processing", Pearson education, 2009.
  - Hartley, R., & Zisserman, A., "Multiple View Geometry in Computer vision". Cambridge university press, 2003.
  - http://d2l.ai/ - Aston Zhang and Zachary C. Lipton and Mu Li and Alexander J. Smola "Dive into Deep Learning", 2020
  - Scientific papers…

Lab sessions

- The laboratory 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

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

Prof. Samuele Salti (Second Part)

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

- Laboratories:
  - Camera Calibration
  - 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 2025
     - 17th of July 2025
     - 8th of September 2025
     - …yet to be defined by the AI course in back office…should be in January or February 2026

IPCV Exam

2. A practical part consisting of two assignments, one for each part of the course (max grade 7)
   - The assignment MUST be completed in groups of 2 (maximum 3) students - you must decide the members of your group yourselves. Assignment completed by one student alone will not be reviewed and will receive 0 points.
   - 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):
     - First deadfline ~half of July 2025 (for scholarship deadline in august)
     - Other three/four deadlines to be defined (will be decided in function of the final exams dates)

   - Double check the submission, it must be in the “submitted” state, not in “draft to submit”

IPCV Exam

- One bonus point will be given if the exam is completed in the academic year of enrolment (only for student of this academic year, i.e., 2024/2025)

- The exam is considered passed if the sum of the two grades is greater than or equal to 18
  - The marks for the written part will be published on AlmaEsami
  - The marks for the assignment we be published on Virtuale

- 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 source code and weights publicly available online.
     - The choice should be based on articles published in the conferences or journals discussed in the next slide
     - We will analyse 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.

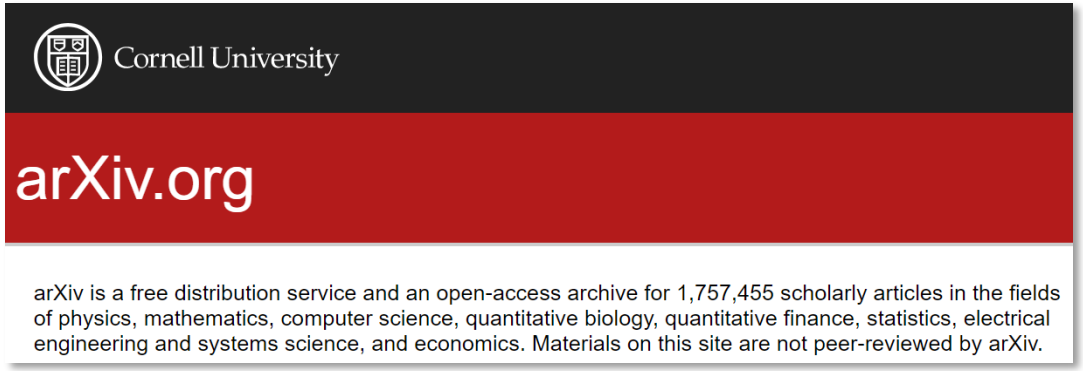
Where to look?

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…

2024

Google Scholar

	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

Computer Vision research is highly influential…

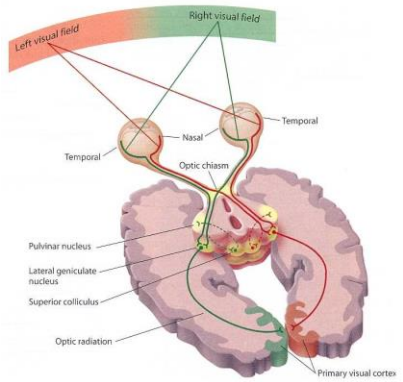
Two days ago…

Google Scholar

	Publication	h5-index	h5-median
1.	Nature	488	745
2.	IEEE/CVF Conference on Computer Vision and Pattern Recognition	440	689
3.	The New England Journal of Medicine	434	897
4.	Science	409	633
5.	Nature Communications	375	492
6.	The Lancet	368	678
7.	Neural Information Processing Systems	337	614
8.	Advanced Materials	327	420
9.	Cell	320	482
10.	International Conference on Learning Representations	304	584

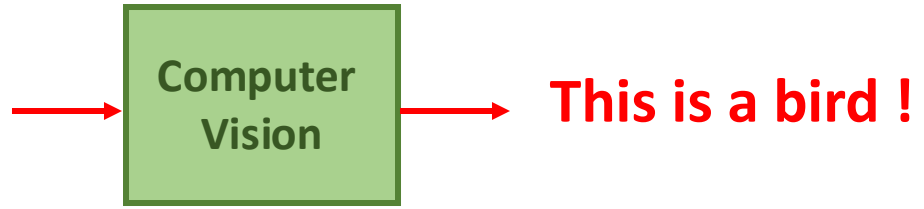
Sponsors @CVPR

What is Computer Vision?

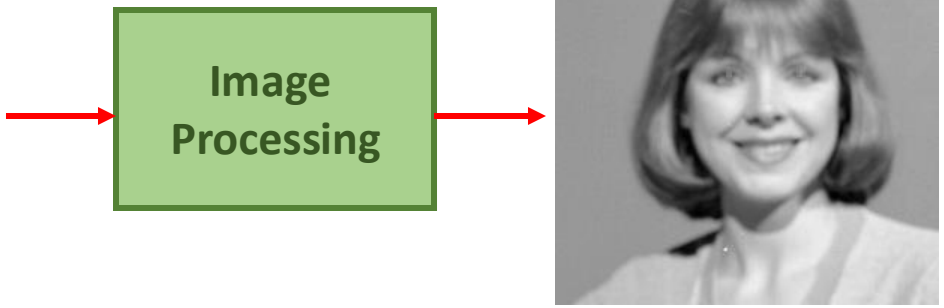


Computer Vision vs. Image Processing

- **Computer Vision** deals with extraction of information from images

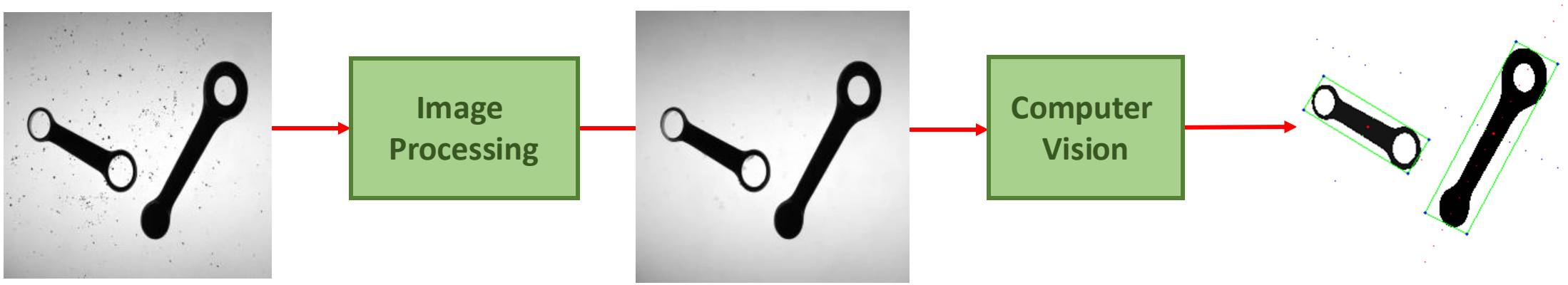


- **Image Processing** aims at improving the quality of images



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)

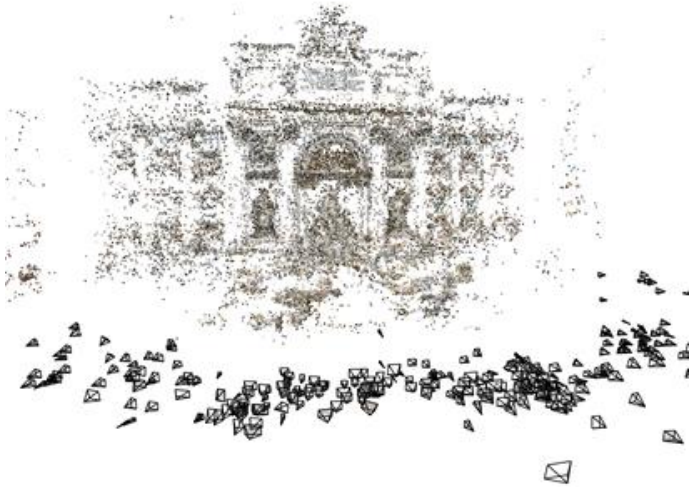
Credit: K. Grauman

What kind of information?

Real-time stereo

Wang et al.

Structure from motion



Snavely et al.

Tracking

Demirdjian et al.

Credit: K. Grauman

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)

Credit: K. Grauman

What kind of information?

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

Credit: K. Grauman

What kind of information?

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

Credit: K. Grauman

What kind of information?

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

Credit: K. Grauman

What kind of information?

- Objects
- Activities
- Scenes
- Locations
- Text / writing
- Faces
- Gestures
- Motions
- Emotions…

Credit: K. Grauman

What kind of information?

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

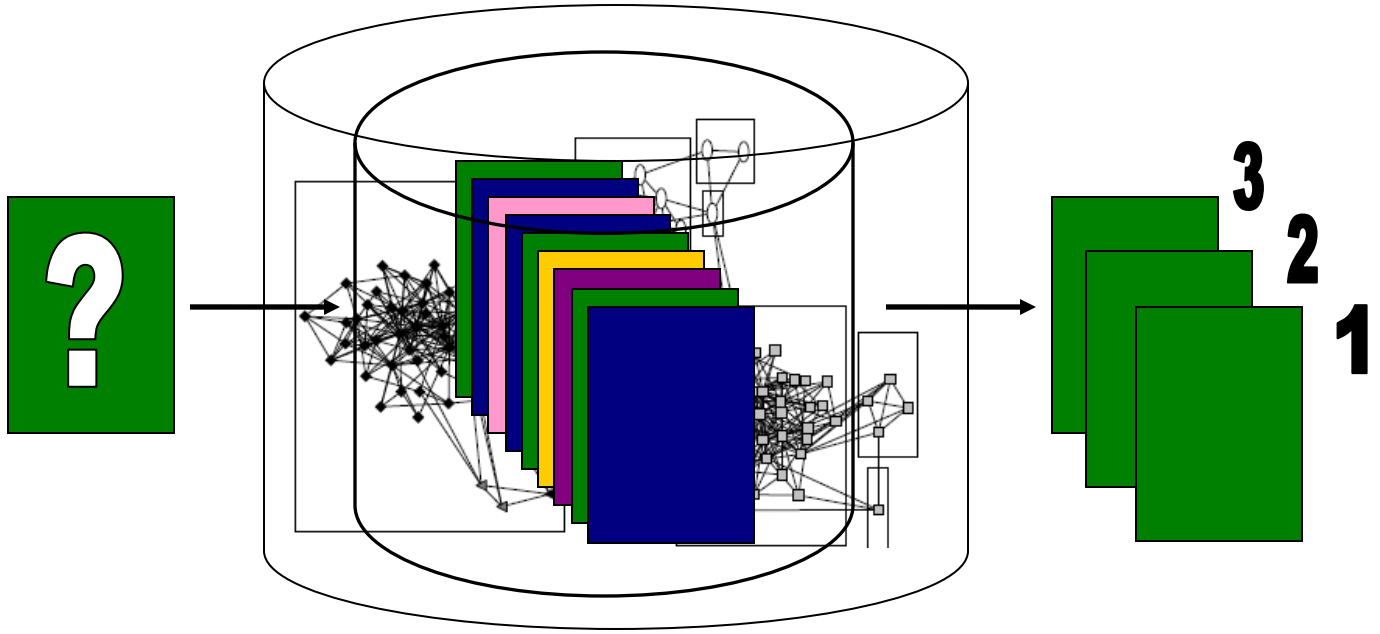
Credit: K. Grauman

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*)

Credit: K. Grauman

What kind of information?



Query

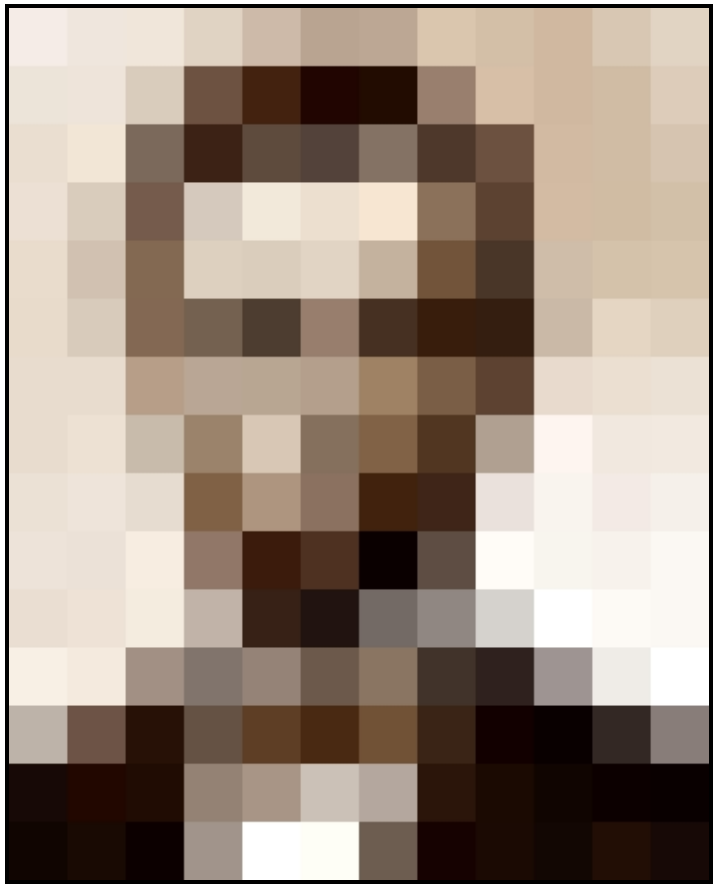
Image or video archives

Relevant content

Credit: K. Grauman

Computer Vision Challenges

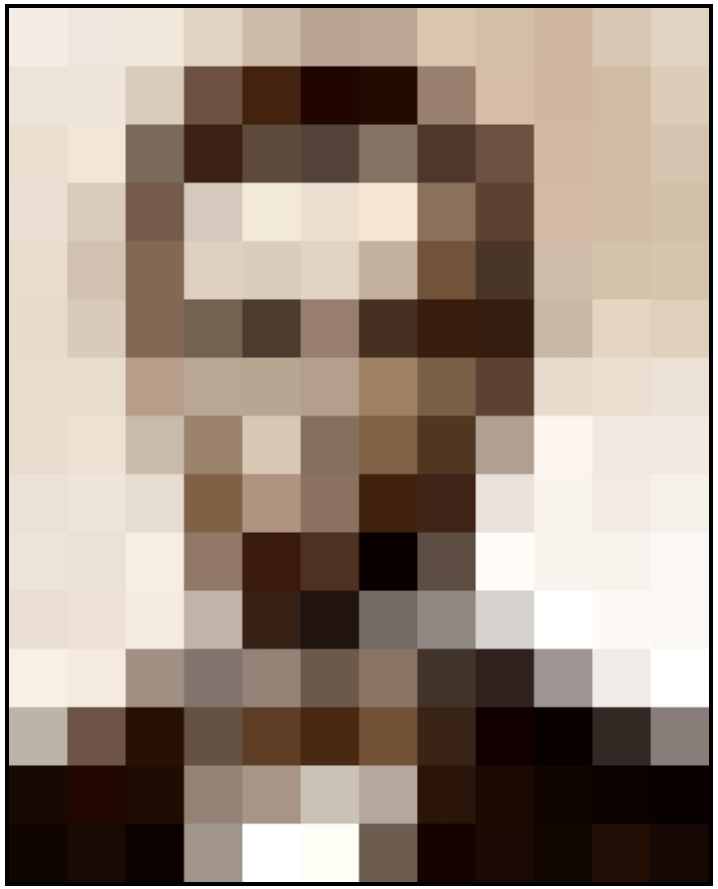
- What the computer sees?



Credit: Larry Zitnick

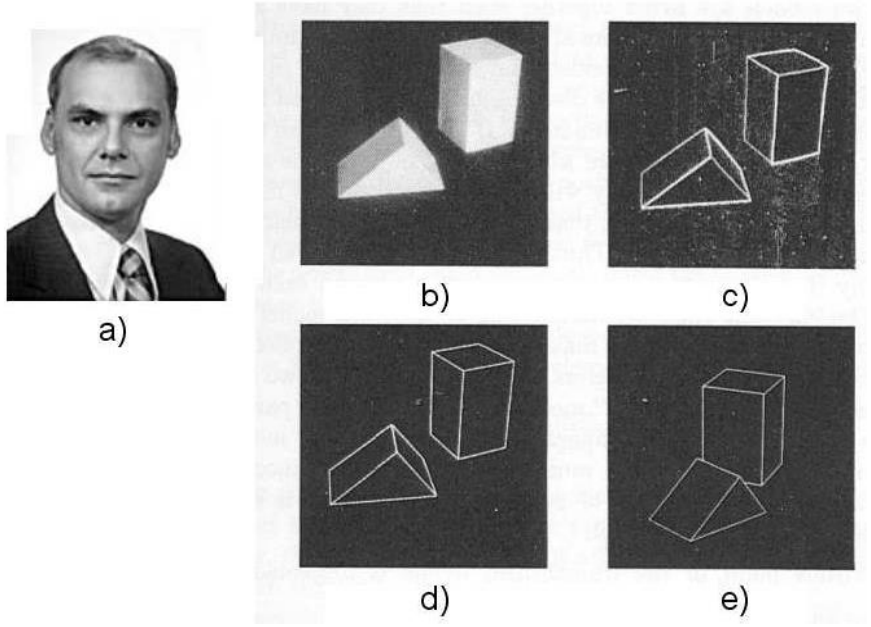
Computer Vision Challenges

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

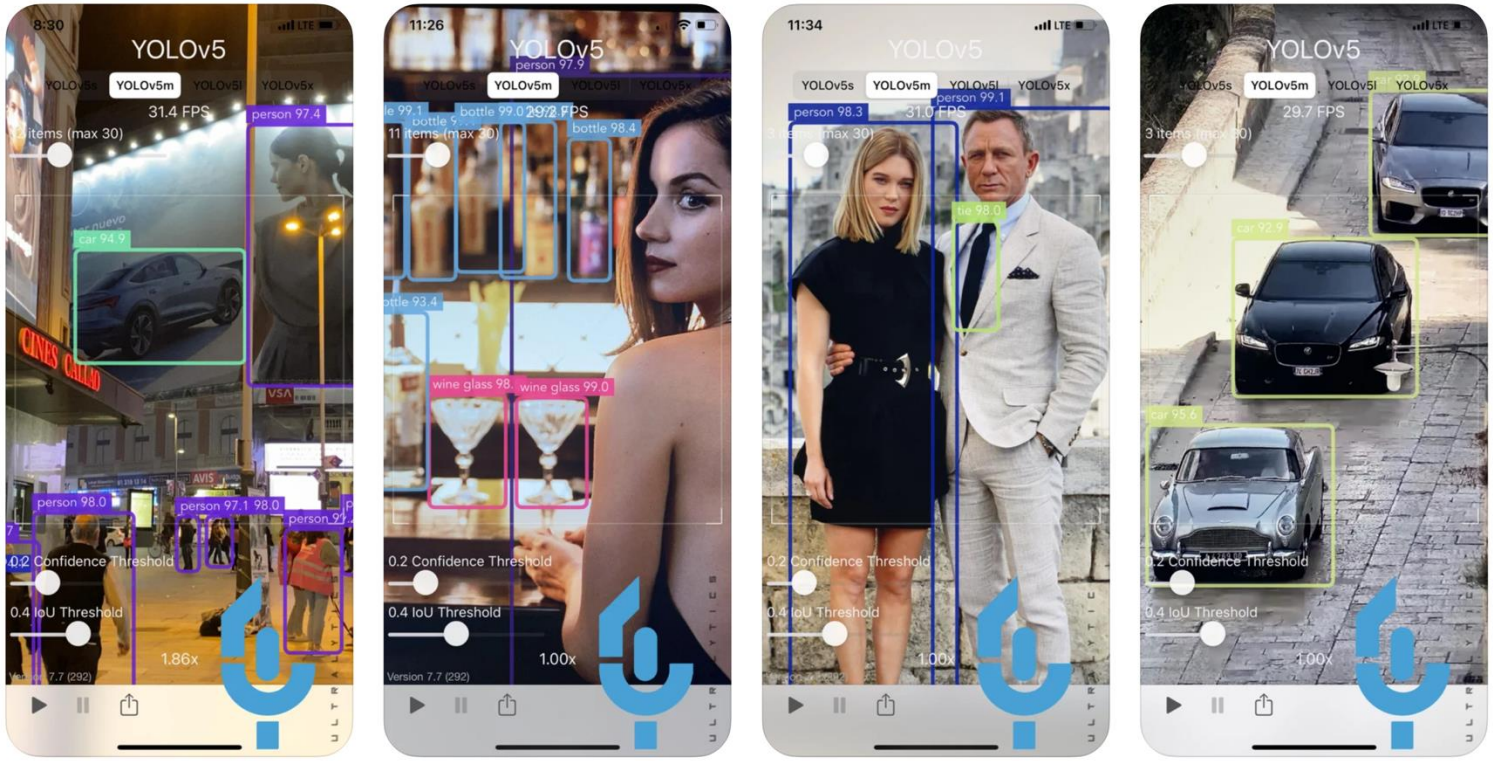


Credit: Larry Zitnick

Quite a successful journey !



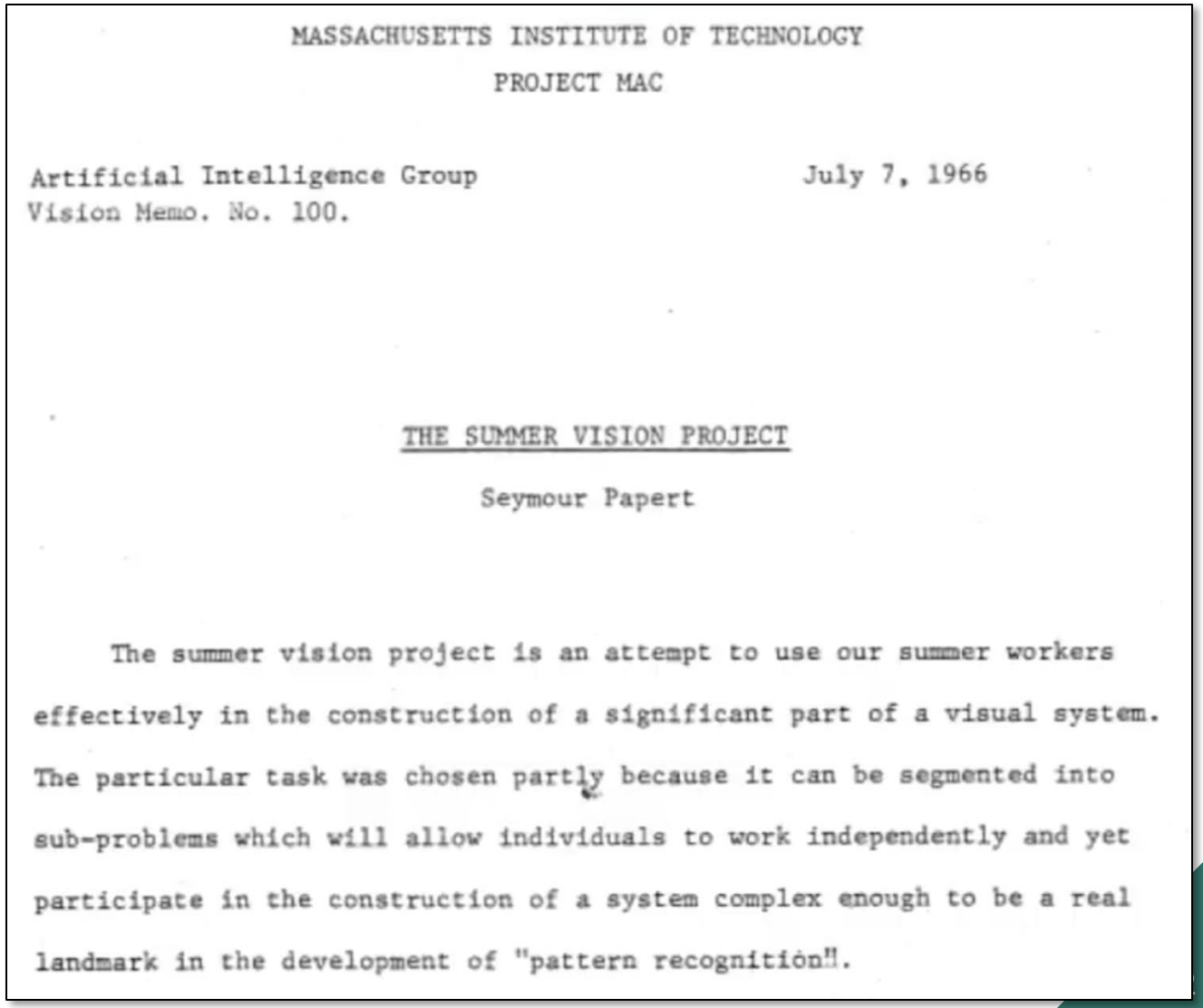
Larry Roberts, PhD thesis, Machine perception of three-dimensional solids, MIT, 1963.



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 :)



MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group — July 7, 1966

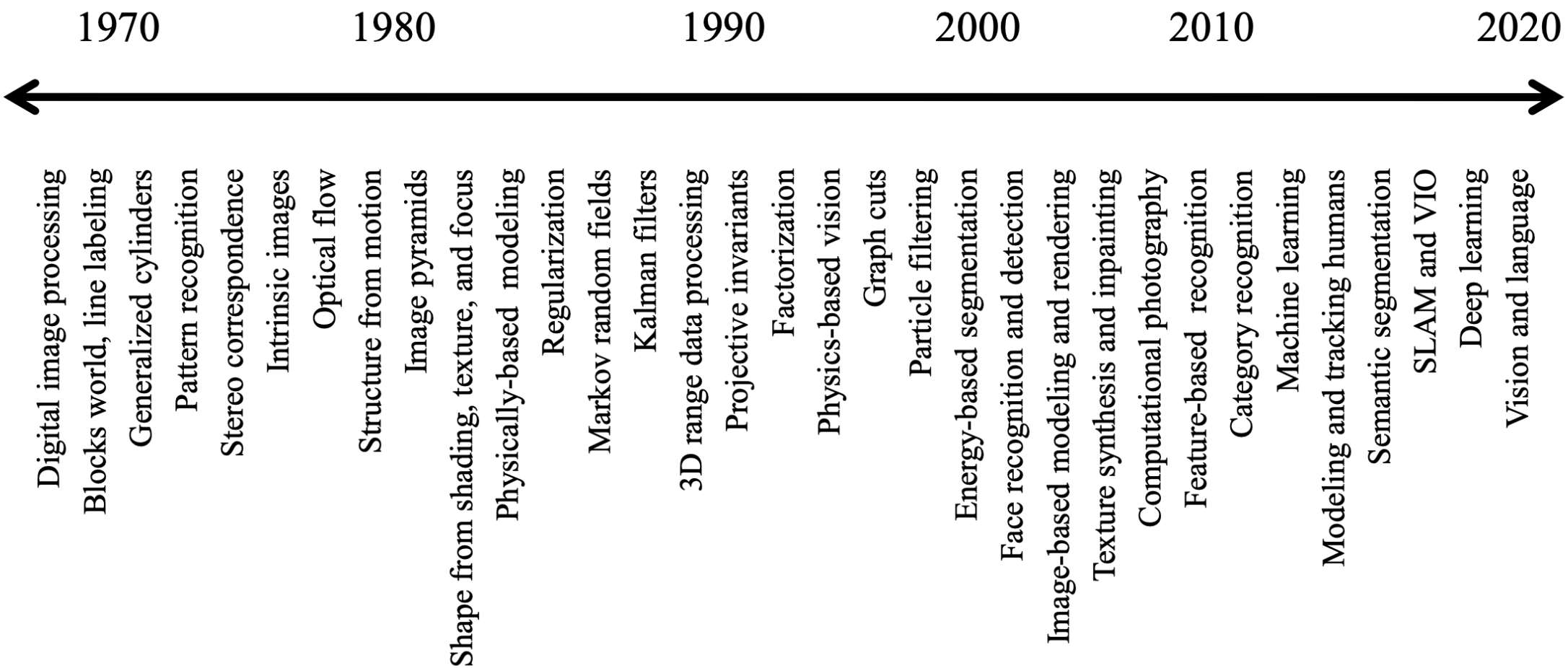
Vision Memo. No. 100.

THE SUMMER VISION PROJECT

Seymour Papert

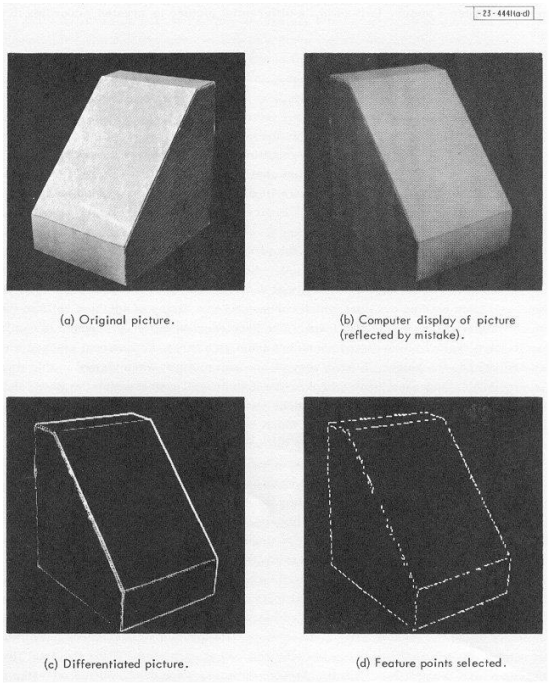
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

Some History…

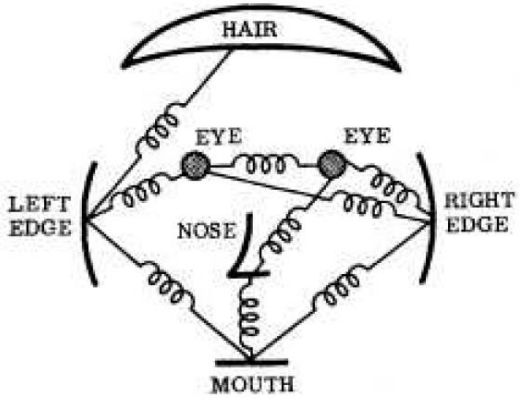


R. Szeliski, "Computer Vision: Algorithms and Applications"

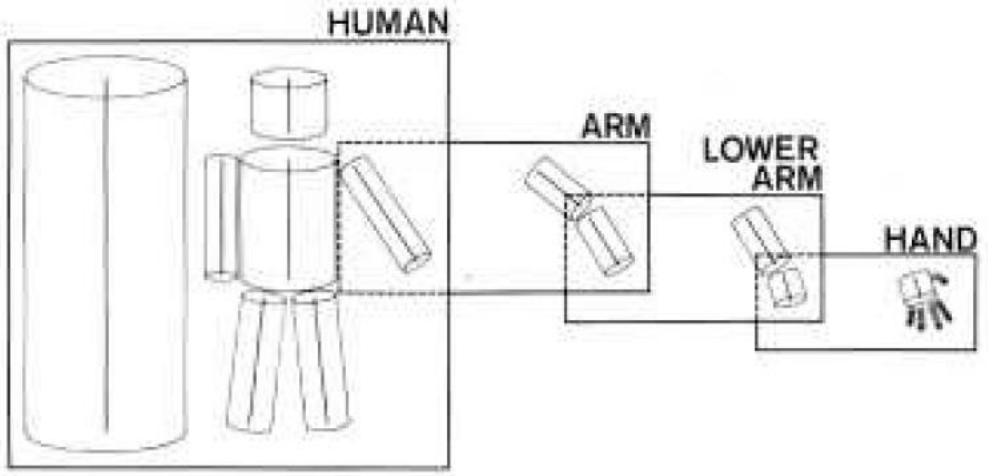
Some History…



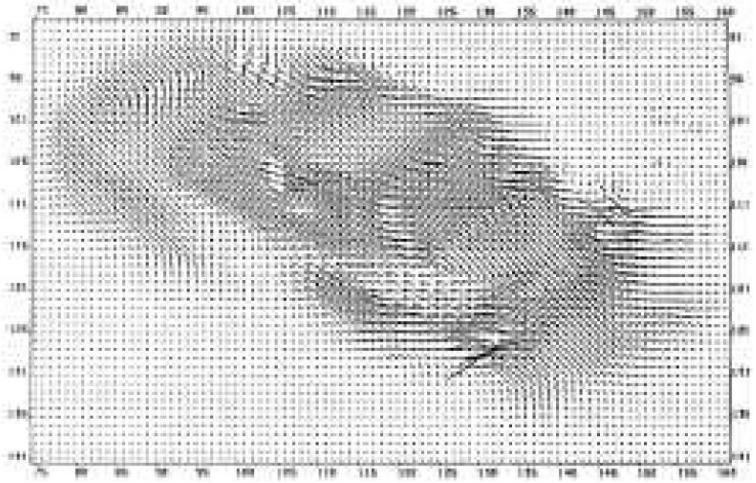
Blocks world
(Roberts 1963)



Pictorial structure
(Fischler and Elschlager 1973)

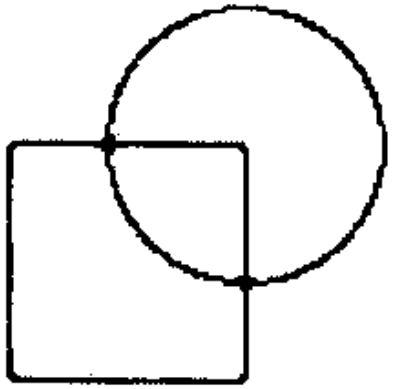


Articulated body model (Marr 1982)



Optical flow (Nagel and Enkelmann 1986)

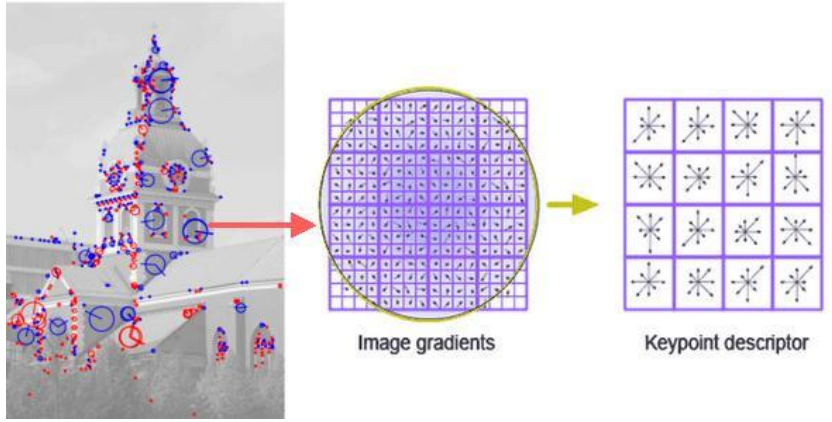
Some History…



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)

R. Szeliski, “Computer Vision: Algorithms and Applications”

Panorama stitching

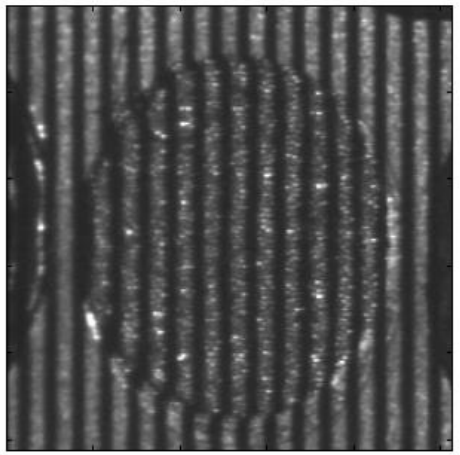
Brown and Lowe, Recognising Panoramas, 2003

Key Process Technology in all Industries

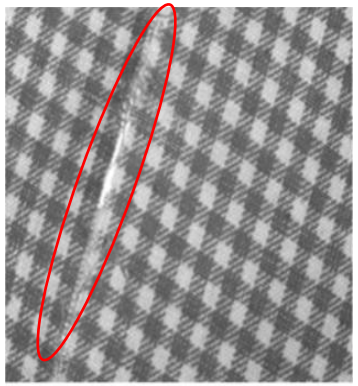
Cognex DataMan (1982)

Pharma & Tobacco

Electronics



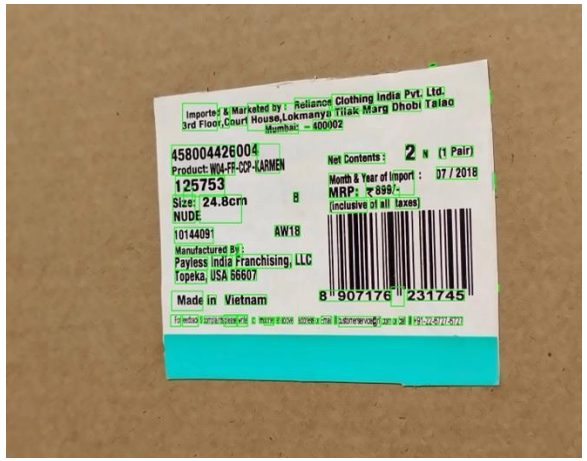
Textile



Food&Bevarage

Automotive

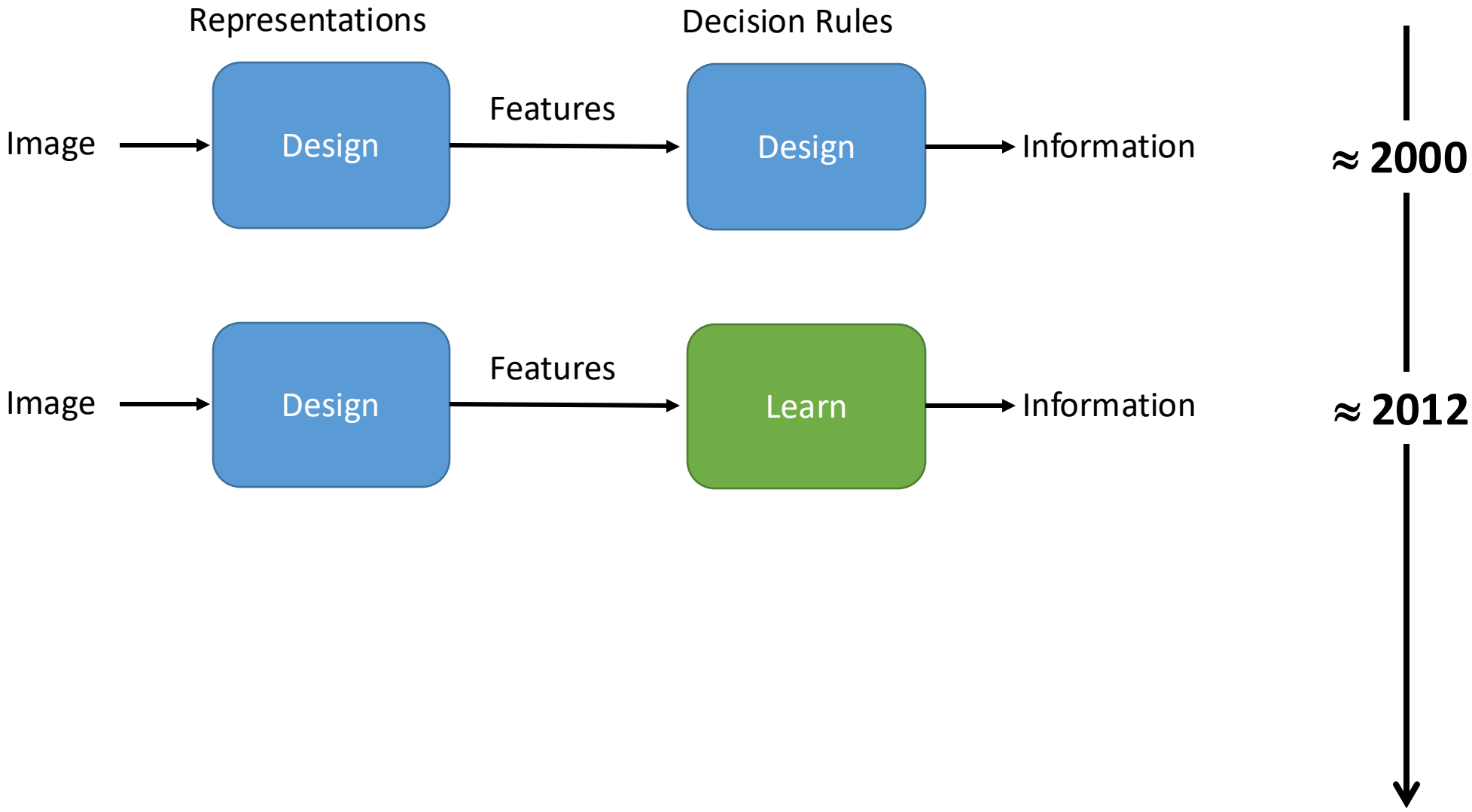
Transportation & Logistics



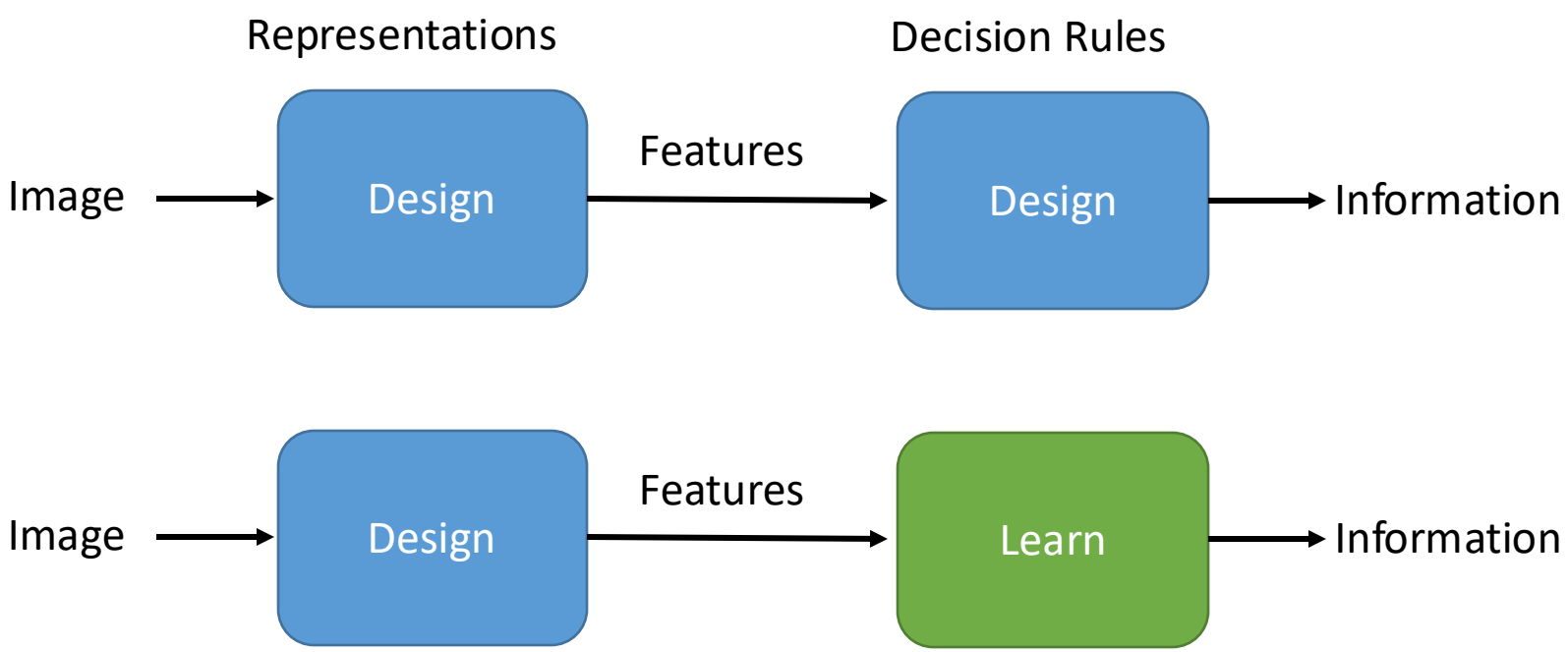
Inspection, Gauging, Guidance, *Tracing*, *Compliance*, *Monitoring*

<u>Process Optimization</u>

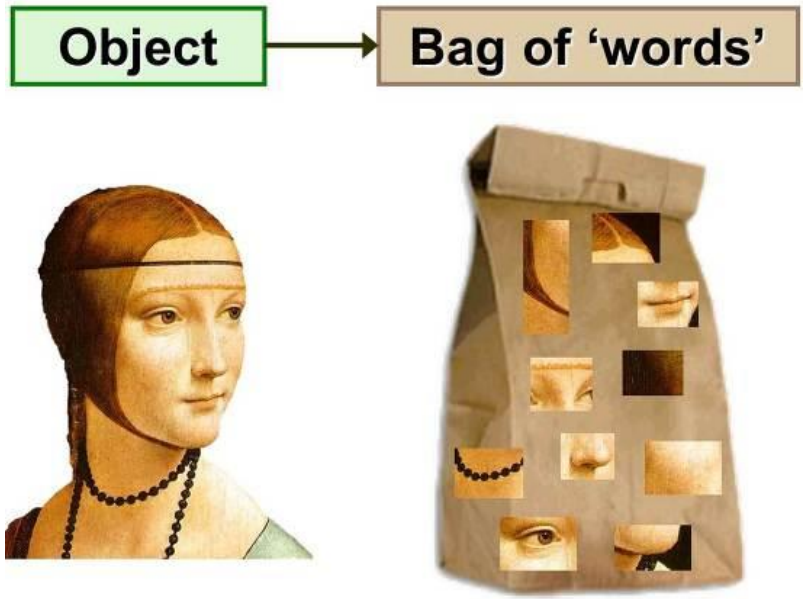
Paradigm Shift



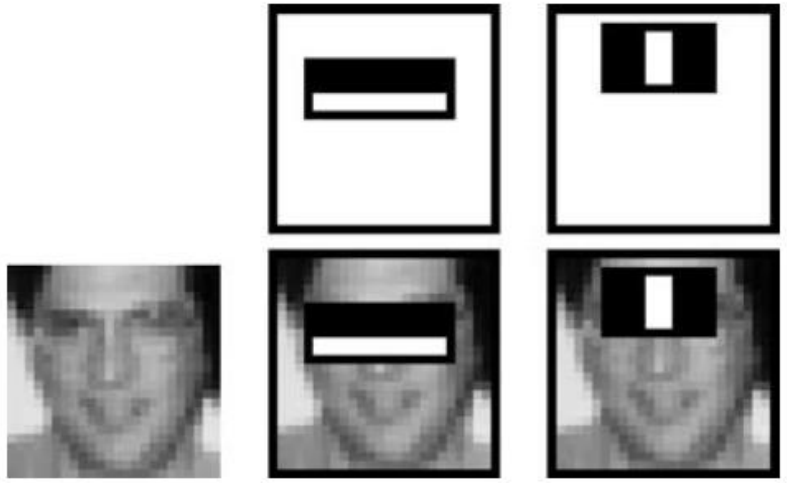
Paradigm Shift (what about data?)



Some History…



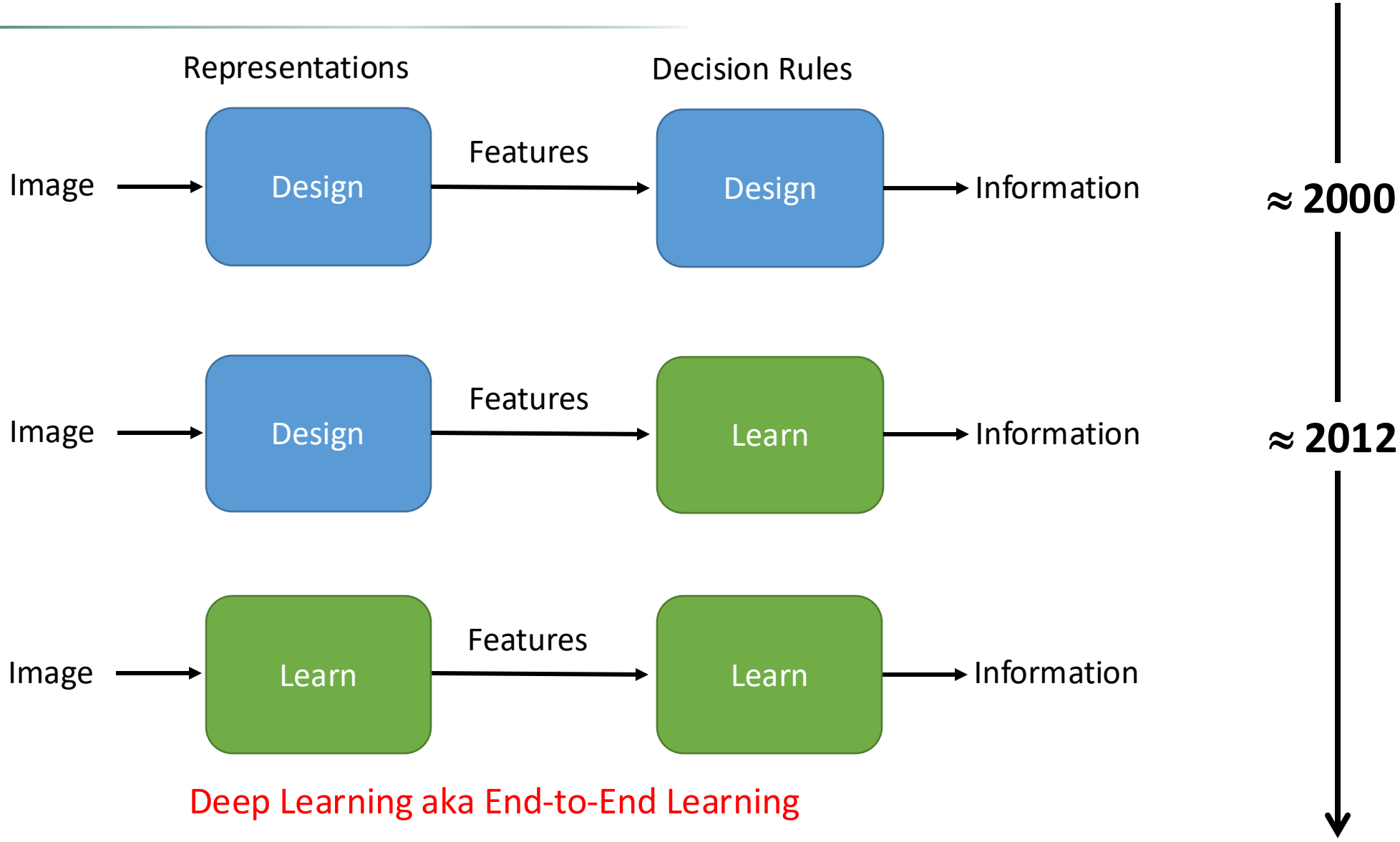
Bag of visual words
(Sivic and Zisserman 2003)



Robust object detection
(Viola and Jones 2003)

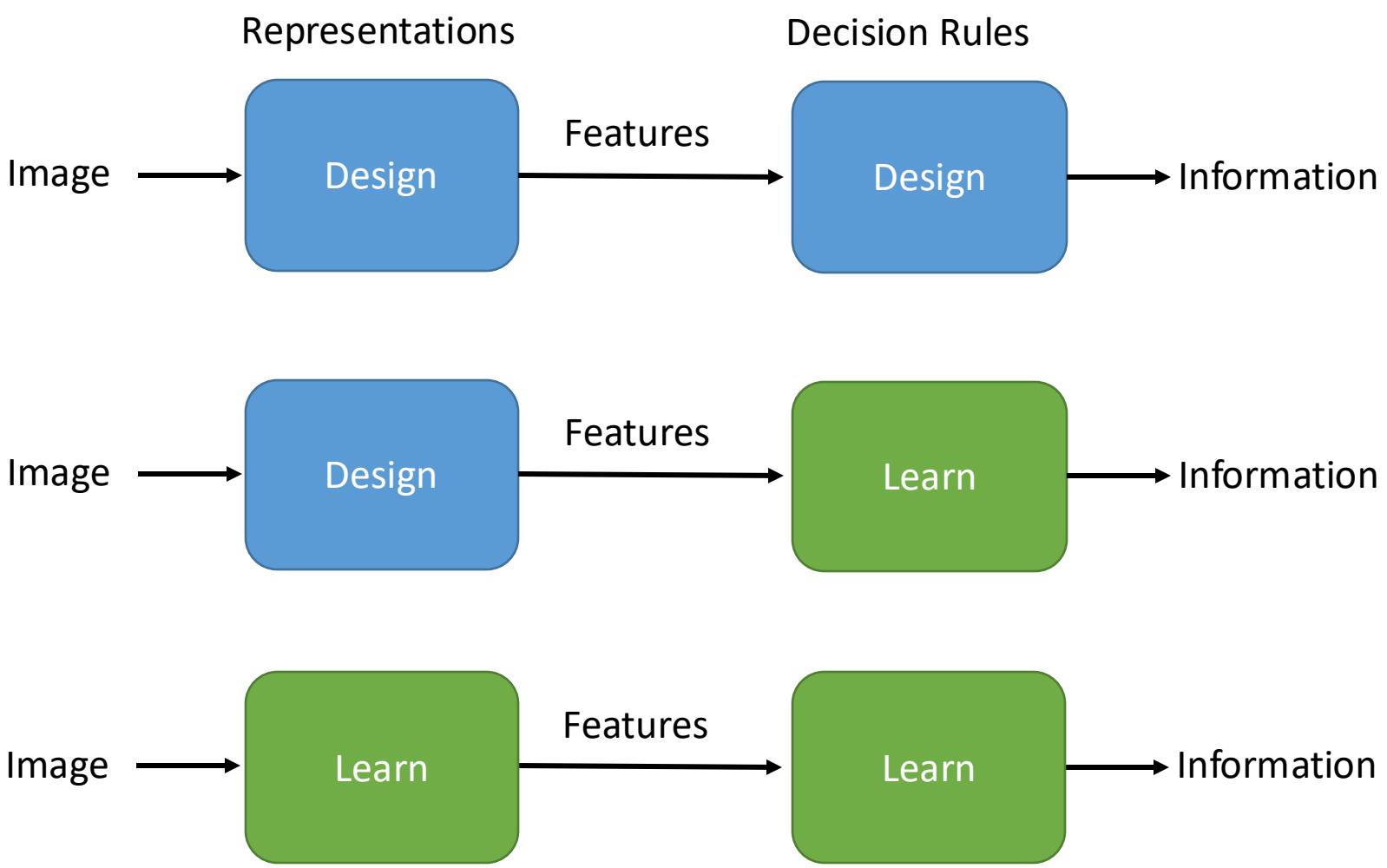
Pedestrian detection
(Dalal and Triggs 2005)

Paradigm Shift

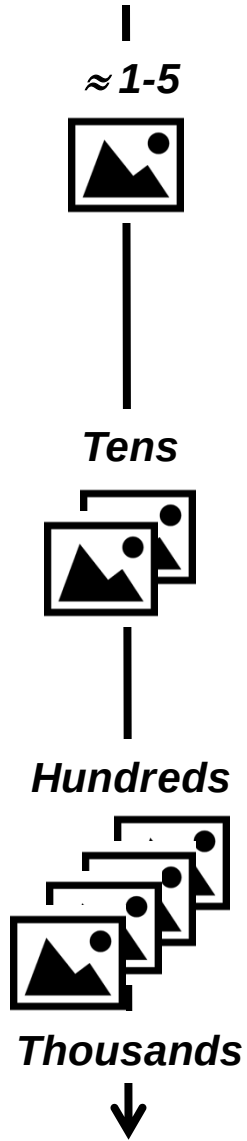


Deep Learning aka End-to-End Learning

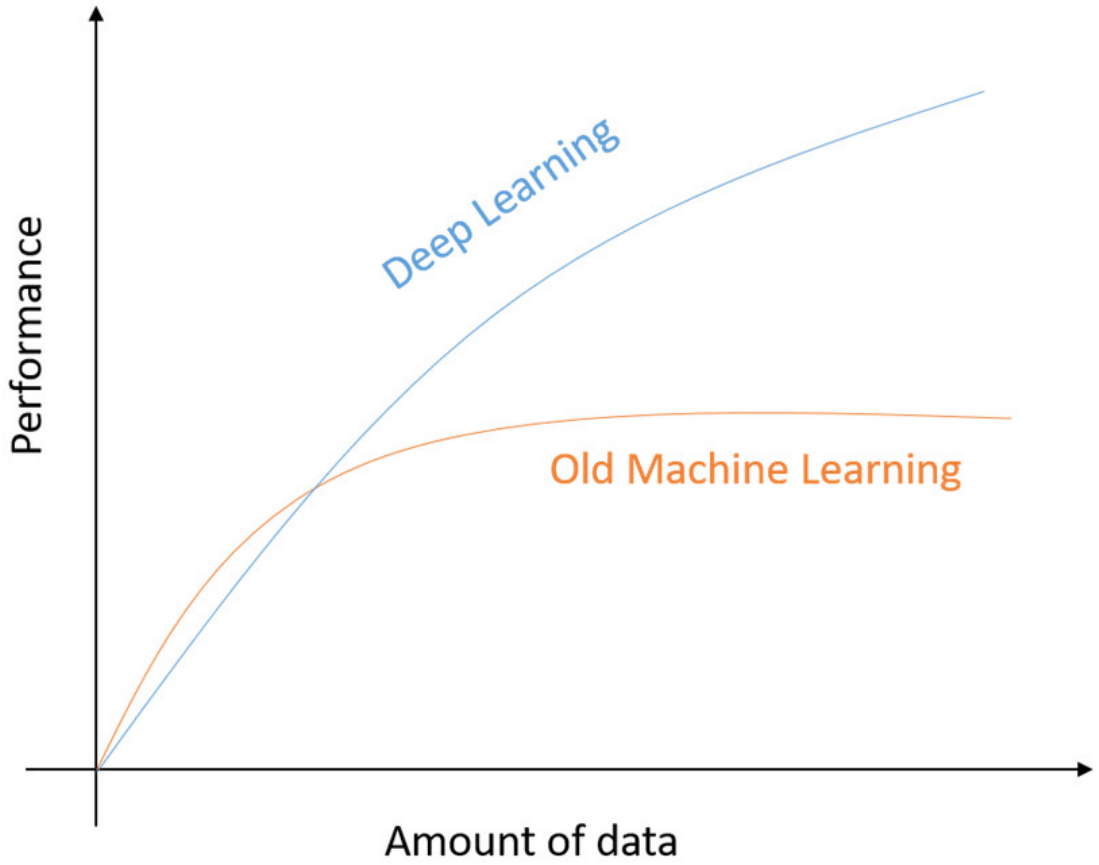
Paradigm Shift (what about data?)



Deep Learning aka End-to-End Learning

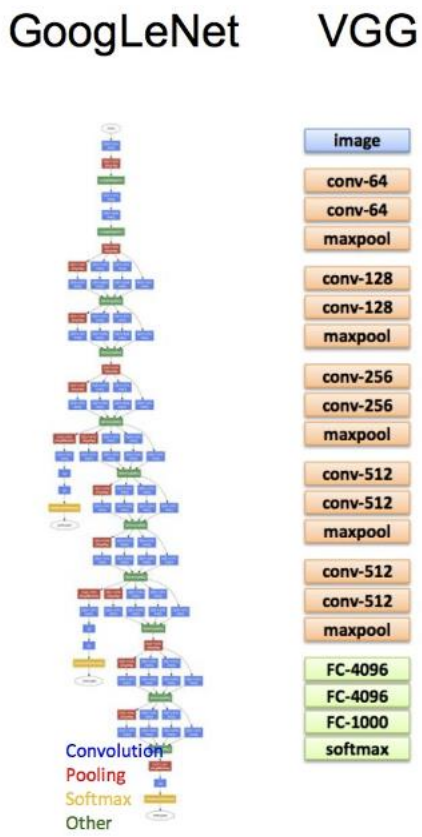


Deep & Machine Learning vs. Data

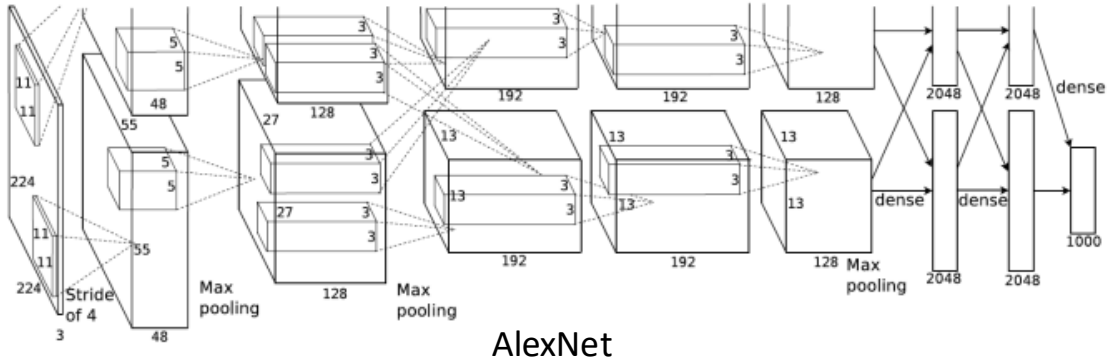


Md. Zahangir Alom et.al. "*State-of-the-Art Survey on Deep Learning Theory and Architectures*", Electronics 2019.

Some History…



[Szegedy arxiv 2014] [Simonyan arxiv 2014]



AlexNet
(A. Krizhevsky et al 2012)

Turing award 2018

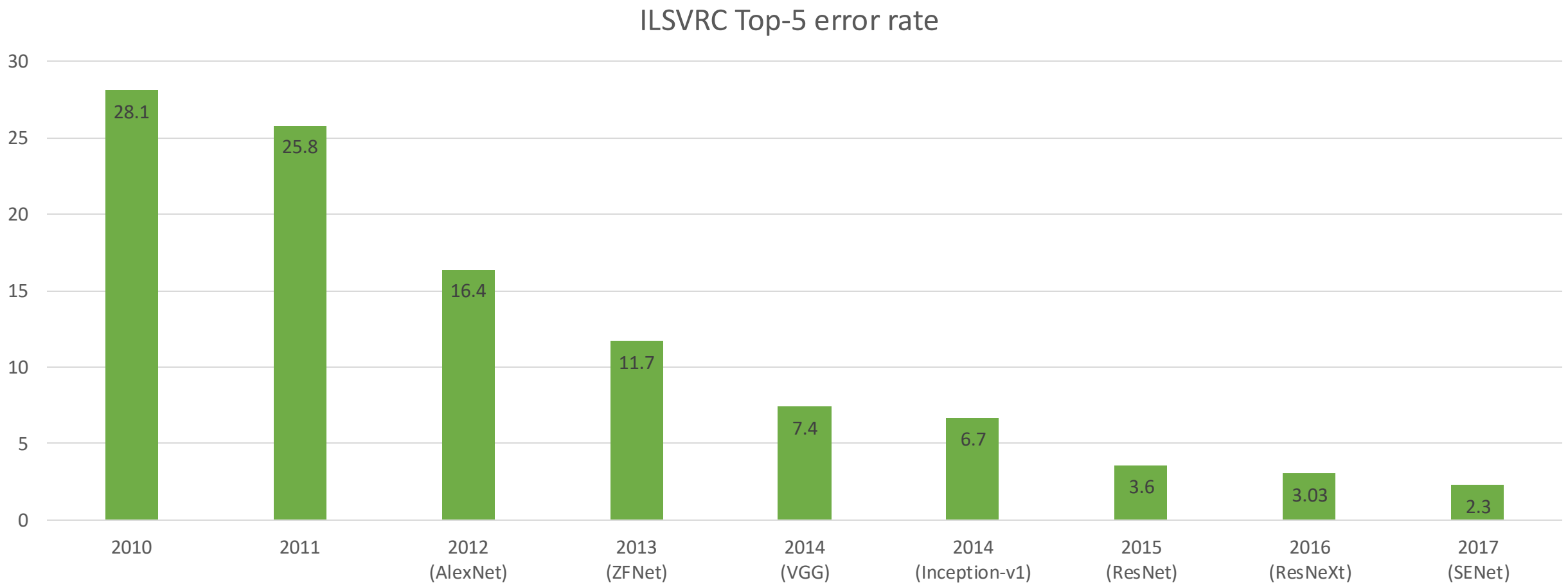


Image-Net and ILSVRC

- ImageNet
  - 22K Categories
  - 14M Images

- Image-Net Large Scale Visual Recognition Challenge
  - 1000 object classes
  - 1431167 images

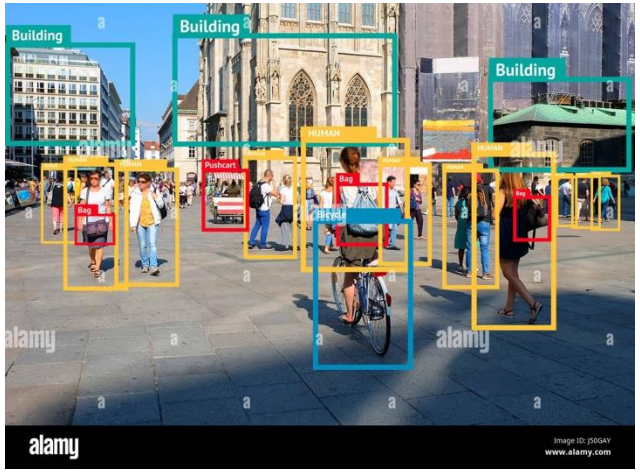
Image-Net and ILSVRC



Deep Learning for Computer Vision

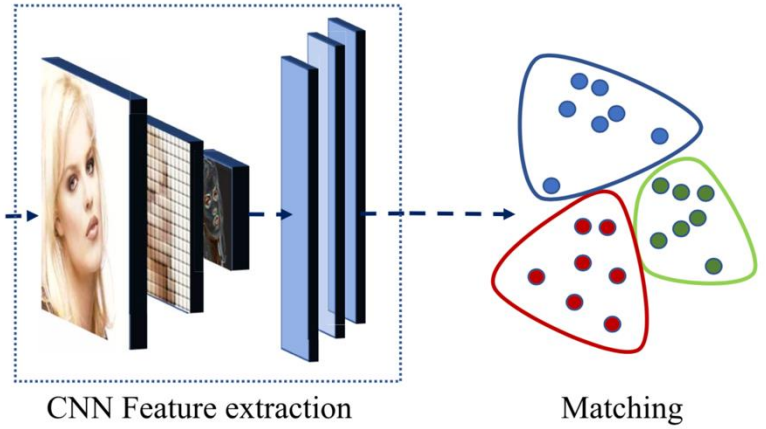
Image Classification

Object Detection and Instance Segmentation



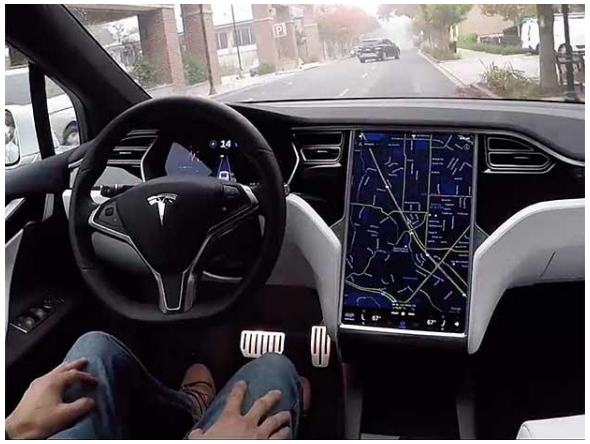
Semantic Segmentation

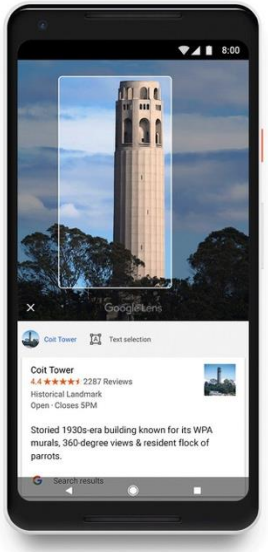
Traffic Sign Recognition

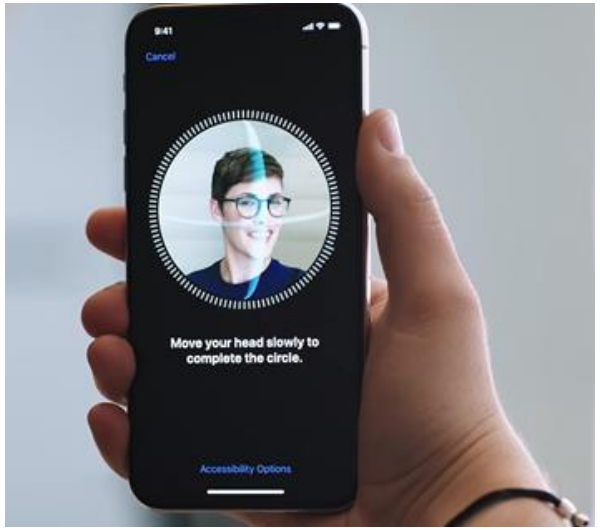


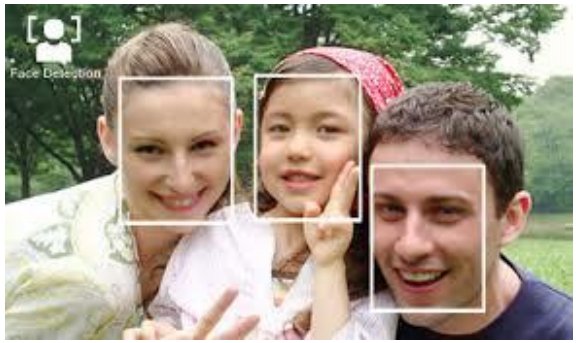
Face Recognition

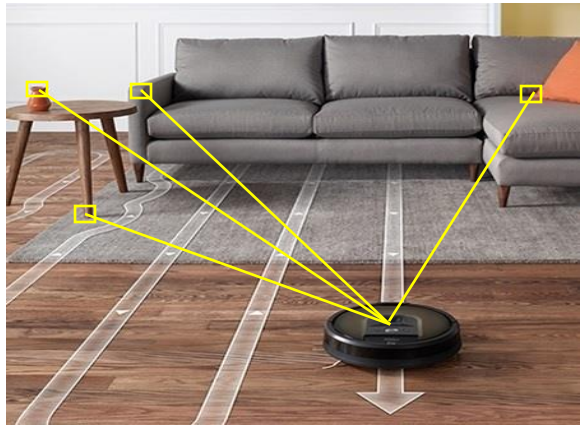
Mass-Market Consumer Products











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»

https://paperswithcode.com/area/computer-vision

Where are we?

Browse State-of-the-Art

12,404 benchmarks 5,363 tasks 155,123 papers with code

Computer Vision

Semantic Segmentation

347 benchmarks

6258 papers with code

Object Detection

398 benchmarks

4451 papers with code

Image Classification

492 benchmarks

4434 papers with code

Representation Learning

16 benchmarks

4396 papers with code

Contrastive Learning

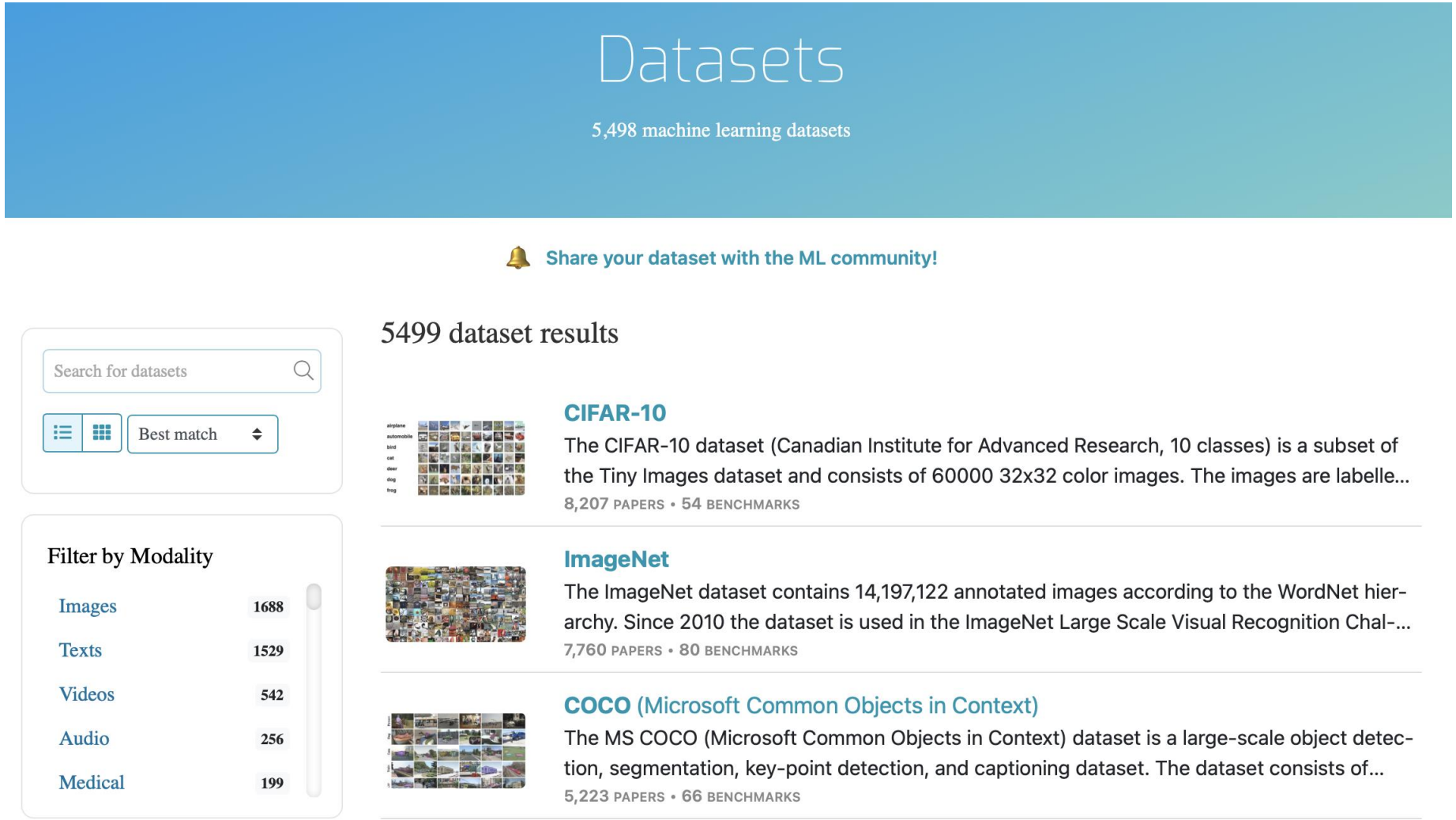
4 benchmarks

2826 papers with code

See all 1934 tasks

Some are duplicates

Also for datasets…



https://paperswithcode.com/datasets

What’s next?